

Smartphone Pressure Collection and Bias Correction Using Machine Learning

CALLIE MCNICHOLAS AND CLIFFORD F. MASS

Department of Atmospheric Sciences, University of Washington, Seattle, Washington

(Manuscript received 23 May 2017, in final form 17 December 2017)

ABSTRACT

Over half a billion smartphones worldwide are now capable of measuring atmospheric pressure, providing a pressure network of unprecedented density and coverage. This paper describes novel approaches for the collection, quality control, and bias correction of such smartphone pressures. An Android app was developed and distributed to several thousand users, serving as a test bed for onboard pressure collection and quality-control strategies. New methods of pressure collection were evaluated, with a focus on reducing and quantifying sources of observation error and uncertainty. Using a machine learning approach, complex relationships between pressure bias and ancillary sensor data were used to predict and correct future pressure biases over a 4-week period from 10 November to 5 December 2016. This approach, in combination with simple quality-control checks, produced an 82% reduction in the average smartphone pressure bias, substantially improving the quality of smartphone pressures and facilitating their use in numerical weather prediction.

1. Introduction

Over the past decade, increasing computational resources have enabled the development of high-resolution convection-allowing numerical weather prediction (Pinto et al. 2015; Seity et al. 2011; Baldauf et al. 2011; Lean et al. 2008). Although, advances in model resolution have fostered more realistic representations of convective systems, they have not produced improvements in forecast location, timing, and intensity (Weisman et al. 2008). Such forecast deficiencies have been at least partly attributed to a lack of spatial and temporal observation density (Mass et al. 2002; Roebber et al. 2002, 2004; Gallus et al. 2005; Sun et al. 2014). Recent studies have demonstrated the value of surface observation density for convection-allowing models (Madaus et al. 2014; Sobash and Stensrud 2015), further motivating efforts to ameliorate this deficiency.

One method of expanding surface observing networks to meet the demands of increasing model resolution is through crowdsourcing, retrieving information from many people, typically through the Internet. Increases in the number of Internet-connected devices with environmental sensors, combined with growth in personal weather station ownership (<http://www.wxqa.com/>), have facilitated an expansion of crowdsourcing efforts in the atmospheric sciences (Muller et al. 2015). Recent

research has examined crowdsourced temperature observations from Nutator personal weather stations (<http://www.netatmo.com>), which have been used to quantify the urban heat island in London, United Kingdom (Chapman et al. 2017), and Berlin, Germany (Meier et al. 2017). Other crowdsourcing studies have focused on smartphones, which are now used by nearly one-third of the world's population (Newzoo 2017). Sensor data from smartphones have been used to estimate temperature distributions (Overeem et al. 2013; Droste et al. 2017), precipitation type and amount (Elmore et al. 2014; De Vos et al. 2017), and surface pressure (Mass and Madaus 2014; Kim et al. 2015; Kim et al. 2016; Hanson and Greybush 2016; Madaus and Mass 2017).

Crowdsourced observations from smartphones offer the potential of extraordinary density and spatial coverage, which could help resolve convective-scale phenomena and enhance high-resolution numerical weather prediction. This potential is demonstrated in Fig. 1, which displays a 2D histogram of pressure observations, retrieved over a 1-h period, from smartphones (acquired with the Weather Channel app) and the Meteorological Data Assimilation Ingest System (MADIS; Miller et al. 2005). MADIS includes observations from the aviation routine weather report (METAR) network, the Citizen Weather Observer Program (<http://www.wxqa.com/>), and dozens of local mesoscale surface observing networks (mesonets).

Corresponding author: Callie McNicholas, cmcnich@uw.edu

DOI: 10.1175/JTECH-D-17-0096.1

© 2018 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the [AMS Copyright Policy](#) (www.ametsoc.org/PUBSReuseLicenses).

2D Histogram of Observation Locations (0000 UTC - 0100 UTC 15 September, 2016)

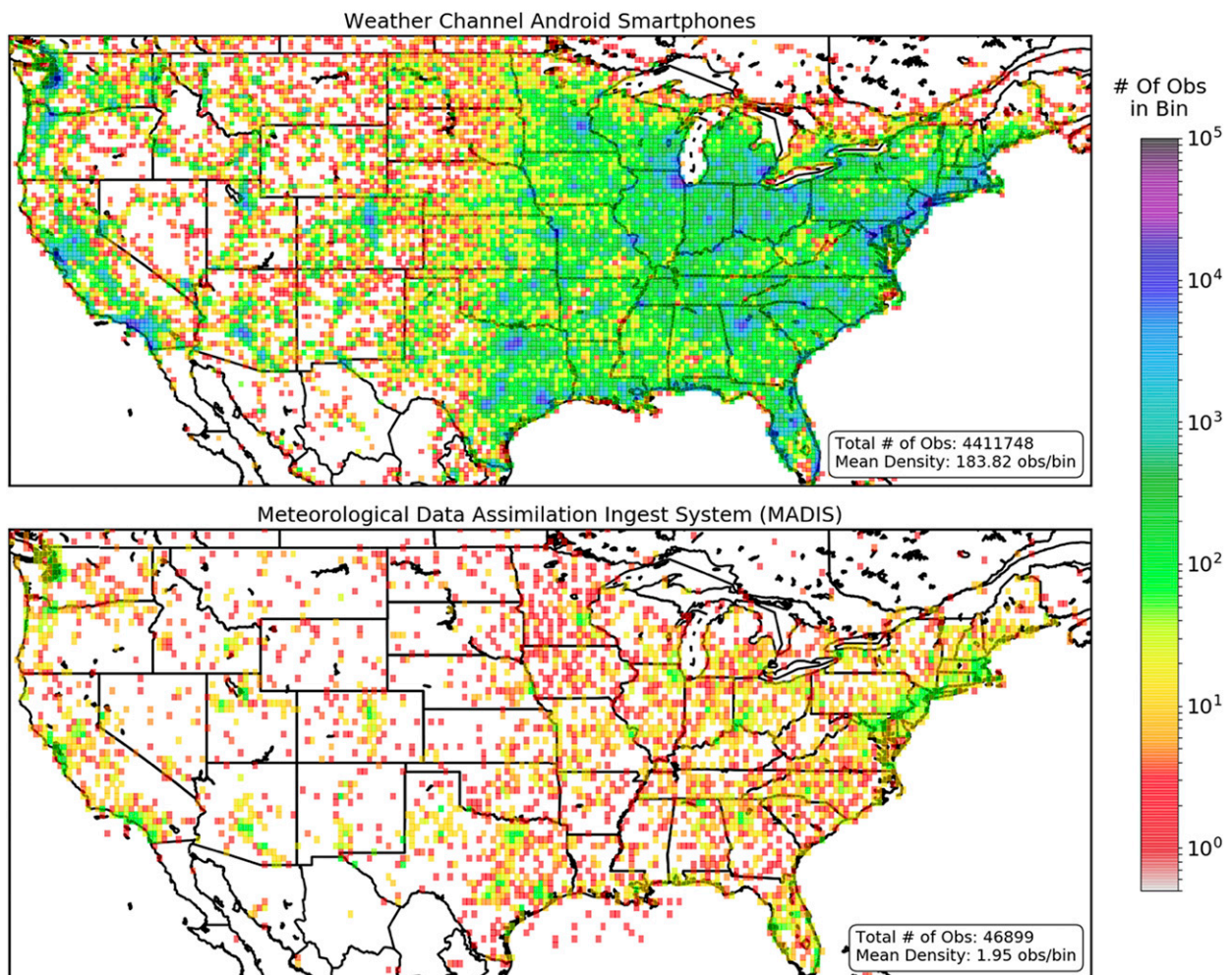


FIG. 1. A 2D histogram of surface pressure observation count between 0000 and 0100 UTC 15 Sep 2016 for (top) Weather Channel Android smartphones and (bottom) MADIS stations. MADIS stations include METAR and Integrated Mesonet stations passing the first three stages of MADIS QC. Observations are binned into square grids with dimensions $0.25^\circ \times 0.25^\circ$ (approximately 20×20 km² at 45° latitude).

The median observation density of smartphone pressures is approximately two orders of magnitude greater than that of the MADIS network. In portions of the northeastern United States, smartphone pressure density exceeded 20 000 observations per 20 km². The density of smartphone pressures in Fig. 1 represents a fraction of the potential total, since the Weather Channel app is not installed on all smartphones and not all smartphones using the Weather Channel app contribute pressure observations each hour. While crowdsourcing pressures from millions of smartphones could vastly improve the density and extent of surface pressure observations, substantial data-quality challenges remain, which is an issue that is explored below.

A major question is whether increased density of pressure observations results in improved analyses and forecasts. The first discussion of the potential of smartphone pressures for numerical weather prediction was provided by [Mass and Madaus \(2014\)](#), which included an example of smartphone pressure assimilation during a convective event in eastern Washington State. [Madaus et al. \(2014\)](#) found a monotonic decrease in domain-averaged analysis error with increasing density of pressure observations from mesonets. They also found that additional observational density improved short-term forecasts of six frontal passage events and one convergence zone event in Washington State. [Hanson and Greybush \(2016\)](#), performing idealized simulations with

synthetic smartphone observations, concluded that if observational uncertainty is well represented, then smartphones pressures can improve model forecasts of surface variables. In the real world, quantifying observational uncertainty for smartphone pressures is complicated by smartphone pressure sensor bias and errors of representativeness.

Madaus and Mass (2017) assimilated smartphone pressures from the PressureNet (<http://www.pressurenet.io/>) and WeatherSignal (<https://www.facebook.com/Weathersignal>) mobile applications during a 72-h convectively active period over the northeastern United States. Validity, statistical, and spatial-consistency quality-control (QC) checks were used to improve data quality, with only one-third of smartphone pressures passing all QC checks. Overall, smartphone pressure assimilation reduced the median 1-h forecast error for surface pressure and 10-m wind by only 0.08 hPa and 0.05 m s^{-1} , respectively. In contrast, the median 2-m temperature 1-h forecast error increased by 0.35 K, with the degradation of temperature forecasts attributed to a lack of observation quality. Nearly half ($\sim 45\%$) of the assimilated smartphone pressures degraded pressure analyses at the location of assimilation, when verification was performed with the assimilated observations.

The study described below builds upon the work of Madaus and Mass (2017) by quantifying and reducing errors in smartphone pressure data. Sources of such errors result from the following:

- 1) Poor collection approaches: improper pressure collection procedures that do not account for sensor internal filtering.
- 2) Inaccurate metadata: for example, inaccurate location information can result in elevation error, especially in regions of complex terrain.
- 3) Sensor bias: systematic sensor errors, resulting from a variety of origins, including soldering issues during sensor installation.
- 4) User behavior: including smartphone speed and locations above/below ground level.

The first goal of this study is to reduce such errors through quality assurance (QA) and bias-correction procedures so that smartphone pressures can be used to describe convective-scale phenomena and enhance convection-allowing numerical weather prediction. An important innovation is the exploration of the potential of machine learning for bias evaluation and correction. A second goal is to evaluate the feasibility of crowdsourcing and quality-controlling pressures from millions of smartphones, since pressure collection from widely used mobile applications such as The Weather Channel app is now a possibility. To achieve each of these goals, a

free Android-based smartphone app (uWx; <http://www.cmetwx.com>) was developed as a test bed for evaluating collection and QA strategies for smartphone pressures.

