Developing a TeraGrid Based Land Surface Hydrology and Weather Modeling Interface

Hsin-I Chang[1], Dev Niyogi[1], Fei Chen[2], Anil Kumar[1,2], Carol Song[1], Lan Zhao[1], Rao S. Govindaraju[1], Venkatesh Merwade[1], Ming Lei[1], Ken Scheeringa[1].
[1]Purdue University, West Lafayette, IN 47906 USA, [2]NCAR, Boulder, CO 80307
iclimate@purdue.edu

1 Introduction

Real world hydrologic and environmental problems are often challenged by several factors such as (i) the curse of dimensionality and scale, (ii) data that tend to be very high-dimensional, heterogeneous, and sparse, and (iii) uncertainty of data availability. In order to address such challenges, the need for a flexible and scalable cyberinfrastructure (CI) has been articulated in many workshops and meetings of the environmental and hydrologic engineering communities. Specifically, one of the stated thrusts of the Consortium of Universities for the Advancement of Hydrologic Science (CUAHSI) is to enable new science though data discovery, delivery, publication, and curation (http://www.cuahsi.org/his.html). The CI needs for the WATer and Environmental Research Systems (WATERS) community have been described in the WATERS Network Cyberinfrastructure Plan (http://www.watersnet.org/plngdocs.html). A CI that supports the incorporation of data directly into a full range of models, calibration of models, and comparisons with observations has been identified as a key element that will lead to improved understanding and predictability.

An integrated multidisciplinary project with participation of computer scientists, environmental engineers, hydrologists, atmospheric scientists, and education specialists is currently underway. Under the title of Cyberinfrastructure for End-to-End Environmental Explorations (C4E4), this multi-disciplinary team utilizes current developments in data engineering: distributed storage, cataloguing, metadata management, data transfer, data mining, and data fusion to address hydrologic and environmental questions, as well as environmental data management and integration in real-world settings. The C4E4 allows researchers to combine heterogeneous data resources with state-of-the-art modeling and visualization tools through a web portal. As a prototype, fusion of data and models has been undertaken by the C4E4 team over the St. Joseph watershed in Northern Indiana. The scientific questions that drive C4E4 are largely related to the effects of human activity on the quality of air, land, and water resources. The computation-intensive models are supported by resources on the TeraGrid.

Apart from fulfilling a research mission, C4E4 also integrates teaching and learning forum. This is achieved by vertical integration of undergraduate, graduate and post doctoral researchers, working and utilizing different components of the project for their research. The C4E4 framework is also being used for the informal science education mission by providing linkages with university extension services. This paper describes the vision, status of ongoing coupling activities, and the next steps for the C4E4/TeraGrid exploratory. The paper builds on a prior paper by Zhao et al. (2007), which outlined the water quality/hydrological modeling interface using a Surface Water Analysis Tool (SWAT). The focus of the current paper is to discuss the meteorological and regional hydrometeorological setup. The capability of coupling models of different scales through shared datasets and easily comparing model output with observations provides an important tool for students who conduct research in hydrometeorological and climate studies.

The integrated vision is presented in section 2. Our past work on SWAT modeling is described in section 3, followed by the data integration activity in section 4. Section 5 lists a summary of models and the design configurations over the C4E4 study region. Section 6 summarizes the educational impact.

2 An Integrated Vision

The C4E4 vision is a CEO:P (Cyberinfrastructure for Environmental Observatories: Prototype) framework to link the feedback between surface hydrology and the water cycle for assessing the impacts of water quantity and water quality (Govindaraju et al., 2008). Our prior efforts have been directed towards the Soil Water Assessment Tool (SWAT) (Zhao et al., 2007) and surface stream flow hydrology using distributed resources. The ongoing efforts are now directed towards the development of capabilities for surface water hydrology as well as for hydrometeorological assessments. An inherent component of this effort is the development of capabilities to assimilate satellite remote sensed products. Our vision for this development is briefly summarized in this section.

