Computational Sustainability and Artificial Intelligence in the Developing World

John Quinn, Vanessa Frias-Martinez, Lakshminarayan Subramanian

■ *The developing regions of the world contain most of the human population and the planet's natural resources, and hence are particularly important to the study of sustainability. Despite some difficult problems in such places, a period of enormous technology-driven change has created new opportunities to address poor management of resources and improve human well-being.*

It might be thought that artificial intelligence techniques or other types of computational methods are irrelevant in countries with few technological resources. As just one example of the possibilities, however, take road traffic in cities. The chaotic and spectacular road congestion that is characteristic of developing-world cities is a microcosm of opportunities for applying AI methods. The problems are mainly caused by inadequate infrastructure (for example, road layouts that have not changed significantly despite decades of economic growth, unsealed or pothole-strewn roads), and a lack of resources to monitor or control traffic (for example, scarce and possibly corrupt traffic police, rolling blackouts affecting traffic lights). Computational solutions might come in the form of ways to cheaply gather real-time data, to advise individuals or emergency vehicles on optimal routes, to dynamically redeploy a limited number of

  ISSN 0738-4602

traffic police, or to analyze possible reconfigurations of the road network to remove bottlenecks. Any such solution must take into account the unique nature of traffic in these places, where the assumptions made in developed-world intelligent transport systems — for example, that drivers travel in the correct direction, and only on the road — might not be valid.

In this and other domains such as health and agriculture, we find that a number of developing-world planning and decision-making challenges boil down to optimization under constraints on the basis of noisy data. Given the right assumptions, computational solutions can be brought to bear on specific cases of this sort, and in this article we describe examples of practical solutions we have applied in Africa, Latin America, and India.

It is unsurprising that computing has not provided many such solutions in these regions until relatively recently. In the mid-1990s in Uganda, for example, conveying data electronically was not easy. Even making a phone call was a privilege restricted to those with access to one of the few phones in the country (run by the national telecoms monopoly), and phoning internationally would often require meeting an exchange operator in advance and paying a bribe in order to have the call put through at a prearranged time. A lack of electricity supply, network infrastructure, or computing hardware made it difficult to deploy any type of computing system, or for it to run reliably, or for anyone to access or benefit from it in any meaningful way.

By contrast, the developing world now contains most of the world's phone owners and Internet users. Just as in the developed world, the penetration of networked devices has led to vast amounts of data, which can reveal a wide range of information that would be very difficult to measure otherwise. From mobility patterns to traffic information these signals expose insights about such societies, providing information relevant to areas like health or urban planning. With few incumbent technological interests, there can also be a lack of red tape to hinder development of new technology, allowing the quick rollout of services such as money transfers by mobile phone — which have yet to be successfully implemented in rich countries to the same extent.

This is not to suggest that the field of computational sustainability in the developing world is now an easy domain to deploy computing ideas, however, and we have seen several well-meaning projects that were ultimately unsuccessful. In a fragile economy, technology deployments generally need to be immediately cost saving or profit making in order to survive. Finding the right set of assumptions can also be difficult: in an engineering approach to abstract away the nonessential parts of a problem, we often find that subtle yet crucial social factors are lost in the process.

We have three suggestions of research topics in which computational sustainability and artificial intelligence can be applied specifically to the developing world.

Intelligence Gathering

Often the data available to developing-world decision makers is both noisy and scarce, and policies on such topics as health or agriculture might be formed on the basis of very weak information. Frequently, information is gathered through expensive surveys or personal interviews; in a region with roads that are impassable any time there is heavy rain, for example, this is difficult. AI techniques allow the possibility of making better inferences from existing data sources, combining many weak signals into a few strong ones, or taking advantages of new data collection possibilities such as mobile crowdsourcing. It can also be possible to produce replacements that are closer to real time (and therefore more actionable) for official statistics in this way (on public health or food security, for example), or to generate new signals altogether that provide insights that were not previously available.

Just as in the developed world, mobile communications and social media generate vast amounts of rich behavioral and social information useful to inform policy makers. However, these type of data sets pose many challenges: (1) privacy: large-scale data sets typically involve millions of citizens whose privacy needs to be maintained, (2) algorithmic: design algorithms that can extract information from terabytes of data, (3) representativity: make sure that the digital traces represent the overall population, especially critical in the developing world with large socioeconomic differences, (4) scale: analysis at urban or national levels might require different techniques, and (5) visualization: intelligent visualization techniques that allow organizations and decision makers with little AI/IT knowledge and budget to understand and explore analytical results.

Compensating for a Lack of Human Experts

Where there is a shortage of skilled personnel it can be useful to automate their decision-making processes. For example, laboratory technicians are often in short supply in poor countries, making it difficult for people to get reliable diagnoses of disease. A similar situation applies to agricultural extension workers, who can recognize viral plant infections and advise farmers on the best course of action, meaning that farmers might not be able to plan effectively. AI techniques have the potential to mitigate those problems, by carrying out automated laboratory tests or providing personalized advice to farmers.

We might also try to amplify the abilities of an expert, rather than replace them. For example, we have personally seen cases in which laboratory technicians in a national referral hospital are expected to carry out microscopical malaria tests on over a hundred blood samples per day. It is simply not possible for one person to carry out that number of tests in such a period of time with any degree of rigor. Given automated tools to help triage samples and direct their attention, however, they might be able to work more effectively.

In areas where there are high degrees of illiteracy, applications of AI can help in other ways. For example, speech-recognition methods might be modified to cope with languages that are underresourced. In AI models of education applied to developing-world contexts, the missing human experts are the teachers; that is, the long-term goal might in fact be to create more human experts rather than to replace them.

Choosing How to Allocate Scarce Resources

A defining characteristic of developing countries is that they have very limited resources, and it is usually not clear how to optimally allocate them (for example, there is some limited budget for sanitation engineering or traffic management, and various spending choices). Specific cases can be framed computationally as optimization problems. In some cases this might be agent-based or adversarial: for example, inspectors traveling between pharmacies checking for counterfeit drugs ideally need travel schedules that are both cost-effective and difficult to predict.

At the macro level, most developing regions are relatively chaotic and very poorly planned; by collecting better intelligence we would like to move policy and action toward data- and learning-driven policy making. The issues of optimal resource allocation are hence closely related to the first challenge of collecting good intelligence in new ways.

