Long-Term Agroecosystem Research in the Central Mississippi River Basin: Introduction, Establishment, and Overview


Abstract

Many challenges currently facing agriculture require long-term data on landscape-scale hydrologic responses to weather, such as from the Goodwater Creek Experimental Watershed (GCEW), located in northeastern Missouri, USA. This watershed is prone to surface runoff despite shallow slopes, as a result of a significant smectitic clay layer 30 to 50 cm deep that restricts downward flow of water and gives rise to a periodic perched water table. This paper is the first in a series that documents the database developed from GCEW. The objectives of this paper are to (i) establish the context of long-term data and the federal infrastructure that provides it, (ii) describe the GCEW/ Central Mississippi River Basin (CMRB) establishment and the geophysical and anthropogenic context, (iii) summarize in brief the collected research results published using data from within GCEW, (iv) describe the series of papers this work introduces, and (v) identify knowledge gaps and research needs. The rationale for the collection derives from converging trends in data from long-term research, integration of multiple disciplines, and increasing public awareness of increasingly larger problems. The outcome of these trends includes being selected as the CMRB site in the USDA–ARS Long-Term Agro-Ecosystem Research (LTAR) network. Research needs include quantifying watershed scale fluxes of N, P, K, sediment, and energy, accounting for fluxes involving forest, livestock, and anthropogenic sources, scaling from near-term point-scale results to increasingly long and broad scales, and considering whole-system interactions. This special section informs the scientific community about this database and provides support for its future use in research to solve natural resource problems important to US agricultural, environmental, and science policy.

Agriculture faces environmental challenges that require increasingly longer-term, broader-scale, and more-integrative transdisciplinary data to inform public policy toward solutions. Several trends in science, in scientific instrumentation, equipment, and technology, and in public awareness converge to make ready access to these and similar data increasingly important.

Science has evolved through three phases into what is sometimes termed the Fourth Paradigm (Gray, 2009). It was at first empirical, with a focus on observing phenomena. Some 200 to 400 years ago, scientists began to develop theoretical models to explain those observations. In the latter half of the 20th century, these analytical models were insufficient to capture the complexity of knowledge about interconnected systems, and numerical simulation models became the norm. This phase was termed the computational phase by Gray (2009). The Fourth Paradigm, data-intensive scientific discovery, from the book of the same name (Hey et al., 2009), describes the changes in data volume, type, documentation, access, and communication to enable data exploration needed for current science. This phase he terms variously eScience, data-intensive science, or data exploration. Important for this series of papers, Gray (2009) emphasized the need for data to be self-described, meaning that metadata, including documentation of methods, units, context of collection, responsible agents, provenance, and so on, must be permanently connected with the data both in storage and in access.

E.J. Sadler, R.N. Lerch, N.R. Kitchen, C. Baffaut, K.A. Sudduth, R.J. Kremer (retired), and E.D. Vories, USDA–ARS, Cropping Systems and Water Quality Research Unit, 269 Agricultural Engineering Building, Univ. of Missouri, Columbia, MO 65211; S.H. Anderson and R.J. Miles, Dep. of Soil, Environmental and Atmospheric Sciences, School of Natural Resources, 302 Anheuser-Busch Natural Resources Building, Univ. of Missouri, Columbia, MO 65211; A.A. Prato, (retired), Dep. of Agricultural and Applied Economics, Univ. of Missouri, Columbia, MO 65211; D.B. Myers, Division of Plant Sciences, 214 Waters Hall, Univ. of Missouri, Columbia, MO 65211; R. Broz, MU Extension, 205 Agricultural Engineering, Univ. of Missouri, Columbia, MO 65211; F.J. Young, USDA–NRCS (retired), 601 Business Loop 70 West, Parkade Center, Suite 250, Columbia, MO 65203-2546. Assigned to Technical Editor Scott Bradford.

Abbreviations: CEAP, Conservation Effects Assessment Project; CMRB, Central Mississippi River Basin; GCEW, Goodwater Creek Experimental Watershed; HUC, hydrologic unit code; LREW, Little River Experimental Watershed; LTAR, Long-Term Agroecosystem Research; LWREW, Little Washita River Experimental Watershed; MOW, Mahantango Creek Watershed; MSE4, Management Systems Evaluation Areas; RCEW, Reynolds Creek Experimental Watershed; SBR, Salt River Basin; WGEW, Walnut Gulch Experimental Watershed.
Technological trends have produced enabling technology that increases the volume of data that can be collected, stored, and accessed. Gray’s (2009) thesis was that software tools for data management, visualization, and analysis lagged behind the need for them. However, anyone who worked with geospatial data before the common availability of modern GIS tools is fully appreciative of their utility for their purpose. Similarly, modern database tools facilitate data management well beyond the capabilities of first-generation programming languages. In a number of disciplines, the term informatics is used to describe both the extensive databases and the informational infrastructure supporting them. Michener and Jones (2012) describe what is needed in the data domain to enable einformatics, including integrative platforms, adoption of standard protocols, and good data stewardship practices, but also point out that sociocultural changes are required. These include promoting informatics literacy, data sharing, and scientific transparency and reproducibility. Knowledge of both physical and sociological changes would be critical for understanding the type and rate of ecological and environmental change.