The heterogeneous modeling system and tools being integrated must be capable of representing and/or analyzing phenomena that are at different scales – urban scale flash flooding to regional scale climate change impacts on stream flow levels. This mandates availability of different tools with varying resolutions (temporal and spatial).The point observations form the first set of data that need to be compiled and made available for the C4E4 domain. A Google Map based interface that provides the surface datasets (Fig. 1) was designed and implemented. The requirement of gridded long term rainfall and surface meteorology was addressed by adopting the North

American Regional Reanalysis (NARR) (Mesinger et al., 2006) output tailored over the Midwest US. To assimilate heterogeneous observations including satellite products, we adopt the Land Data Assimilation System (LDAS) (Mitchell et al., 2000, Chen et al., 2007) framework that can generate surface fields. The need for developing coupled hydrometeorological modeling capabilities was met by configuring a WRF modeling system with telescopic domain covering Midwest and Indiana that can take advantage of the LDAS fields. A common theme in the different studies is the rainfall amounts and distribution. For SWAT runs, the rainfall can be from rain gauges, as well as from the NARR data. For LDAS, the rainfall and soil / vegetation fields can be imported from files prepared for SWAT. The LDAS fields are then used as input to run WRF. The approach used in adopting these modeling tools is based on the choice of spatial and temporal scales needed to resolve the phenomenon of interest.

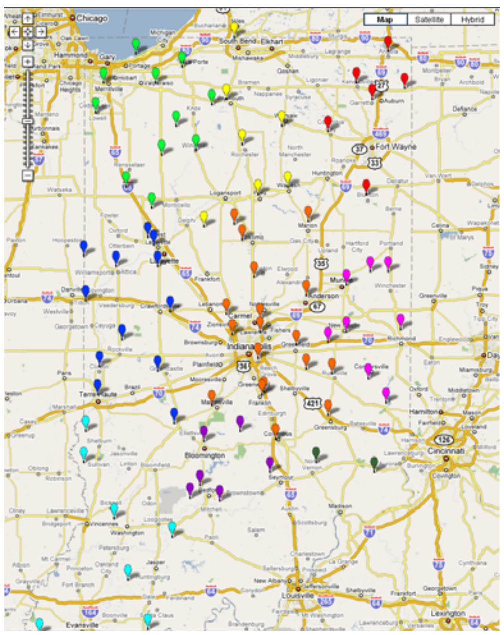


Fig. 1. A Google Map based surface dataset for Indiana (source: Indiana State Climate Office, www.iclimate.org).

We anticipate that as the system is completed, a user will have the capabilities to conduct SWAT based analysis using a choice of point, artificial, or NARR based rainfall products, Pre-configured HRLDAS (High Resolution LDAS) [Chen et al, 2007] setup to generate surface fields for soil moisture, soil temperature, and surface meteorological fields at 1 to 4 km grid spacing over Indiana will also be available by utilizing NARR fields as the meteorological driver. Alternatively, saved HRLDAS output will serve as input to the Weather Research and Forecasting (WRF) model for episodic coupled runs at high resolution (~1 km to 10 km) over the C4E4 region. We envision that the AMSR and MODIS satellite swaths will be integrated within the HRLDAS setup to develop gridded output that will be archived and accessible by users. At a longer time scale, we anticipate developing capabilities to run coupled WRF model and benefit from the WRF Chem air quality modeling efforts as they become available. Variables related to cloud, water vapor and aerosols are provided by the Purdue Terrestrial Observatory satellite, and will be applied to the WRF Chem model. This end-to-end assessment of the environmental state over the C4E4 study area is the broader vision of this multi-model based system.

Fig. 2 outlines the direct linkages that are ingrained within the setup at this stage. The second – order interactions such as using HRLDAS soil moisture fields for SWAT are not shown but are also envisioned as the system matures.

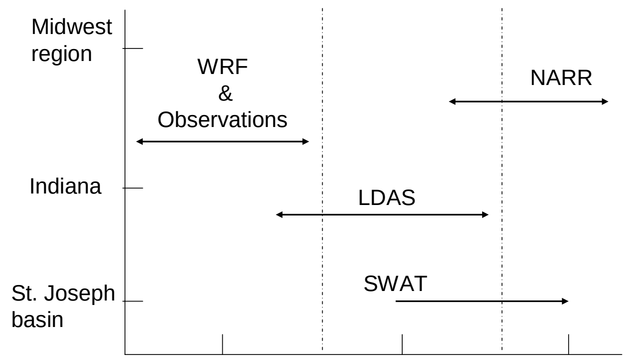


Fig. 2. Multi-model schematic with various spatial and temporal scales.

