Abstract

This paper highlights a spatial interaction entropy maximization (SIEM) model to reproduce and understand past human settlement hierarchy in Central Anatolia (CA) and the Khabur Triangle (KT) during the Middle Bronze Age (ca. 2000 – 1600 BC). We propose applying SIEM to understand which sites and areas would have become prominent in this period by using known archaeological sites as point data and textual evidence for calibration purposes. The model addresses to which extent general factors such as topography, transportation, or social-ecological advantages (e.g., environmental benefits, external contacts, religion, etc.) make locations attractive for trade and settlement and why some archaeological sites become more prominent than others in the period discussed. The modelling results have been checked against the observed data from archaeological survey carried out in the KT and CA in order to explain the best fits. The results show that geography and topography can explain the growth of some main urban centres. In other cases, settlement hierarchies are generally explained by different initial advantages (endogenous or exogenous) for each site. Similar patterns are evident in CA and the KT, where movement constraint is due to political landscapes fragmented into numerous competing polities. In the KT, movement is slightly more restricted than CA, suggesting that warfare intensity is higher in open plains than in hilly and mountainous landscapes. Furthermore, SIEM is demonstrated to be useful in characterizing areas of interactions on different spatial scales in CA and the KT. Overall, the results demonstrate the advantage of SIEM model to enable researchers to account for missing empirical data, to compare settlement and interaction dynamics in different regions and to explain which underlying factors are beyond the reproducible observed patterns.

Keywords

Spatial interaction, entropy maximization, settlement hierarchy, urban systems, Middle Bronze Age, Near East, city-states, networks.
Introduction

This article aims to understand regional settlement systems through the application of a spatial interaction model which can be helpful for understanding past human settlement hierarchy in the Khabur Triangle (KT) and in Central Anatolia (CA) during the Middle Bronze Age (ca. 2000-1600 BC; MBA). In this period, the distribution of settlement sizes in these regions was relatively broad, with numerous small and medium sized sites and only a few large sites. This settlement structure arguably reflects the actual political landscape in the early second millennium, which was divided into several independent city-states (for CA see Veenhof and Eidem 2008, 147-179; Barjamovic 2011, 6; Barjamovic et al. 2012, 48-50; for the KT see Charpin and Ziegler 2003; Veenhof and Eidem 2008, 290-321; Ristvet 2008 and 2012; Palmisano 2015). It is not well understood how such settlement size structures developed on the basis of inter and intra-regional interactions and socio-environmental factors. Therefore, a methodology is needed for understanding the causal logics behind past human settlement size dynamics. To achieve this, we propose applying a relatively novel method to predict which sites and areas would have become prominent in this period by using known archaeological sites as point data and historical information for calibration purposes. The modelling results can be checked against empirical results of archaeological surveys undertaken in the KT and in CA in the past decades. The model below addresses to which extent geography, transportation, external contacts, and socio-economic/environmental factors make locations attractive for trade and settlement and why some archaeological sites become relatively major urban centers in the period discussed. This includes how political and geographic constraints affect regional settlement transformations, while also accounting for uncertainty in the archaeological data. Furthermore, such a method, is also useful to trace networks of relationships between pairs of sites, and consequently to identify regional urban hubs and areas of interaction in a given region at different spatial scales.

This methodology builds on a series models, generally termed as spatial interaction entropy maximization (SIEM), that were originally introduced in the 1960s and 1970s in geography to forecast flows of goods and people in spatial systems (Wilson 1967, 1970, 2008, 2010; see Wilson 2012a for a recent overview), and then applied to archaeological settlements datasets in a series of academic papers published some twenty-five years ago (Rihill and Wilson 1987) and again in the last couple of years (Wilson 2012b; Altaweel 2013 and 2014; Bevan and Wilson 2013; Davies et al. 2014). These simulations have the advantage of explaining how general causal factors (e.g. population pressures, political and territorial divisions, topographical boundaries, etc.), that are difficult to isolate and quantify from the archaeological record, could have affected settlement expansion or contraction in a given geographic setting. Hence, the target of this paper is to present a simple simulation model that not only explores how major settlements emerge, but how such emergence develops at the expense of other sites and because of political circumstances or external factors affecting a region. At a more general level, the results demonstrate that a quantitative model is useful in explaining emergent urban settlement hierarchies across landscapes at different scales.

We first present background to the case studies in the section below. Then, we will introduce and explain the methodology of SIEM and structural dynamics modelling approaches. Subsequently, the modelling results, including outputs from two different possible scenarios, are provided. In particular, in the first scenario we will test the role of geography in shaping urban growth and settlement hierarchy, while in the second scenario we will assess what general factors (e.g. trade, economic and political relevance, etc.) are responsible for the growth of specific areas and/or urban centres and the decline of others. The results show different factors that may have catalysed or diminished urban population growth. The relevance of these results for understanding settlement in the KT and CA during the MBA (ca. 2000 – 1600 BC) are discussed. Finally, conclusions are drawn with regard to the methodology and its potential for understanding the development of settlement hierarchies.
Background

Case Studies and historical background

For the purpose of this project, two different well-defined regions have been chosen (Fig. 1). The first case study is the KT (Fig. 1b), an area located within the Syrian Jazira, measuring some 37,480 km² and extending between the Tigris and Euphrates rivers, bounded by what is today the Syrian/Iraqi border to the east, the Syrian and Turkish border to the north, the Jebel Sinjar and by the Jebel ‘Abd-al-Aziz to the south and the Khabur River to the west. The second case study is CA (Fig. 1a), a region covering a total area of about 200,000 km² between the Pontic Mountains to the north and the Taurus mountains to the south.

The choice of the KT and CA has been stimulated by an interest in how different geographical settings, specifically an open tableland versus a mountainous inland area with a large intermountain river valley respectively, contributed to the development of local settlement systems. The selection has also been influenced by the limited number of regions where a sufficiently high intensity of archaeological excavations and surveys has been conducted and by the need to provide a coherent framework for the settlement systems analysed, given several gaps in the survey record over Upper Mesopotamia and CA for the MBA (ca. 2000-1600 BC). The two areas, however, should not be viewed as fully isolated from each other as testified by the long-distance commercial system set up by the Assyrians in the early second millennium, if not before. Understanding settlement system dynamics in CA and KT during the early second millennium is of pivotal importance, as the first area was divided into several city-states and local monarchies hosting Old Assyrian commercial colonies (Barjamovic 2011, 6), while the second one became criss-crossed by long-distance commercial routes from Aššur to CA (Goetze 1953; Hallo 1964; Oguchi 1999 and Kolinski 2014 for a broad overview) and was characterized by the presence of the Old Assyrian karus Apum and Šubat-Enlil/Šehna, respectively during Kültepe’s lower town level II (ca. 1970-1835 BC) and level Ib (ca. 1835-17th c. BC) periods (see Eidem 2008, 32; Ristvet 2008; Veenhof and Eidem 2008; Barjamovic 2011). The situation changed in the MBA II (ca. 1800 – 1600 BC), when large and centralized territorial states imposed their authority upon numerous and weaker existing political entities. In this period the KT was part of Šamši-Adad I’s kingdom (ca. 1808 – 1776 BC) and subsequently of Zimri-Lim’s kingdom (ca. 1780-1758 BC). In particular, the KT played a prominent role in the international political scenario of the early 18th century BC as the seat of the capital of Šamši-Adad I’s kingdom (ca. 1808 – 1776 BC) at Šubat-Enlil/Šehna (Tell Leilan). In the second half of the 18th century, Anitta was able to impose his power over the southern half of CA, and the texts suggest he took the title of Great King (Barjamovic et al. 2012, 50). Nevertheless, the city-states remained the more stable and longest-lasting political unit, while the larger regional kingdoms were often politically fragile and could last only one generation or dynasty.
Fig. 1. Map showing the case studies and the archaeological surveys carried out in the Khabur Triangle and in central Anatoli
Archaeological excavations and surface surveys carried out across the KT and CA provide the bulk of data about the spatial location and extent of settlement at both regional and local scales, as well as about settlement occupation histories. Nevertheless, the actual available data can be problematic; site densities from surveys carried out in CA are far lower (ranges from 0.4 to 5 sites per 100 sq. km.) than those recorded in systematic and extensive regional surveys performed in the KT (around 10 or more sites per 100 sq. km; e.g. Ristvet 2005; Wright et al. 2007; Ur and Wilkinson 2008; Ur 2010) and just a few have been intensively carried out in the Konya Plain (Baird 1996-2002), in Paphlagonia (Matthews and Glatz 2009), in Gordion (Kealhofer 2005), in the Lower Euphrates basin (Özdoğan 1977), and around Boğazköy (see Fig. 1 and Tables 1-2 for a list of surveys carried out in the KT and CA). In addition, existing publications indicate only the overall extent of mounds but neither the size for a particular chronological phase nor the extent of the surrounding lower town.