2. Background

a. Smartphone pressure sensors

Beginning in 2012 smartphone manufacturers began installing pressure sensors to provide support for global positioning system (GPS) location services, enabling a faster and more precise determination of vertical position. These sensors have enabled new mobile applications, such as indoor navigation and calorie consumption calculation (Bosch Sensortec 2012). Over the past five years, the number of smartphone models with pressure sensors has grown substantially (now over 180), many produced by high-end manufacturers, such as Samsung and Apple. Smartphone pressure sensors are microelectromechanical (MEMS) devices composed of a diaphragm formed on a silicon substrate that bends with applied pressure. Under applied pressure the bending of the silicon substrate causes a deformation in the crystal lattice structure of the diaphragm. This deformation initiates a change in the electrical resistance of the diaphragm, which is used to determine pressure variations.

A typical pressure sensor found in many popular smartphones is the Bosch BMP280, which has an absolute accuracy of $\pm 1 \text{ hPa}$ and a relative accuracy of $\pm 0.12 \text{ hPa}$ (Bosch Sensortec 2018). Since pressure sensors like the BMP280 are highly sensitive and susceptible to random errors and noise, oversampling and an internal infinite impulse response (IIR) filter are often used to suppress high-frequency noise caused by wind, doors–windows opening–closing, and other transient effects. Similar filtering techniques are used in other pressure sensors, such as the DPS310 digital pressure sensor (Infineon 2016).

b. uWx—Pressure collection App

To provide a test bed for evaluating a variety of collection and quality-control procedures for smartphone pressures, an Android-based smartphone app, uWx, was developed. This app is free to download and available in the Google Play store. In the first 6 months after its initial release date, uWx collected more than 15 million pressure observations from over 3000 unique smartphones, with the majority in the Pacific Northwest. Approximately 90% of uWx users contribute pressures at least once per hour. The frequency of pressure collection is user adjustable and can vary from 5 to 60 min, with a default determined by the battery capacity of the smartphone. Specifically, the frequency of pressure

collection by uWx is halved when battery life is under 25%. The pressure collection frequency is increased to every 10 (5) min if a smartphone is in a location under a National Weather Service–issued severe weather watch (warning). Although observation frequency is periodically reduced to conserve power, the median period between pressure observations for all uWx users is 22 min. Subhourly pressure collection is accomplished without sacrificing battery life. On a typical Android device, uWx would take approximately 150 h to drain a 3000-mA h battery. As a testament to the efficiency of uWx, the app has maintained a core of approximately 1000 pressure-collecting users since its initial public release.

3. Methodology

Before discussing methodology, two definitions are provided for clarity. In this study, a *pressure measurement* is defined as an instantaneous atmospheric pressure value reported by the smartphone pressure sensor. A *pressure observation* is an average of a collection of pressure measurements retrieved from a smartphone in a single session.

a. Pressure collection and quality assurance

In the PressureNet app and early versions of uWx, location retrieval was performed prior to pressure retrieval. Once a location estimate was retrieved, the first available pressure measurement was saved and uploaded as a pressure observation to the app server. However, retrieving the first pressure measurement fails to account for the internal IIR filtering performed by the pressure sensor. Common smartphone pressure sensors like the BMP280 employ an IIR filter of the following form (Bosch Sensortec 2018):

$$x_f(t) = \frac{[x_f(t-1) \times (k-1) + x(t)]}{k}. \quad (1)$$

Filtered data $x_f(t)$ is derived from a weighted average of previously filtered data $x_f(t-1)$ and current unfiltered data $x(t)$. In Eq. (1) the filter coefficient k is a unitless constant that modulates the weight of the last reported measurement. In the Android operating system, the filter coefficient is set to four. Between retrievals the sensor operates in sleep mode in which pressure measurements are not made. When pressure retrieval begins, the sensor switches to measurement mode and the last measured pressure is used to initialize the IIR filter.

The step response [e.g., a response to change in pressure between two smartphone pressure–altimeter observations (SPOs)] of different BMP280 filter settings

is displayed in Fig. 2. Based on this figure, if the filter coefficient is four and the step (change) in pressure between two SPOs is 5 hPa, retrieving the first pressure measurement would result in an error equal to 75% of the pressure change (i.e., 3.75 hPa). Since meteorological conditions, smartphone elevation, and observation frequency can vary substantially, the magnitude of the pressure change between observations has a wide range. For this reason it is good practice to extend the sensor listening period to ensure a pressure observation is uninfluenced by IIR filtering.

uWx pressure retrieval is performed in the background every 5–60 min, depending on the set frequency of pressure collection. At the start of pressure acquisition, the pressure sensor is called and measurements are recorded for 15–40 s. The first 10 s of pressure retrieval allow the sensor to “spinup” and account for internal filtering. Pressure measurements are recorded during an additional 5–30-s period while location estimate(s) are retrieved. In uWx pressure measurements are reported at a frequency of 20 Hz to balance power consumption and measurement frequency. A pressure observation is computed by averaging the last 50 pressure measurements reported by the sensor. An estimate of sensor noise is retrieved by computing the standard deviation of the last 50 pressure measurements. For additional details on mobile pressure collection, see Aeolus (<https://github.com/cmacc994/aeolus>), a sample Android app that demonstrates basic pressure collection procedures implemented in uWx.

b. Location retrieval

Android smartphones can retrieve location updates from three sources: GPS, Wi-Fi networks, and cellular networks. It is common for apps, like PressureNet, to exclude the GPS from location retrieval, using instead location updates from Wi-Fi and cellular networks. When Wi-Fi networks are inaccessible, location estimates are retrieved from the cellular network, which can lead to large location errors. In the Android operating system, location accuracy is defined by the radius of 68% confidence, with location errors assumed to be random and normally distributed. Location estimates from the cellular network can have location accuracies ranging from a few hundred meters to several kilometers. In contrast, location estimates from the GPS and Wi-Fi networks typically have location accuracies less than 60 m. Reducing location error is important, since position errors result in ground elevation errors. To reduce horizontal location errors, uWx mandates the use of the GPS. The GPS receiver is called and set to operate in high-accuracy/high-power mode. Network location providers (Wi-Fi and cellular) are also called to assist and hasten location retrieval.

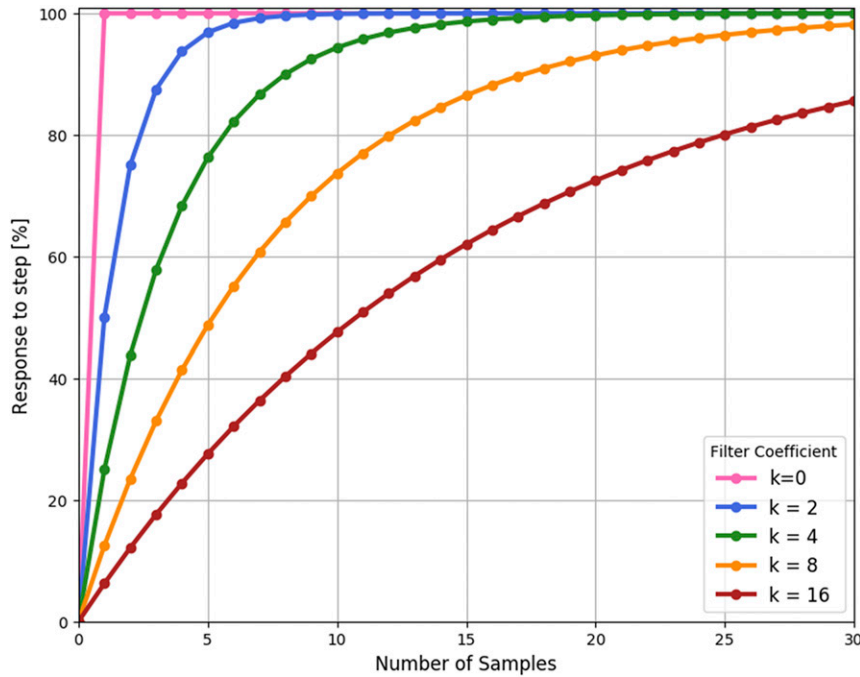


FIG. 2. Step response of a smartphone pressure sensor with an IIR filter and a filter coefficient k , typically hard coded into the sensor device driver.

In uWx ground elevation estimates are found by combining location information (latitude, longitude) with terrain information from a U.S. Geological Survey digital elevation model (DEM), which has a resolution of 30 m and an RMS error of 1.55 m (Gesch et al. 2014). An estimate of elevation uncertainty is computed from the DEM by taking two standard deviations of the nine DEM points closest to the smartphone. The elevation retrieved from the DEM is used to compute altimeter setting (p_{ALT}), which is sea level pressure derived assuming the *U.S. Standard Atmosphere, 1976* (COESA 1976):

$$p_{ALT} = (k_2 z_{sfc} + p_{sfc}^{k_1})^{1/k_1},$$

where $k_1 = \frac{\gamma_s R_d}{g}$ and

$$k_2 = \frac{p_B^{k_1} \times \gamma_s}{T_B}. \tag{2}$$

In Eq. (2), the dry air gas constant and acceleration due to gravity are denoted as R_d ($m^2 s^{-2} K^{-1}$) and g ($m s^{-1}$), respectively. Variables γ_s , p_B , and T_B , represent the lapse rate ($K m^{-1}$), sea level pressure (Pa), and sea level temperature (K) of the *U.S. Standard Atmosphere* (COESA 1976; Duchon 1976). As Eq. (2) demonstrates, elevation errors result in pressure errors, since elevation is used to reduce pressure to sea level. When a smartphone is indoors and most likely to be above or

below ground, the GPS receiver often fails to return an elevation. Signal attenuation indoors can prevent the receiver from achieving a lock on at least four satellites, the number required to retrieve a 3D position fix. Even if an elevation is retrieved from the GPS, the vertical accuracy of GPS is notoriously poor (typically 2–3 times worse than the horizontal accuracy). For this reason, QA procedures in uWx focus on reducing horizontal location errors, since they contribute to errors in pressure and are simpler to correct than vertical location errors.

c. Bias estimation

In uWx bias estimation is performed on a remote app server, using neighborhood altimeter observations from METARs and mesonets in the MADIS network. Nearby MADIS observations are placed into four quadrants spanning the four cardinal directions around the SPO. Interpolation of observations to the smartphone location is performed if at least three quadrants contain two observations within 300 km. A piecewise cubic spline is used to interpolate nearby MADIS observations to the time of the SPO. An inverse distance weighting technique then spatially interpolates nearby MADIS observations to the location of the SPO (Shepard 1968). Cross validation is used to estimate an appropriate power factor for the interpolation, and jackknifing is performed to estimate the uncertainty of the interpolation. In jackknifing each observation is left out during the

