ON ACCOUNTING FOR THE INTERPLAY OF KINETIC AND NON-KINETIC ASPECTS IN POPULATION MOBILITY MODELS[†]

Kalyan S. Perumalla and Budhendra L. Bhaduri
Computational Sciences and Engineering Division
Oak Ridge National Laboratory
Oak Ridge, Tennessee, USA
E-mail: perumallaks@ornl.gov, bhaduribl@ornl.gov

KEYWORDS
Population dynamics, GIS, Parallel Simulation.

ABSTRACT

Several important applications are placing demands on satisfactory characterization of the bi-directional interplay between kinetic and non-kinetic aspects in the mobility of people and commodities. Example applications include emergency planning, energy planning, and policy making, all of which require new holistic approaches for capturing the interplay. Accurate characterization of the interplay requires integration of three distinct components, namely, data, models and computation. The availability of new sources of high-resolution data, and of detailed models together with recent advances in scalable computational methods now permits accurate modeling and simulation. This paper serves to highlight and argue that the interplay can in fact be captured in a high level of detail in simulations, enabled by the availability of new data, models and computational capabilities. Some of the challenges that are encountered in incorporating the interplay are outlined and plausible solution approaches are described in the context of large-scale scenarios involving mobility of people and commodities.

INTRODUCTION

A range of important applications deal with phenomena that involve activities of people in various normal as well as abnormal situations. Normal day-to-day mobility of people is of interest to analyze economic and environmental concerns. Activities and mobility of people under extraordinary circumstances as in emergencies are of interest for proper planning and response. Effects of changes to transportation modes or fuel technologies are of interest for urban planning and alternative energy policy studies. All of these applications have at their core the need to relate mobility with non-mobility tangibles of interest. In other words, there is a need to properly relate the *kinetic* aspects of mobility to *non-kinetic* aspects of people's activities. Kinetics deal with physical aspects of motion while non-kinetics deal with phenomenal or behavioral aspects that influence, and are influenced by, kinetics. This distinction between the two elements is further elaborated as follows.

Kinetics Kinetics of mobility deal with those aspects of motion that are physically related. These include properties such as the length, weight, velocity, acceleration and braking characteristics, along with related vehicular phenomena such as starting/stopping and lane changing. Traditional traffic/mobility simulators are kinetic simulators that adequately take into account the composition of kinetic phenomena of a large number of vehicles on complex road networks. Considerations include ways in which vehicles accelerate, merge, diverge, decelerate, stop, follow, overtake, coast or collide. Physical aspects of vehicles as well as of road networks (e.g., number of lanes, speed limits, types of light controllers, etc) are accounted for in detail. Important effects such as travel time delays and congestion levels are captured as a result of kinetics.

Non-kinetics On a plane separate from pure motion lie the non-kinetic factors, which are behavioral elements and/or resource tangibles in the environment. Some behavioral (non-physical) aspects are typically modeled in existing mobility simulation systems at limited levels of detail. One of the simplest behavioral aspects is the degree of driver aggression. The aggression level determines a driver's acceleration/deceleration traits as well as propensity of a driver to change lanes. Other extended non-kinetic aspects include elements such as individualized dynamic route selection, dynamic modification of trip destinations and intermediate stop selection. Even more elaborate non-kinetic aspects include the inter-relationships of trips among individuals and their relation to resources. Examples include effects of gasoline/energy consumption among localities and levels of exposure to emissions by age groups of interest (e.g., kids, senior citizens). Kinetics exerts influence, and is in turn itself influenced by, these non-kinetic aspects. For example, the speed of the vehicle is clearly influenced by the aggression level of the driver; the travel time taken on one trip can

---

[†] This manuscript has been authored by employees of UT-Battelle, LLC, under contract DE-AC05-00OR22725 with the U.S. Department of Energy. Accordingly, the United States Government retains and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a non-exclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this manuscript, or allow others to do so, for United States Government purposes.

determine the period of stay at one destination and the choice of next destination.

However, existing tools are, to a large extent, focused either on kinetic aspects or primarily on behavioral aspects but not both simultaneously. A range of simulators exist with detailed models of kinetics, including CORSIM, TRANSIMS (Fisher 2000; Meister et al. 2006), VISSIM, PARAMICS (Cameron et al. 1996) and OREMS (Franzese et al. 2001). A complementary set of systems exists for capturing highly detailed non-kinetic aspects, such as SEAS (Chaturvedi et al. 2005) and Repast/Mason (North et al. 2006) and others (De Almedia et al. 2005). In contrast, in our projects, we are developing approaches that adequately incorporate a *combination* of both kinetic and non-kinetic aspects at high fidelity, with relevance across a range of applications.

