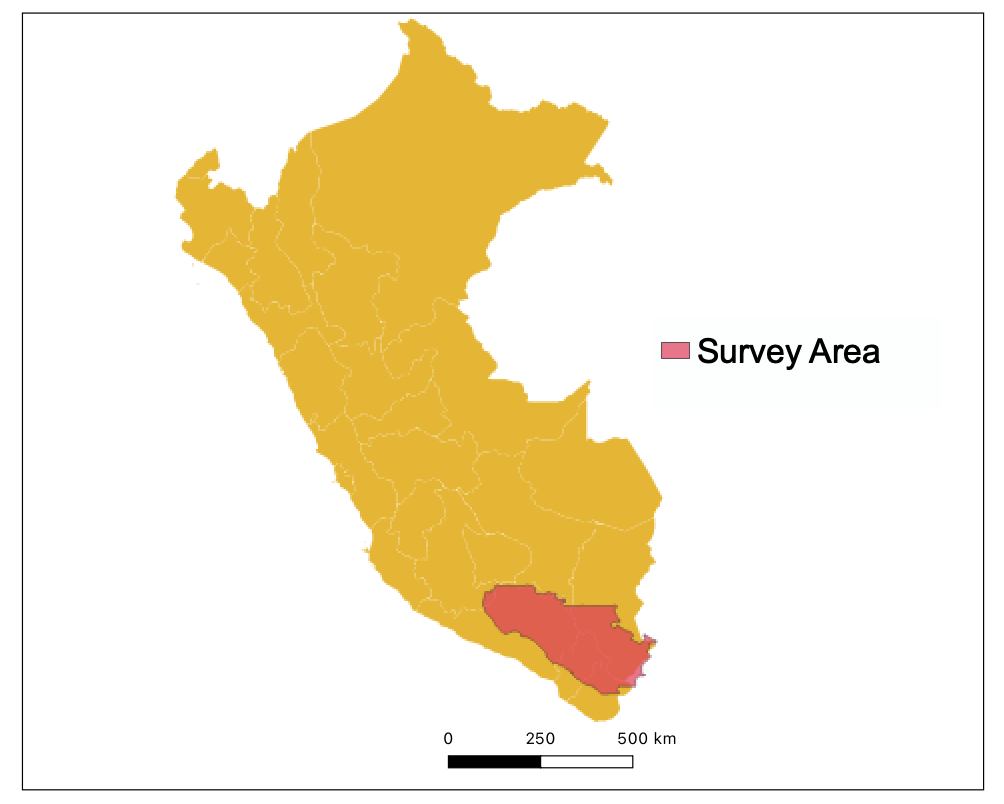
**DS-5999 Final Report**

**Summary**

The remains of historic ranches, or “estancias” as they are called in the local Spanish, dot the landscape of the highland Andes of modern-day Peru. The *Geospatial Platform for Andean Culture, History, and Archaeology (GeoPACHA)* is a cross-institutional campaign to provide a comprehensive documentation of archaeological sites in the Andean region, including the aforementioned estancias.

This report describes a geospatial suitability analysis of the locations of estancias retrieved from the GeoPACHA platform. Attention was focused on two specific types of geospatial attributes: vegetation indices and terrain attributes - both derived from remote sensing technologies and accessed through the *Microsoft Planetary Computer* interface. The results of the analysis show a positive correlation between the density of estancias and the local amount of vegetation, as well as clear tendencies towards particular types of terrain. Further work needs to be done to describe the anthropological motivations behind the patterns presented herein, but some preliminary hypotheses are offered.

**Background**



The area of focus of this analysis is a large area in the south of Peru (figure 1) covering a region of the *Central Andes* which are distinguished from other stretches of the Andes by their breadth and abundance of plateaus - making them more suitable for human habitation than other Andean highlands. Although the survey area is purposely designed to exclude the arid coast and the tropical inland, there is still a very apparent residual trend of wetter and hotter climates as one moves east across the region.

Figure 1: location and extent of survey area within Peru

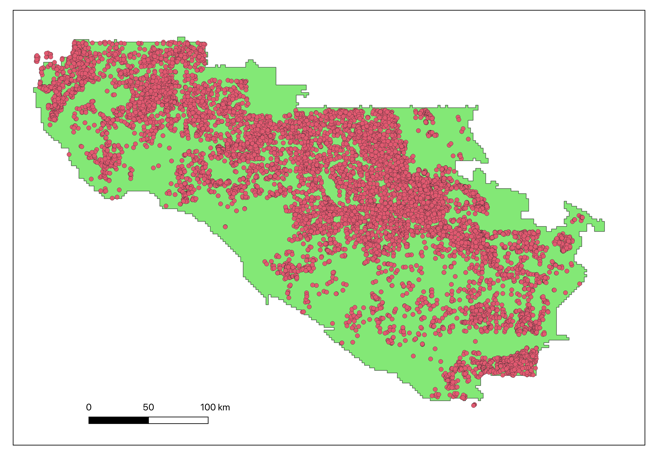
The archaeological sites examined in the GeoPACHA project, although not explicitly temporally filtered, date back to Incan settlement of the area in the 15th century and in limited numbers continue to exist into the modern day, spanning both pre- and post-colonial time periods. The estancias established in the central Andean highlands produce fur, meat, and working animals, and maintained herds of llamas, alpacas, or old-world livestock such as goats, sheep, and cattle. Rather than participating in a centralized economy, estancias were more likely to engage in trade with agricultural settlements within their close locale, remaining within a relatively small ethnic polity. The vast situational diversity of estancias and their fragmented but crucial ties to neighboring agricultural settlements motivates a deeper analysis that can contextualize their enduring role in Andean cultures.

