Geomorphic trajectory and landform analysis using graph theory: a panel data approach to quantitative geomorphology

Abstract

Comparing successive datasets of GIS polygons derived from remote sensing data is a common approach to quantify morphological change. GIS-derived datasets capture instantaneous observations or "snapshots" of the *state* of a system at a given time but do not explicitly capture the temporal sequences needed to characterize system *processes*. Comparisons between these "temporally-naive" datasets can be used to infer properties and trends of the landscape as a whole, but tracking changes in the characteristics of individual landforms (e.g., sandbars, dunes, or other surface features of interest) across snapshots is labor-intensive and infeasible for large or irregular datasets. Using traditional computer-based procedural methods to compare sequences of datasets without knowledge of temporal trajectories introduces several challenges and data artifacts that complicate analysis. We propose a graph-theory approach for processing sequential spatial data to automatically identify and track distinct groups of related landforms or "geomorphic units" across fully- or partially-overlapping snapshots. This approach allows tracking even in cases where landforms fragment, merge, migrate, or become temporarily obstructed from view. The method promotes new panel data analysis opportunities and overcomes three critical limitations of traditional procedural methods of assessing landscape change from spatial data: *(1)* it can generate landscape metrics based on geomorphic units, rather than the arbitrary geographic units of the underlying spatial datasets, *(2)* it distinguishes missing or obstructed observations from changes in the characterization of landforms due to environmental conditions, and *(3)* it automatically generates panel datasets and discriminates between within-landform change and across-landform variation. The panel datasets can be used to upscale feature-level information to system-level metrics and analysis. Furthermore, a graph-theory approach can yield insight on geomorphic change through analysis of the graph structure, and offers a promising approach for geomorphological analyses which retain information on the spatial configuration of geomorphic units. We demonstrate the method with examples from emergent sandbars on the Missouri River.

Keywords

geomorphology, spatial data, panel data, graph theory, GEOBIA

Introduction

Geomorphologists commonly draw polygons to map landforms from remote-sensing imagery. Comparing repeated surveys or collections of "snapshots" of a landscape creates opportunities for quantitative temporal analysis of morphological change. However, independent snapshots represent the state of the system at a single time without any information about the temporal continuity or discontinuity between individual landforms at different times. A collection of snapshots is a *repeated cross-sectional study* of a landscape, in that it captures the characteristics of individual landforms at different points in time but does not provide on its own any information on how characteristics of a given landform differ between snapshots, or even if the same landforms are captured in any given pair of images. Repeated cross-sectional studies can generate landscape-scale metrics (e.g., total number of landforms of each type, total area and spatial density of each type, etc.) and the temporal evolution of these metrics, but cannot identify how this evolution emerges from the trajectories of individual landforms because they do not explicitly track these trajectories. Cross-sectional analyses almost always lack the statistical power to differentiate causal relationships from simple associations (Hilton and

Patrick 1969; Hsiao 1985; Duncan and Kalton 1987; Trivellato 1999). Furthermore, inferences about a population are rarely applicable to individuals or groups within a population (Robinson 1950; Simpson 1951; Selvin 1958; Pearl 2009).

In spatial studies, direct comparison of snapshots is meaningful only when all snapshots have the same spatial extent. Generating landscape metrics from snapshots that inconsistently represent system state (such as when successive snapshots only partially overlap) or contain missing or obstructed observations (due to environmental conditions such as cloud cover) risks confusing actual change in landscape structure with statistical artifacts (Heckman 1977; Reddy and Dvalos 2003; Gelfand et al. 2010). Maintaining consistency in landscape metrics across snapshots may therefore exclude large regions from the entire collection. Furthermore, metrics derived from direct comparisons of snapshots can be misleading when they are derived from landform delineations made under the specific conditions of each snapshot, and do not incorporate information on landform history. These metrics may misrepresent landscape characteristics when prevailing conditions such as season (e.g., presence or absence of vegetation) or hydrology (e.g., temporary inundation) strongly influence the identification and characterization of landforms, or when landscapes are highly dynamic—places where landforms fragment, merge, or migrate within the system (Figure 1).

Figure 1. Satellite imagery of sandbars in the Missouri River. Sandbars fragment, merge and migrate in response to geomorphic processes, but also become partially- or fully-inundated as river flow rises. Drawn polygons may not adequately represent the true boundaries of the landforms, and directly comparing metrics derived from these polygons may not capture the nature of geomorphic processes and interactions between individual landforms. *Image source:* Google Earth. $42°47'$N and $97°08'$W. Accessed January 22, 2018.

Many geomorphic studies rely on expert analysis to identify landforms and track characteristics of individual geomorphic units over time. The competent geomorphologist can intuitively recognize and distinguish between geomorphically-relevant features across snapshots. Explicitly establishing temporal linkages between individual landforms across snapshots transforms repeated cross-sectional data into a *panel* dataset. Panel studies are more powerful than repeated cross-sectional studies because they track changes in characteristics of individuals as well as the overall population over time. They provide stronger statistical power for inferring causal relationships by understanding how population trends emerge from the trajectories of these individuals (Shadish et al. 2001; Frees 2004). However, visually evaluating each landform and manually identifying trajectories is not feasible for large datasets containing hundreds of features. Computer algorithms that use objective criteria to connect geomorphically-relevant features across snapshots are needed to analyze geomorphic change at these larger scales. However, to our knowledge no objective methods for programmatically establishing individual trajectories across large datasets have been described in the geomorphological literature.

We present a robust, automated method for processing extensive geomorphic datasets using graph theory which operates on top of an existing object classification scheme to identify and track geomorphic units—defined as distinct groups of related landforms—across fully- or partially-overlapping snapshots. This method automatically links observations of individual landforms across multiple snapshots, even in cases where landforms fragment, merge, migrate, or become temporarily obstructed from view. The method is applicable to a wide variety of spatial datasets and can distinguish actual landform change from apparent change, where landform size and shape are confounded by variability in observation conditions. The method addresses challenges encountered when managing disparate spatial datasets for landform evolution studies.

This paper provides three incremental contributions to geomorphology and understanding of landform evolution: *(1)* it establishes a robust, objective and repeatable framework for identifying and linking geographic units across a series of snapshots into geomorphic units in order to generate landscape metrics; *(2)* it supports distinguishes missing or obstructed observations from changes in the characterization of landforms due to environmental conditions; and *(3)* it automatically generates panel datasets that capture trajectories of individual landforms and discriminates between within-landform change and across-landform variation. Using this method to generate temporally-explicit data structures also creates opportunities

to leverage graph theory for additional morphological analyses. We demonstrate the method with examples from emergent sandbars in the Missouri River.

Background

As spatial landform data have become more abundant, change-detection methods have grown more common in geomorphological research (James et al. 2012; Napieralski et al. 2013). While many studies have focused on gridded models (e.g., Brasington et al. 2003; Martınez-Casasnovas et al. 2004; White 2006; Wheaton et al. 2010), Geographic Object-Based Image Analysis (GEOBIA) (Lang 2008; Blaschke 2010) also emerged as a powerful tool for landform evolution studies (Drăguţ and Blaschke 2006; Shruthi et al. 2011; Phinn et al. 2012). Despite increasing data availability and maturation of analytical methods, most studies of landscape change are scale-limited, typically falling into two categories: detailed investigation of small study areas generated from fine-scale data such as Lidar, aerial photographs or sub-meter resolution satellite imagery (e.g., Woolard and Colby 2002; Feurer et al. 2008; Heritage et al. 2009; Mandlburger et al. 2015) or studies of landscape trends generated from coarse-scale data such as RADAR or meter/decameter resolution satellite imagery (e.g., Farr and Chadwick 1996; Bishop et al. 2002, 2003). Few geomorphological studies connect fine-scale landform evolution to large-scale landscape change. The landform-landscape scale analyses are usually data-limited, since coverage, completeness and resolution of spatial data are inversely related to the size of a system. For the most part, researchers relying on remotely-sensed data are often forced to choose between fine-resolution satellite imagery with reduced spatial and temporal coverage and coarser-resolution imagery collected at a higher frequency and greater extent.

