Bootstrap based Uncertainty Propagation for Data Quality Estimation in Crowdsensing Systems

Valerio Freschi, Saverio Delpriori, Emanuele Lattanzi, and Alessandro Bogliolo

Department of Pure and Applied Sciences

University of Urbino

Urbino, Italy

Email: {valerio.freschi,saverio.delpriori,emanuele.lattanzi,alessandro.bogliolo}@uniurb.it

Abstract—The diffusion of mobile devices equipped with sensing, computation, and communication capabilities is opening unprecedented possibilities for high-resolution, spatio-temporal mapping of several phenomena. This novel data generation, collection and processing paradigm, termed crowdsensing, lays upon complex, distributed cyberphysical systems. Collective data gathering from heterogeneous, spatially distributed devices inherently raises the question of how to manage different quality levels of contributed data. In order to extract meaningful information, it is therefore desirable the introduction of effective methods for evaluating the quality of data.

In this work, we propose an approach aimed at systematic accuracy estimation of quantities provided by end-user devices of a crowd-based sensing system. This is obtained thanks to the combination of statistical bootstrap with uncertainty propagation techniques, leading to a consistent and technically sound methodology. Uncertainty propagation provides a formal framework for combining uncertainties resulting from different quantities influencing a given measurement activity. Statistical bootstrap enables the characterization of the sampling distribution of a given statistics without any prior assumption on the type of statistical distributions behind the data generation process.

The proposed approach is evaluated on synthetic benchmarks and on a real world case study. Cross-validation experiments show that confidence intervals computed by means of the presented technique show a maximum 1.5% variation with respect to interval widths computed by means of controlled standard Monte Carlo methods, under a wide range of operating conditions. In general, experimental results confirm the suitability and validity of the introduced methodology.

I. INTRODUCTION

CROWDSENSING is a data acquisition paradigm that is becoming increasingly popular thanks to the wide diffusion of mobile smart devices. The pervasive nature of these devices, combined to their sensing, computation and communication capabilities, makes them the ideal candidates as building blocks of distributed cyberphysical systems (hereafter denoted as CPS) to achieve high resolution spatio-temporal sensing of various kinds of physical quantities [1]–[4].

The analysis of recent scientific literature highlights a surge of works related to mobile crowdsensing in many applicative fields. For instance, crowd-based sensing systems have been developed for monitoring road infrastructures, with different aims: evaluation of road surface conditions [5]–[9], control of traffic congestion [10], identification and mapping of traffic regulators [11]. Other applications concern the detection of available parking spots by means of smartphones [12], [13] or mobile ultrasonic sensors [14]. Several proposals have been presented for environmental applications, with a focus on air pollution [15]–[17] and urban noise monitoring [18], [19].

The distributed structure of crowdsensing systems, the heterogeneity of the devices involved and nature of collective contribution (based on involvement of end-users) influence the quality of data and raise the issue of how to effectively deal with it [1]. Sensed data is collected, processed and possibly aggregated for analysis at various levels of a distributed CPS. To fully enhance the exploitability of this wealth of data, quality indicators should be associated to numbers representing a given sensed quantity (enabling for example to discriminate between reliable and unreliable data). Obtaining an estimate of such a quality has a direct impact on inference and decision processes carried on, in cascade, by other components of the system.

As a matter of fact, different features of a crowdsensing system concur, at different levels, to define data quality: *i)* mapping of geolocalized phenomena/events is affected by the spatial and temporal resolution of the monitoring activity; *ii)* sensed quantities are subject to systematic and random errors because of inherent precision and accuracies associated to embedded sensors and because of the different computing and communication capabilities (e.g. different energy levels or channel characteristics); *iii)* users may contribute data corresponding to different trustworthiness levels. Approaches to deal with the problem of data quality mainly focus on outlier identification and filtering [1], collaborative data inconsistency resolution [20], or trust and reputation systems to promote and ensure the identification of corrupted or malicious contributions [21], [22].

However, since data is the result of a direct or indirect measurement activity carried on by sensing devices, handling the uncertainty associated to each of these measurements (as usually done in physical sciences [23]) appears a natural way to deal with the issue of data quality. Indeed, we propose in this paper a methodology that lays on the identification of the given crowdsensing system as a distributed instrument. Measurements provided by terminal devices represent estimates of the value to be assigned to the quantity to be measured (the *measurand*). Hence, we cast the problem of data quality assignment as an evaluation of the uncertainty of the underlying measurement process. In order to be meaningful, the measurand estimate has to be associated to a measure

of uncertainty. Typically, this is achieved by providing the amount of dispersion of measured values (the smaller the dispersion, the smaller the uncertainty). The interaction among error sources determines dispersion around the true (unknown) value of the measurand and, therefore, the measurement error. If the quantity to be measured is the result of the composition of different measurements (e.g. in the case of an indirect measurement), the uncertainties associated to each measurement phase have to be propagated computing their combination [23]. Uncertainty propagation has been recently applied to sensor networks scenarios to derive (together with trust values) a sensing uncertainty metric used for in-network processing and fault detection [24].

Standard approaches to the propagation of uncertainties usually entail making some modeling assumptions on the measurement process, in particular regarding the probability density functions of quantities involved in measurement [23], [25], [26]. These probability functions (and the distribution they describe) are used in two possible ways: *i)* to analytically derive the distribution of the measured variable given its functional dependence from the input quantities; *ii)* to perform Monte Carlo simulations in order to obtain numerical estimate of the measurand's output distribution, given the functional relationships among input probability functions [25]–[27].

Monte Carlo numerical simulations are a viable solution when reasonable assumptions about input distributions hold. However, this is often not the case: for instance, regarding mobile distributed CPS, it is not easy to derive a mathematical model that describes the statistical features of measurements performed by a variety of devices in different places. Moreover, even if adequate statistical models of input variables can be obtained, their complex composition makes it hard to obtain an analytical expression of the output. For instance, the output variable could be the result of a recursive function that provides indirect computation of the measurand (as for the case study that will be detailed in the following). Other problems raise when the temporal dimension is taken into account. As a matter of fact, composition of different measurements during the whole lifetime of a crowdsensing system can be considered a common practice to smooth and update collected information. This poses the challenge of how to manage such an *update of measurements* at given checkpoints.

To overcome the above mentioned issues, we propose in this work to make use of statistical non-parametric bootstrap. Statistical bootstrap is a widely diffused data driven framework for empirically establishing the uncertainty of unknown quantities when modeling techniques and analytical solutions cannot be easily applied [28], [29]. Basically, the bootstrap is a type of Monte Carlo method that approximates the *sampling distribution* by sampling *with replacement* the original observations (i.e. the data on which inference has to be performed). For each resample, the statistic of interest (e.g. the *sample mean*) is computed and stored. The resulting distribution, called *bootstrap distribution*, can be used as a proxy to make inference on the shape and spread of the sampling distribution of the statistic [28]–[31]. The use of bootstrap to explore uncertainty propagation has been recently proposed by Kass *et al.* in the context of analysis of neural data [32] and applied to improve estimation of a blood pressure measurement system [33].

