An Integrated View of Complex Landscapes: A Big Data-Model Integration Approach to Transdisciplinary Science

DEBRA P.C. PETERS, N. DYLAN BURRUSS, LUIS L. RODRIGUEZ, D. SCOTT MCVEY, EMILE H. ELIAS, ANGELA M. PELZEL-MCCLUSKEY, JUSTIN D. DERNER, T. SCOTT SCHRADER, JIN YAO, STEVEN J. PAUSZEK, JASON LOMBARD, STEVEN R. ARCHER, BRANDON T. BESTELMEYER, DAWN M. BROWNING, COLBY W. BRUNGARD, JERRY L. HATFIELD, NIALL P. HANAN, JEFFREY E. HERRICK, GREGORY S. OKIN, OSVALDO E. SALA, HEATHER SAVOY AND ENRIQUE R. VIVONI

The Earth is a complex system comprising many interacting spatial and temporal scales. We developed a transdisciplinary data-model integration (TDMI) approach to understand, predict, and manage for these complex dynamics that focuses on spatiotemporal modeling and cross-scale interactions. Our approach employs human-centered machine-learning strategies supported by a data science integration system (DSIS). Applied to ecological problems, our approach integrates knowledge and data on (a) biological processes, (b) spatial heterogeneity in the land surface template, and (c) variability in environmental drivers using data and knowledge drawn from multiple lines of evidence (i.e., observations, experimental manipulations, analytical and numerical models, products from imagery, conceptual model reasoning, and theory). We apply this transdisciplinary approach to a suite of increasingly complex ecologically relevant problems and then discuss how information management systems will need to evolve into DSIS to allow other transdisciplinary questions to be addressed in the future.

Keywords: cross-scale interactions, machine learning, landscape ecology, data science, earth science

Unanticipated ecological changes and humans' inability to quickly mitigate or adapt to them have caused significant socioeconomic disruptions. Examples include the Dust Bowl of the 1930s; Hurricanes Harvey, Irma, and Maria in 2017; and disease outbreaks, including Zika and dengue (MEA 2005, Jones et al. 2008, Peters et al. 2008, Carpenter et al. 2009). These "ecological surprises" often result when biological processes, spatial heterogeneity in land surface properties, and environmental drivers interact across spatial and temporal scales to influence multiple levels of biological organization that can lead to alternative states of the system (Bestelmeyer et al. 2011, Johnson et al. 2015). Examples of events governed by cross-scale interactions include wildfires that start small (ignition of an individual tree) and spread to encompass increasingly broader spatial extents as a series of thresholds are crossed when dominant processes change through time (Peters et al. 2004a). For other surprising events, such as floods, volcanic eruptions, and hurricanes, the high-intensity impact of the environmental driver is applied over a broadscale to overwhelm fine-scale heterogeneity in land-surface properties and biological processes that homogenizes impacts and responses over large areas (Brokaw et al. 2012). Heterogeneity can subsequently redevelop as environmental drivers act on the underlying landscape template (Romme et al. 2016) or subtle changes in one or more drivers can cause dramatic, nonlinear responses that can be difficult or impossible to reverse if thresholds are crossed (Fagre et al. 2009). Such changes are often unanticipated because the sequencing and patterns of events, their intensity of impact, or rate and pattern of spread and recovery lack historical precedent (Nadeau et al. 2017).

Ecologists, biologists, and Earth-system scientists are increasingly being asked to predict the occurrence of such events, minimize their negative impacts, and promote system recovery. However, predictions and management recommendations are often based on knowledge of a select subset of system components or by extrapolating data from other locations or previous time periods (e.g., grazing management

BioScience 68: 653–669. Published by Oxford University Press on behalf of American Institute of Biological Sciences 2018.
doi:10.1093/biosci/biy069

in the American Southwest in the late 1800s was based on European and midwestern approaches with catastrophic results; Bestelmeyer et al. 2006). Extrapolations based on such simplifications and assumptions are further compromised and invalidated when (a) a new driver is introduced, such as an exotic species, or a key biological component is ignored (Young et al. 2017); (b) nonlinear and cross-scale interactions result in unanticipated emergent behavior that differs from dynamics of individual components or earlier time periods (Peters et al. 2004a, Soranno et al. 2014); (c) spatial processes connect ecosystem components by the flow of material, propagules, energy, or information (spatial contagion) to overwhelm local processes (Peters et al. 2006, Okin et al. 2015); (d) spatial heterogeneity in the land surface template results in variable ecological responses that accentuate or dampen environmental driver effects (Seidl et al. 2016); (e) legacies, lags, thresholds, and feedback loops occur with nonlinear responses to environmental drivers to violate linearity assumptions (Bestelmeyer et al. 2011, Collins et al. 2014); and (f) the mean and extreme values of one or more drivers extend beyond their historical range of variability (i.e., nonstationarity; Milly et al. 2008). If one or more of these conditions occurs, then extrapolation of data or results from one location (or time period) to other locations (or time periods) can lead to high uncertainty and large unexplained error in the results (Miller et al. 2004, Peters et al. 2004b, Dixon Hamil et al. 2016). Thus, alternative approaches are needed to account for these nonlinear, complex dynamics.

New technologies are increasing awareness that the Earth is a complex, interconnected system at many spatial and temporal scales (Heffernan et al. 2014). Dynamics at one scale can influence or have consequences for scales and dynamics of interest to other disciplines through the flow of material, energy, or information (Liu et al. 2015). These interactions can be increasingly quantified or observed directly with recent advances in technology (e.g., Vivoni 2012). For example, genomics, metagenomics, and microbiological techniques reveal fine-scale interconnections among soil nutrients, microbial communities, and the plant microbiome that explain patterns of adaptation and survival observed by plant biologists and ecologists (Wullschleger et al. 2015). The importance of scale and interactions among scales to emergent ecological dynamics has a long history in ecology (e.g., Levin 1992, Gurevitch et al. 2016), but an approach that quantifies relationships among processes, patterns, and drivers—within and across scales—for a system exhibiting complex dynamics is still needed (Stegen 2018).

Advances in sensors, software, and other infrastructure are also allowing more biological and physical processes, environmental properties, locations, and time steps to be detected, quantified, coordinated, and made accessible through the Internet and other sources than at any time in history (Hampton et al. 2013, Michener 2015). Advances in big data sharing, analytics, and related open-source software tools and plug-ins are facilitating open science across disciplines (Hampton et al. 2015). For example, the R statistical language (R Development Core Team 2017) consolidates a variety of user-contributed packages in one coherent data analysis environment; similarly, proprietary software (e.g., the ArcGIS suite of geospatial software) can be used to collate disciplinary contributions into an integrated framework (e.g., Brown 2014).

New technologies are also supporting the need for a transdisciplinary research approach to understand, predict, and manage for these complex dynamics (Plowright et al. 2008, Reid et al. 2010). For example, understanding the multidimensional aspect of dengue spread requires collaboration among physicians, epidemiologists, and ecologists (Vasilakis et al. 2011). Although data availability and open science within disciplines is increasing (Hampton et al. 2013, Soranno et al. 2014), there is an urgent need for cross-disciplinary collaboration among scientists to integrate and leverage existing data and understanding of processes to inform the strategic collection of new data, to guide management in undersampled locations, and to better predict future conditions so as to minimize ecological surprises and their impacts (Reid et al. 2010, NASEM 2016).

We propose that recent advances in technologies and an emphasis on data and metadata standards and sharing within and across disciplines have positioned ecologists to reduce the probability of future ecological surprises. Our goal was to develop an operational transdisciplinary approach that accommodates and facilitates integration of large and diverse types of data and knowledge to (a) reduce the high spatial heterogeneity in sampling frequency, intensity, and quality across the surface of the Earth and fill data or knowledge gaps for underrepresented locations; (b) characterize the nonstationarity of environmental drivers and ascertain the extent to which knowledge of the past can or cannot inform the future; and (c) inform land managers and others in prioritizing locations.

Our transdisciplinary data-model integration (TDMI) approach combines software technologies in human-assisted machine learning to optimize the efficiency of the scientific method with recent advances in analyses of big data, including multimodel comparisons, multivariate dimension reduction, and machine-learning techniques, with data and knowledge from domain experts and other multiple lines of evidence (figures 1 and 2). The approach is supported by a data science integrated system (DSIS), in which data from long-term observations and experimental manipulations, output from analytical and numerical models, and products from sensors and imagery products, are standardized and harmonized to promote integration. As applied to the specific problems that we study, ecological dynamics in drylands, we sought to extend point- or plot-based understanding to landscape, watershed, and regional scales in order to improve land management and prediction capabilities. To accomplish this, our approach integrates biological processes and associated patterns at multiple scales, spatial heterogeneity in the land surface template, and environmental

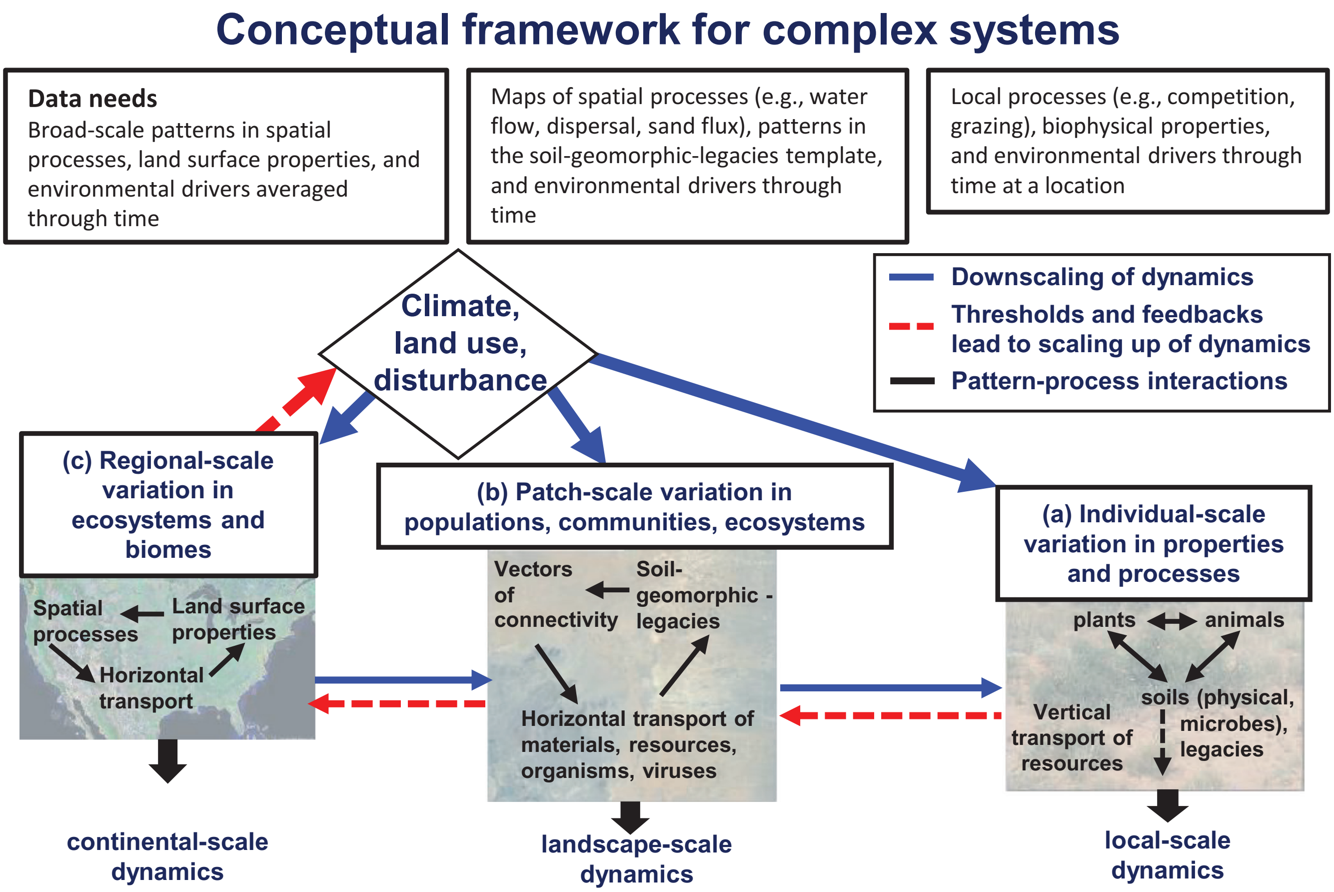

Figure 1. Conceptual framework for complex systems. Spatial scales are linked by processes that either propagate from finer to larger scales (upscaling, red arrows) or overwhelm finer-scale processes and patterns (downscaling, blue arrows). Changes in the pattern–process relationships through time within locations can influence dynamics across locations via cross-scale interactions. Data needs differ for each spatial scale: for example, local processes and biophysical properties at fine scales; maps of spatial patterns in the soil-geomorphic templates, environmental drivers, and spatial processes at landscape scales; and broadscale patterns in spatial processes, land surface properties, and environmental drivers averaged through time at regional to continental scales. Modified from (Peters et al. 2008).

