iMAP: Indirect Measurement of Air Pollution with Cellphones

Murat Demirbas
Computer Sci. & Eng. Dept
University at Buffalo, SUNY

Carole Rudra
Social & Preventive Medicine Dept
University at Buffalo, SUNY

Atri Rudra   Murat Ali Bayir
Computer Sci. & Eng. Dept
University at Buffalo, SUNY

***Abstract*—In this paper, we introduce the cellphone-based indirect sensing problem. While participatory sensing aims at *monitoring of a phenomenon by deploying a dense set of sensors carried by individuals*, our indirect sensing problem aims at *inferring the manifestations of a sparsely monitored phenomenon on the individuals*. The main advantage of the indirect sensing method is that, by making use of existing exposure modeling and estimation methods, it provides a more feasible alternative to direct sensing. Collection of time-location logs using the cellphones plays a major role in our indirect sensing method, while direct sensing at the cellphones is unneeded. We focus on the air pollutant exposure estimation problem as an application of the indirect sensing technique and propose a web-based framework, iMAP, for addressing this problem. We also discuss the information quality (IQ) requirements of indirect sensing in the iMAP framework.**

I. Introduction

Wireless sensor networks (WSNs) [4], [12] are inadequate for large-scale sensing tasks due to the limitations of communication reach, battery life, and the static deployment of the sensors. To overcome these handicaps against scaling, *cellphone-based participatory sensing* has been proposed to enable public and professional users to share their local knowledge [1], [10], [11], [15], [20]. In this approach, sensors on the cellphones gather data, and these data are collected to form a fine-granularity monitoring of a phenomenon in a large-scale area. Example applications include urban planning, public health, cultural identity, creative expression and natural resource management.

In this paper we introduce the cellphone-based indirect sensing problem: a dual and complementary problem of the one that participatory sensing addresses. While participatory sensing aims at *monitoring of a phenomenon by deploying a dense set of sensors carried by individuals*, our indirect sensing problem aims at *inferring the manifestations of a sparsely monitored phenomenon on the individuals*. In other words, the goal of participatory sensing is to enable the fitting of a model on the phenomenon by using the sensor values collected from the cellphones, whereas indirect sensing deals with the converse process: leveraging on an existing model to predict the values individuals have been subject to. For the indirect sensing problem, the time-location logs kept by the cellphones play a major role, while direct sensing at the cellphones is not necessary.

There are several applications of the indirect sensing problem. To keep our discussion concrete, we focus on the air pollutant exposure estimation as a running example. Here the goal is to accurately determine the impact of air pollution on an individual's health. This problem is difficult because many air pollutant concentrations, particularly those related to vehicular traffic, vary as much within cities as they do between cities [7]. To accurately estimate individual air pollutant exposures when studying their health effects, this intraurban variation needs to be taken into account.

The main advantage of the indirect sensing method is that, by making use of existing exposure modeling and estimation methods, it provides a more feasible alternative to direct sensing. The feasibility of the direct sensing approach is limited by the cost, size, and bulk of the different sensors required. For the air pollutant exposure monitoring application, the sensors for $PM_{2.5}$ (fine particulate matter, one of the most hazardous/critical pollutant type for human health [2]) are at least 5lbs in weight and around $5000 in cost. In contrast, indirect sensing is feasible for the air pollutant exposure application due to the prevalence of exposure modeling and estimation methods. In US cities, the regional air pollution monitoring agency provides streaming pollutant information data from several monitoring sites employing $PM_{2.5}$ sensors. In many urban areas, it is possible to use these data to develop an interurban prediction model that allows estimation of location-specific air pollutant exposures. Further, using the time-location logs from a cellphone, it is possible to return an accurate estimation of the air pollutant exposure for the individual.