Motivating Application A principal objective of some of our projects at Oak Ridge National Laboratory (ORNL) is to develop a high performance computing driven geospatial simulation framework that is reusable across high fidelity applications, such as energy impact analysis, or accurate emergency/event planning, all of which is performed from local to regional scales. The simulation model would be utilized to evaluate the impacts of transformational energy optimization strategies such as alternate transportation technologies and energy-aware, travel efficient land use planning. Primary limitations of past attempts have been the lack of high resolution input data, absence of useful "energy and behavior focused" models, and inadequate computational capability. Also, development of memory-full, non-linear fine-grained behaviors at the level of each individual/controller, for a million or more individuals and/or devices, requires unprecedented support for parallel execution and efficiency from the underlying engines, but few such engines exist today.

In the next section, we describe the state-of-the-art and emerging trends in collection and dissemination of geospatial/demographic data at large-scale and high resolution. An initial demonstration is presented of the utility of the detail and scale of such data in accurate tracking of spatio-temporal population distribution in urban areas. Some of the limitations are identified of tools that are built primarily with considerations of kinetics alone. The scalable parallel execution platform required for high-fidelity simulations is presented, followed by highlights of challenges and some of our modeling language-based solution approaches. Finally, a summary and ongoing/future work are presented.

DATA SOURCES

Today, geospatial data collected from remote sensors and from ground-based surveys have become an integral part of the scientific research, planning, policy, and operational missions for local/national/global-scale government agencies. Technological advancement with cost and time efficiencies has prompted an explosion in the volume of spatial data but has also increased spatial and temporal resoluitions. This is also aided, since early 1990s, by community involvement via initiatives such as the National Spatial Data Infrastructure (www.fgdc.gov/nsdi) and the Global Spatial Data Infrastructure initiative (www.gsdi.org). These data sets are useful for modeling and simulation of population and social dynamics, and broadly categorized as follows.

Physiographic Physiographic data include information on natural earth surface features (vegetation and land covers, hydrography, and elevation) and human-made infrastructural features (land use for roads, railways, bridges, tunnels, residential/commercial buildings, etc.). Overall there has been a consistent trend of increasing coverage and resolution of global physiographic data sets, down to 15 meter resolution. The primary example of data assimilation effort in the USA is the Homeland Infrastructure Foundation Level Data (www.hifldwg.org) Working Group that is a coalition of several federal agencies which have collectively produced the Homeland Security Infrastructure Protection Gold (HSIP Gold) geospatial database of eleven critical infrastructures at very high resolution,. Traditional data such as political boundaries and roads, are being augmented with newer data, such as charcaterisitics of thousands of business activities.

Demographic Census has been the traditional source of demographic data. Additionally, intelligent data integration is being used to augment Census data with variation of population over space and time. Geodemographic data at such scales represents a more realistic non-uniform distribution of population. To date, significant advancements have been made in this direction through advanced spatial data integration and modeling through pychnophylictic (Tobler et al. 1997), dasymetric (Mennis 2003), and smart interpolation (Dobson et al. 2000) approaches. Among them, ORNL, as part of its LandScan project, has developed the finest resolution (1 km cell) population distribution model for the entire world, and is currently refining it to an even higher resolution (90 meter cell) for the USA.

Socioeconomic and Behavioral A plethora of household and micro-level data on socioeconomic and activity based behavioral data have been collected and are disseminated primarily by the U.S. Census Bureau and the Bureau of Transportation Statistics (www.bts.gov/data) of the U.S. Department of Transportation. The American Community Survey (ACS) from the U.S. Census is a new nationwide survey based data on general demographic, social, economic and housing characteristics. In addition to ACS, Census also provides a special tabulation project on Census tract-to-tract worker commuting data and worker schedules and time of cummute. Business locations are reported in multiple public (Census) and commercial databases (Dunn and Bradstreet; TeleAtlas; InfoUSA) are locations and number of workers for various industries, hospitals, prisons, restaurants, etc.

DATA-DRIVEN MODELING FOR FIRST-ORDER METRICS

The recent availability of such a range of physiographic, demographic and behavioral data has already started enabling novel modeling capabilities that were hithertofore not possible. Efforts are being focused on gleaning first-order estimates of metrics of interest (e.g., estimation of spatio-temporal population densities over the course of a day) directly from intelligent processing of the available data. Such estimates could be obtained without use of simulations, with detailed non-kinetic data but with simplifying assumptions of kinetics.

