

A method for defining dispersed community territories

Kenneth B. Vernon¹ and Scott G. Ortman¹

¹*Center for Collaborative Synthesis in Archaeology, Institute of Behavioral Science, University of Colorado, Boulder, 1440 15th Street, Boulder, CO 80302*

Draft compiled on 2023-10-07

Abstract

The transition from dispersed to aggregated forms of settlement reflects a critical shift in the relative value of social and primary (food) modes of production. Investigating trade-offs between them, however, requires estimates of the extent of community territories, including their nearby arable land. Here we demonstrate a simple algorithm to do that. Our algorithm is analogous to that used to define core-based statistical areas for the US census, though instead of modern metropolitan areas, we rely on community centers (or areas of known and persistent interaction between unrelated individuals). We provide examples of our algorithm by applying it to archaeological sites in the central Mesa Verde, northern Rio Grande regions, and Cibola regions. Together, these data provide a comprehensive, high resolution record of prehistoric maize farming in the North American Southwest, representing varying levels of aggregation and dispersal through time.

Keywords: Aggregation, Core-Based Statistical Area, Community Detection, Intensification

Target Journal(s): Journal of Archaeological Science

1 Introduction

Urban geographers recognize density as a crucial property of human communities which influences many other properties, including productivity, resource needs, and infectious disease rates (Angel et al. 2021; Bettencourt 2021; Duranton and Puga 2020). The archaeological record of subsistence farming societies contains physical traces of community formations with a wide range of densities, often glossed as “dispersed” or “aggregated” (Birch 2013; Drennan, Berrey, and Peterson 2015; Gyucha 2019). The primary difference between the two has to do with the spatial distribution of residents relative to primary (food) production. In dispersed formations, households are scattered across an area, interspersed with the land that they farmed, thus requiring them to commute to central places for various forms of social production. In aggregated formations, in contrast, households are clustered in a village or town, such that they reside in locations of social production, but must commute to fields for primary production.

From a complex systems perspective, the main way these two settlement morphologies vary is with respect to which costs are being minimized. In the dispersed pattern, costs associated with primary (food) production by the household are being minimized, whereas in the aggregated pattern, it is costs associated with various forms of social production, from government and ritual to economic exchange and warfare. Presumably, choices regarding which sorts of costs to minimize are driven by the relative productivities of primary and social production, which is to say, their relative importance, broadly construed, for human well-being in a given context.

There are many episodes in history where human settlements transitioned from dispersed to aggregated, and presumably this is a signal of a change in the relative value of social vs. primary production for the residents. This can occur for a variety of reasons, including an increase in the social cost of not being aggregated (e.g., due to warfare); a decrease in transport costs for staples (due to animal traction, wheeled vehicles, roads, etc.); an increase in land productivity which changes the balance of costs of transport for primary vs. social production; or an increase in the contribution of exchange to household incomes, which can derive from increases in community size and density. In addition, while one can conceptualize transitions between dispersed and aggregated

forms of settlement in general systems terms, in real human communities a variety of political or ideological factors can keep people from adopting an energetically balanced form of settlement given prevailing conditions. So, there are opportunities for archaeologists to study deviations from general equilibrium conditions in addition to factors that shift the equilibrium in a given context.

For all this work, however, there is a fundamental issue: the identification of community territories. It is relatively straightforward to define the boundaries of aggregated settlements by finding the extent of built space and/or artifactual remains. However, to define the community boundary one needs to know the extent of agricultural land used by the residents, and this is often difficult to determine from the village remains themselves (Varien 1999b). In contrast, it is often quite difficult to determine the boundaries of dispersed communities due to gaps in survey coverage or relatively consistent distributions of farmsteads, but once one has determined the community territory it is straightforward to define the associated agricultural land.

In this paper, we present a method we developed to solve such problems, which enables us to compare the properties and resilience of dispersed vs. aggregated communities on the same landscapes. We draw on three extensive compilations of archaeological data from different portions of the ancestral Pueblo region of the US Southwest, using the locations of what we call community centers as tethers for defining territories that we infer were used by social communities over extended periods of time (Glowacki and Ortman 2012; Varien et al. 2007). In some regions these are sites containing more households than could plausibly be related by reckoned kinship or some form of civic architecture. The key feature of these centers is that they represent locations against which travel time distances to other sites can be compared.

While not strictly required by our algorithm, we make the simplifying assumption that once established, community territories were relatively fixed in space for the duration of their associated sites. This allows us to use the distribution of all recorded residences, regardless of their periods of occupation, in defining community territories. There are several reasons why we believe this is a reasonable simplification. First, there are strong cross-cultural regularities in the time individuals spend in daily travel from and to their residences, so dispersed farming communities tend to be

dispersed across the distance associated with this typical travel time, regardless of the number of households involved (Marchetti 1994). Second, the archaeological record of our study areas supports the idea that community territories did not shrink when their populations aggregated into villages. In some cases, farmsteads were converted to field houses, with some building materials being reused elsewhere and evidence for continued limited activity use (Varien 1999b, 1999a, 2002). In other cases, grid gardens were constructed directly within older rubble mounds by residents of newer, aggregated villages nearby (Gauthier and Herhahn 2005). Third, an analysis that allowed community territories to vary by time step would reduce variation in the density of the resulting units, thus washing out variation in one of the most important social properties one would ideally want to examine. Fourth, in areas where there is strong survey coverage, dispersed residences form clusters, often but not always centered on a few larger settlements, and this clustering is apparent even when the dispersed residences are plotted by time step (Schachner 2012; Varien 2002). Fifth, the effects of population growth for dispersed farming communities are not parallel to the effects of population expansion in aggregated settlements. In the latter, the area of the settlement must grow somewhat because the residential density is already relatively high. But in dispersed settlements, there is often unused land between existing farmsteads that can be settled. This, combined with the strong cross-cultural regularity in typical commute times, leads to much greater consistency in the spatial extent of residents in a dispersed community than in the case of an aggregated community. Finally, given the inherent incompleteness of the archaeological record, it seems most appropriate to use all remains to define what is effectively the maximum extent of community territories. While it was undoubtedly true that individual community territories expanded or contracted over time, the assumption of constant area seems more reasonable than the assumption of constant density.

2 Data

The datasets we apply our algorithm to come from three regions in the US Southwest: Central Mesa Verde (CMV), the Northern Rio Grande (NRG), and the Cibola or Zuni region (CIB) (see Figure 1). The CMV and NRG datasets derive from Phase II of the Village Ecodynamics Project

(Ortman 2016a; Schwindt et al. 2016). The CIB data derive from data syntheses produced by Schachner and Peeples (Peeples 2018; Peeples and Schachner 2012; Schachner 2012). All of these data are now included in cyberSW (Mills et al. 2020). For summaries of the archaeological and environmental context in each region, see Table 1.