Indirect sensing decouples the accurate determination of the effects of a phenomenon on an individual (which is done by using the model and the time-location logs) from the construction of the model (which is done by collecting data via direct sensing at several locations). As a byproduct, indirect sensing eliminates some privacy concerns. The user may choose not to disclose her time-location data, and instead download the model and perform the indirect sensing on her cellphone or PC. This being said, we note that indirect sensing can also be participatory. Several applications are made possible just by publishing time-location data, including urban planning, epidemic monitoring, and social networking applications. In fact, in our proposed iMAP framework below, we consider publishing of time-location data to a secure central web-site for processing.

Contributions. We introduce the indirect sensing problem

978-1-4244-3304-9/09/$25.00 ©2009 IEEE

for cellphone-based sensing. This problem focuses on accurately estimating the value of a modeled phenomenon on an individual using the cellphone-captured time-location logs of the individual. The advantage of this method is to decouple the construction of the model (which can be performed using sparser sensing) from the accurate estimation of the effects of the model on the individual.

We illustrate the indirect sensing method using the air pollutant exposure application as our case study. To this end, we identify "Land Use Regression" (LUR) [19] as a suitable modeling solution. LUR uses geographic information system (GIS) software to measure local traffic, population, and weather characteristics measured at regional air monitoring stations. Next, linear regression models are designed to quantify the relation between the set of these environmental characteristics and the air pollutant concentrations measured at each site. To use the model, the same environmental characteristics are measured at individuals' locations (their geocoded residential addresses) and entered into the model as predictor variables. This results in estimated pollutant concentrations for each user.

We improve over the state-of-the-art for LUR using cellphone-based time-activity data logs. Our method allows us to estimate exposures at multiple locations per person rather than only at the residence. Therefore, this method gives a more accurate determination of the air-pollutant exposure of an individual and solves the time-activity monitoring barrier facing the public health specialists working on the air pollutant exposure problem.

To implement our indirect sensing method, we propose a web-based framework, iMAP, for indirect measurement of air pollution exposure with cellphones. In our proposed framework, cellphones collect time-activity logs for their users and transmit this information to a website opportunistically when an available wireless access point is encountered, or through general packet radio service (GPRS) as a fallback mechanism. The website also pulls GIS information from other websites and calculates individual exposures using the time-location logs. We outline our plans for an air pollutant sensing deployment of iMAP using 50 smartphones, such as iPhones and G1(Google) phones, that have *built-in* GPS modules and wi-fi interface.

The indirect sensing method raises new questions about the information quality (IQ) of the sensing. Since indirect sensing decouples the model from the estimation of the effects of the model on the individual, IQ analysis of the indirect sensing method also reduces to the IQ analysis of the model and that of the time-location logs. We investigate these two questions in the context of the iMAP framework. To address the first question, we consider model validation as a solution. As for the second question, we consider less frequent sampling of the geocoordinates provided that we can meet the IQ constraints of the air-pollution exposure estimation application.

Outline. After introducing the air pollutant exposure estimation problem in Section II, we discuss the LUR method in Section III and present model validation approaches for IQ analysis. In Section IV, we present our proposed web-based framework. We also discuss the IQ analysis of the timelocation logs in the iMAP framework in Section IV. We present further discussion on the applications of indirect sensing in Section V and conclude our paper in Section VI.

II. Air pollutant exposure estimation

Two million premature deaths annually are attributable to air pollutants [9]. Even at lower levels typical of more developed countries, air pollution adversely impacts health across the lifespan. Acute and chronic air pollutant exposures increase risks of cardiovascular and respiratory diseases [8], exacerbate, and perhaps cause, asthma among children [21], and increase risks of neonatal death, low birthweight, and preterm delivery [21], [23].

Estimating air pollutant exposures is difficult for the following reasons. Air pollution is usually highest in urban areas. Many air pollutant concentrations, particularly those related to vehicular traffic, vary as much within cities as they do between cities. To accurately estimate individual air pollutant exposures when studying their health effects, this intraurban variation needs to be taken into account. Since the feasibility of directly measuring individual air pollutant exposures is limited by the cost, size, and bulk of most monitoring equipment, researchers have focused on methods of estimating individual, intraurban air pollutant exposures using modeling. These methods are motivated by the fact that many air pollutant concentrations are strongly related to nearby environmental characteristics such as vehicular traffic, land use, and elevation. By measuring such characteristics, it would be possible to predict the nearby air pollutant concentration with reasonable accuracy.