The main contributions of this work are:

1) A method that frames the problem of assessing data quality in crowdsensing platforms into a formal, technically sound approach, by casting it as an uncertainty evaluation problem.

2) The design, by means of the non-parametric bootstrap, of a data-driven strategy which could be used to get rid of complex interplay among the (potentially many) variables affecting the measure estimate and the related uncertainty propagation.

3) The evaluation of the applicability on a distributed CPS, characterized by periodic update of data and related quality.

The proposed method is validated by means of numerical simulations on synthetic benchmarks, and exemplified on a real-world case study, namely a crowdsensing platform for road surface roughness monitoring. The experimental results highlight the suitability of the presented approach to gauge the uncertainty associated to complex sensing tasks and, consequently, to provide an evaluation of data quality in crowdsensing systems.

The rest of the paper is organized as follows: in Section II we describe the crowdsensing system architecture taken as reference; in Section III we describe the proposed approach, namely the propagation of uncertainties by means of statistical bootstrap; in Section IV we introduce the experimental set up and present performance results; finally, in Section V, we draw concluding remarks.

II. REFERENCE SYSTEM ARCHITECURE

We will often refer, along this article, to a crowdsensing platform called *SmartRoadSense* (hereafter also denoted as SRS), developed to provide quantitative evaluation of roads surface roughness [5]–[8], whose basic structure and features will be described in this section.

The proposed approach is general enough to be exploited also in other analogous contexts, wherever sensing tasks are performed by multiple sensing devices which contribute to estimate the quantity of interest in a specific geo-localized position.

Figure 1 shows the architecture of the SRS platform, which is characterized by the following three layers:

- An app running on users' smartphones during a given car trip. The application makes use of accelerometers to collect and process acceleration values to which the device is subject. The result, representing the estimated roughness of the travelled road in a given point at a given time, is geo-referenced, time-stamped and transmitted to a server by means of radio connectivity.

- A cloud-based back-end service in charge of collecting, aggregating and storing data from multiple users. According to Figure 1, this layer is in charge of two tasks:

  – *Map matching*: georeferenced roughness indexes stored in the database of raw-data (SRS_RAW) are projected on digital cartography maps, specifically
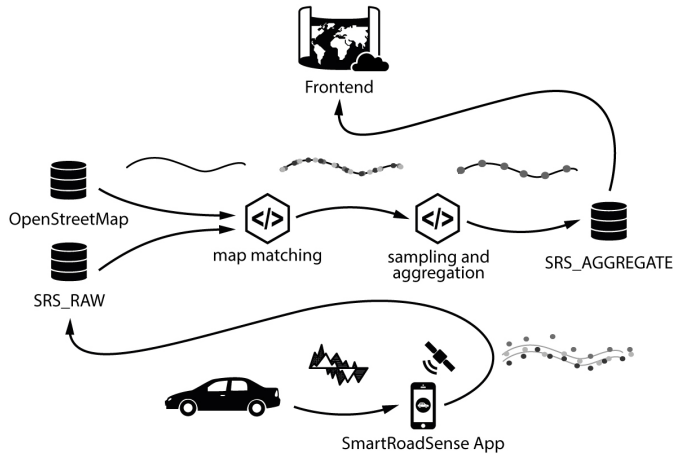
Fig. 1. SmartRoadSense architecture.

OpenStreetMap[1]. Map matching entails the alignment of GPS points to digital cartography maps;
- – *Sampling and aggregation*: data is subsequently aggregated to provide a single evaluation (for a given spatial coordinate) of the roughness index, given the data made available for that point by multiple users. Aggregated data is used to populate the related database (called SRS_AGGREGATE).
- • A front-end service providing visualization capabilities of the geo-referenced information produced by the SRS processing pipeline.

In SRS (and possibly in other crowdsensing systems), the activity can be divided into time epochs, during which data is continuously gathered, processed and aggregated. Segmentation of both space and time dimensions (e.g. through the definition of a bi-dimensional grid and the discretization of the time axis) can be in fact considered a common approach to the design of CPS at different spatio-temporal resolutions [21], [22], [34].

At the end of a given time epoch the system updates current information on the status of measured variables and, in case, it performs some type of composition with data collected in previous epochs. For instance, in SRS an epoch represents a week of monitoring activity. The platform continuously receives values of road roughness from end users. Roads are spatially segmented into landmark points, then all values associated to positions falling within a given range (typically 30m) of a landmark point p are aggregated and concur to the overall roughness index of p (the mean value of contributed points is taken by default). At the end of each week current epoch terminates, and the roughness value of each point p is updated by taking the average between the value of current epoch and the value of previous epoch. This processing inherently implements a form of infinite impulse response filter, the aim of which is to progressively downweigh (through

---

an exponential decay of weights) the contribution of older samples to the value assigned to p. Needless to say, different update rules can be conceived, according to different specific needs.

The above description exemplifies the difficulties that could arise when dealing with uncertainty propagation in these settings, since the measurand (the roughness index of p in SRS) needs to be tracked along its evolution and the corresponding unknown uncertainty subject to possibly complicated transformations.

III. BOOTSTRAP BASED UNCERTAINTY PROPAGATION

To circumvent all the issues related to the propagation of uncertainty in crowdsensing platforms that prevent the adoption of analytical and Monte Carlo methods, we propose in this paper to take advantage of the statistical bootstrap.

Figure 2 provides an overview of the toolflow of the proposed approach when applied to the SRS crowdsensing system, while Table I summarizes the symbols used along the article.

TABLE I
DESCRIPTION OF NOTATIONS AND SYMBOLS.

Symbol	Description
p	Generic geo-referenced landmark (aggregated) point
nw	Number of epochs
$t_0, t_1, ...t_i, ...t_{nw}$	Time epochs
$n_0, n_1, \cdots, n_{nw}$	Size of samples and of resamples at epoch t_i
N_b	Number of bootstrap resampling iterations
$\overline{x}$	Mean value of a generic resample
$\overline{x}_i$	Mean value of resample at epoch t_i
$\overline{y}_i$	Measurand variable computed at epoch t_i
$\mathcal{Y}$	Generic function relating x_i to y_i

As illustrated by Figure 2, data produced by terminal devices at a given time epoch t_i (with $i = 0, 2..., n_{nw}$) are collected into a sample of size n_i. Nonparametric bootstrap is applied to this sample (for each time epoch). Data is sampled with replacement, obtaining N_b resamples, each of size n_i. The statistic of interest (i.e. the mean value $\overline{x}$) is computed from every resample and plugged into the processing block labeled as $\mathcal{Y}$, which represents the functional relationship between all variables that influence the measurand. Concerning SRS, this phase encompasses the propagation of uncertainty along the different time epochs according to the update filter: each mean value $\overline{x}_i$ at time epoch t_i is averaged with the corresponding value y_i computed at time epoch t_{i-1}. The result $y_i = \frac{\overline{x}_i + y_{i-1}}{2}$ is then stored as current value to be composed with a new measurement at next time epoch.

