esa

ECOSPHERE

INNOVATIVE VIEWPOINTS

Agroecosystem research with big data and a modified scientific method using machine learning concepts

M. Susan Moran,[1,†] Philip Heilman,[1] Debra P. C. Peters,[2] and Chandra Holifield Collins[1]

[1]*USDA ARS Southwest Watershed Research Center, Tucson, Arizona 85719 USA*
[2]*USDA ARS Jornada Experimental Range and the Jornada Basin Long Term Ecological Research Program, New Mexico State University, Las Cruces, New Mexico 88003 USA*

Citation: Moran, M. S., P. Heilman, D. P. C. Peters, and C. Holifield Collins. 2016. Agroecosystem research with big data and a modified scientific method using machine learning concepts. Ecosphere 7(10):e01493. 10.1002/ecs2.1493

Abstract. Long-term studies of agroecosystems distributed across the North American continent are providing an extraordinary understanding of regional environmental dynamics. The new Long-Term Agro-ecosystem Research (LTAR) network (organized in 2012) has designed an explicit cross-site research program with multiple U.S. Department of Agriculture (USDA) experimental watersheds, ranges, and forests. Here, we report results from studies using a modified scientific method that includes learning through time that was implemented over the past five years with long-term data from USDA experimental sites in coordination with other networks. The results offer a compelling argument for the LTAR concept of combining site-based expertise with network-wide coordination and collaboration to arrive at more accurate scientific conclusions than possible from individual researchers working alone. Simply put, without site-based expertise and cross-site communication working in parallel to provide input, feedback, and refinement to each subsequent step, similar to the way machine learning works, the interpretations and conclusions of these studies would have been incomplete, if not incorrect. Further, the up-front time commitment to data processing and analytics above the time dedicated to place-based studies increased the productivity of the team and the impact of the research, unlike the common perception that cross-site research is often less efficient. In turn, this approach supported a non-traditional system of credit for co-authors based on citation impact of the journal selected as the publication outlet with less regard for author order. The LTAR network has embraced this modified scientific method in its shared research strategy and common experiment to address the problematic issues of mixed data quality across studies and sites, co-author credit, research efficiency, and scientific impact on data-intensive research. This approach can be combined with other types of collaborative and social media approaches, such as crowdsourcing, to take advantage of the wide range of expertise in the agroecological community as well as other disciplines to move science forward in the time of big data.

Key words: big data; experimental watersheds, ranges, and forests; Long-term Agro-ecosystem Research (LTAR); long-term data; network science; regional-scale analyses.

Received 4 January 2016; revised 28 June 2016; accepted 1 July 2016. Corresponding Editor: Scott L. Collins.

† **E-mail:** sumoran54@gmail.com

Introduction

Data-intensive research is garnering attention in the fields of agriculture, ecology, and environmental science. There is an expectation that data-intensive research will expand scientific discovery and provide unprecedented understanding of environmental problems at large spatial extents across regions to continents (Luo et al. 2011, Michener and Jones 2012, Porter et al. 2012, Hampton et al. 2013, Peters et al. 2014*b*). For such applications, data are often compiled from multiple sites with long-term measurements of diverse ecological and hydrologic variables (referred to here as multilocation time series or MLTS, for convenience). MLTS studies can be used to generalize these locally derived observations from many sites to obtain regional forecasts with inputs to management, thus providing a distinct advantage over individual, place-based data sets with a small realm of inference (Heffernan et al. 2014, Peters et al. 2014*b*).

The challenge in undertaking MLTS studies is defining an overall structure and a specific approach for processing large, complex data sets. There is broad agreement that the traditional scientific method needs to be modified as the volume, velocity, and variety of data (i.e., big data) continue to increase through time as new technologies become available and that these changes need to go beyond open sharing of data (described in Peters et al. 2014*a*). In the traditional scientific method, the complexity of the analyses increases as the data streams increase, and there are time lags between individual discoveries and leaps in knowledge. These lags are associated with the time required for publication of results and time required for others to recreate the analyses and findings even when the data and metadata are readily accessible. Thus, serious consequences can result when a small data approach is used to address complex scientific problems (Hampton et al. 2013).

However, in a modified scientific method, the researchers and the software adopt a machine learning approach where the system maintains a memory of previous analyses and attempts, and this knowledge is used to recommend paths forward at each step along the way. This communication and feedback at each step is critical for MLTS studies because there is often incomplete knowledge by any one researcher for all sites or ecosystems in the analysis. In fact, it is common in MLTS research to begin without a hypothesis or preconceived notion or set of relations among variables but instead with a basic idea resulting in many possible outcomes. It is rare for one individual to have both a specific understanding of temporal mechanics at multiple sites and a broad grasp of the speculative theory to posit a generalization across all sites. Although our long-term goal was to complete an automated machine learning system, we present here an approach based on this modified scientific method where learning was gained and applied by researchers manually.