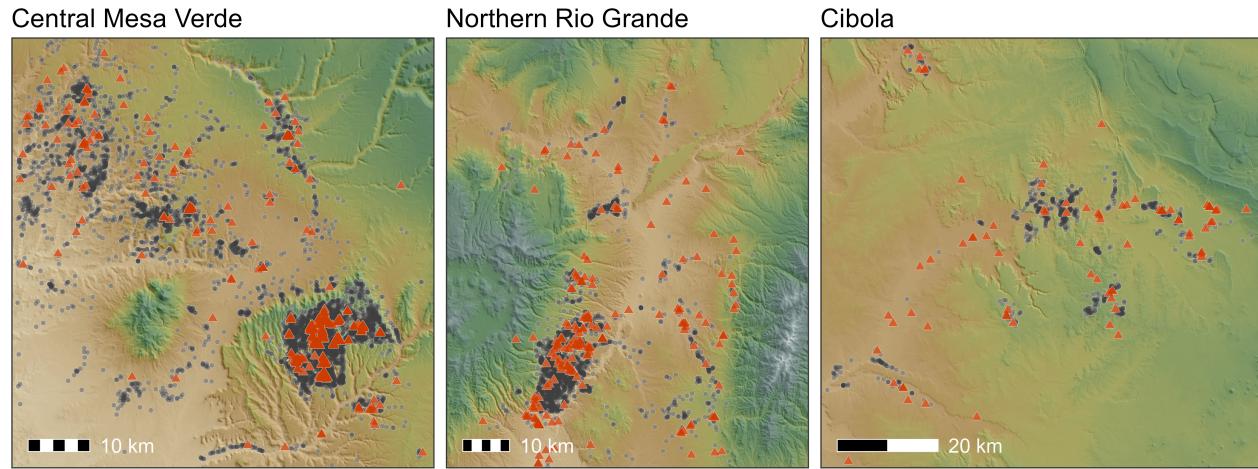


Figure 1. Overview map showing the locations of farmsteads (black dots) and community centers (red triangles) in Central Mesa Verde, Northern Rio Grande, and Cibola.

The CMV area encompasses the southwestern corner of Colorado, from the Mesa Verde cuesta in the south and east to the Dolores River in the north, and the Utah state line in the west. The elevation ranges from roughly 1,399.8 m to 3,032.8 m, with an average elevation of 1,999.0 m ($\sigma = 276.5$ m). SKOPE (Bocinsky et al. 2022) estimates suggest annual precipitation in the CMV study area has averaged around 39.5 cm ($\sigma = 9.1$ cm) over the last 2,000 years, and it puts summer growing degree days for maize around 2,390.1 °F ($\sigma = 97.9$ °F). Across most areas in the region, it is believed that direct precipitation was sufficiently high to support dry farming in most years (Bocinsky and Kohler 2014; Bocinsky et al. 2016).

The NRG study area in north central New Mexico includes basically all of the Rio Grande valley between the Jemez Mountains to the west and the southernmost mountains of the Sangre De Cristo range to the east, with the Rio Grande itself running along the western edge of the valley at the foot of the Pajarito Plateau. The elevation ranges from roughly 1,587.9 m to 3,836.2 m, with an average elevation of 2,220.0 m ($\sigma = 371.3$ m). SKOPE estimates the average annual precipitation in the NRG over the last 2,000 years to be around 40.6 cm ($\sigma = 8.2$ cm), and it puts

Table 1. Regional Summaries

	Area [km²]	Dates [CE]	Archaeology	Environment
CMV	4,556.4	725-1280	Rooms: 67,746 Farms: 5,415 Centers: 512	Precipitation: 39.5 cm (σ : 9.1) Maize GDD: 2,390.1 °F (σ : 97.9) Elevation: 1,999.0 m (σ : 276.5)
NRG	6,958.0	900-1550	Rooms: 46,226 Farms: 2,316 Centers: 181	Precipitation: 40.6 cm (σ : 8.2) Maize GDD: 2,154.4 °F (σ : 102.1) Elevation: 2,222.0 m (σ : 371.3)
CIB	7,424.1	700-1540	Rooms: 27,234 Farms: 713 Centers: 78	Precipitation: 34.3 cm (σ : 7.6) Maize GDD: 2,317.3 °F (σ : 93.8) Elevation: 2,145.9 m (σ : 160.0)

summer growing degree days for maize around 2,154.4 °F ($\sigma = 102.1$ °F). It is argued that direct rainfall on the Pajarito Plateau made dry farming possible there, but in the valley farming required floodwater irrigation along watercourses (Duwe and Anschuetz 2013; Bocinsky and Kohler 2014; Bocinsky et al. 2016; Ortman 2016b).

The CIB region straddles the southern slope of the Colorado Plateau along the border between Arizona and New Mexico. Although this is a rugged terrain by any reasonable standard, it is notably less so than the other study areas. The elevation ranges from roughly 1,771.3 m to 2,777.1 m, with an average elevation of 2,145.9 m ($\sigma = 160.0$ m). SKOPE estimates suggest annual precipitation in the CIB area over the last 2,000 years has averaged around 34.3 cm ($\sigma = 7.6$ cm), and it puts summer growing degree days for maize around 2,317.3 °F ($\sigma = 93.8$ °F). However, no area in the region receives sufficient direct rainfall to support dry farming, so some form of water management was required (Keith W. Kintigh 1985; Muenchrath et al. 2002).

3 Methods

Our clustering algorithm draws inspiration from the procedure used by the United States Office of Management and Budget to define Core Based Statistical Areas (CBSA) (USOMB 2021). We start by defining a core area, in this case a community center location. We then associate neighboring residential sites or farmsteads with this center based on their commute times, and we merge core areas into larger agglomerations based on the proportion of their populations that they share.

Our method may be described as a guided density-based clustering algorithm in that it does not rely on a random set of points when initialized. Instead, the selection of core areas is guided by archaeological data and regional expertise. We think this is a key argument in its favor.

From an archaeological perspective, one can think of our algorithm as combining insights from Varien (1999b) and Reese, Glowacki, and Kohler (2019), specifically Varien’s idea of a community catchment, which is an isochrone defined by a uniform commute time in all directions from a community center, and Reese’s suggestion that a spatial or geographic community can be identified by grouping individuals in terms of the commute distances between them. Our way of combining these ideas is to associate farm sites with specific community centers in terms of their commute distances and then draw the community catchment around those associated farms. We also share with them and others (Lipe and Hegmon 1989; Mahoney, Adler, and Kendrick 2000; Peterson and Drennan 2005) a focus on “geographic” as opposed to “purely social” communities; that is, communities constrained in space and comprised of individuals that can be expected to interact on a regular, even daily, basis.

In broad outline, our proposed community detection algorithm has the following steps:

1. Identify community centers
2. Identify farms within a commute time threshold ($D\text{-}max$)
3. Join communities to their overlapping neighbors (P , $D\text{-}join$)
4. Exclude farms closer to another community center (tie-breaking step)
5. Draw smallest concave hull encompassing all farms, centers, and paths

After step 4, we also apply a filter to remove communities that have less than a minimal number of dispersed farmsteads, specifically four. Although this is a somewhat arbitrary threshold, we note that three vertices are the bare minimum required to define a polygon and measure its area. By removing small communities, we also minimize potential underestimates of total arable land area. In addition, our larger ambition is to understand processes occurring within dispersed farming communities, not necessarily to define all the dispersed communities that once existed in an area. For community centers with few or no farms, there is no meaningful dispersed farming community

to examine, regardless of whether this is due to limited survey coverage or actual past behavior, hence our desire to exclude them.

The primary tuning parameters in our algorithm are P , the proportion of a community population used to join communities; $D\text{-join}$, the commute time required to calculate P ; and $D\text{-max}$, the commute time used to exclude distant farms. These are described in more detail below.