3 PAST WORK FOCUSING ON SWAT MODEL

As an initial step of the C4E4 project implementation, we have designed and developed a distributed workflow framework enabled by TeraGrid resources. The goal of the system is to provide systematic support to run dynamic, end-to-end, data-driven environmental studies which often include data retrieval, transformation, model simulation, data analysis, and visualization. We achieved this goal by first developing reusable service modules that perform specific tasks. These tasks consist of both common operations, such as submitting jobs to the TeraGrid Condor pool, and generating plots for time series data, and operations that are specific to a data source or a model simulation. All the service modules implement standard web services interfaces. They can be easily imported into our workflow runtime environment which is built upon the JOpera project (Pautasso, 2005) developed at ETH Zurich. Furthermore, we designed and implemented runtime workflow status monitoring services and integrated them with the workflow environment. It enables users to receive automatic updates on the execution status of the numerical experiments they have launched.

This workflow framework has been used to help researchers and students perform scientific investigations on the impact of land management practices in complex watersheds using SWAT on the TeraGrid. The SWAT model is a comprehensive watershed management tool developed by USDA Grassland, Soil and Water Research Laboratory. It is a process-based distributed-parameter simulation model consisting of components such as weather, surface runoff, return flow, pond and reservoir storage, crop growth and irrigation, nutrient and pesticide loads, and water transfer. The study using SWAT in the C4E4 project is mainly focused on the St. Joseph watershed in northern Indiana using heterogeneous data input from hydrology, water quality, atmospheric, and other related disciplines. As shown in Fig. 3, researchers are able to construct SWAT modeling pipelines that compose and run SWAT simulations on the TeraGrid Condor cluster, fetch the output, transform, plot, publish the result, and finally send an email notification to the user.

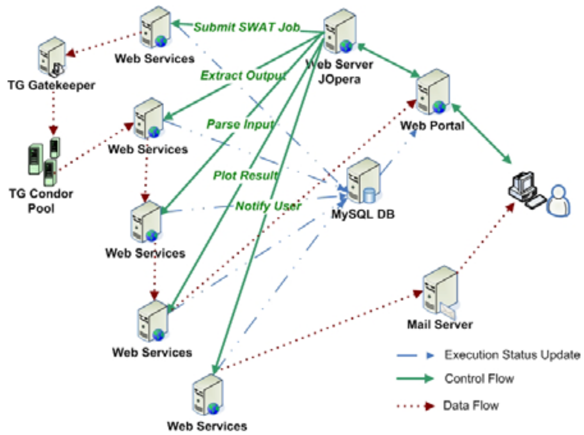


Fig. 3. Distributed SWAT workflow execution (Zhao et al., 2007).

Because SWAT is a conceptual model, calibration of its parameters to reproduce observed stream flow is the first step in setting up the model to make future predictions. Model calibration can take anywhere from few days to several weeks depending on the size of the study area, calibration period and number of targeted parameters. The SWAT workflow using TeraGrid enables parallel calibration of several models to study the effect of extreme events (eg. droughts) and management practices on watershed hydrology. Fig. 4 is a SWAT visualization data product that compares the model stream flow output with the observed monthly values for the St. Joseph River. It demonstrates the framework's capacity to support scientific investigations and visualization experiments. As a result, researchers are able to focus on high level application questions instead of on low level details about data manipulati

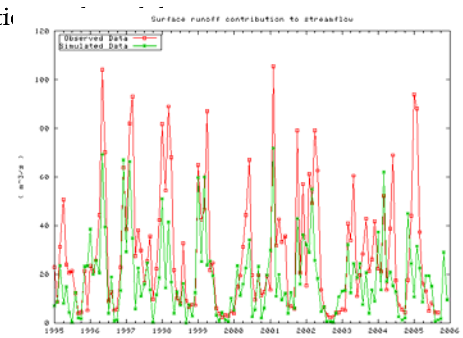
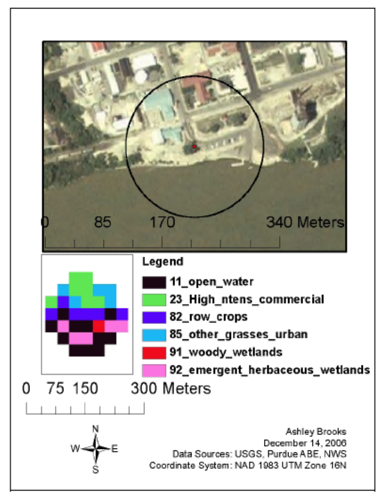


Fig. 4. Visualization of SWAT simulation result: surface runoff contribution to stream flow compared to the measured data downloaded from United States Geological Survey (USGS) National Water Information System (NWIS) website (Zhao et al., 2007).