In the remainder of this article we illustrate some real-world approaches to these challenges in different domains: health, agriculture, transportation, and public policy.

Health

The management of disease is an important part of sustainability, and by considering the resource constraints in developing-world health-care systems we find that many existing computational techniques applied in health can be adapted to the needs of developing nations. We give examples of some work here that range from diagnosis at the level of individuals up to national-scale monitoring systems that take advantage of newly available data sources.

Point-of-Care Diagnosis

The gold standard test for malaria is the analysis of a blood smear under a microscope. This is currently possible only where there is both laboratory equipment and a trained technician to perform the diagnosis. In this method a small blood sample is first taken (usually by finger prick), a glass slide is then prepared using a suitable staining solution, and finally the red blood cells are examined microscopically by an expert to identify whether the characteristic shapes of malaria parasites are visible. While the first two steps are possible by somebody with little training, the final microscopic analysis requires significant experience. Note that other diagnostic tests are available, such as antibody tests, but these have often been found to have a high false positive rate in endemic areas.

A number of researchers have looked at automating the diagnostic process with computer vision techniques. We built on this work by collecting a data set of labeled images taken under field conditions from a Ugandan hospital, in which artifacts and poor staining added to the complexity of the parasite detection problem (figure 1). Constructing a classifier using a set of morphological image features, we were able to obtain usable accuracy, with superior diagnostic performance to antigen-based rapid diagnostic tests, for example (Quinn et al. 2014). Real-time diagnosis was found to be possible even on low-powered Android devices (Mubangizi et al. 2012).

Disease Surveillance

There exist many disease surveillance systems that give public health officials capabilities to monitor and react to epidemic spreadings. Next, we describe one such collective effort in collaboration with researchers in Lahore University of Management Sciences (LUMS), Pakistan, to develop a disease surveillance system for the state of Punjab, Pakistan (Pervaiz et al. 2012a, Ahmad et al. 2013). Punjab-IDSS[1] is a collective research effort to develop an intelligent disease surveillance system that leverages health hotlines and an active mobile health workforce for disease spread tracking, early prediction, and containment of dengue-related epidemics in Punjab, Pakistan. In response to the 2011 dengue outbreak, the Punjab government started a dengue health hotline that enables mobile users to report denguelike symptoms and larvae accumulation and to obtain valuable feedback on preventive measures. A total of 117,470 complaints have been recorded in the system to date. The Punjab Intelligent Disease Surveillance System (IDSS) leverages the health hotline information combined with Internet-based news monitoring tools to provide accurate, real-time dengue epidemic detection at a fine-grained location granularity within a city. To achieve this goal, the system uses a combination of several sophisticated statistical learning algorithms combined with locality-specific dengue propagation models. Our system is also connected to the Dengue Patient Reporting Systems used by major hospitals in Lahore to report suspected and confirmed cases of dengue. Another component of the system is a disease activity tracking system where a team of 1500 mobile health workers

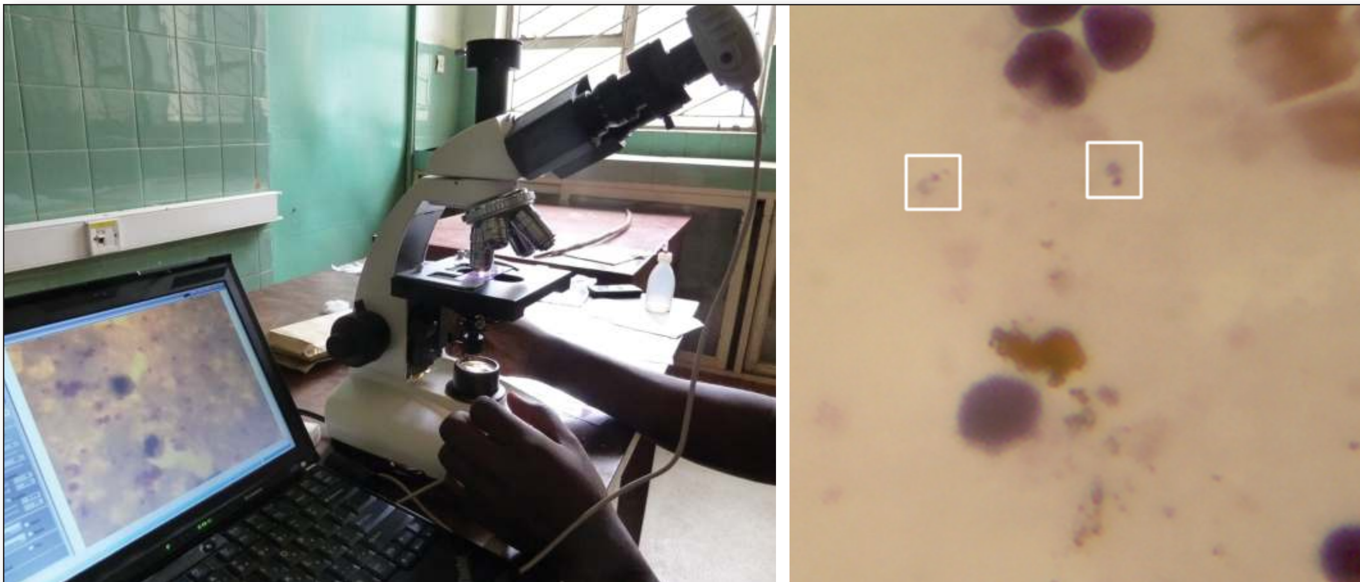

Figure 1. Automating the Diagnostic Process with Computer Vision Techniques.

Left: Real-time malaria diagnosis by capturing an image stream from the eyepiece of a microscope. *Right:* Two *Plasmodium falciparum* examples in a typical field of view of a thick blood smear.

with Android phones conduct statewide disease control activities. For dengue cases detected from the health hotlines, the workers can be mobilized for implementing focused dengue containment strategies. The health workers use the Android devices to gather verifiable proof of larvae accumulation in different localities. We are currently rigorously evaluating the effectiveness of our dengue containment strategies.