drivers using data and knowledge drawn from multiple lines of evidence across multiple spatiotemporal scales. Our TDMI approach goes beyond *extrapolation*, which typically relies on projecting or extending observations into unknown locations (Turner et al. 1989). We focus on functional responses at multiple, interacting scales that link patterns and processes, and allow processes to propagate across scales of spatial heterogeneity (Cadenasso et al. 2007, Peters et al. 2007); flows of material within and among spatial units can overwhelm or attenuate local processes (Okin et al. 2015). Relationships are developed between observed pattern and underlying processes within each scale, and the pattern-process relationships connect scales to provide a mechanistic understanding of dynamics that extend across scales in space or time and lead to cross-scale emergence (Peters et al. 2007, Heffernan et al. 2014).

We apply this approach to a suite of increasingly complex ecologically relevant challenges that culminate in the development of the TDMI approach. Our first example is a simple issue to demonstrate the following concepts: (a) local dynamics: primary production in wet versus dry periods—then build in complexity with expanding spatial extent to illustrate the analytical techniques; (b) landscape-scale dynamics: Great Plains agroecosystems during historic drought—and finally increase the scope of the problem that requires an transdisciplinary team of scientists working with software engineers; (c) regional- to continental-scale dynamics: patterns of animal disease spread across the western United States. In these examples, we use readily available, standardized data, primarily from Web-based sources in our DSIS. These analyses are followed by more challenging applications of our approach, wherein we address questions that integrate short- and long-term data in time and space across a heterogeneous landscape. Many of these data originate from independent, investigator-driven experiments and observational studies with different

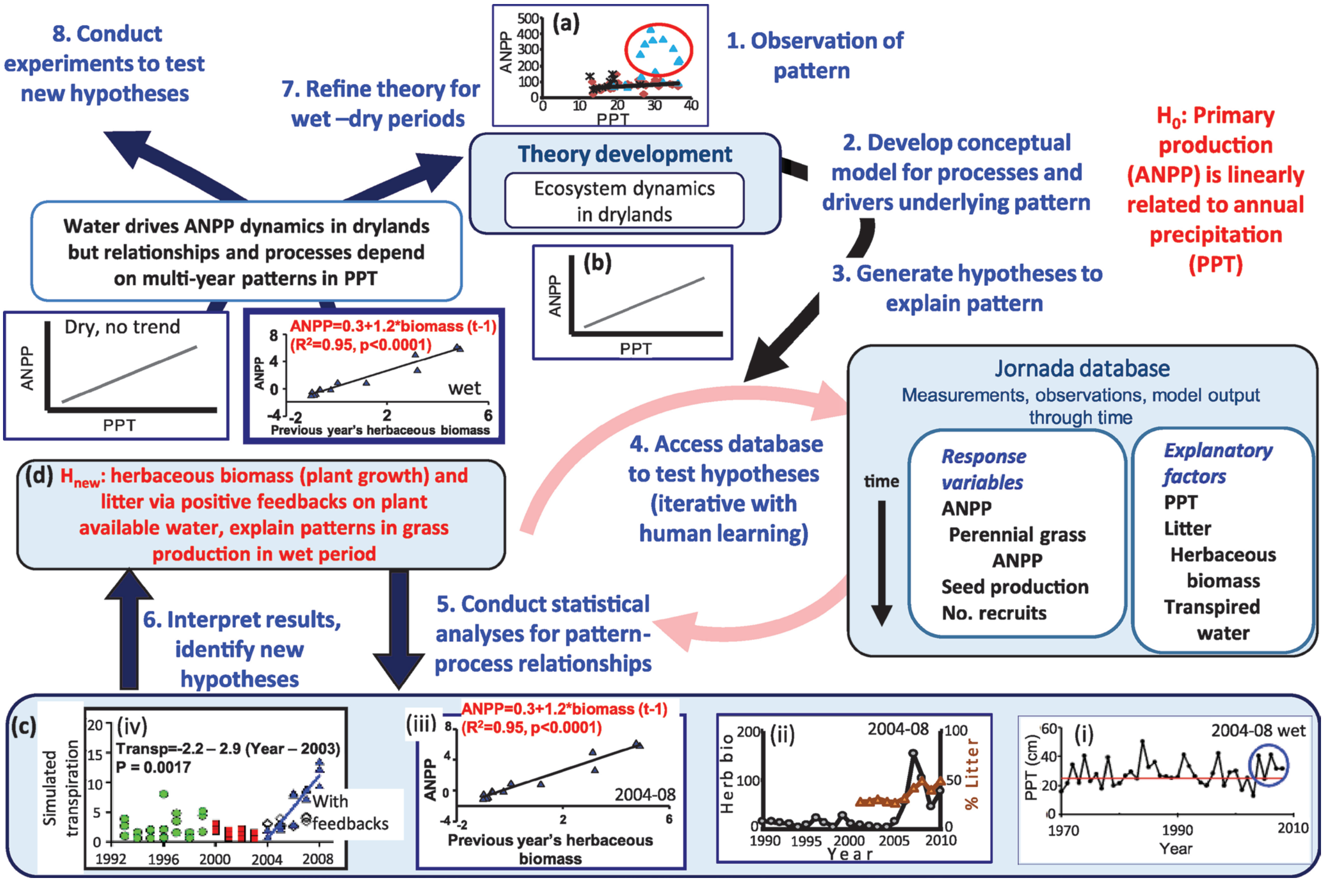

Figure 2. Developing the conceptual basis of transdisciplinary data-model integration: Primary production in wet versus dry periods at local scale. We developed an eight-step framework with relatively simple statistical analyses to explain observations of aboveground net primary production at the Jornada that were remarkably higher than expected based on previous studies and dryland theory. An iterative approach with human learning (light pink arrow) was used to access the long-term Jornada database multiple times to analyze different data sets as different hypotheses about grass demographic processes were tested. A simulation model parameterized with additional Jornada data was used to refine the hypothesis testing and to guide new experiments to be tested in the field. Dryland theory about the relationship between precipitation and primary production was refined on the basis of grass responses in wet periods. See the text for details.

motivations, thus making spatial or temporal integration challenging.

Transdisciplinary data-model integration approach

Our transdisciplinary data-model integration (TDMI) approach is based on a process-based systems approach designed to allow spatial contagion and heterogeneity, temporal nonlinearities, cross-scale interactions, and multiple levels of organization to be evaluated directly. Scales are linked by processes leading to dynamics that can either propagate from finer to broader spatial scales (upscaling, red arrows) or allow for the transfer of material from larger areas to influence and overwhelm processes and patterns at finer scales (downscaling, blue arrows; figure 1). Changes in environmental drivers and pattern–process relationships through time and across space can alter system dynamics within particular locations, and can change dynamics across locations and regions. Our approach derives from hierarchy theory (Allen and Starr 1985) but accommodates cross-scale interactions by identifying conditions wherein broadscale drivers overwhelm fine-scale variability and fine-scale processes propagate nonlinearly to influence broad spatial extents (Peters et al. 2004a). These thresholds, feedback loops, and nonlinear relationships among patterns and processes within and among scales can lead to ecological surprises unless cross-scale interactions and multiscale processes are accounted for.

Our examples focus on a landscape as the functional unit for management decisions. However, the logic of an integrated system can be applied to any spatial extent where information is available at both finer and coarser scales. Because noncontiguous areas may influence the functional unit of interest via the provisioning of materials, information or energy, system dynamics at a given location will also depend on the degree

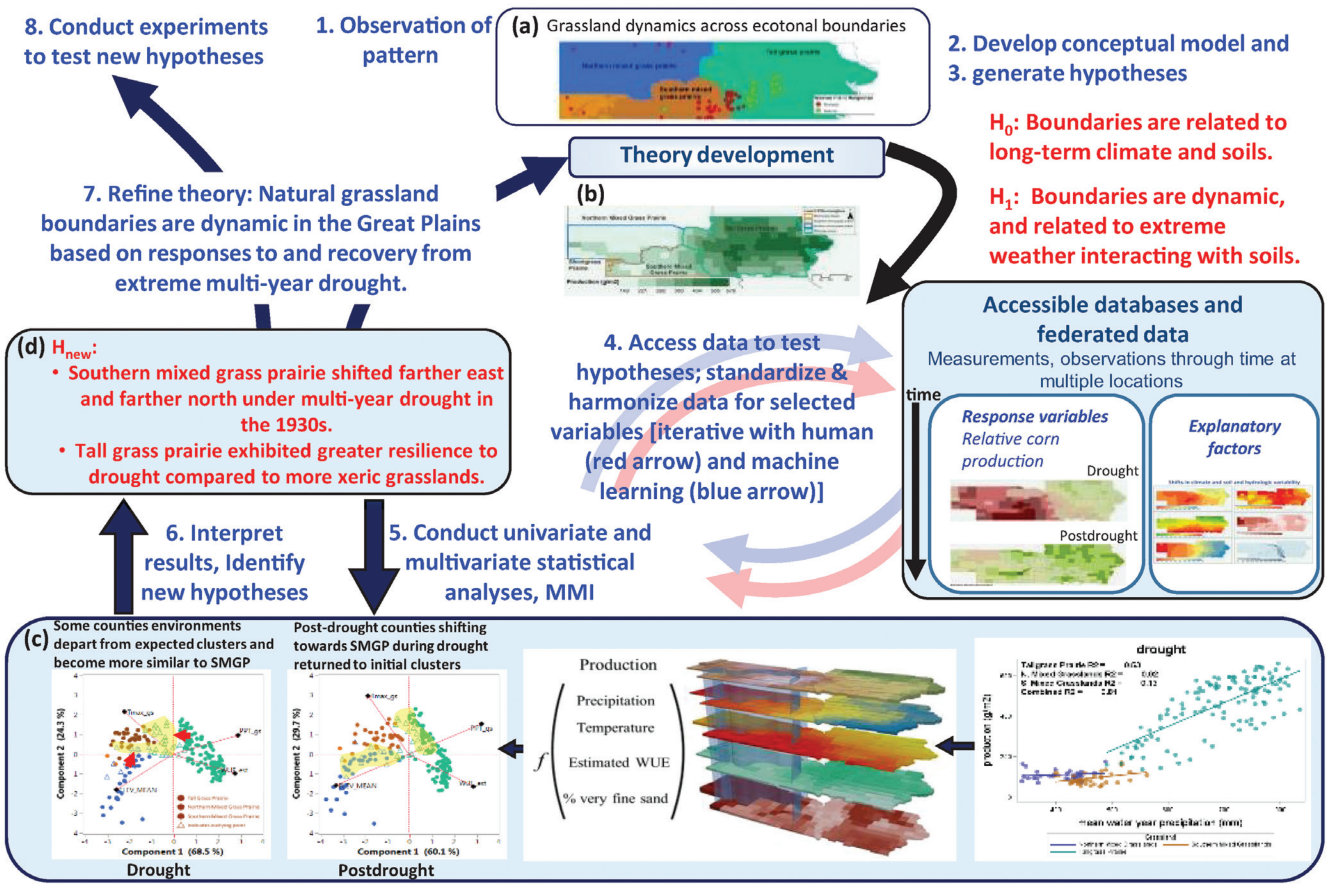

Figure 3. Developing the analytical basis of transdisciplinary data-model integration: Landscape-scale dynamics in Great Plains agroecosystems during historic drought. We used the eight-step framework from figure 2 and developed more detailed computational and statistical analyses beginning in step 4 that include human-guided machine learning (light blue arrow) as well as multivariate analyses, maximum entropy, and multimodel comparisons. In this example, all data were obtained from federated US government sources available on the Internet. The data were first standardized in space and time (county and water year) and then harmonized to the same projection system before statistical analyses were conducted. The results suggested that the western portion of the tallgrass prairie shifted toward the southern mixed-grass prairie during the 1930s drought and shifted back toward tallgrass prairie after the drought (Burruss et al. 2017). Theory about ecotonal boundaries between potential natural grasslands can be refined to be dynamic in response to and recovery from extreme, multiyear drought. This example provides guidance on the data harmonization, multivariate analysis, and multimodel comparison techniques, as well as the role of expert oversight at each step needed for complex regional analyses. See the text for details.

of connectivity and strength of these exogenous influences and how they vary with time and distance.

