235 © IWA Publishing 2005 **Journal of Hydroinformatics** | 07.4 | 2005

# Integrated use of GIS-based field sampling and modeling for hydrologic and water quality studies

Enrique R. Vivoni and Kevin T. Richards

## ABSTRACT

Enhancements to traditional catchment-scale water quality assessments can be realized by leveraging geographical information systems (GIS) for both field data collection and hydrologic and water quality (H/WQ) modeling. In this study, we describe a GIS-based data collection system for geo-referenced environmental sampling utilizing mobile, wireless and Internet technologies. Furthermore, sampled field data is combined with historical measurements within a GIS-based semi-distributed watershed model for simulating water quantity and quality in a large regional catchment. The GIS-based sampling and modeling system is intended to streamline water quality assessments as compared to current practices. We describe an application and field study in the Williams River, New South Wales, Australia designed to assess the impacts of point and non-point source pollution on water quality. Historical data were utilized for calibrating and validating the Hydrologic Simulation Program – Fortran (HSPF) with the BASINS GIS interface over the 1988–2000 period. Results from the study indicate that short-duration, spatially extensive field campaigns provide useful data for enhancing modeling studies based on historical measurements at sparse sites. In addition, the study suggests that the conjunctive use of data collection and modeling is a step towards real-time integration of field data in hydrologic and water quality modeling efforts.

**Key words** | field data collection, geographic information systems, hydrologic modeling, mobile computing, software, water quality

**Enrique R. Vivoni** (corresponding author)
Department of Earth and Environmental Science,
New Mexico Institute of Mining and Technology,
Socorro, NM 87801,
USA
Tel: +1 505 835 5611
Fax: +1 505 835 5634
E-mail: *vivoni@nmt.edu*

**Kevin T. Richards**
East Bay Municipal Utility District,
375 11th Street,
Oakland, CA 94607,
USA

## INTRODUCTION

Traditional catchment-scale water quality assessments are based on monitoring chemical and biological indicators at selected sites to observe long-term trends and develop causal relationships with land-use practices. Due to the spatial distribution of non-point sources, locating the precise origin and triggering of water quality exceedance events is difficult from historical measurements at a few sparse sites. For this reason, Grayson *et al.* (1997) and Eyre & Pepperell (1999) advocate the use of short-duration, spatially extensive water quality sampling campaigns during steady-state conditions. This sampling methodology provides an instantaneous view of the spatial variation of water quality under stable, low flow conditions. However, during storm events when water quality exceedances are typical (Nolan *et al.* 1995), the steady-state “snapshot” sampling strategy is not appropriate for discerning water quality variations within a large regional watershed (Grayson *et al.* 1997).

Diagnosing the spatial distribution of water quantity and quality during storm events requires incorporating field measurements into watershed simulation models. Although water quality modeling studies abound (e.g. Tim *et al.* 1992; Laroche *et al.* 1996; León *et al.* 2001; Engelmann *et al.* 2002), typical applications do not consider spatial variability of point or non-point source pollutants due to the lack of appropriate field data. The conjunctive use of field sampling and hydrologic and water quality (H/WQ) modeling has the potential for providing continuous, spatially distributed predictions under most hydrometeorological conditions, including flooding events. In addition, a distributed H/WQ simulation tool could be used for predicting water quality variations at spatial and temporal scales beyond the field monitoring extent.

Geographical information systems (GIS) provide the spatial context for advancing the state-of-the-art in H/WQ sampling and modeling. Through the explicit representation of geographic location, GIS-enabled technologies capture the spatial variability inherent in sampled environmental parameters or modeled hydrologic processes. Although various studies have integrated H/WQ models and GIS (e.g. Vieux 1991; Sui and Maggio 1999), the use of GIS for both environmental field sampling and hydrologic and water quality modeling has not been addressed previously. Nevertheless, Tim *et al.* (1992) and Rosenthal & Hoffman (1999) demonstrate the use of GIS-enabled H/WQ models for identifying critical regions that require intensive field monitoring.

In this study, we present the integrated use of two technologies for GIS-based H/WQ field sampling and modeling: (1) a new field data collection system that utilizes mobile computers, wireless communications and real-time Internet mapping and (2) a GIS-enabled, semi-distributed catchment model for water quality and flow predictions. In the following, we describe the use of the data collection system during a field study in the Williams River, New South Wales, Australia, as well as a hydrologic and water quality modeling effort based on historical and field-sampled data. The next section describes the new data collection technology, while the third section presents details on the H/WQ model. The fourth section briefly discusses the loose coupling of the GIS-based field data collection and modeling technologies. An application of the GIS-based field data collection and H/WQ modeling for the Williams River is then described in the next section. Finally, the last section discusses the case study results and the potential of GIS-enabled data collection and spatially distributed modeling for enhancing hydrologic and water quality studies.

## FIELD DATA COLLECTION SYSTEM

Field data collection during water quality assessments is a labor-intensive exercise typified by measurement, geo-referencing and transcription errors. Significant delays are expected between the data collection study and subsequent data analysis or modeling. Nevertheless, recent technological advances permit scientists to process, transmit and even map data while still in the field. Vivoni & Camilli (2003) describe a field data collection system that facilitates water quality assessments and data distribution. By “streaming” environmental field data, the system allows multiple teams to gather, share and transmit data simultaneously in tabular and cartographic (GIS) format within a localized field site as well as remote locations via the Internet.

Figure 1 illustrates a schematic of the GIS-based field data collection system for the acquisition, storage, transmission and mapping of geo-referenced environmental data. The primary system objective is to associate sampled water quality parameters with a unique location through the integration of environmental sensors and mobile computers. The system consists of a series of ruggedized hand-held computers and a roving station equipped with a local server, wireless router and connection to a remote web server. The mobile computers contain software used to gather data from a global positioning system (GPS) sensor, a multiparameter water quality sensor, a flow meter, a portable spectrophotometer and bacterial sampling equipment.

Water quality and hydrologic sampling is enhanced through the visualization capabilities provided by various mobile and web-based GIS applications (Figure 2). In order to display cartographic data to the mobile field worker, three mechanisms are implemented: (1) local map generation utilizing mobile GIS software, (2) uploading of a customized mapping service through a web browser and (3) integration with a remote web mapping server. GIS serves to integrate the geo-referenced water quality and hydrologic data with topography, soils, land use and hydrography data for the field site. In addition, GIS data layers describing the watershed land-surface and the hydrologic and water quality sampling locations can be used to establish a predictive H/WQ model that incorporates the spatial variation in basin properties. Furthermore, field data collection has the potential to aid in the selection of the model domain, resolution and parameters.

## HYDROLOGIC AND WATER QUALITY MODELING

Hydrologic and water quality models are useful tools for assessments of point and non-point source pollution.

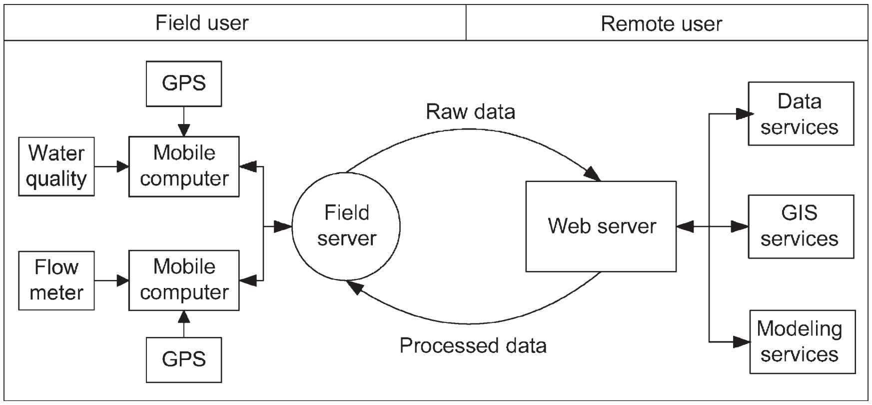

**Figure 1** | Conceptual diagram of the GIS-based field data collection system.

Among existing models, few provide the capabilities for continuous, spatially explicit simulation of hydrologic and transport processes in watersheds. Even fewer models are integrated with GIS, a desirable feature for facilitating spatial data inputs, analysis and simulations (Tim 1996). Among the available H/WQ models, the Hydrologic Simulation Program – Fortran (HSPF) is considered one of the most complete (Laroche *et al.* 1996; Carrubba 2000). The spatial variability in catchment water quality is simulated via discretization of the basin into sub-catchments composed of homogeneous segments contributing to runoff and transport (Jacomino & Fields 1997). Runoff from pervious and impervious land is routed through a channel network within which sediment, nutrients and contaminant transport can be modeled.

