## 2_1_1_Introduction_to_the_Trypillia-Cucuteni_settlement_distribution_(Marco_Nebbia)

The core set of data derives from the publication of all the known Trypillia sites in modern Ukraine. The so-called Trypillia Encyclopaedia is a collection of varied information from different sources such as published scientific papers and grey literature, from excavation reports to unsystematic and sporadic field visits. The collection of data has been built up since the 19th century when the first register and map of archaeological remains in Ukraine has been compiled. At the beginning of the 20th century, an official register of archaeological monuments was collected in 1925 (V.U.A.K) and then updated until 1950, but never published. The first general publication of Trypillia sites was in the middle of the 20th century, with Tatiana Passek's book including 94 entries (Passek 1949; Childe 1951). A decade later, Passek published an updated version which included some sites along the Dniestr in Moldova - a total of 125 Trypillia sites (Passek 1961).

A few years later, the first broad collection of archaeological sites in the country, including 960 Trypillia sites out of a total of 7,000 recorded, was published as *Archaeological Monuments of Ukraine* (Zbenovic et al. 1966). From the 1960s to the early 1990s, a national programme of recording archaeological sites developed a standard protocol of data collection by preparing a document for each site (Passport) which included basic information regarding the type of site (settlement, burial mound, surface scatter), period and dimensions. During this period of investigation, many previously unknown Trypillia sites were found. Counties like Vinnitsa, Cherkassy and Kirovograd remained poorly investigated. In 1971, the first map of 171 Trypillia sites was published within the *Archaeology of the Soviet Republic of Ukraine* volume. Since the fall of the Berlin Wall and the end of the Soviet era, the national programme of archaeological investigations has been decentralized and each county (oblast) developed its own plan for site recording and mapping. In 1995, a series of regional maps were published for the counties of Chervinits, Vinnitsa, Ternopil, Khmelnits and Odessa. This included the plotting of site locations on a map as well as recording other basic information. From this moment, the development of registers of archaeological monuments has been in the charge of each county, with some central monitoring by the Institute of Archaeology in Kiev. The non-systematic and decentralized way of managing the archaeological heritage has led to a level of uncertainty regarding the number of known Trypillia sites, so that some archaeologists (e.g. M. Videiko) have mentioned that these are around 1,500, whereas others (e.g. S. Ryzhov) are more optimistic, with around 5,000 Trypillia sites (Videiko 2004, 564).

Finally, in 2004, a comprehensive collection of all the information regarding known Trypillia evidence was published as an edited volume called the *Encyclopaedia of Trypillia Civilization* (Videiko 2004). The contents of the Encyclopaedia span scientific publications to archive material, from unpublished research and excavation reports to sporadic amateur notes. All the information has been taken from all the above-mentioned sources and updated with new discoveries from 19 counties (oblasts), for a total of 2,042 Trypillia entries, although some Ukrainian archaeologists claim that there are as many as 4,400 (Videiko 2004,

564). Unfortunately, the information collected since the beginning of the 20th century has not been assessed nor evaluated with field visits or excavations but rather taken as granted and reported while the main focus has been devoted to finding more (and possibly bigger!) sites. Meanwhile, as field methodologies and theories advanced and developed, the information collected since the 19th century has not been double-checked or updated with improved recording procedures. Therefore, the Encyclopaedia represents a massive amount of information compiled using a varied range of methods and field procedures that has produced an uneven and inconsistent dataset.

The information reported consists of: 1) a description of the site location, 2) a brief history of site investigations, 3) an estimation of the extent of the site, 4) the chronological horizon of the material found on site, but they are not provided for every site. Conversely, the information is really piecemeal and for only a few entries of the Encyclopaedia do we have full detailed information. Moreover, the inaccuracies, derived from the adoption of old and, by now generally obsolete, field methods and theories, have been transmitted on to the final version of the publication. The compilation of metadata regarding how data have been collected in the field is lacking for almost all the entries. It was therefore, necessary to evaluate the reliability of the information by trying to understand how people recorded sites in the field and elaborated reports.