The current Punjab dengue outbreak detection system raises early warning alerts that are relayed to the Punjab Information Technology Board (PITB) and several hospitals in Punjab. This system is being used by the Punjab government for dengue decision making and they have analyzed and fine-tuned dengue-specific models based on weekly data from Muzaffargarh, Punjab. Another related system developed by the LUMS researchers is FluBreaks (Pervaiz et al. 2012b), a generic disease outbreak detection system that uses Internet search queries and that outperforms Google Flu Trends. We are currently integrating FluBreaks into the Punjab IDSS. In summary, Punjab IDSS is a large-scale collective effort that addresses an extremely challenging problem of designing a fine-grained dengue outbreak detection and containment system in a resource-limited society such as Pakistan.

Combining Disease Surveillance and Diagnosis

The above tasks of estimating the density of an infectious disease in space and time and diagnosing that disease in individuals are generally carried out separately. Informally, doctors may be aware of outbreaks of human disease in particular places or seasonal variations in disease risk, and they may interpret test results accordingly. But the diagnosis is not usually formally coupled with estimates of disease risk.

The tasks of mapping disease density over space and time and of diagnosing individual cases are complementary, however. A "risk map" can be used to give a prior in diagnosis of an individual with a known location. In turn, the results of individual diagnoses can be used to update the map in a more effective way than simply making hard decisions about infection statuses and using summary count data for the update. The potential for combining maps and diagnosis in this way has come about with the possibility of performing diagnosis with networked location-aware devices that can carry out the necessary calculations.

We introduced a probabilistic state space model of malaria spread in the paper by Mubangizi et al. (2012), which incorporated the computer-vision-based system for detecting plasmodium in microscopical blood smear images described earlier. By combining these two tasks, we found the accuracy in each case could be improved as compared to carrying out the tasks in isolation. This is done using dynamic Bayesian networks that represent both the spatial density of disease over time, and the symptoms and infection status of individuals at any time instant (figure 2).

Outbreak Control

In case of a pandemic, the World Health Organiza-

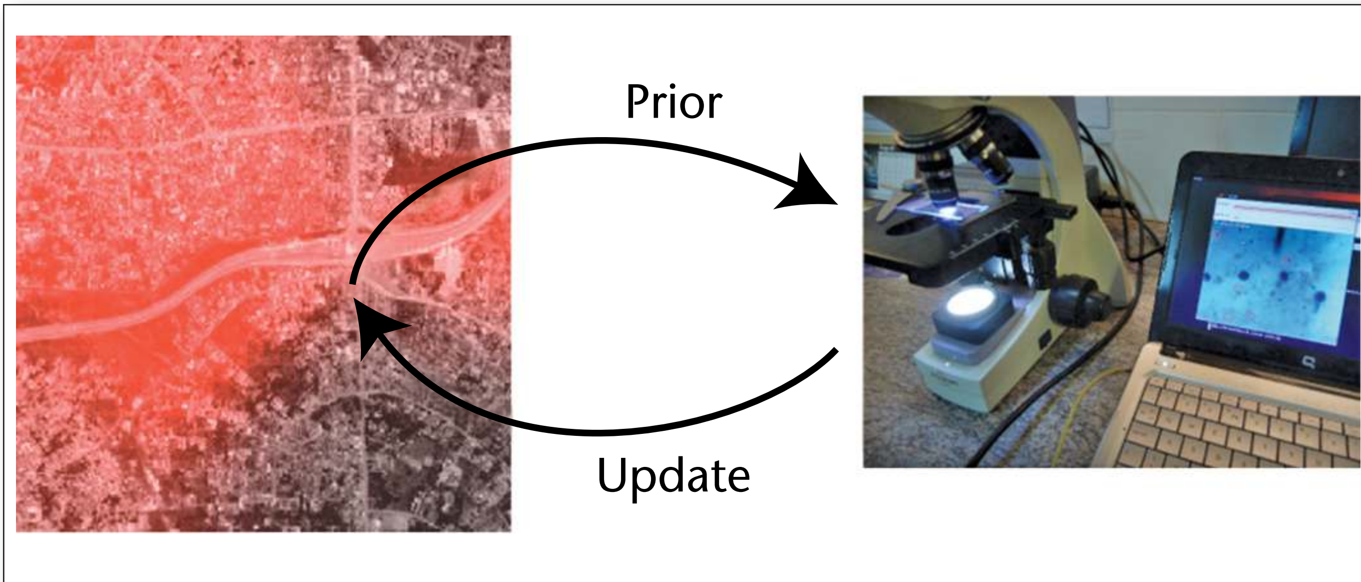

Figure 2. Combining a Probabilistic State Space Model with Computer Vision.

When diagnosis is carried out on a location-aware, networked device, the uncertainty in disease intensity mapping and automated diagnosis can be jointly modeled. This allows both tasks to be carried out more accurately.

tion (WHO) recommends closing educational, government, and business units as a plausible measure to reduce the transmission of a disease. Following these recommendations, governments usually institute policies that aim to reduce individual mobility in order to control an epidemic. Understanding the effectiveness of such mandates becomes critical for the design of successful policies to contain the spread of future epidemics; especially in emerging regions with limited resources and where the economic impact of such measures can be highly negative on the economy. The preventive actions implemented by the Mexican government to control the H1N1 flu outbreak of April 2009 constitute an illustrative example. In fact, the authorities followed the recommendations of WHO and, after raising a medical alert period, they closed all educational and business units to avoid the spreading of the epidemic.

The deficiency of analytical results on the impact of such mandates is mostly due to the lack of large-scale quantitative data about human motion. Such information is typically obtained from census data, which in the case of emerging regions is computed only every 5 to 10 years. Nevertheless, the recent adoption of cell phones by very large portions of the population enables us to capture large-scale quantitative data about human mobility.

AlertImpact (Frias-Martinez, Williamson, and Frias-Martinez 2011; Frias-Martinez, Rubio, and Frias-Martinez 2012) focuses on novel approaches to model and analyze the evolution of an epidemic under different policy scenarios. AlertImpact is based on a novel agent-based model that, instead of using aggregated census data, takes advantage of information extracted from cell phone records to compute the individual mobility and social patterns of a population. Call detail record (CDR) databases are generated when a mobile phone connected to the network makes or receives a phone call or uses a service (for example, SMS, MMS, and so on). From all the data contained in a CDR, our agent-based model uses the encrypted originating number, the encrypted destination number, the time and date of the call, the duration of the call, and the latitude and longitude of the BTS tower used by the originating cell phone number and the destination phone number when the interaction happened. At its core, AlertImpact is an agent-based epidemic model (ABM) that has two main components: (1) a set of agents that are modeled using the information contained in CDRs and (2) a discrete event simulator that models virus propagation using a susceptible-exposed-infectious-recovered (SEIR) model.