Therefore, we can provide only very rough estimates about the empirical extent of MBA sites in the KT and CA, and any results derived from the analyses of the archaeological surveys data have to be interpreted cautiously, as constituting evidence only about the patterns exhibited by relatively large, sedentary farming communities. Nevertheless, the larger and smaller mounds do likely present themselves as relative proxies for sites that were possibly larger or smaller than surrounding settlements. Bearing in mind the above issues, SIEM overcomes the limits imposed by the intrinsic spatial and temporal uncertainty of the archaeological record.

In the KT, relevant survey data include: Meijer (1986), Eidem and Warburton (1996), Lyonnet (2000), Ristvet (2005), Wright et al. (2007), Ur and Wilkinson (2008), and Ur (2010; see Fig. 1b and Table 2). Other nearby surveys (Algaze, 1989; Wilkinson and Tucker, 1995; Ball, 2003) have been left out of the analysis, as these are not as continuous as the others. Within the KT, there are 439 sites that are occupied in the MBA (Fig. 1b). In the eastern KT, the Tell Leilan (Ristvet, 2005) area alone has 157 sites during the MBA. Here, the dominant role of Tell Leilan is clear, which had an area of ca. 90 ha with many surrounding small villages. The dense concentration of settlements around Tell Leilan may be related to its prominent political role as the capital city of Šamši-Adad I’s large territorial kingdom (ca. 1808-1776 BC), and later as the main town of the kingdom of Apum (second half of the 18th century; Charpin 1987; Ristvet 2008). Other major centres include Tell Farfara (ca. 70 ha) and Tell Muhammed Diyab (ca. 35 ha). Along the Wadi Jaghjagh, the main settlements were Tell Brak (ca. 25 ha) and Tell Barri (ca. 9 ha). The long-term political significance of this region may have brought a sort of political stability and security that encouraged dispersed small-scale settlement. In the western KT (i.e., west of the Jaghjagh River), settlements are nucleated and populations were likely concentrated in towns such as Chagar Bazar (ca. 9 ha), Tell Mozan (ca. 35 ha), and Tell Arbid (ca. 7 ha), with the surrounding territory largely devoid of smaller settlements. This suggests that populations were concentrated in few bigger towns surrounded by plains empty of villages, but full of nomads (Ristvet 2012; Ristvet and Weiss 2013, 263-265). Overall, far more settlements and greater diversity of site sizes are found in the east; this could be because the area had more favourable climatic conditions (Evans and Smith, 2006).

Within CA there are 440 sites that were occupied during the MBA (Fig. 1a and Table 1). Other nearby archaeological surveys have been left out of the analysis because these are not as continuous with the others and there are gaps in the archaeological dataset. The settlement system in the Anatolian central plateau is characterized by few large sites such as Kültepe (ca. 50 ha), Acemhöyük (ca. 55 ha), Bůgazköy (ca. 65 ha), Yassihöyük (ca. 25 ha), Varavan Höyük (ca. 25 ha), and Alişar Höyük (ca. 20 ha), with many surrounding small settlements.¹

¹The extents of the sites are just rough estimates based on the sizes of their mounds. Intensive archaeological surveys of the presumed lower town surrounding a mound have never been carried out in Anatolia.
### Table 1. List of archaeological surveys carried out in central Anatolia.

<table>
<thead>
<tr>
<th>Map no.</th>
<th>Season</th>
<th>Reference</th>
<th>Area (sq. km)</th>
<th>Total no. sites</th>
<th>no. MB sites</th>
<th>Sites density (x 100 sq. km)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2000</td>
<td>Bahar 2002</td>
<td>5,825</td>
<td>120</td>
<td>52</td>
<td>2.06</td>
</tr>
<tr>
<td>2</td>
<td>1962, 1965</td>
<td>Brown 1967</td>
<td>31,349</td>
<td>38</td>
<td>14</td>
<td>0.12</td>
</tr>
<tr>
<td>3</td>
<td>2005</td>
<td>Di Nocera 2008-09</td>
<td>1,034</td>
<td>82</td>
<td>29</td>
<td>7.93</td>
</tr>
<tr>
<td>4</td>
<td>1997-99</td>
<td>Dönmez 1999-2000, 2002</td>
<td>23,408</td>
<td>85</td>
<td>32</td>
<td>0.36</td>
</tr>
<tr>
<td>5</td>
<td>1958</td>
<td>French 1970</td>
<td>1,127</td>
<td>51</td>
<td>7</td>
<td>4.5</td>
</tr>
<tr>
<td>6</td>
<td>1993</td>
<td>Gülçür 1995</td>
<td>1,341</td>
<td>61</td>
<td>9</td>
<td>4.54</td>
</tr>
<tr>
<td>8</td>
<td>2008-10</td>
<td>Kulakoglu et al. 2009 - 2011</td>
<td>19,194</td>
<td>87</td>
<td>43</td>
<td>0.45</td>
</tr>
<tr>
<td>10</td>
<td>1997-2001</td>
<td>Matthews and Glatz 2009</td>
<td>7,737</td>
<td>337</td>
<td>19</td>
<td>4.35</td>
</tr>
<tr>
<td>12</td>
<td>1990</td>
<td>Omura 1992</td>
<td>58,847</td>
<td>53</td>
<td>36</td>
<td>0.09</td>
</tr>
<tr>
<td></td>
<td>1991</td>
<td>Omura 1993</td>
<td>6,899</td>
<td>30</td>
<td>11</td>
<td>0.43</td>
</tr>
<tr>
<td></td>
<td>1994</td>
<td>Omura 1996a-b</td>
<td>12,143</td>
<td>54</td>
<td>25</td>
<td>0.44</td>
</tr>
<tr>
<td></td>
<td>1995</td>
<td>Omura 1997</td>
<td>1,634</td>
<td>43</td>
<td>12</td>
<td>2.75</td>
</tr>
<tr>
<td></td>
<td>1996</td>
<td>Omura 1998</td>
<td>1,037</td>
<td>51</td>
<td>8</td>
<td>4.91</td>
</tr>
<tr>
<td></td>
<td>1999-2000</td>
<td>Omura 2000 and 2001a</td>
<td>6,152</td>
<td>66</td>
<td>18</td>
<td>1.07</td>
</tr>
<tr>
<td></td>
<td>2000</td>
<td>Omura 2001b</td>
<td>2,057</td>
<td>64</td>
<td>18</td>
<td>3.11</td>
</tr>
<tr>
<td></td>
<td>2001</td>
<td>Omura 2002</td>
<td>4,555</td>
<td>68</td>
<td>33</td>
<td>1.49</td>
</tr>
<tr>
<td></td>
<td>2002</td>
<td>Omura 2003</td>
<td>1,786</td>
<td>106</td>
<td>10</td>
<td>5.95</td>
</tr>
<tr>
<td></td>
<td>2005</td>
<td>Omura 2006</td>
<td>2,672</td>
<td>46</td>
<td>13</td>
<td>1.72</td>
</tr>
<tr>
<td></td>
<td>2006</td>
<td>Omura 2007a</td>
<td>3,529</td>
<td>40</td>
<td>13</td>
<td>1.13</td>
</tr>
<tr>
<td></td>
<td>2003-06</td>
<td>Omura 2007b</td>
<td>7,988</td>
<td>190</td>
<td>56</td>
<td>2.39</td>
</tr>
<tr>
<td></td>
<td>2007</td>
<td>Omura 2008</td>
<td>1,435</td>
<td>53</td>
<td>20</td>
<td>3.69</td>
</tr>
<tr>
<td>13</td>
<td>1975-76</td>
<td>Özdoğan 1977</td>
<td>369</td>
<td>80</td>
<td>24</td>
<td>21.68</td>
</tr>
<tr>
<td>15</td>
<td>1997-98</td>
<td>Senyurt 1998 and 1999</td>
<td>5,804</td>
<td>53</td>
<td>16</td>
<td>0.91</td>
</tr>
<tr>
<td>17</td>
<td>1988-89</td>
<td>Süel 1989 and 1990</td>
<td>1,440</td>
<td>28</td>
<td>9</td>
<td>1.94</td>
</tr>
<tr>
<td>18</td>
<td>1977</td>
<td>Yakar and Gürsan-Salzmann 1979</td>
<td>21,370</td>
<td>68</td>
<td>15</td>
<td>0.31</td>
</tr>
</tbody>
</table>