4 Data Integration Activity

The C4E4 water quality area is over a watershed. To develop a hydrometeorological study analysis over this region, a broader domain is designed. Therefore, data integration activity was initiated by developing a meta-analysis of the data sources and inventory. This included site visits, photographing the difference sites and developing partnerships with local National Weather Service (NWS) offices. This activity is summarized in a MS thesis by Brooks (2007). Fig. 5a and 5b are examples of the meta data compiled as part of this initiative. The next step in this effort was the development of quality control protocols for the various meteorological data sources. This was achieved by accessing digital forms from various NWS personnel, as well as cross check with the information archived at National Climatic Data Center (NCDC). Fig. 6 shows a station's historical information and quality control that was achieved as part of this dedicated effort. The final step of this activity was to collate this information and develop a Java-based Google Map interface shown in Fig. 1.



(a)

(b)

Fig. 5. (a) 100-meter Mont Vernon DOQQ and land use; (b) a view of the Rochester,IN sensor while facing north. Subsequent images are facing north, south, east and west away from the sensor. (Source: Brooks, 2007).

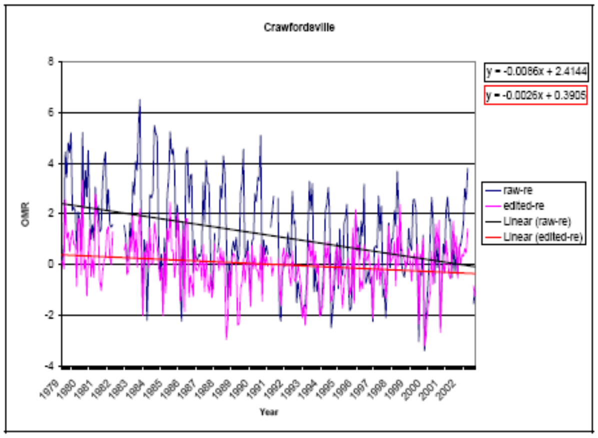


Fig. 6. Crawfordsville's raw and reanalysis trends (Source: Brooks 2007).

5. Hydrometeorological Modeling and Analysis

We have been developing components of a land data assimilation system (HRLDAS) that is representative of the regional characteristics. The C4E4 study region is mostly agricultural landscape, but it also has several medium to large townships. Both of these processes: urban as well as agriculture based are generally grossly represented in the current LDAS setup. So the initial part of our development effort was directed towards testing the agricultural landscape representations using corn, soybean, agriculture/grassland and urban landscape energy and water balance. These developments were tested over a field experimental site over Southern Great Plains and monitoring sites in southern Indiana (Morgan Forest) and Illinois (Bondville) where intensive and detailed field observations were available for testing the framework.

The formulation and stability of HRLDAS system were tested by Chen et al. (2007). They found that the HRLDAS was capable to reproduce observations and regionally consistent energy and water vapor fluxes. In a follow up assessment, LeMone et al. (2008) tested the sensitivity of different surface characteristics on the HRLDAS output. Soil texture, vegetation cover and soil moisture were found to be important input variables in order to develop realistic surface fields. The system was then modified by Niyogi et al. (2006) and Kumar et al. (2008) to include the crops and photosynthesis pathways (C3 and C4 as representative of corn and soybean and grasslands in the C4E4 domain). We found that accurate representation of leaf and canopy stomatal response was important for developing realistic water vapor exchanges between land and the atmosphere. We evaluated the impacts of water vapor formulation on the seasonal evolution of plant transpiration and soil moisture storage. The results produced by HRLDAS using the current simple Jarvis canopy resistance scheme in Noah were compared to the results using GEM. Long-term soil moisture data from the Oklahoma Mesonet and surface sensible heat flux and evaporation data obtained from 10 sites, located at different climate regimes and land-use covers, during the International H2O project (IHOP-02) field experiment are used validate these results. The impact of urban landscapes is still evolving and being added to the system. Initial results were performed over the Southern Great Plains by Niyogi et al. (2006) and Holt et al. (2006) using the coupled modeling framework.