The U.S. Department of Agriculture (USDA) Long-term Agro-ecosystem Research (LTAR) network was organized in 2012 specifically to address that challenge (Robertson et al. 2008, Walbridge and Shafer 2011). The LTAR concept is exceptional in that multiple site personnel *function as a network* with an explicit long-term research program and non-traditional research collaborations based on a system approach. Eighteen experimental sites were chosen for the LTAR network based on expressed criteria, including a track record of productivity; a data record with length, breadth, depth, and overall quality; an existing long-term research facility with support for continued operation for the next 30–50 yr; and a history of partnerships to enhance research, education, and outreach (Walbridge and Shafer 2011). All sites have long-term data available in machine-readable format, and most sites have data available for download from their Web sites (Marks 2001, Bosch et al. 2007, Moran et al. 2008*a, b*, Bryant et al. 2011, Steiner et al. 2014, Sadler et al. 2015) or from Web sites supported by other networks such as Ecotrends (Peters et al. 2013), AmeriFlux (Boden et al. 2013), LTER (Hobbie et al. 2003), and NEON (Kampe et al. 2010, Kao et al. 2012). The LTAR-shared research strategy (SRS) has created a research network with common measurements and a system approach to research (Bryant et al. 2013). This SRS gives a cross-network framework to focus local agroecosystem research conducted in a traditional bottom-up research approach, and a network administrative organizational structure that provides a top-down strategy and constraints operating across all sites. The SRS offers four priority areas of concern, including (1) agroecosystem productivity, (2) climate variability and change, (3) conservation and environmental quality, and (4) socioeconomic viability and opportunities.

The LTAR network has thrust agricultural research into uncharted territory by conducting MLTS studies within the LTAR mandate. Peters et al. (2014*a*) outlined the challenges inherent in using big data for ecology and environmental sciences and suggested a knowledge, learning, analysis system (KLAS) that combines the hypothesis-driven scientific approach with a machine learning method. The traditional scientific approach was modified to accommodate MLTS studies while retaining the credit to individual scientists for original ideas in published papers. The machine learning component of KLAS offers an efficient means to identify data with the largest meaning and to filter out low quality or misleading data. KLAS provides a semiautomated system guided by frequent human input to increase efficiency yet preserve individual scientists' creativity. KLAS also learns through time and builds upon the experience, data inputs, and model outputs of previous users, similar to web-based search engines. This approach contrasts with the current scientific method where new users often recreate published analyses, typically from online source data, before moving forward to new analyses with newly collected data.

This report provides an example of the shift from small, highly controlled data sets to a large, federated data set for agroecosystem research. The example represents a manual implementation of the modified scientific method where new ideas derive from individual scientists while the hypothesis, analytics, data, and conclusions are developed in collaboration with the broader scientific community. This example helps to answer two questions: How will agroecosystem research be implemented in the LTAR network, and conversely, what will the LTAR concept contribute to data-intensive research?

MLTS Example with USDA Experimental Sites

Similar to the KLAS iterative concept (outlined by Peters et al. 2014*a*: their Figure 4), a modified scientific method was implemented with several USDA and LTER experimental sites to study the impact of altered hydroclimatic conditions on vegetation function (where italicized terms refer directly to Fig. 1). The *initial idea* was generated by an individual scientist (termed the *individual*) to investigate the response of aboveground net primary production (ANPP) in grasslands to local precipitation (P) extremes during a period of unusual climatic conditions (the early 21st century) at the annual time step, the decadal temporal scale, and the continental spatial scale. The *precedents* were the predictions and reports of altered hydrologic conditions leading to prolonged warm drought during the early 21st century (IPCC 2013), and the ANPP-to-P relation reported by MLTS studies during the relatively cool, wet conditions of the late 20th century (Huxman et al. 2004).

Preliminary *data* were downloaded as available from Web sites, particularly from Ecotrends (Peters et al. 2013), to lighten the burden on specific sites. NASA Moderate-resolution Imaging Spectroradiometer (MODIS) Enhanced Vegetation Index (EVI) observations were used as a proxy for ANPP over a ground area of approximately 2 × 2 km (Huete et al. 2015, Rocchini et al. 2015). Initially, eight sites were chosen with long-term precipitation and temperature records (1980–2009) and uniformly distributed grassland vegetation extending over an area of 4 × 4 km (Table 1). A report of *preliminary findings* was produced concluding that grassland ANPP was significantly correlated with annual precipitation, and during this prolonged drought, there was a strong ANPP response to both wet and dry years (Fig. 2a).

During the initial telephone conferences (telecons) with *network scientists* (Fig. 1), the *individual* learned a great deal about the data and made many important changes to the study. For example, the San Joaquin Experimental Range (SJQ) site had substantial woodland cover and the Santa Rita Experimental Range (SRE) site was dominated by shrubs rather than grasses; thus for both sites, the MODIS signal was not monitoring the grassland response. Also, the web-downloaded precipitation records for the Desert Experimental Range (DER) and Jornada Experimental Range (JRN) sites were not well associated with the MODIS windows. At these two sites, precipitation was so site specific that it was necessary to use a raingage within a kilometer of the MODIS data, not several kilometers. Thus, the SJQ and SRE sites were dropped from analysis, and precipitation data for the DER and JRN site were replaced with measurements from an alternate raingage provided by the site contacts (Fig. 2b). Next, two sites from the Long-Term Ecologic