The modeling approaches often use the residential address of an individual to estimate air pollutant exposures for the individual [3]. Investigators have attempted to incorporate time-activity data into air pollutant estimation procedures by interviewing study participants regarding their travel schedules [22], filming children to estimate their exposures to indoor sources of pollution (cooking fires) [5], and modeling time-activity patterns in GIS using self-reported travel characteristics [14]. These methods are too costly and time-consuming to apply to large populations. In Section IV, we discuss how we overcome this limitation of modeling approaches using cellphone-based time-activity collection.

Recently, there has been work on using cellphones for air pollution monitoring as part of the participatory urbanism project [18]. This work deployed a simple SMS system that allows one to send a text message containing a zipcode to the system in order to receive the current air quality data for that zipcode. Although this is a very useful service, it has intrinsic limitations because it needs to be user initiated and the granularity of air quality is very low (at the zipcode level). In contrast, our framework aims to provide continuous and passive monitoring of air pollutant exposure of an individual at a much finer granularity.

III. Land use regression

Land use regression (LUR) is a modification of a method developed by Brauer et al [6] to estimate air pollutant exposures in epidemiologic studies. Briefly, the method uses geographic information system (GIS) software to measure local traffic, population, and weather characteristics measured at regional air monitoring stations and then develops linear regression models to estimate the relations between the set of these characteristics and each air pollutant concentration. The same characteristics measured at residential addresses are then entered into the developed models as predictor variables, resulting in estimated pollutant concentrations for each individual.

Specification of models. With each pollutant concentration as the dependent (outcome) variable, LUR fits a multivariable linear regression model. Briefly, this model is constructed so that any variable that is moderately predictive of the PM2.5 concentration is included in the final model. This example shows a model designed to estimate monthly average $PM_{2.5}$ at a specific location.

$$E(Y_{ijk}) = \alpha + \beta_1 * (streetdensity)_{ij} + \beta_2 * (max.temp.)_{ijk} + \beta_6 * (popn.density)_{ij} + \beta_3 * (year)_j + \beta_4 * (month)_{ijk}$$

According to this model, the average monthly $PM_{2.5}$ concentration at a specific point Y_{ijk}, is a function of the following environmental characteristics, where i = location, j = year, k = month

1) street density: the density (m/km^2) of surrounding streets, usually measured within a circular area of 250-1000 m surrounding the point at which $PM_{2.5}$ is measured. Street density is a surrogate for traffic volume, which contributes to local $PM_{2.5}$ concentrations.
2) max. temp.: The average monthly high temperature measured at a nearby weather monitoring station. This variable captures the fact that $PM_{2.5}$ concentrations fluctuate according to weather patterns.
3) popn. density: the density (persons/km^2) living in the surrounding census tract, according to US Census measurements. Population density also influences traffic volume, and thus $PM_{2.5}$ concentration.
4) year: This variable captures time trends in $PM_{2.5}$ concentration.
5) month: This variable captures seasonal changes in $PM_{2.5}$ concentration.

This model is fit at local $PM_{2.5}$ monitoring sites, where both $PM_{2.5}$ and the environmental characteristics can be measured. β coefficients are estimated for each of the environmental characteristics. Then, by plugging into the equation the values for the individual's local environmental characteristics (which can be feasibly measured), individual $PM_{2.5}$ concentrations (which cannot be feasibly measured) can be estimated. The novelty of our proposed iMAP framework (discussed in Section IV) is that it will allow multiple $PM_{2.5}$ estimates per person according to their travel patterns, rather than one estimation based solely on their residential location.

