

# Shared Data Model to Support Environment Sensor Network Data in Hydrologic Models

**M. Kumar<sup>a</sup>, and C.J. Duffy<sup>a</sup>**

<sup>a</sup> *Civil and Environmental Engineering, Pennsylvania State University, University Park,  
Pennsylvania, USA (muk139/cxd11@psu.edu)*

**Abstract:** With the advancement of wireless communication and miniaturization of digital electronics, long term observation of remote hydrologic systems using adaptive sensor networks at high spatio-temporal resolutions and across multiple scales, has become a reality. However, for large spatial scales embedded multi-sensor networks with fine temporal sampling rates, the amount and distribution of data generated by these networks becomes unmanageably large. While the sensor network installation itself is generally supported by basic data management software, in the hydrologic sciences there is little support available to directly incorporate the data generated into the hydrologic model. We contend that a seamless transfer of the observed data to the model can be achieved by developing a shared data model which will standardize storage and management of data both at the sensor base station and the hydrologic model. This will lead to enhanced data transfer integrity and will also result in direct input of the sensor network data to the model in realtime without having to go through intermediate pre-processing steps which are error prone. Here we present the shared Data Model structure along with its design considerations in terms of data types, identification of data-classes, relationships and constraints.

**Keywords:** Sensor Networks; Hydrologic Model; Data Model; CZO-Net.

## 1. INTRODUCTION

In order to advance our understanding of multiscale coupling of hydrologic processes, new observation systems that capture the spatio-temporal dynamics need to be designed. The motivation is improved predictability of the terrestrial water cycle as well as addressing the problem of closing water, energy and solute budgets. Such an observing system can be expected to perform synergistic measurements of the atmosphere (e.g. water vapor, winds, thermodynamics, cloud-radiative forcing, and precipitation), the near-surface (e.g. surface exchange fluxes of heat, momentum and moisture, including transpiration, along with radiation balances, vegetation dynamics, precipitation, and runoff), and the subsurface (e.g. soil moisture, temperature, pressure profiles, water table and baseflow). The installation of such an observing system requires the explicit intersection of the terrestrial scales associated with hillslopes, watersheds, and river basins, with ecological regions, estuaries, subsurface linkages, and meso-scale weather.

A modern observing platform, comprised of distributed “intelligent” sensors with small, automatic, low-cost, energy-efficient, non-invasive, computationally-capable, and communicative sensor nodes, is able to collect long term data from remote locations at scales and resolutions. These systems are already being used in a variety of environmental monitoring applications [Cerpa et. al., 2001; Cardell et. al., 2005, Hart and Martinez, 2006]. On the one hand, each sensor node and type provides a localized measurement of the hydrologic states (much like observations from data loggers), the network reveals information that is more than sum of its parts, since it is able to measure distributed heterogeneity, localized anisotropy and spatially derived variables (e.g. fluxes) by virtue of its node topology. These networks are flexible and robust due to adaptivity of the individual nodes in their work assignment and communication topology in response to changes in

environment conditions (e.g. events), health of the network (e.g. node failures) and project needs (individual nodes are mobile). The network can be designed in different configurations that recognize the natural landscape boundaries and scales where the atmosphere, vegetation and subsurface partitions interact.

Depending on the node density, coverage area, number of states observed (sensors) at each node, and the sampling rate, the amount of heterogeneous data generated can be very large. Typically, the observed data generated by the sensor network will be used in a numerical model as a parameter, forcing, initial state, or as a validation set. This necessitates intensive data development, organization and topological definition of the multi-sensor data vis-à-vis a hydrologic model discretization grid. A seamless transfer of data directly from the sensor network to the model grid can be achieved by the development of a shared “data model” that will provide a standard structure for storage, sharing and exchange of data independent of the software environment and programming languages [McKinney and Cai, 2002].

In this work we discuss the details of a “shared” data model that can be used to directly assimilate observed data generated by the sensor network with the hydrologic model. To guide this discussion we use a sensor network under development at Shale Hills and Shavers Creek Watershed in central Pennsylvania referred to as CZO\_Net. The network employs Crossbow motes® to map hydroclimatic variables such as temperature (atmosphere/ground), relative humidity, incoming solar radiation, barometric pressure, wind speed, wind direction, soil moisture, soil matric potential, and groundwater level. Before discussing the data structure at each nodes and its object oriented classification, the architectural framework of the sensor network is developed.

