AGENT-BASED INTEGRATION OF SENSOR NETWORKS, REMOTE SENSING, AND SIMULATION MODELS

Pereira, Gary M. Department of Geography, San Jose State University, San Jose, CA.

Keywords: agent-based modeling, sensor networks, multi-agent simulation, DDDAS

Understanding geographical and environmental phenomena requires the ability to observe, represent, and model the dynamics of entities and processes at widely varying scales of space and time. Spatially localized sensors and sensor webs can provide information about hundreds of physical, chemical, and biological properties at many time scales for specific locations or mobile entities. Remote sensing and other imaging technologies augment localized sensor data with space-filling observation of spectral and terrain data at generally coarser spatial and temporal resolutions. Computational technologies enable the creation of models of geospatial processes and entities that are used for discovery, analysis, forecast, and decision. These various technologies are brought together when different sensors are used to corroborate, complement, and enhance one another and to calibrate, initialize, populate, and validate models. Increasingly, however, opportunities exist for a mutual communication between models and sensors, with models providing a knowledge base for improvements to the gathering of data, and sensors providing real time grounding and guidance in the evolution of models. Implementation of this sort of mutual integration presents significant conceptual and computational challenges. Protocols range from fully centralized control to more distributed, complex negotiation. Decentralization offers greater flexibility and enhances opportunities for system evolution, but it requires more robust representational structures and computational protocols than the more traditional centralized methods. Extensions to the agent-based modeling paradigm can offer a reasonable foundation from which to address these problems. This report describes a geospatial agent-based methodology for the distributed integration of data sources and models.

Environmental Sensors, Sensor Networks, and Remote Sensing

A comprehensive description of remote sensing and localized Earth sensing technologies is not possible here, but several general developments should be mentioned. Increasingly, remote sensing data of various kinds and at various scales are being integrated, along with elevation and other cartographic data, in processes of data fusion to provide enhanced insight into Earth processes. Many new specialized remote and localized instruments have been deployed that provide a wider perception of atmospheric, terrestrial, and aquatic properties than had previously been attainable. The calibration and processing of remote sensing data with data gathered from localized sensors, flux towers, and mobile field instrumentation has resulted in the derivation of process-relevant, space-filling data sets from the raw data using statistical and physical models. Remote sensing is even being used to infer the presence of conditions suitable for the entities that most remote sensing instruments themselves cannot possible detect directly. For example, remote sensing data are being used as proxy sources for examining the distribution and diversity of animals (Leyequien et al., 2007) and vector-borne diseases (De La Rocque et al., 2004; Tatem et al., 2004) that lie far below the resolving power of the instruments themselves.

Localized Earth sensors are unobtrusively and reliably monitoring a wider variety of conditions than ever before. A partial list of atmospheric variables includes air flow rate and direction, temperature, humidity, light levels, and concentrations of CO, NO, NO_2, O_3, and other gases, aerosols and particulates. Hydrological variables include stream flow rates; the depth, temperature, salinity, pH, conductivity, pressure, and chemical, biological, and optical qualities of water bodies; the height, period, direction, and speed of waves; ocean acoustics; snow accumulation and melting; and glacial dynamics. Geophysical variables monitored by localized sensors include mass movement and seismology. Imaging sensors monitor local activities in the visible and infrared. Webcams are increasingly used to monitor hydrological, atmospheric, and ecological processes. Digital cameras triggered by motion and heat monitor animal activity. GPS-enabled sensors attached to mobile organisms or embedded in currents of air and water log or transmit locational data along with other variables. Robotic devices provide additional spatial versatility and the ability to gather data from locations that are inaccessible or hostile to humans. While many sensors maintain digital data logs on site for later retrieval, others are in periodic or continuous telemetric communication with users. If this communication is bidirectional, users may control the data collection process remotely. The sampling rate of many sensors can be changed to match the temporal scales of current processes, and the view of imaging sensors can be changed. Increasing attention has focused on the use of spatially distributed networks of sensors, actuators, communications, and decision nodes for real-time monitoring of Earth processes. Various approaches have been described within the context of ‘pervasive computing’ (Estrin et al., 2002) and ‘wireless sensor networks’ (Hart and Martinez, 2006; Porter et al., 2005). These approaches are enabled by several rapidly developing technologies, including microelectromechanical systems (MEMS), wireless integrated network sensors (WINS), embedded GPS receivers, and wearable and embedded computing systems. The hallmark of these approaches is the ability of each node to both transmit and receive information, often wirelessly, and to communicate with one another to form flexible routing networks to distant servers. Attention has focused on ‘nanoclients’ (Clarke, 2003), although miniaturization is not a requirement for many Earth monitoring processes (Hart and Martinez, 2006).