In general, the WRF model runs are limited to a short term simulation for studying the different processes. In order to assess longer term feedbacks, the framework relies on HRLDAS fields and NARR products.

5.1 WRF Modeling System

A nested domain with 9, 3 and 1 km horizontal resolution centered at Indiana is designed to carry out simulation using WRF (Weather Research and Forecasting) model. WRF is a fully compressible non-hydrostatic, primitive-equation model with multiple-nesting capabilities to enhance resolution over the areas of interest. This version of the WRF model uses the Eulerian mass coordinate and is referred to as the Advanced Research WRF (ARW). Domain 1 is the coarsest mesh and has 250 x 250 grid points in the north-south and east-west directions, respectively, with a horizontal grid spacing of 9 km. Within Domain 1, Domain 2 is nested with 400 x 400 grid points at 3 km grid spacing. The fine-mesh domain, Domain 3, is 400 x 400 points with 1 km grid spacing and can run 3-way nesting including 1 km domain over Indiana specifically to capture the effect of surface heterogeneity (Fig. 7).

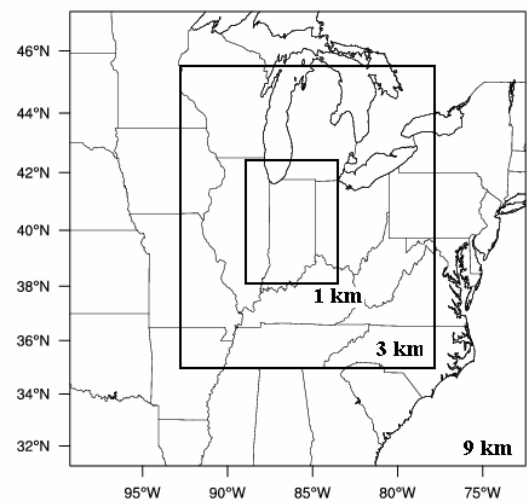


Fig. 7. WRF model domain setup.

The initial and boundary conditions are provided from 6-hourly NCEP FNL 1-degree data and WRFSI (WRF standard initialization) is used to interpolate 1-degree data to model grid and provide initial condition for start date of simulation. The 3 km nested domain gets it initial and boundary condition from coarser domain. There are 31 vertical layers are available in model to define atmospheric structure. The initial and boundary conditions for the large-scale atmospheric fields, sea surface temperature (SST), as well as initial soil parameters (soil water, moisture and temperature) are given by the 1x1 degree obtained from NCEP Global Final Analysis (FNL) and provided the all this initial condition to 27, 9 and 3 km WRF model grid. WRF model can run with Noah land-surface model and have various other schemes designed for cloud physics, radiation, and boundary layer schemes. More information can be found at http://www.mmm.ucar.edu/wrf/users/wrfv2/wrf_model.html .

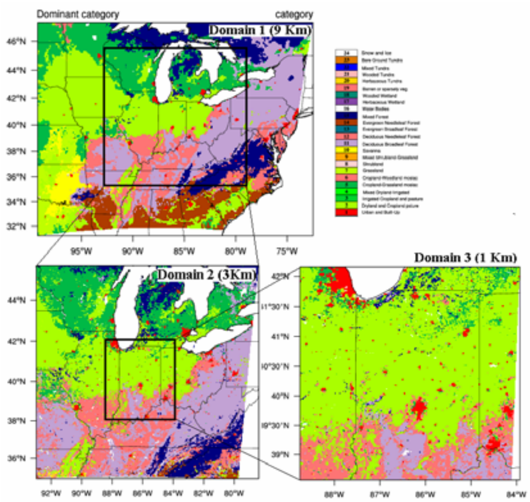


Fig. 8. Land use category in 9km, 3km and 1km resolution.