## 2. ARCHITECTURAL FRAMEWORK OF CZO-NET

The architecture of the network is three “tiered”. At the lowest and least power-intensive layer lie the sensor nodes or “motes”. The sensor nodes observe states in its immediate vicinity and communicate the stored data after “limited” signal preprocessing to the neighboring nodes. The nodes are composed of four primary components [Raghunath et. al. 2002] viz. a) a microcomputer that supports a processor, a memory unit and a controller to

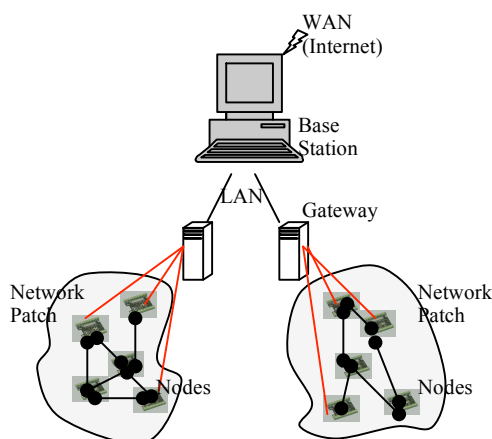


Figure 1: 3-tier sensor network architectural framework

execute power scheduling and communication protocols, b) a transceiver that is essentially a short range radio to transmit and receive data from/to other nodes and gateways, c) a sensing hardware that is a collection of sensors that measures a set of state variables in the immediate vicinity and d) a power supply which can be a battery or a solar/wind power scavenger. The sensor nodes are often deployed in localized cluster patches which communicate between themselves through a gateway. The gateway periodically downloads the data to the remote base station database server through local area network (LAN) connectivity. The logged information is disseminated over the web from the base station over wide area network (WAN). Figure 1 shows the sensor network framework.

### 3. DESIGN CONFIGURATIONS

A generic sensor network design is determined by architectural factors such as fault tolerance; scalability; production costs; sensor network topology; hardware constraints; transmission media; and power consumption [Akyildiz et. al., 1999]. For purposes of hydrologic research in addition to the architectural limitations, the network design would be driven by the science goals. We use the watershed as the organizing principal at the regional scale. Physical properties of the watershed (topography, slope, hydrogeologic and landuse/landcover heterogeneity) and hydrologic process dynamics are the basis for sensor deployment. At each node, sensor systems are deployed in 3D domain, which extends from the base of active groundwater circulation, through the soil, vegetation and the top of the atmospheric boundary layer. At any particular nodal location on the land surface, topology of the sensors is designed to capture the direction and magnitude of boundary fluxes across the faces of a 3D control volume. We note that the numerical model also evaluates the same interfacial fluxes by forming semi-discrete balance equations over a unit discretized domain obtained by integration of coupled process differential equations over the projected control volumes. In this way the model serves as both a conceptual tool that explicitly defines the particular interface for which the instrument should measure the flux, and as a constraint on the overall energy/moisture budget itself. The basic instrument configuration is illustrated in Figure 2, and includes a “whole canopy” micrometeorological tower configuration, boundary-layer profilers, as well as surface, soil and groundwater observations. The flexible design is be able to take advantage of the natural scales of motion for water, energy, and should be adaptable to most physiographic and climatic settings.

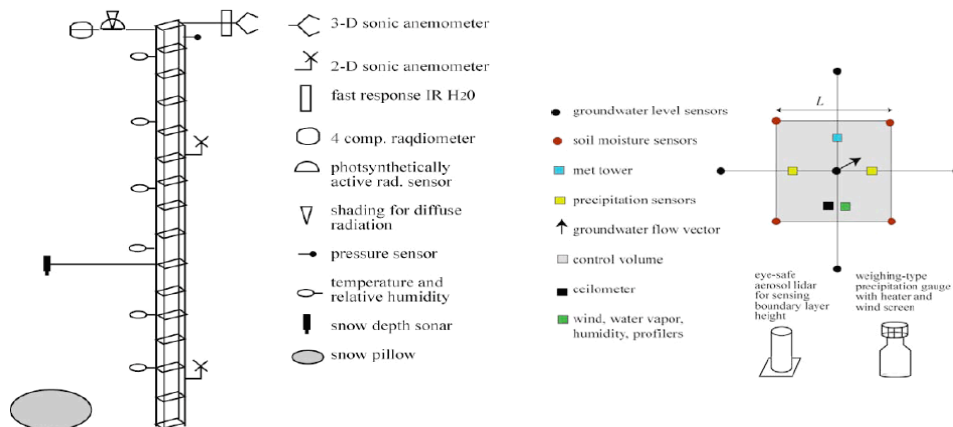


Figure 2: Flux Tower (left) and Subsurface and land-surface instrumentation (right) that constitutes a local sensor array

At the watershed scale, process interactions, hydrogeologic and climatic heterogeneities vary spatially, and thus the placement of sensors would be designed to capture the gradient variability of phenomena of interest.