Studies of dynamic landscape evolution from landform-scale data are particularly constrained by these issues, as they require data with both fine spatial resolution (in order to effectively delineate individual landforms and detect change) and high temporal frequency (in order to adequately capture landform evolution trajectories). Furthermore, managing and analyzing large spatial datasets pose intrinsic challenges in terms of data storage requirements, image processing and normalization to account for atmospheric conditions, and quality assurance of object classification algorithms or manually-digitized landforms (Goodchild and Haining 2004; Renschler et al. 2007; Bishop et al. 2012).

Mid-channel emergent sandbars, common features in sandbed rivers, are an example of a dynamic landform with important landscape-scale implications. On the Missouri River, emergent sandbars provide critical nesting and foraging habitat for the endangered interior population of least tern (*Sternula antillarum*) and the threatened Northern Great Plains population of piping plover (*Charadrius melodus*) (Buenau et al. 2014). Hydrologic variability masks sandbar building, erosion, and migration as river stage variation inundates and, sometimes, divides sandbar features. Superposition of repeated satellite imagery captures, but does not distinguish between, the two types of sandbar change: morphodynamic change (i.e., the erosion, aggradation, migration and true fragmentation of sandbars) and hydrodynamic change (e.g., temporary fragmentation or complete inundation of sandbars during high flows) (Jacobson 2013). Both morphodynamic and hydrodynamic change are important to understanding emergent sandbar habitat (ESH) dynamics, but analysis of long-term landscape trends in ESH requires an approach that can parse these two processes. Separating these effects is non-trivial and requires fine-resolution spatial data with a high observation frequency.

Graph theory is a branch of mathematics concerned with relational data, i.e., groups of discrete concepts or objects (people, communities, buildings, etc.) that interact. Individual objects are represented as "nodes" in a graph, and a relation or interaction between two objects as an "edge" connecting the two objects (nodes). Graphs can model physical connectivity, hierarchies and feedback loops in a wide variety of processes and structures. The mathematics of graph theory naturally support analyses that connect broader system behavior to interactions and behaviors of individual elements, i.e., analyses that aim to 'keep the "whole" in mind while studying the "parts" and vice versa' (Jordán and Scheuring 2004). Graph theory has been widely applied to a variety of problems in computer science, transportation engineering, geography, hydrology, landscape ecology, and the biological and medical sciences. Applications of graph theory in geomorphology are comparatively limited, but a few researchers have applied graph-based approaches to morphological change in natural systems (Perret et al. 1999; Valentini et al. 2007; Pardo-Iguzquiza et al. 2011), landscape structure and connectivity (Rodríguez-Iturbe and Rinaldo 2001; Werner 1993, 1994; Gascuel-Odoux et al. 2011; Aurousseau et al. 2009), and process linkages or cascades (Phillips 2013; Heckmann and Schwanghart 2013; Heckmann et al. 2014). Heckmann et al. (2015) compiled an excellent review of graph-theory approaches in geomorphology, and Phillips et al. (2015) provide a more general review of graph theory in the geosciences.
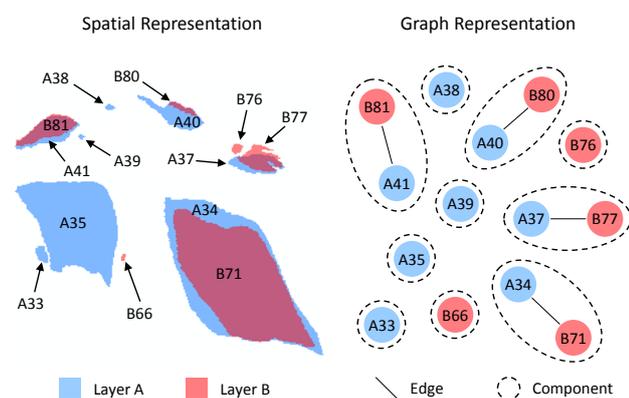
Prepared using **sagej.cls**

Figure 2. *Left:* A spatial union of two layers A and B representing classified sandbars at two different times, with individual features labeled. *Right:* Interpreting the union attribute table AB as an edge list yields a graph representation of feature intersections. The graph components are identified by dashed lines enclosing groups of nodes.

Methodology

The method begins with a collection of N uniquely-named data *layers* representing discrete *features* in some spatial domain D, such as a collection of snapshots in a single geodatabase derived from repeated remotely-sensed images. Depending on the nature of the layers, features may represent physical landforms (e.g., islands, lakes) or contours of continuous data (e.g., zones of high chlorophyll concentration in an aquatic system). A layer ℓ in the collection spans a spatial domain D_ℓ that is equivalent to or contained within the spatial domain of the entire collection ($D_\ell \subseteq D$). Each layer is linked to an attribute table which includes a unique identifier (OBJECTID or FID in GIS software) for each feature in the layer and its bulk properties (e.g., feature area, perimeter, and classification). The combination of the unique layer name and the unique FID of each feature contained within the layer provides a unique identifier for every feature in the entire collection.

Consider two layers in the collection ($A, B \in N$), with spatial domains D_A and D_B, which represent the state of the system at time t_A and a later time t_B, respectively. Further assume that the two layers partially-overlap ($D_A \cap D_B \neq \emptyset$). The spatial union of these two layers ($AB = A \cup B$), shown on the left in Figure 2, generates a collection of features $a_1, a_2, \ldots, a_n$ from A and $b_1, b_2, \ldots, b_m$ from B in the region $D_{AB} = D_A \cup D_B$, where features from A may or may not intersect with multiple features from B and vice versa.

The attribute table of the union layer AB (see supplemental materials) tracks the spatial overlap (or lack thereof) between features in A and features in B. The union table represents a one-to-one spatial relationship between

two features a_i and b_j with a single row in the table, and a many-to-one relationship between features a_j, a_k and b_i with multiple rows. A missing value in either column represents the case where a feature (or a portion of a feature) from one layer did not overlap with a feature in the other layer.

The two-column union attribute table becomes the *edge list* of a graph $G(A, B)$. The unique identifiers in each column (which identify individual features in layers A and B) define *nodes* of the graph, and rows in the table (which identify spatial overlaps between features in A and features in B) are *edges* that connect nodes. A graph *component* is a set of nodes that are connected to each other but not to any other nodes in the graph. The *membership* of a node (a numeric index) identifies which component it belongs to. By definition, a node can only be a member of one component. Two arbitrary features $A_i \in A$ and $B_j \in B$ are members of the same component k if they spatially intersect. The graph $G(A, B)$ corresponding to the union layer AB is shown on the right in Figure 2. Dashed lines enclosing connected nodes indicate the graph components.

The components of $G(A, B)$ embody links between features captured by the spatial union. Because the two layers represent the same system at two different times, the spatial union also infers *temporal* links between features. Feature a_i in A and feature b_j in B have the same membership ($a_i, b_j \in k$) if they are spatially and temporally related (they occupy a similar region of space over the period $t_A \to t_B$). The method populates the original attribute tables of layers A and B with the memberships and combines the attribute tables to produce time series of individual components.

Defining temporal links based on spatial unions requires that two successive observations of a feature at least partially overlap. Therefore, the frequency of observation must be sufficiently high relative to the speed at which a feature migrates. This limitation can potentially be relaxed by using alternative spatial relationships, such as buffered overlaps or minimum distance criteria. Similarly, instances where unrelated migrating features cross paths or occupy the same region at different periods of time can result in erroneous temporal linkages. These cases can be addressed through additional post-processing of the graph and are discussed later.