Developing the conceptual basis of transdisciplinary data-model integration: Primary production in wet versus dry periods at local scale

We first illustrate how the conceptual basis of the TDMI approach was developed to improve understanding of local-scale dynamics and to refine ecological theory in drylands using a suite of long-term data sets from the Chihuahuan Desert Jornada Basin Long Term Ecological Research (LTER) site (32.3 °N; 106.4 °W, 1188 m; hereafter, the Jornada). Jornada landscapes have transitioned from perennial grasslands to dominance by unpalatable, drought-tolerant, woody plants (Buffington and Herbel 1965). This transition exemplifies land transformations in drylands globally, and is often accompanied by soil and nutrient redistribution that fundamentally changes ecosystem processes (e.g., Eldridge et al. 2011). More recently, transitions from shrublands toward grasslands and novel ecosystems are challenging traditional paradigms (Peters et al. 2015). Consequently, concepts and theories of state change dynamics, uncertainties under global change, and macrosystems dynamics (Bestelmeyer et al. 2011, Sala et al. 2012), are more pertinent to these landscapes than desertification dynamics (Reynolds and Stafford-Smith 2002).

Our first step was to graph long-term observations of grass aboveground net primary production (ANPP) in degraded shrublands at the Jornada as a function of annual precipitation (PPT; figure 3a). In step 2, we expected that grass would be a linear function of PPT based on dryland theory (Noy-Meir 1973) and empirical studies from other sites (e.g., Sala et al. 1988, 2012; see our figure 3b). Because long-term Jornada data did not fit the expected pattern (figure 3a, red circle and blue triangles), we sought alternative mechanistic explanations. A closer examination of PPT records revealed that unusually high ANPP values corresponded to a 5-year period of high rainfall (figure 3c.i, blue circle; Peters et al. 2012). In step 3, we used this pattern to generate hypotheses pertaining to demographic processes underlying grass recovery that we evaluated in step 4 using multiple long-term data sets on grass recruitment. In steps 5 and 6, our analyses and results showed that the nonlinear ANPP response could not be explained by seed production, seed germination, or seedling establishment (Peters et al. 2014b). In iterative accesses of the databases, we hypothesized that increases in litter cover associated with increases in perennial grass and other herbaceous biomass would reduce evaporation losses and create a positive feedback to increase plant available water (PAW) to grasses. Additional data from another study in the same location supported this hypothesis by finding abrupt and nonlinear increases of biomass and litter (figure 3c.ii), and a significant relationship between previous year's herbaceous biomass and current year's grass ANPP during the wet period (figure 3c.iii). We then used a soil water simulation model (SOILWAT; Peters et al. 2010) to examine the mechanistic underpinning for our hypothesis. We conducted two simulations with the same input conditions, except one had additional inputs of herbaceous biomass and litter comparable to those that occurred in 2004–2008 (with feedback mechanisms). Because the simulations with increased biomass and litter resulted in an increase in PAW (and transpiration; figure 3c.iv), we have a new hypothesis that can be field-tested in step 8 as a mechanism for grass recovery in wet periods. In addition, we concluded that although the linear relationship between annual PPT and ANPP holds for shrublands during relatively dry and no trend rainfall periods, a new understanding of nonlinear dynamics and refinement of theory is needed in wet periods that accounts for grass establishment and growth over several years (step 7). This new understanding was made possible by a team of ecologists, their collective understanding of relevant processes, the availability of multiple long-term data sets from one location at the Jornada, and a small set of tools (regression analyses, simulation model). More importantly, the sequence of steps in developing the pattern–process relationships in our conceptual model (observe pattern → generate hypothesis → test hypothesis → develop conceptual model → access data base ⇔ develop relationships ⇔ identify new hypotheses → refine theory) with iteration and human learning among steps (shown as pink arrows in figures) became part of the approach employed in our next examples.

Developing the analytical basis of transdisciplinary data-model integration: Landscape-scale dynamics in Great Plains agroecosystems during historic drought

We next describe how the analytical components of the TDMI approach were developed. Here, we sought to determine whether boundaries between natural vegetation types are static and related to climate and fine-scale patterns in soils, as was expected based on ecological theory (e.g., Küchler 1964), or are dynamic and related to other factors associated with extreme weather. We used historic data during (1933–1940) and after (1941–1948) the extreme drought of the 1930s to examine potential dynamics in the boundaries between grassland types in the central Great Plains of the United States. We examined boundaries on potential vegetation maps between the tallgrass prairie (TP) and the southern (SMP) or the northern mixed-grass prairies (NMP), and between the SMP and NMP (figure 4a, step 1). We then developed a conceptual model on the basis of ecotones (step 2) and hypothesized that boundaries were either related to long-term conditions or to variable rainfall and temperature, topographic relief and water redistribution, soil texture, and land use (figure 4b, step 3). Because this drought predated modern agriculture in the United States and the advent of genetic manipulation (Hatfield and Walthall 2015), we used historic corn yield (*Zea mays*) as a surrogate for perennial grass production. Data obtained from online US government sources for explanatory and response variables were standardized in units (county, annual), and maps were harmonized to the same projection system (step 4; Burruss et al. 2017). In step 5, highly correlated variables ($p > .70$) were identified using univariate statistics and subsequently removed from the analysis. Machine learning and multivariate analyses were used to compare multidimensional relationships. A suite of models consistent with our hypotheses was fit to the data. The best model was chosen by a combination of quantitative information criterion in machine learning and subjective expert opinion (i.e., Fieberg and Johnson 2015; see our figure 4c). In step 6, results indicate the western portion of the tallgrass prairie shifted toward vegetation characteristic of the southern mixed-grass prairie during the drought, then shifted back toward tallgrass prairie after the drought (figure 4d). In step 7, we refined our theory that the boundaries between natural grassland subtypes are dynamic and responsive to extreme climatic events. From our TDMI perspective, this example illustrates how data harmonization and multivariate analysis with human learning, and multimodel comparisons with machine learning and expert oversight can be integrated for complex landscape to regional analyses. Machine learning is shown as light blue arrows in all figures.

Developing a transdisciplinary team and a data science integrated system: Regional- to continental-scale dynamics in the patterns of animal disease spread across the western United States

In our third example, we describe how we develop a transdisciplinary team and a DSIS in the TDMI to identify and

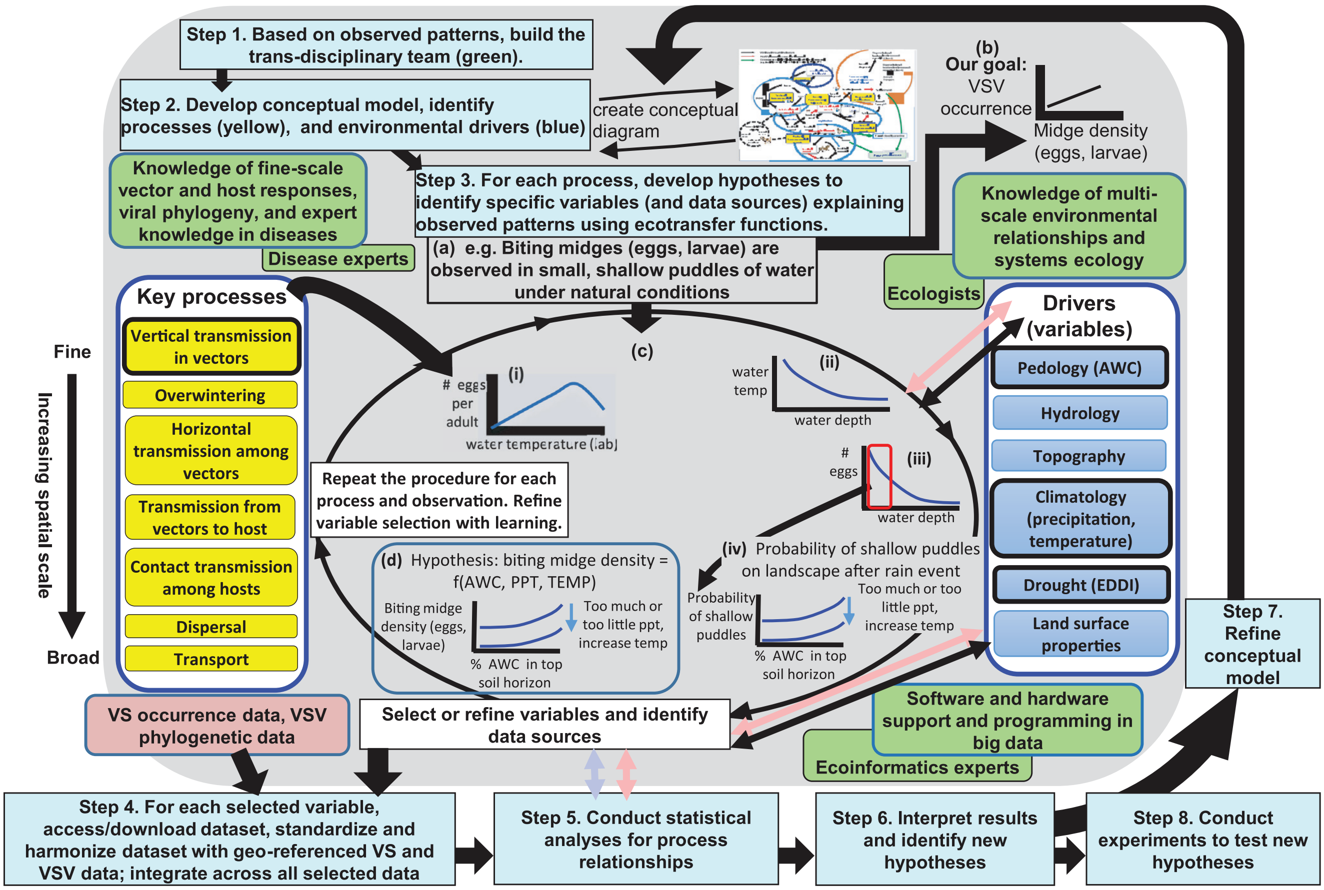

Figure 4. Developing a transdisciplinary team in transdisciplinary data-model integration: Regional- to continental-scale dynamics in the patterns of animal disease spread in the western United States. We developed a geospatial, iterative framework with human and human-guided machine learning based on the eight steps in figure 2 and the computational analyses in figure 3. We used this framework as a predictive approach to integrate large amounts of diverse data in an attempt to explain multiyear patterns in the occurrence of vesicular stomatitis (VS), a vector-borne disease in livestock that occurs across the western United States. Vesicular stomatitis cannot be explained by one or a few biophysical or climatic factors. We enhanced the eight steps from figures 2 and 3 by adding the formation of a transdisciplinary team as part of the initial step and developing a systems diagram (figure 5) as part of step 2. This diagram was needed to show how the processes are related to each other and to their environment (yellow boxes and details in figure 5) before the hypotheses could be generated and variables selected in step 3. The next enhancement was the use of a series of linked pattern–process relationships (i.e., ecotransfer functions) based on the systems diagram combined with data on vector and host responses to their environment and environmental data to develop new hypotheses. This process in step 3 also identified the variables to be selected for analysis from the many possible variables in an environmental driver data set. These data then underwent standardization and harmonization that required close collaboration among scientists and software engineers to ensure that the most appropriate data sets and calculations were used for each variable. This iterative human-learning procedure ensured that data sets or calculations were examined for accuracy, reliability, and comparability with other data sets. A suite of analyses was then conducted depending on the spatial and temporal scale of the question that included human-learning and human-guided machine-learning techniques involving feedback with the transdisciplinary team. See the text for details. This example shows the development of the transdisciplinary approach with mostly federated data that are accessible on the Internet. The data sets within each driver type (blue boxes) have been standardized and harmonized to each other; the challenge is harmonizing across the different types of data sets that are guided by the question.