### Table 2. List of archaeological surveys carried out in the Khabur Triangle.

<table>
<thead>
<tr>
<th>Map no.</th>
<th>Season</th>
<th>Reference</th>
<th>Area (sq. km)</th>
<th>Total n. sites</th>
<th>n. MB sites</th>
<th>Sites density (x 100 sq. km)</th>
</tr>
</thead>
<tbody>
<tr>
<td>19</td>
<td>1988</td>
<td>Eidem and Warburton 1996</td>
<td>193</td>
<td>56</td>
<td>19</td>
<td>29.01</td>
</tr>
<tr>
<td>20</td>
<td>1989-1991</td>
<td>Lyonnet 2000</td>
<td>5,100</td>
<td>161</td>
<td>45</td>
<td>3.15</td>
</tr>
<tr>
<td>21</td>
<td>1976-77; 1979</td>
<td>Meijer 1986</td>
<td>2,296</td>
<td>290</td>
<td>152</td>
<td>12.63</td>
</tr>
<tr>
<td>23</td>
<td>1999-2001</td>
<td>Ur 2010</td>
<td>127</td>
<td>60</td>
<td>9</td>
<td>47.24</td>
</tr>
<tr>
<td>24</td>
<td>1997-98</td>
<td>Ur and Wilkinson 2008</td>
<td>454</td>
<td>83</td>
<td>7</td>
<td>10.28</td>
</tr>
<tr>
<td>25</td>
<td>2002-2003</td>
<td>Wright et al. 2006-2007</td>
<td>1,275</td>
<td>228</td>
<td>74</td>
<td>21.01</td>
</tr>
</tbody>
</table>
In this section, we will first show the results produced by performing rank-size analyses on the two whole study areas and assess comparatively any difference in the observed patterns between them (see Falconer and Savage 1995; Savage 1997). Table 3 and Fig. 2 provide a picture, for each study area, of the most central group of settlement sizes (in hectares). We can see that the midspreads of the KT (the 50% of values between the 3rd and the 1st quartiles; between 1 and 3.1 ha) and CA (between 1 and 2.8) match almost perfectly, and the values of median (1.7 vs. 1.5) differ just minimally. Fig. 3 shows a natural log-scale settlement size and hierarchies' distribution, ranking from largest to smallest, for each study area. A Whitney-Wilcoxon test shows (p-value = 0.09) that there is little difference between the KT and CA in terms of the variability of observed settlement sizes. At first glance, both size distributions appear similarly convex. Therefore, both results in the KT and in CA show a convex distribution for settlement size and rank. In a convex distribution there are many large settlements of roughly the same size in proportion to the number of small settlements. This could indicate population dispersion throughout a given area in sites that are of similar size and thus more competition and less integration between communities (Johnson 1980; Paynter 1982; Wossink 2009, 63-64. Hence, these results possibly indicate fragmented political landscapes of several competing settlement systems and polities little integrated politically and economically. The fragmented political situation in both areas in the MBA was the norm, where city-states fought with each other and shifted alliances for exerting their power over the surrounding areas (for CA see Veenhof and Eidem 2008, 147-179; Barjamovic 2011, 6; Barjamovic et al. 2012, 48-50; Palmisano 2014; for the KT see Charpin and Ziegler 2003; Veenhof and Eidem 2008, 290-321; Ristvet 2008 and 2012; Palmisano 2015).

<table>
<thead>
<tr>
<th>Region</th>
<th>no. sites</th>
<th>minimum</th>
<th>1st quartile</th>
<th>median</th>
<th>mean</th>
<th>3rd quartile</th>
<th>St. dev.</th>
<th>maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Central Anatolia</td>
<td>440</td>
<td>0.1</td>
<td>1</td>
<td>1.5</td>
<td>2.7</td>
<td>2.8</td>
<td>5.47</td>
<td>65</td>
</tr>
<tr>
<td>Khabur Triangle</td>
<td>439</td>
<td>0.1</td>
<td>1</td>
<td>1.7</td>
<td>3.2</td>
<td>3.1</td>
<td>6.62</td>
<td>90</td>
</tr>
</tbody>
</table>

Table 3. Summary of central tendency and dispersion of settlements size (ha) in central Anatolia and in the Khabur Triangle in the Middle Bronze Age.

Figure 2. Box and whisker plot of size (in hectares) of Middle Bronze Age settlements in the Khabur Triangle and in central Anatolia.
Spatial Interaction Models

Methodology

Methods of SIEM have been widely used to forecast urban economic or population growth in spatial systems under conditions of uncertainty (Wilson 1967 and 1970). These models have been used to describe not only urban growth in a given geographical context, but also on smaller scale settings such as the growth of modern retail outlets and particular areas within modern cities (Birkin and Heppenstall 2011; see Wilson 2012 for a broad overview), and for presence-only modelling of species distributions (see Philips et al. 2006). These methods combine Boltzmann’s equations from statistical physics and the ecological models of Lotka and Volterra (Wilson 2008). In this case, several factors such as distance, topography, socio-economic factors (e.g., economic and political relevance, religion, etc.), and movement are incorporated as generalized variables to explain urban transformations. These variables allow one to detect general factors responsible for the growth of specific areas and/or urban centres and the decline of others. Specifically, the aim is to produce simulations which predict the urban layout of a given spatial system, in order to explain under what dynamics certain sites may have acquired relative prominence. The validity of the model will be assessed by the goodness of fit and correspondence between the simulated outputs and the empirical archaeological and textual data.

Entropy maximizing models allow feedback and interaction between settlements and explain how the urban growth of some urban centres/areas may affect surrounding regions. Positive feedback allows major urban centres and regions to grow to a greater extent, while negative feedback diminishes the economic and social capital of other regions (Krugman et al. 1995). In pre-industrial societies, “pull” factors such as geography, environment, transport, economy, and social institutions may have contributed to positive feedback, enabling major urban centres to expand further and simultaneously diminish the population and economic potential of the surrounding regions or centres (Braudel 1995; Batty 2005; Wilson 2012a). However, varied socio-environmental factors could have played a more relevant role in the growth, stabilisation or decline of a given settlement. Put simply, the analysis we present uses a well-established formulation for
spatial interaction modelling, which suggests the general trends and factors that may have affected the urban development and the settlement hierarchies in our case studies. The spatial data required for the model are respectively 439 and 440 sites for the KT and CA. We have estimates of the extent of each site, based on the published archaeological surveys reports (see Tables 1 and 2). These estimated sizes are used to measure how well the simulated populations, used as a proxy to define the sites size in the SIEM, replicate the observed data in the two areas studied. The topographical data are represented by an Aster Global Digital Elevation Model (GDEM) of the study areas (ASTER 2015).

Model Structure

For the purposes of this chapter, a SIEM model of the type already used in other contexts (see Wilson 1967 and 1970; Harris and Wilson 1978; Altaweel 2013 and 2014; Bevan and Wilson 2013; Davies et al. 2014), has been applied to understand which general factors may have affected the growth or the contraction of settlements in the KT and in CA during the MBA (ca. 2000-1600 BC). Here, we define the following variables for each of the sites (see the Appendix for further details about the method):

- $X_i$ = population (e.g. people and/or goods) originating at a given site $i$;
- $Z_j$ = the size of a site or initial advantages/ attractiveness $j$, which regulates flow;
- $\alpha$ = return of attractiveness for site $j$ that leads to migration or movement of people;
- $\beta$ = willingness or travel capability of individuals to travel a given distance to a settlement;
- $d$ = the distance (i.e., cost of travel) between any two sites $i$ and $j$, normalised by the mean of all such distances and using cost surface;
- $e$ = the base natural log.

In summary, the variables above determine how much in-flow of people and/or goods to a specific site $j$ on the basis of its attractiveness ($\alpha$), willingness or ability to travel ($\beta$), and distance ($d$) in relation to the population ($X_i$) of a given settlement. It should be stated that population or flow is used as a relative measure rather than an absolute for sites (i.e., to measure how large one site is relative to others). More precisely, the distance ($d_{ij}$) between each pair of sites is modelled through a matrix of travel movement costs generated by considering the topography and the geographical features (e.g. hills, mountains, rivers, etc.) of a given study area that may have constrained movement (see Fontenari et al. 2005 for the algorithm used; Palmisano 2013, 774-781, for a broad discussion about modelling past human movement). Another way to define $d$ might also take into account social factors (e.g. political or territorial divisions) that may have affected the movement between settlements. The return of attractiveness of a site ($\alpha$) is a general variable used to determine the effect or impact of social-environmental factors (e.g., political, economic, religious, etc.) that made specific settlements more attractive than others (i.e., for migration or commerce). This variable $\alpha$ is set globally for all sites and specifies the scaling of utility as sites size ($Z_j$) varies. Another important factor is the initial size ($Z_j$) of sites ($j$) defining, for instance, each site’s initial advantages (e.g., military power, political dominance, religious prestige) or interaction with sites outside the study area, which would be manifested as an additional flow of goods and people for each site. For the case study purposes, this variable ($Z_j$) is also useful for modelling any sort of external trade contacts between the sites in the two study areas and other regions. Finally, the variable ($\beta$) defining the willingness, the freedom, and the capability of movement is critical to results. As $\beta$ increases, an individual’s preference to travel shorter distances increases for any reason, while as $\beta$ decreases, individuals are able or willing to travel longer distances. This variable may be used for determining general factors that may have favoured (e.g. roads, privileged pathways between settlements) or constrained (e.g. rivers, territorial or political boundaries, warfare) movement between settlements. Therefore, by using SIEM, the most likely set of flows ($S_{ij}$), where flow represents interaction between sites, is then found under specific values of generalized variables.
The model outputs are characterized by two steps: 1) the estimation of interaction flows among sites; 2) the determination of site population \(X\) by measuring its portion relative to simulation size \(Z\). The first step consists of establishing the utility of interaction between any pair of sites \(i\) and \(j\). This results as a cost/benefit calculation, where the benefit is a function of the size of \(j\) and the cost is the physical impedance (e.g., distance) between sites \(i\) and \(j\). At this stage, the variable \(\alpha\) indicates the scaling of utility as the size varies and \(\beta\) defines the strength of the negative effect of travel cost. Therefore, in a given spatial system, where the total flow originating at each site is known, SIEM methods are able to detect the most likely distribution of that total flow among other sites. Put simply, for any given site \(i\), the total out-flow \(X_i\) is shared between each other site \(j\) in proportion to the utility of interaction with \(j\) as perceived by \(i\). In this way, the model finds the flow \(S_{i,j}\) between any such pair of sites.