A third layer can be added to the graph following the same procedure. Assume there is a layer C representing the system state at time $t_C > t_B$ that partially overlaps both layers A and B ($D_A \cap D_B \cap D_C \neq \emptyset$). A spatial union of layer C and the existing union layer AB yields a new union layer $ABC = AB \cup C \equiv A \cup B \cup C$ and a three-column union attribute table. Each pairwise combination of columns in

the union table are edge lists of a graph that links features across layers A, B, and C. New information from layer C is contained in the column pairs (A, C) and (B, C). Appending this new information to the graph $G(A, B)$ potentially revises the component definitions as new linkages are exposed in the supergraph $G(A, B, C) \supset G(A, B)$. The time series for a particular component k may proceed from $t_A \rightarrow t_B, t_A \rightarrow t_C, t_B \rightarrow t_C$, or $t_A \rightarrow t_B \rightarrow t_C$, depending on whether the component contains features from layers A, B, and/or C. The order in which layers are incorporated into the graph does not affect the definition of the graph components. Features in one layer that are not linked to features in other layers are included in the graph as isolated nodes, and as single observations of a unique component in the extracted time series.

The constructed graph is an intermediate data structure that converts a series of system-wide observations (snapshots) to a set of time series of individual features. Our method has some similarities to work by Thibaud et al. (2013) and earlier theoretical development by Del Mondo et al. (2010), but does not require specialized software or database structures, instead using the constructed graph to map the spatio-temporal data to tabular panel data. Phillips et al. (2015) describe a taxonomy for classifying graph-theoretic applications in the geosciences; our method is codified within this framework as a spatially-explicit network analysis of historical contigency, representing distinct spatial features as individual nodes and temporal progression as edges.

The method can be applied to any system where features of interest can be reliably delineated and represented as distinct polygons, and can support virtually any spatial interaction rule (including more complex spatial interactions based on buffers, border sharing, etc.) to identify feature linkages. The method does not alter the underlying landform classification scheme, and therefore cannot resolve deficiencies in the classification or delineation of features; however, the membership information could potentially be incorporated into a post-processing workflow to help identify classification problems. The method relies on simple, automated graph manipulations available in a wide variety of software packages (e.g., the "network" package for R (Butts 2008), the "networkX" module for Python (Schult 2008), the cross-language "igraph" library (Csardi and Nepusz 2006), and others; example R code for recreating the above example is included in the supplemental materials). The method identifies spatio-temporal linkages between discrete observations in a system and can, therefore, upscale feature-scale information to system-scale metrics and analysis.

Example: application to GEOBIA of emergent sandbars

We illustrate the method with an application to emergent sandbars in the Missouri River. The dataset is a collection of spatial layers containing GEOBIA-classified landforms derived from high-resolution satellite imagery. The dataset was developed to map least tern nesting habitat (Strong 2012). Figure 3 shows five snapshots of a section of the Missouri River shortly downstream of Gavins Point Dam (1960 River Mile 808.8 to 794.2). The snapshots are ordered by date and capture flow conditions ranging from 270 m^3 s^{-1} (flow-duration (FD) percentile $> 99\%$) to 1,340 m^3 s^{-1} (FD $\sim 10\%$) (USACE 2013). The GEOBIA delineation and classification of landforms is highly flow-dependent. At the lowest flow (Figure 3B, October 2005) most sandbars are attached to the banks, and one large landform dominates the area. Higher flows fragment the sandbar complexes, disconnect them from the banks, or inundate sandbars completely. The delineated geographic units (polygons) in a given snapshot may in reality represent either intact, contiguous landforms or fragments of a larger landform. However, The GEOBIA classification operates on each snapshot independently and classifies each geographic unit as a separate landform.

Landscape patterns

The graph approach automatically identifies spatio-temporal links between geographic units. Each distinct group of geographic units, represented as a component of the graph, defines a geomorphic unit. We used the procedure outlined in §Methodology to automatically identify geomorphic units based on a simple spatial overlap rule that excluded landforms not classified as either bank-attached (green) or mid-channel (yellow) sandbars (e.g., water, farmland and built structures). The resulting geomorphic units, which we refer to as "sandbar complexes", consist of groups of bank-attached and mid-channel sandbars that are spatially and temporally related. The method extracted relevant attribute data from each layer and linked it to the sandbar complexes.

Habitat patch counts provide important context to habitat area data. The relationship between river flow and sandbar fragmentation is particularly important for least tern reproductive success, because changes in river stage can inundate nests and strand chicks (Smith and Renken 1993; Sidle et al. 1992) as well as modify predation and disturbance risk by changing connectivity between sandbars and channel banks (Schwalbach 1998; Kirsch 1996). As flow increases, some sandbars will fragment while others
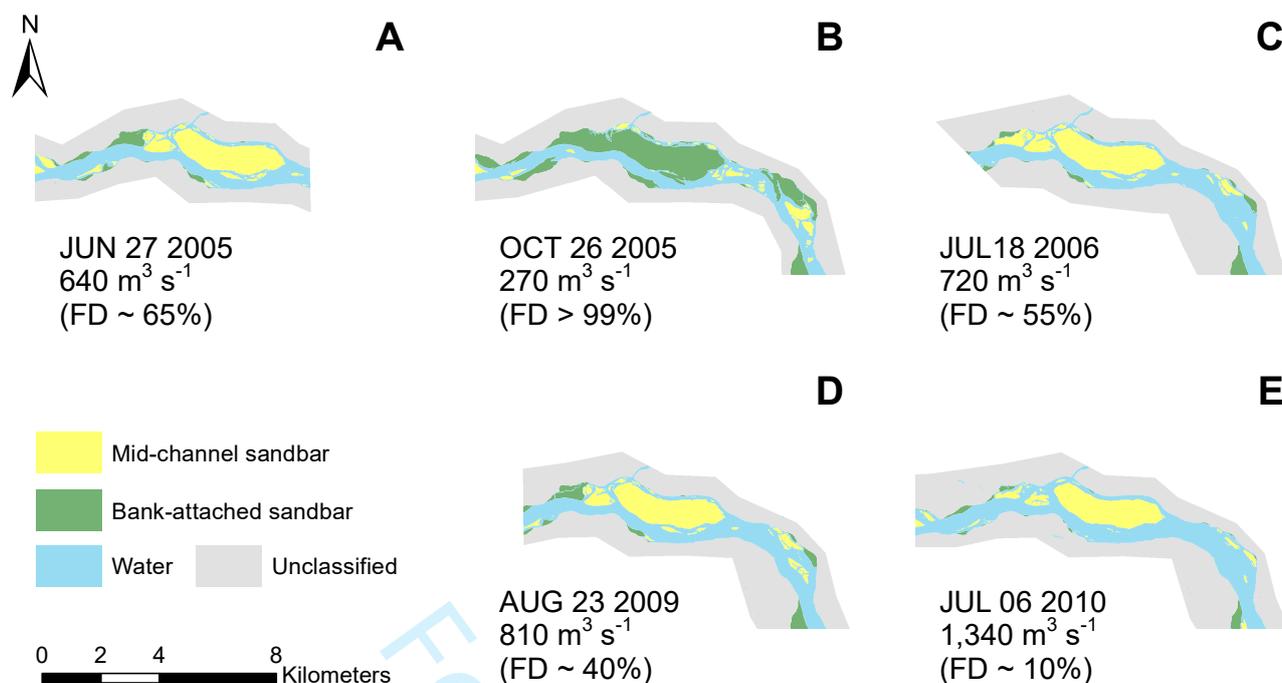
Figure 3. Five snapshots of a section of the Missouri River (1960 River Mile 808.8 to 794.2) spanning June 2005 through July 2010 and flows between 270 and 1,340 cubic meters per second. The flow-duration (FD) percentiles were computed from Gavins Point Dam release records (USACE 2013). Landforms are colored according to a GEOBIA classification scheme.

will become completely inundated. Continued increase in flow will eventually result in total inundation of all sandbars and sandbar fragments, but the degree of fragmentation and inundation varies between sandbars and across the flow regime.