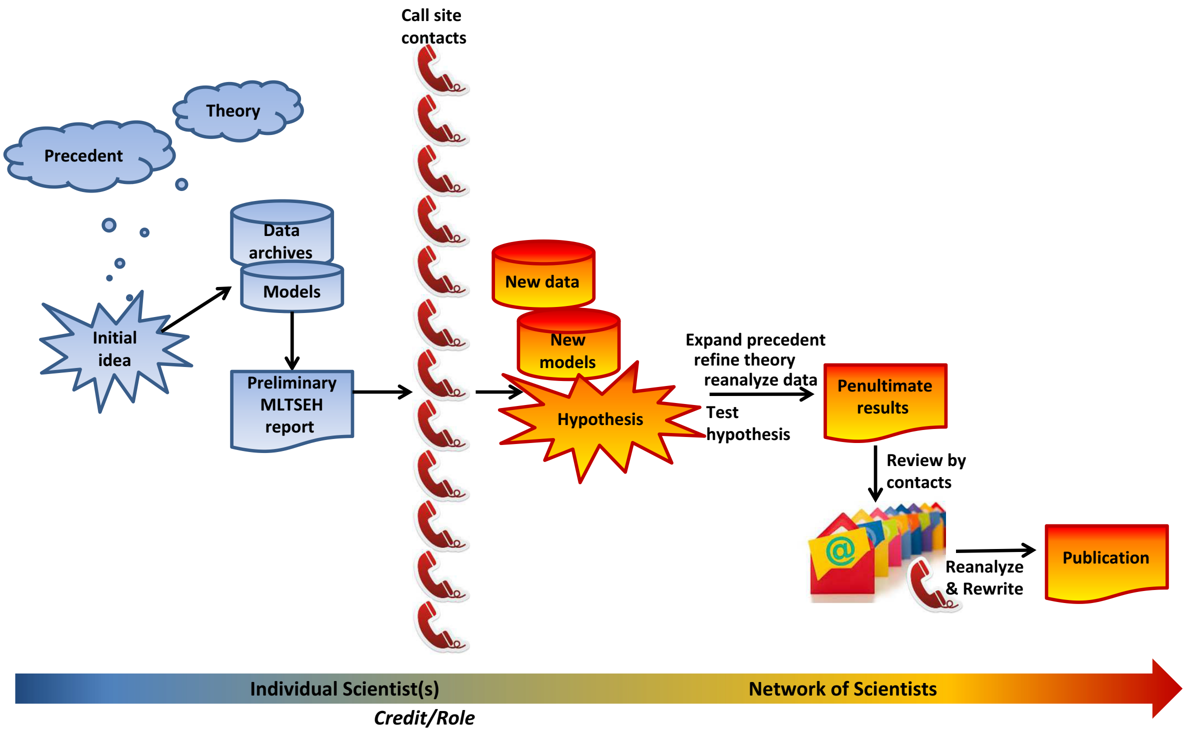

Fig. 1. The modified scientific method for multilocation time-series studies, where the left side is dominated by efforts of an individual, transitioning to full participation by data providers moving from left to right and indicated by a corresponding change in color from blue to red. The implementation of this concept was initiated in 2009 with U.S. Department of Agriculture experimental sites, resulting in the first manuscript (i.e., one of nine) published in 2013.

Research (LTER) network were added (Table 1) followed by telecons with LTER network scientists (Fig. 2c). As a result of discussions with LTER site contacts, the Long-Term Ecological Research program at Konza Prairie (KNZ) and Long-Term Ecological Research program at Sevietta Field Station (SEV) sites were dropped from the analysis because (1) grasslands at KNZ were burned every year, disrupting the ANPP-to-P relation, and (2) the grasslands at SEV covered a finer scale than the coarse MODIS resolution. Some of this place-based information was available in publications, but much of it was not. The result of telecon discussions was the selection of a set of six sites for the analysis that met the study's scientific and technical criteria (Fig. 2d).

As the analysis progressed, a new variable (in situ measurements of ANPP) was deemed necessary to determine the relationship between ANPP and the proxy MODIS EVI. As before, we downloaded ANPP data from web archives for the initial analysis and followed with telecon discussions with site contacts. Again, our results changed due to data updates and new interpretations obtained during telecon discussions. For example, the Central Plains Experimental Range (CPL) site contact recognized that the ANPP data were not well associated with the location of the MODIS window (Fig. 3a), and he offered another unpublished ANPP data set (Fig. 3b). Further, in discussions, we recognized that the in situ ANPP was measured near the peak of the growing season, whereas the MODIS EVI was integrated over the entire growing season. By integrating the MODIS EVI to the date of the in situ ANPP measurement, the relation between in situ ANPP and integrated EVI (iEVI) was improved (Fig. 3c), thus providing better support for the use of iEVI as a surrogate for ANPP (Moran et al. 2014).

The posttelecon *reanalysis* was undertaken by the full team of network scientists. For example, in one case, the MLTS data were transferred in

Table 1. Ten grassland sites considered for the study, including mean annual sum of precipitation (MAP) with standard deviations in parentheses.

Site	MAP (mm)
DER: Desert Experimental Range, Utah	179 (58)
JRN: Jornada Experimental Range, New Mexico	241 (73)
WGE: Walnut Gulch Experimental Watershed, Arizona	305 (91)
CPL: Central Plains Experimental Range, Colorado	381 (91)
SPL: Southern Plains Experimental Range, Oklahoma	587 (165)
LWA: Little Washita River Experimental Watershed, Oklahoma	794 (197)
SJQ: San Joaquin Experimental Range, California	352 (185)
SRE: Santa Rita Experimental Range, Arizona	434 (171)
KNZ: Long-Term Ecological Research program at Konza Prairie, Kansas	807 (221)
SEV: Long-Term Ecological Research program at Sevietta Field Station, New Mexico	182 (74)

Note: Averages represent a 30-yr period between 1980 and 2011 for most sites.

whole from the *individual* to another network scientist for a specific statistical test. Ultimately, a *second report*, in manuscript format, was drafted with co-authors listed in order as follows: First authors were the *individual* and those directly responsible for statistical analysis, and this was followed by the network scientists in alphabetical order.