Sensor Webs

The comprehensive integration of a variety of localized and remote sensing instruments within a single communication and control system has been described in terms of a ‘Sensor Web’ (Liang et al., 2005). Others use terms like ‘environmental sensor network’ to describe such integrated systems of heterogeneous components (e.g., Hart and Martinez, 2006). We will adapt the term Sensor Web, reserving the term ‘sensor network’ to describe communication networks of similar components, as described in the previous section. This latter term is already well established, and ‘wireless sensor networks’ in particular have a long history of specialized development along these lines (e.g., Zhou and Guibas, 2004). The goal of the Sensor Web is to provide a shared communication and information layer for a variety of sensing technologies: airborne and satellite remote sensing, traditional scientific sensors, off-the-shelf sensors like video cameras, microclients, and sensor networks. The natural infrastructure of a Sensor Web is the World Wide Web; many implemented 'environmental sensor networks' listed by Hart and Martinez (2006) use the Internet to transmit a variety of data to central servers. These servers in turn use the Internet to broadcast their results to the world.

Agent-based Modeling and Simulation

The advent of software agents is a natural result of the structured encapsulation of both attributes and functions afforded by object-oriented high-level languages. The general properties of agents include the ability to engage in autonomous behavior, to sense and act upon other agents and their computational environment, and to engage in individualized reasoning. Agent-based programming and design are often used to solve computational tasks like network search, adaptive communication and control, and functional optimization, with few representational requirements. However, they are also being used to simulate economic, social, political, ecological, geographical, environmental, and other real world distributed phenomena, often in the form of a geospatial 'agent based model', or ABM (Epstein, 1999; Kohler and Gumerman, 2000; O'Sullivan and Haklay, 2000; Hare and Deadman, 2004; Goldstone and Janssen, 2005; Bithell, 2006; Breckling et al., 2006; Brown and Xie, 2006). These models are generally implemented in object-oriented languages, and several specialized software packages are now available for the development of ABMs, including Ascape (Brookings Institute, 2007), Swarm (Swarm Development Group, 2007), RePast (ROAD, 2007), and NetLogo (Wilensky, 2007). In an ABM, individual organisms, people, and other geospatial, social, or ecological entities are represented as agents whose behaviors are influenced by their own internal processes, environmental conditions, and the behaviors of other agents. Since they exemplify all the advantages of object-oriented design, agents can be brought into and out of existence and may inherit characteristics of parent agents or of hierarchical classes, an advantage when they are used to represent individual members of different organizations, types, breeds, species, or functional groups.

In an ABM, mobile entities can be represented and tracked in a straightforward way (e.g., Westervelt and Hopkins, 1999). Continuous landscapes are often modeled as space-filling tessellations of areal raster cells or polygons. If each such cell is modeled as an agent, it may respond to and act upon external factors, other cells, and other stationary or mobile agents. This affords an ABM somewhat more flexibility than comparable cellular automata models. There is no fundamental restriction on the geometry of agents, so they may be modeled as points, lines, polygons, cells, or links and nodes of networks. Agents may be generally spatial without a specific geometry (e.g., changing climatic conditions that are spatially homogeneous over the extent of the modeled region), and they may be only vaguely spatial or nonspatial (e.g., executive entities and external observers). A geospatial ABM that performs temporal simulations, or geosimulations, is often termed a 'multi-agent simulation' (MAS). Individual-based models (IBMs), which have been used for some time in biological and ecological research (Grimm and Railsback, 2005; De

Angelis and Mooij, 2005), may be considered to be a type of ABM, although the internal processing of an individual in such models is often more limited than that of an agent. All of this terminology remains in flux, and the literature is rich in competing definitions. It is not necessary to sort it all out here; my only point is that many different approaches to the distributed modeling of individuals or agents have proven successful and enlightening in a variety of domains. ABMs have increasingly been coupled with or embedded within Geographic Information Systems, which provide extensive geodatabase support and computational functionality (Gimblett, 2002; Brown et al., 2005a).