One of the sociocultural changes needed, data sharing, is emerging as a new standard in the publication world. Journals that have incorporated access to supplemental online material are good examples toward this end. This journal’s effort to make high-quality data available for research represents a major advance. Still, Gray’s (2009) point remains valid that a substantial need exists for laboratory information management systems, data collection and visualization tools, and access and communication tools.

It appears that the drivers motivating research have also evolved. Much of the existing literature originated from research that was reductionist in approach, used controlled experimental design, focused on a point scale and short time frame, and was often contained wholly within one discipline. However, we are repeatedly asked in current requests for proposals to do multidisciplinary or transdisciplinary research. That research is often integrative across spatial scales and longer term, as it attempts to explain phenomena that interact among landscape or watershed elements and that are dependent on highly variable weather drivers.

Such research often combines hydrology and ecology, often termed ecohydrology. Moran et al. (2008) asserted that ecohydrology includes feedback mechanisms, gradual trends, and extreme events that are difficult to detect except with long-term hydrological and biological studies. They showed how USDA long-term multisite data were used to understand these key issues, with specific examples illustrating time lag between cause and effects, critical thresholds and cyclic trends, context of rare and extreme events, and mechanistic feedbacks for simulation modeling. The specific example of detecting trends in long-term studies has gained recognition as critical, even to the point of literature assertions that assumptions of stationarity, commonly used in descriptive statistics in climate and water research, are not rigorously applicable (Milly et al., 2008). Another use of long-term, multisite data is to generalize results, obtained in short-term, point-scale studies, that often appear contradictory in the literature (e.g., Ponce Campos et al., 2013). Dozier and Gail (2009) termed such research the science of environmental applications and differentiated it from basic environmental science in that it is need-driven instead of question-driven, and results are scalable, robust, and useful even when incomplete.

These changes in environmental sciences, used broadly, have occurred in response to changes in the scope of problems science is being applied to solve, to changes in public policy to address those problems, and to changes in awareness of both the general public and the scientific community of the problems and the science available to apply to them. In anticipation of and in response to the need for these data, federal research infrastructure at the national and local level has been implemented. Data obtained from that infrastructure must be published into the public domain to be most useful, and both accessibility and metadata must be sufficient for later use.

For any individual database, future users must be able to find a number of types of metadata. They must be able to determine the context of databases that have either similarities or differences from the subject, to allow hypothesis testing across contexts. They must be able to examine information about the physical setting and how that impacts interpretation of data. They must have information about how the land has been and is being used, and how that may affect interpretation of research results. They need information on what research has already been conducted in the physical space, with the data, or that may have particular relevance to their future research.

The Goodwater Creek Experimental Watershed (GCEW) in northeastern Missouri, USA, has developed one such database. The GCEW has been part of a broader USDA–ARS watershed network for more than 40 yr and has a valuable store of long-term watershed hydrological, meteorological, and water quality data. While some of the metadata required to support it exist in the literature, it previously was widely scattered, and other critical information was lacking. Knowledge of the substantial body of published research from GCEW and the surrounding area would enhance future use of these data. To develop the papers that make up this special section in the Journal of Environmental Quality, the authors considered what would be needed to use the data in the database. Therefore, the objectives of this paper are (i) to establish the context of long-term data and the federal infrastructure that provides it, (ii) to describe the GCEW/ Central Mississippi River Basin (CMRB) establishment and the geophysical and anthropogenic context, (iii) to summarize in brief the collected research results published using data from within GCEW, (iv) to describe the series of papers this work introduces, and (v) to identify knowledge gaps and research needs. (References specific to this section are included in the online Supplement S1, Section A.)

### Goodwater Creek Experimental Watershed within the USDA–ARS Watershed Network

During the past 15 yr, repeated calls have been made for a Long-Term Agroecosystem Research, or LTAR, infrastructure (NRC Committee on U.S. Geological Survey Water Resources Research, Water Science and Technology Board, Commission on Geosciences, Environment, and Resources, 1997; NRC Committee on Watershed Management, Water Science and Technology Board, Commission on Geosciences, Environment,
and Resources, National Research Council, 1999; NRC Committee on Opportunities in Agriculture, National Research Council, 2003; Robertson et al., 2008; NRC Committee on Twenty-First Century Systems Agriculture, Board on Agriculture and Natural Resources, Division on Earth and Life Studies, 2010; Glasener, 2010). Compelling arguments were presented in these calls for the value such a network would represent toward answering research questions about long-term trends in landscape-scale environmental impacts of managed agricultural ecosystems. Such research is necessarily multidisciplinary, challenges classical reductionist science, and relies heavily on tools such as process-based watershed-scale models to integrate small-scale process knowledge into increasingly larger and more policy-relevant scales. Particularly compelling is that short-term variation in weather, long-term trends in climate, and both gradual and abrupt shifts in agricultural management combine to require spatially and temporally intensive measurements to be sustained for periods often longer than the careers of single scientists. The duration, in particular, places a substantial value on the quality of the documentation supporting the research. Availability and access to the data can be challenging.