3.1 Identify community centers

Rule: Any site, whether residential or otherwise, with known and persistent interaction between unrelated individuals should count as a community center.

Why persistent interactions? Intuitively, a community center is a place where individuals from different households cross paths with the intention to interact. A simple way to demonstrate that intention is to show that individuals would still be willing to pay some non-negligible cost to visit those centers under various counterfactual changes to local conditions. This serves to distinguish them from merely accidental exchanges or chance encounters. Of course, there can be some path-dependence to this process, with locations of ephemeral or one-off encounters coming to be locations of persistent interaction. The point is that the location will come to be a community center so long as individuals continue to visit that location despite the underlying variance in their circumstances.

Why unrelated individuals? The simple answer is that a community is not a family. As a unit of social organization, it does share something in common with a family, namely, a set of shared interests; but unlike in a family, those shared interests form only a small part of a much larger set of potentially conflicting goals and desires. As a consequence, communities are more susceptible to problems with coordination or collective action. On the other hand, the mere existence of a community center implies some level of coordination - even if that coordination is just about the location of the community center itself! So, community members share less in common than families, but more in common than a purely random sample of the larger population.

We can summarize these ideas by noting that persistent interactions at a location make that location a *center*, and persistent interactions between unrelated individuals at that location make it a *community* center. To make these considerations more useful for defining community bound-

aries, however, we need to operationalize them, to make them applicable to specific archaeological contexts, and that requires a deeper understanding of the archaeology in those contexts. For the two regions represented by the VEP II dataset, we rely on two different but related sets of criteria to identify community centers. In the Central Mesa Verde area, a site is considered a community center if it includes civic architecture, specifically a great kiva or great house, or eight or more pit structures (small kivas) representing eight or more households. In the Northern Rio Grande, a site counts as a community center if it is classified as a village (a site with 50 or more rooms) or town (a site with 500 or more rooms). In the Cibola region, a community center includes at least 50 rooms or civic architecture. The large number of households in these regions are thought to be too large to be plausibly related by reckoned kinship.

3.2 Join farms within a commute time

Rule: For each center, join all farms within commute time, $D\text{-}max$.

Following Reese, Glowacki, and Kohler (2019), we define nearness or proximity in terms of commute time rather than geographic distance. This is done for the obvious reason that the primary opportunity cost associated with daily pedestrian movement is the time spent walking, during which other productive activities cannot take place. There is also a much stronger cross-cultural regularity in the length of time individuals devote to commuting than the distance covered, as the latter is a function of the speed of movement, which is influenced by transport technology (Marchetti 1994). As a result, both the frequency and intensity of interaction at community centers should decay with commute time. The longer it takes to get to a community center, the less frequent and less intense the interactions at that center should be (Varien 1999b; Peterson and Drennan 2005).

While different constraints will naturally arise for different modes of transportation, leading to differences in both the routes taken and the speed and distance covered, for the populations considered here the only available mode of transportation was pedestrian. The biggest obstacle to interaction was thus the landscape itself, in particular its topography - intuitively, the steeper the uphill slope, the slower the hiking speed, the longer the commute time. So, all else being equal,

individuals at a farm that is equidistant (in terms of geographic distance) between two community centers should prefer traveling to the one separated by less rugged terrain.

The concept of commute time is a critical variable in many places in our algorithm, but as a first approximation, we incorporate the idea by setting a commute time threshold, $D\text{-max}$. This variable, in effect, defines an isochrone around each community center, what Varien (1999b) calls a “catchment.” All farm sites that fall within that isochrone get associated with its community center. This threshold also serves as a filter, as we drop all farm sites that are farther than that commute time to any center. Doing so helps to minimize potential overestimates of total arable land area for each community. Here, we have defined $D\text{-max}$ as one hour. This is about twice the median daily commute time observed in cross-cultural studies and it includes roughly 95% of the farmsteads in each region (see Figure 3).

For details of how we estimate commute times over costly terrains, see “Least-cost path analysis,” below.

3.3 Join centers with overlapping neighbors

Rule: For centers C_1 and C_2 with nearby farm populations $N_1 \leq N_2$, if $P \cdot N_1$ are within distance $D\text{-join}$ of C_1 and C_2 , then join C_1 to C_2 .

The larger and more dense farming populations become, the harder it gets to tell them apart and, more importantly, the harder it gets to justify keeping them apart, so we need to articulate a rule that specifies when two communities should be merged into one. For CBSAs, that decision hinges on a distinction between central and outlying counties, or counties where some fraction of the population live in an urban core and counties where some fraction commute to and from a central county on a regular basis. Basically, if the central counties of one CBSA qualify as the outlying counties of another CBSA, then the two CBSAs are merged. The intuition here is that when rates of interaction and exchange between CBSAs are roughly equal to rates within a CBSA, the CBSAs should be merged (USOMB 2021).

We carry that intuition over to our merger rule for dispersed maize farming communities, albeit with some modification to account for the fact that we are working with point locations rather than

administrative boundaries. We assume that if some proportion, P , of the population related to one community center lives within a certain commute time, $D\text{-join}$, of another community center, that those two community centers are effectively parts of the same dispersed community. We do this in the direction of letting larger communities absorb smaller ones.

More precisely, the rule requires that $P \cdot N_1$ be within $D\text{-join}$ of both C_1 and C_2 . This promotes joins when C_1 and C_2 are themselves in close proximity to each other, as the farther apart they are, the smaller the number of sites that overlap within $D\text{-join}$. It also tends to encourage joining two communities when a substantial proportion of the population lives between the two centers. In either case, the result is what Varien (1999b) sometimes refers to as a “macro-community” or “multi-community cluster,” meaning a community organized around multiple community centers.

The rule was partly inspired by the density-based clustering algorithm known as DBSCAN (Ester et al. 1996), as we rely on a moving window (an isochrone, in this case) set around each center, and look for some measure of density to guide join-decisions. As noted above, however, our method does not rely on a random initialization of points, but rather targets community centers. It is also hierarchical in the sense that it exploits additional structure in our data, namely the distinction between community centers and farm sites.

As we cannot measure population directly, however, we use information we have about farm sites, specifically estimates of their room counts, as a proxy. For many sites in the CMV, pit structures or small kivas are more reliably identified than surface rooms. Fortunately, each pit structure is typically the central roofed space of a unit pueblo residence that includes an average of six surface rooms (Adler 1990), so we simply multiply the number of pit structures by six to obtain room estimates for each site. CIB and NRG sites all have room estimates, so no additional assumptions are made about them. Finally, we follow the well-established rule of thumb in US Southwest archaeology that each surface room in a settlement represents a single person, leading to a one to one conversion from rooms to residents (Duwe et al. 2016; Keith W. Kintigh 1985; Lekson 1989; Lipe 1989; Ortman 2016a).

For this analysis, we specify $D\text{-join}$ to be one half hour and P to be 0.8. We acknowledge that these values are somewhat arbitrary. In their defense, they set an extremely high bar, so joining centers that satisfy the rule seems reasonable.

3.4 Exclude farms closer to another community center

Rule: For each farm that falls within $D\text{-max}$ of multiple community centers, associate it with the community center to which it has a shorter commute time.