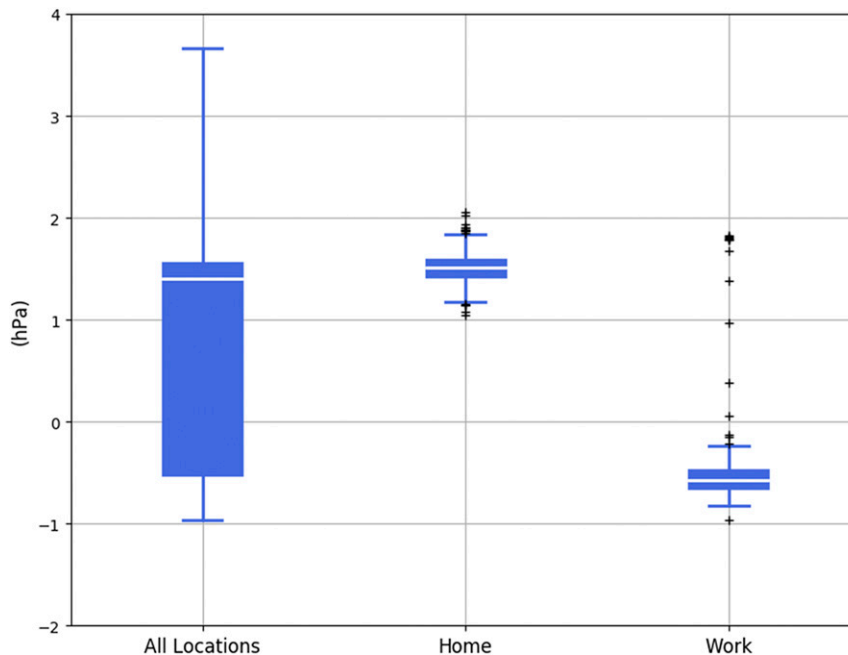


FIG. 3. Distribution of pressure bias for the developer's smartphone for various DBSCAN cluster locations—all locations, home, and work—between 15 Aug and 15 Nov 2016. Median pressure bias is displayed in each box (white bisecting line). A total of 1426 SPOs contributed to this analysis, with 550 SPOs retrieved at the work and 806 SPOs retrieved at home locations.

interpolation, producing a “jackknifed” estimate of the altimeter setting at the smartphone location. A synthetic observation is computed by averaging the jackknifed estimates. The difference between this synthetic observation and the SPO is defined as the pressure bias and is archived for later postprocessing. The uncertainty of the pressure bias is estimated by two standard deviations of the jackknifed estimates. The pressure bias represents the sum of sources of uncertainty, such as sensor bias and elevation errors. To quantify the magnitude and variance of each of these sources of uncertainty, the pressure bias must be decomposed.

d. Clustering analysis

Most smartphones spend considerable time at common locations, such as homes and workplaces, which can serve as de facto observation sites. This fact can be used to gain insights regarding pressure sensor biases. At frequented locations elevation errors are consistent, since users tend to spend significant amounts of time at specific locations (e.g., bedroom, office, etc.) where their elevation above/below ground level is fixed. Evaluating the distribution of pressure bias, at frequented locations, can help reveal the nature of pressure sensor bias, since variance in pressure bias due to elevation errors are minimal. To test this idea, a data mining clustering technique, density-based spatial clustering of applications with noise (DBSCAN;

Ester et al. 1996), was applied. DBSCAN can identify arbitrarily shaped clusters and is robust to outliers.

To test this approach, DBSCAN clustering analysis, performed with 1426 uWx SPOs retrieved from the developer's smartphone between 15 August and 15 November 2016, yielded two clusters corresponding to the developer's home and work locations. SPOs retrieved at the home location were taken from a ground-level apartment, while SPOs retrieved at the work location were taken mostly at a sixth floor office, approximately 16 m above ground level. Figure 3 highlights the distribution of pressure bias for SPOs retrieved at both frequented locations (i.e., home and work) and at all locations. Since SPOs within the home cluster were retrieved at ground level, the magnitude of the sensor bias is well approximated by the median pressure bias at the home cluster (1.51 hPa). When the smartphone was within the work cluster, 16 m above ground level, the pressure bias decreased by 2.08 hPa, from 1.51 to -0.57 hPa, with the 16-m vertical elevation error compensated by the positive (1.51 hPa) bias of the pressure sensor. For home and work clusters, the distribution of pressure bias has less spread than for all locations, with the interquartile range (IQR) of pressure bias decreasing by an order of magnitude, from 2 hPa (all locations) to 0.2 hPa, at home/work. These results suggest that the sensor bias of the developer's smartphone was

relatively unchanged during the 3-month period analyzed. While one smartphone is not representative of all phones, DBSCAN analyses performed on other smartphones (not shown) consistently show a significant decrease in the variance of pressure bias at clustered locations, suggesting that smartphone sensor biases are generally conservative in time.

e. Pressure change retrieval

In previous work by Madaus and Mass (2017), smartphone pressure change observations (SPCOs) were computed during postprocessing without the aid of unique identifiers. Observations collocated in space and separated in time were used to compute pressure change. There was no consideration of location accuracy, local terrain variance, and smartphone motion, since such information was not available. To improve the quality of SPCOs, pressure change estimation in uWx is performed only when a smartphone is not experiencing any substantial motion (e.g., while walking, biking, sitting in a moving vehicle). Substantial motion is detected by a software-based *significant motion sensor*, which utilizes data from an accelerometer to determine small-scale phone motions not resolved by the GPS (Android 2017a). To further supplement the GPS and significant motion sensor, battery information, such as charging status and charging state, are also collected by uWx (Android 2017b). Phones charging more quickly via an ac adapter (e.g., 1.5-A wall charger) are likely stationary, while smartphones charging more slowly through a USB charger (e.g., 0.5-A car charger) may not be. Combining data from the significant motion sensor, battery, and GPS enables a more robust evaluation of smartphone movement. Limiting the retrieval of SPCOs to stationary smartphones helps filter out spurious estimates of pressure change unrelated to atmospheric motions. Tables 1a and 1b outline the requirements for SPCO estimation in uWx. Pressure change is computed from stationary smartphones over fixed intervals of time, typically over 15 min. When elevation uncertainty is small, the requisite location accuracy and the maximum allowable distance between observations is relaxed.

f. Bias correction using random forest machine learning

Predicting and correcting pressure biases, which reflect a complex relationship between sensor values and 3D location, demands a dynamic approach, capable of adapting to user-driven variability. For this reason a machine learning approach to bias correction was tested to determine whether pressure biases could be predicted and corrected using data derived from smartphone sensors and GPS hardware.

TABLE 1. (a) Temporal and (b) spatial prerequisites for pressure change estimation. In (b) Estd refers to elevation uncertainty, which is defined as two standard deviations of the nine USGS DEM grids closest to the smartphone.

TABLE 1a.

| Pressure change (min) | Acceptable time difference between pressure observations (min) |
|-----------------------|--|
| 15 | 10–20 |
| 30 | 20–30 |
| 60 | 50–70 |
| 180 | 165–195 |
| 360 | 345–375 |
| 720 | 705–735 |

TABLE 1b.

| Estd (m) | Max location accuracy (m) | Max acceptable distance (m) |
|----------|---------------------------|-----------------------------|
| 0–1 | 60 | 120 |
| 1–2 | 45 | 90 |
| 2–3 | 30 | 60 |
| >3 | 30 | 30 |

While a variety of machine learning algorithms were initially evaluated for predicting smartphone pressure biases, the random forest algorithm (Breiman 2001) was ultimately selected for pressure bias prediction because of its efficiency, simplicity, and diagnostic capabilities. Random forests utilize a form of ensemble decision tree learning in which learning is achieved by subsetting–splitting input data based on information gain or variance reduction until further splitting is not possible or provides no added value. Random forests overcome the limitations of decision trees, which often lack robustness and suffer from overfitting, by implementing “bagging” or bootstrap aggregation (Breiman 1996) and the random subspace method (Ho 1998). In a random forest, an ensemble of decision trees is created by selecting random samples with replacement from a training dataset. Within each tree, at each candidate split, the randomly sampled variable (feature) that minimizes the mean-squared error of the prediction is chosen as the feature to split. By employing bagging (randomizing observations) and the random subspace method (randomizing features), random forests produce an ensemble of uncorrelated trees that, when averaged, produce predictions with small mean-squared error and low bias.

The ensemble nature of random forests precludes direct interpretation of the tree learning process. Nevertheless, random forests can be used to evaluate the importance of input features. Consider a feature *f*. For each node in a tree that splits on *f*, the variance reduction of the node is weighted by the number of training observations that reached the node. This weighted variance reduction estimate is summed for all nodes in the tree that split on *f*, providing an estimate of

the importance of f for a single tree. This process is repeated for all trees in the ensemble so that an average across the ensemble of trees can be computed, producing an estimate of the importance of f for the entire random forest. In the examination of random forests, feature importance is used to discern which features contribute the most to pressure bias predictions.

In this study random forests were trained on data retrieved between 15 August and 9 November 2016. Since the behavior of each phone is unique, a random forest was generated for each smartphone. To balance performance with computational cost, random forests were initialized with 100 trees. Smartphones with fewer than 50 observations during the 12-week period were not used for bias correction and discarded. Random forests were trained only on SPOs with location accuracies under 60m and absolute pressure biases less than 10hPa. Limiting random forest training to SPOs with modest pressure biases prevented random forests from being trained with SPOs retrieved from aircraft and high-rise buildings. Higher thresholds were evaluated and found to degrade random forest performance.

The input matrix for each random forest included 14 variables collected in real time by uWx (Table 2). These variables are described as features, in accordance with the lexicon of the machine learning community. Once trained, random forest bias predictions were produced for uWx observations retrieved between 10 November and 5 December 2016. The predicted pressure bias was subtracted from uWx SPOs to produce a debiased SPO. MADIS altimeter setting estimates, computed using the interpolation approaches noted above, were used to compute verification pressures at the smartphone locations. The difference between the predicted and true pressure bias is referred to as the *bias prediction error*.

Machine learning with random forests allows pressure biases to be predicted and corrected in real time, using only smartphone sensor and GPS data. One consequence of using random forests is that pressure bias predictions will not exceed the range of pressure bias observed during training, which is at most ± 10 hPa. The magnitude of SPO bias is unknown prior to bias

TABLE 2. Description of features used to train random forests (ancillary features are italicized).

| Acronym | Feature |
|---------------|---|
| phoneid | Pseudorandom unique identifier generated when app is first opened |
| latlon | Concatenation of latitude and absolute value of longitude |
| <i>lux</i> | <i>Illuminance from photodetector (light sensor)</i> |
| <i>elev</i> | <i>Ground elevation from USGS DEM</i> |
| <i>estd</i> | <i>Two standard deviations of the nine nearest USGS DEM grids</i> |
| pres | Average of last 50 pressure measurements |
| alts | Altimeter setting |
| <i>pstd</i> | <i>Standard deviation of last 50 pressure measurements</i> |
| <i>acc</i> | <i>Location accuracy</i> |
| <i>speed</i> | <i>GPS detected speed</i> |
| <i>sat</i> | <i>Number of GPS satellites used in location fix</i> |
| <i>fix</i> | <i>Did the GPS get a location fix? (Boolean)</i> |
| <i>loc</i> | <i>Was the smartphone indoors or outdoors? (Boolean)</i> |
| <i>motion</i> | <i>Was the significant motion sensor triggered? (Boolean)</i> |

correction. Thus, SPOs cannot be filtered prior to bias correction and some SPOs with large biases that exceed the range of bias seen during training will undergo bias correction. Since random forests cannot adequately predict pressure bias in such situations, QC techniques have been developed to remove outliers from bias-corrected observations.