Numerous investigators have utilized HSPF for H/WQ modeling in a variety of watersheds with different climatic, geographic and physiographic characteristics (e.g. Srinivasan *et al.* 1998; Bergman & Donnangelo 2000; Socolofsky *et al.* 2001). Recently, Carrubba (2000) and Engelmann *et al.* (2002) utilized HSPF in the context of the Better Assessment Science Integrating Point and Non-point Sources (BASINS) toolkit. BASINS is a watershed management tool that integrates GIS and a variety of environmental models for point and non-point pollution assessments (US EPA 2001). In addition, BASINS provides access to watershed, water quality and meteorological databases for the continental United States. Whittemore & Beebe (2000) and Endreny (2002) discuss the advantages and disadvantages of utilizing BASINS for interfacing hydrologic models.

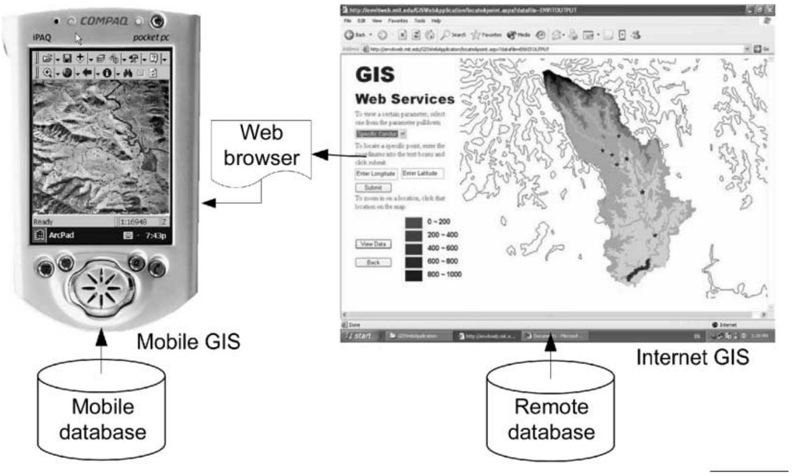

**Figure 2** | Field data mapping utilizing mobile GIS from ESRI ArcPad and GIS.NET Web Service application.

In particular, a weakness identified by these authors is the difficulty in applying the toolkit to watersheds outside the United States or with alternative data sources.

While model discretization in HSPF is based on sub-catchments, the use of the BASINS toolkit permits assignment of spatially varying parameter values across a basin. For applications with a high model resolution represented via a large number of internal sub-catchments, the semi-distributed HSPF model can depict realistic variations in land use, hydrography and terrain (Duda *et al.* 2001; Endreny 2002; Endreny *et al.* 2003). Sub-catchment partitioning also permits stream hydrograph and water quality simulations at individual internal basins (e.g. In *et al.* 2003), as well as spatial analysis of H/WQ model output (e.g., Johnson *et al.* 2003). Thus, the semi-distributed model is amenable to GIS-based analysis despite its lack of representing hydrologic and transport processes over domains in a fully distributed fashion (see Ewen *et al.* 2000; León *et al.* 2001).

Figure 3 illustrates a schematic of the BASINS toolkit and the HSPF model for GIS-based H/WQ modeling. Incorporated directly into ArcView GIS, BASINS provides tools for watershed delineation, stream extraction and sub-catchment selection based on a digital elevation model (DEM). Specification of HSPF model parameters for land-use segments and stream reaches can be performed within a BASINS project. However, utilizing the HSPF model, importing data and analyzing model results requires launching the WDM utility, WinHSPF and GenScn applications, respectively. WinHSPF is an object-oriented user interface for HSPF developed to create and modify HSPF input files (Duda *et al.* 2001). Modification of HSPF parameters is achieved through the various graphical facilities within WinHSPF. Through the loose coupling of BASINS and WinHSPF, H/WQ modeling in watersheds is significantly facilitated (Endreny 2002).

## INTEGRATED USE OF GIS-BASED TECHNOLOGIES

To improve H/WQ studies in large regional basins, an efficient means is sought for monitoring and simulating the spatial and temporal variations in hydrologic and water quality parameters. The integrated use of GIS-based sampling and modeling during a H/WQ study is a means for (1) efficiently gathering field data at sites critical to the model application, (2) merging field data with historical records at selected model-specific sites, (3) corroborating existing maps through field verification for model input, (4) enhancing the geospatial details of the model application via field observations and (5) conducting continuous and spatial simulations during and beyond the sampling period. The proper integration of the GIS-based sampling and modeling technologies would allow the field experimentalist and the computer modeler to better understand the watershed system under study (e.g. Siebert & McDonnell 2002).

While the tight coupling between the GIS-based field sampling and modeling is desirable, a reasonable first step towards this goal is to utilize the two technologies in a loosely coupled fashion. Since the data collection and simulation efforts are conducted within a GIS system, a seamless integration is possible of the geospatial coverages, field observations and data, historical records, model forcing and model output (Figure 3). Loose coupling between the two GIS-based technologies is achieved via: (1) utilizing geospatial watershed data verified in the field to setup the H/WQ model, (2) developing a modeling domain based on sampling sites from historical and spatially extensive studies, (3) using geospatial data to force the continuous H/WQ model and (4) comparing spatial model output to the historical and field data at numerous sampling sites. Furthermore, the integrated use of GIS-based field data in the H/WQ model can lead to enhancements in both activities by exposing deficiencies in the sampling plan or in the model simulation. These deficiencies can be corrected through an iterative process of refinement of the field data collection and model application. For example, field data can be used for updating model parameters and the modeling results can be used to target sampling efforts.

## CASE STUDY

The use of GIS for field data collection and H/WQ modeling is demonstrated during a field campaign and modeling exercise in the Williams River located in New South Wales, Australia (Figure 4). The catchment

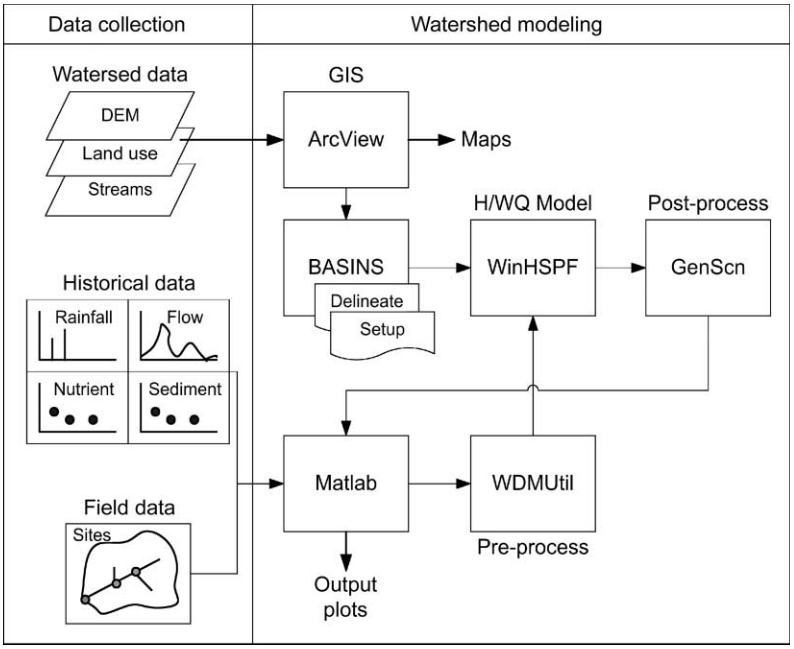

**Figure 3** | Conceptual diagram of the coupled field data collection and hydrologic and water quality (H/WQ) modeling using BASINS and WinHSPF.