When ABMs model real world phenomena, they are usually initialized with representations of environmental conditions and constraints, as well as agent attributes and reasoning methods. The environmental conditions take the form of variable and parameter values that are ultimately derived from remote and localized sensors. For example, remote sensing data might provide, with minimal processing, the vegetation environment within which mobile herbivore agents roam. Within the model, the vegetation environment can be represented as a tessellation of agents, each cell responding to current conditions within itself, its neighbors, and the atmosphere to generate its dynamic growth pattern; this feature is common in cellular automata models as well. Within an ABM, each mobile herbivore internally computes and acts upon decisions to find and consume vegetation in its perceived environment, in order to garner energy for metabolic, locomotive, defensive, and reproductive tasks. An extended ecosystem of multiple vegetative, herbivore, and carnivore species might thereby depend fundamentally on the initialization provided by the remote and localized sensing data.

The output of any simulation model may be compared with temporal or spatiotemporal sets of previously available data, but they usually operate in isolation from external sources of new real-time data. There are unavoidable problems with any validation process. Incompatible models may seem to be valid even if they operate on the basis of entirely different assumptions, exemplifying the problem of equifinality. If only a single final point of temporal comparison is available, a model may appear to be valid even if the interim dynamics are not checked. A model may appear to be invalid even if the dynamics are realistically modeled, if the domain dynamics are chaotic or subject to random influences, path-dependence, and lock-in. But if a model is communicating with external data, as all living things do, then it has an endless source of information with which to improve, adapt, or even evolve, in changing environments. Such a model would be useful in time-critical situations like disaster assessment and management. A dynamic validation strategy using continuously available sensor data may even help to assess a model of chaotic processes, which are accessible to short-term prediction.

Dynamic Data Driven Application Systems

The 'dynamic data driven application system' (DDDAS) concept allows sensory data to dynamically validate and modify models, while also allowing simulation and decision models to dynamically drive sensing strategies (Ouyang et al., 2007). Spurred by the failure of weather models to accurately predict the track and magnitude of a major snowstorm in March 2000, and of a fire simulation to model the propagation of a fire near the Los Alamos National Laboratory in May 2000, the National Science Foundation began funding new ways to improve model performance based on dynamically changing data (NSF, 2000). In a DDDAS, sensors might be controlled from within the model to change sampling rates, turn different sensors on or off, and dynamically allocate communication channels, data storage, and other resources. Earth sensing and modeling DDDAS are already in advanced stages of development (e.g., Patrikalakis et al., 2004; NSF, 2006).

If both observed and simulated phenomena are allowed to interact with one another in a dynamic spatiotemporal model, representations of the 'real' and the 'simulated' become entangled. Simulated changes in the representation of observed phenomena must be distinguished from their initial values. If further periodic observation is available, then both the observed and simulated dynamics can monitored, and divergences can be used to improve both the model and the data gathering strategy dynamically. An example of how this might work in a system of centralized control is shown in Figure 1.

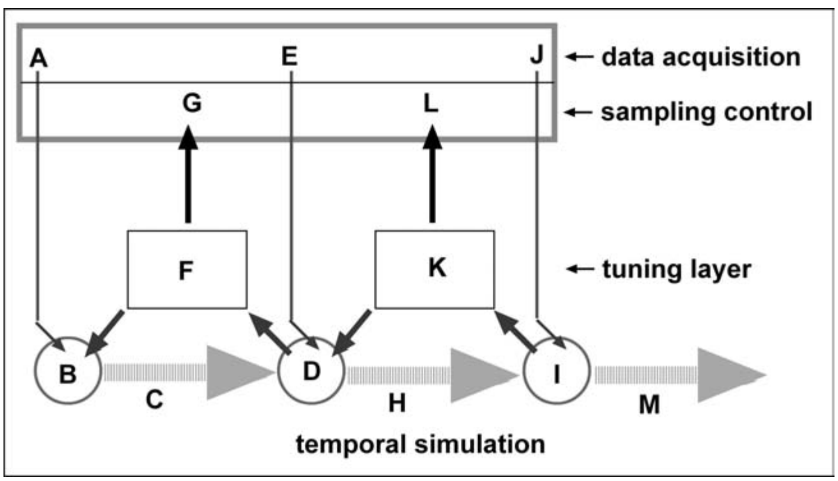

Figure 1. Scheduling sequence of a DDDAS for adaptive sampling and simulation.