The first stage of QC employs simple validity checks to remove prominent outliers (i.e., altimeter setting < 890 hPa, or > 1100 hPa). The second stage involves a statistical check that removes outliers exceeding four standard deviations from the mean of the observational dataset. Statistical outlier thresholds are modified to adjust for skewness using the split-histogram technique outlined in McNicholas and Turner (2014). The third and final stage of QC is a spatial consistency check that utilizes a radial basis function in the form of a thin plate spline (Duchon 1976). Thin plate splines are ideal for fitting pressure observations, as they produce smooth surfaces and lack free tuning parameters. Once a spline is fit to the observations, outliers are determined by evaluating Eq. (3):

$$\text{SPO} = \begin{cases} \text{Outlier} & \text{if } |\hat{x}_{11} - x_{11}| \geq 4\sigma_G, \\ \text{Inlier} & \text{if } |\hat{x}_{11} - x_{11}| < 4\sigma_G, \end{cases} \quad \text{where} \quad \mathbf{G} = \begin{bmatrix} \hat{x}_{00} & \hat{x}_{10} & \hat{x}_{20} \\ \hat{x}_{01} & \hat{x}_{11} & \hat{x}_{21} \\ \hat{x}_{02} & \hat{x}_{12} & \hat{x}_{22} \end{bmatrix}. \quad (3)$$

If the difference between the spline surface (\hat{x}_{11}) and SPO (x_{11}) exceeds four standard deviations of \mathbf{G} , a

matrix of spline altimeter setting estimates centered on the grid containing the SPO (\hat{x}_{11}), the SPO is rejected.

TABLE 3. Comparison of crowdsourcing pressure apps.

| | uWx | PressureNet | WeatherSignal | Weather Channel |
|---|-------|-------------|---------------|-----------------|
| Phones contributing at least one estimate of pressure change (each hour; %) | 90–95 | 10–15 | — | 10–15 |
| Mean location accuracy (m) | 60.3 | 567 | 471 | 527 |
| Median location accuracy (m) | 21.6 | 36 | 36 | 42 |
| Mean observation frequency (h min) | 50 | 3 31 | — | 3 24 |
| Median observation frequency (h min) | 22 | 3 20 | — | 2 18 |
| Phones submitting pressures on an hourly basis (%) | 91.6 | 11.6 | — | 8.2 |
| Monthly observation count ($\times 10^6$) | 3.5 | 100 | 10 | 2000 |

4. Results

a. Comparison with other pressure collection apps

Previous work by [Madaus and Mass \(2017\)](#) utilized crowdsourced SPOs from the PressureNet and WeatherSignal apps. PressureNet stopped collecting SPOs in late 2015, and in August 2016 the Weather Company began acquiring SPOs using the Weather Channel app. PressureNet, WeatherSignal, and the Weather Channel app do not perform in-app QA techniques (on the smartphone). A comparison between uWx and the three SPO providers is shown in [Table 3](#), which presents statistics derived from hourly data retrieved over a 3-day period from 15 to 18 September 2016. Estimates of pressure change from PressureNet and the Weather Channel app were computed by collecting observations falling within the time windows outlined in [Table 1a](#). For two observations falling within an acceptable time window, an estimate of pressure change was retrieved only if the distance between the observations did not exceed 60 m and the location accuracy of each observation did not exceed 60 m. Since a unique identifier was not provided by WeatherSignal, pressure change and observation frequency could not be computed for that data provider.

Although uWx collects the fewest SPOs, it outperforms PressureNet, WeatherSignal, and the Weather Channel app in all other categories. The mean location accuracy of SPOs collected by uWx is nearly an order of magnitude smaller than the mean location accuracy of SPOs from the other providers. This is a consequence of uWx mandating the use of the GPS during location retrieval. By collecting pressures in the background (even when the device is offline), uWx can collect SPOs at nearly 5 times the frequency of the other SPO providers. As a result over 90% of smartphones in the uWx network submit SPOs every hour. Because of frequent pressure collection, uWx can retrieve SPCOs from over 90% of smartphones in the network at any given time. By comparison only 10%–15% of phones, on average, contribute SPCOs from the PressureNet and Weather Channel apps. While reduced observation frequency

contributes to this result, the poor yield is primarily a consequence of their failure to mandate the use of the GPS when retrieving a location estimate. When the use of the GPS is not mandated, the primary location provider can vary over time. This is problematic, since alternating between accurate (GPS–Wi-Fi) and inaccurate (cellular network) location estimates often results in the appearance of smartphone movement, even when the smartphone is stationary.

b. Pressure change

Since pressure change is not influenced by time-invariant pressure bias, postprocessing of uWx SPCOs was not performed. In-app QA procedures were largely successful in filtering observations unsuitable for pressure change estimation. The success of these procedures is illustrated in [Fig. 4a](#), which displays a sequence of 1-h pressure changes collected from uWx smartphones in real time. In this figure, an isallobaric pressure wave propagates northward from the southern suburbs of Portland, Oregon, to Seattle, Washington, between 0600 and 1000 UTC 24 October 2016. The ability of uWx to capture coherent 1-h pressure change perturbations, with magnitudes as small as 0.3 hPa, demonstrates the effectiveness of uWx pressure change collection. Furthermore, uWx 1-h pressure change observations compare well with observations from MADIS mesonet and METAR analyses ([Fig. 4b](#)).

The need for increased spatial and temporal observation density for resolving convective-scale phenomena and enhancing convection-allowing numerical weather prediction was a major motivation of this work. uWx SPCOs demonstrate the ability of smartphones to capture convective structures. For example, [Fig. 5](#) displays a sequence of 15-min pressure change maps during a convective event near Seattle, Washington. In this event uWx SPCOs captured sudden pressure changes and sharp isallobaric gradients induced by a convectively driven cold pool and wake low. By the end of the period (2326 UTC), the leading and trailing edges of the cold pool were made visible by sharp pressure rises north of the city and weak pressure rises within and

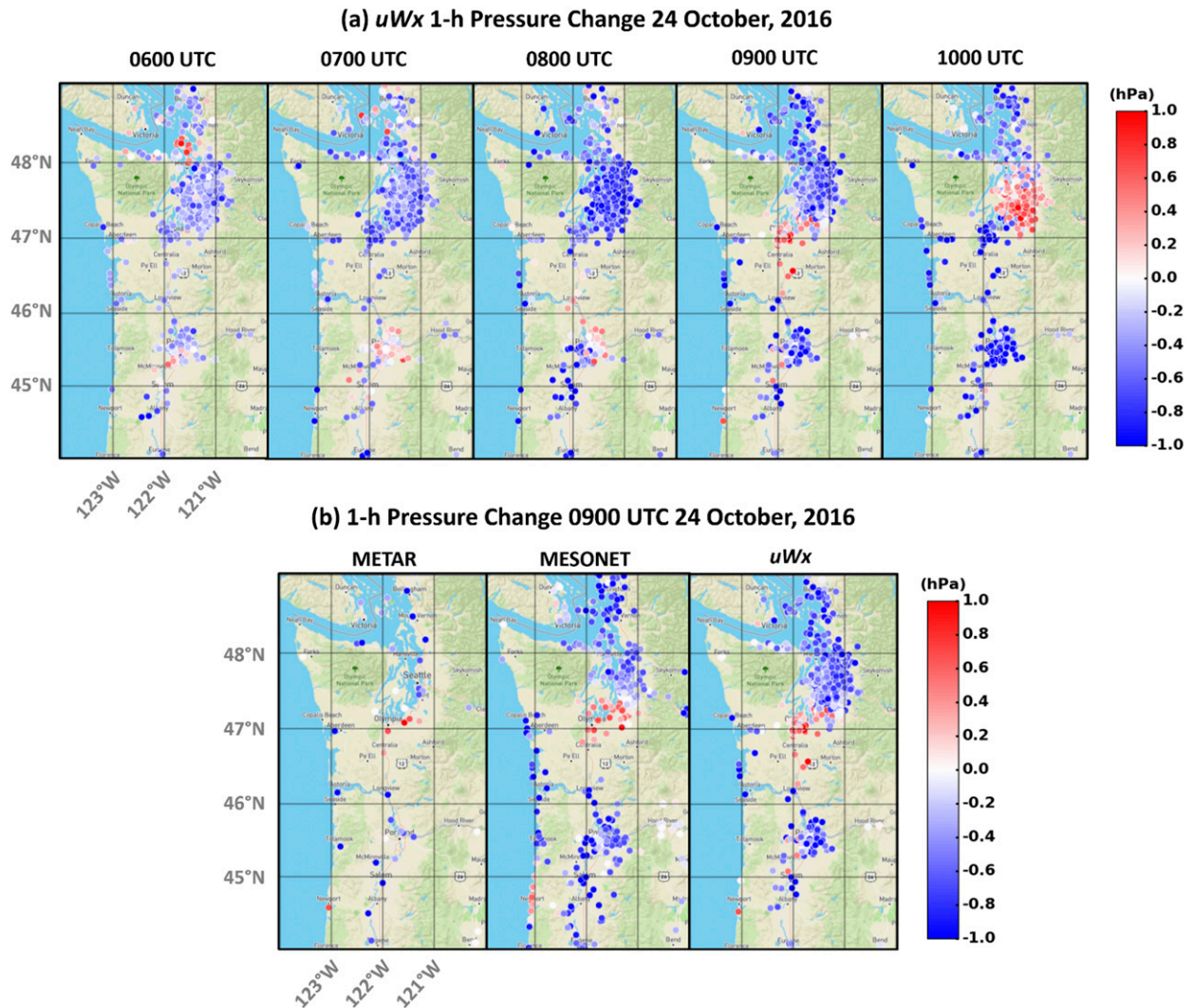


FIG. 4. (a) The 1-h pressure change observations from stationary uWx smartphones for 0600–1000 UTC 24 Oct 2016. Shading indicates pressure rises (red) and falls (blue) over the previous hour. (b) Comparison between 1-h pressure change at 0900 UTC 24 Oct 2016 from stationary uWx smartphones and METAR and Integrated MESONET stations in the MADIS network.

south of the city. The ability of SPCOs to capture the spatial structure and magnitude of mesoscale (Fig. 4) and convective (Fig. 5) phenomena demonstrates their potential as a diagnostic forecasting tool. This is especially true in developing countries, where smartphone penetration is high and meteorological assets, such as surface and radar networks, are sparse or nonexistent.

c. Bias prediction

As outlined above, random forests were trained on uWx data from 15 August to 9 November 2016. Over this period approximately 2325 smartphones contributed SPOs. Random forests were generated for 1978 smartphones that were eligible for bias correction, having retrieved at least 50 SPOs over the 12-week training

period. These random forests were used to predict and correct SPO biases from real-time uWx data, retrieved between 10 November and 5 December 2016. QC checks, also noted above, were applied after bias correction.