Using the anonymized CDRs collected during the H1N1 outbreak in Mexico, AlertImpact shows that the restricted mobility due to the government mandates reduced by 10 percent the peak number of individuals infected by the virus and postponed the peak of the pandemic by two days (figure 3).

Food Security

The processes of food production and supply in some developing countries are fragile and easily affected by changes in climate or economy. It is useful to be able to anticipate the threats to sustainabil-

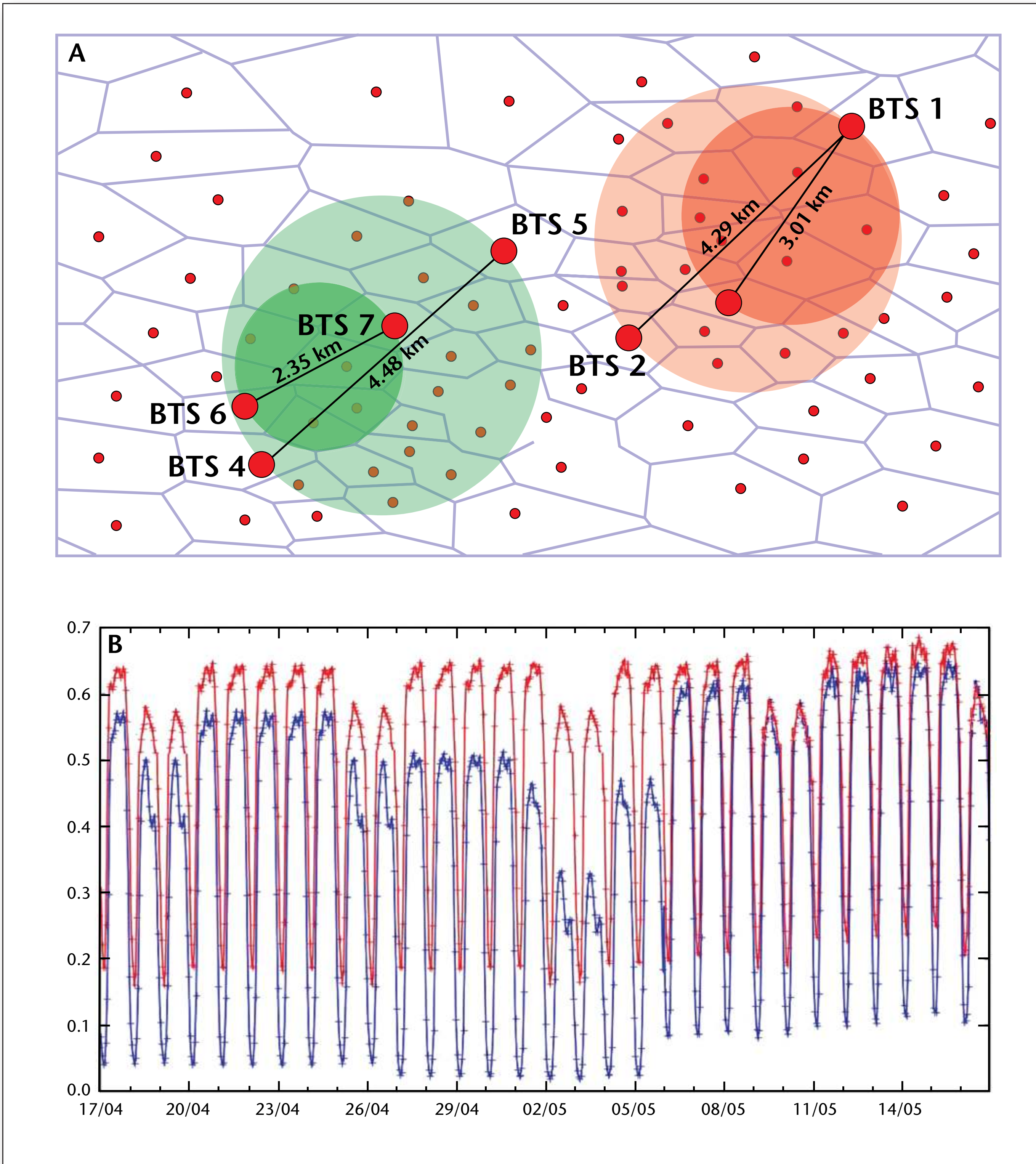

Figure 3. Restricted Mobility Due to Government Mandates.

(a) Changes in individual mobility due to government mandates. The BTSs represent the cellular towers and the polygons the coverage area of each tower. During the government mandates (darker colors), we observe a decrease in the individual diameter of mobility (b) Aggregated mobility of urban citizens during the outbreak (blue line) and during a normal period or baseline (red line). We can observe a decrease in mobility during the period when the government measures to prevent mobility were taken.

ity, for example, by measuring the spread of viral diseases in staple crops or tracking the degradation of farmland. Because good quality data is difficult to come by in these contexts, we again find that inference methods are particularly useful to make the most of the resources that do exist.

Crop Disease Monitoring

The economies of many developing countries are dominated by an agricultural sector in which small-scale and subsistence farmers are responsible for most production, utilizing relatively low levels of agricultural technology. As a result, disease among staple crops presents a serious risk, with the potential for devastating consequences. It is therefore critical to monitor the spread of crop disease, allowing targeted interventions and foreknowledge of famine risk. Currently, teams of trained agriculturalists are sent to visit areas of cultivation and make assessments of crop health. A combination of factors conspire to make this process expensive, untimely, and inadequate, including the scarcity of suitably trained staff, the logistical difficulty of transport, and the time required to coordinate paper reports.