On occasion a farm may fall within $D\text{-max}$ of multiple community centers that are not joined into a single community, which makes them members of multiple communities. For the dispersed farming communities we aim to identify, this would imply that multiple communities could simultaneously farm the same plot of land, which is unrealistic, so we need a rule for breaking ties. Because we are already relying so heavily on commute time, the simplest solution is to associate a farm that falls within $D\text{-max}$ of multiple community centers with the center and, thus, the community that it is closer to, and this is what we have elected to do. We can justify this step on similar grounds as well, with the safe assumption that individuals should be more likely to interact the nearer they are in time and space. We also note that this tie-breaking rule ensures that community polygons are continuous, that they do not contain any holes, which would be quite odd considering these community boundaries are supposed to represent the spatial extent of their arable land. It also encourages some separation between communities, preventing them from overlapping to any significant degree.

3.5 Draw community boundary

Rule: For each community, draw a polygon that encompasses all farms, community centers, and commute paths using a concave hull.

So far, we have only grouped site points into clusters or communities. The goal, however, is to define their spatial extent. That means, for each set of site points, that we need some way of drawing a polygon that encompasses all of them. This can be done in a number of ways. The simplest strategy would be to find the centroid of the set of points and draw a circle with the

smallest radius that still includes them all. This would be utterly arbitrary, however, and frankly unrealistic, not representing any sort of meaningful boundary given the landscape. An alternative would be to use the convex hull of the set of points, but this tends to exaggerate the total area of the community, especially when the distribution of points is suggestive of a concave shape, like the letter ‘C’. A concave hull would handle peculiar shapes, but it would also introduce the opposite problem, restricting the area of the community to an unreasonable degree. Our somewhat brute force way of threading this needle is to identify paths with shortest commute times between every pair of farms in a community, extract the vertices from those paths, and then incorporate those into the set of points used to define the concave hull. The result is a polygon that is concave as well in the dimension of time, meaning the paths with shortest commute times between any two farms never leaves the community.

In some cases, this may lead to overlapping community boundaries, though the amount of overlap should be minimal given the way that we cluster points, particularly when there are many densely distributed points. To some extent, of course, this sort of overlap is both unavoidable and expected as even in the modern world there are disagreements about the exact locations of legal and political boundaries.

It should be noted, too, that we also dilate each concave hull, briefly expanding it with a positive buffer, then shrinking it with a negative buffer having a slightly smaller size. This should have only a minimal effect on the total area of each polygon and also serves to smooth out noise along the edges.

3.6 Least-cost path analysis

The key to our path analysis is to take information about the terrain in our study regions and convert it first into meaningful estimates of hiking speed and then into meaningful estimates of travel time. Fortunately, attempts to model hiking speed across a range of slopes have advanced considerably in recent years, most notably in Campbell et al. (2019). Drawing on a large, crowd-sourced fitness tracker data set consisting of approximately 420,000 individual hikes, jogs, and trail runs from nearly 30,000 individuals in Salt Lake City, UT, Campbell et al. (2019) offer what is

easily the most empirically robust estimate of hiking speed to date. An interesting feature of their analysis - though one less important for the current study - is its sensitivity to variance in the physical and athletic prowess of individuals, which allows for estimates of different speeds at a given slope for different deciles of the sample. For simplicity, our application of Campbell's hiking function relies on coefficients from the first decile of individuals in their sample (the slowest ten percent of hikers) and assumes that everyone in the areas of our analysis were largely equal in their walking speed and endurance. This makes their walking speeds comparable to, but not the same as, what would be estimated by Tobler's hiking function ([Tobler 1993](#)).

We take the following steps to calculate least-cost travel times in our study regions. We first download a digital elevation model for each study area from the USGS National Elevation Dataset using the FedData package in R ([Bocinsky 2020](#)). The original resolution of these gridded data is 30 m, but we aggregate grid cells by a factor of 3, making them closer to 100 m in resolution. Aggregation is a familiar speed-up technique for least-cost analysis on large grids containing tens of millions of cells. We believe the resulting loss in precision owing to such aggregation is reasonable given that our goal is not so much to find specific paths as it is to determine relations between the locations that these paths connect.

The next step is to convert the aggregated elevation data into a graph in which each node represents a grid cell and each edge represents a straight line between adjacent grid cells in the Moore neighborhood (the eight adjacent grid cells). Edge weights are first defined by the slope between adjacent grid cells. Those slope estimates are then fed to Campbell's hiking function to derive estimates of hiking speed along each edge. The inverse of the hiking speed (the pace) is then multiplied by the distance between adjacent grid cell centers to estimate travel time.

Next, we associate each farm and community center with the grid cell that contains it, and hence to a specific node in the graph. In effect, we treat all the sites falling into a grid cell as a single site. We then apply a bi-directional form of Dijkstra's algorithm to the entire set of nodes with associated sites, as this algorithm significantly reduces processing time when the desired outcome involves every combination of start and end points ([Sturtevant and Felner 2018](#)). This gives us two

datasets: (i) a dense distance matrix with least cost travel-times from all farms and community centers to all farms and community centers and (ii) a set of linestring geometries whose vertices are the grid cells traversed on shortest paths.

For simplicity, we average travel times to and from each origin and destination point, thus making the distance matrix symmetric, which effectively treats the round-trip commute as the relevant cost-distance. All of the clustering steps in our algorithm are implemented as operations on this matrix. All analyses, including the cost-distance modeling and our implementation of the algorithm, are conducted in the R programming language and environment ([R Core Team 2023](#)). For more details, please see Supplementary Materials, which include all code necessary to reproduce these analyses and the figures in the text.

4 Results

The goals of this analysis are (i) to identify dispersed farming communities based on the distribution of farmsteads and their proximity to community centers and then (ii) to regionalize those communities by giving them meaningful boundaries. This was done by developing a simple algorithm inspired by the concept of a CBSA combined with previous work by Varien ([1999b](#)) and Reese, Glowacki, and Kohler ([2019](#)). Using parameter values D-join = 30 minutes, D-max = 1 hour, and P = 0.8, our algorithm identified 85 communities in CMV, 55 communities in NRG, and 21 communities in CIB (see Figure 2). The mean areas of these communities are 11.56 km^2 ($\sigma = 7.54$, $\eta = 10.5$) in CMV, 7.20 km^2 ($\sigma = 5.56$, $\eta = 5.86$) in NRG, and 8.80 km^2 ($\sigma = 7.66$, $\eta = 5.93$) in CIB; the mean room counts (for all community centers and farms) are 781.62 ($\sigma = 991.89$, $\eta = 420.0$) in CMV, 627.73 ($\sigma = 594.33$, $\eta = 443.0$) in NRG, and 747.71 ($\sigma = 573.92$, $\eta = 555.0$) in CIB; and the mean room densities are 74.92 ($\sigma = 67.1$, $\eta = 52.66$) in CMV, 138.58 ($\sigma = 150.07$, $\eta = 92.5$) in NRG, and 158.04 ($\sigma = 179.00$, $\eta = 92.32$) in CIB. We emphasize that the means in several cases deviate to a considerable degree from median values (denoted above with the Greek letter η), indicating skew in the data and the potential presence of large outliers. This is illustrated

in the boxplots in Figure 4, which show the distributions of these and other variables of interest that can be derived from our analysis and also used to evaluate its merits (see also Table 2).