In this hypothetical sequence, we begin with data (A) from a sensor initializing the simulation (B). The model iterates a number of times to produce a simulation sequence (C), pausing at checkpoint (D). At this point, the simulated values are compared with sensor values (E) corresponding to the same temporal sequence. If the values correspond well enough, the simulation proceeds (H). If the values do not correspond, then an adaptive tuning mechanism (F) attempts to reconcile the real and simulated dynamics by modifying model parameters or structure. It also may instruct the sensor to change its sampling strategy (G), to better inform the validation process or provide more informative data. The D-F-B-C sequence can occur repeatedly as required before moving on with the simulation, and the whole process repeated further on (e.g., I-K-D-H). Data buffering and storage would serve to synchronize the real and simulated time. Even in this rough conceptual sequence with a single data source, it is apparent that an application system of this kind would have to satisfy a number of difficult scheduling requirements. The question presents itself: would such a top-down strategy really be the most appropriate way to implement a DDDAS, or might a bottom-up, distributed strategy of negotiation between independent computational agents provide better results?

DDDMAS: Removing Barriers between the Model and the World

Given the advantages of agent-based distributed modeling for disaggregated, dynamic entities, as well as the evident difficulties of centrally managing the dynamics of an environmental DDDAS, it would seem that an agent-based approach to the integration of sensors and models might provide a less complex basis for more adaptive scheduling. Instead of introducing a processing layer between sensors and models, complex pathways of control and communication might be more easily managed if all simulated entities, sensors, observers, evaluation, and decision are implemented as agents. Some of these agents are representational of simulated entities and real sources of external data, while others perform computational functions and represent aspects of the system itself. The use of 'geo-agents' for performing computational functions in distributed GIS was described in the first Digital Earth symposium (Luo et al., 1999; Wu et al., 1999), and agent-based methods of organizing and processing geospatial information for purposes other than simulation modeling have subsequently been described elsewhere (e.g., Rahimi et al., 2002). The ability of agent-based computation to represent entities in a simulation as well as perform tasks internal to the system suggests that agent-based approaches should provide a comprehensive basis for implementing DDDAS.

In an agent-based model (or more appropriately a 'multi-agent system', since the model itself is only part of the entire system) that implements a DDDAS, the software system used to implement it (e.g., NetLogo or RePast) provides the high-level scheduling of the functional dynamics of all such agents. Each agent decides its own schedule, based on the characteristic temporal scale of the entity or process it represents, as well as the current state of other agents (Figure 2). In order to acknowledge that the sort of system proposed here is a derivation of both DDDAS and MAS, it might be termed 'dynamic data driven multi-agent simulation', or DDDMAS.

A situation illustrating the utility of this approach for anticipatory and proactive data collection in conservation biology is portrayed in Figure 3. Here we have a landscape representing the major spatial influences on caribou migration routes – in this case, the digital elevation model is portrayed. The circular forms represent the sensing radius of individual ground vibration sensors tuned to detect the signature of passing caribou. These sensors form a wireless network providing real-time communication of data into the model. The small black circle represents the site of remotely controlled video cameras on a hilltop. These cameras are in bidirectional communication with the model, and their viewsheds can be controlled to view multiple possible migration routes. Caribou are simulated as mobile agents, and they each behave in accordance with certain behavioral rules (e.g., a general migratory direction, a preference for low to medium slopes and edible vegetation, and a tendency to herd together with other caribou). Some diversity or stochastic variability among individual animals results in the simulation of multiple possible routes through the landscape. Data from the ground sensors inform the model.