Through the manuscript review, co-authors were able to correct further errors and offer mechanistic interpretations based on their knowledge of the site. For example, one co-author directed us to a recently submitted, unpublished manuscript explaining the lag in ANPP response to precipitation after prolonged drought. This resulted in a final interpretation based on life history in desert and plains grasslands with outliers explained by mortality and response lags (Fig. 4). This contrasted sharply with the preliminary conclusions drawn without the expertise and interpretations provided by site contacts (Fig. 2a).

By *transitioning* from the *individual* to the network of scientists, it was inevitable that other research would emerge from the effort put into the initial data-intensive process. In this example, two other studies were initiated to determine how ANPP across biomes responded to altered hydroclimatic conditions forced by the contemporary drought in Southern Hemisphere and Northern Hemisphere. For these two studies, sites were expanded to include 12 USDA long-term experimental sites in the conterminous United States and Puerto Rico and 17 similar sites in the Australian continent over a range of precipitation regimes and vegetation types from grassland to forest (Ponce Campos et al. 2013, Zhang et al. 2013).

The *final conclusions* of this effort were published in three manuscripts, including the original grassland study used as an example above. Ponce Campos et al. (2013) reported that all biomes, from grassland to forest, showed a shared capacity to tolerate low annual precipitation, and with continuing warm drought, grasslands would likely sustain significant mortality. Zhang et al. (2013) showed that extreme precipitation patterns reduced the sensitivity of ANPP to total annual rainfall by 20% and extreme rainfall patterns were as important as total annual precipitation in understanding vegetation processes. Moran et al. (2014) concluded that North American grasslands will undergo predictable, but regionally distinct, responses to the prolonged warm, dry conditions which are characteristic of climate change. Other network scientists have since taken the lead on new MLTS studies; for example, a study is being considered using EVI to predict cattle performance and the red meat yield from grazing lands.

This section should make it clear that conclusions reached through the modified scientific method with MLTS data were far more complete and compelling than using the traditional scientific method with only the *individual* and a place-based data set (Table 2). First, the method allowed developments of regionally solid, scientifically sound hypotheses across time and space. Second, misinterpretations due to changing measurement protocol, unpublished ecologic and hydrologic events, and historical site management were avoided (Figs. 2–4). By following the modified scientific method, the data, methods, analysis, interpretations, and conclusions evolved from the original idea to several reports (Ponce Campos et al. 2013, Zhang et al. 2013, Moran et al. 2014). Fundamentally, without personal connections with each network scientist, the MLTS interpretations and conclusions would have been incomplete, if not incorrect.

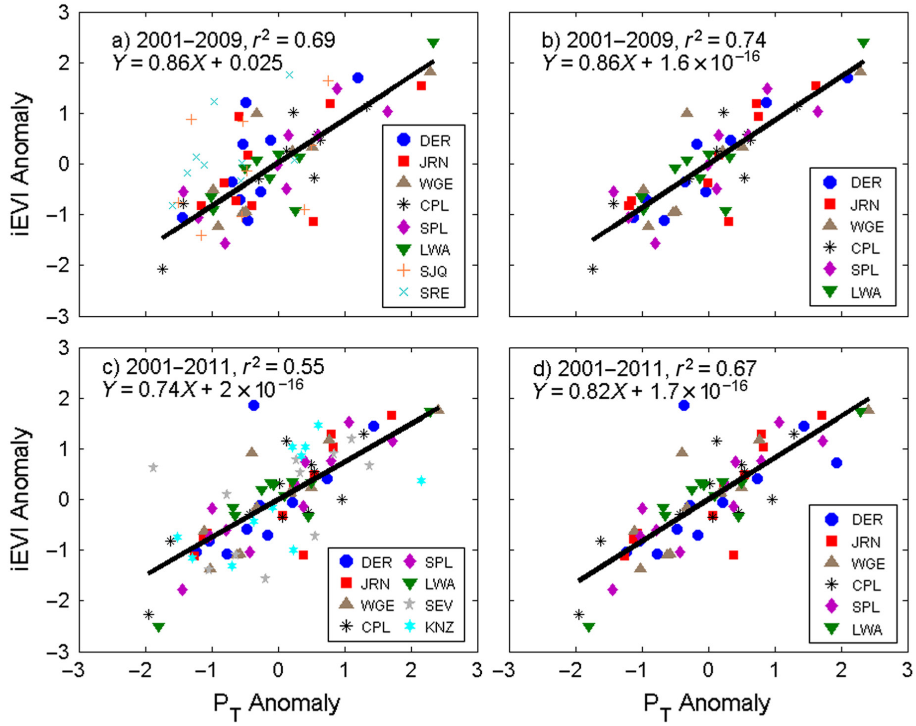

Fig. 2. A chronological history of study sites and data used to investigate the relation between annual precipitation (P_T) and aboveground net primary production (ANPP) using the Moderate-resolution Imaging Spectroradiometer integrated Enhanced Vegetation Index (iEVI) as a proxy for ANPP, where the annual anomaly is the number of standard deviations from the study-period mean at each site (Moran et al. 2014). Over this history, the study period was 2001–2010 for (a) and (b) and 2001–2011 for (c) and d). (a) Results with initial eight sites *before telecons* with eight site contacts. (b) Results *after telecons* with Santa Rita Experimental Range (SRE) and San Joaquin Experimental Range (SJQ) sites dropped and updated precipitation for Desert Experimental Range (DER) and Jornada Experimental Range (JRN). (c) Results with the addition of Long-Term Ecological Research program at Konza Prairie (KNZ) and Long-Term Ecological Research program at Sevietta Field Station (SEV) sites, *before telecons* with KNZ and SEV contacts. (d) Results a*fter telecons* with KNZ and SEV site contacts, including only sites that met the study's scientific and technical criteria.