Estimation of model validity. As we mentioned in the Introduction, the IQ analysis of the indirect sensing method reduces to that of the model and the timelocation log sampling. In order to perform an IQ analysis of the model, we can measure the accuracy of the LUR model using the following two methods. First, accuracy is measured by using the model to estimate $PM_{2.5}$ concentrations at the $PM_{2.5}$ monitoring sites. An accurate model will show little difference between these estimates and the measured $PM_{2.5}$ concentrations at the monitoring sites. Identification of specific locations with poor fit will allow us to reevaluate the model components and design, if necessary. Second, the best test of model accuracy is its ability to predict concentrations at an independent set of $PM_{2.5}$ monitoring sites: that is, sites that were not used to fit the model. To this end, we will use the regression model estimated from all but one of the $PM_{2.5}$ monitoring sites to predict the remaining site's monthly concentrations, for each pollutant separately. This procedure (the $N-1$ method) is conducted for each of the monitoring sites and the results compared using a correlation coefficient. We plan to use both approaches to evaluate and quantify the fit of our model.

IV. A web-based framework for iMAP

Land use regression is cost-efficient and fairly easy to implement. Its main limitation is that it is used only to estimate exposures at an individual's residence. The resulting exposure estimate is inaccurate because it does not take into account a person's time-activity patterns, i.e. where he or she travels over time, to examine the health effects of air pollution.

Here we propose a novel method to address this fundamental limitation of land use regression. GPS-equipped cellphones can provide a valuable source of time-activity information with low burden to individuals and low cost to researchers. We will periodically record the cellphone's location (and therefore, typically, an individual's location) and use the resulting geocoordinates (latitude and longitude) to measure an individual's time-activity patterns. Then, instead of relying on a person's residence to estimate his or her air pollutant exposures, we will estimate exposures based on her changing locations throughout the time period. A particularly important benefit of our approach is the ability to capture locations during commuting and traveling, when air pollutant exposures are usually highest [3].

Our goal is to demonstrate the feasibility of incorporating time-activity data measured with GPS-equipped cellphones into a land use regression model designed to estimate individual exposures to $PM_{2.5}$. Henceforth, we use the term cellphone-based estimates to describe the exposure estimates created from such a model. By using cellphone-based rather than residence-based estimates, we can greatly improve the accuracy with which we are able to estimate $PM_{2.5}$ exposures in epidemiologic and clinical studies of the health effects of air pollution.

In this proposed iMAP (indirect measurement of air pollution) framework, we aim to do the following three tasks:

1) to collect time-activity data from local volunteers using GPS-equipped cellphones.
2) to incorporate these data into a land use regression model to produce cellphone-based $PM_{2.5}$ exposure estimates, and compare these estimates to residential-based $PM_{2.5}$ exposure estimates.
3) to design algorithms to determine the optimal efficiency of this process so that it will be feasible to use in large study populations.

Task 1: time-activity data collection. We will recruit 50 volunteers for a three-month observational study. Eligible volunteers will be Buffalo, NY community members who are willing to provide their residential address and allow us to intermittently collect geocoordinates from their cellphones throughout a three-month period. Volunteers can choose to use a cellphone that we will loan to them or, if they own a cellphone that meets study requirements, use their own for the study. Those who borrow a cellphone from us will be able to use their own subscriber identity module (SIM) cards and thus will be able to retain their current phone numbers and network providers during the study.

We will use smart-phones for our study. The built-in GPS system will enable us to record the geocoordinates of the phone intermittently (for instance, at every 10 minutes) with high precision. These wi-fi capable smart-phones can automatically connect to wireless access points in residential or public areas. We will use this feature to daily transmit encrypted geocoordinate data to our central processing computer (henceforth called the webportal) without any cost to the study participants or involvement of the cellphone service provider. If there is no such wi-fi connection opportunity for more than 24 hours, we will program the smart-phone to transmit encrypted geocoordinate data to the webportal by using its general packet radio service (GPRS) capability as a fallback. Because these smart-phones are equipped with Microsoft Windows Mobile 6 operating system, the time and effort associated with software development for the cellphones will be significantly reduced. Using the Mobile 6.0 SDK and embedded Visual C++ 4.0 platforms, we will be able test our applications in a PC environment before deploying them in the field.