Automatically grouping geographic units (polygons) into geomorphic units (sandbar complexes) makes the analysis metrics more meaningful. Figure 4 compares the geographic unit count (from the traditional, temporally-naive method) to the count of geomorphic units (new method) across the range of flows. The naive method, which counts geographic units without geomorphic context, cannot differentiate between fragmentation and inundation of sandbars. Therefore, it reports *more* sandbars as flow increases and inundates these features (because partial inundation fragments the geomorphic units into multiple polygons) and shows significant variation in the count data at higher flows (due to the co-occurrence of fragmentation and inundation). The naive method cannot differentiate between changes within landforms (fragmentation) and changes across the landscape (inundation) because it does not account for the spatio-temporal links between geographic units, and this important spatial aspect is lost when landscape descriptors are generated from geographic units. Grouping sandbar fragments by their underlying geomorphic unit captures the total inundation of these geomorphic units. The new method generates a reasonable, monotonically decreasing
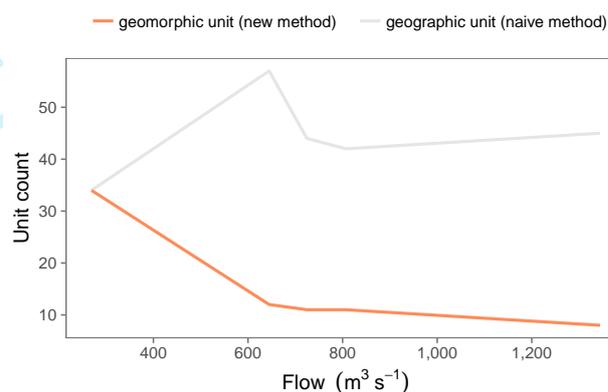

Figure 4. Counts of geographic units compared to geomorphic units across the range of flows. At the lowest flow, the counts are identical. Unlike the new method, the naive method cannot distinguish between fragmentation and inundation of geomorphic units.

trend, where the unit count drops consistently as higher flows inundate sandbar complexes.

The new method retains all polygon information. Changes in landscape configuration and the distribution of size, abundance, and classification of geographic units can be explored within the context of the underlying geomorphic units. Unlike the naive method, the new method identifies exactly *which* units fragment or inundate, and to what extent. Figure 5 (top) compares the distribution of geographic-unit areas (naive method) to the distribution of geomorphic-unit

areas (new method) across the range of flows. The (log-transformed) distribution of geographic-unit area exhibits a strong positive skew at higher flows compared to the distribution of geomorphic-unit area, primarily because fragmentation of geomorphic units is not accounted for. In contrast, the distribution of geomorphic-unit area is largely (log)normal and actually develops a weak negative skew at higher flows. This is because smaller geomorphic units tend to become completely inundated as flows increase while larger units persist but become fragmented. Figure 5 (bottom) shows that a small minority of geomorphic units account for the majority of fragmentation in the study area. Temporally-naive methods incorrectly consider these fragments as individual observations; in reality, multiple geographic units correspond to the same geomorphic unit. Statistical models based on geographic units may have erroneously high measures of significance and statistical power because the assumption that geographic units are independently and identically distributed (*i.i.d.*) is incorrect. Models based on geomorphic units will generally be more realistic despite the apparent loss of statistical power (since grouping geographic units by geomorphic unit effectively reduces the sample size).

Missing data

Achieving complete coverage of a large study area often requires multiple adjacent images. Collections of remotely-sensed imagery often have different extents and include portions of images that are obstructed by clouds, glare or other image artifacts, as shown in Figure 6. Generating landscape metrics (such as described in §Landscape patterns) under these conditions can be challenging with traditional methods. Researchers commonly either *(a)* report normalized landscape metrics to account for differing extents and extrapolate findings to regions outside the area of overlap, or *(b)* splice multiple images together *ad hoc* to generate datasets with comparable extent. The latter option can be prohibitively time-consuming for large datasets, and is generally unsuitable for dynamic landscapes because adjacent images may not provide a consistent representation of the system state due to changing environmental conditions.

The new method overcomes these issues by linking information between individual geomorphic units to generate datasets that contain the maximum set of available data for each unit. Images with differing extents are handled automatically as discussed in §Methodology. When two images have different extents, the spatial union will by definition contain regions where there is no spatial
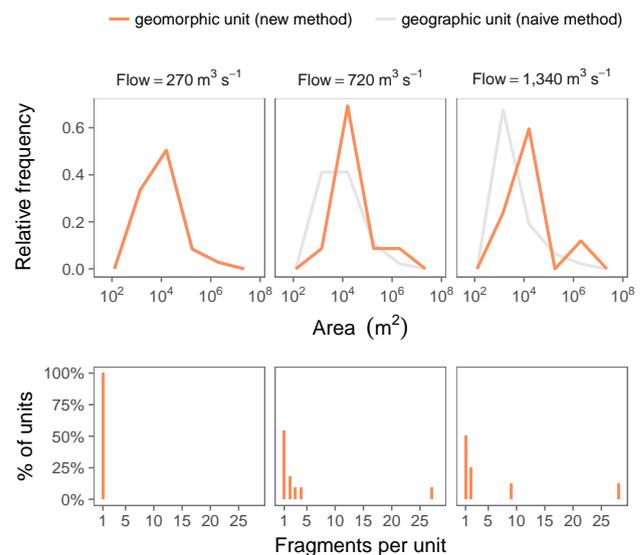

Figure 5. *Top:* Relative frequency of geographic-unit areas compared to geomorphic unit-areas. At the lowest flow, the distributions are identical. The distribution of geographic-unit area develops a strong positive skew as sandbars fragment at higher flows, but the distribution of geomorphic-unit area actually develops a weak negative skew as flow increases. This is because smaller geomorphic units tend to become completely inundated as flow increases, while larger geomorphic units persist as multiple fragments. *Bottom:* Geomorphic-unit fragmentation across the range of flows. At the lowest flow, all geomorphic units are single landforms. A minority of geomorphic units become highly fragmented and are responsible for the majority of the observed increase in geographic units. Most geomorphic features in the study area do not fragment at all prior to becoming inundated, or fragment into only 2-3 geographic units.

overlap between the images. The method also generates a *geomorphic footprint* for each geomorphic unit, a polygon that defines the coverage of each geomorphic unit over the range of snapshots. The geomorphic footprint is used as a post-processor to identify cases where a sandbar complex is obstructed (e.g., by detecting whether it shares a border with classified cloud objects) or truncated at the edge of an image (e.g., by testing if the geomorphic footprint extends past the boundary of a snapshot). These occurrences are flagged in the data and the corresponding nodes are removed from the graph. Figure 6 illustrates how geomorphic footprints are used to handle different types of missing data.

The geomorphic footprint can also be used to detect a third type of missing observation: the case where a geomorphic unit is completely inundated or has otherwise evolved beyond the object classification scheme. In the case where a geomorphic footprint is unobstructed and fully contained by a snapshot, but no geographic units are present within the footprint (i.e., there are no delineated polygons in a region in the snapshot, but one or more polygons are present in the same region in other snapshots), the method
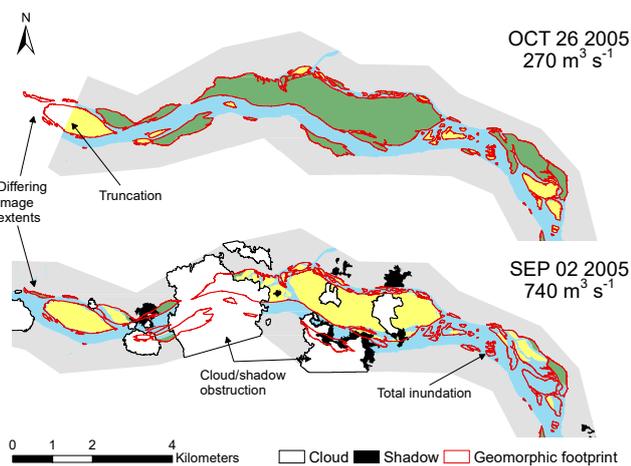
Figure 6. Common causes of missing spatial data. The geomorphic footprints (red outline) are constructed from all available images and used to identify cases where geomorphic units are truncated at the edges of images, obscured by clouds and cloud shadow, or are completely inundated.