5.2 HRLDAS Modeling System

The high-resolution land data assimilation system (HRLDAS) is capable of capturing heterogeneity of soil and vegetation at a spatial scale between 1 and 10 km. In order to drive the HRLDAS model, the model was initialized with static land use and soil texture fields as well as time-varying meteorological fields such as a forcing field. The static fields were made with the help of a WRF Pre-processing System (WPS); the outer two domains had a 9 km and 3 km horizontal resolution, and the inner grid had a 1 km resolution for the Indiana (USA). The WPS system monthly vegetation field, terrain, land-water masks, landuse, and soil texture maps were interpolated and assigned surface fields at each grid point of the HRLDAS. WRF and HRLDAS have similar output variables, such as pressure, temperature, heat fluxes, precipitation and radiation. The uncoupled HRLDAS can be used for long-term simulations. The input data for WRF model are: 4-km hourly NCEP Stage-II rainfall, 1-km land use type and soil texture maps, 0.5 degree hourly GOES downward solar radiation, 0.15 degree AVHRR vegetation fraction and T, q, u, v, from model based analysis (EDAS). HRLDAS output data will include: long term evolution of multi-layer soil moisture and temperature, surface fluxes, and runoff.

The HRLDAS modeling framework is designed to capture high resolution spatial and temporal surface parameters. We are currently running HRLDAS for 2001-2002 and 2005-2007 simulation years. From Fig. 8, we can see clearly the surface heterogeneity differences for 9, 3 and 1 km. The HRLDAS system can also ingest variety of satellite products. Efforts are underway to provide MODIS land fields to represent time varying vegetation changes over Indiana. The HRLDAS analyses are offline meteorology driven and are designed for monthly to seasonal/annual scales. We can conduct simulation using HRLDAS for any domain and resolution depending upon choice of study from local to regional scales. Fig. 9 shows the surface temperature verification from (a) WRF coupled with Noah LSM using USGS land use, (b) WRF coupled with Noah-GEM LSM using USGS land use, and (c) MODIS satellite land use data.

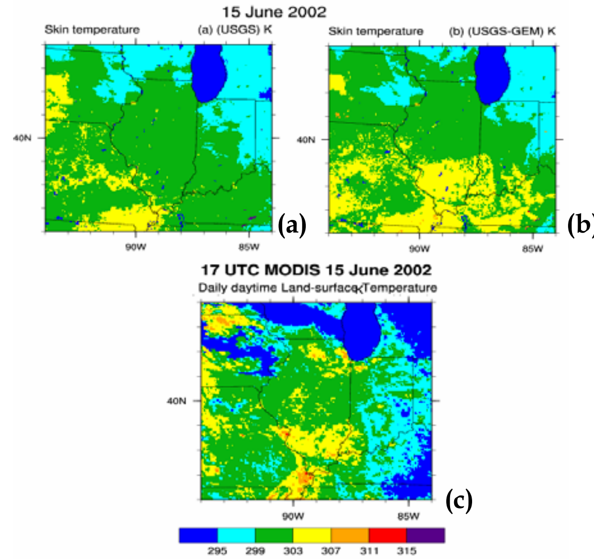


Fig. 9. WRF land surface temperature verification based on (a) Noah LSM with USGS land use, (b) Noah-GEM LSM with USGS land use, and (c) MODIS surface temperature observations.