In the second step, the interaction flows are used to update each site’s size \((Z_j)\), which consists of determining the growth/decline of each site under the calculated set of flows. Thus, for each site \(j\) the total inward flow \((D_j)\) is found by summing \(S_{i,j}\) over all \(i\), and then this value is compared with the site’s current size \(Z_j\). If the total flow is less than \(Z_j\), the current size is unsustainable and the site shrinks, whereas a surplus of inward flow leads to the growth of \(j\). Having made these calculations and adjusted \(X_i\) and \(Z_j\) accordingly, the model proceeds to the next time-step and the process is repeated. At the end of the simulation, the portion of \(X\) for settlements is compared to estimated portion of site size from surveys. In other words, population is used as a proxy for comparing to survey site size and to measure how well model parameters enable simulated population to match survey results. For any individual simulation, we run the model for a set number of discrete steps \(\delta t\).

The flow \(S_{i,j}\) between each pair of nodes \(i\) and \(j\) is calculated using the following formula:

\[
S_{i,j} = X_i \frac{Z_j^\alpha e^{-\beta d_{ij}}}{\sum_k Z_k^\alpha e^{-\beta d_{ik}}}
\]

(1)

These flows are summed to give the total incoming flow \(D_j\) to each site \(j\):

\[
D_j = \sum_i S_{i,j}
\]

(2)

This incoming flow is used to calculate \(Z_j\) at the next time step, with \(\gamma\) used to control the speed of change and \(k\) a constant that can be used to scale \(Z_j\), \(Z_j^{[t+\delta t]}\), using:

\[
Z_j^{[t+\delta t]} = Z_j^t + \epsilon (D_j - kZ_j^t)
\]

(3)

Next, \(X_i^{[t+\delta t]}\) for the following time step is determined by taking the corresponding \(Z_i^{[t+\delta t]}\) value, normalized for the total of \(Z_i^{[t+\delta t]}\) for all sites, and rescaling \(n\) so that sum of all \(X_i^{[t+\delta t]}\) continue to have the same mean as the simulation start and population is adjusted for the next simulation time for each site \(i\):
Then the model goes back to (1) for the next time step and continues until the end of the simulation. This mostly follows Davies et al. (2014), with this method being slightly different. For the given model, simulations are run for 100 time ticks, which is enough time to determine what the results of interactions leads to in settlement structures, as the model results reach equilibrium at that point. It is important to point out that the SIEM is a deterministic model, which simply means that unique input leads to unique output for each model run. In other words, the result is established straightforwardly from a series of initial parameters, so that there are no random elements in the model and the simulation can be performed just once for each parameter setting.

Therefore, we will make use of simulations in order to explore any possible variation in model outputs based on parameter choices, manipulation of the underlying dataset, and synthesis of results. More precisely, the model involves three general parameters ($\alpha$, $\beta$, and $Z_j$) that can be modified to give different sets of initial conditions. Results are then able to provide insights about how human settlement hierarchies may have developed in the KT and CA in the early second millennium BC, giving us a way to compare the regions. Each simulation will provide as observable output, the final site population ($X$), upon which to base the following assessments: 1) comparison between the numerical estimates of simulated values and the observed site’s size by using statistical tests; 2) the extent to which the largest sites in the empirical data are found to have a high population in the simulated runs; 3) mapping of outputs and evaluation of their historical validity.

**Nystuen-Dacey (N - D) graph**

One of the outputs from SIEM is a Nystuen-Dacey graph (N-D), which derives from the simulated flow matrix $S_{ij}$ (goods and people) between $i$ and $j$ sites (Nystuen and Dacey 1961). These relationships represent the subordination of some sites with respect to others, as determined by the structure of their interaction. Hence, for each SIEM’s simulation, the construction of the N-D graph ($G$) follows the steps below:

1. For each site $i$, the destination of its largest flow is found; that is, $j$ for which $S_{ij}$ is maximal.
2. If $j$ is a smaller site than $i$ (so $i$’s main flow is to a smaller site), $i$ is regarded as being independent and no subordination relation is given.
3. If, on the other hand, $j$ is larger than $i$, then $i$ is determined to be subordinate to $j$ and a directed edge from $i$ to $j$ is added to $G$.

After carrying out this process, we have a graph of sites, for which each site is either independent or has exactly one link to another site (see Fig. 4). A Nystuen-Dacey (N-D) graph is useful for identifying the “degree of contact between city pairs and it provides a quantitative basis for grouping cities” (Nystuen and Dacey 1961, 29). This allows a practitioner to assess the direction and magnitude of flows $S_{ij}$ (of goods and people) both on local and regional scales and to identify dominant nodal centres that are the main collectors of flows from the surrounding hinterland (Nystuen and Dacey 1961, 31). In other words, a N-D graph provides a visual and quantitative support to trace the spatial distribution of flows between pairs of sites and consequently to identify areas and sub-areas of interactions in a given region. We will, therefore, apply graph theory centrality measures (e.g. degree centrality, closeness centrality, betweenness centrality) to N-D network to assess the network centrality of settlements, detect main regional urban hubs and the interaction with their respective surrounding rural hinterlands on smaller local scales. In addition, a Markov cluster algorithm (MCL) will be applied to N-D network in order to detect clustering of

$$X_i^{t+\Delta} = n \frac{Z_{ij}}{\sum_k Z_{kj}^{t+\Delta}}$$  \hspace{1cm} (4)
nodes and, therefore, areas and sub-areas of interaction in our two study areas. This algorithm will allow us to turn the graph into a probability (transition) matrix, where the values of arcs represent the probability of goods and people to flow from a node to another one (see van Dongen 2000; Enright et al. 2002).

Figure 4. Example of hierarchical directed Nystuen-Dacey (N-D) network. The node a is the biggest one in the network and, therefore, it receives flow from all surrounding nodes. On the contrary, the node e is the smallest one and arcs from it depart towards all other nodes.

Results

Based on the above model, in this section the results of two different scenarios will be assessed. The first scenario will provide a baseline case where the values of all variables will be equal for all sites, so that it will be possible to test the role of geography and transport in shaping urban growth and settlement size structures. In the second scenario, we will account for any effects and factors not explicitly included in the model that will produce a distribution of site sizes and ranks similar to the empirical data (e.g., foreign contacts and trade from outside the study areas may have affected urban growth). This will allow us to test which values of the general variables (α, β, and Z) are required for specific sites that recreate urban layouts similar to those known from the archaeological and textual evidence. The advantage of this model fitting approach is in its explanatory power, which simply consists of identifying what are the general factors (social or environmental) that could have shaped the observed settlement patterns detectable by archaeological data and written sources.