(1260 km$^2$) is located within the lower Hunter Valley, about 30 km north of Newcastle. Although sparsely populated, the watershed provides approximately 80% of the drinking water for Newcastle through a water resources infrastructure that includes the Chichester Dam, Seaham Weir and Grahamstown reservoir. Since the Williams River is an important water supply, various measures are underway for protecting the watershed from non-point source pollution.

## Catchment description

The Williams River basin is characterized by a mountainous, forested region in the north and a series of low, rolling hills, used for cattle grazing, in the central and southern portions. The catchment has recently been confronted with a variety of water quality issues including eutrophication and blue–green algae blooms (Hunter Water Corporation 2001). Non-point source pollution in the watershed has been attributed to runoff from pasture and grazing lands, especially during storms (e.g. Nolan *et al.* 1995). Due to rainfall and runoff intermittency in the Williams River (Thomas & Henderson-Sellers 1991), the triggering of water quality exceedence events is closely linked to the storms that induce highly variable runoff (Croke & Jakeman 2001).

The warm, temperate climate in the Williams basin is modulated by an orographic effect imposed by the Barrington Tops mountain range and a maritime influence due to its coastal proximity (Wooldridge *et al.* 2001*b*). Both effects induce spatial variability in rainfall within the watershed, as demonstrated by rain gauge records for two locations in the catchment (Figure 5(a)). The northern, mountain region receives considerably more rainfall than the central and southern portions due to the orographic effect. This rainfall variability translates into a higher proportion of runoff attributed to the upper catchment areas during wet and dry years (Wooldridge *et al.* 2001*a*). Combined with the spatial variability in landscape properties (e.g. soils, land use, topography, hydrography), the rainfall variations across the catchment lead to differences in the observed runoff within the stream gauging network (Figure 5(b)).

## Watershed and hydrometerological data

On-going efforts to monitor the Williams River for blue–green algae and nutrients provided important sources of

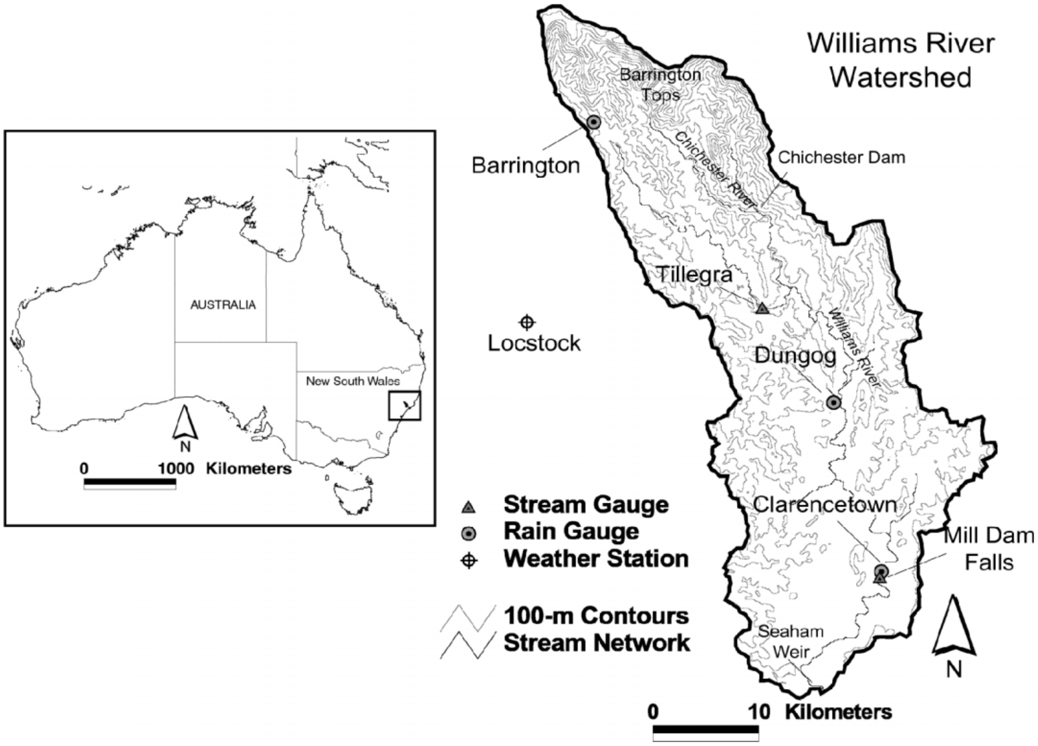

**Figure 4** | Location of the Williams River catchment, New South Wales, Australia. Topography from 25-m DEM displayed as 100-m contour intervals. Locations of selected continuous-recording stream, weather and rain gauges also displayed.

data utilized in this H/WQ field sampling and modeling exercise. In particular, a detailed GIS database containing topography, vegetation, soil units, land use and hydrography was available prior to the field study (Krause *et al.* 1997). GIS data describing the basin land surface are useful for delineating watershed boundaries and the stream network, selecting field sampling sites and obtaining appropriate parameters for H/WQ models. Among these, only the watershed topography and land-use classifications are strictly required for executing HSPF (Figure 6).

In addition to the GIS layers, one weather station, three rain gauges, two stream gauges and three water quality stations are within or in close proximity to the Williams basin (Figure 4). Daily weather data from the Lostock Dam (1988–2002), daily rainfall at Barrington, Dungog and Clarencetown (1988–2002), daily discharge data at Tillegra and Mill Dam Falls (1988–2002) and water quality data at Tillegra, Mill Dam Falls and Boags Hill (1994–2001) from various sources were utilized in this study. The historical water quality data included both indicator parameters and complete chemical analysis.

## Field campaign

A spatially extensive field campaign was carried out for sampling the water quality and flow in the Williams River during summer conditions (January, 2002). Little rainfall and low flow conditions were observed in the watershed. Three field teams consisting of six to seven members were deployed at separate river cross sections. Each team was equipped with two ruggedized hand-held computers, associated environmental sensors and GIS databases. Team members were responsible for water quality and flow sampling, operation of sensors, field instruments and mobile computers, wireless data transmission and cartographic displays. The field teams traveled independently visiting sampling locations chosen based upon: (1) data requirements for H/WQ modeling, (2) existing historical