Figure 7. *Left:* Aggregate plot of relationship between landform area and flow. Aggregated data can document broader landscape trends, but information on variation between landforms is lost. *Right:* The individual geomorphic unit trajectories. Arrows indicate the trajectory of individual units with respect to flow. When the trajectories are known, broader landscape trends can be decomposed into variation between individual geomorphic units and system-wide trends.

can be used to add a "virtual observation" to the dataset to maintain continuity of information. Virtual observations are generated by performing an intersection between a layer and the geomorphic footprint. A footprint that is completely contained and unobstructed in the intersection, but does not contain any features, indicates a virtual observation. Virtual observations can be useful for analyzing landform change or directly modeling inundation likelihood. For example, virtual observations can be assigned zero-value areas and appended to the panel dataset to represent sandbars that are completely inundated.

The simplest way to define the geomorphic footprint is to use all available layers, i.e., the geomorphic footprint will be the combined extent of all geographic units from all layers that share the same membership. However, the geomorphic footprint can be defined using other criteria, and can be adaptive. For example, the geomorphic footprint for an emergent sandbar might be defined as the feature boundary identified at the next lowest discharge, and not require that the entire footprint of the sandbar at the lowest observed discharge be captured by every snapshot. For features that migrate, the geomorphic footprint might be defined as the minimum bounding area of the initial and final location of the feature, ensuring that every snapshot captures the path of migration without any obstructions. The most appropriate strategy for defining the geomorphic footprint will depend on the nature of the landforms and processes being investigated.

The data generated with the new method include all available observations for each geomorphic unit. For the snapshots shown in Figure 3, less than half of the identified geomorphic units are captured in all five snapshots. Coincident observations of geomorphic units can
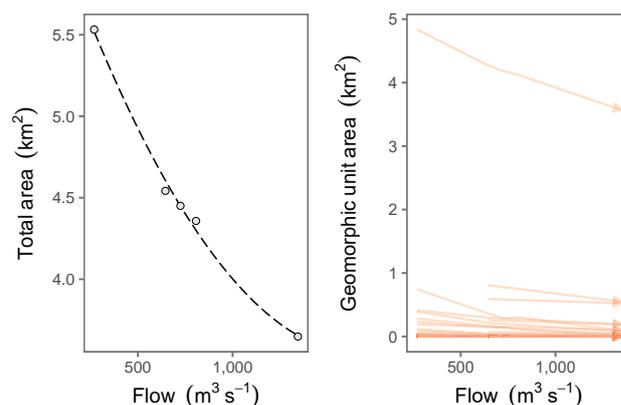
be easily extracted from the data to generate consistent landscape metrics across images, for example by filtering out geomorphic units that were not captured by all images in the collection. However, doing so can result in considerable loss of data (as occurs in traditional naive analyses). In this case, 54% of geomorphic units would need to be excluded in order to derive consistent landscape metrics for all five snapshots. Over 13% of geomorphic units were captured by three or fewer snapshots. The next section describes an alternative approach to analyzing landscape structure that incorporates all available data.

Geomorphic unit modeling

The new method generates datasets that can be directly incorporated into *panel data models*. A panel data model is a class of statistical model where repeated observations of one or more variables are grouped within a (possibly multilevel) hierarchy (Fitzmaurice et al. 2012). Panel studies generally have more explanatory power than cross-sectional studies because they are able to control for time-invariant differences between individual entities and explicitly incorporate the temporal sequence of events (Menard 2002; Ployhart and Vandenberg 2010). Furthermore, a number of strategies exist for imputing missing observations in panel datasets, which can improve the accuracy of parameter estimates (Enders 2010; van Buuren 2012; Ibrahim and Molenberghs 2017).

Automatically "connecting the dots" between observations to specify trajectories of individual geomorphic units supports a broader range of analyses. Without information on trajectories, a practical analysis of geographic unit data

would focus on documenting broader landscape trends, e.g., by totaling the sandbar area for each discharge. For instance, the change in total area relative to the area observed at the minimum flow could regressed against "excess" flow (i.e., the increase in flow above the minimum observed flow), as shown in Figure 7 (left). When the trajectories of individual landforms are known, these broader landscape trends can be further decomposed into variation between individual geomorphic units and system-wide trends. Figure 7 (right) shows the trajectories of individual geomorphic units.

Visual comparison of the aggregate trend to the individual trajectories reveals two important drawbacks of the traditional aggregate analysis. First, the overall trend is driven almost entirely by the largest geomorphic unit. The combined initial area of the remaining features is less than one square kilometer, compared to the largest geomorphic unit's area of almost five square kilometers. Changes in smaller geomorphic units are overshadowed by larger units as a result of aggregation. While this may be qualitatively apparent from visual inspection of the snapshots, quantifying the influence of each geomorphic unit on the aggregate trend is only possible when their trajectories are explicitly known. Second, the aggregate analysis requires that the region of analysis be restricted to the area of overlap for all snapshots. As a result, a significant number of geomorphic units must be excluded from the analysis of the overall trend.

Knowledge of individual trajectories can be used to improve aggregate analysis by, for example, informing binning or grouping strategies (e.g., by identifying geomorphic units with similar properties or trajectories) or estimating missing observations from the individual trajectory of a geomorphic unit in order to expand the region of analysis. The individual trajectories can also be used to guide sensitivity analyses of the aggregate trend and quantify the influence of each geomorphic unit on the model fit, such as by applying leave-one-out analyses where trajectories of individual geomorphic units are excluded and the aggregate trend is recomputed (Burt et al. 2009). Alternatively, deeper insights into system behavior can be gained by explicitly representing the trajectories of individual geomorphic units in system-level models. However, explicitly modeling individual geomorphic units can introduce additional challenges, and such models may have reduced reliability and statistical power relative to an aggregate model if the observation sample size of each geomorphic unit is small (Carroll and Pearson 1998; Kadmon et al. 2003; Hernandez et al. 2006).

Explicit knowledge of trajectories also allows researchers to conduct analyses that are difficult to develop from aggregate data. Figure 8 shows two analyses based on
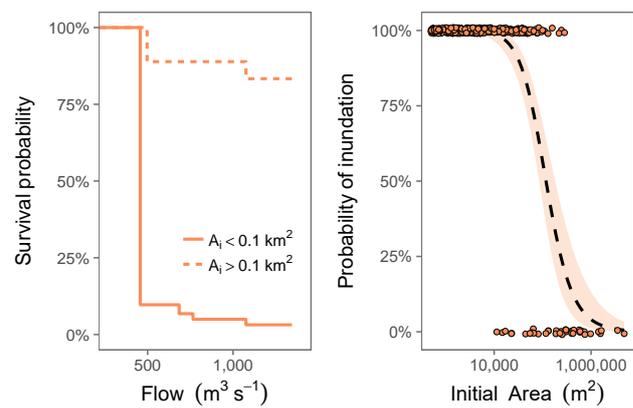

Figure 8. *Left:* Empirical survival plot of geomorphic units, broken into two roughly-equal groups based on initial area (i.e., area observed at the lowest flow). *Right:* Logistic regression of probability of inundation as a function of initial area.

sandbar inundation. The left panel shows empirical survival curves for geomorphic units, split into two roughly-equal groups based on their initial area (i.e., area observed at 270 m^3 s^{-1}). In this example, survival refers to complete inundation of a sandbar in response to increasing discharge. The right panel shows a different approach that models probability of inundation as a logistic function of initial area. Both of these analyses rely on geomorphic unit trajectories to capture the inundation of individual geomorphic units. A comparable analysis is not easily formulated for aggregate data or geographic unit data, and would likely suffer from the issues discussed in the previous sections.

Inferring geomorphic structure from graph structure

Constructing a panel dataset using our methodology requires only the most basic premises of graph theory—namely, the node-edge representation of data and the concept of graph components. The previous sections demonstrate how, even in its basic form, a graph-theory approach can mitigate several common issues with landform analysis. However, once the temporally-explicit data structure is generated, higher-level graph theory can yield insight of geomorphic change through analysis of the component graph structure. These techniques offer a promising avenue for geomorphic analyses which retain information on the spatial configuration of geomorphic units.