Survey resources can be used much more efficiently by performing data collection with mobile devices and by directing survey progress through the application of AI techniques (figure 4). We deployed such a system for monitoring viral disease in cassava in Uganda (Quinn, Leyton-Browne, and Mwebaze 2011). Diagnosis of plant disease can be automated using images taken by a camera phone, enabling data collection by survey workers with only basic training. The classification uses a simple set of color and shape features that are feasible to extract on a mobile device (Aduwo, Mwebaze, and Quinn 2010). For classification of cassava mosaic disease, we found 96 percent AUC to be achievable in this way. This allows us to build up a real-time map of crop disease. Furthermore, we can introduce an active learning problem in which survey teams can be dynamically directed to the most informative areas. Since the standard categorizations of plant disease levels are in terms of ordinal categories, we find that Gaussian process ordinal regression (Chu and Ghahramani 2005) is an effective spatial density model on which to base the analysis.

Identifying Drought and Agricultural Trends in Every Locality

Agriculture forms the backbone of several emerging economies. In the past few years, several agrarian regions have been severely affected, due to a combination of several factors including climate, lack of water availability, soil infertility, and so on. However, in reality, many policy makers and the general public are often unaware of the status of agricultural conditions across different localities within their countries.

We have built a location summarization system that leverages information available on the web to summarize important climate and agricultural trends in a specific location (Chakraborty and Subramanian 2011). Such information could potentially be useful knowledge to both raise awareness about specific trends as well as for policy makers in learning about locations with problematic agricultural conditions.

The system automatically constructs a location-specific climate and agricultural information aggregation and summarization portal based on disparate information sources from the web. Given a location, the system searches the web for information concerning different parameters in influencing agriculture and climate and presents a summary of relevant information. Our system is built around three key ideas. First, we (manually) identify target topics of interest within climate and agriculture (such as soil, water) and construct a list of appropriate search queries that comprehensively describe the different aspects of the target topic. Second, for each target topic (such as soil or water), we download the top search result pages and perform information extraction on the textual content of these pages. The information extraction process aims to extract the critical textual snippets that can capture the key trends within the target area. Finally, we perform information summarization where the goal is to identify key trends corresponding to each target topic. We have tailored standard information retrieval techniques to address these problems. This summarized information on the location can be utilized to detect different problems and infer possible remedies from it. Hence, the aim is to highlight the important as well as lesser-known facts, thereby increasing the availability of knowledge. Clearly, availability of knowledge can lead to detection and potentially prevent any catastrophes.

Prediction of Food Insecurity from Remote Sensing Data

Satellite images, and features derived from them such as the Normalized Difference Vegetation Index, have long been used for early warning of food shortages. This gives an overall prediction of food insecurity in an area, though in a heterogeneous population it does not directly predict which sectors of society or households are most at risk. We used information on 3094 households across Uganda collected between 2004–2005 combined with remote sensing images taken at 10-day intervals in the same period to model probabilistically the relationship between calorific intake per person in a household, satellite NDVI and rainfall estimate data, and demographic features such as land size, household size, and livestock ownership (Quinn, Okori, and Gidudu 2010). We showed that adding demographic information about households to satellite observation data gives better accuracy in making predictions at a household level. The increase

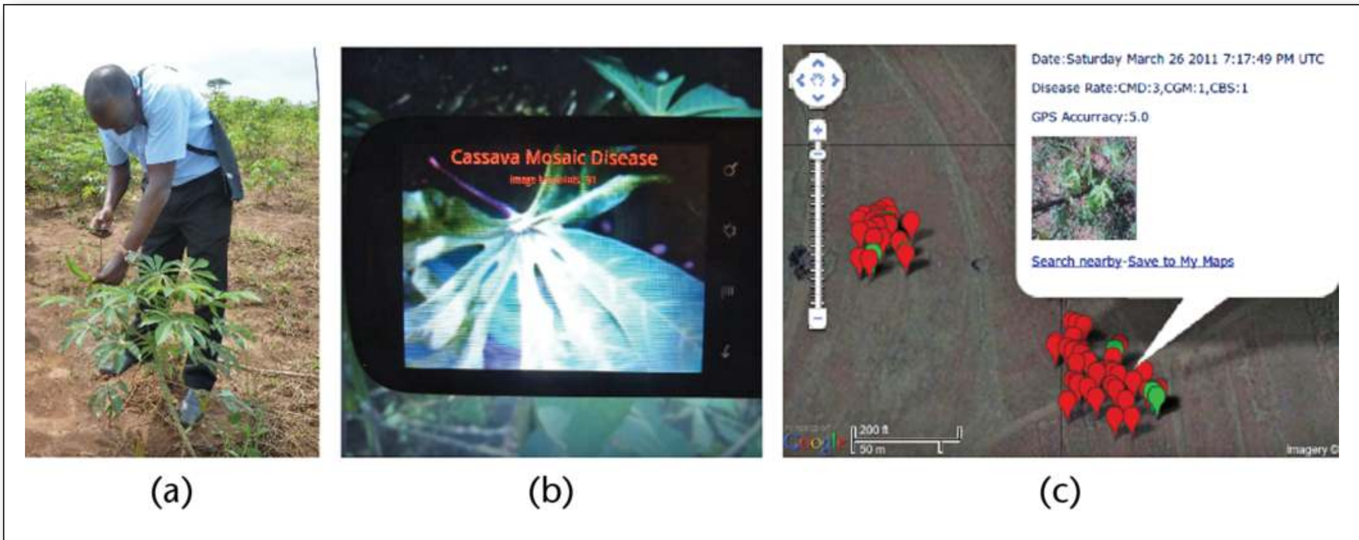

Figure 4. Performing Data Collection with Mobile Devices and Directing Survey Progress with AI Techniques.

(a) Mobile-phone based survey of cassava field; (b) Software on the phone detects cassava mosaic disease from leaf appearance; (c) Data collected with the phone is instantly uploaded to the web.

in the specificity of the predictions has the potential for administrators or aid agencies to take more targeted action than would be possible with only a general famine warning for some area.

Cropland Disappearance

Croplands worldwide are in decline. Degradation of arable land is a cause for concern, especially in developing countries where agriculture, including subsistence farming, makes up a significant percentage of economic output. In developing regions, urban population is increasing, leading to expansion of cities and development of new cities or townships. Often these expansions are done on arable lands. Apart from urban expansion, industrial developments are often done on agricultural land resulting in loss of arable lands. On many occasions, these acquisitions are unplanned and unauthorized. Such loss of arable land can have huge impact, particularly for agrarian economies. Not only can it affect the lives and livelihoods of the population who are directly dependent on agriculture, it can directly affect food security because of reduced production. Apart from human-led development, changing climate is also leading toward a change in the land pattern.