Scenario 1: The Benefit of Geographic Location

In this scenario, the aim is to identify which sites could have taken advantages of their geographic location and to which extent site attractiveness (α) and willingness to travel or capability of movement (β) could have affected urban growth and settlement structure in the assessed areas. In this context, as α increases, feedback to site attractiveness is increased, while the increase of β indicates more constraints to movement and then less capability to travel for long distances. More precisely, in terms of human behaviour, we can say that at a given β value, the attractiveness of a site A to a site B several kilometres away, is half what it would be if site B was immediately next to site A. In this scenario, initial values including size (Z), population (X), attractiveness (α), and capability of movement (β) are equal for all sites at the beginning of the simulation², and an

² Specifically, size (Z) and population (X) are set to 1 for all sites.
incremental changes of parameters is done to the $\alpha$ and $\beta$ values. This is done to see how variations of site attractiveness and movement impedance affect populations and if certain sites consistently appear as relatively larger or smaller settlements. The simulation ends when the population and size results are considered stable, resulting in runs being 100 simulation ticks long. In this scenario, $\alpha$ is incrementally increased to 10 and $\beta$ to 1 (step intervals of 0.01), with $\beta$ values greater than 1 sometimes causing simulations to fail. These simulations are not intended to produce realistic site hierarchies, but instead to detect which sites could have benefited from their geographical location by reproducing, at least, the known settlements size distribution. This requires a measure of the linear correlation between the simulated population, used as a proxy to measure site sizes in the model, and the observed settlement size distributions, using a least squares correlation. We calculated the correlation of simulated site sizes to observed site size estimates, having first sorted each set by size. Then, two grids of search of parameters were produced in order to explore the goodness of fit for various configurations for CA and the KT (see Fig. 5a-b). Each grid shows fit values ranging between 0-1, where 1 is a perfect fit, for the values of $\alpha$ and $\beta$. The parameters values exhibiting the strongest least square correlation coefficients are given in Table 4, while the modelled settlement sizes are mapped in the Fig. 6a-b.

![Anatolia and Khabur Triangle Heat Maps](image)

**Figure 5.** Heat map showing least-square correlation in central Anatolia (a) and in the Khabur Triangle (b) under different $\alpha$ and $\beta$ conditions. The greyscale values represent correlation values, with dark grey representing the better fit and light grey the worse fit results.

<table>
<thead>
<tr>
<th>Region</th>
<th>Attractiveness ($\alpha$)</th>
<th>Travel ($\beta$)</th>
<th>Least-square correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Central Anatolia</td>
<td>2.1</td>
<td>0.06</td>
<td>0.96</td>
</tr>
<tr>
<td>Khabur Triangle</td>
<td>2.1</td>
<td>0.62</td>
<td>0.98</td>
</tr>
</tbody>
</table>

**Table 4.** Parameter values giving the highest least square correlation when compared with observed data.
The resulting plots are similar for both CA and the KT and differ just slightly (Fig. 5a-b). For both study areas, the best fit are found for low values of $\alpha (= 2.1)$, while the value of $\beta (= 0.6)$ in the KT is higher than the value of $\beta (= 0.06)$ in CA (Table 4). This suggests that in the KT the movement was more constrained than in CA. Although in CA the best fit are for relatively low $\beta$ and thus easy movement, it is also possible to get reasonable fit for results when $\alpha = 1.1$ and $\beta$ is between 0.44 and 0.50 (Fig. 5a). This is explained by the fact that when $\alpha$ is generally low, settlement size distributions matching the empirical record can be reproduced either through easy movement of goods and people in a given area or through movement restriction preventing population from flowing freely into local urban centres. Thus, in the case of CA it is not to exclude that the known settlement size distribution could be due to more movement restriction ($\beta = 0.44-0.50$). In the KT there is also good fit when $\beta$ is between 0.10-0.15, with $\alpha = 2.1$ (Fig. 5b). This apparent contradiction is explained by the fact that lower values of $\beta$ allow easy movement (of people and goods) between sites, which results in an even distribution of population when $\alpha$ is low. Hence, no site easily gets a large population and the resulting settlement sizes are comparable to the observed data. Put simply, very easy movement (low $\beta$) or more difficult movement (high $\beta$) create very comparable settlement structures when $\alpha$ is low.
Figure 6. Mapped output from scenario 1 for central Anatolia and the Khabur Triangle, with parameter settings as indicated in Table 4. Blue indicates larger relative site size under the model.
The strong correlation values for both study areas show that the model has reproduced the settlements sizes distributions at a relatively accurate level. Although these results show good agreement in site sizes distribution, the sites known for large estimated size in the observed data tend not to be those which the model forecasts to be large. The failure to identify precise sites does not necessarily imply that the model has not identified more general geographical areas, which would be expected to feature large sites. Of course, in reality, there are various factors that lead to the dominance of certain sites (e.g., earlier settlements could initially be larger, exogenous factors, socio-environmental factors may favour one site vs. others), whereas only geographic location is included here. If the analysis is widened, though, to consider situations where the dominant simulated site is not one of the true largest sites, but is in close proximity to one, it is possible to find the extent to which the localities of large sites are identified. In CA, the area to the north of Lake Tuz Gölü and Kızılirmak River between the Bozok and the Haymana Plateau results as the more likely to attract a greater portion of population (e.g., people and goods; Fig. 6a). Therefore, these parameter sweeps provide an output where known big centres such Altilar Höyük, Yassihöyük and Varavan Höyük acquire high population values, while other known large sites such as Açemhöyük, Alişar Höyük, Bogazköy, and Kültepe do not become prominent. If we have a look at the results in the KT, the area along the Wadi Jaghjah seems to be advantageous for settlement urban growth. In this case, it seems that only Tell Brak could have benefited from its geographical location (Fig. 6b). The present scenario shows how, in CA and in the KT, the geographical location may have benefited the growth of some known large MBA urban settlements (Altilar Höyük, Yassihöyük, Varavan Höyük, and Tell Brak). However, it is not sufficient to explain the development of other important urban sites (e.g., Açemhöyük, Alişar Höyük, Bogazköy, Kültepe, Tell Leilan, Tell Mozan, Tell Mohammed Diyab), even though these might have also been influenced to some lesser extent by their local position within CA and in the KT.

**Scenario 2: Reproducing Settlement Hierarchies**

The results of the previous scenario show that known large MBA sites, in general, do not emerge in a model based solely on local geographical location. Therefore, to distinguish such sites and reproduce settlement hierarchies more comparable with the known urban layout (from the archaeological and textual evidence) occurring in CA and the KT during the MBA (ca. 2000-1600 BC), some modifications to our initial model are required. Unlike the previous scenario, where initial conditions for sites size ($Z_j$) are set to 1 for all sites, in what follows we will instead differentiate the initial size ($Z_j$) for each site accordingly to the known estimated sizes about the sites in question. In this way the model setup already begins by favouring certain sites at the expenses of others. In more concrete terms, this could be regarded as accounting for any effects not explicitly included in the model (e.g. political, religious prominence, pre-existing trade contacts, etc.), which give to a site possible advantages or disadvantages relative to other sites in terms of population growth. In fact, the variable size ($Z_j$) is used as a relative proxy to regulate the flow of goods and people into a specific site. The greater the initial size value ($Z_j$) the more attractive a site is to settle (Davies et al. 2014, 145). Our approach is, therefore, to use the MBA site estimated sizes as an alternative to separate parameterisation of different factors (e.g. political prominence, inter-regional trade, favouruable environmental conditions) that have a material effect on the real-world outcome but are not included in the baseline scenario 1 (i.e., only incorporating geographical location). Furthermore, because $Z_j$ values are different for each site, the values of $\alpha$ and $\beta$ that make simulation results match with the empirical data differ from the previous scenario. For instance, a site with high initial advantage ($Z_j$) does not need high $\alpha$ to become large. Therefore, if a site is already large or has advantages over other sites, then this scenario tests how $\alpha$ and $\beta$ could enable settlement hierarchies. Put simply, in this second scenario, simulations are intended to show which $\alpha$ and $\beta$ values are needed to reproduce both the known settlement hierarchy and the site size distribution known from the observed archaeological survey data.

To do this, we will measure the least squares linear correlation between the simulated and observed sites size distributions and the rank-order correlation between the modelled and real settlements rankings using Spearman's rank correlation. Therefore, perfect rank order relationship is assigned a value of 1, while for no correlation is assigned a value of 0. It is important to point out
that a rank order relationship “does not reflect actual magnitude of the rankings for settlements size but only the order of the rankings” (Drennan 2009, 210). For this reason, least squares and Spearman’s rank correlation cannot be compared directly. In fact, high least squares correlation values do not necessarily correspond to high Spearman’s correlation values and vice versa. In order to avoid this problem, we produced two new heat maps by averaging the least squares and Spearman’s correlations for each parameters configuration (Fig. 7a-b).

The parameters values giving the strongest correlations in terms of settlements’ rank and size are given in Table 5, while the modelled settlement sizes are mapped in the Fig. 8a-b. The resulting good fit shown by the heat maps indicates that the model has successfully reproduced the observed data. The best model fit for both CA and KT is with identical alpha values (0.9 for both case studies), and similar β values (respectively 0.6 and 0.7; Table 5). Furthermore, Fig. 7b shows that for the KT the good fit results are with slightly higher β values in comparison with the CA’s ones (Fig. 7a).

Figure 7. Heat map showing least square and Spearman’s correlations averaged in central Anatolia (a) and in the Khabur Triangle (b) under different α and β conditions. The greyscale values represent correlation values, with dark grey representing the better fit and light grey the worse fit results.