Discussion

The LTAR network is prepared to implement a modified scientific method for MLTS research (Fig. 1) combining bottom-up site-based expertise and top-down network-wide coordination. By design, the LTAR network addresses the common concerns related to MLTS research including the mixed quality of data archives, allocation of credit for research, publication delay, and research impact.

Data quality

There is a general concern in the scientific community about the mixed quality of streaming data for MLTS research. The KLAS machine learning component offers "an objective way to flag, correct, or delete poor quality (e.g., missing or out of range values, inadequate metadata) or unimportant data (e.g., extraneous or repetitive information) identified during the analysis" (Peters et al. 2014*a*). This is a first step that should receive wide acceptance. The LTAR network promotes personal connections to allow access to and analysis of source data based on the premise that even data of the highest technical quality are of poor quality when used incorrectly. In this example, data quality was explored through direct communication with site contacts. Thus, the serious misinterpretations and wrong

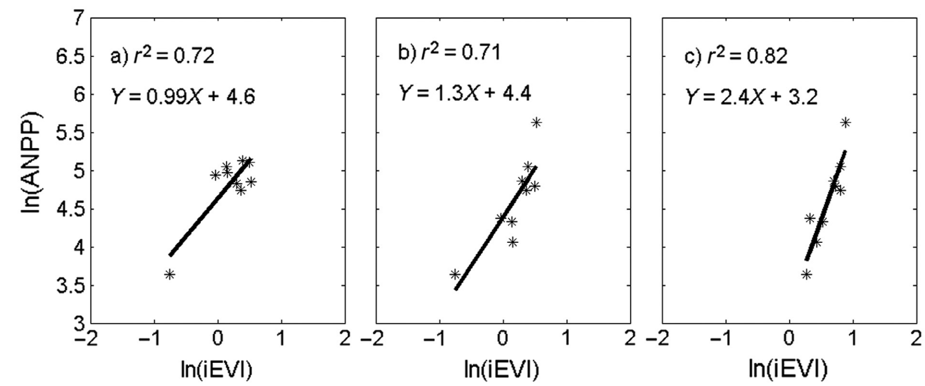

Fig. 3. A chronological history of data and methods used to determine the relationship between in situ measurements of aboveground net primary production (ANPP) and Moderate-resolution Imaging Spectroradiometer integrated Enhanced Vegetation Index (iEVI) at the Central Plains Experimental Range (CPL) site. (a) Results with ANPP downloaded from the web *before telecon* with CPL contact. *After telecon:* (b) Results with new ANPP data and (c) results with new methods, integrating EVI to the date of the in situ ANPP measurement (rather than over the whole growing season). Final results (c) (from Moran et al. 2014) can be compared with preliminary conclusions drawn without the expertise and interpretations provided by the site contact (a).

conclusions listed in the example above were avoided.

Role/credit

The traditional scientific method upholds a system of credit given to the *individual* through publication in journals. In this MLTS example, the authorship order started with whomever had the original idea, followed by those who worked most closely with that *individual*, and finally, all the network scientists in alphabetical order. The first author retained the traditional credit for the idea and the leadership. With the modified scientific method, co-authors also receive credit for teamwork and/or leadership in developing the hypothesis, analysis, interpretation, and writing.

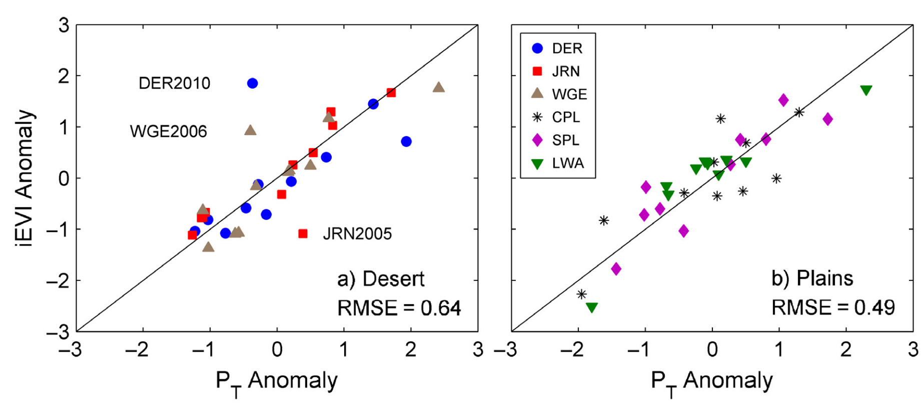

Fig. 4. Published results using the modified scientific method showing the relation between measured and modeled standardized integrated Enhanced Vegetation Index ($iEVI_S$) where sites and years for which the difference between measured and modeled $iEVI_S$ exceeded ±1σ are labeled. RMSE is the root-mean-squared error of the difference between measured and modeled $iEVI_S$ and P_{TS}. Final results here (from Moran et al. 2014) can be compared with preliminary conclusions drawn without the expertise and interpretations provided by site contacts (Fig. 2a).