A long and severe data cleaning process assessed the accuracy as well as reliability of the information reported and resulted in a strict selection of "usable" data. The entries that have been deemed sufficiently reliable for the research have been plotted on the map (Fig. XX1).

A total of 499 sites (from the complete record of 2,042 sites contained in the Encyclopaedia) have been considered and mapped with a location accuracy that ranges from few metres to 2 km, depending on the details provided in the descriptions (ADS LINK TO 2_1_3_Trypillia_spreadsheet). Fig. 2_1_1 (ADS LINK TO 2_1_2_IMAGES/2_1_1_Encyc_sites_complete) displays the overall distribution of known Trypillia sites in modern Ukraine. The figure itself shows how site densities changes across the whole territory of occupation, and the data cleaning process confirmed that these differences are actually reflecting diverse research intensity rates. In fact, the two counties which are archaeologically best investigated are now Vinnitsa and Cherkassy. For this reason, all the spatial analysis and interpretations have been conducted diachronically and mostly at a large scale.

The database represents the data that have been used for the research on Trypillia settlement patterns and the table (Table 1) here displayed reports the type of information that has been recovered from the Encyclopaedia and how they have been organised.

Table 1 *List of all the fields of the attribute table of the Trypillia database.*

| FIELD | DESCRIPTION |
|---|---|
| **ID** | Unique ID identification number. |
| **Name** | Name of the nearest village or main watercourse (as it appears in the Encyclopaedia). |
| **Oblast** | Name of the county where the site is located. |
| **Region** | Name of the municipality where the site is located. |
| **Phase** | Trypillia phase attributed to the site (ADS YORK LINK Table 1 in SECTION 2_2_1_REPORT) for Trypillia relative chronology). |
| **Area (ha)** | Site area as reported in the encyclopaedia for the majority of the sites and corrected where possible. |
| **A** | Boolean value of presence/absence. |
| **BI** | Boolean value of presence/absence. |
| **BII** | Boolean value of presence/absence. |
| **CI** | Boolean value of presence/absence. |
| **CII** | Boolean value of presence/absence. |
| **Remote_sensing** | Level of certainty for site visibility on satellite imagery (from 0 to 1). |
| **Stage_code** | A numerical value for the assigned Trypillia phase. |
| **Location_certainty** | Level of certainty for the location assigned to the point (from 0 to 1). |
| **Elevation** | Elevation of the point derived from the SRTM data (30m) in metres. |
| **Notes** | Notes on the pottery group assigned to the site (if present) and other general notes. |
| **Annotations** | Annotation on where the area value has been |

derived from that specific site.

For many years, the focus of the research on Trypillia megasites remained at the site level, trying to gather more information and data for the whole settlement. Only recently have archaeologists started to look at these large settlements in the wider contest of the Southern Bug-Dnieper interfluve 'system' or "Western Trypillia Culture" (WTC) (Diachenko 2010; Diachenko and Menotti 2012: 2015). Manzura studied Cucuteni-Trypillia settlement dynamics within the framework of the "colonisation" of the North Pontic steppe territory, but referred to megasites (or "super-centres") as evidence of a shift from an egalitarian tribal system towards a more complex societal organization in control of restricted resources from intensified production and exchange (Manzura 2005).

A new scale of analysis for the understanding of the nature of megasites has been proposed by exploring the spatial relationship that these large settlements have with the coeval network of smaller sites. Diverse lines of investigation have been pursued in order to further our knowledge on the settlement strategies of Trypillia sites and megasites.

The analysis of the settling strategies that sites and megasites followed, allowed us to infer that there does not seem to be an environmentally-linked reason why Trypillia people decided to settle and develop such massive sites in the SBD interfluve (ADS LINK TO 2_1_2_IMAGES /2_1_2_Megasites_locations); therefore, we can argue that this analysis suggests that there must be other reasons why megasites developed mainly in the Southern Bug – Dnieper territory, perhaps more related to the social rather than the environmental sphere.