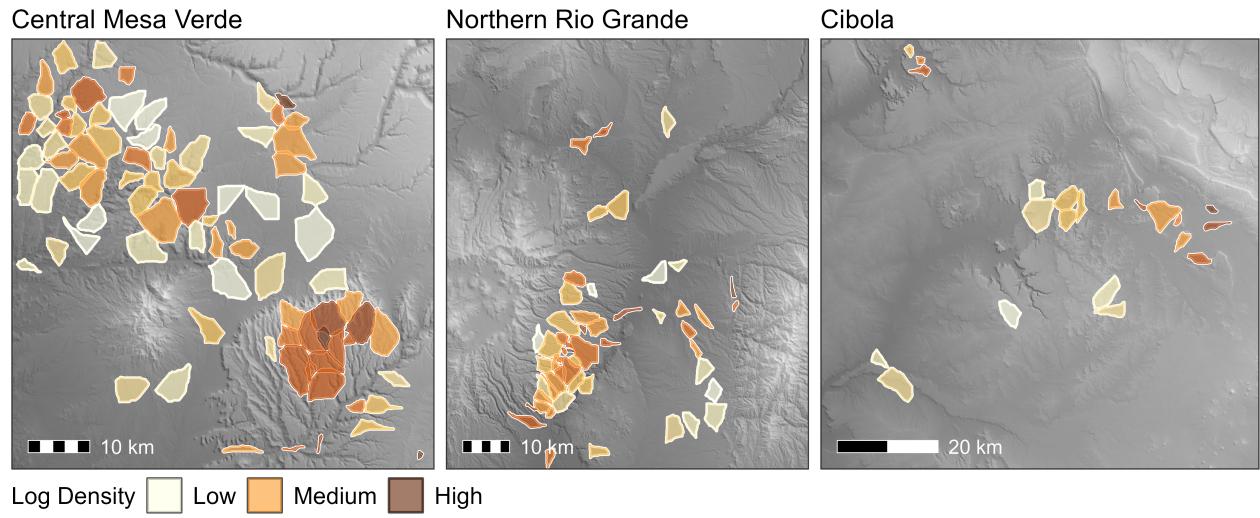


Figure 2. Map showing the size, shape, and spatial distribution of community polygons. Color represents the log density of rooms (the number of rooms per square kilometer), with lighter colors indicating lower values and darker colors indicating higher values. Note that densities in each region were re-scaled to have the same range, so this figure cannot be used to evaluate absolute differences between regions!

In addition to showing the community polygons, Figure 2 also offers a simple demographic profile for each region based on the spatial distribution of all rooms (their count per square kilometer) across communities irrespective of time period. This provides a simple demonstration of the utility of this method. For visualization purposes, the log of the density is used, as it helps to normalize the distribution and, in particular, to rein in large outliers that swamp the variance. It is important to note, too, that values in each region are re-scaled to be in the same range, so the figure cannot be used to estimate absolute differences between regions. Still, the relative differences within each region can be enlightening, but even they must be interpreted with caution.

Not surprisingly, virtually all of the Mesa Verde cuesta communities show high densities. The community that includes Goodman Point and Shield Pueblo in the central area of the CMV region just north of Ute Mountain also has a relatively high density, as do the communities in the far northwest that include Turkey House, Lowry, and Cottonwood Ruin. The Dolores area community around Grass Mesa Village in the northeast is also high density. In the NRG, the highest density

Table 2. Community Summaries

	Area (km2)	Rooms (N)	Farms (N)	Centers (N)	Commute (mins)	Room Density (N/km2)
CMV	η : 10.5	: 420	: 28	: 1	: 21.61	: 52.66
	μ : 11.56	: 781.62	: 59.29	: 6.01	: 22.35	: 74.92
	σ : 7.54	: 991.89	: 85.78	: 13.33	: 8.46	: 67.1
NRG	η : 5.86	: 443	: 20	: 2	: 16.59	: 92.5
	μ : 7.2	: 627.73	: 39.53	: 2.62	: 17.41	: 138.58
	σ : 5.56	: 594.33	: 37.63	: 2.75	: 7.21	: 150.07
CIB	η : 5.93	: 555	: 19	: 2	: 18.52	: 92.32
	μ : 8.8	: 747.71	: 28.24	: 2.29	: 19.64	: 158.04
	σ : 7.66	: 573.92	: 21.96	: 1.55	: 9.11	: 179

community is a narrow band that lies in a shallow canyon running south from Cundiyo on the east side of the valley. The thin east-west running community near San Ildefonso Pueblo is also high density. Note that it crosses the Rio Grande, following the contours of the Pojoaque River drainage. Just west of there on the Pajarito Plateau, there is a high density community that includes the Sandia center. South of Sandia there are two additional high density communities, one near the center of the Pajarito Plateau and another further south along the Chiquito Rio northwest of Cochiti Lake. In the CIB area, the three densest communities can be found in the far eastern edge of the study area near El Morro and Tinaja. One dense community just to the west of there includes the Ramah School center. Another lies in the far northwest near Manuelito.

5 Discussion

Our approach offers several advantages over previous efforts to define dispersed community territories. First, it handles variation in survey coverage by seeking to define communities we have evidence for, rather than all communities that may once have existed. Second, it handles variation in the degree to which surveyors have lumped or split architectural remains in defining sites ([Keith W. Kintigh 2003](#)). This is a special problem in the CMV. The tradition of site survey recording within the Mesa Verde cuesta has generally involved treating each room block as a separate site, even when there is a continuous artifact scatter between them. The tradition in the Montezuma Valley, in contrast, has been to include all room blocks within a single artifact scatter as part of

a single site. Our method is robust to these differences in site definition, enabling properties of the resulting communities to be compared more directly across these areas. Finally, our method focuses on travel time, a strong regularity in human affairs, and can be applied to landscapes with dramatically different topographies. The resulting community territories take obvious natural topographic barriers to interaction into account, but also allow territories to cross such boundaries when this is suggested by the data themselves.

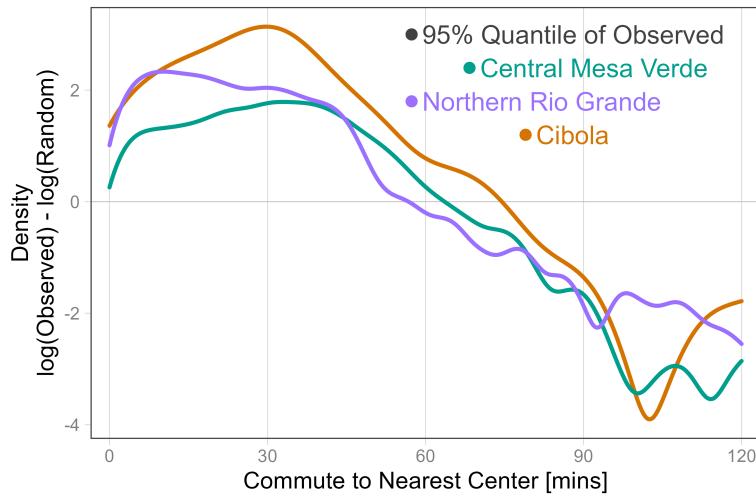


Figure 3. Figure shows the difference between the log density of commute times (in minutes) from observed farms and random locations to the nearest community center for each region, along with the location of their 95% quantile (this is just the 95% quantile for observed farms, not random points).