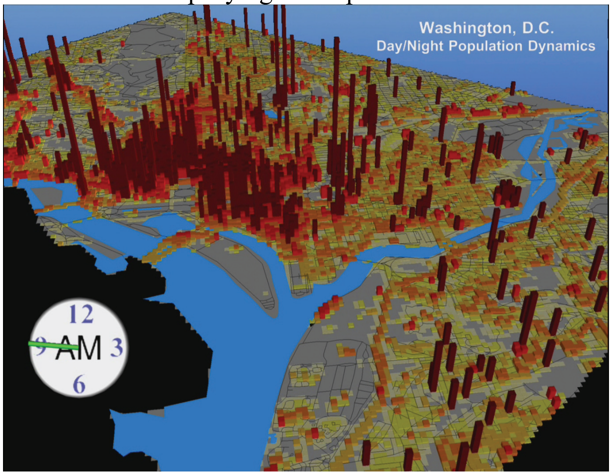

Figure 1: Snapshot of Spatio-temporal Distribution of Population in a Typical Day in Downtown Washington D.C., USA

Figure 1 shows a snapshot from a video of series of frames depicting three-dimensionally the distribution of population over an average day+night period in the Washington D.C. area. The height of vertical bars represents the number of people estimated to be in the corresponding spatial cell. A highlight of this work is the fine granularity of cell size, which is at least two orders of magnitude finer than that present in plain Census based databases. The analysis is performed at an unprecedented spatial resolution using some of the state-of-the-art population modeling techniques(Bhaduri 2003). The output of this model is useful as a first-order metric in applications such as emergency planning in which first responders need to be dispatched roughly proportionately with the expected population density across geographical regions.

Nevertheless, while the purely data-driven model has been a great leap forward in certain applications, it is inadequate for other applications. While being better than the past, it still needs improvements via simulations for better accuracy. This is an example of situation where there is more non-kinetic modeling but insufficient kinetics are modeled. For example, not all people reach their destinations at the same time (synchronously) in real life, but the model makes such simplifying (incorrect) assumptions in the absence of accurate combination of demographics with kinetics.

LIMITATIONS OF KINETICALLY-FOCUSED APPROACHES

On the other end of the spectrum are systems that are concerned about the kinetic measures of interest without full regard to the realism of the phenomena that drive the kinetics. These include the concepts of origin-destination matrices, turn probabilities and driver aggression levels, which unfortunately interfere with accurate capture of kinetic and non-kinetic interplay.

Origin-Destination Matrices Traditionally, Origin-Destination (OD) matrices have been used to generate dynamic behavior and initiate activity in the network. OD matrices are typically generated prior to simulation, and the generated matrices are fed to the simulation as input. A statically-determined set of trips is specified as the number of vehicles traveling between an ordered pair of source and destination points. The trips are generated to statically match desired spatio-temporal distributions of network activity. OD matrices are useful when the focus is on individual lanes, congestion, traffic light performance and such. They are easy to use, and hence employed by the majority of traffic simulators. However, they are not necessarily the most accurate with respect to other applications of interest such as in studying interaction of individuals with non-traffic elements such as resource consumption.

OD matrices are inherently static in nature and cannot accommodate the effect of changing network conditions on vehicular trip behavior and other individual behavioral complexities. Consequently, they can induce artificial effects in the simulation that are not truly representative of real-life conditions. For example, an OD matrix might encode a trip from work place to home, in order to statically suit the assumed network loading. However, when dynamic aspects, such as gasoline level in the car, are taken into account, the same trip might be split into a trip from work place to a gas station and then from gas station to home. Similarly, the converse might be true: a trip from work place to a party and then back home might be short circuited into a trip from work place directly home if the time spent on the road due to congestion becomes excessive and consequently the individual dynamically makes the decision to skip the party. These examples are meant to illustrate that static approaches such as OD matrices are inadequate to account for kinetic and non-kinetic interplay, and hence are unresponsive to behavioral aspects of individuals in response to changing network conditions.

Turn Probabilities Turn probabilities are another approximation often used to induce background traffic and experiment with vehicle populations. Here, retaining the identities of individual vehicles or retaining overall consistency of their mobility activity is not of particular concern. Vehicles are introduced onto the network and mobility produced by inducing random turns whenever vehicles reach an intersection. Turn probabilities are varied to suit the desired traffic patterns, but individual vehicular behavior is not of particular interest. Also, artifacts induced by such approximations are often ignored (such as a vehicle continuously circling a block as an artificial/unintentional consequence of a chain of turn probabilities resulting in such circling pattern).