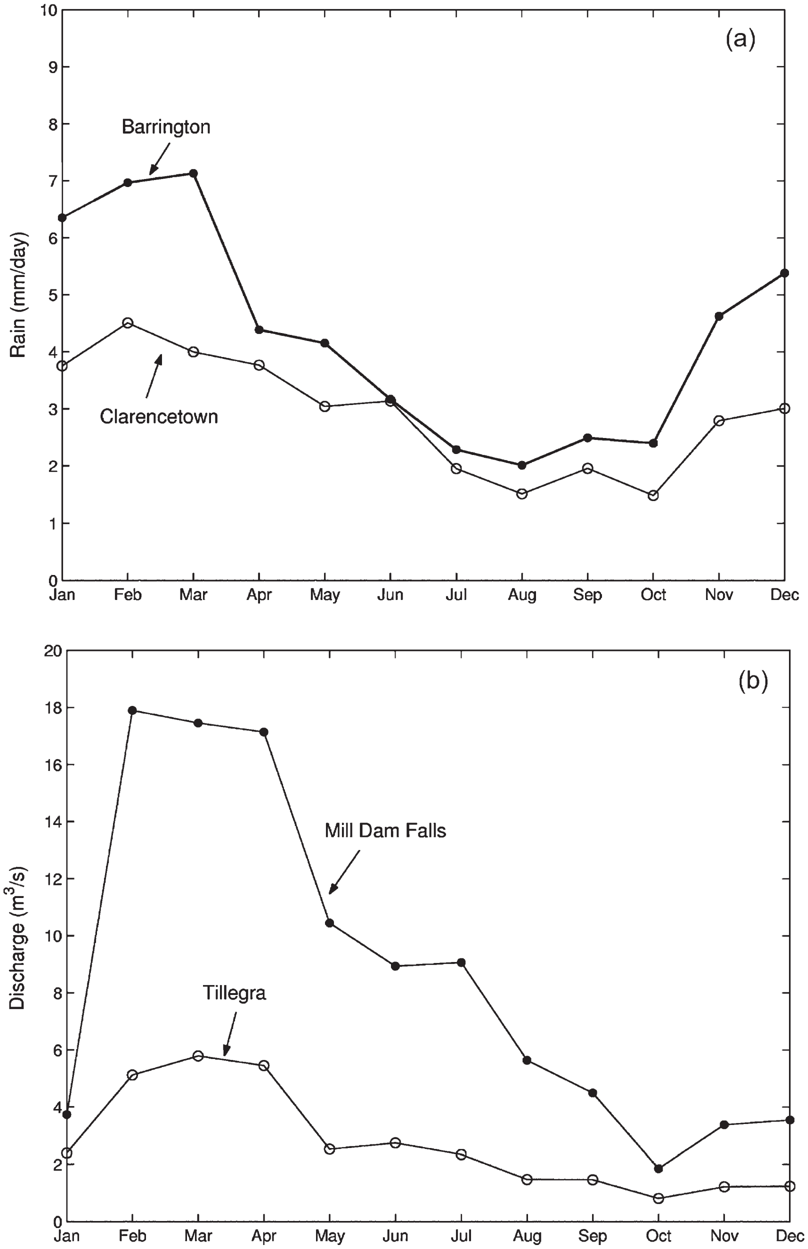

**Figure 5** | Spatial variability in rainfall and runoff within the Williams River catchment. (a) Mean monthly rainfall for the Barrington and Clarencetown rain gauges over the 1988–2000 period (mm $d^{-1}$). (b) Mean monthly discharge for the Mill Dam Falls and Tillegra streamflow gauges for the period 1988–2000 ($m^3 s^{-1}$). Figure 4 illustrates the location of rain and streamflow gauges.

data for validation and (3) potential nutrient sources in the catchment. In addition, the sampling teams corroborated existing GIS coverages with field observations to ensure their appropriate use in the H/WQ model application.

Figure 7 illustrates an example of the GIS-based field data collection undertaken simultaneously by three field teams near the Tillegra gauging station. Upon gathering geo-referenced flow and water quality samples at each location, each team transmitted data to the laptop field server onboard a roving van. The wireless communications network allowed the teams to be spaced within a few kilometers of the roving van, depending on line-of-sight and signal strength. From this field server, geo-referenced data from all the field teams was periodically transmitted back to

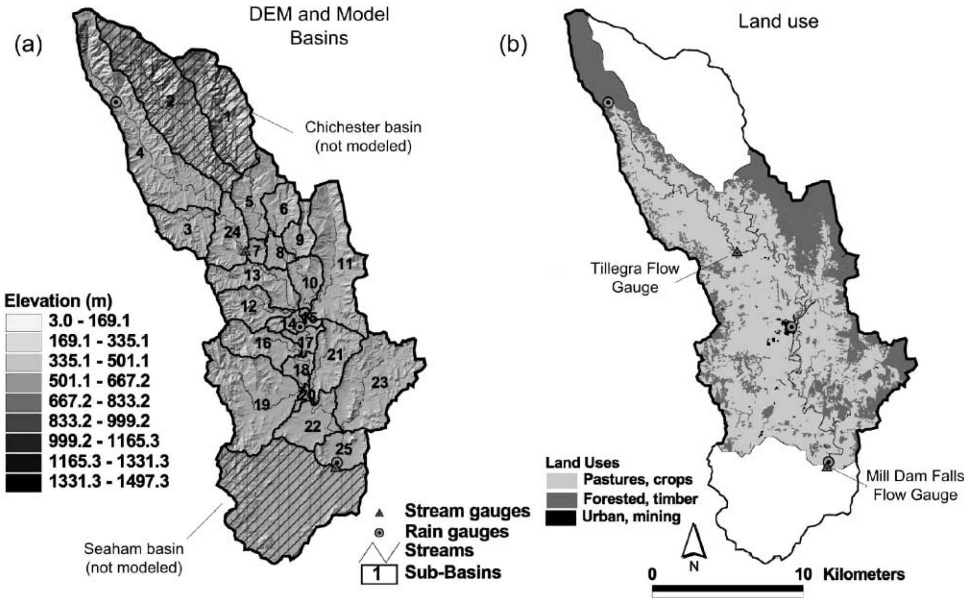

**Figure 6** | Hydrologic and water quality modeling domain for the Williams River basin. (a) Stream network and subcatchment delineation achieved within BASINS based on the 25-m DEM. (b) Land use assigned to each basin within BASINS and WinHSPF.

the mobile computers as a map display indicating the spatial variability of the sampled parameters. During the field study, this was achieved using a remote web mapping service (Figure 2). The compiled data set from the individual teams was subsequently used to determine the locations for delineated sub-basins and verifying model output within the H/WQ application.

Low-flow, snapshot sampling at various locations within the Williams River and its tributaries permitted analysis of the spatial variation in water quality and flow parameters. As an example, Figure 8 illustrates the variation of fecal coliform counts as a function of distance to the watershed outlet for three stream reaches. Comparisons are shown to the mean, maximum and minimum levels measured at Boags Hill during summer conditions over the period 1988–2001. Analogously, the spatial variation in nutrient concentrations, pH, conductivity, dissolved oxygen, temperature, turbidity and discharge were obtained for this low-flow, summer period and compared to historical data from various sites (not shown). In all cases, the results of the field campaign were within the parameter ranges obtained during previous monitoring studies. The spatially extensive database of field sampled H/WQ parameters was subsequently used for the watershed model application using the semi-distributed HSPF model with BASINS.

## Hydrologic modeling

Prior to watershed simulations, a perceptual model of the basin hydrology was developed from observational evidence obtained during the field campaign, which aided in the selection of HSPF modules, parameter ranges, input coverage data and model forcing. For example, the sharp rainfall gradient in rain gauge observations is accounted for with spatially variable rainfall input in the model. In addition, water diversions at the Chichester Dam and Seaham Weir are accounted for by isolating these sub-basins from the main watershed (Figure 6). The historical record and field campaign data also served to identify which basin locations were critical to include in the model structure to obtain output relevant to the observations. The BASINS and WinHSPF tools were subsequently utilized to set up, calibrate and verify the HSPF simulations over the 1988–2000 period.

### Model setup

The BASINS toolkit provides various GIS functions for delineating streams and basin boundaries, assigning land use and specifying rainfall and stream gauges. The stream network was derived from the 25-m DEM using a constant-area threshold of 14 km$^2$, chosen based on comparisons to the hydrography extracted from aerial photographs.