Because our algorithm requires that we calculate the commute time between each farm and community center, we can use it to evaluate potential spatial interactions between them. Figure 3 tries to get at this idea by showing the difference between (i) the log probability density of commute times from observed farms to their nearest centers and (ii) the log probability density of commute times from random locations to their nearest centers, along with the 95% quantiles of the observed distributions. This is analogous to what Reese, Glowacki, and Kohler (2019) refer to as the “null difference cost distance.” One can interpret it as showing where commute times differ from what one might expect were farm sites distributed at random across the landscape, making it similar to the Kullback-Leibler Divergence - basically, what a point process model like MaxEnt tries to estimate (Elith et al. 2011; Merow, Smith, and Silander Jr 2013). Values much larger than zero

indicate that there are more farms located at those commute times from centers than you would expect by chance. Conversely, values less than zero indicate that there are less farms located at those locations than one might expect by chance. In either case, these differences tell us that there is positive spatial covariance between farmsteads and nearby centers at commute times of less than approximately one hour (the value we assigned to $D - \max$).

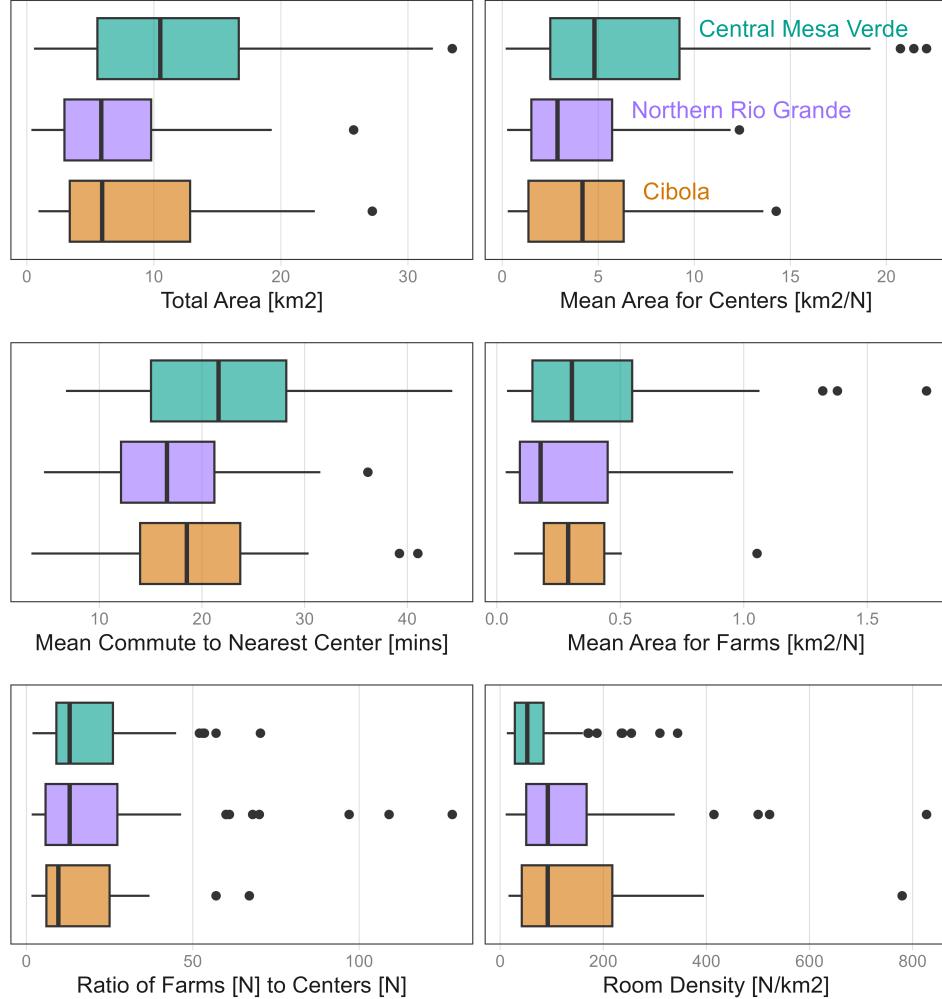


Figure 4. Figure shows for each region the distribution of relevant variables across communities.

The shape of the responses are not, however, identical. We attribute this to three factors: sampling intensity, local topography, and regional settlement patterning. For instance, the large spike in the CIB region at or around the 30 minute point along with the longer tail (and the larger 95% quantile) is at least partially owing to the fact that block survey in that area is limited to a few disconnected locations. The area is also much flatter than in the CMV and NRG regions, making

the background commute distribution much more uniform, at least within the surveyed areas. We probably could have accounted for the limited sampling by sampling random background points closer to the observed site locations, but that is not really the point of this exercise.

Perhaps more interesting though is the fact that these regions represent different settlement patterns over different periods of time. In the CMV area, the pattern of settlement was largely one of growth and dispersal, with local populations reaching an asymptote after which they spread out into neighboring areas (Schwindt et al. 2016). Nearly all centers in the CMV area arose within farming clusters, but they were relatively small compared to those in the other study areas, and occurred only to a limited degree in the period leading up to the Great Drought in 1280 CE (and to a lesser degree the period leading up to and including much of the 10th century), with no large aggregations occurring afterwards (Varien 1999b; Glowacki and Ortman 2012; Ortman 2016b). In the NRG region, communities had a few dispersed farms early on, more in the period immediately following 1280 (mostly concentrated in the northern Pajarito Plateau), and then very large aggregations after that (Ortman 2016a, 2016b). Communities in the CIB region had almost no dispersed farming for most of the sequence, mostly skipping over that pattern and moving straight into large aggregations as populations surged in the period after 1280 (Keith W. Kintigh 1985; Schachner 2012; Peeples 2018), as suggested by the relatively small ratio of farms to centers (see Figure 4). As in the CMV area, aggregations in the NRG and CIB regions sometimes developed within dispersed farming communities, but unlike in the CMV, many also seem to have been established in areas away from where dispersed farms had been previously located (Schachner 2015; Ortman 2016b). This would explain why our algorithm dropped 29 of the 77 centers in the CIB region and 37 of the 181 centers in the NRG but only 1 of the 512 centers in the CMV.

It should be emphasized that this is perhaps the biggest limitation of our algorithm, that it simply cannot define community boundaries when the relationship between community centers and dispersed farms is not obvious, either because the center is geographically isolated from the farms or because there is no center at all. We note, however, that our concept of a community center does

not technically require the presence of an archaeological site. That is just our way of identifying the locations of centers in the archaeological record.

In (Reese, Glowacki, and Kohler 2019), the area of the community is allowed to expand and contract, with the average area per farm remaining relatively constant on the Mesa Verde cuesta, in the range 0.15 to 0.27 km² per farm. In our algorithm, the total area of the community remains fixed, but the density is allowed to change. With the tuning parameter values we chose, the algorithm put about 50% of CMV communities in the range 0.2 to 0.55 km² per farm, with similar spreads for CIB and NRG. That, however, represents a minimum possible level in our analysis and would likely be greater as only a proportion of the farms were actually occupied at a given time. Still, this means that the communities we have defined are more likely to meet the 0.14 km² acre threshold required for a household to subsist on maize (Benson 2011; Bocinsky and Varien 2017). But, when we look at the area per room (the inverse of the room density), the picture is a little different. For the Mesa Verde region, the median area per room is roughly 0.04 km², which puts it at the upper end of the range reported in (Reese, Glowacki, and Kohler 2019). In the NRG and CIB regions, however, the median area per room is closer to 0.01 km², with a first quartile closer to 0.005 km². Again, these are minimum thresholds assuming every room is occupied at the exact same time, so there was likely more elbow room at any given time. Still, it does suggest that the area for farming in those regions may have been more circumscribed than in the CMV, which is also indicated by the boxplots showing total area in Figure 4. That would imply that levels of intensification (*sensu* Boserup 1965) in those regions were likely greater, too, with aggregation and social integration being one way that intensification probably played out.