Driver Aggression Levels The aggressiveness with which the drivers of vehicles drive has marked effect on the kinetic behaviors observed on the network. Typically, aggressiveness is modeled in terms of propensities towards changing lanes or intensities of acceleration/braking. However, aggression levels are not necessarily static, but are influenced by non-kinetic aspects. As an example, the aggression level of an individual changes and adapts to how late or early the individual is for his/her next destination.

In general, with respect to aggressiveness, non-kinetics can in fact be viewed as higher-order extensions of kinetics. While velocity and acceleration are first and second order kinetics, aggressiveness consistitues the third order. Similarly, changes over time in aggressiveness is realized as the fourth order derivative. Static level of aggressiveness is individual-specific and is typically supported by most kinetics-only simulators. However, fourth-order changes (changes in levels of aggressiveness at runtime) are dynamic and adapt to changing network and other environmental conditions. Such dynamics are needed in order to capture the kinetics and non-kinetics interplay accurately.

Entity Identities We believe that artifacts induced by approximations such as OD matrices and turn probabilities, while tolerable for certain applications, are in fact not only avoidable but also remedied by using a different modeling and computation approach. The framework should start with individuals modeled from ground up and their identities retained through out the simulation. Threads of activity are modeled for each and every individual, and individual identities are maintained through out entire duration of simulation.

SCALABLE EXECUTION OF MODELS

We are using the SCATTER simulation system (Perumalla 2006) for parallel execution of detailed transportation models at the individual level. SCATTER – Scalable Tool for Transportation and Energy/Emergency Research – is a parallel discrete event-based simulator that: (a) employs microsimulation models at vehicular level, (b) optimizes the data structures for modeled entity representations to minimize memory usage, (c) uses as a discrete event modeling formulation to simulate kinetic models faster, and (d) structures the models from the outset to enable parallel execution. It not only provides models with great amount of detail in the kinetic aspects, but also provides a software framework for interfacing non-kinetic aspects with the kinetic models. Its design has overriding goals of memory and speed efficiency from the outset, and, specifically, scalability via parallel execution. The use of discrete event modeling techniques as well as parallel simulation methods, provides for fast runtime speed.

Some of the kinetic aspects modeled for road network include vehicle position, velocity, acceleration/decelaration, traffic light controllers, multiple lanes per road segment, parking capabilities, and common kinetics-related behavior models such as “car-following” and “lane changing.” Congestion and cumulative blocking/queueing across inter-dependent chain of intersections is also modeled. Shortest-path routing is supported as well as a framework to add other dynamic route selection algorithms at the level of each individual vehicle. It does not suffer from any artifacts such as artificial cycling of vehicles on the road network as induced by other approaches based on turn-probabilities.

For parallel execution, SCATTER uses a unique partitioning approach to divide the simulated road network segments in a special way in order to expose execution concurrency and circumvent the tight coupling of simulation-time among models of inter-connected road intersections. SCATTER has been demonstrated on benchmark road networks to scale to over one million vehicles simulated per processor. It currently runs on a cluster of workstations platform connected by network, and does not rely on availability of shared-memory across processors.

MODELING LANGUAGE PRIMITIVES FOR COMPLEX ACTIVITIES

When modeled at the level of individuals, trips can be fairly complex to model. Mechanisms for specifying threads of activity at the level of each individual need to be addressed. New techniques are needed for the user to specify the precise interaction of kinetics with non-kinetics (e.g., gasoline levels, merging of individuals, etc.). It is also necessary to efficiently represent and execute a large number of activity threads (millions).

In order to correctly model the kinetic and non-kinetic interplay, every modeled unit (vehicle/individual) needs to retain its state and identity across trip steps and across multiple trips that it undertakes. The modeled unit should be capable of adapting and reacting in an autonomous fashion dynamically at runtime. It is insufficient, for example, to artificially introduce vehicle injections and vehicle removals at different points of the road network without retaining and reflecting the states of the removed vehicles into the injected ones.

As a means for expessing generalized trip specifications, we use a language for trip behavior.

Complex sequences of trips are specified using the language, in which the kinetic aspects are specified in conjunction with non-kinetics. Trips can be specified at the level of each individual. Each trip, in its most basic form, consists of a sequence of destinations and periods of stay at each destination. Most interestingly, the destinations and stay periods can depend and vary greatly on both kinetic and non-kinetic aspects of the trip. It is these aspects that are accommodated in the language. In general, a trip plan consists of several trip steps. Each step is a pair: <destination, stay period>. The language is designed to support the complexity inherent in the selection of destination and determination of stay periods in trips, as elaborated next.