Table 2. Comparison of the modified scientific method (a network of scientists and multilocation time series [MLTS] data) with the traditional scientific method (only the *individual* and a place-based data set).

Metric	Traditional scientific method with place-based time-series data	Modified scientific method with MLTS data and a network of scientists
Data quality	Research is based on place-based measurements made by the investigator or mined data from archives of *unknown* quality	Data are mined from archives of *unknown* quality. With the full participation of data providers (Fig. 1), measurement protocols were clarified, new data were obtained where necessary, and new models were implemented. The original precedent was expanded, the theory was revisited, and the data were subsequently reprocessed to test the new hypothesis
Role/credit	*Individual* receives credit for original idea, analysis, interpretation, and writing. Data providers are acknowledged	*Individual* receives credit for original idea; network scientists receive co-authorship and credit for teamwork and/or leadership in developing the hypothesis, analysis, interpretation, and writing. In one case, co-author received institutional credit for publication impact without regard for author order
Efficiency	U.S. Department of Agriculture promotion expectation is about 1.5 publications/yr, for a total of six first-author publications in 4 yr	In this example, after 4 yr, there were four manuscripts submitted by network scientists, four spinoff publications, and one book chapter, for a total of nine reports, all with different first authors
Impact (Appendix S1)	Simulation of the average number of citations over a 15-yr record for place-based research resulted in a total of 122 citations/ publications (n = 26)	The same simulation for MLTSHE research resulted in a total of 313 citations/ publications (n = 8). Although simulation results are limited in scope and not conclusive, the difference is a substantial 2.5 times increase in citations

For example, one author who was no higher than 9th on the four papers reported in this example was able to use this work as one of eight accomplishments in a promotion package because the intellectual effort was considered balanced. This is exemplified in the acknowledgment by Ponce Campos et al. (2013) in the Journal *Nature*: "G.E.P.C., M.S.M. and A.H. conceived the study, assembled the data, and produced the preliminary results. The remaining authors collected and analyzed data, and contributed to the interpretation of results. All authors contributed to writing the paper. Statistical analyses were performed by G.E.P.C." Using the modified scientific method, it is reasonable to consider a non-traditional system of credit for co-authors based on publication impact without regard for author order.

Efficiency

There is no denying that MLTS research requires an up-front time commitment to data processing and analytics above the time dedicated to local, place-based measurements. In this example with USDA and LTER sites, the time from idea to publication for the first manuscript was 4 yr (2009–2013, Ponce Campos et al. 2013). Yet, this time investment led to two more publications (Zhang et al. 2013, Moran et al. 2014), one manuscript submission (M. A. Ross et al., *unpublished manuscript*), four spinoff publications using a similar MLTS approach with different data/ hypothesis/locations (Ma et al. 2013, Zhang et al. 2014, Hottenstein et al. 2015; Barnes et al. 2016), and one book chapter (Huete et al. 2015), which all were attributable to the mutual effort expended on the initial database. For studies following the first effort, the personal connection had already been made and data analyses were better understood, both leading to greater efficiency. Following the first effort, the method (Fig. 1) was still followed with fewer phone calls and more email communication, but the two reviews were retained. That is, one preliminary report was used to agree on precedent, theory, models, data, approach, hypothesis; this was followed by reanalysis and a second review of penultimate results leading to publication. With the

LTAR network in place, data quality and familiarity at the member sites will be building and resulting in better analytics.

Impact

Multilocation time series studies are not only pioneering but also compelling, because they are based on long-term measurements across many biomes in a natural setting. As a result, MLTS studies can be published in the highest impact journals for broad scientific impact. In our example, the three manuscripts published since 2013 already have a total of more than 65 citations. The publication by Zhang et al. (2013) was selected as an American Geophysical Union Research Spotlight in Eos Earth and Space Science News, and results from Ponce Campos et al. (2013) were the subject of several University and Government press releases. For one senior scientist, the number of citations for MLTS publications was 2.5 times greater than for place-based research in ecohydrology (Fig. 5, Appendix S1). The LTAR network is populated through a selection process based partly on a track record of productivity and a history of partnerships to enhance research, education, and outreach. Bringing such sites together with a common overall structure and a suggested specific approach has the potential to result in the highest scientific impact.

Of the four concerns brought up here, data quality is certainly the most troubling. Every effort must be made to avoid wrong results in data-intensive studies due to poor data quality that are propagated through citation and lead to further errors in follow-on studies. Though unintentional, the consequences of data misuse are dire. With attention to data-intensive studies by authors, reviewers, and data providers, LTAR and other such networks will arrive at more accurate and complete scientific conclusions.

Concluding Remarks

This analysis substantiates a workable modified scientific method to conduct agroecosystem research in existing and future networks of experimental sites. The method provides a flexible structure to transform an idea to a hypothesis and come to conclusion with the full participation and valuable expertise of the data providers.