<table>
<thead>
<tr>
<th>Region</th>
<th>Attractiveness (α)</th>
<th>Travel (β)</th>
<th>Least-square correlation</th>
<th>Spearman’s rank correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Central Anatolia</td>
<td>0.9</td>
<td>0.6</td>
<td>0.99</td>
<td>0.98</td>
</tr>
<tr>
<td>Khabur Triangle</td>
<td>0.9</td>
<td>0.7</td>
<td>0.98</td>
<td>0.98</td>
</tr>
</tbody>
</table>

Table 5. Parameter values giving the highest combination in terms of Least-square's correlation and Spearman’s rank correlation when compared with observed data.
Figure 8. Mapped output from scenario 2 for central Anatolia and the Khabur Triangle, with parameter settings as indicated in table 5. Blue indicates larger relative site size under the model.
Visual inspection of the results in Fig.6a-b shows, at a first glance, that both CA and the KT are characterized by significant clustering of sites. This could be the result of the uneven intensity of the archaeological surveys carried out in the areas or related to the contemporaneousness of the modelled sites that lie within a wide lifetime range. In fact, it is plausible that only a subset of the modelled locations were in existence during the MBA. In addition, since every site represents a source of flow (e.g. people and goods), a dominant site situated in an area with high site density has a larger potential in-flow and probability of reaching a large size than a similarly dominant site located in a sparse region. To deal with this problem and the uncertainty in the original dataset, we make use of bootstrap statistics technique. This consists of performing 500 simulations, where, for each simulation, we have randomly removed a subset of \( n \) of the sites according to four different probability settings (0.05, 0.15, 0.25, and 0.5). When all simulations are complete, the mean modelled population size of sites across all the simulations is used. This method is based on the assumption that each sampling represents a possible 'state of the world', and that averaging in this way accounts for uncertainty about the composition of the original dataset. So, we carried out 5 simulations for each of the KT and CA's datasets and with \( n \) of sites and parameters as specified in the Table 6. The results show good consistency between the simulation results performed with the original whole dataset and the different subsets of samples for both CA and the KT. The results are statistically significant in all cases (p-value < 0.01). Put simply, sites which are found to be large under the original scenario 2 algorithm remain large under this random sampling.

<table>
<thead>
<tr>
<th>Region</th>
<th>Probability of removing sites</th>
<th>n. of sites</th>
<th>Least - square 's correlation</th>
<th>Spearman's rank correlation</th>
<th>Attractiveness (( \alpha ))</th>
<th>Travel (( \beta ))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Central Anatolia</td>
<td>0</td>
<td>440</td>
<td>0.99</td>
<td>0.98</td>
<td>0.9</td>
<td>0.7</td>
</tr>
<tr>
<td></td>
<td>0.05</td>
<td>418</td>
<td>0.96</td>
<td>0.98</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.15</td>
<td>374</td>
<td>0.96</td>
<td>0.98</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.25</td>
<td>330</td>
<td>0.95</td>
<td>0.81</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>220</td>
<td>0.94</td>
<td>0.98</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Khabur Triangle</td>
<td>0</td>
<td>439</td>
<td>0.98</td>
<td>0.98</td>
<td>0.9</td>
<td>0.6</td>
</tr>
<tr>
<td></td>
<td>0.05</td>
<td>417</td>
<td>0.94</td>
<td>0.98</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.15</td>
<td>373</td>
<td>0.96</td>
<td>0.98</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.25</td>
<td>329</td>
<td>0.95</td>
<td>0.98</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>219</td>
<td>0.94</td>
<td>0.98</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 6. Table showing Pearson and Spearman's correlations under different probability settings of an averaging and random sampling system.
Connectivity and Interaction at the Regional Scale

In this section, we utilize Nystuen-Dacey (N-D) graphs resulting from scenario 2 of SIEM to understand areas of interactions on different spatial scales in CA and the KT. To do so, we will first perform some modifications to the original structures of the N-D network outputted by the SIEM model. As already said in section 3.3, the N-D graph shows the amount of flow $S_{ij}$ between any pair of sites $i$ and $j$ in terms of hierarchical relationships (see Fig. 4). Therefore, this network is directed (each line connecting nodes is an arc) and each arc is weighted according to flow value between pairs of nodes. Because N-D network shows all possible connections between pairs of nodes, its original structures outputted from SIEM results as a cloud of thousands of arcs between nodes in our two study areas. As a result, even far away pairs of nodes will be connected by arcs with flow values very close to 0. In order to avoid such a problem, we will apply a cut-off threshold to arc values. Hence, arcs having values below this threshold will be dropped from the analysis, allowing to focus on more intense flows $S_{ij}$ between pair of nodes and assess the strength of interaction between settlements on the basis of the threshold value applied. For the purposes of our research objectives, we will adopt the values + 1 and + 2 standard deviation as thresholds (see Tables 7 and 8 for assessing how the N-D graph changes according to different threshold values). Given these assumptions, we will first calculate the network centrality of settlements in both CA and the KT, and then we will assess possible clustering of nodes in sub-networks or sub-areas of interactions.

<table>
<thead>
<tr>
<th>Arc value threshold</th>
<th>Min. arc value</th>
<th>Max. arc value</th>
<th>No. nodes</th>
<th>No. arcs</th>
<th>No. components</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>0.1e-05</td>
<td>15261</td>
<td>440</td>
<td>96,578</td>
<td>1</td>
</tr>
<tr>
<td>+ 1 std. dev.</td>
<td>&gt; 134</td>
<td>15261</td>
<td>369</td>
<td>1,986</td>
<td>1</td>
</tr>
<tr>
<td>+ 2 std. dev.</td>
<td>&gt; 244</td>
<td>15261</td>
<td>196</td>
<td>538</td>
<td>2</td>
</tr>
</tbody>
</table>

Table 7. Nystuen-Dacey graph structure in central Anatolia according to the arc value threshold.

<table>
<thead>
<tr>
<th>Arc value threshold</th>
<th>Min. arc value</th>
<th>Max. arc value</th>
<th>No. nodes</th>
<th>No. arcs</th>
<th>No. components</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>0.1e-05</td>
<td>456</td>
<td>439</td>
<td>96,141</td>
<td>1</td>
</tr>
<tr>
<td>+ 1 std. dev.</td>
<td>&gt; 1</td>
<td>456</td>
<td>173</td>
<td>532</td>
<td>2</td>
</tr>
<tr>
<td>+ 2 std. dev.</td>
<td>&gt; 14</td>
<td>456</td>
<td>120</td>
<td>135</td>
<td>5</td>
</tr>
</tbody>
</table>

Table 8. Nystuen-Dacey graph structure in the Khabur Triangle according to the arc value threshold.
Network Centrality and Clustering

The first aim of this section is to identify which sites could have played a pivotal role in the network of flows $S_{ij}$ (of goods and people) occurring in CA and the KT. To do so, we will measure network centrality of nodes in two different scenarios (see the properties of graphs in Tables 7 and 8): 1) a N-D graph with arcs (or edges) flow values higher than $+1$ standard deviation; 2) a N-D network with flow values higher than $+2$ standard deviation. This will allow us to assess different intensity of interactions on different spatial scales within our two case studies. In the scenario 1, we calculated the indegree of sites in a directed N-D network. The results in CA (Fig. 9a) and the KT (Fig. 10a) unsurprisingly match with the results outputted in the scenario 2 of the SIEM model for both study areas: only a few main urban centres collect the flow (of goods and people) from the surrounding hinterland. The N-D graph is useful because provides a visual inspection of the spatial distribution and magnitude of interactions between main nodal centres and secondary settlements. In CA, Bögazköy seems to play a pivotal role by receiving significant flows even from distant settlements. Unlike CA, in the KT the interactions occur on a smaller scale and a shorter distance. In addition, the western part of the KT shows many isolated and unconnected nodes and a separated component around Tell al-Fakhkhariya. The limit of a directed N-D network is that our analysis is limited to assess the strength of interaction between pairs of nodes so that the network diameter is one. As our network may represent the flow of goods and people as result of interaction (e.g. taxes, trade, migration, agricultural surplus supply) between sites, it is worth also assessing how sites interact at different spatial scales and what are the strongest links. In addition, a node could have played a crucial role as an intermediary to the transmission of goods in the network. Those two aspects can be assessed by respectively calculating the closeness centrality and the betweenness centrality in our N-D graph. To do so, we need to convert it into an undirected graph, where the ties (edges) between nodes are two-way.