analyze the environmental factors and biological processes explaining the invasion of North America by vesicular stomatitis (VS) virus, a vector-borne, zoonotic RNA virus that affects livestock. Vesicular stomatitis is a reportable disease for national and international trade reasons, and a database exists beginning in the year 2000 (*www.aphis.usda.gov*). We used this data set within our TDMI approach to explore the complexity of the VS disease system (vector–host–virus–environment), and to identify the processes and environmental variables governing its spatial or temporal patterns. Use of the TDMI in this context would test its utility in transdisciplinary, continental-scale complex systems in which diverse data and knowledge are available but not integrated.

Vesicular stomatitis is the most commonly reported livestock vesicular disease in the Americas (Rodríguez 2002), occurring in greater than 1.1 million square kilometers (km^2) of the western United States from 2004 to 2016. In ruminants, VS resembles foot-and-mouth disease (FMD), a devastating animal disease absent in the United States, requiring rapid reporting and differential diagnostics of VS cases. Vesicular stomatitis has been primarily studied by epidemiologists with the assistance of veterinarians and local, state, and federal authorities. Control of VS outbreaks is primarily based on animal quarantines and strategies for managing exposure to insect vectors (primarily black flies and biting midges). Spatial patterns of disease have been related to one or a few biophysical and climatic factors (Rodríguez et al. 1996, McCluskey et al. 2003), but VS occurrence is expected to be strongly influenced by the biology and genetics of its vectors and hosts interacting with their environment (climate, soil, hydrology, vegetation). Understanding the VS system thus requires an approach integrating (a) transdisciplinary scientific expertise, (b) very large, heterogeneous databases over time across the western United States to account for insects biologies, viral phylogenetics, disease occurrence in livestock, environmental heterogeneity in time and space; and (c) technical expertise for data harmonization, integration, and analysis.

Guided by our research on data-model integration at the Jornada (figure 3), our multimodel, iterative analytical approaches (figure 4), and recent developments in machine learning (Peters et al. 2014a), we developed a geospatial, iterative TDMI approach featuring human-guided machine learning to coherently integrate (a) fine-scale process-based data and understanding of vector and host responses to a pathogen and the local environment; (b) georeferenced disease incidence and virus phylogenetic data; and (c) fine-scale patterns of climate, hydrology, topography, soils, vegetation, and host density over a multidecadal time period across the continental extent of the disease (figure 5). This approach provides an objective method for evaluating the potential importance of individual and interacting environmental variables within and across scales (local, landscape, region) when environmental data are available but little is known about the ecology of a complex system. Iterative human learning among team members occurs at steps 2–4, whereas human-guided machine learning occurs at step 5. Machine-learning methods are used to identify structure in complex, often nonlinear data in order to generate accurate predictive models (Olden et al. 2008). Although parts of the modeling process are automated through machine learning, our human-guided approach uses expert knowledge to specify how the data are represented and to identify the variables to be included in the model as the iterations proceed, in an extension of the process described in (Olden et al. 2006).

Step 1: Transdisciplinary team formation and synthesis of disease system understanding. Our transdisciplinary team consists of scientists with expertise in the phylogenetics of VS, the biology of the insect vectors, and equine (host) epidemiology interacting with ecologists with expertise to account for interactions between the disease components and their environment (an ecohydrologist, a range scientist, and a landscape ecologist). Ecoinformatics experts provide software and hardware expertise to standardize, harmonize, and analyze the diverse data sets. A systems ecologist integrates the scientific and technological components of the VS system.

Step 2: Synthesis of disease system understanding. Based on the team's expertise, general relationships were formulated to explain spatial heterogeneity and temporal variability in VS across the spatiotemporal occurrences of the disease. A conceptual model of the disease system was developed to describe how the six processes related to VS transmission and dispersal are interrelated (figure 2). This diagram was instrumental in synthesizing the known information about VS and highlighting knowledge gaps. The development of this conceptual model and associated diagram were refined through time using human learning as the team worked through the problem.

Step 3: Hypothesis development and variable and data-set identification. The conceptual model was used to develop specific hypotheses to explain the relationship between environmental drivers and biotic factors related to host density (number of animals, number of ranch or farm properties). Online, open-access databases were identified for each driver and mined to retrieve information at the temporal and spatial resolution, extent, and duration congruent with disease occurrence data and questions being addressed by the science team. Weekly climatic data were selected as the finest temporal scale of resolution to account for uncertainty in the date of occurrence of infection compared with the reported date and a lack of detailed knowledge about certain aspects of insect biology. Temporal aggregations (e.g., seasonal averages of climatic variables) were conducted for the hypotheses to be tested. The spatial resolution (1 km × 1 km) was selected to capture uncertainty in the exact location where each animal became infected. The spatial extent was the 10-state region in the western United States for the entire duration where the disease occurred, including the prior year to account for time lags and legacy effects (2003 to 2016). We strategically selected variables for analysis by developing hypotheses that linked biological (e.g., biting-midge egg density) and environmental variables (e.g., soil water–holding capacity) needed to explain vector or host lifecycle stages under natural conditions (e.g., biting-midge eggs are found in small, shallow puddles of water) that were expected to contribute to disease spread (figure 5). These "pattern–process relationships" are similar to "pedotransfer functions" that predict difficult to measure soil properties from readily available soil profile data (e.g., Sequeira et al. 2014). New hypotheses were specified

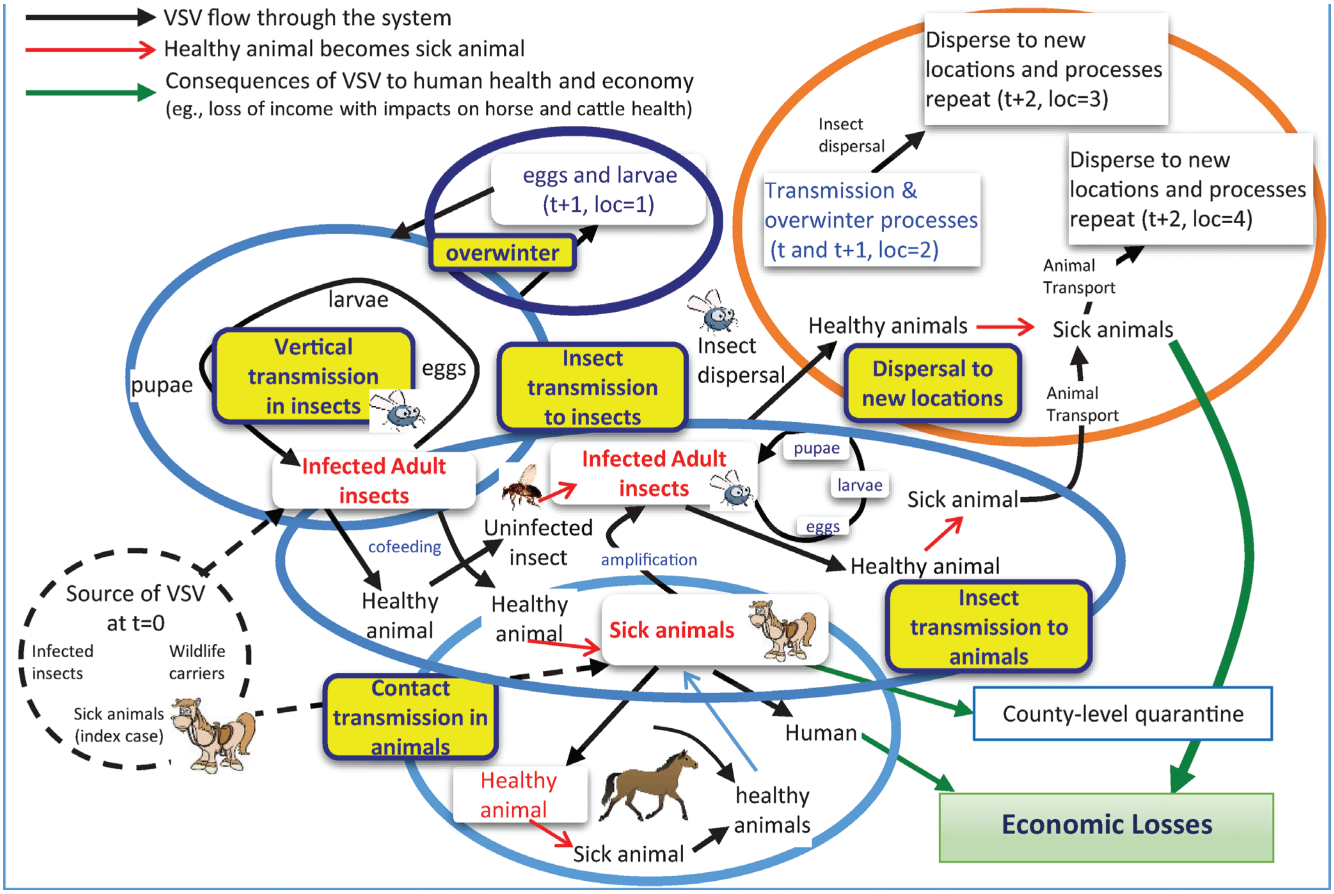

Figure 5. Systems diagram. A conceptual model of the disease system was developed to describe how the six processes related to disease transmission (yellow boxes) are related to each other. The different colored circles show either temporal (dark blue: overwintering) or spatial processes (light blue: local; red: spatial dispersal). The red arrows indicate change from healthy vector or host to infected vector or host. The green arrows indicate losses due to county-level quarantine and economic losses. For many diseases, including vesicular stomatitis (VS), experimental data of vector or host response to drivers are most often examined under controlled conditions in a laboratory (e.g., Walton et al. 1987, Kramer et al. 1990, Cupp et al. 1992). These data-rich studies provide potential responses to a climatic driver under laboratory conditions. Geographic patterns in the disease and its genetic lineage under natural environmental conditions (www.usda.aphis.gov) were used to develop relationships that resulted in the initial systems diagram that was refined through time with more data.

through model-building exercises based on expert knowledge of existing data on vector and host responses to their environment combined with environmental data deemed parsimoniously relevant. For example, the STATSGO2 soils database contains more than 70 attribute tables with more than 850 variables. We selected two variables (available water holding capacity, AWC, and clay content) in the surface horizon because midge larvae development occurs in small puddles on the soil surface whose occurrence are related to these variables (Mullens 1989). This approach was repeated for each process with each driver and factor to determine the variables to be analyzed (figure 5, yellow and blue boxes). The processes vary in spatial scale from very fine (less than one millimeter; vector to vector transmission) to very broad (kilometers; dispersal by insects and transport by vehicles), and interactions with variables across scales can lead to complex, nonlinear dynamics (figure 5). This procedure was iterative with learning between the scientists and the software engineers as online data sources were examined for their resolution, and additional data sources were obtained as needed.