One approach to an optimal sensor deployment capable of heterogeneous sampling in localized region of the watershed uses Delaunay triangulation. Figure 3 shows three different potential configurations of the sensor network design where sensor node placements reflect a) the boundary between characteristic hydrodynamic descriptors like hypsometry/ vegetation/ soil property, b) a nested local zone of interest, and c) a new state such as a river. One representative example of the latter case is measurement of stream temperature [Troch, 2008] in Valles Caldera, NM. We note that a higher nodal density in any of the shown sensor mesh configurations can be hierarchically obtained by application of “incremental-insertion” algorithm [Lawson, 1977], assuming that the rest of the architectural constraints are satisfied.

We assume that any site chosen would have a completed (or anticipated) digital watershed survey available for soils, geology, vegetation, high resolution topography, in addition to the hydroclimatic database.

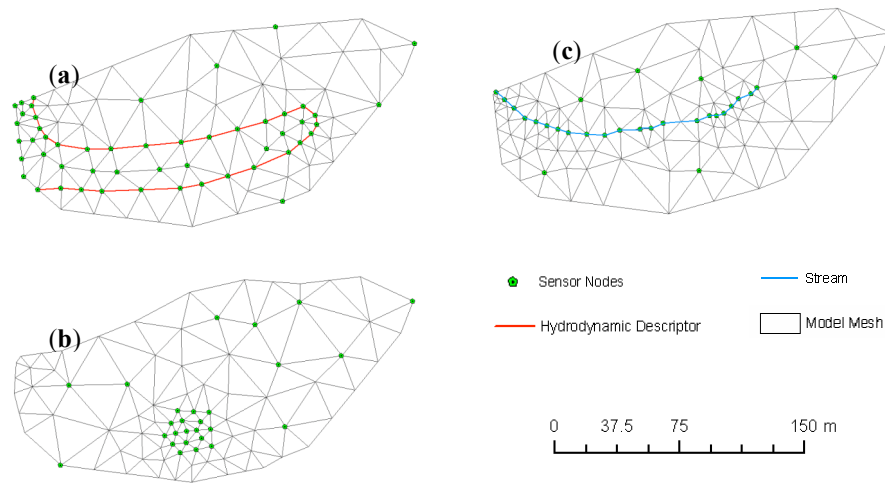


Figure 3: Three different sensor network configurations mapped on Shalehills Watershed

#### 4. SENSOR NETWORK DATA MODEL

The first prerequisite to optimal data model design is accurate assessment of all types of data and data formats. As shown in Figure 2, each sensor node measures a range of data types including precipitation, net solar radiation, wind speed, relative humidity, air and ground temperature, soil moisture and matric potential. Nonetheless, all the observations are essentially time series. We also note that for some data types, multiple observations are needed at precise separation in order to calculate derived fluxes (such as for ground water head and soil moisture) or for uncertainty estimation (in case of precipitation). The topology of the network is mapped by tracking its neighbors. The sensor network data model is shown in Figure 4. The designed data model follows the standard object oriented representation in UML 2.0. Data types are first organized into different classes. The classes interact with each other through standardized relationship definitions. The advantages of using this strategy are the potential to incrementally enrich the data model, the ability to construct complex objects (extensibility), robustness, and adaptability to changing hydrologic conditions by using different instances of a single object (reusability), and by using the constructs of inheritance, polymorphism and encapsulation [McKinney and Cai, 2002].

Generalization relationship between any two classes means that one of the classes (Child class) is derived from the other (Base class). This relationship is inherent to object-oriented modeling through the “inheritance” mechanism. This relationship markedly simplifies and clarifies the data model and minimizes redundancy in definitions, access and storage. Generalization is denoted by a solid line with a closed arrowhead pointing to the super class. Figure 4 shows that Solar Radiation, Precipitation, Temperature etc. inherit the properties of “Time Series” class.