A substantial contribution toward such an LTAR network existed in both federal and university infrastructure. The federal context is quite relevant to the current paper. The USDA–ARS has a current network of approximately 28 watersheds, experimental ranges, and other research facilities that have been established to collect long-term physical, chemical, and biological data on many aspects of agriculture at the watershed or landscape scale (Walbridge and Shafer, 2011). Periods of record in the ARS network range from 5 to 98 yr, with a number exceeding 30 yr and thus quite valuable for analyses dependent on long-term variation. The first two of these, the North Appalachian Research Watershed near Coshocton, OH, and the Riesel Watershed near Temple, TX, were established in the 1930s to study field-to-farm-scale impacts of agriculture (Harmel et al., 2007, 2014).

During the 1960s, six regional watersheds were established that had been authorized by Senate Bill 59 (USDA, 1959; Farrell, 1995). These include the Walnut Gulch Experimental Watershed (WGEW) in southeast Arizona to represent the semiarid southwest (Moran et al., 2008), the Reynolds Creek Experimental Watershed (RCEW) near Boise, ID, to represent the mountain northwest (Marks, 2001), the Little River Experimental Watershed (LREW) in southern Georgia to represent the Gulf-Atlantic Coastal Plain (Bosch et al., 2007), the Mahantango Creek Watershed (MCW) to represent upland agricultural watersheds in the Appalachian Valley and Ridge Physiographic Province (Bryant et al., 2011), the Little Washita River Experimental Watershed (LREW) to represent the southern Great Plains (Garbrecht et al., 2007; Steiner et al., 2008b, 2014), and the Goodwater Creek Experimental Watershed (GCEW) to represent the north-central region (described below, starting with the 1959 Senate action).

Of these six, five have had database documentation paper series published. Marks (2001) introduced nine other papers for RCEW, Bosch et al. (2007) was the first of five for LREW, Moran et al. (2008) introduced 19 for WGEW, comprising both technical notes on data and research papers using them, Bryant et al. (2011) was the first of four for MCW, and Steiner et al. (2014) was the first of 12 for LWREW. The sixth, this series for GCEW, is described below. Two additional ARS watersheds, established in the 1990s, have shorter periods of records but are substantially documented. Hatfield et al. (1999) was the first of seven research articles based on the North Walnut Creek Watershed near Des Moines, IA, which represents the central Iowa and Minnesota Till Prairies and Des Moines Lobe physiographic regions of the Corn Belt. Locke (2004) introduced 17 articles with primary focus on the Beasley Lake Watershed, an oxbow lake that is a cutoff meander of the Big Sunflower River in the Lower Mississippi River floodplain.

As a result of their extensive infrastructure, sustained commitment toward maintenance, online availability and accessibility, and extensive documentation of the databases, all of these watersheds were considered as candidates for membership in an LTAR network as described above. In fall 2011, ARS announced its intent to organize selected outdoor laboratories into an ARS LTAR network (Walbridge and Shafer, 2011). The selection of the initial 10 members, made after a competitive process with external peer review of applications, was formally announced in September 2012, and an additional 8 were named in 2014 (http://ars.usda.gov/ltar). One of these 18, the CMRB LTAR, is anchored on the GCEW and is augmented by collateral infrastructure developed by the ARS in Columbia, MO, and by their cooperators. (References specific to this section are included in Supplement S1, Section B.)

**Goodwater Creek Experimental Watershed Establishment and Context**

**Senate Bill 59: 1971–1990**

In 1959, the US Senate (USDA, 1959) called for soil and water research to be done that represented numerous physiographic regions of the United States, with the rationale that spatial variation in soil resources, weather, and agricultural management made regional research results more relevant than the transfer of research data from different regions. Included in this plan were the several regional hydrologic watersheds mentioned above. In 1961, the North Central Hydrologic Laboratory was authorized as the fourth of these, after the northwest in Boise, ID (RCEW), south central in Chickasha, OK (LREW), and southwest in Tucson, AZ (WGEW). Placement in the northeastern Missouri region leveraged the historical ARS runoff and erosion research at the small plot scale at the Midwest Claypan Experimental Farm near McCredie, MO (Jamison et al., 1968). A nested design at four scales was selected, starting in the headwaters of Goodwater Creek near Centralia, MO, and following Goodwater Creek to the Young’s Creek watershed outlet (Fig. 1). Weirs were placed at 12, 31, and 73 km² along the main stem, with the USGS station at Young’s Creek as the fourth scale at 189 km². However, the Young’s Creek site was discontinued by USGS during construction of the three weirs, and the plan was reduced to the three-scale nested design, with initial data obtained 1 Apr. 1971. Initial research emphasis was on rainfall-runoff relationships, and the rainfall and streamflow measurements were supported by evaporation pan and daily extremes of air temperature and humidity data. Occasional research focused on groundwater elevation, sediment delivery, and nitrate water quality.