Step 4: Data harmonization and integration with a data science integration system. After identifying variables and their data sets, harmonization was needed to facilitate integration. First, the VS disease occurrence data were converted from a geographic coordinate system to an equal area projection system (e.g., Albers Equal Area Conic) to ensure cell size remained the same (1 km^2) throughout the large spatial extent (approximately 1.1 million km^2) of the study area. All other variables underwent harmonization to match the projection, geographic origin, and cell size of this VS occurrence base map. For the raster data (e.g., gridded PPT at a 4 km × 4 km resolution), harmonization consisted of resampling maps to 1 km × 1 km. Vector data (e.g., points, lines, and curves) were converted to raster, harmonized to the base layer, and then translated into distance maps. Polygons were rasterized by calculating average properties, then harmonized to the base layer. All spatial data were manipulated with ArcGIS v.10.3 that assisted in this harmonization procedure. Ultimately, the variable selection procedure resulted in 472 raster layers, which were collated into a harmonized data cube as a DSIS to enable calculations and predictions to be carried out in geographical space. For analyzing processes local to the VS occurrences, tabular, as opposed to geographic, databases within the DSIS were devised. Tabular data sets preserved fine-scale (weeks or months prior to a VS incident) temporal relationships between VS occurrence and environmental variables without masking by arbitrary classifications (e.g., month or season) at broad spatial scales. These tabular data were extracted from raster maps by calculating the mean of values extracted from a 100-point grid centered on a VS occurrence location that covered a cell size (4 km^2) to characterize the environment surrounding a VS occurrence. The harmonization of geographic and tabular data enabled analyses to be carried out in a consistent manner across spatial and temporal scales. Harmonization required close collaboration among scientists and software engineers to ensure the most appropriate data sets and calculations were used for each variable. This was an iterative human-learning procedure where data sets or calculations were examined for accuracy, reliability, and comparability with other data sets.

Step 5: Analysis and interpretation of results. Data cube and occurrence data were used to construct a landscape- to regional-scale species distribution model by a machine-learning maximum entropy approach (Phillips et al. 2004). The model was evaluated in tandem by exploratory data analysis with team expertise using the package MaxentVariableSelection in R (Jueterbock 2015) to control model complexity, avoid collinearity among predictor variables, and optimize parameters. Although information-theoretic approaches can guide this evaluation (Warren and Seifert 2011), the use of experts to identify biologically meaningful variables is paramount (Fourcade et al. 2017). The number of candidate variables in tabular data sets for multivariate analysis was reduced to 20 for local-scale analysis using an iterative process to identify those with the strongest univariate relationship to VS occurrence and avoid collinearity among predictors more than 70%. Model performance was assessed using the corrected Akaike information criterion (AIC), which provides a relative measure of model quality considering fit and complexity (Peters et al. 2017).

Step 6: Test new hypotheses and feedback to conceptual model. Our TDMI approach was successful in identifying vectors of importance in different outbreak phases (e.g., black flies in the first years, then biting midges as the expansion spreads; Peters et al. 2017). The next step will be to conduct focused experiments to test these new hypotheses emerging from the TDMI approach, and to improve our understanding of the system by refining the vector–host–virus–environment interrelationships in the conceptual model (figure 2). Because our approach harmonizes and analyzes a large suite of diverse biotic and environmental data across multiple scales, we generated new insights into a complex system in which vector–host–environment relationships in the western United States were largely unknown. This new cross-scale understanding in space and time was only possible through the transdisciplinary team of scientists and software engineers of working together.

Developing a knowledge landscape map integrated with a data science integrated system

Here, we extend the TDMI approach described above to see whether it can be used to integrate diverse, long-term environmental data with detailed process-based data and knowledge spanning multiple levels of organization obtained from disparate locations to create a fully integrated "knowledge landscape map." This map would integrate multiple lines of evidence from specific study locations and time periods in a process-based approach by accounting for spatial heterogeneity in patterns and temporal nonlinearities in processes at multiple interacting scales.

We chose the Jornada to test this application of the TDMI approach for four reasons. First, because the Jornada has been a US Department of Agriculture research site since 1912 and a National Science Foundation Long Term Ecological Research Site since 1982, there are many long-term data sets of the biota and environmental variables available (*www.jornada.nmsu.edu/lter*), and these are complimented by data from diverse long-term experiments, simulation model outputs, imagery products, and onsite expertise for collaboration summarized in (Havstad et al. 2006, Peters et al. 2015). Second, developing general principles from the Jornada will be applicable to a large portion of the Earth's surface because this landscape is representative of drylands that cover 40% of the land surface of the Earth (Reynolds and Stafford-Smith 2002). Third, the Jornada landscape presents a significant challenge to the TDMI because it is spatially heterogeneous and temporally variable with thresholds, feedback mechanisms, lags, and legacies (Bestelmeyer et al. 2011, Sala et al. 2012, Monger et al. 2015). If the TDMI works here, it should also work for other complex terrestrial landscapes.

The fourth reason is that there is tremendous need for this type of landscape integration of data and knowledge at the Jornada. The primary questions addressed at the Jornada relate to the causes of variation in, and restoration options to mitigate, the long-term transformation from perennial grasslands to shrublands. Research approaches have involved "disentangling" the landscape into finer-grained components based on landforms, soil properties, and soil–vegetation relationships, and developing state-transition models and management strategies on the basis of these units (e.g., Monger and Bestelmeyer 2006). This simplification has resulted in classifying the Jornada Basin according to soil-geomorphic units that exhibit variation in vegetation and soils across a hierarchy of scales that differ with respect to the two dominant physical transport vectors (wind, water) at broad scales (figure 6a, 6b; Monger 2006). A third vector, animals, interacts with wind and water to redistribute resources and create multiscale patterns (the red circle in figure 6b insert). The resulting five ecosystem types represent Chihuahuan Desert ecosystems, and are the basis for the stratified sampling used in many Jornada studies (e.g., Peters et al. 2010, Rachal et al. 2015). This disentangling approach has facilitated a deep understanding of the processes and drivers governing patterns at fine to intermediate spatial and temporal scales (Havstad et al. 2006).

Although this approach has led to many insights into controls on function by ecosystem type, our ability to apply this knowledge to other locations or time periods is constrained on four fronts common to long-term research sites worldwide: (1) It ignores connectivity among geomorphic units that might drive change and lead to ecological surprises (Okin et al. 2015). For example, geomorphic units can be connected by large-scale processes of wind or water that generate different dynamics in different locations (Stewart et al. 2014). On the piedmont slope sand sheet, sand is deposited by large-scale wind processes from locations farther west resulting in long-term shrub coppice dune development whereas on the piedmont slope bajada, the Doña Ana mountains block the wind such that water redistribution is the predominant driver of vegetation change leading to shrub dominance associated with sheet erosion and an absence of dunes (Gibbens et al. 2005, Monger 2006). (2) Inferences developed for specific ecosystems must be reliably extended to other locations, either at the Jornada or at other sites only cautiously. (3) Studies established to sample or monitor representative areas of interest for particular reasons (individual or programmatic), do not represent the local, landscape or regional spatial heterogeneity, and sampling locations are not strategic or coordinated through time. Accordingly, more than 60% of studies at the Jornada have focused on the basin floor sand sheet (figure 6c). Long-term observations or monitoring of vegetation, soil, animals, and climate are distributed across a larger part of the Jornada than many of the experiments (figure 6d), often because observations were located programmatically to provide baseline data for all researchers. Large parts of the Jornada remain undersampled and undercharacterized relative to other intensively studied locations. This disparity is true for all research sites, but is magnified as site extent increases. (4) Studies are an eclectic collection of short- and long-term ad hoc investigations with inconsistent response and explanatory variables, methods (including sampling frequency or intensity), timings and durations. This limits their utility for comparison, synthesis, and integration among studies. Accordingly, the standardization and harmonization processes in the TDMI approach will be more challenging than when using federated data from Internet sources (e.g., the VS example above).

Below, we describe an approach using well-defined data from multiple ecosystem types through time, and illustrate how general principles can be spatially distributed and applied to new locations. We then describe how a DSIS will be needed to support a landscape knowledge map for large, heterogeneous, long-term study sites such as the Jornada.

Estimating aboveground net primary production across the Jornada Basin using our integrated approach

In our final example, we sought to extend our previous result of perennial grass recovery on one location (the basin floor sand sheet) to the entire Jornada. We first translate our simple example (figure 3) into our transdisciplinary approach (figure 7) to address two goals: (1) improve our understanding of the processes controlling location-specific patterns through time to predict future dynamics, and (2) extend these findings to other locations using a process-based understanding of intensively studied locations. For the latter, we used perennial grass ANPP data from creosotebush and tarbush shrubland ecosystems on the piedmont slope bajada and transition zone geomorphic units, respectively (figure 7f). We then develop a grass recovery index based on the slope of the relationship between ANPP through the wet 2004–2008 period for each of the nine shrubland locations using a soil property (AWC) and initial perennial grass biomass in 2004 (figure 7g). This index was then used to estimate ANPP in each year of the wet period for the entire Jornada Basin (figure 7h) that can be compared with satellite products for the same time periods. This estimate of perennial grass recovery based on a mappable soil property and initial grass biomass is not specific to an ecosystem type.

This approach is not simple extrapolation because it accounts for thresholds in grass recovery through time that occurred between a drought (2000–2003) and the wet period (2004–2008). In addition, the traditional classification system of ecosystem types had to be abandoned for an approach that allowed a continuous change in cover across the landscape. There could also be thresholds in space that are not associated with ecosystem types. These nonlinearities in time and space cannot be addressed in simple extrapolation methods.

This relatively simple example reveals how our TDMI approach can integrate multiple lines of evidence to produce a landscape knowledge map for a spatially heterogeneous research site, accounting for thresholds, legacies, and lags (figure 7d, 7e, 7g). This example was possible because the response data (ANPP) and the explanatory data (PPT, AWC, and initial grass biomass) were readily standardized and harmonized for these research locations. All of these data are core data sets that are georeferenced, collected, and postprocessed by the Jornada LTER in the same way through time; thus, integration techniques focus on harmonized data sets.