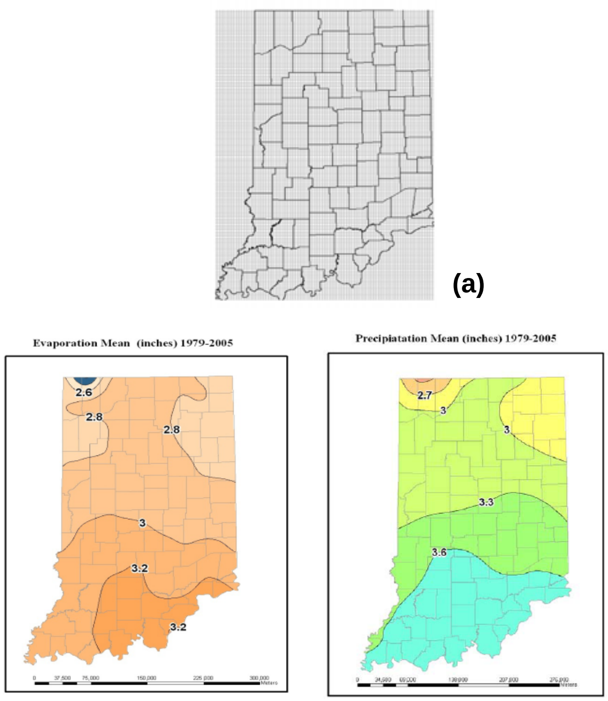
5.3 NARR

The third analysis system that has been integrated for the C4E4 cyber exploratory is a subset of the North America Regional Reanalysis (NARR) dataset. The NARR products are on Eta 221 grid (32km spatial resolution, Fig. 10a) at 29 vertical pressure levels, and include all in situ observations used in NCEP/NCAR Global Reanalysis project, and additional precipitation data, TOVS-1B radiances, profiler data, land surface and moisture data, etc. The output analyses are 3-hourly with additional 9 variables in the 3-hour forecasts to reflect accumulations or averages. These products are downloaded for the C4E4 domain and ideally suited for annual to decadal scale analysis. Fig. 10 shows the NARR grid points. A wide range of reanalysis fields (data) are available for each of the grid point with the C4E4 domain. The corresponding maps for annual mean evaporation and precipitation maps from the NARR fields from 1979 to 2007 are also shown (Fig. 10b, c). These data and mapping capabilities are available with a daily temporal resolution for the C4E4 region.

6 Educational Impact

Multidisciplinary graduate students working in the project have been able to get hands-on experience with applied high performance computation and use of the TeraGrid. The project has led to development of research topics for utilizing multi-model simulations, as well as understanding the urbanization effect on the hydrometeorological cycle. Formal classroom lectures have been designed building on this exploratory for hydrology, weather and climate, and land surface modeling courses. Results from some of the ongoing research tasks with active student involvement have been shown in this paper. We expect the exploratory to become a major source for engaging and developing multidisciplinary graduate education and training experience.

This cyber exploratory is modular and has active graduate and postdoctoral researcher involvement in its development, testing and implementation. The SWAT component of this multimodel framework was adopted as first step of the study, and has been tested over TeraGrid for the C4E4 purpose by a graduate student. Three independent but parallel subprojects, each involving a different graduate student and one post-doc, are underway to test and implement WRF, LDAS and NARR products over the C4E4 region.



(b) (c)

Fig. 10. (a) available NARR grid points from the C4E4 domain, the NARR annual mean maps for (b) evaporation and (c) precipitation.

The NARR products have been processed and analyzed, and are available over the C4E4 domain. A follow up activity will be developing the front end/GUIs for assessing the various NARR data via a pull down menu and pre-configured domain.

The LDAS has been tested for the Midwest region for its ability to reproduce annual and short term surface energy balance and hydrological fluxes as well as surface and subsurface hydrothermodynamic fields. Steps are underway to make a preconfigured setup available for the C4E4 domain. The LDAS analysis at this point will be done offline, i.e. will not be available to every user to make HRLDAS run over the TeraGrid. However, we plan to develop and keep HRLDAS fields and make them available to the users as they access C4E4 website.

The WRF model is a community tool available for download from NCAR. Current efforts are directed toward testing wide range of physical parameterizations available in WRF over the C4E4/Midwest domain. It is planned that a preconfigured WRF will be available for generating atmospheric fields for the graduate students to conduct high level if-then assessments. The users will have the ability to display and plot model outputs from a pre-selected pull down menu. This will balance the need for user centric model run capabilities and computational efficiency, particularly as the system is initially being deployed for graduate student training purpose.

In the next phase, we anticipate each of the tools (SWAT, WRF, LDAS and NARR data) would be thus configured over the C4E4 domain and accessible via TeraGrid. In a subsequent phase, efforts will also be directed towards wider dissemination of the tools to the broader education community for multiscale multimodel hydrometeorological analysis.