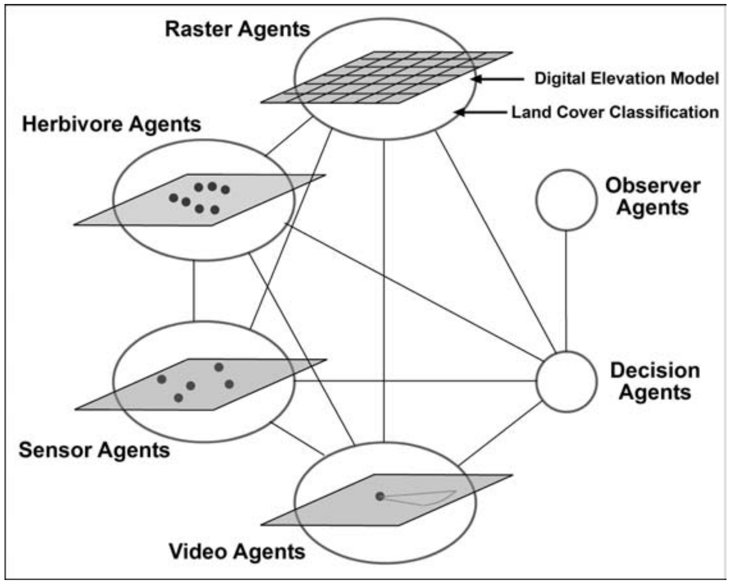

Figure 2. Schematic description of a DDDMAS. Oval agent types are spatial. Circular agent types are nonspatial. Sensor and video agents are portals to current external data. Raster agents are portals to remote sensing and other space-filling data. A global scheduler (not portrayed) provides primary synchronization. Links describe potential bidirectional communication between agents of different types.

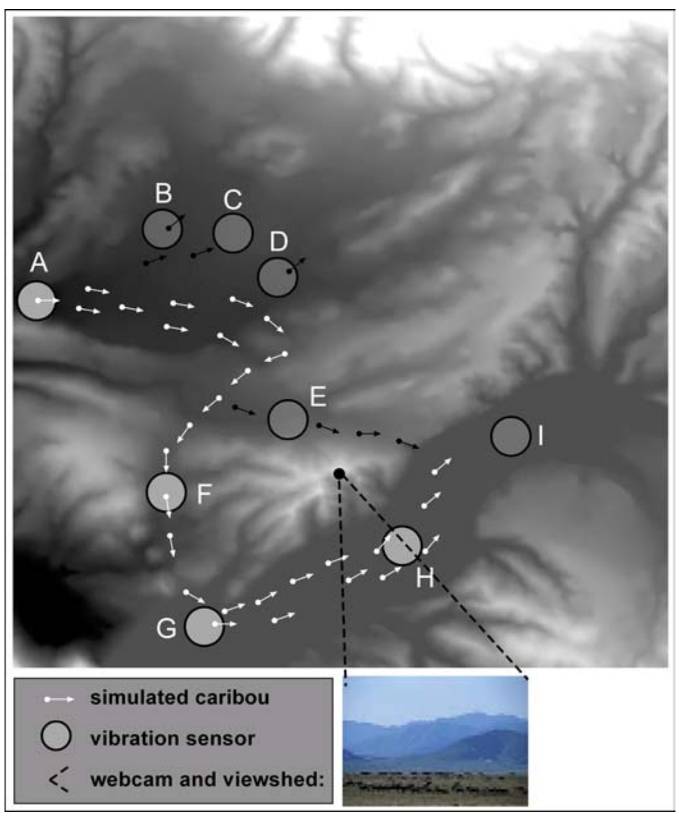

Figure 3. Agent-based adaptive modeling and observation of caribou migration, deliberately underpopulated for visual clarity.

In this particular case, sensor A signals the passage of caribou, triggering the model to simulate multiple possible routes through the portrayed portion of the landscape, with the general migratory direction from left to right. Sensors B, C, D, and E do not subsequently indicate the presence of caribou, but sensor F does. The model therefore removes simulated caribou agents (indicated in black) that do not conform to the incoming data, and validates those that pass sensor F, as reinforced by sensor G (these agents are indicated in white). The simulation suggests a route that passes through one of the possible viewsheds of the video cameras, and subsequently directs the cameras to increase their frame rates and adjust directionality, in anticipation of caribou. Cameras that cannot view this route might be turned off, in order to provide greater bandwidth for the cameras engaged by the model. If the system performs properly, the caribou are observed before they trigger ground sensor H, which provides additional validation. The richness of data provided by the video cameras, possibly aided by human users or vision algorithms, can then inform the model regarding numbers of individuals as well as their spatiotemporal behavior, possibly modifying the behavioral rule base of caribou agents in the model.