In 1991, the Management Systems Evaluation Areas (MSEA) were established as part of the President’s Water Quality Initiative; one of those areas was in GCEW (Ward et al., 1994). The USDA–ARS, the University of Missouri, USGS, USEPA, and other partners enabled continuous water quality measurements at the three stream weirs and added substantial focused research on management system effects on water quality at selected areas within GCEW. Weirs equipped with flow measurement and water samplers were placed at edges of three fields draining areas ranging from 7 to 35 ha. A three-replicate, 30-plot design of 0.34-ha plots, each plot encompassing a toposequence of soils, was placed adjacent to the largest of these fields. Selected plots were instrumented for seasonal measurements of pesticide and dissolved nutrient water quality from surface discharge and root-zone leaching. An automated weather station was placed at the site to provide both daily and diurnal data for common weather parameters. As the primary focus of the MSEA project was on groundwater impacts of agricultural production, groundwater well nests were placed in the fields and elsewhere in GCEW for a total of 88 groundwater monitoring wells. Later in this period, funding constraints prompted the decommissioning of the two smaller stream weirs and the two smaller field weirs, leaving only the GCEW outlet and the 35-ha Field 1 in operation (Fig. 1).


In 2003, the ARS and NRCS partnered to establish the Conservation Effects Assessment Project (CEAP) to determine effectiveness of cost-share practices under the Conservation Title of the Farm Bill (Mausbach and Dedrick, 2004). The GCEW was chosen as one of the ARS benchmark watersheds (Richardson et al., 2008) to support evaluations at the scale of the Mark Twain Lake watershed, of which Goodwater Creek is a headwater stream (Fig. 2). As part of that project, autosamplers were deployed at eight USGS flow measurement sites and four additional sites (Lerch et al., 2008). Two of the latter were within GCEW’s parent 10-digit hydrologic unit code (HUC) watershed, Long Branch (480 km²), which includes a USGS flow site (Lower Long Branch). Thus, measurements at GCEW were augmented with further nested and paired measurements to the 10-digit scale. The two added were at the historical Young’s Creek site (189 km²) and in parallel up the main branch, which was named Upper Long Branch (185 km²). All of these flow through the Lower Long Branch site into Mark Twain Lake. In parallel with the Long Branch scale were South Fork Salt River (589 km²), Elk Fork (517 km²), and Middle Fork (865 km²) in the Middle/South Fork Salt 8-digit HUC. In the North Fork Salt 8-digit HUC were Crooked Creek (212 km²), Otter Creek (228 km²), North Fork at Shelbina (1183 km²), and Black Creek (286 km²). In the Lower Salt 8-digit HUC were...
Lick Creek (260 km²), which flows into Mark Twain Lake, and the outflow of Mark Twain Lake. Research conducted during the CEAP project focused on testing nutrient, pesticide, and sediment water quality and evaluating the effects of beneficial management practices. As part of the ongoing research on the 35-ha field watershed, a precision agricultural system (PAS) study was begun in 2004 (Lerch et al., 2005; Kitchen et al., 2005a). In 2005, a cooperative project involving University of Missouri and USDA–ARS scientists enabled both further research and stakeholder involvement in watershed management within GCEW.

As part of the national CEAP effort, a Web-based database system was developed to house and deliver hydrologic, weather, and water quality data from participating locations. The resulting database, named STEWARDS, for Sustaining The Earth’s Watersheds Agricultural Research Database System (Sadler et al., 2008; Steiner et al., 2008a, 2009a,b) is the current public access point for the GCEW database, and while data may also be delivered through some additional databases yet to be developed, STEWARDS is permanent.

Long-Term Agroecosystem Research: 2012 and Forward

In 2012, GCEW was selected as one of the 10 initial ARS members of the LTAR network, expanded in 2014 to 18 sites. This designation initiated a new era for long-term research in the region and broadened the mission from its focus in the 1990s on hydrology and water quality to the ecological implications of agricultural production. The LTAR network allowed for continued expansion of scale both outward and inward. It explicitly encompasses broader scales both in the region and across the continent, and it is intended to develop process-level understanding at fine scales. It also formally links the CMRB LTAR with other LTAR nodes to address continental-scale problems. (References specific to this section are listed in Supplement S1, Section C.)