The assessment of how the emergence of megasites affected the variability of site sizes per each phase showed how, despite being statistical outliers (ADS LINK TO 2_1_2_IMAGES /2_1_3_settlement_size_distributions), megasites did not influenced remarkably the overall site size hierarchy across the three main phases of their development, as shown by the GINI coefficient calculated for site areas (ADS LINK TO 2_1_2_IMAGES / 2_1_4_GINI_coefficient_of_site_size_variability). Furthermore, the graph (ADS LINK TO 2_1_2_IMAGES /2_1_5_megasites_and_sites_count) shows how the number of megasites increased alongside the number of smaller coeval settlements.

The exploratory analysis of the distribution patterns of Trypillia sites across the landscape, how they changed at different scales and how they changed in time, led to the definition of a certain threshold that can represent the scale at which megasites can be considered "Central" places for the surrounding settlements. A two-fold approach called PPA (Point Pattern Analysis) has explored the first- and second-order properties of the distribution of Trypillia settlements. In brief, the first-order characteristics describe the average point density (or intensity) and distribution patterns across the whole region under investigation; whereas the second-order characteristics describe the density of points relatively to their internal spatial organization, reflecting internal interactions of attraction of inhibition (Bevan et al. 2013, 31).

A simple visual assessment of the five distribution maps of Trypillia sites shows an increasing level of settlement clustering until phase CI – especially with the appearance of the first megasites in phase BI - and a trend to a more dispersed settlement pattern in phase CII (ADS LINK TO 2_1_2_IMAGES /2_1_6_Sites_per_phase). Moreover, by phase BII, the full Dniester-Dnieper interfluve is occupied by Trypillia settlements and remarkable further expansions are not visible until phase CII. During phase CI (the period of maximum size of megasites), the level of site clustering reaches its peak, but the areas occupied remain broadly the same. It is only with phase CII, with the end of megasites, that a remarkable overall dispersal can be observed and the Trypillia area of influence reaches its maximum extent.

Consequently, the combination of Incremental Global Moran's I (Moran 1950)[1] and then Anselin's Local Moran's I (Anselin 1995)[2] showed that there is a global and developing nucleation process of Trypillia settlements during the first four phases A, BI, BII, CI, whereas the last CII phase is characterized by a return to a more dispersed settlement pattern. On a local scale, the 'centrality' of megasites is measured by a statistically significant difference in site size at a scale of approximately 100 km. This threshold seems to be consistent for around 1,000 years, without being affected by the increasing number of sites and their development and nucleation. This statistical scale of emerging megasites could suggest the human scale at which these settlements were seen as exceptionally big and therefore "special" places.

Overall, we can argue that, the distribution of Trypillia megasites concentrated in the Southern Bug-Dniestr interfluve for a social, rather than an environmentally-linked, reason; that their appearance did not particularly affect the overall site size hierarchy, which was quite stable for almost 1,000 years; and finally, that the "centrality" of megasites can be identified within approximately 100 km distance, which could represent their social mega-hinterland.

---

[1] *One of the most commonly used methods of analysing spatial autocorrelation of a given value is Moran's I index, developed by Patrick Moran (1950). The index describes the combined behaviour of point location (=site location) and point value (=site size) and whether the pattern is random, clustered or dispersed. The analytical tool is inferential, thus meaning that its interpretation is based on a null hypothesis, which in this case states that the pattern is random. Most software packages calculate the Moran's I value and the z-score and p-value in order to evaluate the statistical significance of the test. Moran's index calculates the global spatial autocorrelation of the data, thus measuring only the overall clustering of a given point dataset.*

[2] *A further development of the test has been suggested by Luc Anselin (1995), who proposes the "local indicators of spatial association" (LISA) in order to statistically evaluate the clustering occurring in a local spatial unit, starting from the principle that if there is no statistical evidence for global clustering this does not exclude the possibility of local clustering happening. Moreover, Anselin's local Moran's I statistics allows for the differentiation of clusters of low and high values and of spatial outliers. In this way it is possible for instance to identify a high value (in this case a big site = megasite) within a neighbourhood of low values (smaller Trypillia settlements) and measure the statistically significant scale of the neighbourhood. In other words, the data will set the scale at which a site becomes statistically 'mega' compared to its neighbouring sites.*