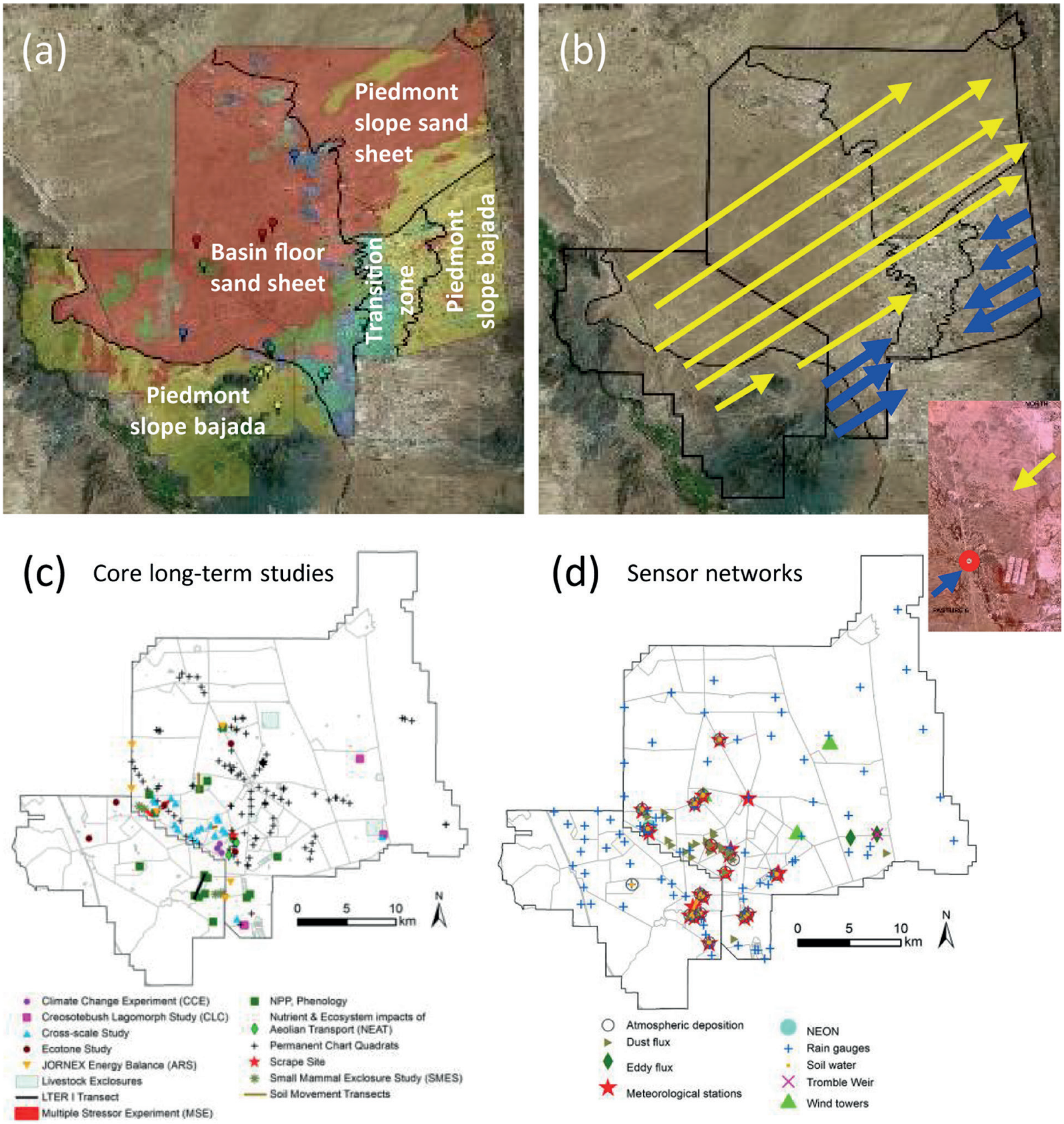

*Figure 6. Spatial heterogeneity in the Jornada landscape. (a) Four soil-geomorphic units exhibit variation in vegetation and soils across a hierarchy of scales: basin floor sand sheet, piedmont slope sand sheet, transition zone, and piedmont slope bajada. (b) Each unit differs with respect to the two dominant physical transport vectors (wind, yellow arrows; water, blue arrows) at broad scales. Insert: All three vectors (including animals; red circle) redistribute resources to result in multiscale patterns across the landscape. Feedback mechanisms between spatial variation in vegetation result in ecosystem types with different dominant vegetation and soil properties. Five ecosystem types occur that generally correspond to soil-geomorphic units (a): Upland grasslands dominated by black grama (*Bouteloua eriopoda*; green) and mesquite-dominated shrublands (*Prosopis glandulosa*; red) are found primarily on the basin floor and piedmont slope sand sheets; creosotebush-dominated shrublands (*Larrea tridentata*; yellow) occur on the piedmont slope bajada; tarbush-dominated shrublands (*Flourencia cernua*; light green) occur in the transition zone; and playa grasslands dominated by tobosa grass (*Pleuraphis mutica*; blue) occur on low-lying areas throughout the landscape. (c) Long-term experiments and observational studies at the Jornada have been located based on study-specific objectives; most experiments occur on the sand sheet. (d) Observational and sensor networks at the Jornada are primarily programmatic and are spatially distributed to cover the heterogeneity of the research site.*

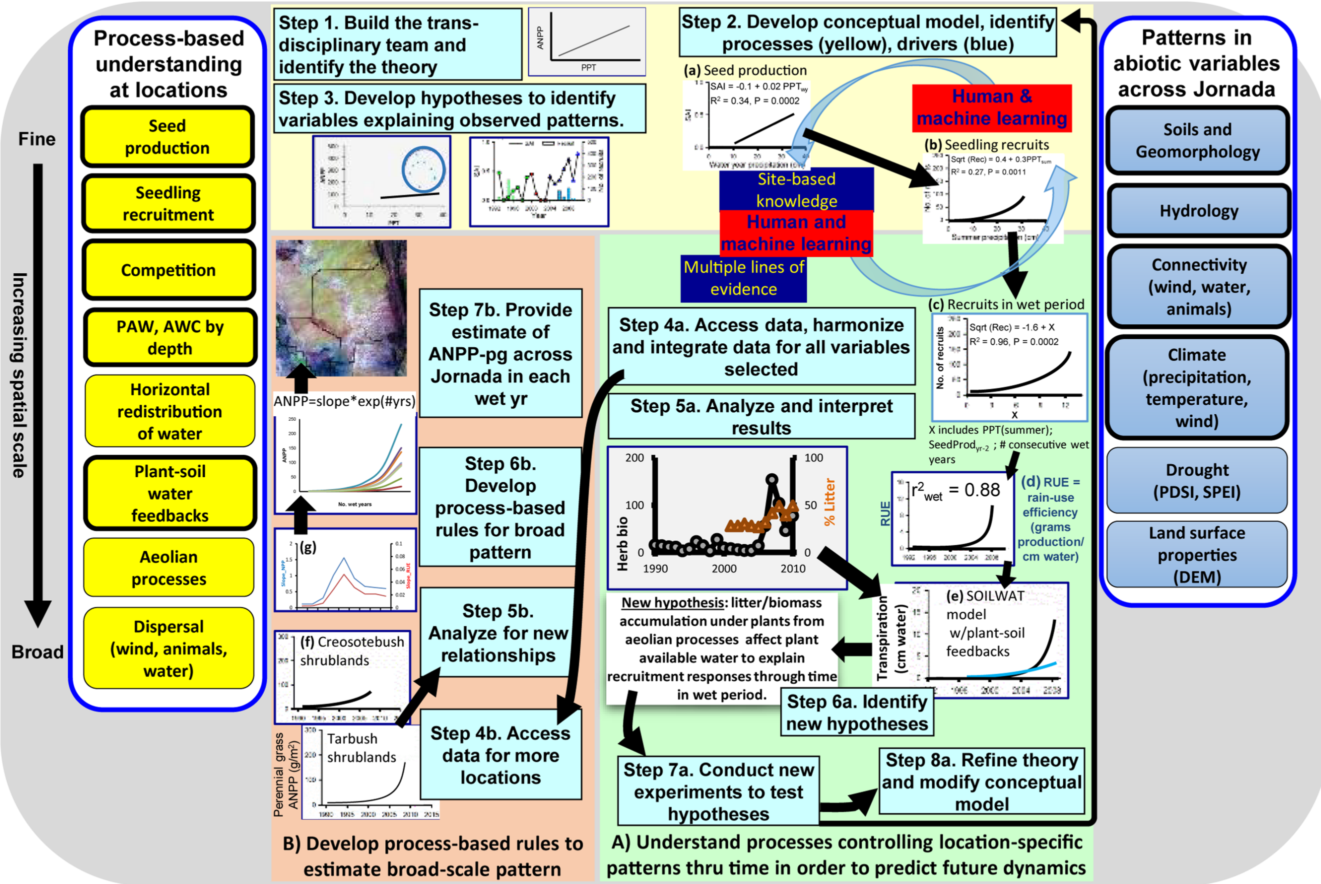

Figure 7. Estimating aboveground net primary production (ANPP) across the Jornada Basin using our transdisciplinary data-model integration approach. We extended our previous result of perennial grass recovery on one location (figure 2) to the entire Jornada using our approach in order to (a) to predict future dynamics at this location, and (b) extend these findings to other locations across the landscape. In both cases, human and human-guided machine learning are used in the data acquisition and analysis phases and when developing the best ecotransfer functions for this system. To make predictions (a), we use our process-based understanding to develop relationships between precipitation and ANPP (in this example) during wet and dry periods that can be used under future climatic conditions. To extend our results to other locations (b), we used similar perennial grass ANPP data through time from two other ecosystem types (creosotebush shrublands and tarbush shrublands) on other locations. We developed an index of grass recovery based on a soil property and initial grass biomass that is not specific to an ecosystem type on the Jornada and can be tested at other dryland locations. See the text for details.

A general transdisciplinary data-model integration approach with multiple lines of evidence in a data science integrated system

More complex questions will require the integration of multiple lines of evidence that are first standardized and harmonized for the entire Jornada Basin. For example, observations in parts of the Jornada landscape suggest that wind and water interact within and across scales to influence ecosystem dynamics (Okin et al. 2018). In other parts of the landscape, an integration of hydrological and ecological processes across scales is needed for reconstructing the current mosaic of plant community patterns (Vivoni 2012). For instance, shrublands on piedmont slopes promote runoff into channels, which can lead to the transport of water to augment downslope precipitation, to influence vegetation state transitions and productivity in adjacent sites, and to confound local precipitation–production relationships (Schreiner-McGraw and Vivoni 2017). At a long-term research site such as the Jornada, these multiple lines of evidence include data from long-term observations and experimental manipulations, output from analytical and numerical models, and products from sensors and imagery products that form the basis of a DSIS. Based on our examples (figures 2–7), an eight-step TDMI approach with human-guided machine learning and multiple lines of evidence is proposed for addressing three common phenomena: alternative states, global change, and cross-scale interactions (figure 8). Theoretical frameworks can organize the knowledge base needed for understanding system behavior and pattern–process linkages within and across scales. The key to the