Geophysical Setting and Anthropogenic Changes

Essential to interpreting agroecosystem watershed research is a consideration of context, in particular, natural and anthropogenic history from which relevance can be derived. Such context is needed for the GCEW and the associated Salt River Basin (SRB) in northeastern Missouri. This section provides a brief description of the geophysical setting and anthropogenic changes for the SRB that serve as metadata for understanding past and future agroecosystem watershed research. The soil landscapes of the SRB arise from (i) underlying sedimentary bedrock; (ii) layers of moderately dissected glacial tills that define the visible landforms; and (iii) soil profiles formed in till, pedisediments, and loess. The loess is a relatively continuous mantle of variable thickness except on areas of significant slope or instability. Key soil features correlated with landscape include loess thickness, smectitic clay mineralogy, depth to an argillic (claypan) horizon, clay content within the argillic horizon, and depth to a paleosol. These are important hydrologic drivers for watershed vulnerability. In just 200 years, human activity has greatly altered landscape resources. Greatest change to the landscape occurred during the first half of 20th century, when intense rainfall followed moldboard plow tillage and/or extended drought periods, and the resultant valley sedimentation and systematic channelization destabilized stream channels throughout the basin. The soil and water conservation response initially identified poor practices and developed alternatives to prevent further degradation. Later efforts restored lost ecosystem function. Undoubtedly conservation measures have helped, but signs of ongoing impairment persist within the region.

Readers who need more details on the geophysical setting and anthropogenic changes to the landscape are referred to the detailed discussion included in Supplement S2. Potential users of the data described in this series would have difficulty interpreting the data without the knowledge contained within it, which is critical metadata.

Geospatial Data Sources and Documentation

Geospatial datasets provide important information for understanding watershed and hydrological processes. Examples include point locations of measurement sites such as rain gauges, topographic databases important in flow routing, and various soil and land cover datasets important for modeling and interpreting the interaction of agroecosystems with hydrology. Associated with the ongoing monitoring of hydrology and water quality in GCEW and the larger SRB, numerous geospatial datasets have been developed. There are three classes of geospatial data: (i) GIS layers in the STEWARDS online database, (ii) publicly available data relevant to GCEW and CMRB, and (iii) locally developed or value-added geospatial data. As these geospatial data are not subject to the standard query in STEWARDS, they are described in Supplement S1, Section D.

Research Results from GCEW/CMRB

Little documentation of the watershed as a whole has been published, and only limited site descriptions have been collected into internal documents. The collection of papers introduced here serves to complete the documentation. On the other hand, an extensive body of literature has been published from individual studies and those using selected data from GCEW. Supplement S3 provides a brief overview of that literature, organized broadly under headings of hydrological studies, water quality studies, economic studies, studies characterizing soil resource variability, studies of managing resource variability, studies of precision conservation or targeting conservation practices, and sociological studies. These published reports either use data contained within the database described throughout this series or support the understanding of the transport processes that produced the data contained therein. Potential users of the database would necessarily require knowledge of precursor research, both to gain awareness of findings and to avoid redundant effort.

Introduction to the Special Section

The papers in this special section comprise three types. This paper serves the entire collection as narrative metadata, meaning that the information is required to interpret all of the data in the database, but that it is not sufficiently structured to be readily stored in the database itself, except as images of these papers. The next four papers (Sadler et al., 2015; Baffaut et al., 2015b; Lerch et al., 2015b, 2015c) document specific data themes of weather, flow, herbicides, and nutrients. The final four papers (Lerch et al., 2015a; Kitchen et al., 2015; Sadduth et al., 2015; Baffaut et al.,
2015a) document research that used data described in the first five papers. As described above, this paper documents the establishment of the watershed, lists sources of geospatial data not logically included in one of the four data papers, and identifies research gaps. Supplement S2 to this paper describes the underlying geomorphology of the region, without which neither the subsurface nor surface processes could be interpreted. Supplement S2 also describes the anthropogenic changes made to the landscape since western settlement, which have dramatically altered the soil resource base through erosion and altered the hydrology through the residual effects of streambank erosion and channelization of streams and rivers. In addition, it describes the historical evolution of production practices and conservation efforts that have and continue to impact processes controlling ecosystem services in the region. Supplement S3 briefly lists research already published using the data or supporting its interpretation in the GCEW/CMRB context.

Each of the data papers describes the data, the methods used, the quality assurance procedures and confidence levels, prior research relevant to the data, the possible future uses to which it might be put, and how to access the data in the online STEWARDS database. Sadler et al. (2015) describe the GCEW weather record, including locations and descriptions of GCEW instruments, quality assurance/quality control procedures specific to the weather data, and external resources for comparable data. These weather data have been and will be critical for hydrologic analyses, crop growth models, hydrologic models, and climate change analyses. Baffaut et al. (2015b) describe the flow data available from stream channels and weirs, field weirs, and plot flumes. These data have proven critical for hydrologic, water quality, and climate change analyses, as well as for comparison with model outputs in calibration and validation of watershed hydrologic models. Surface water quality data for pesticides are described by Lerch et al. (2015b), including some specific sampling, handling, and analytical procedures. Those data have been extensively used to document water quality concerns, inform best management practice choices, and inform national policy, particularly for atrazine. Substantial success in generalizing explanatory models to account for known factors has been achieved there, including scaling atrazine transport from 0.0034 to >1000 km². Substantial effort has been directed, with some success, toward explanatory models for decision support in management practice placement as well. Similarly, Lerch et al. (2015c) describe data for nutrients in surface and groundwater.