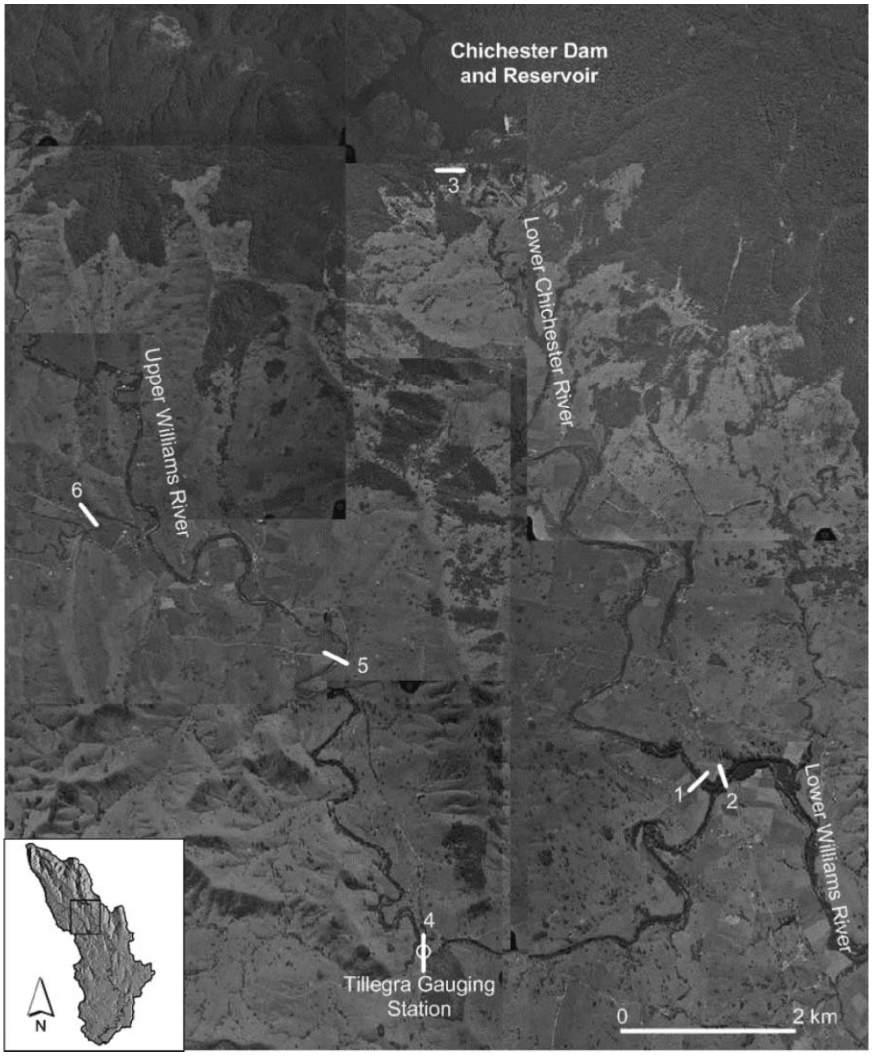

**Figure 7** | Sampling locations in the northern Williams River watershed overlaid on aerial photography. Multiple teams measured water quality and quantity parameters at six cross sections in this region, including the Tillegra gauging station.

The river geometry required to run the HSPF routing module (RCHRES) was computed for each stream reach in BASINS based on the spatially variable topographic model. In addition, the watershed was divided into sub-basins for semi-distributed modeling. Sub-basin delineation resulted in 25 watersheds north of Mill Dam Falls (1000 km$^2$), as shown in Figure 6. The Mill Dam Falls and Tillegra stream gauges were added as sub-basin outlets to ensure model comparisons to flow observations at these sampling locations.

HSPF is primarily parameterized through the use of land-use cover (Figure 6). Basin land use was grouped into five major classes: forests, grazing, crops, urban areas and water bodies. BASINS facilitates the specification of the perviousness for each land use which determines the use of the pervious (PERLND) or impervious (IMPLND) HSPF modules. Impervious segments were assigned using the criteria of directly connected impervious (DCI) areas (Bergman & Donnangelo 2000). Given its rural nature, DCI or impervious areas are infrequent in the watershed. Nevertheless, low permeability soils determined from soil county maps (Matthei 1995; Henderson 2000) were identified and correlated with distance to the streams to determine DCI areas. This resulted in a spatial map of land use and perviousness used to parameterize each sub-basin in the model.

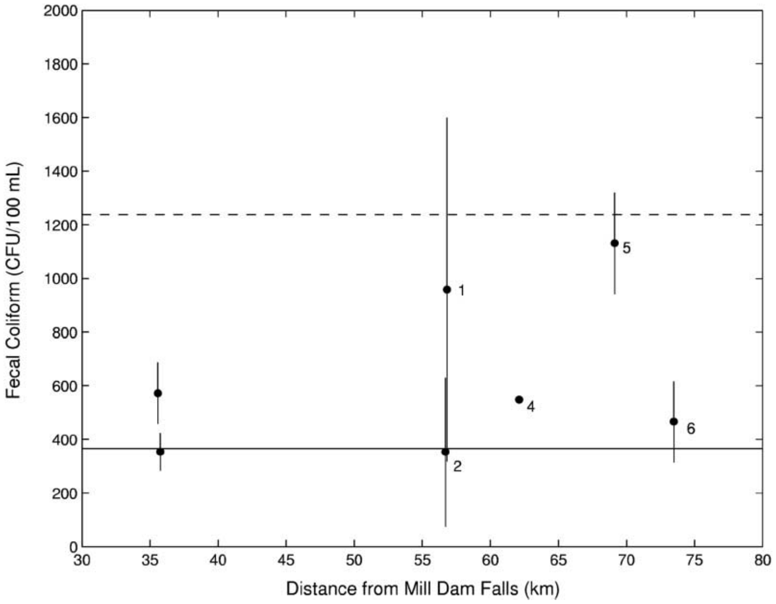

**Figure 8** | Spatial variability in fecal coliform sampling within the Upper Williams River. Mean cross sectional fecal coliform counts (CFU/100 mL) as a function of river distance (km) from the outlet (Mill Dam Falls). Vertical bars represent ±1 standard deviations. Solid horizontal line is the mean value of fecal coliform sampling during January over the 1996–2001 period. Dashed horizontal line represents +1 standard deviation from historical January mean. Numbering refers to sampling sites shown in Figure 7.

To summarize, the GIS-based HSPF application in the Williams River watershed utilized the spatial distribution of topography to delineate sub-basins and river reaches, the distribution of rainfall obtained from three rain gauge stations as model forcing and spatial variability in land cover to parameterize the pervious and impervious HSPF model components. In addition, the field sampling locations, river gauge stations and historical records were utilized to ensure model output at corresponding measurement sites.

## Model calibration

The calibration of the HSPF hydrologic predictions was performed for the Williams River upstream of Mill Dam Falls. The flow observations at Mill Dam Falls were divided into a calibration period from 1988–1995 and a verification period from 1996–2000. Initialization errors associated with the model storages were accounted for during a three year "spin-up" time in the calibration sequence. The model verification exercises were performed at both Mill Dam Falls and Tillegra, an interior gauging station, to assess parameter transferability across catchment scales (Figure 6).

A step-wise approach for calibrating the storage and flux parameters in HSPF was utilized (US EPA 2000*a*). First, the annual water balance was adjusted using the deviation of runoff volume ($D_v$) as a performance measure by tuning parameters that control evapotranspiration (LZETP), groundwater loss (DEEPFR) and infiltration rate (INFILT). A similar procedure was used for the seasonal water balance by dividing the year into a wet season (January–June) and a dry season (July–December). Finally, the daily storm hydrographs for a number of events permitted adjustments of the runoff mechanisms by manipulating the groundwater baseflow recession (AGWRC), infiltration (INFILT) and baseflow evaporation (BASETP) parameters (Table 1). For this last step, both the deviation of runoff volume ($D_v$), defined as

$$D_v = \frac{V_o - V_s}{V_o} \tag{1}$$

where $V_o$ is the observed runoff volume and $V_s$ is the simulated runoff volume, and the Nash–Sutcliffe efficiency coefficient ($E$), defined as

$$E = 1 - \frac{\sum_{i=1}^{N}(Q_o - Q_s)^2}{\sum_{i=1}^{N}(Q_o - \bar{Q})^2} \tag{2}$$

**Table 1** | Calibrated HSPF parameter values for the pervious fraction of the Williams River watershed over the 1988–1995 period

| Parameter (units) | Description | Range | Default value | Calibration value |
|---|---|---|---|---|
| INFILT ($\mathrm{mm\,h^{-1}}$) | Soil infiltration rate index | 0.025–12.7 | 4.06 | 2.54 |
| AGWRC ($\mathrm{d^{-1}}$) | Groundwater recession rate | 0.85–0.99 | 0.98 | 0.80 |
| LZETP | Lower zone ET index | 0.1–0.9 | 0.1 | 0.1–0.4 |
| DEEPFR | Fraction of groundwater loss | 0–0.5 | 0.1 | 0.2 |
| BASETP | Fraction of groundwater ET | 0–0.2 | 0.02 | 0.05 |

where $Q_o$ is the observed discharge, $Q_s$ is the simulated discharge and $\bar{Q}$ is the mean of the observed discharge, were utilized to gauge model performance (ASCE 1993).