Prospects for DDDMAS

Comprehensive networks of perception at all geographical scales, as well as inductive detection and representation of patterns with the aid of computational models, are beginning to improve our scientific understanding of natural and social phenomena. Cellular automata and agent-based models in particular, by representing knowledge about entities at their characteristic, disaggregated scales, reveal the emergence of structures and dynamics at wider scales. This has led to strategies that seek to match patterns in the observed world and emergent patterns in models (Grimm et al., 2005). The complexity and diversity of the real world remain challenging for modelers, who try to be parsimonious in their representations. Computational models are not often successfully predictive, but if they yield results that are somehow similar to observations, then they may be useful in revealing structural and dynamical processes that help to explain the real world. This trend can be seen for example in the use of agent-based models of land cover change (Parker et al., 2005). However, path-dependence remains troubling for modelers, since the contingent particularities on the ground and in the models may result in widely varying results over time (e.g., Brown et al., 2005b). DDDMAS, in providing real-time correction of model evolution, can help overcome the problem of path dependence.

If models are electronically integrated with a continuous perception of the world, opportunities present themselves for automated model adaptation and evolution. Evolutionary algorithms, swarm optimization, and other methods are being explored in the context of distributed modeling. The data that such methods are applied to are usually assumed to be static and stable, but if the available data change over time, then the nature of inference and decision changes. Such a development in thought, decision, and planning would provide greater sensitivity to the real nature of the world, and provide a better foundation for ecologically sound, sustainable policies in the future (Rammel, van den Bergh, 2003).

References

Bithell, m., Brasington, J., and Richards, K. (2006). Discrete-element, individual-based and agent-based models: Tools for interdisciplinary enquiry in geography? *Geoforum* in press.

Breckling, B., Middelhoff, U., Reuter, H. (2006). Individual-based models as tools for ecological theory and application: Understanding the emergence of organisational properties in ecological systems. *Ecological Modelling* **194**, 102-113.

Brookings Institution. (2007). Ascape 1.9 (Software). http://www.brook.edu/es/dynamics/mode/s/ascape

Brown, D. G., Riolo, R., Robinson, D. T., North, M., Rand, W. (2005a). Spatial process and data models: Toward integration of agent-based models and GIS. *Journal of Geographical Systems* **7**, 25-47.

Brown, D. G., Page, S., Riolo, R., Zellner, M., Rand, W. (2005b). Path dependence and the validation of agent-based spatial models of land-use. *International Journal of Geographical Information Science* **19**, 153-174.

Brown, D. G., Xie, Y. (2006). Spatial Agent-based Modelling. *International Journal of Geographical Information Science* **20**, 941-943.

Clarke, K. C. (2003). Geocomputation's future at the extremes: high performance computing and nanoclients. *Parallel Computing* **29**, 1281-1295.

De Angelis, D.L. and Mooij, W.M. (2005). Individual-based modeling of ecological and evolutionary processes. *Annual Review of Ecology and Systematics* **36**, 147–168.

De La Rocque, S., Michel, V., Plazanet, D., Pin, R. (2004). Remote sensing and epidemiology: examples of applications for two vector-borne diseases. *Comparative Immunology, Microbiology and Infectious Diseases* **27**, 331-341.

Epstein, J. M. (1999). Agent-Based Computational Models and Generative Social Science. *Complexity* **4**, 41-60.

Estrin, D., Culler, D., Pister, K., Sukhatme, G. (2002). Connecting the Physical World with Pervasive Networks. *IEEE Transactions on Pervasive Computing* **1**, 59-69.

Gimblett, R. H. (2002) *Integrating Geographic Information Systems and Agent-Based Modeling Techniques for Simulating Social and Ecological Processes.* Oxford University Press, New York.

Goldstone, R. L., Janssen, M. A. (2005). Computational models of collective behavior. *Trends in Cognitive Sciences* **9**, 424-430.

Grimm, V. et al. (2005). Pattern-Oriented Modeling of Agent-Based Complex Systems: Lessons from Ecology. *Science* **310**, 987-991.

Grimm, V., Railsback, S.F. (2005). *Individual-based Modeling and Ecology*. Princeton University Press, Princeton.