To characterize and monitor the change in land pattern in a given locality, we have developed a system that uses satellite image data over several years measured at fine-grained granularities to monitor land change over the years (Chakraborty et al. 2012). Our system leverages a combination of well known computer vision and image processing algorithms to compute the change in land pattern based on satellite data retrieved from Google Earth, which offers a large consolidated corpus of satellite images across the globe including historical information. Given a location or a geographical area, our system can access the latest available satellite image in addition to earlier images available and classify the images as cropland, developed, forest, or barren. Following this classification process, the tool computes the total amount of change of pattern in the region and also the type of change (for example, cropland changed to developed land).

Transportation

As described earlier, the problems regarding transportation in the developing world provide many opportunities for the application of computational techniques. The unique characteristics of developing-world traffic, combined with the budget constraints of city planners, often make the usual approaches to traffic management in developed countries inappropriate. Therefore new methods are required to address these problems.

Modeling Commuting Patterns

Commuting matrices characterize the transitions of a population between different geographical regions representing the origin and destination of a route. These matrices are key for a variety of fields, including transportation engineering and urban planning. Up to now, these matrices have been typically generated from data obtained from surveys. Nevertheless, such approaches typically involve high costs that limit the frequency of the studies, especially in low-resource regions like developing countries. At

Figure 5. Collecting Traffic Flow Data with Camera Phones.

Left: Traffic congestion in Kampala. Center: Low-cost, solar, camera-phone-based traffic monitoring unit. *Right:* Video analysis of traffic flow from solar unit.

the same time, cell phones can be considered one of the main sensors of human behavior due to its ubiquity and, as such, a pervasive source of mobility information at a large scale both in developed and emerging regions.

We have proposed a new technique for the estimation of commuting matrices using the mobility data collected from the pervasive infrastructure of a cell phone network: Call detail records (Frias-Martinez, Soguero, and Frias-Martinez 2012). Our goal is to show that we can construct cell-phone-generated matrices that capture the same patterns as traditional commuting matrices, but at a much lower cost. In order to do so we use optimization techniques in combination with a variation of temporal association rules. The resulting commuting matrices computed from CDRs constitutes an effective solution to complement traditional approaches. Our experimental evaluation and validation has showed that we can compute commuting matrices with a high level of accuracy using CDRs, and as a result our CDR-generated matrices can be used for the same purposes as traditional matrices, which typically are much more expensive to compute. As cell phone infrastructure becomes yet more pervasive, we envision a future in which developing regions will be able to gather and understand transportation information without the need to carry out expensive surveys or use unaffordable technologies and in collaboration with telecommunication companies.

Vision-Based Road Traffic Congestion Monitoring

Due to poorly planned road networks, a common feature of many developing regions is the presence of small critical areas that are common hot spots for congestion; poor traffic management around these hot spots potentially results in elongated traffic jams. More information about traffic congestion patterns would enable better use of existing infrastructure in resource-constrained cities. Collecting real-time congestion information with current technologies in use is expensive, prohibitively so in many developing countries. The congested and chaotic nature of traffic in these regions can invalidate certain conventional approaches, for example, any that make assumptions that vehicles travel in fixed lanes.

We have developed a simple automated image-processing mechanism for detecting the congestion levels in road traffic by processing CCTV camera image feeds (Jain, Sharma, and Subramanian 2012; Jain et al. 2012). Our algorithm is specifically designed for noisy traffic feeds with poor image quality. Based on live CCTV camera feeds from multiple traffic signals in Kenya and Brazil, we show evidence of this congestion collapse behavior lasting long time periods across multiple locations. To partially alleviate this problem, we present a local decongestion protocol that coordinates traffic signal behavior within a small area and can locally prevent congestion collapse, sustaining time variant traffic bursts. Using a simulation-based analysis on simple road network topologies, we have shown that our local decongestion protocol can enhance road capacity and prevent congestion collapse in localized settings (Jain, Sharma, and Subramanian 2012).

In a related project deployed in Uganda (Nakibuule, Ssenyange, and Quinn 2013), we found solar-powered units built around camera phones to be effective in collecting traffic flow data. The use of such hardware drastically cuts the cost of collecting congestion information compared to conventional roadside CCTV systems or other traffic sensors such as induction loops. To calculate speeds of traffic flow with this system, we need to calibrate the camera projection then use keypoint matches to identify