Figs. 9b and 10b show closeness centrality in both CA and the KT. In the latter one, it is, as might be expected, to note that sites in the middle of the area are more central within the network, with the exception of five sites around Tell al-Fakhkhariya that belong to a separate component. On the other hand, in CA the pattern differs and there is not a continuous gradient of change to the far periphery of the region, but a tendency to generate sub-regional spheres of interaction that may be explained, to some extent, in terms of topographical isolation and terrestrial connectivity. The betweenness centrality (Figs. 9c and 10c) in both study areas match with the SIEM model’s outputs (Fig 6a-b). This shows that the major urban centres were not only collector of flows but also pivotal hubs to the transmission of goods in the network. If we use a Markov cluster (MCL) algorithm for detecting the clustering of nodes, the first step is to use a weighted directed N-D network. The resulting outputs by using the MCL algorithm show two different situations for CA and the KT (Figs. 9d and 10d). In CA we have 6 different clusters, where the biggest one encompasses almost all the area. This could be due to the role played by Bögazköy in the model, which has strong ties with far away sites given its relative dominance. Other small clusters result around Yassihöyük (Ankara), Varavan Höyük, Altılar Höyük, Açemhöyük, and Kayalipinar Harabesi. This method could be useful to roughly estimate possible areas of interaction on a larger scale and longer distance. It is interesting to note that the all other clusters are located beyond the Kızılirmak bend, which could suggest that this river could have played both as physical and cultural/political barrier. On the other hand, in the KT the resulting scenario appears more fragmented: twelve less-defined different clusters distributing around the major centres. These could reflect the higher degree of competition among city-states in an intensively inhabited area.

---

3 A component is a portion of the network where all nodes are connected by at least one line (arc or edge).
4 The network diameter is the largest distance, in terms of lines, between two nodes.
Figure 9. Nystuen-Dacey network’s centrality results for central Anatolia (+1 standard deviation): a) indegree; b) closeness centrality; c) betweenness centrality; d) Markov cluster. The lines are coloured according to flow, ranging from light to dark red.
Figure 10. Nystuen-Dacey network’s centrality results for the Khabur Triangle (+ 1 standard deviation): a) indegree; b) closeness centrality; c) betweenness centrality; d) Markov cluster. The lines are coloured according to flow, ranging from light to dark red.
In scenario 2, we perform the same analysis carried out above by making use of a higher arc (or edges) value threshold (+2 standard deviation) for the N-D network in both CA and the KT (see Tables 7 and 8 for the thresholds used in this second scenario). This will allow us to assess if and where stronger interaction occur on smaller local scales. In this second scenario, the N-D graph for CA and the KT will be respectively composed of 196 and 120 nodes and will be divided into 2 and 5 components. Of course, the in degree of the main nodal centres decreases in comparison with the one in the first scenario as there are fewer vertices in the N-D network. Nevertheless, the general pattern matches with the outputs from scenario 1 for both case studies (Figs. 11a and 12a). The closeness centrality results differ with the previous scenario as in both areas there is a more marked general tendency to divide the networks in sub-areas. In CA, we can detect two bigger sub regions with high closeness centrality: one in the northern part around Boğazköy and Kösele Tepesi, and one in the central part where is located Yassihöyük (Fig. 11b). In the KT two bigger sub-regions are in the areas around Tell Brak and Tell Leilan (Fig. 12b).

As for the indegree, the betweenness centrality in both areas roughly matches with the outputs obtained in the first scenario (see Figs. 11c and 12c). The MCL algorithm’s outputs show a different set of circumstances in CA in comparison with the results from scenario 1. CA is now divided into nine different sub-regions and now Boğazköy’s cluster is completely confined within Kizilirmak bend (Fig. 11d). This scenario shows a blurry spatial distribution of strong local interactions between main urban centres and their respective hinterlands. In the KT, the resulting pattern is similar to the one detected in the first scenario (Fig. 12d).
Figure 11. Nystuen-Dacey network’s centrality results for central Anatolia (+2 standard deviation): a) indegree; b) closeness centrality; c) betweenness centrality; d) Markov cluster. The lines are coloured according to flow, ranging from light to dark red.
Figure 12. Nystuen-Dacey network’s centrality results for the Khabur Triangle (+ 2 standard deviation): a) indegree; b) closeness centrality; c) betweenness centrality; d) Markov cluster. The lines are coloured according to flow, ranging from light to dark red.
Hollow Ways and Connectivity in the Khabur Triangle

There is little evidence for pre-Classical roads in Upper Mesopotamia and central Anatolia, with the exception of the network of hollow ways, which stretches for a total of over 6000 km. We can also apply network centrality measures to the palimpsest of these informal landscape features (tracks or paths) generated by human or animal movement, which stretches for a total of over 6000 km in the KT (see Wilkinson 1993, 1994, and 2003; Wilkinson et al. 2010; Ur 2003 and 2009; Altaweel 2008). These archaeological features represent a privileged way to understand connectivity in the area and they can be useful to further infer flow of people and goods between settlements in the KT. The analysis carried out in this section starts from the assumption that hollow ways created in the third millennium BC were probably still used in the early second millennium BC. In fact, the hollow ways that could be originally dated to the MBA, on the basis of the occupation period of the sites from which they depart, are few and, therefore, insufficient for this kind of analysis (see Fig. 13a). Thus, the existing hollow ways have been converted into an undirected network, where the nodes (settlements) are connected, where possible, by edges (hollow ways). The resulting network is composed by 200 nodes connected by 248 edges. The degree of nodes show high values for the main urban centres in the area: Tell Brak, Tell Muhammed Diyab, Tell Leilan, Tell Farfara, and Tell Hamidiya (Fig. 13b). Unfortunately, we do not have preserved hollow ways around Tell Mozan. Even so, where the preserved hollow ways allow calculation of a network centrality measure, the resulting pattern shows that the biggest urban centres in the area have high degree values. This confirms that hollow ways could have been the main axes on which the flow of goods and people transited and the results offer good evidence of the interaction of the main nodal centres with their surrounding rural hinterland (villages and hamlets). The closeness centrality unsurprisingly show that sites locate in the middle of the area appear to be more central and, hence, there is a continuous gradient of change to the far east and west of the KT (Fig. 13c). The betweenness centrality values show that main urban centres such as Tell Leilan, Tell Farfara and Tell Brak also act as crucial go-between hubs in the connectivity system (Fig. 13d). In particular, it seems that a continuous axis of interconnected hollow ways departs from Tell Brak and proceeds to Tell Leilan via Tell Farfara.

Figure 13. Spatial distribution of hollow ways in the Khabur Triangle. In red, a) the hollow ways likely formed during the Middle Bronze Age (ca. 2000 – 1600 BC). Hollow ways network’s centrality results: b) indegree; c) closeness centrality; d) betweenness centrality.
Discussion

The results above demonstrate the utility of a SIEM model for exploring how geographical settings and unspecified social, political, and environmental factors may have affected the urban growth and the settlement hierarchies in CA and in the KT during the MBA (ca. 2000-1600 BC). The advantage of this modelling approach is that it enables researchers to account for missing empirical data and to reproduce outputs matching the known historical and archaeological evidence for explaining which generalized phenomena (e.g., geographical location, political or religious importance, trade contacts, etc.) may have caused settlements growth, stability or decline. On the other hand, the weakness of the present method is that we do not know which specific factors caused the observed results. Hence, the outputs may be used for highlighting general settlement hierarchy patterns and for providing general explanations, or those caused by multiple factors, regarding the development of past human settlement hierarchies.

The simulation’s results show how local geography alone may have played an important role in determining why some settlements in CA and the KT such as Attilar Höyük, Varavan Höyük, Yassihöyük and Tell Brak became larger than others. Perhaps somewhat unexpectedly, Kültepe does not appear large in the simulation results. This is due to the fact that its prominence could be better explained on a larger interregional spatial scale rather than with the local Anatolian geography assessed. Nevertheless, local geographic location alone cannot explain past human settlement hierarchies, as local interactions seems to not make known major MBA centres large. In particular, the KT, due to its geographic location and lack of natural borders, does not provide relative isolation, possibly maximizing exogenous influences from more distant settlements and regions. This area has mostly been a buffer-zone between political entities based outside the area, and so changes that occurred in this region in the early second millennium BC should be regarded in the perspective of international developments involving the neighbouring areas.