The final four papers contribute original research in this collection but also serve to illustrate some of the research questions that can be answered using the data in this database. Kitchen et al. (2015) report groundwater nutrient levels and their hydrogeologic controls. They show that NO₃–N levels in groundwater vary considerably from field to field and from well to well, and appear to be more a function of past management, capacity of the till to buffer changes, hydrogeologic variability found among wells, and localized activity of biological processes rather than of specific cropping practices during the GCEW research period. Lerch et al. (2015a) discuss the loads obtained when the stream concentration and flow data are considered jointly, and illustrate the magnitude of the nutrient transport issues for the region. In general, the undrained lands in the central claypan region, because of their higher surface runoff and less interflow than the drained lands in the northern and eastern Corn Belt, are generally considered less of a contributor to continental-scale nutrient transport issues such as Gulf hypoxia, but nutrient loads to receiving water bodies have posed challenges for water treatment plants because of algae blooms. Sudduth et al. (2015) compare three methods of determining the water quality in one of those receiving bodies, Mark Twain Lake, which is source water for the majority of the NE Missouri region. Both proximal and aircraft hyperspectral observations were compared with direct sampling at the times of observation. Remote sensing methods provided good results when calibration relationships were established using water samples obtained on the same date. Finally, Baffaut et al. (2015a) describe the calibration and validation of the SWAT model in the GCEW and illustrate the utility of the tool for hypothesis testing. Supplemental material to the latter paper also includes the survey material and results (Murphy et al., 2010) that provide the information necessary to determine land management in the watershed. (The full list of articles for the special section is provided in Supplemental Material S1, Section E.)

Knowledge Gaps and Research Needed

One of the fundamental needs is knowledge at the whole system level. Therefore, this section is organized from a whole-system perspective, borrowing from engineering practice. In control theory, the current state of a system is described by a group of state variables (see, e.g., Brogan, 1985). In the discipline of system dynamics (Forrester, 1961, 1971), the state variables are generally described as stocks or levels of the component of interest, with flows or rates of transport among components being of critical interest in modeling systems. Future states of the system depend on the current state and those changes, either directly or through feedback loops. Invoking conservation of mass or energy, balance equations for the state variable are then the sum of the rates of change. These principles are the foundation of dynamic simulation models, a side benefit of which is that they often illustrate what is yet unknown.

Depending on the system, state variables can be few or many, easy or very hard to measure, and can vary over many orders of magnitude in their impact on the system. For a landscape or watershed that is managed for production of food, feed, fuel, fiber, and forest products overlaid on recreational, municipal, and industrial activities, the system is acted on by numerous driving forces, and the list of state variables can be unmanageable. Therefore, research usually focuses on those that have relatively larger impacts on the state of the system. In a managed agroecosystem such as the GCEW/CMRB, experience has shown that water, nitrogen (N), phosphorus (P), potassium (K), carbon (C), sediment, and energy (and temperature for physical systems) are fundamentally important for the description of the system’s state and its capacity for production. In addition, contaminants such as herbicides or pesticides, aerosols, particulates, or pathogens may affect the system’s capacity to deliver other ecosystems services, particularly in supporting the health and quality of life of the organisms within the systems. Therefore, knowledge is
required of the quantity of the state variables, the forms each can take, the transformations among those forms, the coupling of many of the state variables, and the fluxes in space and time both within the system and across the system boundary. In general, quantitative information on these balances is lacking to some degree for all the state variables listed and is essentially unknown for other state variables that may prove important as more knowledge is developed.

Historically, most watershed research was established to examine the water balance, with emphasis on stream flow and rainfall. The magnitude of the evaporation flux and the strong link between water and energy suggested that measurements of evaporation, air temperature, windspeed, humidity, and/or solar energy were required. Consequently, the knowledge base is reasonably broad for these well-known weather variables. However, other fluxes of material or energy have not been studied for as long or in such detail. Recent interest in carbon balance and climate change has raised awareness of importing C into the watershed and transformations of carbon-based energy stocks at the watershed scale, but quantitative information remains scarce. The scarcity is well illustrated, for example, by attempts to analyze system-level energy dynamics by systems ecology or life-cycle analysis.