The geocoordinates collected from the cellphones will be stored in the webportal in a suitable format for future analysis. We will also design our webportal to automatically retrieve publicly available data from government websites measuring local current $PM_{2.5}$ concentrations and vehicular traffic. These data will be used in building and implementing exposure models (see below). The final product of this first task will be a dataset of geocoordinates that will be used in land use regression models.

Task 2: cellphone-and residential-based $PM_{2.5}$ exposure estimation. We will construct and employ land use regression models to estimate individual $PM_{2.5}$ exposures within the Buffalo, NY region. First, using geographic information system (GIS) software, we will collect and map $PM_{2.5}$ measurements taken hourly at eight monitoring sites administered by the regional air pollution monitoring agency. We will also collect and map data on regional environmental characteristics such as annual traffic volume, real-time traffic conditions, population and street density, land use, and altitude. All of these data are publicly available at no cost from governmental web sites, such as N.Y. State Department of Environmental Conservation website, N.Y. State Department of Transportation website, N.Y. State Transportation Federation, and U.S. Census Bureau Geography Division website. Next, we will measure these environmental characteristics at each of the $PM_{2.5}$ monitoring sites using GIS. We will then choose the set of characteristics that best predicts the $PM_{2.5}$ concentrations measured at the monitoring sites. These procedures will be based on stepwise least-squares linear regression and will be performed using statistical software. Finally, using statistical software, we will apply the land use regression models to estimate $PM_{2.5}$ exposures among the study volunteers.

We will produce residence-based $PM_{2.5}$ estimates by fitting the models with environmental characteristics measured at each participant's residence. We will then produce cellphone-based $PM_{2.5}$ estimates by fitting multiple models per person, each using environmental characteristics measured at a location reported by her cellphone. As a baseline, we will measure a person's location once every ten minutes over the three month enrollment period. This amounts to approximately 13,000 $PM_{2.5}$ exposure estimates per individual. We will daily average these estimated exposures to produce a time-activity integrated $PM_{2.5}$ exposure estimate, i.e. our cellphone-based estimate. We will also be able to derive exposure estimates during more specific intervals, for instance during weekday rush hours or weekdays and weekends separately.

Task 3: efficiency improvement of our $PM_{2.5}$ estimation procedures. Here we consider IQ analysis of timelocation log sampling. To use this method in future larger studies, we must develop an efficient automated process to collect and integrate the data while still meeting the minimum IQ constraints of the air-pollution exposure estimation application. For instance, if we aim to estimate average 90-day exposures among 50 people using geocoordinates sampled every 10 minutes, we must calculate 648,000 exposure estimates. A study population of 5,000 (typical for this field) would require 64,800,000 estimates. We will improve the efficiency of our method in several ways. For example, we will design an algorithm to calculate exposure estimates on a "need to know" basis. The webportal will store each person's last location and the corresponding exposure estimate. If the current location is the same or very near to the last location, then we need not re-estimate the exposure but can use the previously stored estimate. We will explore the ways in which our assumptions (such as proximity of locations) impact our exposure estimates.

Another way to improve efficiency is to less frequently collect geocoordinates from the cellphone while it is less mobile. This method would allow us to sample and record location data more frequently when a person is commuting than when she is at home each night, for instance. We model this idea in terms of the following algorithmic problem. Let each measurement correspond to a point in some high dimensional space. In the case of iMAP, each point can lie in a three dimensional space where the latitude, longitude and time form the three dimensions. Given an input parameter D, the problem is to partition the points into as few clusters as possible such that any two points in each cluster have an Euclidean distance of at most D. This parameter D will be determined by the application: for example, in iMAP this will be the measure of the mobility of the cellphones. Given such a clustering, the cellphone needs to report just one measurement from each cluster (by a suitable choice of D, all the points in a cluster will be "equivalent").