The spatial union governs the structure of the graph. When features are relatively persistent, it is useful to initially graph the union of all available layers, generating a graph with edges representing spatial relationships between features in all pairwise combinations of overlapping layers. For example, if the spatial layers represent features within a
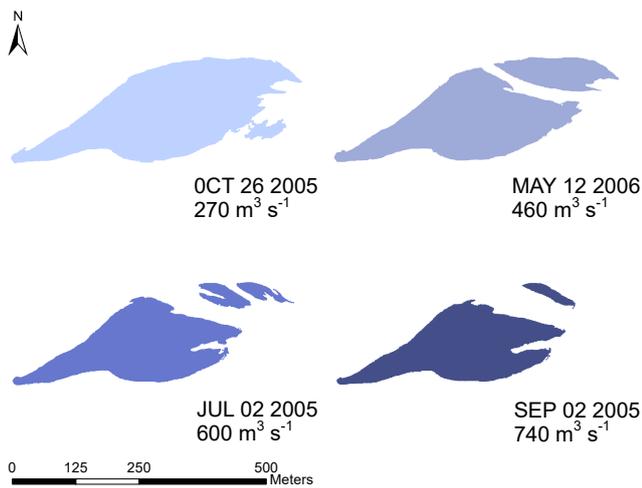
Figure 9. A single sandbar component observed at four distinct flows.

system at different times, the graph of the union of all layers will capture feature evolution across multiple time scales, i.e., it will be a "multi-scale graph" of the data set. Appropriately-constructed two-dimensional GIS data typically preclude overlap between features in a single layer; therefore, multi-scale graphs will not contain edges between features from the same layer. When the order of the unions has meaning (for example, if the layers are time-ordered) a directed graph can be constructed where edge directions indicate the transition from one layer to the next. A graph produced in this way will have no mutual edges (e.g., time proceeds in only one direction) and no cycles, i.e., the graph is a *directed acyclic graph* (DAG). DAGs have a number of useful properties (Thulasiraman and Swamy 1992) that can be leveraged for geomorphological analysis.

Consider the example of a single sandbar complex observed at four different times in Figure 9. The multi-scale graph representing the sandbar complex generated from the ordered sequence of images A, B, C, and D (where the layers could be ordered in chronological sequence, or from high flow to low flow, etc.) contains edges from nodes in A to nodes in B, B to C, and C to D as well as edges from A to C, A to D, and B to C. If edges outside of the sequence $A \rightarrow B \rightarrow C \rightarrow D$ are removed, the resulting graph represents the progression of change, referred to here as a *progression graph*. The multi-scale graph expresses the presence or absence of a path between any two nodes in the progression graph, i.e., it is the *reachability graph* of the progression graph. A variety of methods can extract a progression graph from a reachability graph, such as longest path algorithms (which are relatively fast for DAGs).

The edge directions of a reachability graph are defined entirely by the union order, whereas the *non-directional*

structure (i.e., the presence/absence of nodes or edges) is not. Regardless of the ordering used for unions, the non-directional structure of the reachability graphs will be the same for all progression graphs (they will be *isomorphic*). Therefore, it is not necessary to repeat spatial union operations with different orderings (which could be computationally intensive) in order to produce different progression graphs. Instead, the edge directions of a reachability graph can be manipulated directly to produce different progression graphs.

Figure 10 shows the relationship between the non-directional reachability graph of the sandbar complex in Figure 9 and two progression graphs, one ordered by time and the other ordered by flow. Both progression graphs have the same reachability graph when edge direction is ignored. If the sandbar does not erode or build over time, the geomorphic structure of the sandbar at any given time will be a function of flow only. The flow-progression graph will be a "true tree" (i.e., the graph will contain a single "root node" and every node will have at most one incoming edge) like the flow-progression graph in Figure 10. Conversely, the time-progression graph of the same sandbar complex will not be a true tree, unless by coincidence river flow increases monotonically over the time interval of the image collection.
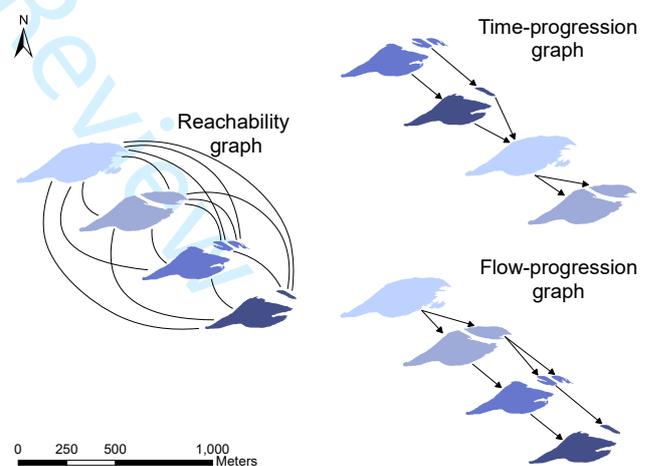

Figure 10. Graph representations of the sandbar complex in Figure 9. *Left:* The reachability graph, with edge directions omitted. *Top right:* the time-progression graph. *Bottom right:* the flow-progression graph.

Identifying cases where morphodynamic change has occurred but is confounded by hydrodynamic change can be challenging. §Geomorphic unit modeling described statistical approaches to modeling the relationship between flow and sandbar change. However, the component graphs can also be used to identify change. Morphodynamic change can be distinguished from hydrodynamic change by
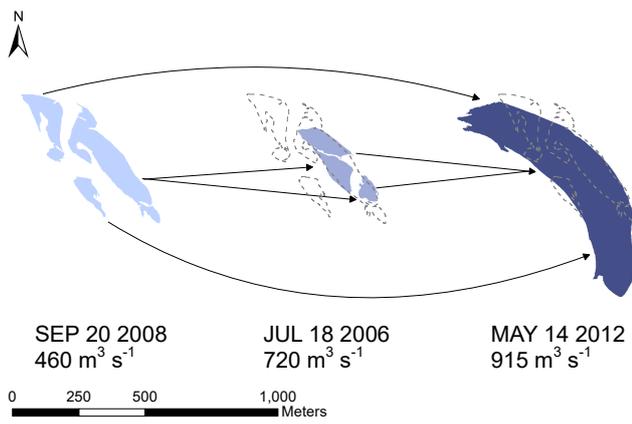
Prepared using **sagej.cls**

Figure 11. Flow-progression graph of a sandbar complex. The flow-progression graph is not a true tree and suggests morphodynamic change between 2008 and 2012. However, for the 2006–2008 interval the flow-progression graph is a tree.

Figure 12. *Top:* The reachability graph of a sandbar complex based on a union of five layers. Although no features were captured in the 2006 layer, landform persistence is assumed (virtual observation) based on overlap between other layers, and the features from each layer are linked as a single geomorphic unit. *Bottom:* The reachability graph of a sandbar complex based on a memory of approximately one year. Because no features were detected in 2006, no connection is made between the features detected in 2005 and those detected in 2007. As a result, the features from 2005 are grouped into one geomorphic unit while the features from 2007 and 2008 are grouped into a separate geomorphic unit.

evaluating the structure of the flow-progression graph. If the flow-progression graph is not a true tree, this indicates morphodynamic change. The sandbar complex in Figure 11 shows change that is not fully explained by hydrodynamics. The sandbar complex consolidated and grew between 2008 and 2012, likely the result of high flows in 2010 and the historic 2011 flood (Cowman 2012; Gusman et al. 2015). These significant morphodynamic changes result in a flow progression graph that is not a tree.