Nationally important issues such as eutrophication of streams and fresh-water bodies and hypoxia in saltwater systems have placed an emphasis on estimating the N and P balances at scales from the field to the watershed, in particular the export of N and P through streamflow. For N especially, transformations among various forms complicate the assessment. At both the local and continental scale, the N balance is not well closed. In many contexts, quantifying the denitrification flux appears to be a critical research target. The sediment-bound component for P means that erosion and associated sediment transport and the dissolution equilibria and dynamics of the dissolved reactive P form are also important. While the sediment itself is conserved, the fluxes from upland sources through multiple downstream deposits of varying residence times, interacting with the transformations among forms of P and sometimes N in particular, render the knowledge regarding associated fluxes inadequate. The prominence of N, P, and K fertilizer in agricultural and other (lawn and garden) production means that estimates of the watershed-scale import fluxes are critical. Those estimates are usually made by large-area average expectations of producer behavior and not by direct knowledge or measurement.

As scale is increased from point to field, and then to multiple fields interspersed with nonfield areas, values for the local state variables, already temporally variable, become increasingly spatially variable, and many more interactions among landscape components can occur and become increasingly important. Knowledge of the local values of some state variables is available from reductionist research at point or field scale, and from the possibility of local knowledge of producer practice, but the increasingly variable producer behavior as scale increases becomes difficult to observe or predict. Many more operational, tactical, and strategic decisions are made by the increasing number of producers operating on increasing numbers of management units under increasing sets of economic, resource, and other conditions. Difficulty in accounting for producer behavior at each point forces large-scale analyses to increasingly rely on broad-average estimates of behavior. At some scales, this reliance is justified by the improved success of models at larger scale. In other cases, factors unimportant (and safely assumed negligible) at a smaller scale become large enough that they must be taken into account at larger scales.

For the state variables listed above, the first knowledge gap is the inventory of forms and fluxes that need to be examined, including both in-system fluxes and transport in and out of the system by natural or human forces. The second step would be to qualitatively assess which fluxes are large enough that they must be quantified, and the third step is to quantify them. None of these steps have yet been systematically pursued for most of the state variables listed, although the focus on N in the discussion above about denitrification indicates that partial knowledge exists for some. However, a research need exists to assess all terms of the N balance. General knowledge of the water balance is likely the best, assuming the anthropogenic extraction of groundwater, use of transported municipal water, release of wastewater, and transport of gray water for in-field spraying are small relative to rainfall, evaporation, and streamflow. Even then, effects of the many management decisions on those water balance terms at the producer level are poorly quantified. Nutrient balances in general have focused on the field scale but must be considered at the watershed scale. In particular, anthropogenic inputs, transformations, and outputs of N and P must be assessed for their impact on the larger-scale landscape and watershed. Knowledge of the mass balance of sediment is nearly limited to rates of in-field erosion and watershed outlet transport. Some quantification of channel processes and their role in the mass balance has occurred, but more is needed. Difficulties in achieving life-cycle analyses (LCA) in this area illustrates that little research has been done on the energy balance of the combined agriculture–urban–undeveloped landscape system, and it would appear that anthropogenic energy fluxes for nonagricultural activities should not be assumed negligible. Clearly, substantial research needs exist for comprehensive studies on energy balance terms and the effect of producer management decisions on the energy fluxes.

The next step, which should be prioritized to those major state variables identified, would be to determine the effect of common production practices on the states and fluxes of those state variables. For practices offering scope for reduced impact, development of alternatives would be an appropriate research topic. For important fluxes recalcitrant to source reduction approaches, research could focus on mitigative approaches.

In many cases, prior research such as described above informs both the existing knowledge and the identification of knowledge gaps. Many such examples exist, and the discussion of knowledge gaps should be considered in concert with the existing research to guide future work. For example, in the geophysical context of the GCEW and CMRB, the anthropogenic impacts of stream channelization described in Supplement S2 pose complications in understanding and predicting the effects of riparian ecosystems on bank stability. Channelization was shown to have altered the stable-channel characteristics and, in particular, caused dynamic effects that propagated headcuts up channels. It is unknown how far headcuts have moved, but Willett et al. (2012) showed that historical channelization (age >>30 yr) continues to cause
high streambank erosion on tributaries upstream from the channelization. Clearly, scope exists for research to explain such results and evaluate the hypothesis of the channelization having a legacy impact long after and at some distance from the actual channelization.

One of the values of long-term data is that the period of record is sufficient to test hypotheses regarding temporal trends that are hidden by short-term variability. This is increasingly important in the presence of long-term trends in temperature and the anticipated effects of those trends on rainfall suggesting that rainfall intensity may increase and rainfall frequency may decrease. This outcome of climate change, coupled with the effect of higher temperatures on evaporation, poses a challenge to agricultural production and, in particular, places a premium on robust, resilient cropping systems. Ongoing research in GCEW/CMRB is addressing these issues, but additional scope exists for analyses of the datasets described in this series. The expected effect of climate change is that rainfall intensity and frequency patterns will shift, and these data could be critical for research on those characteristics. Many research efforts will use basin- or watershed-scale simulation models for what-if scenario testing, and the datasets described herein on weather, land use, geophysical characteristics of the soil, and producer behavior are well suited to drive the models. The streamflow and water quality data described here would provide tests of model performance during calibration and validation. Associated crop yield data would help in tests as well.