Association is the most common relationship in a class diagram. Associations can connect classes both in time and in space. They are denoted by an optional arrowhead on one end of the line. An Association linkage without an arrowhead is a bi-directional Association, which means that both of the connecting classes are aware of the relationship. Single ended arrowhead relationships are unidirectional Associations that link the classes in which only one knows about the relationship. The class from which the arrow invocation emerges is the class which has knowledge of the relationship. One other type of association that has been implemented in the developed data model is Reflexive association. This

linkage represents the association of the class to itself. This essentially means that another instance of class is associated with the present one. We note from Figure 4 that each Time Series class is associated to a “Sensor Node” class which is essentially the location at which it is observed.

Aggregation relationships explain the interaction of individual parts/components (Simple Objects) to a Complex Object. The relationship is denoted by a white diamond (for the Aggregate class) on one end of the link and arrow (for the “part” class) on the other . Sensor Node Aggregate to form Sensor Patch.

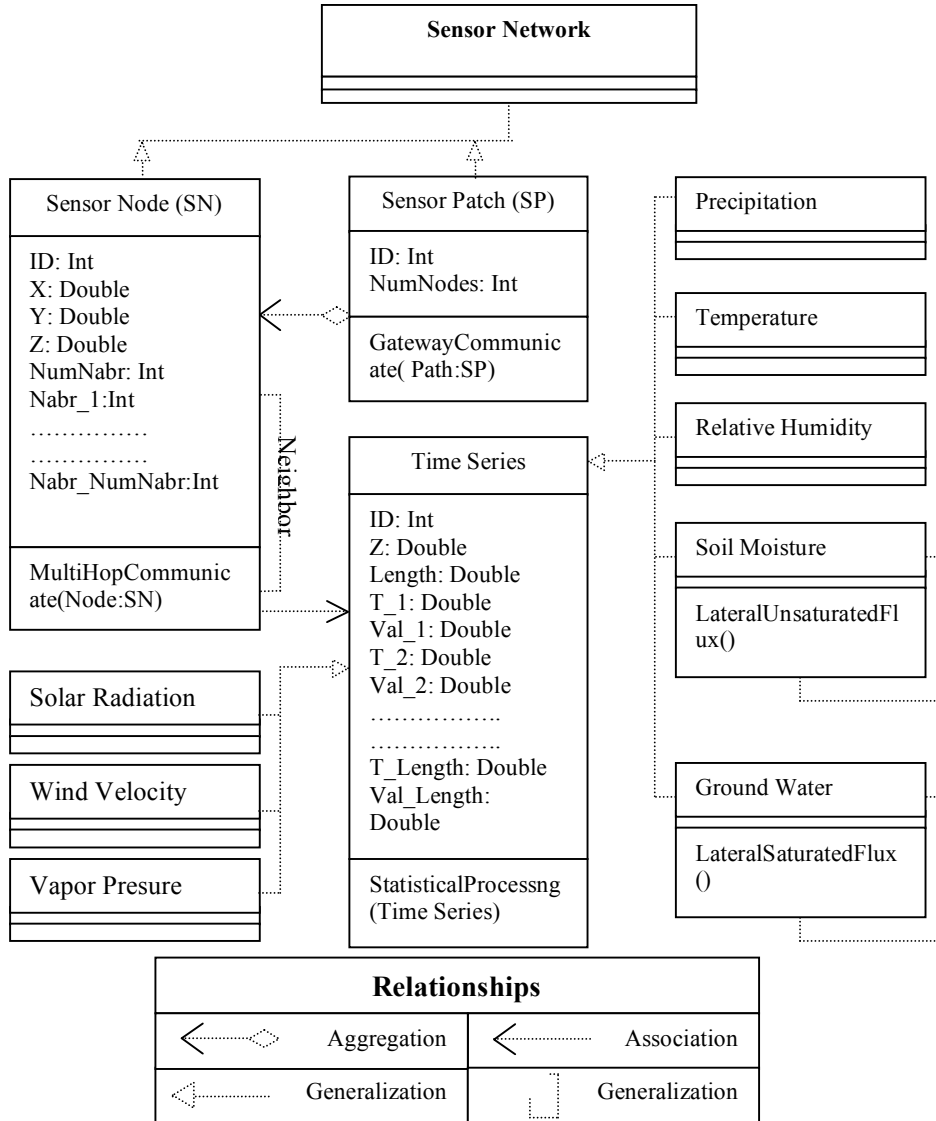


Fig. 4: Sensor Network Data Model in UML 2.0. Note that all the hydroclimatic data measured at the sensor node is “associated” to it. The operators in the bottom compartment for each individual class are basic schema processing that are either possible at the node itself or at the base station.