6 Conclusion

The primary goal of this analysis was to derive the boundaries of dispersed maize farming communities from the point locations of archaeological sites. To that end, we developed a method that leverages the difference between farmsteads and community centers to cluster sites into communities and define the boundaries of those communities. This was inspired by the concept of a CBSA,

along with important contributions from Varien (1999b) and Reese, Glowacki, and Kohler (2019). An important difference between our approach and that of (Reese, Glowacki, and Kohler 2019) is that our community boundaries are static, though crucially this allows us to look at changes in density through time. So, the next step in our analysis will be to determine the number of rooms inhabited during each time step. A variety of methods are possible for this, but in future studies we intend to utilize uniform probability density analysis of pottery assemblages from each site (Ortman 2016a) to apportion rooms to time slices. This will allow us to capture the changing properties of both dispersed and aggregated communities using consistent measurements. The properties that can be compared include:

- Community area (agricultural land in the case of aggregated settlements);
- Settlement area (village area for aggregated communities, community area for dispersed communities);
- Community population by time step;
- Average distance between individuals (square root of community area / 2 for dispersed communities, and square root of village area / 2 for aggregated communities);
- Environmental conditions (captured by importing community territories into the paleoCAR niche modeling tool in <https://www.openskope.org/>);
- Topography (the intersection of community territories with a DEM);
- Properties of civic architecture found within the community, by time step; and
- Nearest neighbor distances to adjacent community centers.

The ability to compare these properties for communities of varying scales and densities is crucial for evaluating the relative costs and benefits of these properties for human welfare.

7 Acknowledgements

Portions of this research were supported by a grant from the National Science Foundation (#2213921, to Ortman).

References

- Adler, Michael Allan. 1990. “Communities of Soil and Stone: An Archaeological Investigation of Population Aggregation Among the Mesa Verde Region Anasazi, AD 900–1300 (Order No. 9116107).” PhD thesis, University of Michigan. <https://www.proquest.com/dissertations-theses/communities-soil-stone-archaeological/docview/303851225/se-2>.
- Angel, Shlomo, Patrick Lamson-Hall, Alejandro Blei, Sharad Shingade, and Suman Kumar. 2021. “Densify and Expand: A Global Analysis of Recent Urban Growth.” *Sustainability* 13 (7). <https://doi.org/10.3390/su13073835>.
- Benson, Larry V. 2011. “Factors Controlling Pre-Columbian and Early Historic Maize Productivity in the American Southwest, Part 1: The Southern Colorado Plateau and Rio Grande Regions.” *Journal of Archaeological Method and Theory* 18 (1): 1–60. <https://doi.org/10.1007/s10816-010-9082-z>.
- Bettencourt, Luis M. A. 2021. *Introduction to Urban Science: Evidence and Theory for Cities as Complex Systems*. Cambridge, MA: The MIT Press.
- Birch, Jennifer. 2013. *From Prehistoric Villages to Cities: Settlement Aggregation and Community Transformation*. New York, NY: Routledge.
- Bocinsky, R. Kyle. 2020. *FedData: Functions to Automate Downloading Geospatial Data Available from Several Federated Data Sources*. Manual. <https://CRAN.R-project.org/package=FedData>.
- Bocinsky, R. Kyle, A Gillreath-Brown, K Kintigh, A Kinzig, T Kohler, A Lee, B Ludascher, T McPhillips, C Nguyen, and C Pritchard. 2022. “Synthesizing Knowledge of Past Environments (SKOPE).” Web Application. SKOPE: Synthesizing Knowledge of Past Environments. 2022. <https://doi.org/10.5281/zenodo.6522974>.
- Bocinsky, R. Kyle, and Timothy A Kohler. 2014. “A 2,000-Year Reconstruction of the Rain-Fed Maize Agricultural Niche in the US Southwest.” *Nature Communications* 5: 5618.

- Bocinsky, R. Kyle, Johnathan Rush, Keith W. Kintigh, and Timothy A. Kohler. 2016. "Exploration and Exploitation in the Macrohistory of the Pre-Hispanic Pueblo Southwest." *Science Advances* 2 (4): e1501532. <https://doi.org/10.1126/sciadv.1501532>.
- Bocinsky, R. Kyle, and Mark D. Varien. 2017. "Comparing Maize Paleoproduction Models with Experimental Data." *Journal of Ethnobiology* 37 (2): 282–307. <https://doi.org/10.2993/0278-0771-37.2.282>.
- Boserup, Ester. 1965. *The Conditions of Agricultural Growth: The Economics of Agrarian Change Under Population Pressure*. Chicago, Illinois: Aldine.
- Campbell, Michael J., Philip E. Dennison, Bret W. Butler, and Wesley G. Page. 2019. "Using Crowdsourced Fitness Tracker Data to Model the Relationship Between Slope and Travel Rates." *Applied Geography* 106: 93–107. <https://doi.org/10.1016/j.apgeog.2019.03.008>.
- Drennan, Robert D, C Adam Berrey, and Christian E Peterson. 2015. *Regional Settlement Demography in Archaeology*. Bristol, CT: Eliot Werner Publications.
- Duranton, Gilles, and Diego Puga. 2020. "The Economics of Urban Density." *Journal of Economic Perspectives* 34 (3): 3–26. <https://doi.org/10.1257/jep.34.3.3>.
- Duwe, Samuel, and Kurt F. Anschuetz. 2013. "Ecological Uncertainty and Organizational Flexibility on the Prehispanic Tewa Landscape: Notes from the Northern Frontier." In *From Mountain Top to Valley Bottom: Understanding Past Land Use in the Northern Rio Grande Valley, New Mexico*, edited by B. J. Vierra. Salt Lake City, UT: University of Utah Press.
- Duwe, Samuel, B. Sunday Eiselt, J. Andrew Darling, Mark D. Willis, and Chester Walker. 2016. "The Pueblo Decomposition Model: A Method for Quantifying Architectural Rubble to Estimate Population Size." *Journal of Archaeological Science* 65 (January): 20–31. <https://doi.org/10.1016/j.jas.2015.10.011>.
- Elith, Jane, Steven J. Phillips, Trevor Hastie, Miroslav Dudík, Yung En Chee, and Colin J. Yates. 2011. "A Statistical Explanation of MaxEnt for Ecologists." *Diversity and Distributions* 17 (1): 43–57. <https://doi.org/10.1111/j.1472-4642.2010.00725.x>.