The attained distribution of values assigned to y is the output bootstrap distribution which can be studied to obtain information about its center, shape and spread. While the center of the output bootstrap distribution represents the estimate of the statistic under study (the mean value in this specific case), the shape provides effective information about the type of distribution and, finally, the spread conveys information about the output uncertainty (which is what we are searching for). Needless to say, bootstrap resampling does not lead to any

---
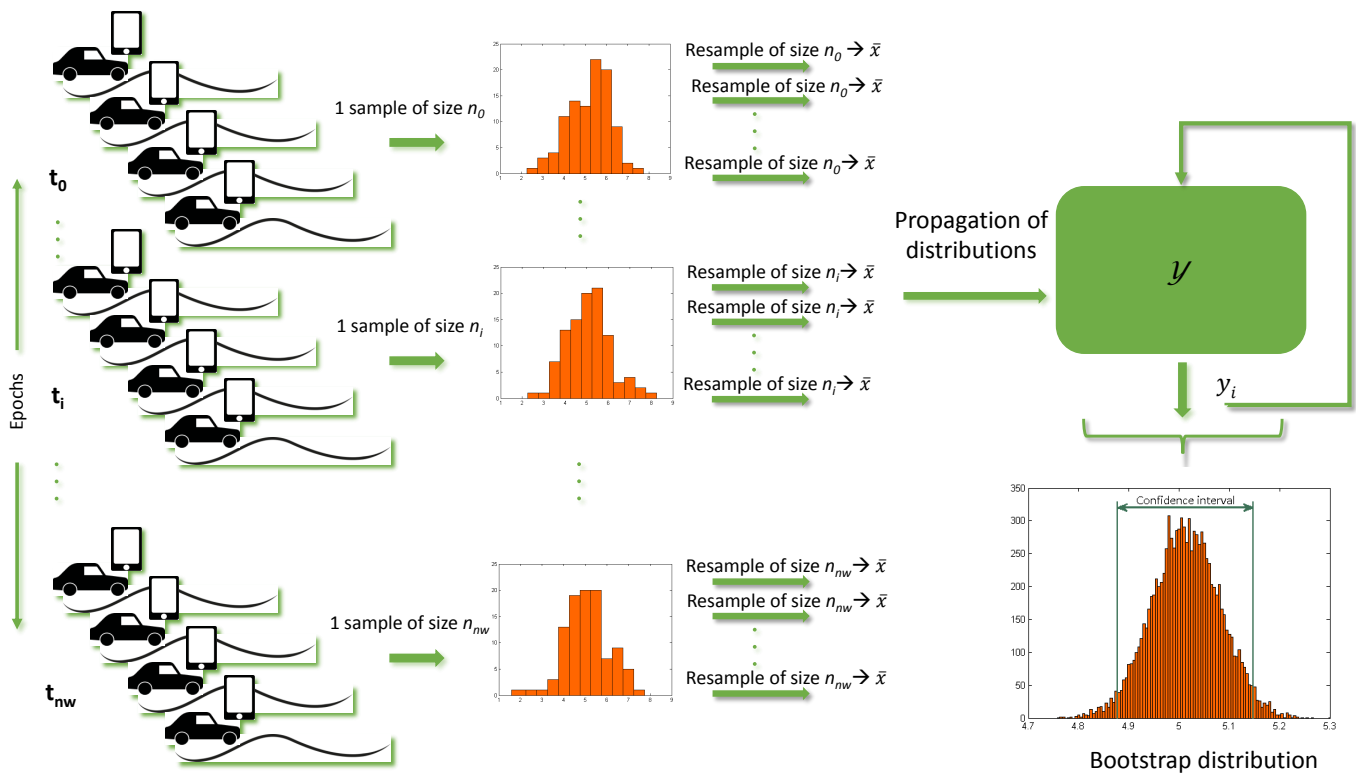
[1] http://www.openstreetmap.org

Fig. 2. Toolflow of the proposed approach.

improvement in the accuracy of the estimate of the statistic, since this clearly depends on the accuracy of the initial sample from which all resamples are derived. Nevertheless, important information regarding the sampling distribution is encoded in the bootstrap distribution and this is exactly what we are interested to exploit to infer the uncertainty of measurement. The histogram placed in right-bottom corner of Figure 2 illustrates this concept. Indeed, bootstrap distribution values are finally sorted and samples representative of a given percentile are extracted to provide a confidence interval CI to be used as uncertainty estimate.

It is worth noticing that other statistics might be evaluated according to the presented method. In fact, one of the major points of strength of the bootstrap is its flexibility in handling different types of statistics, which has to be contrasted with the difficulties faced to derive analytical results (for example using asymptotic arguments). If we were interested, for instance, at investigating the uncertainty associated to the median value (instead of the mean) the same approach would remain valid and we would only need to change the computation of the mean value $\overline{x}$ from each resample with the respective median value.

The main algorithmic steps of the proposed approach can be summarized as follows:

1) **for** $k = 1$ to N_b
   - **for** $i = 2$ to nw
     – Sample with replacement the observation vector collected during time epoch t_i
     – Compute mean value $\overline{x}_i$ (or any statistics of interest) of the bootstrap resample at time t_i

   – Update measurand y_i according to $\mathcal{Y}$. In SRS, $y_i = (\overline{x}_i + y_{i-1})/2$
2) Extract 95% confidence interval from bootstrap distribution

The whole process represented by the pipeline reported in Figure 2 and by the above described pseudo-code is repeated for each crowd-based measurement (i.e. for each aggregated point in the SRS example setting). This raises the question of the scalability of the system, which should be taken into consideration when a huge number of uncertainty evaluations have to be carried on. While a detailed discussion of this topic is out of the scope of the present article, it should be remarked the inherent parallelism of the proposed approach. In fact, uncertainty intervals associated to different geo-localized points can be computed independently from each other. Therefore, in principle, they can be split in many processing tasks that can be autonomously executed in parallel, potentially mitigating the impact of computational burden.

IV. EXPERIMENTAL RESULTS

To validate the introduced technique, several experiments have been conducted:

- First, a set of synthetic benchmarks has been devised to compare the bootstrap based uncertainty evaluation against a standard Monte Carlo method, under the assumption of knowing the input probability distributions, needed to to run the Monte Carlo experiments.
- Second, a sensitivity analysis has been performed to evaluate the dependence of the results from the number

of bootstrap resamplings (N_b), allowing to explore the tradeoff between accuracy and computational complexity.

- Third, an experiment has been conducted to simulate the case of a system measuring a time-varying quantity.
- Last, the uncertainty of a measurement within the SRS crowdsensing platform has been computed to show the applicability to a real world use-case.