The operation that is carried out on each Class Object is shown in the lowest compartment. Operations such as SignalProcessing() on each Time Series is carried out before the sensor nodes use MultiHopCommunication() protocols to direct it through the gateway to Base

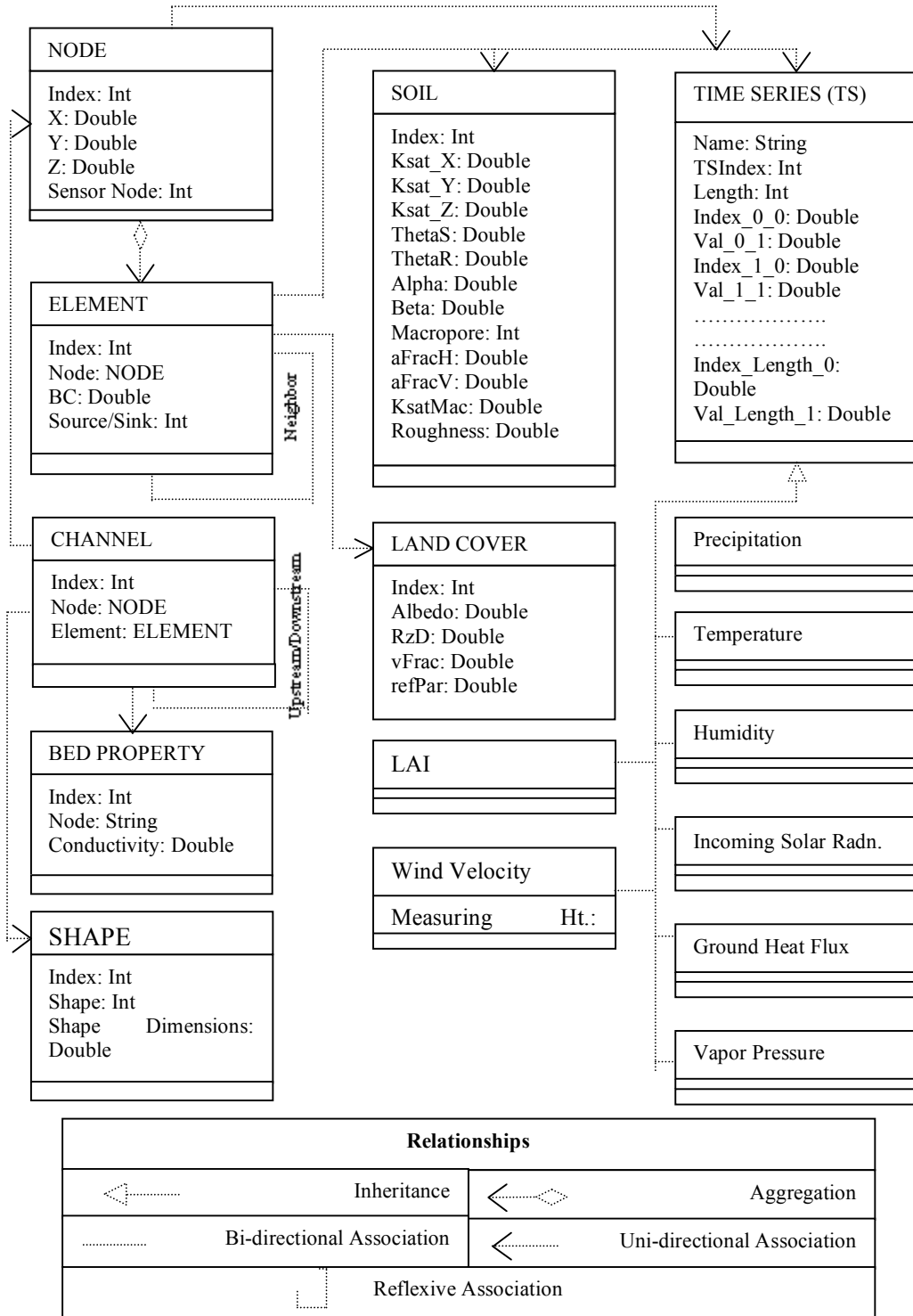


Figure 5: Shared Data Model for sensor network and hydrologic model. All the sensor node attributes and relationships are assigned to mesh nodes.