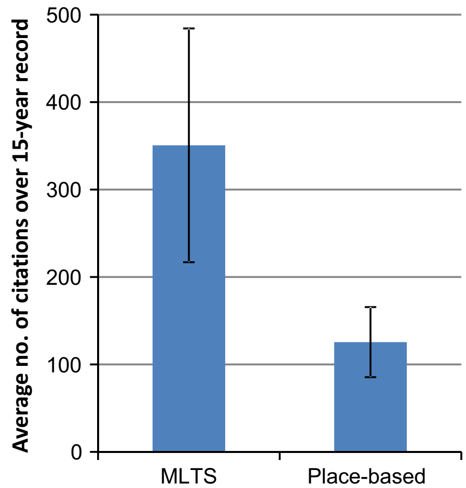

Fig. 5. A simulation of the average number of citations over a 15-yr record for a set of multilocation time series publications (n = 8) and place-based research publications (n = 26) for one co-author. See details in Appendix S1.

In this example, conclusions drawn *without* the suggested site-based expertise and network-wide participation (left site of Fig. 1, pretelecon) were found to be incomplete and, at times, incorrect (exemplified in Figs. 3, 4). Yet, there is a trend toward such data-intensive studies conducted without site contacts and with conclusions drawn based on established theory and simple regression analyses. The latter method does not elicit confidence in results and has little chance of providing new theory related to, for instance, the altered hydroclimatic conditions characteristic of the 21st century.

In this example, we have documented that the modified scientific method can be efficient and productive while providing reasonable credit and responsibility for all. For one scientist, productivity increased with an increasing number of contacts rather than the common perception that cross-site research might be less efficient (Appendix S1). Further, this method has already inspired a non-traditional system of credit for co-authors based on publication impact without regard for author order.

One of the issues in a modified scientific method is determining where the initial recommendations come from to "seed" the system. In our manual example, the initial recommendations came from the researchers' expertise and knowledge with the individual systems, but in a fully automated system, alternative approaches will be needed. Crowdsourcing is one way to obtain collective knowledge from the community of researchers that can be used to generate the initial set of recommendations (Silberzahn and Uhlmann 2015). Asking a group of researchers to analyze the same set of data to address the same question will provide a ranking of the alternative approaches that can be used based on how commonly they were used, and the sequence of steps each researcher used to solve the problem. This information can then be used to start the recommendations in the machine learning program. Utilizing the combined and varied strengths of our collective community as well as ideas from other scientific disciplines will be instrumental in moving ecological research forward in this time of big data (Peters et al. 2014*a*).

Finally, the research results using this method have been recognized for solid scientific contributions as measured by publication in high-impact journals, high citation records, and recent awards. The LTAR network has embraced this modified scientific method in its SRS and Common Experiment to address the problematic issues of data quality, credit, efficiency, and impact in data-intensive research. The initial success expressed here with USDA experimental sites bodes well for the LTAR and other such networks going forward.

Literature Cited

Barnes, M. L., M. S. Moran, R. L. Scott, G. E. Ponce-Campos, and B. Mitra. 2016. Spatial patterns of vegetation response to climate variability across the American Southwest. Ecosphere 7:e01339. http://dx.doi.org/10.1002/ecs2.1339

Boden, T. A., M. Krassovski, and B. Yang. 2013. The AmeriFlux data activity and data system and evolving collection of data management techniques, tools, products and services. Geoscientific Instrumentation Methods and Data Systems 2:165–176.

Bosch, D. D., J. M. Sheridan, R. R. Lowrance, R. K. Hubbard, T. C. Strickland, G. W. Fevereisen, and D. G. Sullivan. 2007. Little River experimental watershed database. Water Resources Research 43:W09470.

Bryant, R., K. Havstad, P. Heilman, P. Kleinman, T. B. Moorman, M. S. Moran, J. L. Steiner, and T. Strickland. 2013. Long Term Agroecological Research Network, shared research strategy. USDA-ARS. https://www.ars.usda.gov/ARSUserFiles/np211/LTAR_SRS_Final_29Sept2015.pdf

Bryant, R. B., T. L. Veith, G. W. Feyereisen, A. R. Buda, C. D. Church, G. J. Folmar, J. P. Schmidt, C. J. Dell, and P. J. A. Kleinman. 2011. USDA-ARS Mahantango Creek watershed, Pennsylvania, United States: physiography and history. Water Resources Research 47:W08701.

Hampton, S. E., C. A. Strasser, J. J. Tewksbury, W. K. Gram, A. E. Budden, A. L. Batchellar, C. S. Duke, and J. H. Porter. 2013. Big data and the future of ecology. Frontiers in Ecology and the Environment 11:156–162.

Heffernan, J. B., et al. 2014. Macrosystems ecology: understanding ecological patterns and processes at continental scales. Frontiers in Ecology and the Environment 12:5–14.

Hobbie, J. E., S. R. Carpenter, N. B. Grimm, J. R. Gosz, and T. R. Seastedt. 2003. The US long term ecological research program. BioScience 53:21–32.

Hottenstein, J. D., G. E. Ponce-Campos, J. M. Yanes, and M. S. Moran. 2015. Impact of varying storm intensity and consecutive dry days on grassland soil moisture. Journal of Hydrometeorology 16:106–117.

Huete, A., G. Ponce-Campos, Y. Zhang, N. Restrepo-Coupe, X. Ma, and M. S. Moran. 2015. Chapter 1: Monitoring photosynthesis from space. Pages 3–22 *in* P. S. Thenkabail, editor. Remote sensing handbook. Volume 2. Land resources monitoring, modeling, and mapping with remote sensing. Taylor and Francis, Boca Raton, Florida, USA.