The emergence of data and data-driven research as described earlier in Gray’s (2009) Fourth Paradigm places substantial new requirements on data documentation. Recall from the rationale above that data need to be self-described, meaning that all metadata—documentation of methods, units, context of collection, responsible agents, provenance, and so on—must be permanently associated with the data. Michener and Jones (2012) cite the need for integrative platforms, adoption of standard protocols, good data stewardship practices, promotion of informatics literacy, data sharing, and scientific transparency and reproducibility. Again referring to the rationale, Gray (2009) cited a need for laboratory information management systems, data collection and visualization tools, and access and communication tools. Collectively, the trends in data management and environmental research illustrate that research skill sets and infrastructure are not keeping pace with the growth in volume and types of data. Considerable effort must be devoted to standardized, general conceptual organization, or ontologies, of environmental databases. Standardized formal structures, or schema, would then be necessary. It would require a great deal of training in the relevant skills to accomplish this, but the benefits in data archiving and communication would be substantial. The research need in this effort is the development of suitable standardized ontologies and schema, plus the transport formats. Efforts toward this end exist in many disciplines, at scales up to and including global standards, usually the International Organization for Standardization (ISO). This process is neither rapid nor easy, but laboratories holding datasets such as described in this series would be advised to monitor and if possible contribute to the efforts.

The database described herein was developed from a watershed infrastructure, so the data are focused on hydrologic observations. The LTAR network encompasses all of agroecosystem research, and that broadening of scope defines a substantial research need for other aspects of agroecosystem research, primarily in observing biological elements of ecology that have not been a focus of agricultural research in the GCEW/CMRB context. Within the context, there has been some cooperative research on agroforestry and in riparian forest effects on streams and water quality, but little else involving the 10 to 20% of the watersheds that are forested. As the focus of the research unit was cropping systems and water quality, little focus has been placed on livestock systems, except for the observable effects on the water quality endpoint. There appears to have been little research emphasis on wildlife within the GCEW, although university cooperators have research infrastructure in place to provide such information. There is also a cooperative research focus on native prairie ecological systems, which has been used both as a baseline index for comparison with managed systems and as a guide to the introduction of bioenergy feedstocks in the region. Clearly, a substantial need exists for adding these complementary research foci to the existing research infrastructure in the GCEW and CMRB contexts.

The above knowledge gaps have been discussed from the local perspective. Many of the same gaps exist to some degree in the broader LTAR network. These have been addressed by the scientific leadership of the LTAR network in the LTAR Shared Research Strategy (Bryant et al., 2013), and the reader is directed there for details. However, it is worth noting here that a great opportunity exists for leveraging the suite of long-term databases that exists across LTAR sites toward answering questions at regional and continental scales that cannot be answered using a long-term database from any single location.

**Summary and Conclusions**

This paper, the first in the series, documents GCEW data within the context of the broader US watershed infrastructure, documents the specific development of GCEW, and discusses the gaps in knowledge that should be addressed. Metadata in Supplemental online materials (Supplement S2) describe the genesis and geophysical setting, describe the anthropogenic changes that have been made to the landscape, and (Supplement S3) list the research that has been conducted within the study area or with the data described in the series. The series of four data papers document the GCEW database, illustrate the scope of the data, the methods used to collect and transform it, the level of confidence users can have in it, and the range of potential research uses to which it can be applied. The research papers document four uses of the data and provide original research results.

The first five papers in the series provide a comprehensive documentation to the database, both serving notice that the data exist and sufficiently supporting the data with specific methods and metadata, and sufficient physical, historical, and research context that adds substantial value to the data itself. Substantial opportunity for collaborative research at the local and continental scale is represented in the database, as suggested in the knowledge gap identification and research needs discussion. Research needs include (i) quantifying watershed
scale fluxes of N, P, K, sediment, and energy, (ii) accounting for fluxes involving forest, livestock, and anthropogenic sources, (iii) scaling from near-term point-scale results to increasingly long and broad scales, and (iv) considering whole-system interactions through process modeling, systems ecology, or life-cycle analyses. Substantial progress is needed in defining database ontologies suitable for holding the complexity of data obtained in research sites such as GCEW/CMRB, and success in that endeavor would contribute to leveraging the GCEW/CMRB node into the larger LTAR and other networks. Together, the papers in this special section support the use of the GCEW database by scientists and others for research to solve these and other natural resource problems important to US agricultural, environmental, and science policy.

References


Glaser, K. 2010. It’s time for a long-term agroecosystem research (LTAR) network. CSA News, CSA, Madison, WI.