A. Synthetic benchmarks

The rationale of these experiments was to assess the suitability of the proposed approach in terms of accuracy of the confidence interval. The proposed bootstrap-based uncertainty propagation has been validated by comparing it with standard Monte Carlo uncertainty propagation (SMC, for short), assuming the knowledge of input probability density functions. We recall that, while this is an assumption that has to be made if one wants to apply standard Monte Carlo propagation, it cannot be taken for granted. The bootstrap propagation technique, conversely, doesn't rely on any type of knowledge of input data, rather it performs a data-driven Monte Carlo simulation by drawing the so called *pseudo-observations* from the vector of initial observations and generating from it (through resampling) all the information needed for inference tasks.

Three types of distributions have been considered, covering a wide spectrum of possible statistical configurations, namely: a Gaussian distribution of mean $\mu = 5$ and standard deviation $\sigma = 1$, a uniform distribution taking values in the $[4, 6]$ interval, and a Rayleigh distribution with scale parameter $b = 5$.

We included the Gaussian distribution because of its role in statistics and error distributions [28]. As well, we chose the uniform distribution since it is often studied in uncertainty evaluations of measurements [25]. Finally, we also took into consideration the Rayleigh distribution because it is an example of asymmetric distribution, which adds to the significance and coverage of our experiments.

For what concerns the bootstrap based uncertainty propagation, we generated 100 points (drawn from each of the input distributions) representing the observations. Sampling with replacement has been performed with $N_b = 10^5$ replications, mean values have been computed and given as input to the propagation pipeline representing the update process described in Section III: the mean of observed values at each epoch has been averaged with the mean of observed values at previous epoch. Three sets of experiments were performed, simulating a time horizon of, respectively 2, 10 and 25 epochs (on a system like SmartRoadSense, characterized by weekly updates, this means simulation on an interval spanning from half a month to around half a year). The approximate 95% confidence interval has been computed by taking the 0.025 and 0.975 quantiles of the resulting bootstrap distribution.

Regarding Monte Carlo simulations, for each type of distribution we generated 100 points, took the mean value and propagated it according to the same rule (i.e. mean of current epoch averaged with the updated value y computed at previous epoch) on the same time horizons (i.e. 2, 10, and 25 epochs). The whole process has been repeated for 10^5 trials, leading to

a distribution of values from which the average value and an estimated 95% confidence interval have been computed.

All the experiments have been repeated for 10 runs. Results are represented by the average of the following values: *i)* the mean value at the end of the propagation process (representing the estimate of the measurand); *ii)* the lower bound of the 95% confidence interval; *iii)* the upper bound of the 95% confidence interval. In Figures 3, 4, and 5 we reported histograms providing a comparison of the performance of both methods according to the above mentioned metrics for, respectively, Gaussian, uniform and Rayleigh input distributions. For each figure, histograms denote the mean value estimate, together with error bars encoding the confidence intervals for each simulated epochs horizon (2, 10, 25). As a reference term, we also computed the values (represented as star markers in Figures 3, 4, and 5) that would be obtained for the measurand if no bootstrap were applied, but only a simple composition of the observations were made epoch by epoch.
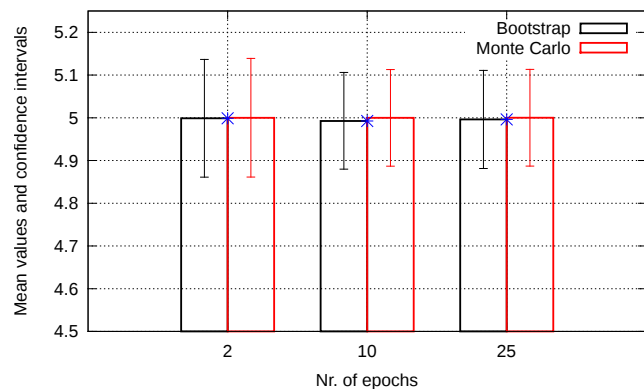

Fig. 3. Synthetic benchmarks: mean values and confidence intervals. Normal distribution ($\mu = 5, \sigma = 1$).
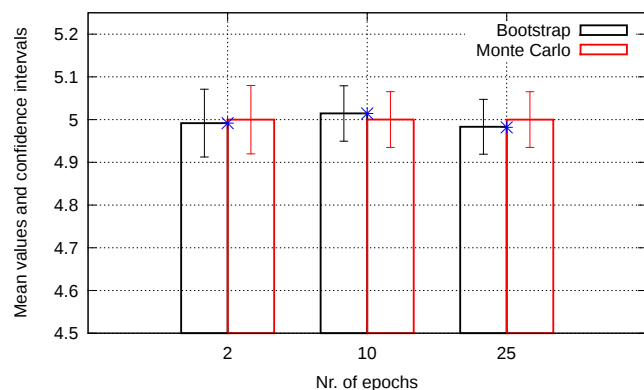

Fig. 4. Synthetic benchmarks: mean values and confidence intervals. Uniform distribution ($a = 4, b = 6$).

Results provide evidence of a very good agreement between the standard Monte Carlo approach and the proposed bootstrap uncertainty propagation method.

In particular, the width of confidence intervals obtained with our method are within a 1.5% deviation from the intervals estimated by means of SMC, with a maximum 0.15% relative
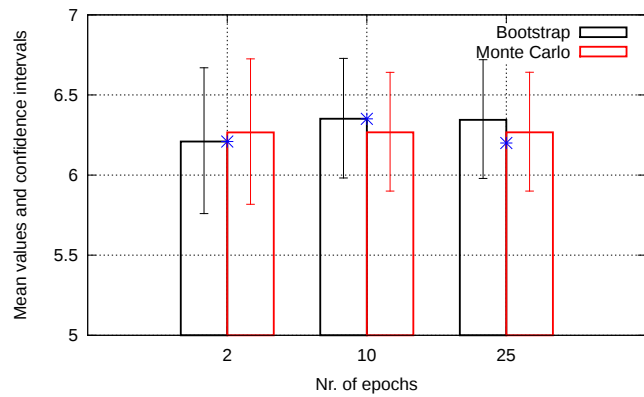
Fig. 5. Synthetic benchmarks: mean values and confidence intervals. Rayleigh distribution ($b = 5$).

error on the value of the lower bound and a 0.13% relative error on the upper bound for the Gaussian input.

In case of uniform input distribution, we obtained confidence intervals whose width differs at most for a 1.7% from that of SMC, while lower bounds of the intervals are within 0.32% from their SMC counterparts, and upper bounds fall within a 0.35% range.

Finally, the analysis of experimental results with input following a Rayleigh distribution showed a 0.81% maximum difference between the confidence interval widths of the proposed approach and those of SMC. The maximum relative error obtained by the proposed approach w.r.t. the SMC method, amounts to 1.38%, for lower limits and 1.31% for upper limits.

As expected, our technique doesn't lead to any improvement in the accuracy of estimates, which clearly depends on the accuracy of the initial observation vector (an inherent feature of resampling techniques). This justifies the differences seen with uniform distribution and, in particular, with the Rayleigh distribution.