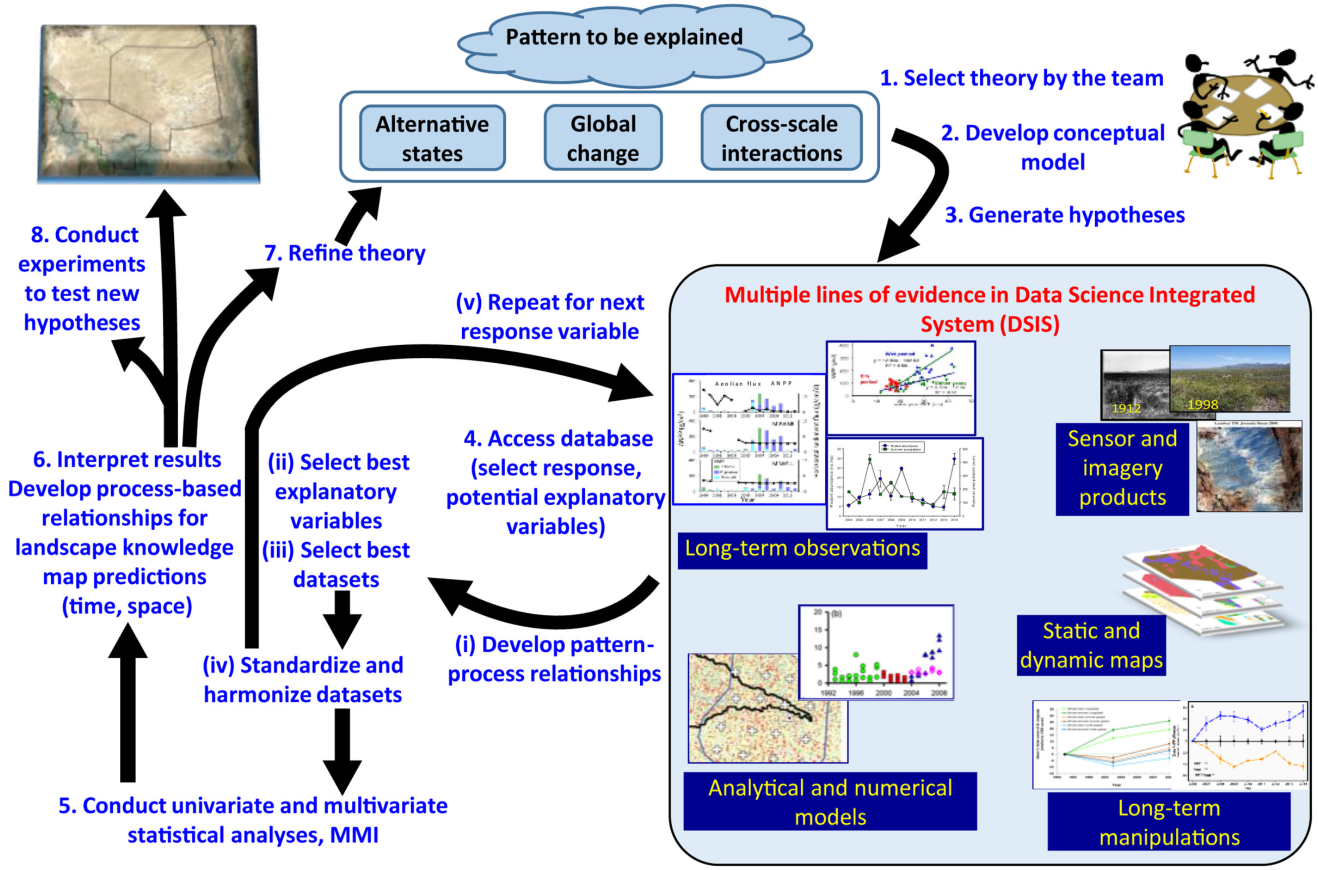

Figure 8. General transdisciplinary data-model integration approach for transdisciplinary research. Our approach based on spatiotemporal modeling of cross-scale interactions coupled with human and human-guided machine learning integrates (a) biological processes, (b) spatial heterogeneity in the land surface template, and (c) environmental drivers using data and knowledge drawn from multiple lines of evidence (i.e., short- and long-term observations, experimental manipulations, analytical and numerical models, products from imagery, conceptual model reasoning, and theory) in a data science integrated system. The eight steps that follow from a pattern that needs to be explained are similar regardless of the complexity of the problem.

TDMI approach is (a) site-based knowledge, (b) expertise in multiple disciplines, and (c) a systems perspective that enables an understanding and quantification of cross-scale interactions through the development of pattern–process relationships (i.e., ecotransfer functions) at multiple, interacting scales and levels of biological organization.

Retooling the information management system into a data science and integration system

We envision a DSIS that broadens traditional information management systems that store and make accessible data and metadata in a standard format to include suites of harmonized explanatory and response variables for each study location. Studies typically include response and explanatory variables defined by investigator- rather than site-specific objectives, thus limiting their broader utility. A standard set of response or explanatory variables will be needed for landscape integration based on knowledge of the system that goes beyond the environmental drivers. For example, depth to a restrictive layer is a better explanatory variable for perennial grass resilience during drought than surface soil texture (Herbel et al. 1972, Browning et al. 2012). Thus, depth to calcium carbonate should be measured or estimated at all research locations in drylands and made accessible as part of the DSIS. In addition, derived data products and pattern–process relationships developed during each TDMI procedure should be maintained as part of the DSIS. Maintaining a developmental history of the integrated landscape data sets, similar to that of the knowledge learning analysis system (Peters et al. 2014a), will allow users to build on previous users' experiences and extend location-based understanding to other areas or to future points in time.

Conclusions

New technologies are increasing awareness that the Earth is a complex, interconnected system at many interacting spatial and temporal scales, and that understanding, predicting, and managing the dynamics of complex systems requires a transdisciplinary, data-intensive approach. Our premise was that technological advances in analytical, numerical, and behavioral approaches combined with extensive sources of readily available federated and locally available data make a transdisciplinary approach increasingly feasible and tractable for application to multiscale, complex ecological problems. The TDMI approach proposed here exploits a wide array of existing data and expert knowledge to solve specific ecological problems, and directs new research efforts in productive and strategic directions. We successfully applied this approach to a suite of examples to demonstrate how point- or plot-based understanding can be more reliably extended to landscape scales and therefore can better inform decisions relevant to land management. We also developed a general landscape knowledge map with a corresponding data science integration system that can be used for a more diverse collection of questions. Future developments will focus on making predictions with greater certainty, advancing toward the goal of minimizing ecological surprises and averting unintended consequences. The steps we proposed could be refined and used for a wide range of large-scale ecological, social, and Earth-system problems at other terrestrial research sites.

Acknowledgments

This work was supported by the US Department of Agriculture–Agricultural Research Service Current Research Information System Projects at the Jornada Experimental Range (no. 6235-11210-007), the Plum Island Animal Disease Center (no. 8064-32000-058-00D), the Center for Grain and Animal Health Research (no. 8064-32000-058-00D, #3020-32000-008-00D), the Rangeland Resources and Systems Research Unit (no. 3012-21610-001-00D), and the National Laboratory for Agriculture and the Environment (no. 5030-11610-003-00D). Funding was provided by the National Science Foundation to New Mexico State University for the Jornada Basin Long Term Ecological Research Program (nos. DEB 12-35828 and DEB 14-40166). We thank Mr. Darren James and Dr. Geovany Ramirez of NMSU for additional analyses in support of this study. We thank Dr. Matthew Pinch and Elizabeth Vera for manuscript production support. We thank the US Department of Agriculture Office of the Chief Scientist for support of DPCP. The VSV Project is a USDA ARS Grand Challenge Project.

References cited

Allen TBH, Starr TB. 1985. Hierarchy: Perspectives for Ecological Complexity. University of Chicago Press.

Bestelmeyer BT, Brown J, Havstad K, Fredrickson EL. 2006. A holistic view of an arid ecosystem: A synthesis of research and its applications. Pages 354–368 in Havstad KM, Huenneke LF, Schlesinger WH, eds. Structure and Function of a Chihuahuan Desert Ecosystem: The Jornada Basin Long-Term Ecological Research Site. Oxford University Press.

Bestelmeyer BT, et al. 2011. Analysis of abrupt transitions in ecological systems. Ecosphere 2: 1–26.

Brokaw N, Crowl T, Lugo A, McDowell W, Scatena F, Waide R, Willig MR. 2012. A Caribbean Forest Tapestry: The Multidimensional Nature of Disturbance and Response. Oxford University Press.

Brown JL. 2014. SDMtoolbox: A Python-based GIS toolkit for landscape genetic, biogeographic and species distribution model analyses. Methods in Ecology and Evolution 5: 694–700.

Browning DM, Duniway MC, Laliberte AS, Rango A. 2012. Hierarchical analysis of vegetation dynamics over 71 years: Soil–rainfall interactions in a Chihuahuan Desert ecosystem. Ecological Applications 22: 909–926.

Buffington LC, Herbel CH. 1965. Vegetational changes on a semidesert grassland range from 1858 to 1963. Ecological Monographs 35: 139–164.

Burruss ND, Peters DPC, Yao J. 2017. Integrating historic datasets to inform ecotonal boundaries between great plains grasslands. Paper presented at the 102nd Annual Meeting of the Ecological Society of America; 6–11 August 2017, Portland, Oregon.

Cadenasso ML, Pickett STA, Schwarz K. 2007. Spatial heterogeneity in urban ecosystems: Reconceptualizing land cover and a framework for classification. Frontiers in Ecology and the Environment 5: 80–88.

Carpenter SR, et al. 2009. Science for managing ecosystem services: Beyond the Millennium Ecosystem Assessment. Proceedings of the National Academy of Sciences 106: 1305–1312.

Collins SL, et al. 2014. A multiscale, hierarchical model of pulse dynamics in arid-land ecosystems. Annual Review of Ecology, Evolution, and Systematics 45: 397–419.

Cupp EW, Maré CJ, Cupp MS, Ramberg FB. 1992. Biological transmission of vesicular stomatitis virus (New Jersey) by *Simulium vittatum* (Diptera: Simuliidae). Journal of Medical Entomology 29: 137–140.

Dixon Hamil K-A, Iannone BV III, Huang WK, Fei S, Zhang H. 2016. Cross-scale contradictions in ecological relationships. Landscape Ecology 31: 7–18.

Eldridge DJ, Bowker MA, Maestre FT, Roger E, Reynolds JF, Whitford WG. 2011. Impacts of shrub encroachment on ecosystem structure and functioning: Towards a global synthesis. Ecology Letters 14: 709–722.

Fagre DB, et al. 2009. Climate Change Science Program: Thresholds of Change in Ecosystems. US Geological Survey. Synthesis and Assessment Product no. 4.2.

Fieberg J, Johnson DH. 2015. MMI: Multimodel inference for models with management implications? Journal of Wildlife Management 79: 708–718.

Fourcade Y, Besnard AG, Secondi J. 2017. Paintings predict the distribution of species, or the challenge of selecting environmental predictors and evaluation statistics. Global Ecology and Biogeography 27: 245–256.

Gibbens RP, McNeely RP, Havstad KM, Beck RF, Nolen B. 2005. Vegetation changes in the Jornada Basin from 1858 to 1998. Journal of Arid Environments 61: 651–668.

Gurevitch J, Fox GA, Fowler NL, Graham CH. 2016. Landscape demography: Population change and its drivers across spatial scales. Quarterly Review of Biology 91: 459–485.

Hampton SE, Strasser CA, Tewksbury JJ, Gram WK, Budden AE, Batcheller AL, Duke CS, Porter JH. 2013. Big data and the future of ecology. Frontiers in Ecology and the Environment 11: 156–162.

Hampton SE, et al. 2015. The Tao of open science for ecology. Ecosphere 6: 1–13.

Hatfield JL, Walthall CL. 2015. Meeting global food needs: Realizing the potential via Genetics × Environment × Management Interactions. Agronomy Journal 107: 1215–1226.

Havstad KM, Huenneke LF, Schlesinger WH, eds. 2006. Structure and Function of a Chihuahuan Desert Ecosystem: The Jornada Basin Long-Term Ecological Research Site. Oxford University Press.

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

Herbel CH, Ares FN, Wright RA. 1972. Drought effects on a semidesert grassland range. Ecology 53: 1084–1093.

Johnson PTJ, Ostfeld RS, Keesing F. 2015. Frontiers in research on biodiversity and disease. Ecology Letters 18: 1119–1133.

Jones KE, Patel NG, Levy MA, Storeygard A, Balk D, Gittleman JL, Daszak P. 2008. Global trends in emerging infectious diseases. Nature 451: 990–993.

Jueterbock A. 2015. R package MaxentVariableSelection: Selecting the Best Set of Relevant Environmental Variables along with the Optimal Regularization Multiplier for Maxent Niche Modeling. R Foundation for Statistical Computing. (4 June 2018; *https://cran.rproject.org/web/packages/MaxentVariableSelection/index.html*)

Kramer WL, Jones RH, Holbrook FR, Walton TE, Calisher CH. 1990. Isolation of arboviruses from *Culicoides* midges (Diptera: Ceratopogonidae) in Colorado during an epizootic of Vesicular Stomatitis New Jersey. Journal of Medical Entomology 27: 487–493.

Küchler AW. 1964. Manual to Accompany the Map: Potential Natural Vegetation of the Conterminous United States. American Geographical Society.

Levin SA. 1992. The problem of pattern and scale in ecology: The Robert H. MacArthur Award Lecture. Ecology 73: 1943–1967.

Liu J, et al. 2015. Systems integration for global sustainability. Science 347 (art. 1258832).

McCluskey BJ, Beaty BJ, Salman MD. 2003. Climatic factors and the occurrence of Vesicular Stomatitis in New Mexico, United States of America. Reviews in Science and Technology 22: 849–856.

[MEA] Millennium Ecosystem Assessment. 2005. Ecosystems and Human Well-Being: Synthesis Report. Island Press.

Michener WK. 2015. Ecological data sharing. Ecological Informatics 29: 33–44.