We also add two key constraints that iMAP imposes to this clustering problem formulation. First, we would like to cluster a measurement as soon as we receive it (in other words, the algorithm can make only one "pass" at the data). Second, the cellphone has limited resources so the algorithm should work with limited memory and time constraints [1]. These two constraints are *exactly* what are addressed by *data stream algorithms* [17]. We plan to use techniques from data stream algorithms to solve this clustering problem.

We note that in general the minimum IQ requirements of the timelocation log sampling is entwined with that of the model. Thus, given the model, it is possible to further relax the granularity of the timelocation log sampling in places where the model estimates little variation.

V. Discussion

Here we consider other applications and extensions of indirect sensing with cellphones.

Our indirect sensing method is also well suited to epidemic monitoring and control applications. In this case, the model tracks the dissemination patterns of an illness approximately. People who have become infected may choose to disclose their time-activity logs to a centralized secure/anonymized webportal. Users may occasionally synchronize their time-activity logs with the webportal to check if they are under risk for any illnesses. This technique could especially help with monitoring and early diagnosis of sexually transmitted diseases as it respects the privacy of the users. For this application, bluetooth-based connectivity logs can also be maintained at the cellphones in order to improve the accuracy of the modeling and prediction.

A suitable high-level abstraction for time-activity logs is that of a time-activity *trace*. Here the idea is to mine the common *subsequences* from a user's time-activity log, label them as a trace data-structure, and represent the time-activity log as a concatenation of a small number of traces. Some examples of traces may include home-to-work, work-to-home, home-to-markets, Saturday-night, Sunday night, etc. It is well known that humans exhibit a fair repetition of traces in their daily, weekly, and monthly travel patterns [13]. These higher level trace abstractions (calculated via data-mining without any need for user input) may simplify calculation of LUR for sequences of time-activity inputs. If such a sequence corresponds to a trace with minor differences, the pre-computed LUR value for the trace can be used with minor adjustments. The trace concept may also be useful for proactive and selective notification of threats in an area to the users whose likely traces in the near future intersects the area.

[1]This is especially important as we would like our client software on the cellphones to use up as little resources so as not to interfere with the primary tasks of the cellphone.

VI. Concluding remarks

We introduced the indirect sensing problem for cellphone-based sensing which focuses on accurately estimating the value of a modeled phenomenon at the level of an individual. We illustrated the indirect sensing method using the air pollutant exposure application as our case study. The advantage of this method is to divorce the construction of the model (which can be performed using sparser sensing) from the accurate estimation of the effects of the model on the individual. We outlined our proposal for a web-based framework for indirect sensing. Our iMAP framework employs LUR and improves on the current-state-of-the-art by using the cellphones to passively collect time-location logs for the users and refine the estimated risk factor accordingly.

We plan to develop and deploy our iMAP framework using 50 smartphones in the Buffalo, NY area. Our future work will include investigation of alternative modeling techniques for the air pollution exposure problem. We will consider the Kriging [16] method as an alternative to LUR and compare and contrast their characteristics using iMAP. We will also investigate efficient solutions to the online clustering problem we formulated in Section IV in order to enable adaptive sampling that optimizes both the granularity and efficiency of the time-activity logging process.

References

[1] T. Abdelzaher, Y. Anokwa, P. Boda, J. Burke, D. Estrin, L. Guibas, A. Kansal, S. Madden, and J. Reich. Mobiscopes for human spaces. volume 6, April - June 2007.

[2] P. Bruckmann A.D. Kappos, T. Eikmann, and et al. Health effects of particles in ambient air. *Int. J. Hyg. Environ. Health*, 207(4):399–407, 2004.

[3] S. D. Adar and J. D. Kaufman. Cardiovascular disease and air pollutants: evaluating and improving epidemiological data implicating traffic exposure. *Inhal. Toxicol.*, 19(1):135–149, 2007.