Station. In order for sensor network data to be used seamlessly in hydrologic modeling, the sensor data model constructs –classes and relationships- need to be supported in the hydrologic model data structure. By generating mesh decomposition using points and lines as constraints (shown in Figure 3, more details in Kumar et. al. [2008]), nodes and edges of the triangles in unstructured mesh decomposition of the model domain automatically

represents the sensor nodes and its neighbors respectively (shown in Figure 3). This means that the relationships and classes corresponding to each sensor node can be directly transferred to the data structure, relational attributes and topology information associated with discretized unstructured domain nodes.

## **5. SHARED “HYDROLOGIC MODEL”–“SENSOR NETWORK” DATA MODEL**

The developed data model is shown in Figure 5. The classes identified to describe the hydrologic system and processes are: Node, Element, Channel, Soil and Time Series. Each node is uniquely identified by its coordinate location and a sensor node ID if it exists. The data model supports Aggregation, Uni-directional Association, Reflexive Association and Generalization relationships between the objects. An Element class represents a discretized triangular element in 2D and a prismatic element in 3D and is defined by six nodal locations listed in clockwise direction at two levels. The prismatic element has five neighbours- three on the sides and one at the top and bottom. We note that neighbours of an element also belong to an Element class and this recursive relationship is captured by Reflexive association. A Channel class is defined by the two end nodes and neighbouring elements on the either side of channel. Each channel segment is also composed of an upstream and downstream channel segment which is captured by a Reflexive association. Channel is also Bi-directionally associated to each Element. Bi-directionality ensures that both Element and Channel is aware of this topological relationship. These relations are fundamentally important for spatial integrity of the hydrologic modeling framework. Each Element class is also associated with Soil class and Time Series. This ensures proper, clean and efficient assignment of properties to each Element. Similarly Channel is associated to Bed Property and Shape classes. Soil Class contains several attribute fields such as for Hydraulic conductivities and van-Genuchten equation parameters. We note that Precipitation, Temperature, Humidity, Incoming Solar Radiation, Ground Heat Flux, Vapour Pressure, LAI, Vegetation Fraction, Wind Velocity, Time dependent boundary conditions and the observed and simulated state variables are all instances or child objects to the Time Series Class. We note that the shared data model (shown in Figure 5) supports all the data types and the relationships that are used to store sensor the sensor network data (shown in Figure 4).

## **6. CONCLUSIONS**

This paper presents the design and details of a shared data model which supports coupling of sensor network data and a hydrologic model. The data model incorporates representation of a wide range of data types, feature objects and relationship between classes. The data model is rich yet flexible in terms of its extensibility and simplicity. The conceptualization and characterization of this coupling strategy can be used with other physically distributed models and can well be extended to management, visualization and decision support tools.

## **REFERENCES**

- Akyildiz, I.F., Su, W., Sankarasubramaniam, Y., and Cayirci, E., *Wireless sensor networks: a survey*, Computer Networks, Amsterdam, Netherlands, 1999
- Cardell, R., Smettem, K., Kranz, M., and Mayer, K., *A Reactive Soil Moisture Sensor Network: Design and Field Evaluation*, International Journal of Distributed Sensor Networks, pp. 149 – 162, 1, 2, 2005
- Cerpa, A., Elson, J., Estrin, D., and Girod, L., *Habitat Monitoring: Application Driver for Wireless Communications Technology*. In Proceedings of the ACM SIGCOMM Workshop on Data Communications in Latin America and the Caribbean, pages 20–41, 2001
- Hart, J.K. and Martinez, K., *Environmental Sensor Networks: A revolution in the earth system science?*, Earth-Science Reviews 78, 177–191, 2006
- Kumar, Bhatt, Duffy, *An efficient domain decomposition framework for accurate representation of geodata in distributed hydrologic models*, International Journal of GIS, 2008 (submitted).

- Lawson, C. L., Software for C1 Surface Interpolation, Mathematical Software III (John R. Rice, editor), pages 161-194. Academic Press, New York, 1977.
- McKinney, D.C., and Cai,X.,Linking GIS and water resources management models: an object-oriented method. Environmental Modeling and Software. v17, 5, 413-425. 2002
- Raghunathan, V., Schurgers, C., Park. S, and Srivastava, M.B., Energy-aware wireless microsensor networks," IEEE Signal Processing Magazine, 19, 2 , 40-50, March 2002
- Troch, 2008, <http://www.hwr.arizona.edu/~surface/fieldsites/valles.html>