Miller JR, Turner MG, Smithwick EAH, Dent CL, Stanley EH. 2004. Spatial extrapolation: The science of predicting ecological patterns and processes. BioScience 54: 310–320.

Milly PCD, Betancourt J, Falkenmark M, Hirsch RM, Kundzewicz ZW, Lettenmaier DP, Stouffer RJ. 2008. Stationarity is dead: Whither water management? Science 319: 573–574.

Monger HC. 2006. Soil development in the Jornada Basin: Structure and function of a Chihuahuan Desert ecosystem. Pages 81–106 In Havstad KM, Huenneke LF, Schlesinger WH, eds. Structure and Function of a Chihuahuan Desert Ecosystem: The Jornada Basin Long-Term Ecological Research Site. Oxford University Press.

Monger HC, Bestelmeyer BT. 2006. The soil-geomorphic template and biotic change in arid and semi-arid ecosystems. Journal of Arid Environments 65: 207–218.

Monger HC, Sala OE, Duniway MC, Goldfus H, Meir IA, Poch RM, Throop HL, Vivoni ER. 2015. Legacy effects in linked ecological–soil–geomorphic systems of drylands. Frontiers in Ecology and the Environment 13: 13–19.

Mullens BA. 1989. A quantitative survey of *Culicoides variipennis* (Diptera: Ceratopogonidae) in dairy wastewater ponds in southern California. Journal Medical Entomology 26: 559–565.

Nadeau CP, Urban MC, Bridle JR. 2017. Climates past, present, and yet-to-come shape climate change vulnerabilities. Trends in Ecology and Evolution 32: 786–800.

[NASEM] National Academies of Sciences Engineering and Medicine. 2016. Big Data and Analytics for Infectious Disease Research, Operations, and Policy: Proceedings of a Workshop. National Academies Press.

Noy-Meir IC. 1973. Desert ecosystems: Environment and producers. Annual Review of Ecology and Systematics 4: 25–51.

Okin GS, Heras MM, Saco PM, Throop HL, Vivoni ER, Parsons AJ, Wainwright J, Peters DPC. 2015. Connectivity in dryland landscapes: Shifting concepts of spatial interactions. Frontiers in Ecology and the Environment 13: 20–27.

Okin GS, Sala OE, Vivoni ER, Zhang J, Bhattachan A. 2018. The interactive role of wind and water in functioning of drylands: What does the future hold? BioScience. doi:10.1093/biosci/biy067

Olden JD, Poff NL, Bledsoe BP. 2006. Incorporating ecological knowledge into ecoinformatics: An example of modeling hierarchically structured aquatic communities with neural networks. Ecological Informatics 1: 33–42.

Olden JD, Lawler JJ, Poff NL. 2008. Machine learning methods without tears: A primer for ecologists. Quarterly Review of Biology 83: 171–193.

Peters DPC, Pielke RA, Bestelmeyer BT, Allen CD, Munson-McGee S, Havstad KM. 2004a. Cross-scale interactions, nonlinearities, and forecasting catastrophic events. Proceedings of the National Academy of Sciences 101: 15130–15135.

Peters DPC, Herrick JE, Urban D, Gardner R, Breshears D. 2004b. Strategies for ecological extrapolation. Oikos 106: 627–636.

Peters DPC, Bestelmeyer BT, Herrick JE, Fredrickson EL, Monger HC, Havstad KM. 2006. Disentangling complex landscapes: New insights into arid and semiarid system dynamics. BioScience 56: 491–501.

Peters DPC, Bestelmeyer BT, Turner MG. 2007. Cross-scale interactions and changing pattern–process relationships: Consequences for system dynamics. Ecosystems 10: 790–796.

Peters DPC, Groffman PM, Nadelhoffer KJ, Grimm NB, Collins SL, Michener WK, Huston MA. 2008. Living in an increasingly connected world: A framework for continental-scale environmental science. Frontiers in Ecology and the Environment 6: 229–237.

Peters DPC, Herrick JE, Monger HC, Huang H. 2010. Soil–vegetation–climate interactions in arid landscapes: Effects of the North American monsoon on grass recruitment. Journal of Arid Environments 74: 618–623.

Peters DPC, Yao J, Sala OE, Anderson J. 2012. Directional climate change and potential reversal of desertification in arid and semiarid ecosystems. Global Change Biology 18: 151–163.

Peters DPC, Havstad KM, Cushing J, Tweedie C, Fuentes O, Villanueva-Rosales NC. 2014a. Harnessing the power of big data: Infusing the scientific method with machine learning to transform ecology. Ecosphere 5: 1–15.

Peters DPC, Yao J, Browning DB, Rango A. 2014b. Mechanisms of grass response in grasslands and shrublands during dry or wet periods. Oecologia 174: 1323–1334.

Peters DPC, Havstad KM, Archer SR, Sala OE. 2015. Beyond desertification: New paradigms for dryland landscapes. Frontiers in Ecology and the Environment 13: 4–12.

Peters DPC, Rodriguez L, McVey DS, Elias EH, Pelzel-McCluskey A, Derner JD, Yao J, Pauszek SJ, Schrader TS, Burruss N. 2017. OOS 7-10—Towards a theory of ecological catastrophes based on cross-scale interactions: Insights from long-term data. Paper presented at the 102nd Annual Meeting of the Ecological Society of America; 6–11 August 2017, Portland, Oregon.

Phillips SJ, Dudík M, Schapire RE. 2004. A maximum entropy approach to species distribution modeling. Paper presented at Twenty-First International Conference on Machine Learning; 4–8 July 2004, Banff, Alberta, Canada.

Plowright RK, Sokolow SH, Gorman ME, Daszak P, Foley JE. 2008. Causal inference in disease ecology: Investigating ecological drivers of disease emergence. Frontiers in Ecology and the Environment 6: 420–429.

R Development Core Team. 2017. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. (4 June 2018; *www.R-project.org*)

Rachal DM, Okin GS, Alexander C, Herrick JE, Peters DPC. 2015. Modifying landscape connectivity by reducing wind driven sediment redistribution, Northern Chihuahuan Desert, USA. Aeolian Research 17: 129–137.

Reid WV, et al. 2010. Earth system science for global sustainability: Grand challenges. Science 330: 916–917.

Reynolds JF, Stafford-Smith M. 2002. Global desertification: Do humans create deserts? Pages 1–22 in Stafford-Smith M, Reynolds JF, eds. Do Humans Create Deserts? Dahlem University Press.

Rodrıguez LL. 2002. Emergence and re-emergence of vesicular stomatitis in the United States. Virus Research 85: 211–219.

Rodríguez LL, Fitch WM, Nichol ST. 1996. Ecological factors rather than temporal factors dominate the evolution of vesicular stomatitis virus. Proceedings of the National Academy of Sciences 93: 13030–13035.

Romme WH, Whitby TG, Tinker DB, Turner MG. 2016. Deterministic and stochastic processes lead to divergence in plant communities 25 years after the 1988 Yellowstone fires. Ecological Monographs 86: 327–351.

Sala OE, Parton WJ, Joyce LA, Lauenroth WK. 1988. Primary production of the Central Grassland region of the United States. Ecology 69: 40–45.

Sala OE, Gherardi LA, Reichmann L, Jobbágy E, Peters DPC. 2012. Legacies of precipitation fluctuations on primary production: Theory and data synthesis. Philosophical Transactions of the Royal Society B 367: 3135–3144.

Schreiner-McGraw AP, Vivoni, ER. 2017. Deep observations in arid piedmont watersheds and linkages to historical conditions in the Chihuahuan Desert. Ecosphere 8 (art. e02000).

Seidl R, Donato DC, Raffa KF, Turner MG. 2016. Spatial variability in tree regeneration after wildfire delays and dampens future bark beetle outbreaks. Proceedings of the National Academy of Sciences 113: 13075–13080.

Sequeira CH, Wills SA, Seybold CA, West LT. 2014. Predicting soil bulk density for incomplete databases. Geoderma 213: 64–73.

Soranno PA, et al. 2014. Cross-scale interactions: Quantifying multi-scaled cause–effect relationships in macrosystems. Frontiers in Ecology and the Environment 12: 65–73.

Stegen JC. 2018. At the nexus of history, ecology, and hydrobiogeochemistry: Improved predictions across scales through integration. mSystems 3: 1–8.

Stewart J, Parsons AJ, Wainwright J, Okin GS, Bestelmeyer BT, Fredrickson EL, Schlesinger WH. 2014. Modeling emergent patterns of dynamic desert ecosystems. Ecological Monographs 84: 373–410.

Turner MG, Dale VH, Gardner RH. 1989. Predicting across scales: Theory development and testing. Landscape Ecology 3: 245–252.

Vasilakis N, Cardosa J, Hanley KA, Holmes EC, Weaver SC. 2011. Fever from the forest: Prospects for the continued emergence of sylvatic dengue virus and its impact on public health. Nature Reviews Microbiology 9: 532–541.

Vivoni ER. 2012. Spatial patterns, processes and predictions in ecohydrology: Integrating technologies to meet the challenge. Ecohydrology 5: 235–241.

Warren DL, Seifert SN. 2011. Ecological niche modeling in Maxent: The importance of model complexity and the performance of model selection criteria. Ecological Applications 21: 335–342.

Walton TE, et al. 1987. Epizootic vesicular stomatitis in Colorado, 1982: Epidemiologic and entomologic Studies. American Journal of Tropical Medicine and Hygiene 36: 166–176.

Wullschleger SD, Breen AL, Iversen CM, Olson MS, Näsholm T, Ganeteg U, Wallenstein MD, Weston DJ. 2015. Genomics in a changing arctic: Critical questions await the molecular ecologist. Molecular Ecology 24: 2301–2309.

Young HS, Parker IM, Gilbert GS, Sofia Guerra A, Nunn CL. 2017. Introduced species, disease ecology, and biodiversity–disease relationships. Trends in Ecology and Evolution 32: 41–54.

Debra P. C. Peters (deb.peters@ars.usda.gov), Emile H. Elias, Scott Schrader, Jin Yao, Brandon T. Bestelmeyer, Dawn M. Browning, and Jeffrey E. Herrick are affiliated with the US Department of Agriculture, Agricultural Research Service, Jornada Experimental Range Unit and the Jornada Basin Long Term Ecological Research Program, in Las Cruces, New Mexico. N. Dylan Burruss and Heather Savoy are affiliated with New Mexico State University, Jornada Experimental Range Unit, and Jornada Basin Long Term Ecological Research Program, in Las Cruces, New Mexico. Luis L. Rodriguez and Steven J. Pauszek are with the US Department of Agriculture, Agricultural Research Service, Plum Island Animal Disease Center, in Orient Point, New York. D. Scott McVey is affiliated with the US Department of Agriculture, Agricultural Research Service, Center for Grain and Animal Health Research, Arthropod-Borne Animal Diseases Research Unit, in Manhattan, Kansas. Angela M. Pelzel-McCluskey and T. Jason Lombard are with the US Department of Agriculture, Animal and Plant Health Inspection Service, Veterinary Services, in Fort Collins, Colorado. Justin D. Derner is with the US Department of Agriculture, Agricultural Research Service, Rangeland Resources and Systems Research Unit, in Cheyenne, Wyoming. Steven R. Archer is affiliated with the School of Natural Resources and the Environment at the University of Arizona, in Tucson. Colby W. Brungard and Niall P. Hanan are affiliated with the Department of Plant and Environmental Sciences, Jornada Basin Long Term Ecological Research Program, New Mexico State University, in Las Cruces. Jerry L. Hatfield is with the US Department of Agriculture, Agricultural Research Service, National Laboratory for Agriculture and the Environment, in Ames, Iowa. Gregory S. Okin is with the Department of Geography at the University of California, Los Angeles. Osvaldo E. Sala is affiliated with the School of Life Sciences at Arizona State University, in Tempe. Enrique R. Vivoni is affiliated with the School of Earth and Space Exploration and the School of Sustainable Engineering and the Built Environment at Arizona State University, in Tempe.