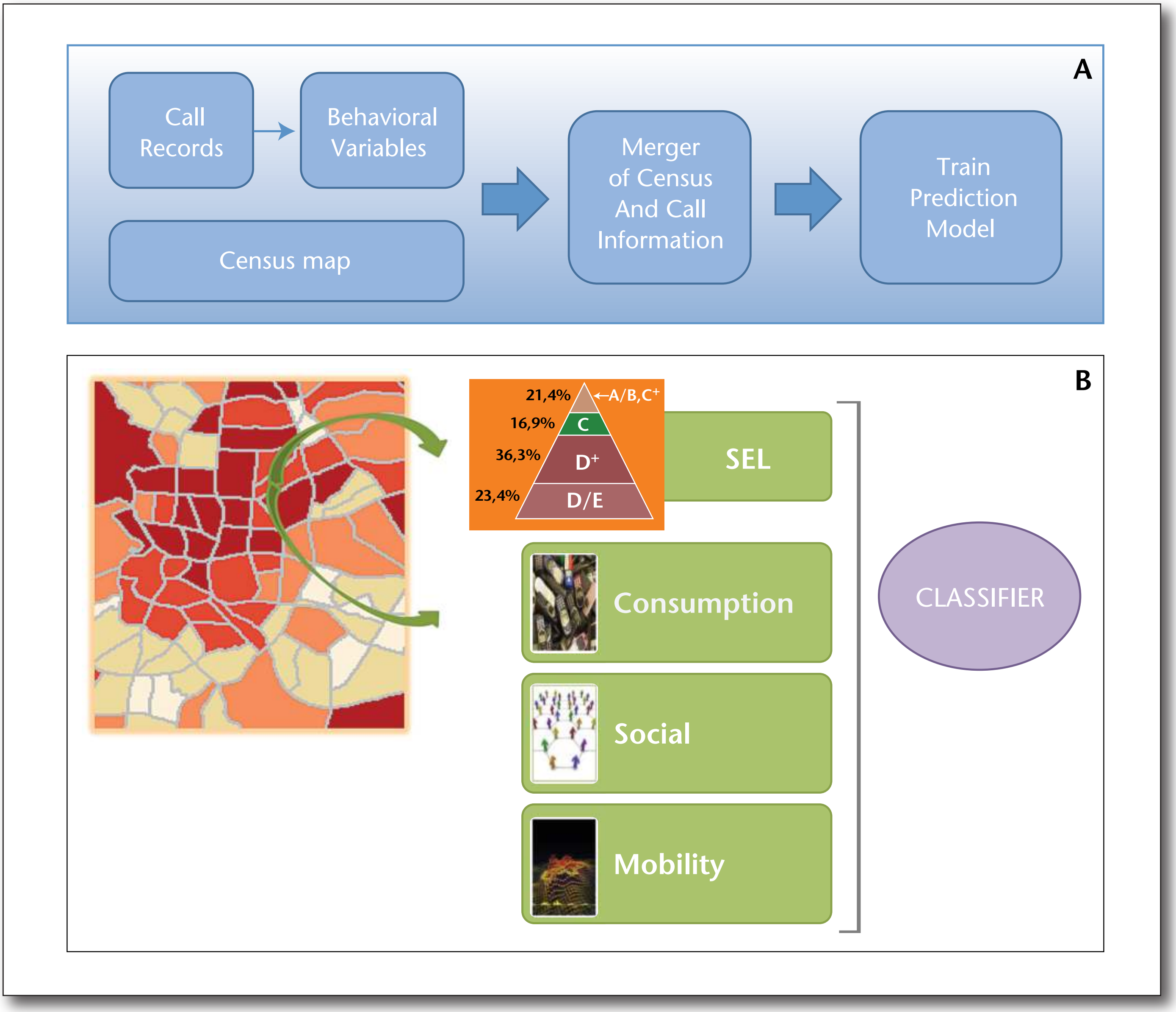

Figure 6. CenCell.

(a) Architecture of CenCell. (b) We build a supervised classifier using socioeconomic information from census maps with the consumption, social, and mobility variables computed from call detail records.

motion. We find it necessary to classify each moving patch in the image stream to distinguish vehicles from nonvehicles, given the amount of visual clutter in this setting. The system is pictured in figure 5.

Social Economics and Policy

As we have already discussed, the analysis of call detail records at large-scale is especially relevant to emerging regions where, due to limited budgets, surveying or gathering social information is often challenging and comes at a high cost. For example, socioeconomic maps contain important indicators regarding the status of households at urban and national scales. Computing these maps is critical given that many policy decisions made by governments and institutions are based upon socioeconomic information. For that purpose, national statistical institutes (NSIs) conduct censuses every 5 to 10 years and typically require a large number of enumerators to carry out interviews gathering information pertaining the main socioeconomic characteristics of each household. However, the entire interview process is highly expensive, especially for budget-constrained regions.

To overcome this issue, we have designed CenCell, a new tool for governments and policy makers that facilitates computing affordable census maps by decreasing the number of geographical areas that need to be interviewed by the enumerators (Frias-Martinez, Rubio, and Frias-Martinez 2012; Frias-Martinez et al. 2012). The tool is designed to allow institutions to approximate the census information of areas not covered by the enumerators using anonymized CDRs gathered by telecommunication companies.

At its core, CenCell consists of a battery of supervised (SVMs and random forests) and unsupervised (EM clustering) techniques that determine the socioeconomic level of a region based on the average consumption, mobility, and social network patterns of its citizens computed from their calling records. We have empirically evaluated CenCell with millions of cell phone records from urban citizens and we have shown that it correctly determines the socioeconomic levels computed by the NSIs with high accuracies (Soto et al. 2011). Thus, CenCell significantly decreases the workload of the enumerators that carry out the interviews and as such, allows us to reduce the budget allocated for the computation of census maps.

Additionally, CenCell also allows us to investigate the statistical relationships between socioeconomic levels and different consumption, mobility, and social network patterns (figure 6). Such input can help decision makers understand, among other things, the effect of socioeconomic factors on the way citizens commute (through mobility variables) or on their social connections (through social network variables) at large scale (Frias-Martinez and Virseda 2013, Frias-Martinez et al. 2013, Frias-Martinez and Virseda 2012).

Conclusions

In this article we have outlined the ways in which computational sustainability and artificial intelligence methods can be applied to problems in the developing world and shown practical examples from a number of different domains. The technological conditions that have made such work feasible have only arisen recently, and new developments are continually providing further opportunities for computational methods to improve sustainability and well-being in poorly resourced parts of the world.

Note

1. See punjab-idss.org.

References

Aduwo, J.; Mwebaze, E.; and Quinn, J. 2010. Automated Vision-Based Diagnosis of Cassava Mosaic Disease. In *Proceedings of the Workshop on Data Mining in Agriculture* (DMA). Berlin: Ibai Publishing.

Ahmad, T.; Rehman, N. A.; Pervaiz, F.; Kalyanaraman, S.; Safeer, M. B; Chakraborty, S.; Saif, U.; and Subramanian, L. 2013. Characterizing Dengue Spread and Severity Using Internet Media Sources. In *Proceedings of the 4th Annual Symposium on Computing for Development* (ACM DEV). New York: Association for Computing Machinery.

Chakraborty, S., and Subramanian, L. 2011. Location Specific Summarization of Climatic and Agricultural Trends. In *Proceedings of the 20th International Conference on the World Wide Web* (WWW). New York: Association for Computing Machinery.

Chakraborty, S.; Dalton, S.; Nyarko, Y.; and Subramanian, L. 2012. Computing the Disappearance of Crop Land Using Satellite Images. Paper presented at the 3rd International Conference on Computational Sustainability, July 4–6, Copenhagen, Denmark.

Chu, W., and Ghahramani, Z. 2005. Gaussian Processes for Ordinal Regression. *Journal of Machine Learning Research* 6(7): 1019–2005.