Once the accuracy of the bootstrap based uncertainty propagation has been assessed and demonstrated to be consistent with that of Monte Carlo approaches (that assume prior knowledge of statistical distribution of input), we turned our attention to other types of experiments. We indeed analyzed the effect of the number of resampling iterations (N_b) on the system performance, by computing 95% confidence intervals with the proposed algorithm for different values of N_b (namely $N_b = 10^2, 10^3, 10^4, 10^5$) along a time horizon of 10 epochs. Input observations were randomly drawn from a normal distribution ($\mu = 5, \sigma = 1$).The results obtained over 100 runs are plotted in Figures 6a, 6b, and 6c for, respectively, the left bound of the 95% confidence interval, the mean estimate, and the right bound of the 95% confidence interval.

Experiments highlight the variation of the mean estimate and of the confidence intervals as the number of resamplings changes. In particular, albeit not markedly significant (maximum variations are within a 1.7% range), the effect of Monte Carlo random fluctuations across the different runs is clear: higher values of N_b correspond to lower variations across the runs, in accordance to known results in the bootstrap theory [35]. It took on average 36.5s to compute the confidence
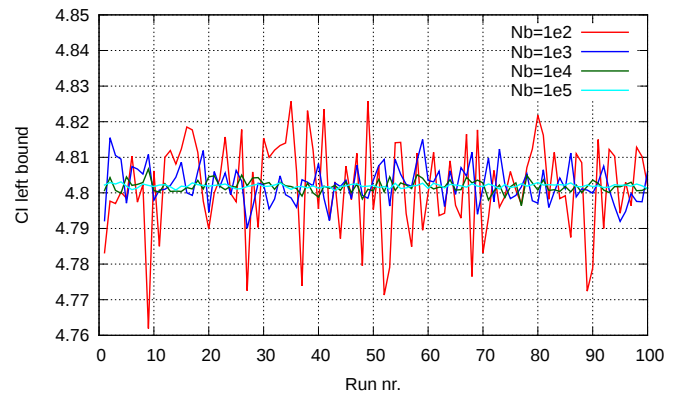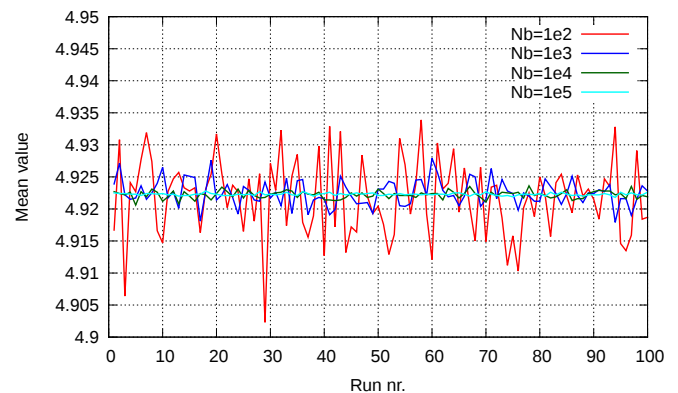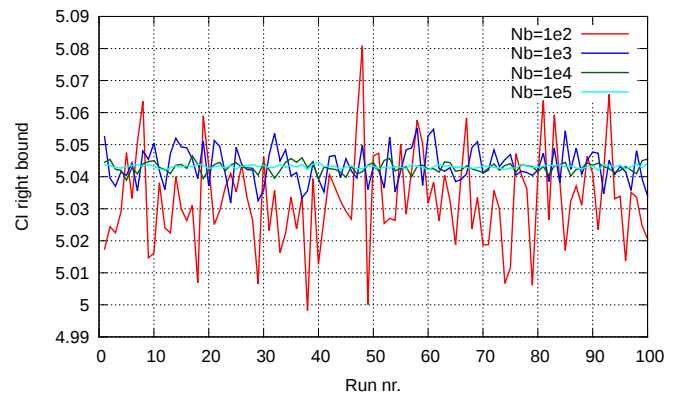
Fig. 6. Sensitivity analysis: left bound of 95% confidence interval (a), estimated mean values (b), right bound of 95% confidence interval (c), for different number of bootstrap resamples across 100 runs. Normal distribution ($\mu = 5, \sigma = 1$).

intervals for a single run when $N_b = 10^5$, and 0.0365s when $N_b = 10^2$. Experiments have been performed on an Intel® i7 CPU, with a 2.80GHz frequency clock and 8GB RAM, running a Matlab® implementation of the bootstrap-based approach.

This empirical evidence confirms the potential for alleviating the computational workload by lowering the number of resampling iterations without severely affecting the accuracy. Conversely, when mitigation of stochastic fluctuations is an issue, N_b should be increased.

The final experiments on synthetic data have been designed to test the proposed approach on a wide time interval during which the value of the measurand is subject to dynamic change. This experimental set up has been conceived to model situations when a potential drift of the physical quantities has to be monitored and tracked by the crowdsensing system. In SRS, for instance, the road surface could progressively deteriorate and, at a given point, could be subject to maintenance. The effect of this possible evolution has been evaluated by simulating a piecewise linear dynamics of the measurand along different time epochs. In particular, input data was generated according to three types of statistical distributions as follows:

- *Gaussian distribution*: observations were randomly generated with a mean value linearly increased at each epoch (from $\mu = 5$, at epoch 0 to $\mu = 7$, at epoch 50) and then set back to $\mu = 5$ for the remaining 50 epochs. Standard deviation was kept constant for the whole simulation ($\sigma = 1$).
- *Uniform distribution*: data was generated by taking values uniformly at random in an interval that was progressively shifted from $[4, 6]$, at epoch 0 to $[6, 8]$, at epoch 50. From epoch 51 to epoch 100 values were drawn again uniformly in the $[4, 6]$ interval.
- *Rayleigh distribution*: input values were taken from a Rayleigh distribution whose scale parameter b was linearly changed from $b = 5$ (at epoch 0) to $b = 7$ (at epoch 50) and then set again to $b = 5$ (from epoch 51 to epoch 100).

These observations have been then used as input for the uncertainty propagation processing pipeline based on the non-parametric bootstrap. Following the previously described experiments, the update at each epoch was performed by taking the average between the measurand estimate at current epoch and the one at the previous epoch.

Plots of the mean value and error bars representing the associated confidence intervals are reported in Figures 7, 8, and 9 to illustrate the results for, respectively, the normal, uniform, and Rayleigh distribution. As expected, the system can effectively cope with a changing input, by dynamically tracking its evolution. Thanks to the proposed approach, the estimates of the measured variables and the corresponding confidence intervals can be also effectively updated.