Figure 6 displays the distribution of uWx altimeter MAE at different stages of postprocessing. Altimeter MAE was computed, for each smartphone, from uWx SPOs retrieved between 10 November and 5 December 2016. To ensure a sufficient sample size, altimeter MAE was computed only for smartphones that retrieved at least 50 SPOs over the 25-day analysis period. After bias correction the median altimeter MAE decreased by 76% from 1.61 hPa to 0.38 hPa. When QC checks were applied to bias-corrected SPOs, the skewness of

uWx 15-minute Pressure Change, 4 May 2017

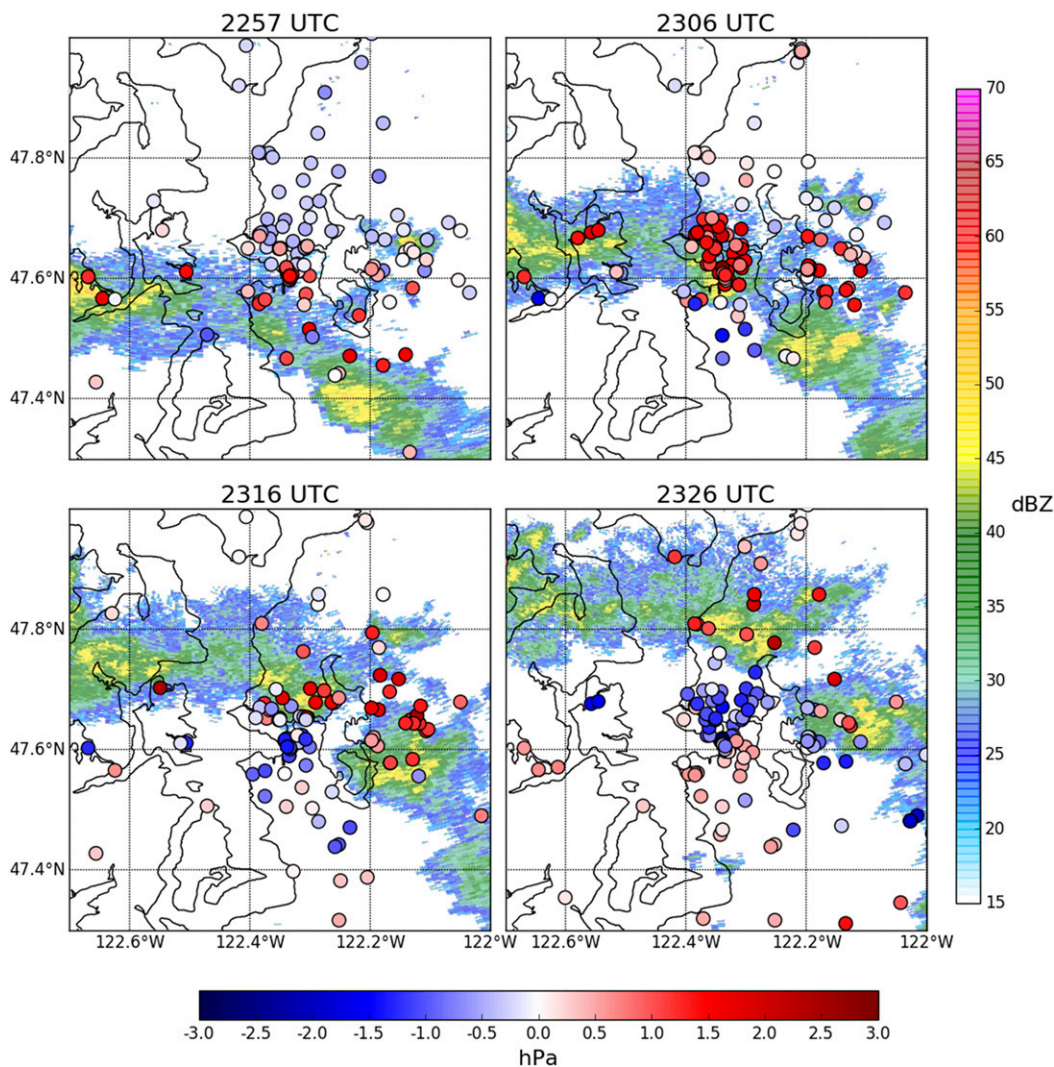


FIG. 5. The 15-min pressure change from stationary uWx smartphones during a convective event over Seattle. All pressure change observations are binned into 10-min windows centered at the time of radar reflectivity retrieval, displayed at the top each subplot. Shading indicates that pressure rises (red) and falls (blue) occurred over the last 15-min.

altimeter MAE was substantially reduced and the median MAE decreased to 0.28 hPa. Overall, bias correction and QC checks reduced the median and mean altimeter MAE by 82% (from 1.61 to 0.28 hPa) and 85% (from 2.23 to 0.33 hPa), respectively.

Bias correction accounted for most of the reduction in altimeter MAE, confirming the ability of random forests to predict and correct pressure biases caused by elevation error and sensor bias. The success of random forest bias predictions can be partly attributed to the fact that many SPOs are retrieved at frequented locations, where elevation errors are consistent and pressure biases are more predictable. The conservative nature of sensor

biases also contributes to the success of random forest bias predictions, which would suffer if the bias of individual smartphone sensors was highly variable.

Figure 7 displays the spatial distribution of uWx altimeter MAE used in Fig. 6. Since smartphones are mobile, altimeter MAE is plotted at the most frequent location of each phone. Prior to bias correction, altimeter MAE is highly variable throughout the domain. Post bias correction, a dramatic reduction in altimeter MAE is observed, especially in the Seattle–Tacoma region QC checks filter out poor-quality SPOs, enabling altimeter MAE to become small and nearly uniform over the Seattle–Tacoma region. In rural regions reductions in altimeter MAE are more

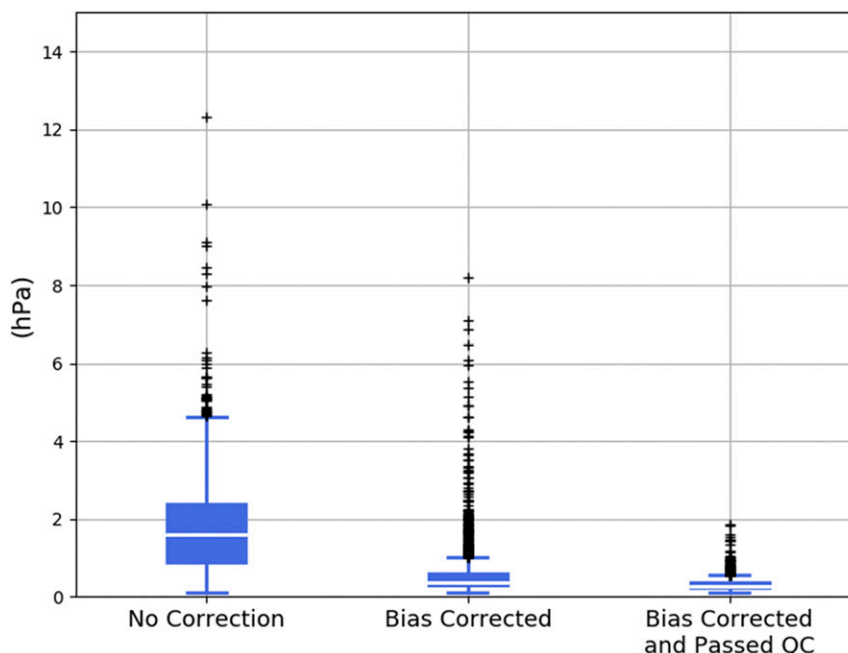


FIG. 6. the distribution of uWx altimeter MAE at different stages of postprocessing: no correction, bias corrected, and bias corrected and passed QC. The altimeter MAE was computed for each smartphone from uWx SPOs retrieved between 10 Nov and 5 Dec 2016. Altimeter MAE was computed only for smartphones that retrieved at least 50 SPOs over the 25-day analysis period. Approximately 1974 smartphones satisfied this criterion.

modest, as QC checks are more lenient in rural regions where observation density is low.

In numerical weather prediction, observations are typically binned in fixed time windows for input into ensemble data assimilation systems. To evaluate the quality of SPOs for assimilation, SPOs retrieved between 10 November and 5 December 2016 were organized into hourly bins. The MAE of SPOs, at different stages of postprocessing, was computed for each hourly bin (Fig. 8a). It is important to clarify that in this figure, unlike the previous two, SPO error is not averaged for each smartphone. Instead the MAE is computed for all SPOs (from many smartphones) retrieved within a given hour. Thus, Fig. 8a essentially displays the domain-averaged SPO error for each hour. In this figure the median hourly altimeter MAE, prior to bias correction, was 1.83 hPa. Bias correction and QC checks reduced the median hourly altimeter MAE to 0.51 and 0.3 hPa, respectively. Over the 25-day period, the examined hourly altimeter MAE of bias-corrected SPOs increased by $0.0045 \text{ hPa day}^{-1}$. Bias-corrected and quality-controlled SPOs' hourly altimeter MAE increased at a slower rate of $0.003 \text{ hPa day}^{-1}$. These findings suggest that random forests should be retrained at least once per month to prevent increases in hourly altimeter MAE from exceeding 0.1 hPa. The positive trend in hourly altimeter

MAE may be a result of a gradual sensor drift. The trend also may be attributed to the retrieval of smartphone pressures at new locations, unseen during training, where the value of ancillary features has no parallel in the training dataset or where the pressure bias falls outside the range of observed pressure biases during training.

In the time series of altimeter MAE (Fig. 8a), some peaks are observed, mainly in the early evening, when the number of SPOs is greatest. These peaks are the result of outliers that skew the distribution of SPO error to the right. Random forests are trained only on SPOs with absolute pressure biases less than 10 hPa. Thus, when an SPO is retrieved in a high-rise building in downtown Seattle, substantial pressure biases, exceeding 10 hPa, can result. Although infrequent, SPOs are retrieved from commercial aircraft and light rail, and thus they also contribute to the peaks observed in the hourly altimeter MAE of bias-corrected and uncorrected SPOs. Although random forests were never trained to predict large pressure biases, SPOs with substantial pressure biases still undergo random forest bias correction, since at the time of bias correction the true pressure bias of each SPO is unknown.