Hare, M., Deadman, P. (2004). Further towards a taxonomy of agent-based simulation models in environmental management. *Mathematics in Computers and Simulation* **64**, 25-40.

Hart, J. K., Martinez, K. (2006). Environmental Sensor Networks: A revolution in the earth system science? *Earth-Science Reviews* **78**, 177-191.

Kohler, T.A., Gumerman, G.J. (2000). *Dynamics of Human and Primate Societies: Agent-based Modeling of Social and Spatial Processes*. Oxford University Press, Oxford, pp. 5–76.

Liang, S., Croitoru, A., Tao, C. V. (2005). A distributed geospatial infrastructure for Sensor Web. *Computers & Geosciences* **31**, 221-231.
Leyequien, E., Verrelst, J., Slot, M., Schaepman-Strub, G., Heitkönig, I., Skidmore, A. (2007). Capturing the fugitive: Applying remote sensing to terrestrial animal distribution and diversity. *International Journal of Applied Earth Observation and Geoinformation* **9**, 1-20.
Luo, Y., Wang, X., Cong, S., Zhang, N., Li, M., Xu, Z., Lu, Z. (1999). A Study on the Model of Agent-based Distributed GIS. In: *Proceedings of the First International Symposium on Digital Earth*, Beijing, China.
NSF. (2000). Dynamic data driven application systems. National Science Foundation, USA. Workshop report. http://www.nsf.gov/cns/dddas/dd.das.work_shop_rprt.doc.
NSF. (2006). Dynamic data driven application systems. National Science Foundation, USA. Workshop report. http://www.dddas.org/NSFWorkshop2006.html
O'Sullivan, D., Haklay, M. (2000). Agent-based models and individualism: is the world agent-based. *Environment & Planning A* **32**, 1409–1425.
Ouyang, Y., Zhang, J. E., and Luo, S. M. (2007). Dynamic data driven application system: Recent development and future perspective. *Ecological Modelling*, in press.
Parker, D. C., Manson, S. M., Janssen, M A., Hoffmann, M. J., Deadman, P. (2002). Multi-agent system models for the simulation of land-use and land-cover change: A review. *Annals of the Association of American Geographers* **93**, 314-337.
Patrikalakis, N. M., McCarthy, J. J., et al. (2004). Towards a Dynamic Data Driven System for Rapid Adaptive Interdisciplinary Ocean Forecasting. In *Dynamic Data Driven Application Systems*, F. Darema, ed., Kluwer Academic Publishers, Netherlands.
Porter, J., Arzberger, P., Braun, H., et al. (2005). Wireless Sensor Networks for Ecology. *BioScience* **55**, 561-572.
Rahimi, S., Cobb, M., Ali, D., Paprzycki, M., Petry, F. (2002). A knowledge-based multi-agent system for geospatial data conflation. *Journal of Geographic Information and Decision Analysis* **6**, 67-81.
Rammel, C., van den Bergh, J., 2003, Evolutionary policies for sustainable development: adaptive flexibility and risk minimizing. *Ecological Economics* 47, 121-133.
ROAD. (2007). RePast S (Software). RePast Organization for Architecture and Design, Chicago, IL. http://repast.sourceforge.net.
Swarm Development Group. (2007). Swarm (Software). http://www.swarm.org.
Tatem, A. J., Goetz, S. J., Hay, S. I. (2004). Terra and Aqua: new data for epidemiology and public health. *International Journal of Applied Earth Observation and Geoinformation* **6**, 33-46.
Westervelt, J.D., Hopkins, L.D. (1999) Modeling mobile individuals in dynamic landscapes, *International Journal of Geographical Information Science* **13**, 191-208.
Wilensky, U., 2007. NetLogo 3.1.4. (Software). http://ccl.northwestern.edu/netlogo/. Center for Connected Learning and Computer-Based Modeling, Northwestern University, Evanston, IL.
Wu, S., Huang, X., and Li, Q. (1999). GeoAgents. In: *Proceedings of the First International Symposium on Digital Earth*, Beijing, China.
Zhou, F., Guibas, L. (2004). *Wireless Sensor Networks: An Information Processing Approach*. Morgan Kaufmann Publishers, San Francisco, CA.