Table 2 illustrates the model performance prior to and after calibration as quantified by $D_v$ and $E$. Results suggest that the HSPF model is adequately reproducing streamflow during the 1988–1995 calibration period, including the storm and interstorm conditions. Figure 9(b) compares the observed and simulated discharge at Mill Dam Falls during 1995, after the three-year spin-up period, given the meteorological forcing in Figure 9(a). Note that, although the March flood peak is overestimated, the time to peak, storm runoff volume, baseflow recession and low flow behavior are reproduced well. Model performance over the 1991–1995 period was similar to the results shown for 1995.

## Model verification

Verification of the calibration parameter values was performed for the stream flow at Mill Dam Falls and Tillegra for the period 1996–2000. Figure 10 illustrates the verification results for the year 2000 at Mill Dam Falls. As compared to the calibration, a slightly higher $D_v$ and improved $E$ are observed (Table 2). The peak flow, time to peak, baseflow recession and low-flow conditions are simulated appropriately over the verification period for the 1000 km$^2$ region (Figure 6).

Given that model calibration parameter values were applied uniformly for the pervious and impervious fractions determined from spatial land-cover data, an important test of the model is whether simulations at internal points are appropriate. One method employed here is comparing the simulated and observed hydrographs at the internal Tillegra gauge during the verification period (Figure 11). The results of this “blind” verification suggest the model is over-estimating runoff in the Tillegra sub-basin (195 km$^2$), due to the bias introduced by using high rainfall values at Barrington for the entire sub-basin. Despite this, model performance is adequate, considering that observations at Tillegra were not used during calibration and that rainfall from three rain gauges is used to force the watershed model in a spatially explicit fashion. Model performance over

**Table 2** | Statistical measures of HSPF model performance during the calibration and verification periods for the Mill Dam Falls and Tillegra stream gauges

| | Calibration | | Verification | |
|---|---|---|---|---|
| Gauge/period | $D_v$ | $E$ | $D_v$ | $E$ |
| *Mill Dam Falls* | | | | |
| 1991–1995 | −0.076 | 0.16 | – | – |
| 1995 | 0.043 | 0.26 | – | – |
| 1996–2000 | – | – | −0.020 | 0.65 |
| 2000 | – | – | −0.043 | 0.79 |
| *Tillegra* | | | | |
| 1991–1995 | 0.262 | 0.32 | – | – |
| 1995 | 0.397 | 0.35 | – | – |
| 1996–2000 | – | – | 0.301 | 0.24 |
| 2000 | – | – | 0.271 | 0.34 |

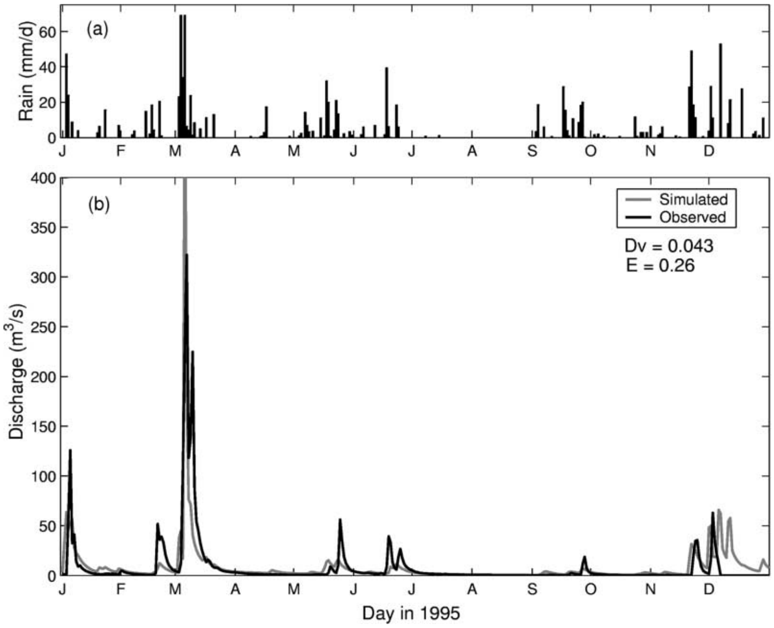

**Figure 9** | Streamflow calibration for the Williams River. (a) Measured rainfall at Clarencetown. (b) Simulated (gray) and observed (black) discharge at Mill Dam Falls. The deviation of runoff volume ($D_V$) and Nash–Sutcliffe efficiency ($E$) for 1995 is shown.

the entire verification period was similar to the results shown for 2000 in Figures 10 and 11.

### Water quality modeling

The calibration and verification of the hydrologic component of the semi-distributed HSPF model for the Williams River watershed is a prerequisite for the simulation of transport processes in the catchment. Water quality modeling using HSPF has received much less attention than hydrologic simulations, although notable examples exist (e.g. Laroche *et al.* 1996; Engelmann *et al.* 2002). In this research, we have focused on simulating the bacterial loading, particularly fecal coliform, to the Williams River basin from the grazing activities in the catchment. Water quality exceedence events in the catchment have been related directly to coliform concentrations produced by cattle grazing near streams (Nolan *et al.* 1995).

Although historical fecal coliform data are available, current estimates of the amount of cattle grazing activity, its temporal variability and spatial distribution are insufficient to accurately ascertain the bacterial input to the catchment. Given appropriate source data, the HSPF model can be utilized with the existing hydrologic calibration in predicting bacterial transport (In *et al.* 2002). Specifically, the input of fecal coliform from cattle near streams can be modeled using point sources along the stream network, while non-point sources can be accounted for with overland flow. A recent tool developed by the US EPA facilitates the estimation of fecal coliform loading based on land cover, catchment area and bacteria sources including septic tanks and cattle grazing in streams (US EPA 2000*b*). On-going efforts are focused on modeling bacterial transport for the Williams River based on the historical records at selected sites and augmented with the spatial snapshot of field sampled fecal coliform counts (Figure 8) with the calibrated HSPF model.

## DISCUSSION

In this study, we illustrate the integrated use of a GIS-based field data collection system and a semi-distributed model for hydrologic and water quality assessments in a large regional watershed. Geographic information systems facilitated the collection of field data among a set of teams dispersed throughout the basin. Favorable comparisons between the field samples to historical records suggest the field

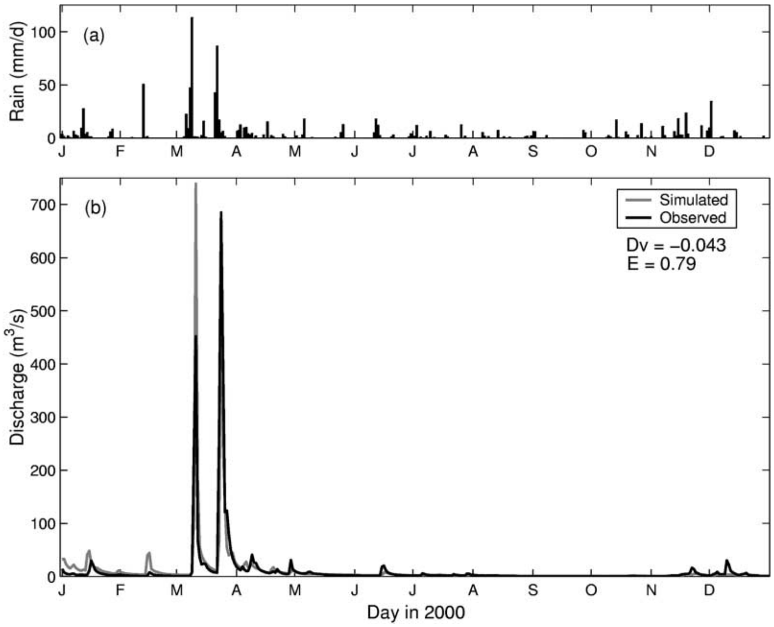

**Figure 10** | Streamflow verification at outlet gauge. (a) Measured rainfall at Clarencetown. (b) Simulated (gray) and observed (black) discharge at Mill Dam Falls.