B. Case study: SmartRoadSense

In order to exemplify the practical applicability of our proposal, we applied the bootstrap-based method to a dataset extracted from the SmartRoadSense project [36]. Data refer to a road segment in Italy composed of 10 monitored points, each one aggregating from 12 to 30 measurements across two adjacent weeks (week 18 and 19, corresponding to the period from May 2, 2016 to May 15, 2016) of the SRS monitoring activity. The main features of the dataset are reported in Table II: with respect to the aggregated point indicated in column 1, we reported in column 2 the week (epoch) the values refer to, in column 3 the number of points aggregated, and in columns 4 and 5 their mean value and standard deviation.
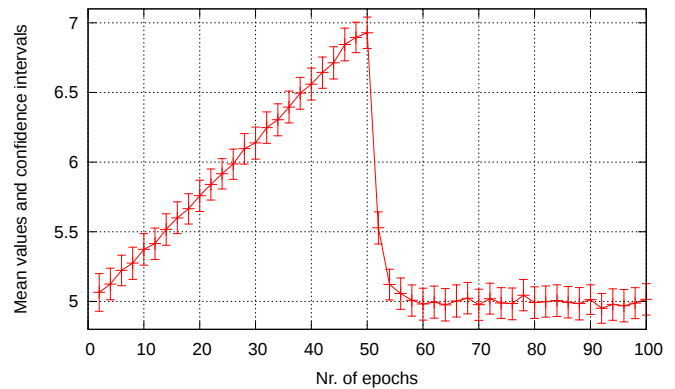

Fig. 7. Dynamic input analysis: mean values and confidence intervals as function of number of simulated updates. Normal distribution. Nr. of bootstrap resamples = 10^3.
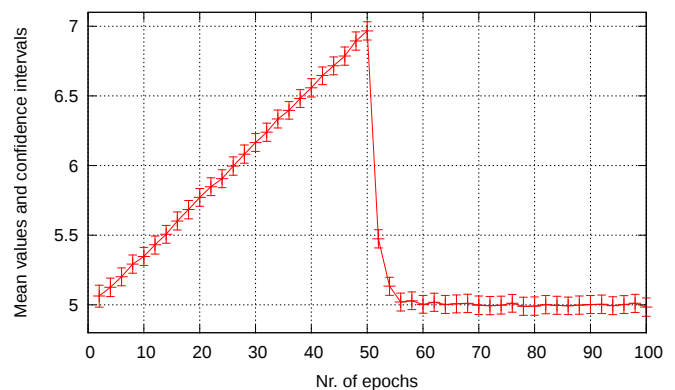

Fig. 8. Dynamic input analysis: mean values and confidence intervals as function of number of simulated updates. Uniform distribution. Nr. of bootstrap resamples = 10^3.
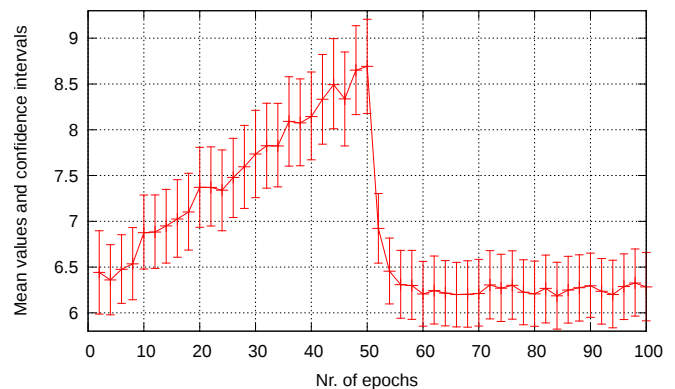

Fig. 9. Dynamic input analysis: mean values and confidence intervals as function of number of simulated updates. Rayleigh distribution. Nr. of bootstrap resamples = 10^3.

For each point of the dataset we applied the bootstrap-based uncertainty propagation method to compute the 95% confidence interval at the end of the period spanning the two weeks of observations. The number of bootstrap resampling iterations was set to $N_b = 10^4$.

Results are reported in Figure 10, plotted as a histogram

TABLE II

SRS DATASET: NUMBER OF RAW POINTS, MEAN VALUE AND STANDARD
DEVIATION FOR EACH AGGREGATED VALUE AND MONITORED WEEK.

Aggregated point	Week	Count	Avg	Stdev
P_1	18	12	0.1587	0.0217
	19	13	0.1318	0.0314
P_2	18	30	0.2278	0.1764
	19	23	0.2017	0.1702
P_3	18	30	0.2023	0.1093
	19	23	0.2421	0.1680
P_4	18	23	0.5260	0.5422
	19	30	0.5055	0.5580
P_5	18	18	0.4093	0.4577
	19	28	0.5954	0.6407
P_6	18	18	0.2312	0.2084
	19	25	0.3863	0.4178
P_7	18	23	1.1319	1.0053
	19	32	0.9361	0.7257
P_8	18	23	1.6216	0.8599
	19	24	1.0610	0.6417
P_9	18	13	1.9805	0.8552
	19	9	1.2424	0.5309
P_{10}	18	10	1.9473	0.8892
	19	10	1.6059	0.5432

with error bars, overlying the piece of map that includes the road under investigation. Each bar is associated to an aggregated point whose roughness index is expressed through a color map (green for low roughness values, red for high roughness levels).
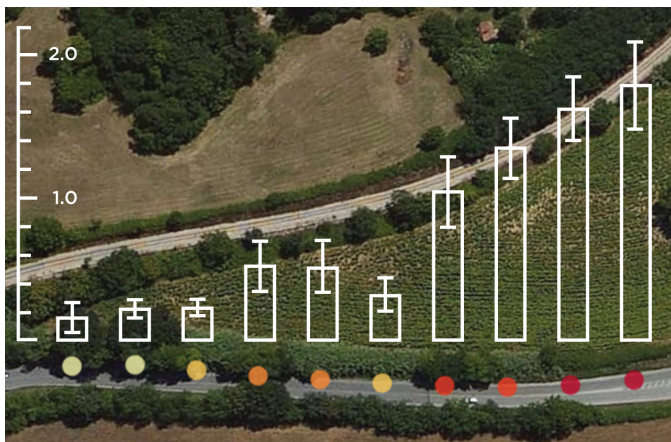

Fig. 10. Case study: mean values and confidence intervals at [2.5% 97.5%] for 10 points taken from a monitored road in SRS. Nr. of bootstrap resamples $= 10^4$.