Figure 8b displays the hourly count of uWx SPOs, at different stages of postprocessing. There is a diurnal cycle in the observation count, since fewer SPOs are retrieved overnight. At night smartphones are more

Spatial Distribution of uWx Altimeter MAE (10 November - 5 December, 2016)

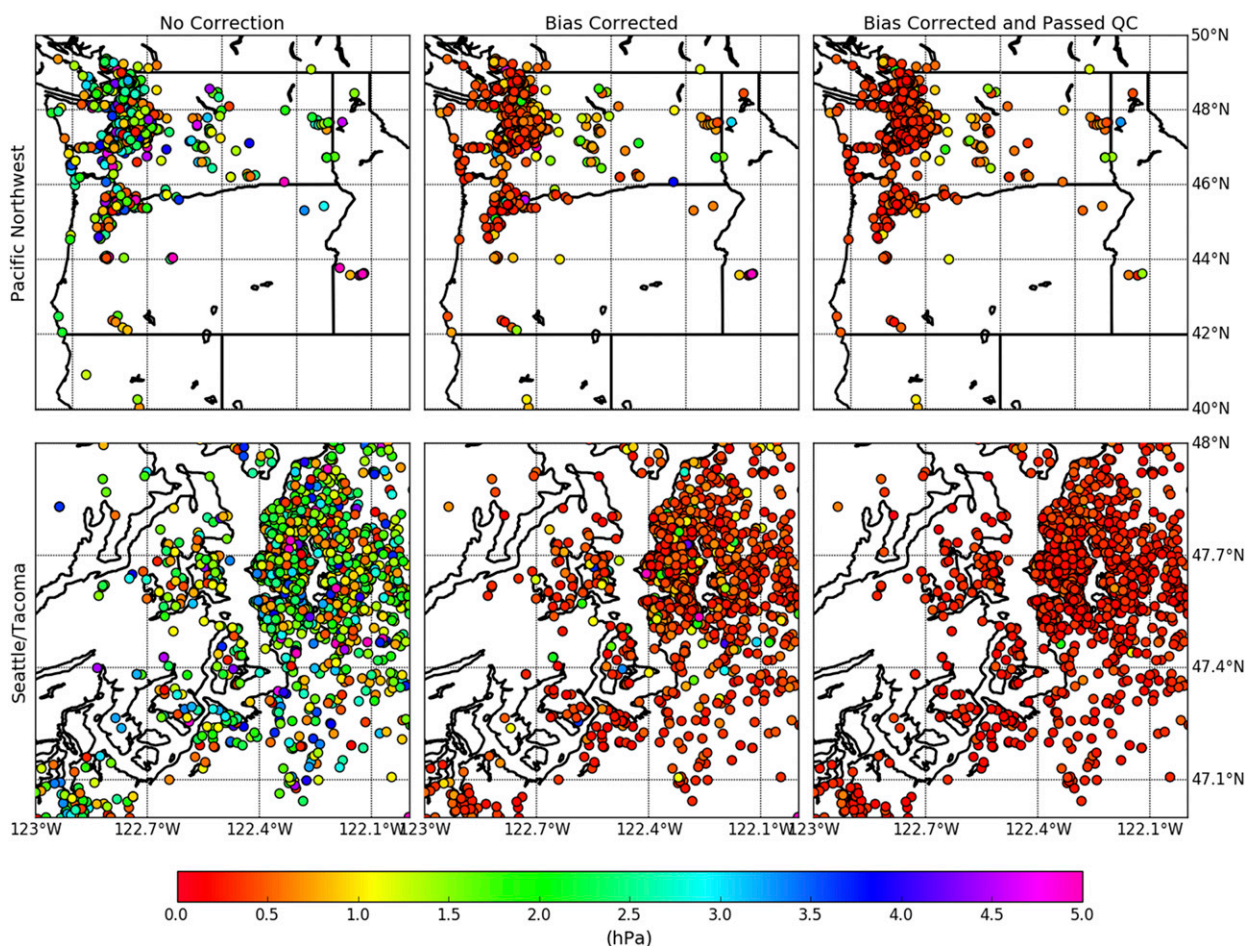


FIG. 7. The spatial distribution of uWx altimeter MAE at different stages of postprocessing. The altimeter MAE was computed for each smartphone from uWx SPOs retrieved between 10 Nov and 5 Dec 2016. Those 1974 smartphones used to compute the distribution of uWx altimeter MAE in Fig. 6 were used in this analysis.

likely to be stationary. If a smartphone is stationary and not charging, the Android operating system will reduce the frequency of background tasks, including pressure collection. On average 3469 SPOs are retrieved each hour. Of these SPOs approximately 96.8% (3359) undergo bias correction. Approximately 83.9% of SPOs (2910) undergo bias correction and pass all QC checks. In Madaus and Mass (2017), two-thirds of PressureNet and WeatherSignal SPOs were rejected during post-processing. By retaining four-fifths of all SPOs, uWx can substantially reduce the MAE of SPOs without sacrificing quantity for quality.

d. Feature importance

The previous section illustrated the effectiveness of random forests in reducing altimeter MAE. To gain

more insight into how random forests predict pressure biases, feature importance was evaluated for each uWx random forest. The distribution of random forest feature importance, for SPOs retrieved between 15 August and 9 November 2016, is displayed in Fig. 9a. It is notable that no single feature dominates (i.e., no median importance ≥ 0.5). The best predictor of pressure bias, horizontal location (latlon), has a median importance of 0.18 on a normalized scale from 0 to 1. It is not surprising that all features exhibit relatively low importance, since random forests, by design, randomize the selection of features at each step in each tree of the forest. This method is designed to reduce overfitting by preventing strong predictors of pressure bias, such as horizontal location (latlon), from dominating decision trees in the random forest.

uWx Error Statistics (10 November - 5 December, 2016)

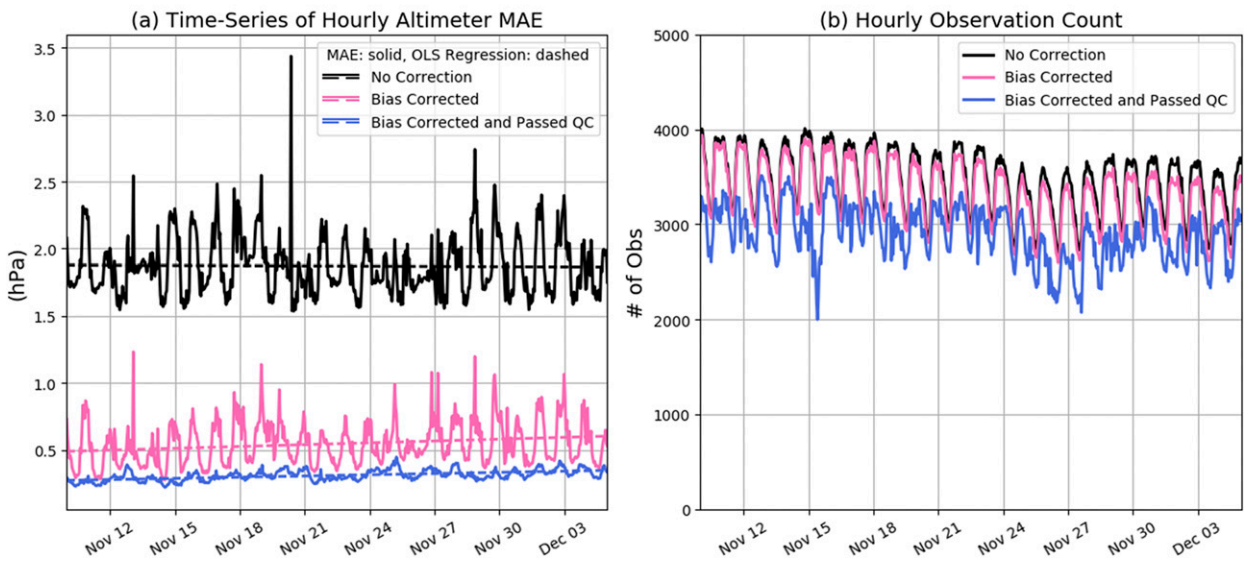


FIG. 8. (a) Time series of uWx hourly altimeter MAE at different stages of postprocessing over the period spanning 10 Nov–5 Dec 2016. Hourly altimeter MAE is computed from SPOs collected each hour. Ordinary least squares (OLS) regression was performed for each hourly altimeter MAE time series. (b) Time series of uWx hourly observation count at different stages of postprocessing. Hourly observation counts represent the number of SPOs retrieved within a given hour.

The rank (order) of feature importance is more significant than its normalized value, since the ranking of features reveals which features contribute the most to random forest bias predictions. A frequency plot of

feature rank, for each feature, is displayed in Fig. 9b. For 30% of all uWx smartphones, the most important feature for predicting pressure bias is horizontal location (latlon). Since elevation errors are a function of

uWx Feature Importance (15 August - 9 November, 2016)

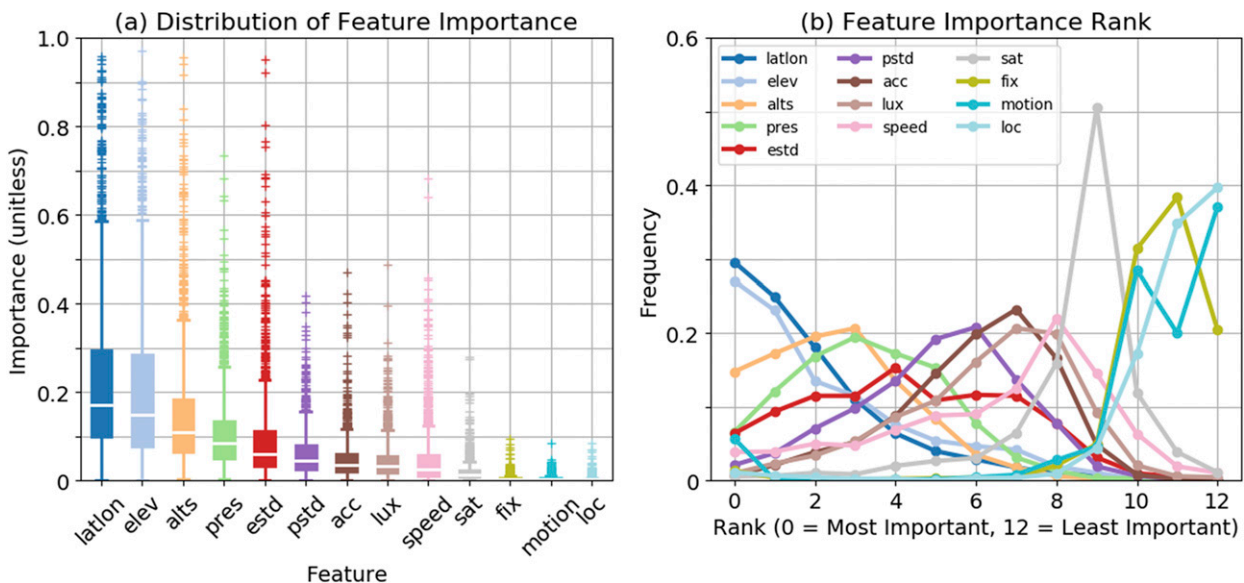


FIG. 9. (a) Feature importance of uWx variables (features) used to train random forests over the period 15 Aug–9 Nov 2016. For each phone, feature importance is estimated using the random forest generated for that phone. Each box plot displays the distribution of the importance of feature *X*, for all uWx phones. Median importance is displayed in each box plot (white bisecting line). (b) Frequency plot of feature rank from most important (rank = 0) to least important (rank = 12). For all uWx phones, feature importance was sorted and arranged in descending order. The frequency at which each feature was ranked *X* is plotted on the vertical axis.

horizontal location and ground elevation, it is not surprising that both features (latlon and elev) are the top-ranked feature for over 50% of smartphones. It is interesting to note that for a small fraction of uWx smartphones, features such as elevation variance (estd), significant motion (motion), and GPS speed (speed) are the top ranked feature. The first four top-ranked features (latlon, elev, alts, and pres) are correlated and thus convey similar information content. If only one or two of these features were used, a single feature could dominate over the others. While seemingly repetitive, the inclusion of these four features contributes to the robustness of random forests by increasing the diversity of individual trees.

e. Sensitivity to training period and feature selection

To evaluate how the size of the training sample affects bias prediction error, random forests were trained on uWx data of variable sample sizes. Specifically, a collection of random samples was retrieved from uWx smartphones that collected at least 1000 observations between 15 August and 15 November 2016. Approximately 1139 uWx smartphones met this criterion. The samples collected from these smartphones were used to train and verify random forests during fourfold cross validation. The predicted pressure bias and true pressure bias, calculated using the MADIS interpolation technique outlined above, were used to compute the RMS error of bias predictions. This error was computed from verification data not used during training. Figure 10 displays the RMS error of predicted pressure biases for uWx smartphones as a function of sample size. As the sample size increased, the skewness of the RMS bias prediction error decreased. With more samples available for training, random forests were more robust to outliers.