campaign was effective in determining basin water quality during low-flow conditions. Expanding the scope of a field study to storm periods is more difficult since the watershed is inherently unsteady during these events. In the absence of field data, a well-calibrated H/WQ model can be used to discover the spatial and temporal variation of runoff or bacterial transport during storm events. We follow this approach by calibrating the hydrologic components of a GIS-based H/WQ model driven by rainfall data and verified with discharge observations.

Continuous simulation models, such as HSPF, provide a tool by which better assessments of water quality conditions can be made in large watersheds (Srinivasan *et al.* 1998). All too often, however, predictive H/WQ models do not examine water quantity and quality conditions in a geographic context. In this study, we have incorporated the spatial variability in catchment properties into the semi-distributed HSPF model, in particular with regard to rainfall forcing, topography, hydrography and land cover. Model calibrations at Mill Dam Falls, the basin outlet, were verified with internal simulations conducted for an internal stream gauge at Tillegra. This "blind" verification technique is a useful means for assessing the distributed performance of a hydrologic model and provides an indication of the reliability of calibrated parameters based on land cover and soils coverage (e.g. Reed *et al.* 2004). Given the basin extent and hydrologic comparisons at the outlet and internal gauging stations, the HSPF model performance is considered adequate and comparable to previous studies (e.g. Srinivasan *et al.* 1998; Carrubba 2000; Bergman & Donnangelo 2000; Engelmann *et al.* 2002).

The semi-distributed model application in the Williams River was enhanced through the use of the historical records of hydrologic flows and water quality parameters in the basin and via the spatially extensive, synoptic field campaign taken during a low-flow period. The "snapshot" surveys of stream discharge at multiple locations were used to assess the spatial flow variations within the channel network and compare these to the model results at multiple nested sites. While longer term records were used in the model calibration and verification exercise, the field campaign provided a measure of model performance at a particular instance in time. Similarly, the water quality parameters measured during the field campaign (Figure 8) can play a critical role in calibrating and verifying the bacterial transport simulations conducted in HSPF.

As discussed previously, the GIS-based technologies utilized in this study are loosely coupled as a means for integrating field data and semi-distributed modeling. Tighter coupling of the two technologies is desirable for several

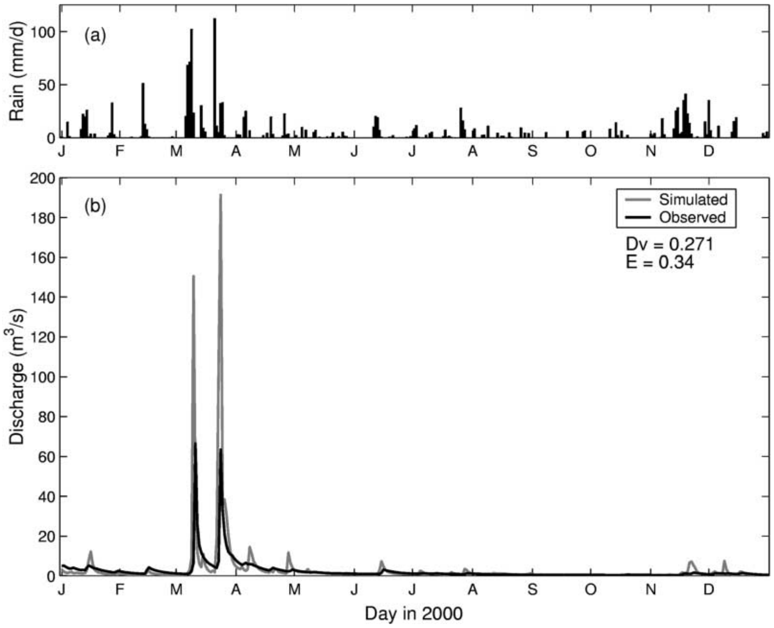

**Figure 11** | Streamflow verification at internal stream gauge. (a) Measured rainfall at Barrington. (b) Simulated (gray) and observed (black) discharge at Tillegra.

reasons, including the potential for feedback between field data collection and continuous modeling. A fully coupled system can best be achieved by implementing a web-based H/WQ modeling service that remotely uses the data from the GIS-based data collection system (Figure 2). Such a development would provide the capability for using *in situ*, real-time data in a continuous model operating off-site (Figure 1). Short-term, extensive field studies can be complemented with a network of telemetered stations continuously collecting data at a select number of sites, which can be transmitted via the Internet (e.g. Christensen *et al.* 2002; Flieschmann *et al.* 2002). While appealing, a fully coupled, GIS-based system for H/WQ data collection and simulation is currently limited by the development of web-based modeling services with the capacity to utilize telemetered data streams. The application of real-time modeling services to hydrologic and water quality assessments would greatly facilitate the process of collecting, displaying and analyzing both field observations and model output. Given the spatial nature of applications, fully distributed models (e.g. Ewen *et al.* 2000; León *et al.* 2001) that explicitly account for spatial variability are an attractive possibility for coupled data collection and modeling systems.

## CONCLUSIONS

This paper describes the use of geographic information systems for field data collection and semi-distributed basin modeling in water quantity and quality applications. GIS provides the spatial context for gathering, displaying and analyzing geo-referenced field data obtained using a mobile, wireless system. GIS also serves as an interface for a spatially explicit hydrologic and water quality model that simulates basin processes during storm and interstorm periods. In the Williams River case study presented here, we utilized the GIS-based system to (1) select sampling sites based on historical data and model requirements, (2) map a suite of hydrologic and water quality parameters collected at various locations, (3) superimpose field samples on land use, soils, topography and stream coverages, (4) field verify the spatial correctness and descriptions of historical maps and (5) set up, parameterize and view the output from a semi-distributed hydrologic and water quality model. Results indicate that the field campaign obtained data consistent with low-flow records while enhancing the spatial distribution of sampled parameters. Through a model calibration and verification exercise, the watershed simulation tool was shown to reproduce the hydrologic response in the basin in reaches beyond the field sampling locations and over both low-flow and flood periods. In particular, the model has the capability to simulate water quality exceedances during storms, which are difficult to sample simultaneously over large watersheds.

## ACKNOWLEDGEMENTS

This research was funded through the MIT/Microsoft Alliance project at the Massachusetts Institute of Technology, Cambridge, MA. We acknowledge the support of Richard Camilli, Sheila Frankel, Eric Adams, Dara Entekhabi, Daniel Sheehan and Mario Rodríguez. In addition, the assistance of the ENVIT Student Group at MIT is greatly appreciated. We also acknowledge Garry R. Willgoose, Jetse D. Kalma and Andrew Krause, University of Newcastle, NSW as well as Ian Charlton and Simon Mckee from the Department of Land and Water Conservation, NSW for providing useful data. We thank two reviewers who provided useful suggestions on an earlier manuscript.

## REFERENCES

American Society of Civil Engineers 1993 Criteria for evaluation of watershed models. *J. Irrig. Drainage Engng.* **119** (3), 429–442.

Bergman, M. J. & Donnangelo, L. J. 2000 Simulation of freshwater discharges from ungaged areas to the Sebastian River, Florida. *J. AWRA* **36** (5), 1121–1132.