A variety of TeraGrid resources are being examined to make these tools available to larger community. These education resources, including models, analysis tools and tutorials, will be made available through a web-based science gateway with middleware that enables the models and tools to use the TeraGrid resources at the back end. The science gateway program in the TeraGrid project has provided support to a number of community based science gateways (e.g., nanoHUB, GridChem, BioPortal, and LEAD) and has been successful in bringing the powerful resources to the broader user communities. A prototype gateway, C4E4, is already developed, which supports a web enabled SWAT model, workflow, and aggregated data access. Additional models, analysis tools, tutorials, presentations will be added targeting use in classes for graduate students. A TeraGrid community account will be used to support gateway users, further lowering the barrier of entry.

The multi-scale models provide an excellent learning experience for the students for topics such as offline versus coupled models, optimizing resolution versus model domain and run duration, sensitivity of different input variables to the model output, and the uncertainty in the model results as a function of variance in the model input as well as parameterizations. In terms of informal science education, we have designed a Google Map interface for data access and weather visualization permanently archived and mirrored at the Indiana State Climate Office (www.iclimate.org). As part of this and a broader climate education initiative, the Indiana State Climate Office now compiles hourly and daily climate data from sources including the cooperative station network, local airports, and Purdue agricultural research centers. Various weather variables are now available, including air temperature, precipitation, snowfall, dewpoint, wind, barometric pressure, sky and present weather conditions, solar radiation, soil temperature, and others. We are also exploring ways of transferring this information to middle and high school teachers that seek to understand climate change via visualization modules and results from the

TeraGrid based model results will provide one avenue for generating local examples.

ACKNOWLEDGMENT

The C4E4 project is supported by the National Science Foundation under Grant No: 0619086. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. Study also benefited in parts by NSF-ATM 0233780 (S. Nelson), NASA–THP (J. Entin NNG05GQ47G and NNG06GH17G), NASA-IDS (NNG04GL61G J. Entin and G. Gutman), NASA LCLUC Program (G. Gutman).

REFERENCES

[1] Brooks, A. V., "Assessment of the Spatiotemporal Impacts of Land Use Land Cover Change on The Historical Climate Network Temperature Trends in Indiana", MS Thesis (D. Niyogi and M. Baldwin - Co-Chair), 2007.

[2] Chen, F., and coauthors, "Description and Evaluation of the Characteristics of the NCAR High-Resolution Land Data Assimilation System" J. Appl. Meteorol. and Clim. 46, 694-713, DOI: 10.1175/JAM2463.1, 2007.

[3] Govindaraju, R.S. and coauthors. "A vision of cyberinfrastructure for end-to-end environmental explorations (C4E4)", ASCE J. Hydrologic Eng., accepted March 2008.

[4] Holt, T., and coauthors, "Effect of Land - Atmosphere Interactions on the IHOP 24-25 May 2002 Convection Case", Mon. Wea. Rev., 134, 113 – 133, 2006.

[5] Kumar, A., and coauthors, "Using photosynthesis-based canopy resistance model and new MODIS-based data to improve the presentation of vegetation transpiration in the Noah land surface model", 22nd Conf. on Hydrology and 20th Conf. on Climate Variability and Change, AMS Annual Meeting, New Orleans, LA, 2008.

[6] LeMone, M. A., and coauthors, "Adding Horizontal Heterogeneity as a Criterion for Evaluating a Land-Surface Model", Mon. Wea. Rev., accepted January 2008.

[7] Mesinger, F., and coauthors, "North American regional reanalysis", Bull. Amer. Meteor. Soc., 87, 343–360, 2006.

[8] Mitchell, K., and coauthors, "The collaborative GCIP Land Data Assimilation (LDAS) Project and supportive NCEP uncoupled land surface modeling initiatives", Preprints, 15th Conf. on Hydrology, Long Beach, CA, Amer. Meteor. Soc., 1–4, 2000.

[9] Niyogi, D., and coauthors, "Urban and land surface effects on the 30 July 2003 mesoscale convective system event observed in the Southern Great Plains", J. Geophys. Res., 111, D19107, doi:10.1029/2005JD006746, 2006.

[10] Pautasso, C., "JOpera: An Agile Environment for Web Service Composition with Visual Unit Testing and Refactoring", VL/HCC, 2005.

[11] Zhao, L., and coauthors, "Interweaving Data and Computation for End-to-End Environmental Exploration on the TeraGrid", Proceedings of the TeraGrid 2007 Conference, Madison, WI, June 2007.