Even for sandbar complexes that erode or build without significant change to geomorphic structure, progression graphs can still be informative. For instance, area-progression graphs can be constructed based on the total area of the sandbar complex at each snapshot in time. If the flow-progression graph is not isomorphic to the area-progression graph, this indicates that there are changes in the sandbar area that are not consistent with hydrodynamics alone. Alternatively, the graph structure can be used to identify specific periods during which morphodynamic change occurred. For example, the flow progression graph of the sandbar complex in Figure 11 is a true tree for 2006–2008 interval, indicating that the observed change during this period is fully consistent with flow variability. Therefore, one can reasonably conclude from the graph that the majority of morphodynamic change recorded in the available data occurred during the 2008–2012 interval. Testing for tree structures and isomorphism can be used as a preliminary screening method for separating sandbars that undergo significant morphodynamic change over a given time period from "stable" sandbars that do not. Most graph manipulation software packages include algorithms for programmatically identifying tree structures and comparing graphs (or subsections of graphs, i.e., *subgraphs*) for isomorphism.
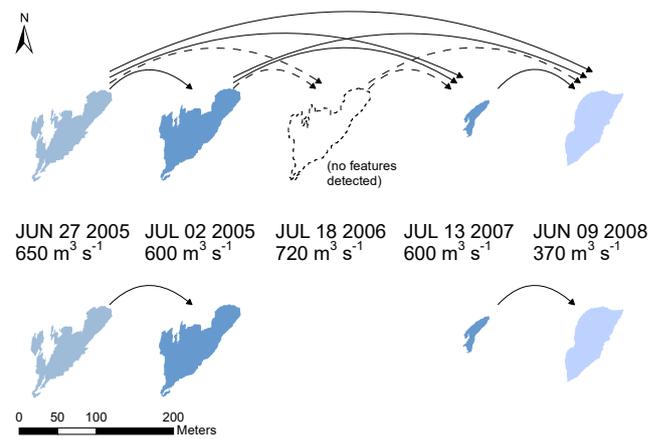
Sandbars can become fully submerged in one snapshot and re-emerge in a later snapshot when flows recede. The geomorphic footprint is a critical tool for linking observations across high-flow snapshots and separating hydrodynamic change from morphodynamic change. The simplest way to construct the geomorphic footprint is to perform a union of all available snapshots. However, using all available layers to construct the geomorphic footprint assumes features that disappear and reappear in the same region are the same geomorphic unit, as shown in Figure 12 (top). This assumption may not be true in landscapes where landforms undergo rapid morphodynamic change and persist for relatively short periods of time. In such cases, defining the geomorphic footprint from all available layers may erroneously link a newly-created landform to a landform that occupied the region at an earlier time but has since disappeared (e.g., due to erosion). Migrating features that cross the same region at different times may also be erroneously identified as belonging to the same geomorphic unit. Linkages between features across layers can be corrected by incorporating the concept of *memory* (Brierley 2010) to modify geomorphic unit definitions.

The relationship between a progression graph and the reachability graph provides a mechanism for imposing a finite memory on geomorphic units. Edges can be dropped from the reachability graph if they exceed a specified time threshold. Alternatively, the edges of a progression graph can be modified and the reachability graph can be reconstructed.

Incorporating a threshold for landform memory into the reachability graph may change how features are grouped into geomorphic units, as shown in Figure 12 (bottom). These changes will affect the resulting panel data and geomorphic footprints, and the structure of any derived progression graphs.

Conclusion

We described a simple and flexible graph-based method for generating panel datasets from geomorphological spatial data. The method provides a number of advantages over traditional temporally-naive procedural methods that operate on geographic units without differentiating between within-landform change and between-landform variation. We used GEOBIA-derived data on emergent sandbars to provide examples of generating landscape descriptors, handling missing data and developing statistical models using the new method, and described promising avenues for geomorphological analyses that leverage graph structure. The approach is readily applied to a wide variety of dynamic landscapes and types of spatial data, and overcomes significant challenges in managing and analyzing spatial data in large-scale studies of geomorphic change.

References

Aurousseau P, Gascuel-Odoux C, Squividant H, Trepos R, Tortrat F and Cordier M (2009) A plot drainage network as a conceptual tool for the spatial representation of surface flow pathways in agricultural catchments. *Computers & Geosciences* 35(2): 276–288.

Bishop M, James L, Shroder J and Walsh S (2012) Geospatial technologies and digital geomorphological mapping: concepts, issues and research. *Geomorphology* 137(1): 5–26.

Bishop MP, Shroder JF, Bonk R and Olsenholler J (2002) Geomorphic change in high mountains: a western himalayan perspective. *Global and Planetary change* 32(4): 311–329.

Bishop MP, Shroder JF and Colby JD (2003) Remote sensing and geomorphometry for studying relief production in high mountains. *Geomorphology* 55(1): 345–361.

Blaschke T (2010) Object based image analysis for remote sensing. *ISPRS journal of photogrammetry and remote sensing* 65(1): 2–16.

Brasington J, Langham J and Rumsby B (2003) Methodological sensitivity of morphometric estimates of coarse fluvial sediment transport. *Geomorphology* 53(3): 299–316.

Brierley G (2010) Landscape memory: the imprint of the past on contemporary landscape forms and processes. *Area* 42(1): 76–85. DOI:10.1111/j.1475-4762.2009.00900.x.

Buenau K, Hiller T and Tyre A (2014) Modelling the effects of river flow on population dynamics of piping plovers (Charadrius melodus) and least terns (Sternula antillarum) nesting on the missouri river. *River Research and Applications* 30(8): 964–975.

Burt J, Barber G and Rigby D (2009) *Elementary statistics for geographers*. Guilford Press.

Butts CT (2008) network: a package for managing relational data in r. *Journal of Statistical Software* 24(2).

Carroll S and Pearson D (1998) The effects of scale and sample size on the accuracy of spatial predictions of tiger beetle (cicindelidae) species richness. *Ecography* 21(4): 401–414.

Cowman T (2012) Impacts of the 2011 flood on the missouri river channel. *Quality on Tap, South Dakota Association of Rural Water Systems* 7: 6–9.

Csardi G and Nepusz T (2006) The igraph software package for complex network research. *InterJournal* Complex Systems: 1695.

Del Mondo G, Stell J, Claramunt C and Thibaud R (2010) A graph model for spatio-temporal evolution. *Journal of Universal Computer Science* 16(11): 1452–1477.

Drăguţ L and Blaschke T (2006) Automated classification of landform elements using object-based image analysis. *Geomorphology* 81(3): 330–344.

Duncan G and Kalton G (1987) Issues of design and analysis of surveys across time. *International Statistical Review/Revue Internationale de Statistique* : 97–117.

Enders C (2010) *Applied missing data analysis*. Guilford Press.

Farr TG and Chadwick OA (1996) Geomorphic processes and remote sensing signatures of alluvial fans in the Kun Lun mountains, China. *Journal of Geophysical Research: Planets* 101(E10): 23091–23100.

Feurer D, Bailly JS, Puech C, Le Coarer Y and Viau AA (2008) Very-high-resolution mapping of river-immersed topography by remote sensing. *Progress in Physical Geography* 32(4): 403–419.

Fitzmaurice G, Laird N and Ware J (2012) *Applied longitudinal analysis*. 2 edition. John Wiley & Sons.

Frees E (2004) *Longitudinal and panel data: analysis and applications in the social sciences*. Cambridge University Press.

Gascuel-Odoux C, Aurousseau P, Doray T, Squividant H, Macary F, Uny D and Grimaldi C (2011) Incorporating landscape features to obtain an object-oriented landscape drainage network representing the connectivity of surface flow pathways over

rural catchments. *Hydrological Processes* 25(23): 3625–3636.

Gelfand A, Diggle P, Guttorp P and Fuentes M (2010) *Handbook of spatial statistics*. CRC press.

Goodchild M and Haining R (2004) Gis and spatial data analysis: Converging perspectives. *Papers in Regional Science* 83(1): 363–385.

Gusman J, Tripolitis V, Jenkins C, Svendsen C and Pridal D (2015) Sandbar growth and decay on the Missouri River during the high flows of 2010 and the historic 2011 flood. Consultant report to the U.S. Army Corps of Engineers, WEST Consultants.

Heckman J (1977) Sample selection bias as a specification error (with an application to the estimation of labor supply functions). NBER Working Paper 172, National Bureau of Economic Research, Cambridge, MA, USA.

Heckmann T, Hilger L, Meßenzehl K, Hoffmann T, Schwanghart W, Götz J and Buckel J (2014) Network analysis of sediment cascades derived from digital geomorphological maps–a comparative study of three catchments in the Austrian and Swiss Alps. In: *EGU General Assembly Conference Abstracts*, volume 16. p. 11452.