Carrubba, L. 2000 Hydrologic modeling at the watershed scale using NPSM. *J. AWRA* **36** (6), 1237–1246.

Christensen, V. G., Rasmussen, P. P. & Ziegler, A. C. 2002 Real-time water quality monitoring and regression analysis to estimate nutrient and bacteria concentrations in Kansas streams. *Wat. Sci. Technol.* **45** (9), 205–219.

Croke, B. F. W. & Jakeman, A. J. 2001 Predictions in catchment hydrology: an Australian perspective. *Marine Freshwater Res.* **52**, 65–79.

Duda, P., Kittle, J., Gray, M., Hummel, P. & Kinerson, R. S. 2001 WinHSPF – an independent, fully-integrated component of a comprehensive modeling system. In *Proc. AWRA Annual Spring Specialty Conference*. American Water Resources Association. Middleburg, VA, pp. 23–28.

Endreny, T. A. 2002 BASINS toolkit for hydrological monitoring, modeling and assessment. *Hydrol. Process.* **16**, 1331–1335.

Endreny, T. A., Somerlot, C. & Hassett, J. M. 2003 Hydrograph sensitivity to estimates of map impervious cover: a WinHSPF case study. *Hydrol. Process.* **17**, 1019–1034.

Engelmann, C. J. K., Ward, A. D., Christy, A. D. & Bair, E. S. 2002 Application of the BASINS database and NPSM model on a small Ohio watershed. *J. AWRA* **38** (1), 289–300.

Ewen, J., Parkin, G. & O'Connell, P. E. 2000 SHETRAN: distributed river basin flow and transport modeling system. *J. Hydrol. Engng.* **5** (3), 250–258.

Eyre, B. D. & Pepperell, P. 1999 A spatially intensive approach to water quality monitoring in the Rous River catchment, NSW, Australia. *J. Environ. Mngmnt.* **56**, 91–118.

Flieschmann, N., Staubmann, K. & Langergraber, G. 2002 Management of sensible water uses with real-time measurements. *Wat. Sci. Technol.* **46** (3), 33–40.

Grayson, R. B., Gippel, C. J., Finlayson, B. L. & Hart, B. T. 1997 Catchment-wide impacts on water quality: the use of 'snapshot' sampling during stable flows. *J. Hydrol.* **199**, 121–134.

Henderson, L. E. 2000 *Soil Landscapes of the Dungog 1:100,000 Sheet*. Department of Land and Water Conservation. NSW, Australia.

Hunter Water Corporation 2001 *Water Quality of the Williams River. Annual Report*. Newcastle, Australia.

In, S., Brannan, K., Mostaghimi, S., Cho, J. & Gupta, R. 2002 Modeling fecal coliform bacterial loading from an urbanizing watershed. In *Proc. Annual AWRA Conference, Philadelphia, PA*, American Water Resources Association. Middleburg, VA, p. 107.

In, S. J., Brannan, K. M. & Mostaghimi, S. 2003 Simulating hydrologic and water quality impacts in an urbanizing watershed. *J. AWRA* **39** (6), 1465–1479.

Jacomino, V. M. F. & Fields, D. E. 1997 A critical approach to the calibration of a watershed model. *J. AWRA* **33** (1), 143–154.

Johnson, M. S., Coon, W. F., Mehta, V. K., Steenhuis, T. S., Brooks, E. S. & Boll, J. 2003 Application of two hydrologic models with different runoff mechanisms to a hillslope dominated watershed in the northeastern US: a comparison of HSPF and SMR. *J. Hydrol.* **284** (1–4), 57–76.

Krause, A. K., McCabe, M. & Kalma, J. D. 1997 The development of a terrain-based erosion hazard map for the Williams River catchment in eastern New South Wales. In *Proc. MODSIM, Hobart, Australia*, Modelling and Simulation Society of Australia and New Zealand. Canberra, Australia, pp. 464–469.

Laroche, A.-M., Gallichand, J., Lagace, R. & Pesant, A. 1996 Simulating atrazine transport with HSPF in an agricultural watershed. *J. Environ. Engng.* **122** (7), 622–630.

León, L. F., Soulis, E. D., Kouwen, N. & Farquhar, G. J. 2001 Nonpoint source pollution: a distributed water quality modeling approach. *Wat. Res.* **35** (4), 997–1007.

Matthei, L. E. 1995 *Soil Landscapes of the Newcastle 1:100,000 Sheet*. Department of Land and Water Conservation. NSW, Australia.

Nolan, A. L., Lawrance, G. A. & Maeder, M. 1995 Phosphorus speciation in the Williams River, New South Wales: eutrophication and a chemometric analysis of relationships with other water quality parameters. *Marine Freshwater Res.* **46**, 1055–1064.

Reed, S., Koren, V., Smith, M., Zhang, Z., Moreda, F., Seo, D.-J. & DMIP participants 2004 Overall distributed model intercomparison project results. *J. Hydrol.* **298** (1–4), 27–60.

Rosenthal, W. D. & Hoffman, D. W. 1999 Hydrologic modeling/GIS as an aid in locating monitoring sites. *Trans. ASAE* **42** (6), 1591–1598.

Siebert, J. J. & McDonnell, J. J. 2002 On the dialog between experimentalist and modeler in catchment hydrology: use of soft data for multicriteria model calibration. *Wat. Res. Res.* **38** (11), 1241.

Socolofsky, S., Adams, E. E. & Entekhabi, D. 2001 Disaggregation of daily rainfall for continuous watershed modeling. *J. Hydrol. Engng.* **6** (4), 300–309.

Srinivasan, M. S., Hamlett, J. M., Day, R. L., Sams, J. I. & Petersen, G. W. 1998 Hydrologic modeling of two glaciated watersheds in Northeast Pennsylvania. *J. AWRA* **34** (4), 963–978.

Sui, D. Z. & Maggio, R. C. 1999 Integrating GIS with hydrological modeling: practices, problems and prospects. *Comput., Environ. Urban Syst.* **23**, 33–51.

Thomas, G. & Henderson-Sellers, A. 1991 An evaluation of proposed representations of subgrid hydrologic processes in climate models. *J. Climate* **4**, 898–910.

Tim, U. S. 1996 Emerging technologies for hydrologic and water quality modeling research. *Trans. ASAE* **39** (2), 465–476.

Tim, U. S., Mostaghimi, S. & Shanholtz, V. O. 1992 Identification of critical non-point pollution source areas using geographical information systems and water quality modeling. *Wat. Res. Bull.* **28** (5), 877–887.

US EPA 2000a *BASINS Technical Note 6: Estimating Hydrology and Hydraulic Parameters for HSPF*. *EPA-823-R-00-012*. US EPA, Washington, DC.

US EPA 2000b *Bacteria Indicator Tool User's Guide*. *EPA-823-B-01-003*. US EPA, Washington, DC.

US EPA 2001 *BASINS Version 3.0 User's Manual*. *EPA-823-B-01-001*. US EPA, Washington, DC.

Vieux, B. 1991 Geographic information systems and non-point source water quality and quantity modeling. *Hydrol. Process.* **5**, 101–113.

Vivoni, E. R. & Camilli, R. 2003 Real-time streaming of environmental field data. *Comput. Geosci.* **29**, 457–468.

Whittemore, R. C. & Beebe, J. 2000 EPA's BASINS Model: Good Science or Serendipitous Modeling. *Journal of the American Water Resources Association* **36** (3), 493–499.

Wooldridge, S., Kalma, J. & Kuczera, G. 2001a Parameterization of a Simple Semi-distributed Model for Assessing the Impact of Land-use on Hydrological Response. *Journal of Hydrology* **254**, 16–32.

Wooldridge, S. A., Franks, S. W. & Kalma, J. D. 2001b Hydrological Implications of the Southern Oscillation: Variability of the Rainfall-runoff Relationship. *Hydrological Sciences* **46** (1), 73–88.