To provide some further detail about the sensitivity of the method with respect to the number of resampling iterations, we computed 95% confidence intervals for point P_{10} of the SRS dataset with different values of N_b (i.e. $Nb = 10^2, 10^3, 10^4, 10^5$) across a set of 100 runs. Interestingly, the analyzed point clearly represents an example of a small sample size being composed of 10 measurements (in each of the two weeks). Results of this experiment are illustrated in Figures 11a, 11b, and 11c for, respectively, the left bound of the intervals, the mean estimate, and the right bound. Stochastic Monte Carlo variations are, as for the synthetic benchmarks,
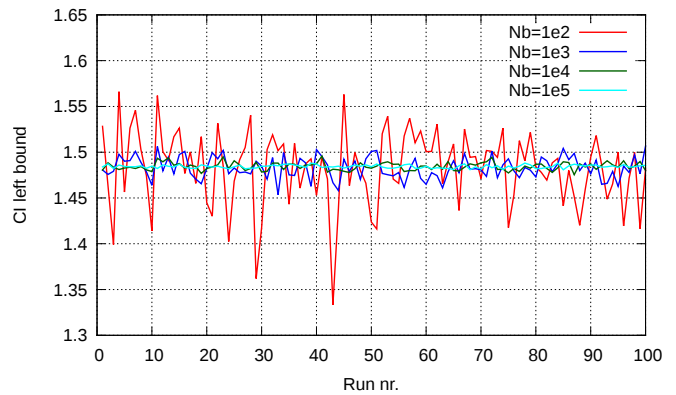
(a)
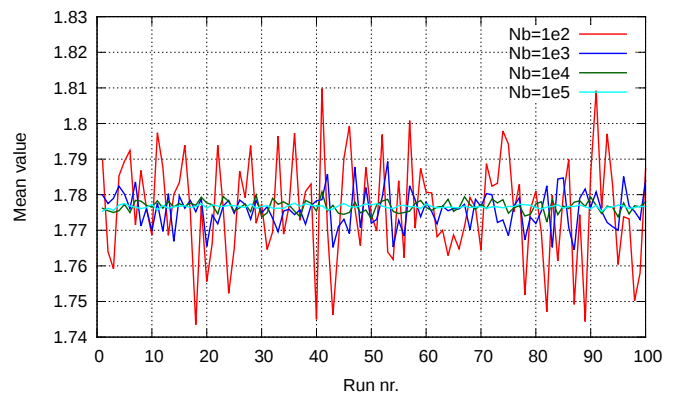
(b)
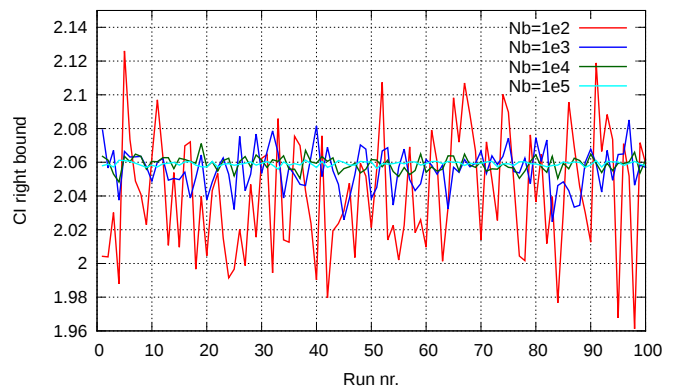
(c)

Fig. 11. Sensitivity analysis: left bound of 95% confidence interval (a), estimated mean values (b), right bound of 95% confidence interval (c), for different number of bootstrap resamples across 100 runs. SRS dataset, point P_{10}.

significantly compressed in a small range when $N_b \geq 10^4$. It is worth noticing a higher variability of the results from run to run for low values of N_b (up to 17.4% for the left bound, $N_b = 10^2$), with respect to the experiments performed on synthetic benchmarks experiments, plausibly because of the effect of the small sample size.

On average, the confidence intervals for a single run were computed in 6.8s, when $N_b = 10^5$. The same task, when $N_b = 10^2$, was completed approximately three orders of magnitude faster, i.e. around 7ms (timing results refer to the

same hardware configuration and implementation used for synthetic benchmarks).

V. Conclusions

Given the growing diffusion of pervasive crowdsensing systems, the possibility of assigning quality indicators to sensed quantities is getting increasing importance. In this paper, we proposed to frame the crowdsensing activity as a distributed measurement process, where multiple end-user devices contribute to the estimate of the quantity to be measured.

The problem of associating a quality value to monitored quantities is therefore cast to evaluating the uncertainty of the estimate within a propagation framework. The complex relationships among involved variables and the difficulty to ascertain their statistical features hinders the possibility of applying either analytic techniques or standard Monte Carlo simulations. We therefore introduced the use of non-parametric bootstrap (a statistical tool that exploits resampling to generate pseudo-observations from input data) to drive the uncertainty propagation process on a purely data-driven basis, without the need to resort to any modeling assumption on the measurement process.

Extensive experimental results demonstrate the effectiveness of the method, in terms of its mathematical consistency, accuracy, and flexibility. Likewise, as a case study, we performed experiments with data from a road surface monitoring crowdsensing platform, the results of which provide a simple, yet significant, demonstration of the potential of the proposed method in terms of its applicability to real-world crowdsensing systems.

Acknowledgment

References

[1] B. Guo, Z. Wang, Z. Yu, Y. Wang, N. Y. Yen, R. Huang, and X. Zhou, "Mobile crowd sensing and computing: The review of an emerging human-powered sensing paradigm," *ACM Computing Surveys (CSUR)*, vol. 48, no. 1, p. 7, 2015.

[2] X. Hu, T. H. S. Chu, H. C. B. Chan, and V. C. M. Leung, "Vita: A Crowdsensing-Oriented Mobile Cyber-Physical System," *IEEE Transactions on Emerging Topics in Computing*, vol. 1, no. 1, pp. 148–165, June 2013.

[3] R. K. Ganti, F. Ye, and H. Lei, "Mobile crowdsensing: current state and future challenges," *Communications Magazine, IEEE*, vol. 49, no. 11, pp. 32–39, 2011.

[4] N. D. Lane *et al.*, "A survey of mobile phone sensing," *Communications Magazine, IEEE*, vol. 48, no. 9, pp. 140–150, 2010.

[5] G. Alessandroni, A. Bogliolo, A. Carini, S. Delpriori, V. Freschi, L. C. Klopfenstein, E. Lattanzi, G. Luchetti, B. Paolini, and A. Seraghiti, "Demo Abstract: Mobile Crowdsensing of Road Surface Roughness," in *MobiSys 2015, Proceedings of the 13th International Conference on Mobile Systems, Applications and Services*. ACM, 2015, pp. 439–439.

[6] G. Alessandroni, L. C. Klopfenstein, S. Delpriori, M. Dromedari, G. Luchetti, B. D. Paolini, A. Seraghiti, E. Lattanzi, V. Freschi, A. Carini, and A. Bogliolo, "SmartRoadSense: Collaborative Road Surface Condition Monitoring," in *UBICOMM 2014, The Eighth International Conference on Mobile Ubiquitous Computing, Systems, Services and Technologies*. IARIA, 2014, pp. 210–215.

[7] S. Delpriori, V. Freschi, E. Lattanzi, and A. Bogliolo, "Efficient Algorithms for Accuracy Improvement in Mobile Crowdsensing Vehicular Applications," in *UBICOMM 2015, The Ninth International Conference on Mobile Ubiquitous Computing, Systems, Services and Technologies*. IARIA, 2015, pp. 145–150.