Heckmann T and Schwanghart W (2013) Geomorphic coupling and sediment connectivity in an alpine catchment–exploring sediment cascades using graph theory. *Geomorphology* 182: 89–103.

Heckmann T, Schwanghart W and Phillips J (2015) Graph theory–recent developments of its application in geomorphology. *Geomorphology* 243: 130–146.

Heritage GL, Milan DJ, Large AR and Fuller IC (2009) Influence of survey strategy and interpolation model on dem quality. *Geomorphology* 112(3): 334–344.

Hernandez P, Graham C, Master L and Albert D (2006) The effect of sample size and species characteristics on performance of different species distribution modeling methods. *Ecography* 29(5): 773–785.

Hilton T and Patrick C (1969) Cross-sectional versus longitudinal data: an empirical comparison of mean differences in academic growth. *ETS Research Report Series* 1969(1).

Hsiao C (1985) Benefits and limitations of panel data. *Econometric Reviews* 4(1): 121–174.

Ibrahim J and Molenberghs G (2017) Missing data methods in longitudinal studies: a review. *TEST* 18: 1–43.

Jacobson R (2013) Riverine habitat dynamics. In: Shroder JF (ed.) *Treatise on Geomorphology*. San Diego: Academic Press. ISBN 978-0-08-088522-3, pp. 6–19. DOI:https://doi.org/10.1016/B978-0-12-374739-6.00318-3.

James LA, Hodgson ME, Ghoshal S and Latiolais MM (2012) Geomorphic change detection using historic maps and dem

differencing: The temporal dimension of geospatial analysis. *Geomorphology* 137(1): 181–198.

Jordán F and Scheuring I (2004) Network ecology: topological constraints on ecosystem dynamics. *Physics of Life Reviews* 1(3): 139–172.

Kadmon R, Farber O and Danin A (2003) A systematic analysis of factors affecting the performance of climatic envelope models. *Ecological Applications* 13(3): 853–867.

Kirsch E (1996) Habitat selection and productivity of least terns on the Lower Platte River, Nebraska. *Wildlife Monographs* (132): 3–48.

Lang S (2008) Object-based image analysis for remote sensing applications: modeling reality–dealing with complexity. *Object-based image analysis* : 3–27.

Mandlburger G, Hauer C, Wieser M and Pfeifer N (2015) Topobathymetric lidar for monitoring river morphodynamics and instream habitatsa case study at the pielach river. *Remote Sensing* 7(5): 6160–6195.

Martınez-Casasnovas J, Ramos M and Poesen J (2004) Assessment of sidewall erosion in large gullies using multi-temporal dems and logistic regression analysis. *Geomorphology* 58(1): 305–321.

Menard S (2002) *Longitudinal research*, *Quantitative Applications in the Social Sciences*, volume 76. 2 edition. SAGE Publications, Inc.

Napieralski ID Jacob andBarr, Kamp l and Kervyn M (2013) Remote sensing and GIScience in geomorphology. In: Bishop M and Shroder J (eds.) *Treatise on Geomorphology*. Elsevier, pp. 187–227.

Pardo-Iguzquiza E, Durán-Valsero J and Rodríguez-Galiano V (2011) Morphometric analysis of three-dimensional networks of karst conduits. *Geomorphology* 132(1): 17–28.

Pearl J (2009) *Causality: Models, Reasoning and Inference*. 2 edition. Cambridge University Press.

Perret J, Prasher S, Kantzas A and Langford C (1999) Three-dimensional quantification of macropore networks in undisturbed soil cores. *Soil Science Society of America Journal* 63(6): 1530–1543.

Phillips J (2013) Networks of historical contingency in earth surface systems. *The Journal of Geology* 121(1): 000–000.

Phillips J, Schwanghart W and Heckmann T (2015) Graph theory in the geosciences. *Earth-Science Reviews* 143: 147–160.

Phinn SR, Roelfsema CM and Mumby PJ (2012) Multi-scale, object-based image analysis for mapping geomorphic and ecological zones on coral reefs. *International Journal of Remote Sensing* 33(12): 3768–3797.

Ployhart R and Vandenberg R (2010) Longitudinal research: The theory, design, and analysis of change. *Journal of Management*

36(1): 94–120.

Reddy S and Dvalos L (2003) Geographical sampling bias and its implications for conservation priorities in africa. *Journal of Biogeography* 30(11): 1719–1727. DOI:10.1046/j.1365-2699.2003.00946.x.

Renschler C, Doyle M and Thoms M (2007) Geomorphology and ecosystems: challenges and keys for success in bridging disciplines. *Geomorphology* 89(1): 1–8.

Robinson W (1950) Ecological correlations and the behavior of individuals. *American Sociological Review* 15(3): 351–357.

Rodríguez-Iturbe I and Rinaldo A (2001) *Fractal river basins: chance and self-organization*. Cambridge University Press.

Schult DA (2008) Exploring network structure, dynamics, and function using networkx. In: *Proceedings of the 7th Python in Science Conference (SciPy)*. pp. 11–15.

Schwalbach M (1998) *Conservation of Least Terns and Piping Plovers Along The Missouri River and Its Major Western Tributaries in South Dakota*. Masters thesis, South Dakota State University.

Selvin H (1958) Durkheim's Suicide and problems of empirical research. *American journal of sociology* 63(6): 607–619.

Shadish W, Cook T and Campbell D (2001) *Experimental and quasi-experimental designs for generalized causal inference*. 2 edition. Wadsworth Publishing.

Shruthi RB, Kerle N and Jetten V (2011) Object-based gully feature extraction using high spatial resolution imagery. *Geomorphology* 134(3): 260–268.

Sidle J, Carlson D, Kirsch E and Dinan J (1992) Flooding: Mortality and habitat renewal for least terns and piping plovers. *Colonial Waterbirds* 15(1): 132–136.

Simpson E (1951) The interpretation of interaction in contingency tables. *Journal of the Royal Statistical Society. Series B (Methodological)* : 238–241.

Smith J and Renken R (1993) Reproductive success of least terns in the mississippi river valley. *Colonial Waterbirds* : 39–44.

Strong L (2012) Extending a prototype knowledge and object-based image analysis model to coarser spatial resolution imagery: an example from the Missouri River. *Proceedings of the 4th GEOBIA* : 530–535.

Thibaud R, Del Mondo G, Garlan T, Mascret A and Carpentier C (2013) A spatio-temporal graph model for marine dune dynamics analysis and representation. *Transactions in GIS* 17(5): 742–762.

Thulasiraman K and Swamy M (1992) *Acyclic Directed Graphs*, chapter 5.7. New York, NY, USA: John Wiley & Sons, Inc. ISBN 0-471-51356-3, p. 118.

Trivellato U (1999) Issues in the design and analysis of panel studies: a cursory review. *Quality & Quantity* 33(3): 339–351.

USACE (2013) Hydrologic statistics. Technical report, Missouri River Basin Water Management Division, U.S. Army Corps of Engineers, Omaha, NE.

Valentini L, Perugini D and Poli G (2007) The "small-world'" topology of rock fracture networks. *Physica A: Statistical Mechanics and its Applications* 377(1): 323–328.

van Buuren S (2012) *Multiple imputation*. Interdisciplinary Statistics Series. Chapman & Hall/CRC, pp. 25–52.

Werner C (1993) Ridge/channel path interdependence in drainage basins. *Geografiska Annaler. Series A. Physical Geography* : 73–81.

Werner C (1994) Explorations into the formal structure of drainage basins. *Earth Surface Processes and Landforms* 19(8).

Wheaton JM, Brasington J, Darby SE and Sear DA (2010) Accounting for uncertainty in dems from repeat topographic surveys: improved sediment budgets. *Earth Surface Processes and Landforms* 35(2): 136–156.

White R (2006) Pattern based map comparisons. *Journal of Geographical Systems* 8(2): 145–164.

Woolard JW and Colby JD (2002) Spatial characterization, resolution, and volumetric change of coastal dunes using airborne lidar: Cape hatteras, north carolina. *Geomorphology* 48(1): 269–287.