Destination Selection and Modification A destination can be specified, as with traditional approaches, using the identifier of a specific intersection. Additionally, a destination can also be specified using non-kinetic criteria. The first factor in the criteria is the *type* of destination (e.g., residence, work place, grocery store, gas station, hotel, etc.). This factor clearly leaves ambiguity that needs to be resolved at simulation time and cannot be fully determined statically before simulation. The second factor is a set of *qualifiers* on the destination type (e.g., cheapest gas price, least congested store, hotel with rooms available, etc.). This second factor is useful in resolving the choice of the exact destination during actual simulation. Note that a location is potentially still needed in addition to the destination type and qualifiers. The location (e.g., geographical coordinates or an intersection identifier) specifies the vicinity in which the selection criteria are applied at runtime. Each destination, thus, is a triple: <destination type, destination location, destination qualifiers>.

Once a vehicle reaches its current destination, the destination becomes its current source, and a new destination is generated by invoking its dynamic destination selection rules. The next destination could either be specified as part of the original trip, or the trip could specify a dynamic choice with a special wild card value as destination. This results in a callback to the unit (vehicle) for generation of next destination. Note that the destination is distinct from the series of intermediate nodes through which the vehicle gets routed in order to reach the destination. The routing itself is partly kinetic (e.g., shortest path) and partly non-kinetic (e.g., scenic/preferred route, etc.).

Stay Period Determination Stay period is the amount of time to elapse while the vehicle is parked at its current destination. The stay period could be relative, or it could also be an absolute time for departure from that node to its next destination in the trip. Absolute times are useful to encode behaviors that are resilient to delays experienced by traffic due to network conditions (e.g., closing time of a restaurant or a factory).

Vehicle-People Interaction While trip behavior is possible to generalize to include kinetic and non-kinetic aspects, difficulty arises in specifying identities of individuals when trips merge/diverge (e.g., bus stops, school pickups, parties, errands, etc.). While high behavioral fidelity is possible with data availability, it also brings the need for accommodating entity identities and tracking them correctly along their kinetic movement. We are exploring ways in which to express the drop-off and pick-up of school children to/from their schools, while retaining the distinctions of modeled state values and trip specificactions of the adult individuals involved in the drop-off/pick-up.

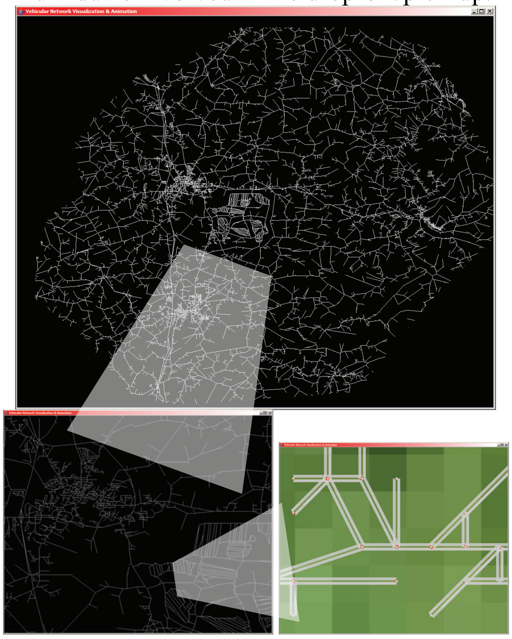

Figure 2: Portion of Actual Road Network of a Region of Interest

Implementation in SCATTER During initialization using the network configuration file, every node/intersection in the road network is tagged with its appropriate geographic information, such as its residential nature or type of work place. This information is later used at runtime to consult for the satisfaction of vehicle destination criteria.

Currently supported destination types include ordinal values for HOME, WORK, GAS, GROCERY and HOTEL. This set has been defined for current purposes focused on emergency scenarios in which interplay of vehicular kinetics with critical commodities such as gasoline, water and canned food needs to be tracked. The destination is a unique identifier of any intersection in the network, in which case the vehicle is routed towards that specific destination. Alternatively, if the intersection identifier is negative, it is interpreted as the negative of the identifier of a node near which the destination type should be explored by the system (one block away, in current implementation).