Huxman, T. E., et al. 2004. Convergence across biomes to a common rain-use efficiency. Nature 429: 651–654.

Intergovernmental Panel on Climate Change (IPCC). 2013. Climate change 2013: the physical science basis. Pages 1–1535 *in* T. F. Stocker, D. Qin, G.-K. Plattner, M. Tignor, S. K. Allen, J. Boschung, A. Navels, Y. Xia, V. Bex, and P. M. Midgley, editors. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, UK.

Kampe, T. U., B. R. Johnson, M. Kuester, and M. Keller. 2010. NEON: the first continental-scale ecological observatory with airborne remote sensing of vegetation canopy biochemistry and structure. Journal of Applied Remote Sensing 4:043510.

Kao, R. H., et al. 2012. NEON terrestrial field observations: designing continental-scale, standardized sampling. Ecosphere 3. http://dx.doi.org/10.1890/ES12-00196.1

Luo, Y., K. Ogle, C. Tucker, S. Fei, C. Gao, S. LaDeau, J. S. Clark, and D. S. Schimel. 2011. Ecological forecasting and data assimilation in a data-rich era. Ecological Applications 21:1429–1442.

Ma, X., et al. 2013. Spatial patterns and temporal dynamics in savanna vegetation phenology across the North Australian Tropical Transect. Remote Sensing of Environment 139:97–115.

Marks, D. G. 2001. Introduction to special section: Reynolds Creek experimental watershed. Water Resources Research 37:2817.

Michener, W. K., and M. B. Jones. 2012. Ecoinformatics: supporting ecology as a data-intensive science. Trends in Ecology and Evolution 27:85–93.

Moran, M. S., D. P. C. Peters, M. P. McClaren, M. H. Nichols, and M. B. Adams. 2008*a*. Long-term data collection at USDA experimental sites for studies of ecohydrology. Ecohydrology 1:377–393.

Moran, M. S., et al. 2008*b*. Preface to special section on fifty years of research and data collection: U.S. Department of Agriculture Walnut Gulch experimental watershed. Water Resources Research 44:W05S01.

Moran, M. S., et al. 2014. Functional response of U.S. grasslands to the early 21st century drought. Ecology 95:2121–2133.

Peters, D. P. C., K. M. Havstad, J. Cushing, C. Tweedie, O. Fuentes, and N. Villanueva-Rosales. 2014*a*. Harnessing the power of big data: infusing the scientific method with machine learning to transform ecology. Ecosphere 5. http://dx.doi.org/10.1890/ES13-00359.1

Peters, D. P. C., H. W. Loescher, M. D. SanClements, and K. M. Havstad. 2014*b*. Taking the pulse of a continent: expanding site-based research infrastructure for regional- to continental-scale ecology. Ecosphere 5. http://dx.doi.org/10.1890/ES13-00295.1

Peters, D. P. C., et al. 2013. Long-term trends in ecological systems: a basis for understanding responses to global change. Technical Bulletin No. 1931. U.S. Department of Agriculture, Washington, D.C., USA. http://www.ars.usda.gov/is/np/LongTermTrends/LongTermTrendsIntro.htm

Ponce Campos, G., et al. 2013. Ecosystem resilience despite large-scale altered hydro climatic conditions. Nature 494:349–352.

Porter, J. H., P. C. Hanson, and C.-C. Lin. 2012. Staying afloat in the sensor data deluge. Trends in Ecology and Evolution 27:121–129.

Robertson, G. P., et al. 2008. Long-term agricultural research: a research, education and extension imperative. BioScience 58:640–645.

Rocchini, D., et al. 2015. Potential of remote sensing to predict species invasions: a modeling perspective. Progress in Physical Geography 39:283–309.

Sadler, E. J., et al. 2015. Long-term agroecosystem research in the Central Mississippi River Basin: introduction, establishment, and overview. Journal of Environmental Quality 44:3–12.

Silberzahn, R., and E. L. Uhlmann. 2015. Many hands make tight work. Nature 526:189–191.

Steiner, J. L., P. J. Starks, J. D. Garbrecht, D. N. Moriasi, X. J. Zhang, J. M. Schneider, J. A. Guzman Jaimes, and E. Osei. 2014. Long-term environmental research: the Upper Washita River experimental watersheds, Oklahoma, USA. Journal of Environmental Quality 43:1227–1238.

Walbridge, M. R., and S. R. Shafer. 2011. A Long-Term Agro-Ecosystem Research (LTAR) Network for agriculture. Pages 26–30 *in* L. Koss, editor. Proceeding of the Fourth Interagency Conference on Research in the Watersheds: Observing, Studying, and Managing Change. Bureau of Land Management, Fairbanks, Alaska, USA.

Zhang, X., M. S. Moran, X. Zhao, S. Liu, T. Zhou, and G. E. Ponce Campos. 2014. Impact of prolonged drought on rainfall use efficiency using MODIS data across China in the early 21st century. Remote Sensing of Environment 150:188–197.

Zhang, Y., et al. 2013. Extreme precipitation patterns reduced terrestrial ecosystem production across biomes. Journal of Geophysical Research-Biogeosciences 118:148–157.

Supporting Information

Additional Supporting Information may be found online at: http://onlinelibrary.wiley.com/doi/10.1002/ecs2.1493/supinfo