Several experiments were performed to examine the impact of feature selection on random forest bias prediction. These experiments utilized the same set of uWx smartphones used to test the sensitivity of random forests to training size. For the baseline experiment, feature selection was limited to horizontal location and elevation—the two most frequently top-ranked features for bias prediction (Fig. 9b). Additional experiments expanded feature selection to include pressure, altimeter setting, and all available features. For each experiment, random forests were trained and cross validated over the 3-month period spanning 15 August to 15 November 2016. The average percentile error of bias prediction was computed by averaging the Nth percentile of bias prediction error for all uWx random forests (smartphones), trained with a particular feature set. The average percentile error difference between random forests trained with each feature set is displayed in Fig. 11. A negative

error difference implies bias prediction errors were reduced relative to the baseline experiment.

The addition of pressure, altimeter setting, and ancillary features improves the average percentile error of bias prediction, most notably in the right tail of the distribution. Substituting pressure and elevation with altimeter setting yields weaker improvements in bias prediction than including pressure and elevation, even though altimeter setting is a function of both. As noted previously, training random forests with fewer features results in greater similarity among trees, which contributes to overfitting. Adding additional features increases the diversity of trees within the random forest, reducing correlations between trees and ensuring a more robust fit to outliers. For this reason, the greatest reductions in bias prediction error are observed at higher percentiles when random forests are trained using all features, even those with low overall importance.

5. Discussion

Convection-allowing numerical weather prediction demands observing networks with high spatiotemporal density for initialization and verification. uWx demonstrated that, through in-app QA procedures and bias correction, the quality of SPOs could be substantially improved without sacrificing their observation density. This result is encouraging for the utility of SPOs in numerical weather prediction, which has been limited by poor data quality (Madaus and Mass 2017). Prior attempts at improving the quality of SPOs overlooked sources of SPO error and relied on pressure data from preexisting surface observing networks (Kim et al. 2015; Kim et al. 2016; Madaus and Mass 2017). This study is the first to analyze sources of SPO error and demonstrate that, after a short training period, SPOs can be bias corrected in real time, using only data retrieved from the smartphone.

In numerical weather prediction, data assimilation systems require estimates of observational uncertainty to weight the influence of observations on the model. While considerable effort has been made in quantifying model uncertainty, less attention has been given to observational uncertainty. In ensemble data assimilation, pressure error variances are often assumed to be static (Wheatley and Stensrud 2010; Madaus et al. 2014; Sobash and Stensrud 2015; Madaus and Mass 2017), since error statistics for individual observations are typically unknown. In uWx, sensor noise, elevation uncertainty, pressure bias uncertainty, and the bias prediction RMS error can be used to estimate an observation error variance, unique to each SPO. By quantifying uncertainty, uWx SPOs can be weighted by their relative accuracy,

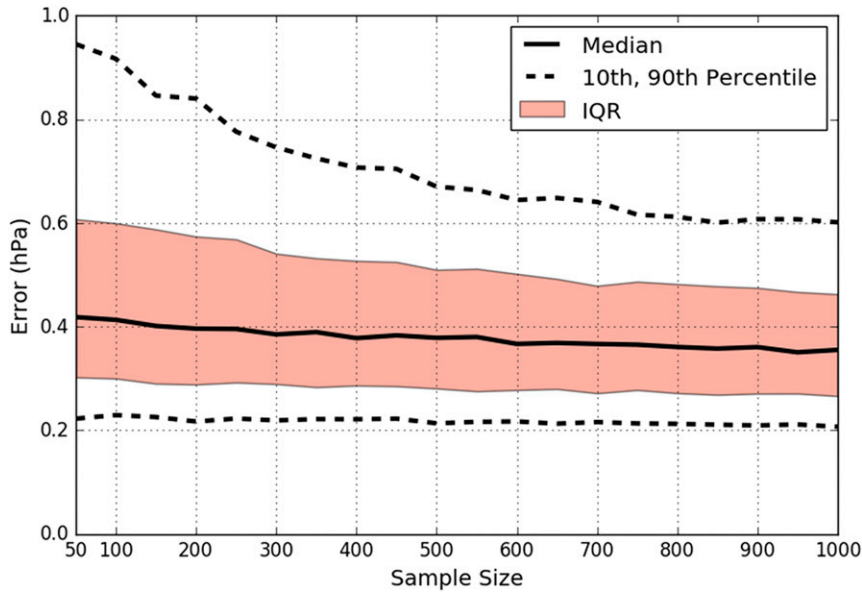


FIG. 10. RMS phone bias prediction error for random forests as a function of cross-validation sample size for 1139 smartphones. Shown are the median (solid line), the 10th (lower dashed line) and 90th percentiles (upper dashed line) of the bias prediction error, and the IQR of bias prediction error (shaded region).

enhancing data assimilation and increasing their potential for numerical weather prediction.

Bias correction using machine learning represents the biggest computation challenge for large-scale pressure

collection. Currently, uWx generates random forests for every smartphone on a remote server during post-processing. On a midrange Intel CPU (e.g., Intel Xeon E5-2620 version 2 at 2.10 GHz), it takes ~5 s to train a

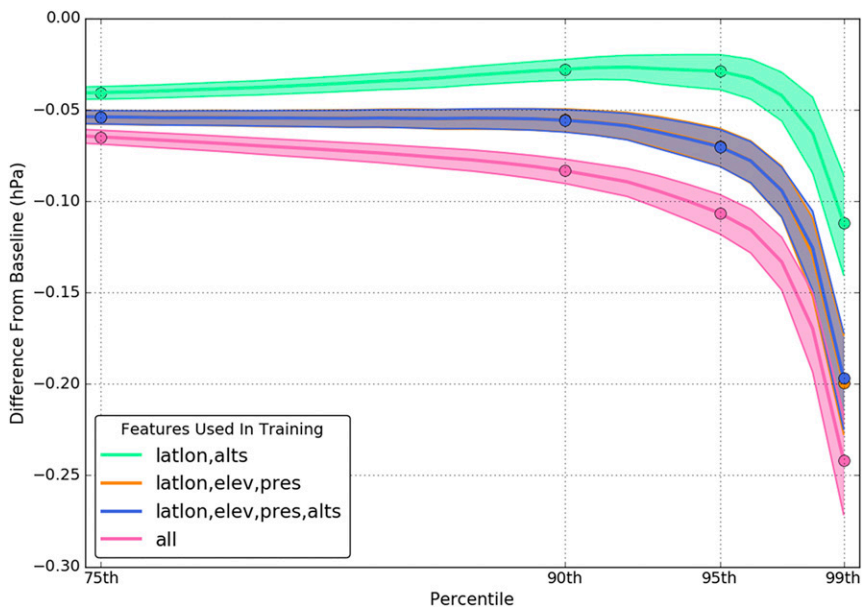


FIG. 11. Sensitivity of pressure bias prediction error to feature selection. Percentiles of pressure bias prediction error are computed for each smartphone. The N th percentile of pressure bias prediction error is averaged across 1139 uWx smartphones. This procedure is performed for random forests trained with different sets of features. As a baseline, random forests are trained using only location data (latlon and elev). The average percentile error difference between random forests trained with each feature set and this baseline is displayed above.

random forest, over a 3-month period, on a single CPU thread. At this rate it would take ~ 1400 processors to train a million random forests in an hour. One possible solution involves leveraging the massive parallel architecture of graphics processing units (GPUs) to perform bias correction for many smartphones simultaneously. Another more scalable approach would involve performing in-app bias correction by embedding the random forest algorithm into the app code. While this approach is promising, the practicality of performing machine learning in app remains unclear and is a topic of future research.

6. Conclusions

Pressure observations acquired from millions of smartphones have the potential to provide dense surface observations around the world. Previously, inconsistent data quality from smartphone pressures undermined the applications of such observations in numerical weather prediction. However, if the data quality can be improved, the smartphone pressures could enhance numerical weather prediction by providing unprecedented observational coverage and density. This study attempts to confront this challenge by developing new approaches to quantify uncertainty and to reduce error in smartphone pressures. To act as a test bed for pressure collection and QA procedures, a crowdsourcing pressure app, uWx, was developed. Among the innovations in uWx was the extension of the sensor listening period, which mitigated error caused by a failure to account for sensor filtering. Location accuracy was also improved by mandating the use of the GPS during location retrieval, reducing the magnitude of position and elevation errors.

Although QA procedures were successful in mitigating error associated with collection and location problems, pressure biases persisted as sensor bias and vertical elevation errors remained unaccounted for. Since traditional QC techniques failed to produce the desired results, a machine learning approach using random forests for postprocessing smartphone pressures was developed. This approach, in combination with simple QC checks, reduced the mean smartphone pressure bias by 82%. Remarkably, improvements in data quality were gained at little expense to data quantity. On average, 84% of SPOs underwent bias correction and passed all QC checks.

As more meteorological sensors are embedded in devices, vehicles, buildings, and other items, the potential for crowdsourced weather observations will grow. This study demonstrates that pressure biases in crowdsourced data from mobile platforms can be predicted and removed, facilitating their use in numerical weather prediction. This work also demonstrates the substantial promise of machine learning techniques for enhancing the value of observations. In a sister study, ensemble data assimilation experiments with

crowdsourced uWx pressures will be used to quantitatively examine whether SPOs can improve mesoscale forecasts.