Figure 2 shows portions of an actual network being used as a complex scenario for testing SCATTER. The network is intended to reflect the complexity inherent in both the physiographic as well as behavioral complexity of a real-life scenario. The network would be subject to analysis under a range of scenarios: normal operation (planning, prediction, optimization), and/or effects under emergencies (hurricanes, evacuations) or cross-impact policy analysis (energy usage, hybrid fuel vehicles, emission levels).

SUMMARY

A range of applications exhibit the need for accurately incorporating the interplay between kinetic and non-kinetic aspects in the mobility of population and commodities, in normal, extraordinary, or futuristic scenarios. These applications are poised to account for this interplay by benefiting from a confluence of advances along three fronts, namely, data, models, and computation. There is a rich set of sources emerging for physiographic and demographic data. Behavioral modeling is being better understood with the availability of socioeconomic/behavioral data. Advances in computer simulation methods are enabling scalable execution of high-fidelity models at the scale of millions of individuals. In this context, some of our projects at ORNL are building upon our unique geospatial data and modeling capabilities that include high resolution geographic data sets of US residential population, population dynamics, and additional transportation/energy/emissions data. A scalable simulation framework is being built using a parallel discrete event modeling paradigm, and the activity pattern models are planned to be based on existing programs, such as the American Community Survey from U.S. Census Bureau and other literature on purpose specific trip rates, trip lengths, and number of stops per day/per circuit.

REFERENCES

Bhaduri, B. (2003). Watch Us Move. Popular Science.

Cameron, G. D. B., et al. (1996). "PARAMICS, Parallel Microscopic Simulation of Road Traffic." Journal of Supercomputing **10**(1): 25-53.

Chaturvedi, A., et al. (2005). Bridging Kinetic and Non-kinetic Interactions over Time and Space Continua. Interservice/Industry Training, Simulation and Education Conference.

De Almedia, C. M., et al. (2005). "GIS and Remote Sensing as Tools for the Simulation of Urban Land-Use Change." International Journal of Remote Sensing **26**(4): 759-774.

Dobson, J., et al. (2000). "LandScan: A Global Population Database for Estimating Populations at Risk." Photogrammetric Engineering and Remote Sensing **66**(7): 849-857.

Fisher, K. M. (2000). "Transims is Coming!" Public Roads **63**(5): 49-51.

Franzese, O., et al. (2001). A Methodology for the Assessment of Traffic Management Strategies for Large-scale Emergency Evacuations. 11th Annual Meeting of ITS America.

Meister, K., et al. (2006). A Comprehensive Scheduler for a Large-scale Multi-agent Transportation Simulation. International Conference on Travel Behaviour Research.

Mennis, J. (2003). Generating Surface Models of Population using Dasymetric Mapping. Professional Geographer. **55:** 31-42.

North, M. J., et al. (2006). "Experiences Creating Three Implementations of the Repast Agent Modeling Toolkit." ACM Transactioins on Modeling and Computer Simulation **16**(1): 1-25.

Perumalla, K. S. (2006). A Systems Approach to Scalable Transportation Network Modeling. Winter Simulation Conference, IEEE.

Tobler, W. R., et al. (1997). "World Population in a Grid of Spherical Quadrilaterals." International Journal of Population Geography **3**: 203-205.

AUTHOR BIOGRAPHIES

KALYAN S. PERUMALLA is a senior researcher in the Computational Sciences and Engineering Division at ORNL. He also holds an adjunct faculty appointment with the College of Computing, Georgia Tech. He received a Ph.D. in Computer Science from Georgia Tech (1999). Dr. Perumalla has over 10 years of research and development experience in the area of parallel and distributed systems, and has published widely on associated topics. He has built several research prototype systems for distributed debugging, network modeling, interoperable simulations and parallel optimization, many in use by researchers worldwide. He has served as co-PI on multiple federally-funded projects on scalable parallel/distributed systems. He can be reached via email at perumallaks@ornl.gov, and his homepage is at www.ornl.gov/~2ip.

BUDHENDRA BHADURI is a Group Leader of the Geospatial Information Systems and Technology group at ORNL. His research interests are geospatial sciences and information infrastructures. He received his Ph.D. in Earth and Atmospheric Sciences from Purdue University (1998). Dr. Bhaduri's expertise includes Geographic Information Systems, Remote Sensingand a variety of topics in earth and environmental sciences, and he has published extensively in these areas. He leads the team that creates and maintains LandScan, the leading world population database at high-resolutions. He can be reached via email at bhaduribl@ornl.gov and his homepage is at www.ornl.gov/sci/gist/bios/bio_bhaduri.html.