[8] V. Freschi, S. Delpriori, L. C. Klopfenstein, E. Lattanzi, G. Luchetti, and A. Bogliolo, "Geospatial Data Aggregation and Reduction in Vehicular Sensing Applications: the Case of Road Surface Monitoring," in *Proceedings of 9th IEEE International Conference on Connected Vehicles, ICCVE 2014*. IEEE, 2014, pp. 711–716.

[9] J. Eriksson, L. Girod, B. Hull, R. Newton, S. Madden, and H. Balakrishnan, "The Pothole Patrol: Using a Mobile Sensor Network for Road Surface Monitoring," in *Proc. of the 6th international conference on Mobile systems, applications, and services*. ACM, 2008, pp. 29–39.

[10] A. Thiagarajan *et al.*, "VTrack: Accurate, Energy-aware Road Traffic Delay Estimation Using Mobile Phones," in *Proc. of the 7th ACM Conference on Embedded Networked Sensor Systems*. ACM, 2009, pp. 85–98.

[11] S. Hu, L. Su, H. Liu, H. Wang, and T. F. Abdelzaher, "Smartroad: Smartphone-based crowd sensing for traffic regulator detection and identification," *ACM Trans. Sen. Netw.*, vol. 11, no. 4, pp. 55:1–55:27, Jul. 2015.

[12] S. Nawaz, C. Efstratiou, and C. Mascolo, "Smart Sensing Systems for the Daily Drive," *IEEE Pervasive Computing*, vol. 15, no. 1, pp. 39–43, Jan 2016.

[13] ——, "ParkSense: A Smartphone Based Sensing System for On-street Parking," in *Proceedings of the 19th Annual International Conference on Mobile Computing & Networking*, ser. MobiCom '13. New York, NY, USA: ACM, 2013, pp. 75–86.

[14] S. Mathur, T. Jin, N. Kasturirangan, J. Chandrasekaran, W. Xue, M. Gruteser, and W. Trappe, "ParkNet: Drive-by Sensing of Road-side Parking Statistics," in *Proceedings of the 8th International Conference on Mobile Systems, Applications, and Services, MobiSys '10*. New York, NY, USA: ACM, 2010, pp. 123–136.

[15] D. Hasenfratz, O. Saukh, C. Walser, C. Hueglin, M. Fierz, T. Arn, J. Beutel, and L. Thiele, "Deriving high-resolution urban air pollution maps using mobile sensor nodes," *Pervasive and Mobile Computing*, vol. 16, pp. 268–285, Dec 2014.

[16] B. Predic, Z. Yan, J. Eberle, D. Stojanovic, and K. Aberer, "ExposureSense: Integrating daily activities with air quality using mobile participatory sensing," in *Pervasive Computing and Communications Workshops (PERCOM Workshops), 2013 IEEE International Conference on*, March 2013, pp. 303–305.

[17] K. Aberer *et al.*, "OpenSense: open community driven sensing of environment," in *Proc. of the 1st International Workshop on GeoStreaming (IWGS '10)*, 2010, pp. 39–42.

[18] L. Ruge, B. Altakrouri, and A. Schrader, "SoundOfTheCity-Continuous noise monitoring for a healthy city," in *Pervasive Computing and Communications Workshops (PERCOM Workshops), 2013 IEEE International Conference on*. IEEE, 2013, pp. 670–675.

[19] R. K. Rana, C. T. Chou, S. S. Kanhere, N. Bulusu, and W. Hu, "Earphone: An End-to-end Participatory Urban Noise Mapping System," in *Proceedings of the 9th ACM/IEEE International Conference on Information Processing in Sensor Networks, IPSN '10*. New York, NY, USA: ACM, 2010, pp. 105–116.

[20] E. Miluzzo, C. T. Cornelius, A. Ramaswamy, T. Choudhury, Z. Liu, and A. T. Campbell, "Darwin Phones: The Evolution of Sensing and Inference on Mobile Phones," in *Proceedings of the 8th International Conference on Mobile Systems, Applications, and Services, MobiSys'10*. New York, NY, USA: ACM, 2010, pp. 5–20.

[21] K. L. Huang, S. S. Kanhere, and W. Hu, "Are you contributing trustworthy data?: the case for a reputation system in participatory sensing," in *Proceedings of the 13th ACM international conference on Modeling, analysis, and simulation of wireless and mobile systems*. ACM, 2010, pp. 14–22.

[22] ——, "On the need for a reputation system in mobile phone based sensing," *Ad Hoc Networks*, vol. 12, pp. 130–149, 2014.

[23] J. Taylor, *Introduction to error analysis, the study of uncertainties in physical measurements*. University Science Books, 1997.

[24] E. Asmare and J. A. McCann, "Lightweight Sensing Uncertainty Metric - Incorporating Accuracy and Trust," *IEEE Sensors Journal*, vol. 14, no. 12, pp. 4264–4272, Dec 2014.

[25] M. Cox, P. Harris, and B.-L. Siebert, "Evaluation of measurement uncertainty based on the propagation of distributions using Monte Carlo simulation," *Measurement Techniques*, vol. 46, no. 9, pp. 824–833, 2003.

[26] M. Cox, "Propagation of distributions by a Monte Carlo method, with an application to ratio models," *The European Physical Journal Special Topics*, vol. 172, no. 1, pp. 153–162, 2009.

[27] Joint Committee for Guides in Metrology, "Evaluation of measurement data - Guide to the expression of uncertainty in measurement (GUM)," 2008.

[28] J. A. Rice, *Mathematical Statistics and Data Analysis*. Duxbury Press, 1995.

[29] A. Kulesa, M. Krzywinski, P. Blainey, and N. Altman, "Points of significance: sampling distributions and the bootstrap," *Nature methods*, vol. 12, no. 6, pp. 477–478, 2015.

[30] B. Efron, "Bootstrap methods: another look at the jackknife," *Annals of Statistics*, vol. 7, pp. 1–26, 1979.

[31] B. Efron and R. Tibshirani, *An introduction to the bootstrap*. Chapman & Hall;, 1993.

[32] R. E. Kass, U. T. Eden, and E. N. Brown, *Analysis of Neural Data*. New York, NY: Springer New York, 2014, ch. Propagation of Uncertainty and the Bootstrap, pp. 221–246.

[33] S. Lee, C. H. Park, and J. H. Chang, "Improved Gaussian Mixture Regression Based on Pseudo Feature Generation using Bootstrap in Blood Pressure Estimation," *IEEE Transactions on Industrial Informatics*, vol. PP, no. 99, pp. 1–1, 2015.

[34] D. Zhao, H. Ma, L. Liu, and X.-Y. Li, "Opportunistic coverage for urban vehicular sensing," *Computer Communications*, vol. 60, pp. 71–85, 2015.

[35] T. C. Hesterberg, "What teachers should know about the bootstrap: Resampling in the undergraduate statistics curriculum," *The American Statistician*, vol. 69, no. 4, pp. 371–386, 2015.

[36] SmartRoadSense, "http://smartroadsense.it/," 2016. [Online]. Available: http://smartroadsense.it/