[4] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci. A survey on sensor networks. *IEEE Communications Magazine*, 2002.

[5] B. Barnes, A. Mathee, and K. Moiloa. Assessing child time-activity patterns in relation to indoor cooking fires in developing countries: a methodological comparison. *Int. J. Hyg.. Environ. Health*, 208(3):219–225, 2005.

[6] M. Brauer, G. Hoek, P. van Vliet, K. Meliefste, P. Fischer, U. Gehring, J. Heinrich, J. Cyrys, T. Bellander, M. Lewne, and B. Brunekreef. Estimating long-term average particulate air pollution concentrations: application of traffic indicators and geographic information systems. *Epidemiol.*, 14:228–239, 2003.

[7] D. Briggs. Exposure assessment. In P. Elliot, editor, *Spatial epidemiology: methods and applications*. Oxford University Press, 2000.

[8] R. D. Brook. Is air pollution a cause of cardiovascular disease? updated review and controversies. *Rev. Environ. Health*, 22(2):115–137, 2007.

[9] G. H. Brundtland. Reducing risks to health, promoting healthy life. *JAMA*, 288(16):1974, 2002. From the World Health Organization.

[10] J. Burke, D. Estrin, M. Hansen, A. Parker, N. Ramanathan, S. Reddy, and M. B. Srivastava. Participatory sensing. 2006.

[11] A. T. Campbell, S. B. Eisenman, N. D. Lane, E. Miluzzo, and R. A. Peterson. People-centric urban sensing. In *WICON '06: Proceedings of the 2nd annual international workshop on Wireless internet*, page 18, 2006.
[12] D. Estrin, R. Govindan, J. S. Heidemann, and S. Kumar. Next century challenges: Scalable coordination in sensor networks. In *Mobile Computing and Networking*, pages 263–270, 1999.
[13] J. Ghosh, S. J. Philip, and C. Qiao. Sociological orbit aware location approximation and routing (solar) in manet. *Ad Hoc Networks*, 5(2):189–209, 2007.
[14] J. Gulliver and D.J. Briggs. Time-space modeling of journey-time exposure to traffic-related air pollution using gis. *Environ. Res.*, 97(1):10–25, 2005.
[15] A. Kansal, M. Goraczko, and F. Zhao. Building a sensor network of mobile phones. *IPSN*, pages 547–548, 2007.
[16] G. Matheron. Principles of geostatistics. *Economic Geology*, (58):1246–1266, 1963.
[17] S. Muthukrishnan. Data streams: Algorithms and applications. *Foundations and Trends in Theoretical Computer Science*, 1(2), 2006.
[18] E. Paulos, R. J. Honicky, and E. Goodman. Sensing atmosphere. *Workshop position paper for the Sensing on Everyday Mobile Phones in Support of Participatory Research at ACM SenSys 2008*, 2007.
[19] P. H. Ryan and G. K. LeMasters. A review of land-use regression models for characterizing intraurban air pollution exposure. *Inhal. Toxicol.*, 19(1):127–133, 2007.
[20] A. Santanche, S. Nath, J. Liu, B. Priyantha, and F. Zhao. Senseweb: Browsing the physical world in real time. *IPSN*, (377–378), 2006.
[21] J. A. Sarnat and F. Holguin. Asthma and air quality. *Curr. Opin. Pulm. Med.*, 13(1):63–66, 2007.
[22] K. Sexton, S.J. Mongin, J.L Adgate, and et al. Estimating volatile organic compound concentrations in selected microenvironments using time-activity and personal exposure data. *J. Toxicol. Environ. Health A.*, 70(5):465–476, 2007.
[23] N. Sorensen, K. Murata, E. Budtz-Jorgensen, P. Weihe, and P. Grandjean. Prenatal methylmercury exposure as a cardiovascular risk factor at seven years of age. *Epidemiology*, 10(4):370–375, 1999.