Frias-Martinez, E.; Williamson, G.; and Frias-Martinez, V. 2011. An Agent-Based Model of Epidemic Spread Using Human Mobility and Social Network Information. In *Proceedings of the 3rd International IEEE Conference on Social Computing*. Piscataway, NJ: Institute of Electrical and Electronics Engineers.

Frias-Martinez, V., and Virseda, J. 2013. Cell Phone Analytics: Scaling Human Behavior Studies into the Millions. *Journal of Information Technologies and International Development* 9(2): 35–50.

Frias-Martinez, V.; Soguero, C.; Josephidou, M.; and Frias-Martinez, E., 2013. Forecasting Socioeconomic Trends with Cell Phone Records. In *Proceedings of the 3rd Annual Symposium on Computing for Development* (ACM DEV). New York: Association for Computing Machinery.

Frias-Martinez, V., and Virseda, J. 2012. On the Relationship Between Socio-Economic Actors and Cell Phone Usage. In *Proceedings of the 3rd International IEEE Conference on Information and Communication Technologies and Development*. Piscataway, NJ: Institute of Electrical and Electronics Engineers.

Frias-Martinez, V.; Rubio, A.; and Frias-Martinez, E. 2012. Measuring the Impact of Epidemic Alerts on Human Mobility Using Cell-Phone Network Data. Paper presented at the Second Workshop on Pervasive Urban Applications, Newcastle, UK, June 19.

Frias-Martinez, V.; Soguero, C.; and Frias-Martinez, E., 2012. Estimation of Urban Commuting Patterns Using Cellphone Network Data. Paper presented at the ACM SIGKDD Workshop on Urban Computing, Beijing, China, August 12. dx.doi.org/10.1145/2346496.2346499

Frias-Martinez, V.; Soto, V.; Virseda, J.; and Frias-Martinez, E., 2012. Computing Cost-Effective Census Maps from Cell Phone Traces. Paper presented at the Second Workshop on Pervasive Urban Applications, Newcastle, UK, June 19.

Jain, V.; Sharma, A.; and Subramanian, L. 2012. Road Traffic Congestion in the Developing World. In *Proceedings of the 2nd Annual Symposium on Computing for Development* (ACM DEV). New York: Association for Computing Machinery.

Jain, V.; Dhananjay, A.; Sharma, A.; and Subramanian, L. 2012. Traffic Density Estimation from Highly Noise Image Sources. Paper Presented at the 91st Annual Meeting of the Transportation Research Board of the National Academies. January 22–26, Washington, DC.

Mubangizi, M.; Ikae, C.; Spiliopoulou, A.; and Quinn, J. A., 2012. Coupling Spatiotemporal Disease Modeling with Diagnosis. In *Proceedings of the Twenty-Sixth AAAI Conference on Artificial Intelligence.* Palo Alto, CA: AAAI Press.

Nakibuule, R.; Ssenyange, J.; and Quinn, J. A. 2013. Low Cost Video-Based Traffic Congestion Monitoring Using Phones as Sensors. In *Proceedings of the 3rd Annual Symposium on Computing for Development* (ACM DEV). New York: Association for Computing Machinery

Pervaiz, F.; Ahmed, T.; Rehman, N. A.; Saif, U.; and Subramanian, L. 2012a. Punjab-IDSS: Dengue Surveillance, Early Detection and Containment. Paper Presented at the 91st Annual Meeting of the Transportation Research Board of the National Academies. January 22–26, Washington, DC.

Pervaiz, F.; Pervaiz, M.; Rehman, N. A.; and Saif, U. 2012b. FluBreaks: Early Epidemic Detection from Google Flu Trends. *Journal of Medical Internet Research* 14(5):e125 dx.doi.org/10.2196/jmir.2102

Quinn, J. A.; Andama, A.; Munabi, I.; Kiwanuka, F. N. 2014. Automated Blood Smear Analysis for Mobile Malaria Diagnosis. In *Mobile Point-of-Care Monitors and Diagnostic Device Design,* ed. W. Karlen and K. Iniewski. Boca Raton, FL: CRC Press.

Quinn, J. A.; Leyton-Brown, K.; and Mwebaze, E. 2011. Modeling and Monitoring Crop Disease in Developing Countries. In *Proceedings of the Twenty-Fifth AAAI Conference on Artificial Intelligence.* Palo Alto, CA: AAAI Press.

Quinn, J. A.; Okori, W.; and Gidudu, A. 2010. Increased-Specificity Famine Prediction Using Satellite Observation Data. In *Proceedings of the 1st Annual Symposium on Computing for Development* (ACM DEV). New York: Association for Computing Machinery

Soto, V.; Frias-Martinez, V.; Virseda, J.; and Frias-Martinez, E. 2011. Prediction of Socioeconomic Levels Using Cell Phone Records. In *Proceedings of the 19th International Conference on User Modeling, Adaptation and Personalization,* Lecture Notes in Computer Science. Berlin: Springer.

John Quinn is a data scientist at the United Nations Global Pulse lab in Kampala, a senior lecturer in computer science at Makerere University, and a visiting lecturer at the University of Sheffield. He has a BA in computer science from the University of Cambridge (2000) and a Ph.D. from the University of Edinburgh (2007). At Makerere he runs the AI-DEV research group, which studies the use of artificial intelligence techniques in health, agriculture, and resource-allocation problems.

Vanessa Frias-Martinez is an assistant professor in the iSchool at the University of Maryland. She received her M.Sc. and Ph.D. degrees in computer science from Columbia University. Frias-Martinez is interested in social computing, with a focus on the intersection between big data, policy, and social development. From 2009 to 2013, she was a researcher in the Data Mining and User Modeling Group at Telefonica Research in Madrid, Spain.

Lakshminarayanan Subramanian is an associate professor in the Computer Science Department in the Courant Institute of Mathematical Sciences at New York University (NYU). His research interests are in the areas of networks, distributed systems, and computing for development. He coleads the Networks and Wide-Area Systems (NeWS) research group and the Center for Technology and Economic Development (CTED) at NYU. He is a recipient of the NSF CAREER Award (2009), IBM Faculty Awards (2009, 2010), Google Faculty Award (2013), C.V. Ramamoorthy Award at the University of California, Berkeley, and Microsoft Research Challenge Award on Cellphones for Healthcare (2008). He is a cofounder of Entrupy Inc.