- Ester, Martin, Hans-Peter Kriegel, Jorg Sander, and Xiaowei Xu. 1996. “A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise.” In *Proceedings of the 2nd ACM International Conference on Knowledge Discovery and Data Mining (KDD)*, 226–31. AAAI Press. <https://cdn.aaai.org/KDD/1996/KDD96-037.pdf>.
- Gauthier, Rory P., and Cynthia Herhahn. 2005. “Why Would Anyone Want to Farm Here?” In *The Peopling of Bandelier: New Insights from the Archaeology of the Pajarito Plateau*, edited by R. P. Powers, 27–34. Santa Fe, NM: School of American Research Press.
- Glowacki, Donna M., and Scott G. Ortman. 2012. “Characterizing Community-Center (Village) Formation in the VEP Study Area.” In *Emergence and Collapse of Early Villages: Models of Central Mesa Verde Archaeology*, edited by Timothy A. Kohler and Mark D. Varien. Berkeley, CA: University of California Press.
- Gyucha, Attila. 2019. *Coming Together: Comparative Approaches to Population Aggregation and Early Urbanization*. Albany, NY: State University of New York Press.
- Kintigh, Keith W. 1985. *Settlement, Subsistence, and Society in Late Zuni Prehistory*. Tucson, AZ: University of Arizona Press.
- Kintigh, Keith W. 2003. “Coming to Terms with the Chaco World.” *KIVA* 69 (2): 93–116. <https://doi.org/10.1080/00231940.2003.11758487>.
- Lekson, Stephen H. 1989. “Kivas?” In *The Architecture of Social Integration in Prehistoric Pueblos*, edited by William D. Lipe and Michelle Hegmon, 161–67. Occasional Papers 1. Cortez, CO: Crow Canyon Archaeological Center.
- Lipe, William D. 1989. “Social Scale of Mesa Verde Anasazi Kivas.” In *The Architecture of Social Integration in Prehistoric Pueblos*, edited by William D. Lipe and Michelle Hegmon, 53–71. Occasional Papers 1. Cortez, CO: Crow Canyon Archaeological Center.
- Lipe, William D., and Michelle Hegmon. 1989. “Historical Perspectives on Architecture and Social Integration in the Prehistoric Pueblos.” In *The Architecture of Social Integration in Prehistoric Pueblos*, edited by William D. Lipe and Michelle Hegmon, 15–34. Occasional Papers 1. Cortez, CO: Crow Canyon Archaeological Center.

- Mahoney, Nancy M., Michael A. Adler, and James W. Kendrick. 2000. "The Changing Scale and Configuration of Mesa Verde Communities." *KIVA* 66 (1): 67–90. <https://doi.org/10.1080/00231940.2000.11758422>.
- Marchetti, C. 1994. "Anthropological Invariants in Travel Behavior." *Technological Forecasting and Social Change* 47 (1): 75–88. [https://doi.org/10.1016/0040-1625\(94\)90041-8](https://doi.org/10.1016/0040-1625(94)90041-8).
- Merow, Cory, Matthew J. Smith, and John A. Silander Jr. 2013. "A Practical Guide to MaxEnt for Modeling Species' Distributions: What It Does, and Why Inputs and Settings Matter." *Ecography* 36 (10): 1058–69. <https://doi.org/10.1111/j.1600-0587.2013.07872.x>.
- Mills, Barbara J., Sudha Ram, Jeffery J. Clark, Scott Ortman, and Matthew A. Peeples. 2020. "cyberSW Version 1.0." *Archaeology Southwest*.
- Muenchrath, Deborah A, Maya Kuratomi, Jonathan A. Sandor, and Jeffrey A. Homburg. 2002. "Observation Study of Maize Production Systems of Zuni Farmers in Semiarid New Mexico." *Journal of Ethnobiology* 22 (1): 1–33.
- Ortman, Scott G. 2016a. "Uniform Probability Density Analysis and Population History in the Northern Rio Grande." *Journal of Archaeological Method and Theory* 23 (1): 95–126. <https://doi.org/10.1007/s10816-014-9227-6>.
- . 2016b. "5 Discourse and Human Securities in Tewa Origins." *Archaeological Papers of the American Anthropological Association* 27 (1): 74–94. <https://doi.org/10.1111/apaa.12075>.
- Peeples, Matthew A. 2018. *Connected Communities: Networks, Identity and Social Change in the Ancient Cibola World*. Tucson, AZ: University of Arizona Press.
- Peeples, Matthew A., and Gregson Schachner. 2012. "Refining Correspondence Analysis-Based Ceramic Seriation of Regional Data Sets." *Journal of Archaeological Science* 39 (8): 2818–27. <https://doi.org/10.1016/j.jas.2012.04.040>.
- Peterson, Christian E., and Robert D. Drennan. 2005. "Communities, Settlements, Sites, and Surveys: Regional-Scale Analysis of Prehistoric Human Interaction." *American Antiquity* 70 (1): 5–30. <https://doi.org/10.2307/40035266>.

- R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Manual. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Reese, Kelsey M., Donna M. Glowacki, and Timothy A. Kohler. 2019. “Dynamic Communities on the Mesa Verde Cuesta.” *American Antiquity* 84 (4): 728–47. <https://www.jstor.org/stable/26818404>.
- Schachner, Gregson. 2012. *Population Circulation and the Transformation of Ancient Zuni Communities*. Tucson, AZ: University of Arizona Press.
- . 2015. “Ancestral Pueblo Archaeology: The Value of Synthesis.” *Journal of Archaeological Research* 23: 49–113.
- Schwindt, Dylan M., R. Kyle Bocinsky, Scott G. Ortman, Donna M. Glowacki, Mark D. Varien, and Timothy A. Kohler. 2016. “The Social Consequences of Climate Change in the Central Mesa Verde Region.” *American Antiquity* 81 (1): 74–96. <https://doi.org/10.7183/0002-7316.81.1.74>.
- Sturtevant, Nathan, and Ariel Felner. 2018. “A Brief History and Recent Achievements in Bidirectional Search.” *Proceedings of the AAAI Conference on Artificial Intelligence* 32 (1). <https://doi.org/10.1609/aaai.v32i1.12218>.
- Tobler, Waldo R. 1993. “Three Presentations on Geographical Analysis and Modeling: Non-isotropic Geographic Modeling, Speculations on the Geometry of Geography And, Global Spatial Analysis.” 93 (1). National Center for Geographic Information and Analysis.
- USOMB, United States Office of Management and Budget. 2021. *2020 Standards for Delineating Core Based Statistical Areas*. *Federal Register Vol 86 Pp 37770-37778*. Vol. 86. <https://www.federalregister.gov/documents/2021/07/16/2021-15159/2020-standards-for-delineating-core-based-statistical-areas>.
- Varien, Mark D. 1999a. “Regional Context: Architecture, Settlement Patterns, and Abandonment.” In *The Sand Canyon Archaeological Project: Site Testing*, edited by Mark D. Varien. Cortez, CO: Crow Canyon Archaeological Center.
- . 1999b. *Sedentism and Mobility in a Social Landscape: Mesa Verde & Beyond*. Tucson, AZ: University of Arizona Press.

- . 2002. “Persistent Communities and Mobile Households: Population Movement in the Central Mesa Verde Region, A.D. 950 to 1290.” In *Seeking the Center Place: Archaeology and Ancient Communities in the Mesa Verde Region*, edited by Mark D. Varien and R. H. Wilshusen, 163–84. University of Utah Press.
- Varien, Mark D., Scott G. Ortman, Timothy A. Kohler, Donna M. Glowacki, and C. David Johnson. 2007. “Historical Ecology in the Mesa Verde Region: Results from the Village Ecodynamics Project.” *American Antiquity* 72 (2): 273–99. <https://doi.org/10.2307/40035814>.