Acknowledgments. The authors also acknowledge the Weather Company (IBM) for its generous financial support of this research and its provision of pressure data from this app. Peter Neiley, of the Weather Company, was instrumental in our acquisition of Weather Channel smartphone pressure data. Support was also provided by a grant from the NOAA CSTAR program through Grant NA10OAR4320148AM63. The authors would like to acknowledge the uWx users, past and present, who made this research possible

REFERENCES

- Android, 2017a: Motion sensors. Google, accessed 28 April 2017, https://developer.android.com/guide/topics/sensors/sensors_motion.html.
- , 2017b: Monitoring the battery level and charging state. Google, accessed 28 April 2017, <https://developer.android.com/training/monitoring-device-state/battery-monitoring.html>.
- Baldauf, M., A. Seifert, J. Förstner, D. Majewski, M. Raschendorfer, and T. Reinhardt, 2011: Operational convective-scale numerical weather prediction with the COSMO model: Description and sensitivities. *Mon. Wea. Rev.*, **139**, 3887–3905, <https://doi.org/10.1175/MWR-D-10-05013.1>.
- Bosch Sensortec, 2012: Digital barometric pressure sensor BMP280 from Bosch Sensortec sets market reference: Improved performance, smallest footprint. Bosch Press Release, 3 pp., <http://www.bosch-presse.de/pressportal/de/en/digital-barometric-pressure-sensor-bmp280-from-bosch-sensortec-sets-market-reference-41912.html>.
- , 2018: BMP280 digital pressure sensor. Bosch Datasheet, 49 pp., https://ae-bst.resource.bosch.com/media/_tech/media/datasheets/BST-BMP280-DS001-19.pdf.
- Breiman, L., 1996: Bagging predictors. *Mach. Learn.*, **24**, 123–140, <https://doi.org/10.1023/A:1018054314350>.
- , 2001: Random forests. *Mach. Learn.*, **45**, 5–32, <https://doi.org/10.1023/A:1010933404324>.
- Chapman, L., C. Bell, and S. Bell, 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. *Int. J. Climatol.*, **37**, 3597–3605, <https://doi.org/10.1002/joc.4940>.
- COESA, 1976: *U.S. Standard Atmosphere, 1976*. NOAA, 227 pp.
- De Vos, L., H. Leijnse, A. Overeem, R. Uijlenhoet, 2017: The potential of urban rainfall monitoring with crowdsourced automatic weather stations in Amsterdam. *Hydrol. Earth Syst. Sci.*, **21**, 765–777, <https://doi.org/10.5194/hess-21-765-2017>.
- 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. *J. Atmos. Oceanic Technol.*, **34**, 1853–1866, <https://doi.org/10.1175/JTECH-D-16-0150.1>.
- Duchon, J., 1976: Splines minimizing rotation-invariant semi-norms in Sobolev spaces. *Constructive Theory of Functions of Several Variables*, W. Schempp and K. Zeller, Eds., Lecture Notes in Mathematics, Vol. 571, Springer, 85–100, <https://doi.org/10.1007/BFb0086566>.

- Elmore, K. L., Z. L. Flamig, V. Lakshmanan, B. T. Kaney, V. Farmer, H. D. Reeves, and L. S. Rothfus, 2014: mPING: Crowd-sourcing weather reports for research. *Bull. Amer. Meteor. Soc.*, **95**, 1335–1342, <https://doi.org/10.1175/BAMS-D-13-00014.1>.
- Ester, M., H.-P. Kriegel, J. Sander, and X. Xu, 1996: A density-based algorithm for discovering clusters in large spatial databases with noise. *KDD'96: Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, E. Simoudis, J. Han, and U. Fayyad, Eds., AAAI Press, 226–231.
- Gallus, W. A., Jr., J. Correia, and I. Jankov, 2005: The 4 June 1999 derecho event: A particularly difficult challenge for numerical weather prediction. *Wea. Forecasting*, **20**, 705–728, <https://doi.org/10.1175/WAF883.1>.
- Gesch, D. B., M. J. Oimoen, and G. A. Evans, 2014: Accuracy assessment of the U.S. Geological Survey National Elevation Dataset, and comparison with other large-area elevation datasets—SRTM and ASTER. U.S. Geological Survey Open-File Rep. 2014-1008, 10 pp.
- Hanson, G. S., and S. J. Greybush, 2016: Impact of assimilating surface pressure observations from smartphones on regional, convective-allowing ensemble forecasts: Observing system simulation experiments. M.S. thesis, Dept. of Meteorology and Atmospheric Science, The Pennsylvania State University, 47 pp.
- Ho, T. K., 1998: The random subspace method for constructing decision forests. *IEEE Trans. Pattern Anal. Mach. Intell.*, **20**, 832–844, <https://doi.org/10.1109/34.709601>.
- Infineon, 2016: DPS310—Digital pressure sensor: Digital barometric pressure sensor for portable devices. Infineon Technologies AG Doc. IFX-sch1406115644540, 41 pp., https://www.infineon.com/dgdl/Infineon-DPS310-DS-v01_00-EN.pdf?fileId=5546d462576f34750157750826c42242.
- 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. *J. Sens.*, **2015**, 245498, <https://doi.org/10.1155/2015/245498>.
- Kim, Y.-H., J.-H. Ha, Y. Yoon, N.-Y. Kim, H.-H. Im, S. Sim, and R. K. Y. Choi, 2016: Improved correction of atmospheric pressure data obtained by smartphones through machine learning. *Comput. Intell. Neurosci.*, **2016**, 4, <https://doi.org/10.1155/2016/9467878>.
- Lean, H. W., P. A. Clark, M. Dixon, N. M. Roberts, A. Fitch, R. Forbes, and C. Halliwell, 2008: Characteristics of high-resolution versions of the Met Office Unified Model for forecasting convection over the United Kingdom. *Mon. Wea. Rev.*, **136**, 3408–3424, <https://doi.org/10.1175/2008MWR2332.1>.
- Madaus, L., and C. Mass, 2017: Evaluating smartphone pressure observations for mesoscale analyses and forecasts. *Wea. Forecasting*, **32**, 511–531, <https://doi.org/10.1175/WAF-D-16-0135.1>.
- , G. Hakim, and C. Mass, 2014: Utility of dense pressure observations for improving mesoscale analyses and forecasts. *Mon. Wea. Rev.*, **142**, 2398–2413, <https://doi.org/10.1175/MWR-D-13-00269.1>.
- Mass, C. F., and L. Madaus, 2014: Surface pressure observations from smartphones: A potential revolution for high-resolution weather prediction? *Bull. Amer. Meteor. Soc.*, **95**, 1343–1349, <https://doi.org/10.1175/BAMS-D-13-00188.1>.
- , D. Ovens, K. Westrick, and B. A. Colle, 2002: Does increasing horizontal resolution produce more skillful forecasts? *Bull. Amer. Meteor. Soc.*, **83**, 407–430, [https://doi.org/10.1175/1520-0477\(2002\)083<0407:DIHRPM>2.3.CO;2](https://doi.org/10.1175/1520-0477(2002)083<0407:DIHRPM>2.3.CO;2).
- McNicholas, C., and D. D. Turner, 2014: Characterizing the convective boundary layer turbulence with a high spectral resolution lidar. *J. Geophys. Res. Atmos.*, **119**, 12 910–12 927, <https://doi.org/10.1002/2014JD021867>.
- Meier, F., D. Fenner, T. Grassmann, M. Otto, D. Scherer, 2017: Crowdsourcing air temperature from citizen weather stations for urban climate research. *Urban Climate*, **19**, 170–191, <https://doi.org/10.1016/j.uclim.2017.01.006>.
- Miller, P. A., M. F. Barth, and L. A. Benjamin, 2005: An update on MADIS observation ingest, integration, quality control and distribution capabilities. *21st Int. Conf. on Interactive Information and Processing Systems for Meteorology, Oceanography, and Hydrology*, San Diego, CA, Amer. Meteor. Soc., J7.12, https://ams.confex.com/ams/Annual2005/techprogram/paper_86703.htm.
- Muller, C. L., L. Chapman, S. Johnston, C. Kidd, S. Illingworth, G. Foody, A. Overeem, and R. R. Leigh, 2015: Crowdsourcing for climate and atmospheric sciences: Current status and future potential. *Int. J. Climatol.*, **35**, 3185–3203, <https://doi.org/10.1002/joc.4210>.
- Newzoo, 2017: Top 50 countries by smartphone users and penetration. Newzoo, accessed 28 April 2017, <https://newzoo.com/insights/rankings/top-50-countries-by-smartphone-penetration-and-users/>.
- 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. *Geophys. Res. Lett.*, **40**, 4081–4085, <https://doi.org/10.1002/grl.50786>.
- Pinto, J. O., J. A. Grim, and M. Steiner, 2015: Assessment of the High-Resolution Rapid Refresh model's ability to predict mesoscale convective systems using object-based evaluation. *Wea. Forecasting*, **30**, 892–913, <https://doi.org/10.1175/WAF-D-14-00118.1>.
- Roebber, P. J., D. M. Schultz, and R. Romero, 2002: Synoptic regulation of the 3 May 1999 tornado outbreak. *Wea. Forecasting*, **17**, 399–429, [https://doi.org/10.1175/1520-0434\(2002\)017<0399:SROTM>2.0.CO;2](https://doi.org/10.1175/1520-0434(2002)017<0399:SROTM>2.0.CO;2).
- , —, B. A. Colle, and D. J. Stensrud, 2004: Toward improved prediction: High-resolution and ensemble modeling systems in operations. *Wea. Forecasting*, **19**, 936–949, [https://doi.org/10.1175/1520-0434\(2004\)019<0936:TIPHA>2.0.CO;2](https://doi.org/10.1175/1520-0434(2004)019<0936:TIPHA>2.0.CO;2).
- 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. *Mon. Wea. Rev.*, **139**, 976–991, <https://doi.org/10.1175/2010MWR3425.1>.
- Shepard, D., 1968: A two-dimensional interpolation function for irregularly-spaced data. *ACM'68: Proceedings of the 1968 23rd ACM National Conference*, ACM, 517–524, <https://doi.org/10.1145/800186.810616>.
- Sobash, R. A., and D. J. Stensrud, 2015: Assimilating surface mesonet observations with the EnKF to improve ensemble forecasts of convection initiation on 29 May 2012. *Mon. Wea. Rev.*, **143**, 3700–3725, <https://doi.org/10.1175/MWR-D-14-00126.1>.
- Sun, J., and Coauthors, 2014: Use of numerical weather prediction for nowcasting convective precipitation: Recent progress and challenges. *Bull. Amer. Meteor. Soc.*, **95**, 409–426, <https://doi.org/10.1175/BAMS-D-11-00263.1>.
- Weisman, M. L., C. Davis, W. Wang, K. W. Manning, and J. B. Klemp, 2008: Experiences with 0–36-h explicit convective forecasts with the WRF-ARW model. *Wea. Forecasting*, **23**, 407–437, <https://doi.org/10.1175/2007WAF2007005.1>.
- Wheatley, D., and D. Stensrud, 2010: The impact of assimilating surface pressure observations on severe weather events in a WRF mesoscale system. *Mon. Wea. Rev.*, **138**, 1673–1694, <https://doi.org/10.1175/2009MWR3042.1>.