Because location data alone were not sufficient to explain the dominance of most known large sites, in the second scenario it was necessary to include a certain amount of known information about the sites in question. This process required the manipulation of initial conditions of some model’s parameters. Particularly, it was found that closer matches tended to be associated with differentiated values of initial size $Z_j$ for each site, a parameter controlling a wide range of different beneficial factors (e.g., political or religious prominence, trade contacts, etc.) which can be related specifically to the nature of the data. The results generally show that other non-geographic factors do make specific known sites larger, as sites become relatively large through modifications to initial size $Z_j$ and incremental changes of parameters to the $\alpha$ and $\beta$ values. The fact that for CA and the KT the best fits are found in very similar parameters configurations, particularly in terms of $\alpha$ and $\beta$, implies that there is no marked difference in the settlement size structures (Table 5). Both areas show relatively high $\alpha$ and $\beta$ values that, translated to the discussed historical data, reveal politically fragmented landscapes characterized by the presence of several competing peer polities aiming to exert their influence over their surrounding hinterland. As a consequence, larger percentage of the population is concentrated in the few larger centres as the willingness of travel or the capabilities of movement are restricted (high $\beta$ values), thereby leading to the establishment of larger local sites that also exert their power over their surrounding hinterlands and, therefore, absorb more flow from nearby sites (i.e., which reflects higher $\alpha$). In the KT, higher $\beta$ values indicate less movement capability and willingness of travel across the landscape in comparison with CA (see Fig. 7b). This could reflect political circumstances occurring in the KT, where the largest sites were constrained in a smaller plain area in which the lack of marked topographical features (e.g., wide rivers, mountain ranges) could have further enhanced competition (e.g. warfare which reflects higher $\beta$) between large city-states of comparable size and political prominence (Eidem 2000, 257; Charpin and Ziegler 2003; Veenhof and Eidem 2008, 290-321; Ristvet 2008 and 2012; Palmisano 2015). In particular, it seems that Tell Brak, Tell Leilan, and Tell Mozan in the MBA would need greater exogenous effects or initial advantages to enable them to consistently reach a relatively greater size. This could come in the form of Śamši-Adad I making Tell Leilan (the ancient Šubat-Enlil) an important political capital (Charpin 1987; Eidem 2008; Ristvet 2008), Tell Brak being the seat of the “Lady of Nagar” (Oates et al. 1997, 141), and local or external benefits that Tell Mozan, Tell Muhammad Diyab and Tell Farfara may have relative to other sites.
In CA, the constraints placed on movement due to factors such as political and territorial divisions could have made individuals travel shorter distances and may have also concentrated the flow of people from surrounding rural communities into few large local urban centres (see Bachuber 2012, 576-578). This may reflect well the central Anatolian political landscape fragmented into numerous independent city-states during the Old Assyrian Colony Period (ca. 1970-1710 BC; cf. Barjamovic 2011, 6; Barjamovic et al. 2012, 44-49). In CA, the geographical location does not explain alone why sites such as Açemhöyük, Alişar Höyük, Bogazköy and Kültepe became prominent in the early second millennium. The results show that the urban growth of those sites could be related to trade, external contacts, and other general advantageous factors (e.g. political prominence, military power, religious prestige, etc.). The simulation’s outputs are particularly interesting for Kültepe, which requires high values of initial $Z_i$ to start becoming large. This reflect the international character of this site, which hosted an Old Assyrian kārum and was one of the main hubs of the commercial trade network set up by the Assyrians in Upper Mesopotamia and CA (see Barjamovic 2008 and 2011; Veenhof 2008). Put simply, the urban development of Kültepe cannot be explained in terms of local interaction within CA, but it may be the result of external contacts and long-distance trade activities with other regions and settlements. In fact, the archaeological evidence from Kültepe’s lower town (level II and Ib) such as cylinder seals and balance pan weights respectively belonging to different regional styles and weight systems, Khabur ware and Syrian Bottles show the involvement of Kültepe in long-distance contacts with Syria and Northern Mesopotamia (see Aubet 2013; Ascalone and Peyronel 2006, 401-421; Emre 1999; Oguchi 1997; Özgüç 2006; Özgüç-Tunca 2001). Another important site whose urban development could be related to external contacts and trade is Açemhöyük. The role played by this centre during the MBA is reflected from the two palaces Sarikaia and Haṭipler (level 3-4) that have yielded archaeological evidence (e.g. seals, clay bullae, pottery, etc.) showing long-distance contacts with upper Mesopotamian Amorite dynasties (cf. Özgüç 1980, 67; Özgüç-Tunca 2001, 128). This site has long been identified with Purušhaddum (cf. Forlanini 2008, 65-66; Veenhof and Eidem 2008) but recently Barjamovic has identified it with Ulama, the seat of a an Assyrian wabartum (Barjamovic 2011, 411).

Furthermore, in this paper we have shown how a graph-theoretical approach applied using N-D networks represents a valid complement for visual examination of interactions among settlements within a given study area. The network centrality measures applied to N-D networks, with different strength of relationship between nodes, allowed us to detect spatial interactions on both regional and local scale. The results in scenarios 1 and 2 have showed that the major settlements in both CA and the KT played a pivotal role both as nodal centres collecting flows (of goods and people) from their respective rural hinterland (e.g. villages, farmsteads, hamlets) and as go-between flow of goods between settlements. This latter aspect is emphasized by the presence of several interconnected sub-regions (or networks) developing around each main urban centre, which suggests a two-tiered spatial scale of interaction in a N-D network: 1) main nodal centre connected to its peripheral hinterland; 2) inter-connections between distinct sub-regions. The first kind of interaction can be described as the flow that main urban centres collected by their rural hinterlands through agricultural surplus and work-force. This is quite evident for the KT, where the hollow ways could be interpreted as the main vectors over which goods and people transited between settlements. The second kind of interaction occurs at larger spatial scales and can be explained as trade and political relationship (or influence) among different and competing city-states. In CA, Bogazköy seems to have long distance strong relationships in the area within of the Kızılırmak River, which could have played as physical and political boundary. Other smaller sub-regions emerge around the other main urban centres to the south of the Kızılırmak and of the Delice Rivers: Açemhöyük, Altılar Höyük, Kültepe; Sevket Tepesi; Varavan Höyük, Yassihöyük (Kirşehir), and Yassihöyük (Ankara). In the KT, the situation is more fragmented and the areas is divided into

---

5 The MBA rural hinterland of the central Anatolian plateau is archaeologically elusive and under-investigated. No intensive archaeological surveys of the agricultural settlements surrounding major archaeological sites have ever been carried out for the MBA. The early second millennium farming hinterland is well attested in some textual evidence (see Forlanini 1992, 176; Dercksen 2008, 139; Barjamovic 2011, 232-235).
smaller sub-networks and isolated components. This is the results of a more heated competition and conflict between the polities occurring in the area, as already showed by the higher $\beta$ values in the SIEM model. In Anatolia, past human dynamics and interactions could have been strongly shaped by the landscape, where the movement was presumably funnelled into narrow natural corridors in the wide alluvial valleys. By contrast, in the KT the landscape, based on terrain slope, seems to not have affected past human movement and, instead, the flow of people, animals and goods could have been funnelled into the evident hollow ways. Of course, given the volatile political situation in the area, the routes could have been more circuitous and shifted considerably over short periods (see Branting et al. 2013, 143).

**Conclusion**

The present work has demonstrated the benefits of a well-established modelling approach to help explain how geographical settings and other factors may have shaped settlement hierarchy and areas of interaction across the KT and CA in the MBA (ca. 2000 – 1600 BC). In particular, the advantage of SIEM models is that they enable researchers to account for missing empirical data and to reproduce patterns of settlement growth, stability, or decline that matches known historical and archaeological evidence (e.g., geographical location, political or religious importance, trade contacts, etc.). In practical terms, these models are also useful for forecasting general areas where larger or smaller sites are to be expected. The application of graph theory, such as N-D graphs, also provided powerful insights about past human movement dynamics and interaction both on local and regional scales and to identify pivotal nodal centres. On the other hand, problems arising while performing our analyses have highlighted the limits of the research present here as well, thereby outlining some potential new lines of enquiry. For example, comparisons between archaeological and textual data have highlighted the constraints currently imposed by temporal uncertainties and the geographic scantiness of data for the first two centuries of the early second millennium (ca. 2000-1800 BC). Apart from the case of Kültepe’s lower town, which has yielded a well-defined chronological sequence, most sites have offered only a coarser temporal resolution. Therefore, a diachronic development of political and trade landscapes occurring in the early second millennium BC cannot be offered at a detailed scale, but only broadly treating the MBA timespan as a whole. Furthermore, our consideration of known sites’ sizes uses estimated extent, while it may be more prudent to draw values from a weighted distribution. Further use of bootstrap techniques would be a useful way of overcoming some of these problems, although not entirely without further primary fieldwork. A future stage of this research should, therefore, be the creation of models capable of generating more explicit and testable hypotheses at a larger spatial scale than the two case studies considered so far. This might involve more detailed spatio-temporal data, a precise definition of the initial conditions of the spatial models built, and the integration of extra archaeological settlement data and various classes of material culture. Overall, the use of SIEM models and other statistical techniques have provided useful solutions to some of the uncertainty in archaeological datasets and useful insights into observed archaeological patterns at a more general level.

**Acknowledgements**

This research was supported by a Grand Challenges small grant from the UCL and Faculty in the Arts and Humanities. We also thank Andrew Bevan and the two anonymous reviewers who provided thoughtful comments, feedbacks and suggestions to improve the quality of this work.

**References**


Fontenari, S., Franceschetti, S., Sorrentino, D., Mussi, F., Pasolli, M., Napolitano, M., and Flor, R., 2005. r.walk. GRASS GIS.