Figure 2: Location of estancias within the survey area

Although consisting only a small part of GeoPACHA’s scope, the platform has collected a dataset of over 9,500 estancias within the survey region. Nearly all of these locations have been determined through manual search of satellite imagery, where typically only the footprints of their foundations remain intact and recognizable. The remote nature of their detection precludes more intimate/thorough methods of data collection, meaning that spatial coordinates compose the entirety of the dataset. Critically, there is no apparent way to determine the chronology of archaeological sites, so 80-year old and 600-year-old estancias must be considered together in our analysis.

**Methodology**

The analysis can be divided into the investigation of two factors: vegetation and topography. More complex relationships considering vegetation and topography together were explored, but ultimately did not yield any more significant insight than the examination of both factors independently. Analysis was facilitated by the Microsoft Planetary Computer, a remote server that is mounted onto a data center containing petabytes of remote sensing data made available through convenient APIs. Analysis was conducted in Python with the assistance of many packages within Python’s geospatial analysis ecosystem. The most prominently used of these packages were GeoPandas, XArray, Shapely, and PyProj, which provide a powerful suite of utilities that enable Python to be used as a powerful alternative to dedicated Geographic Information Systems software.

Vegetation data was derived from Moderate Resolution Imaging Spectroradiometer (MODIS) satellites, a class of sensors designed to collect regular and high-quality climate measurements on a global scale. Parallel analysis was conducted using two vegetation indices, the Normalized Difference Vegetation Index (NDVI) and the Enhanced Vegetation Index (EVI).

Both indices are based on the reflectance behavior of vegetation, absorbing red light and emitting near-infrared (NIR) light. Because the quantities of red and NIR light are otherwise very closely coupled, a difference in these two different bands of light strongly indicates vegetation. EVI improves upon the NDVI formula by correcting for atmospheric resistance and canopy background with the addition of several additional coefficients.

Terrain data was derived from the Copernicus digital elevation map (DEM). The Copernicus DEM is a product of data collected from the TanDEM-X satellite using synthetic-aperture radar. The DEM provides elevation mapping at a 30 meter resolution (1 elevation value per 30 x 30 m square) within an accuracy of 2 meters of altitude. The differences in elevation between neighboring cells was used to calculate the grade and aspect of the surface at the locations of each estancia.

For both analyses, vegetation and terrain data were extracted not only to estancia coordinates, but also to a set of negative locations generated uniformly across the survey area. Comparing the distribution of values for estancias locations against randomly generated locations shows how the distribution of estancias differed from the overall survey area. This approach offers the most practical way to discover interesting patterns, but is far from perfect. For example, although large bodies of water are removed from the survey area, it is possible a negative location was generated on top of a smaller body of water or a river. In those cases, a statistical difference will emerge because estancias are not built on surfaces of water, which is not a particularly deep anthropological insight. However, the analysis presented here is believed to be significant notwithstanding these subtleties.

The Python code developed over the course of this project has been compiled into a small python library made available to the Vanderbilt Spatial Analysis Research Lab (SARL) to facilitate similar analyses across a variety of archaeological domains. In particular, several utilities have been developed for using Microsoft Planetary Computer to extract geospatial data efficiently to a set of coordinate points.

**Vegetation Findings**

Both NDVI and EVI indices were extracted to positive and negative points for each month within the range from January 2018 to December 2020, in order to investigate seasonal differences in vegetation quantities. The results demonstrate a significant correlation between higher amounts of vegetation and a higher density of estancias. figure 3shows the average NDVI of estancias vs the average NDVI of the survey area over the period of 36 months. EVI shows a visually identical pattern, although the correlation is slightly but consistently less.

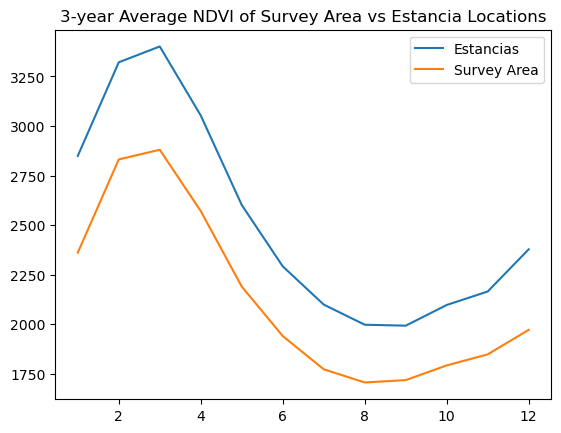


Figure 3

It is interesting that the average NDVI of estancias seems to be a near-fixed amount higher than the NDVI of the survey area, given that NDVI is a unitless ratio – so what a difference of NDVI’s represents isn’t immediately apparent. Because of the fixed difference, the proportionally greater amount of vegetation around estancias is greatest in the winter and early spring from June to November.

The high vegetation around NDVIs is intuitively understood as ranchers preferring locations with enough greenery for their livestock to graze. Indeed, it is a known pattern that high-altitude estancias are often located on highland marshes/springs known as *bofedales*. The novelty in this analysis is the seasonal aspect. For example, the locations where estancias are built are not just greener in the relatively lush summer, but have enduring vegetation even through the dry winter. That there are locations with year-round grazing is key to the feasibility of these establishments. Lastly, Andean ranching has a history of rotating herds between pastures, and this finding of enduring greenery hints that this tradition is motivated more by a need to avoid overgrazing rather than certain pastures becoming seasonally barren.

**Terrain Findings**

Grade and aspect were derived from DEM data, and then elevation, grade, and aspect were extracted to positive and negative points. Significant differences between random distribution and the estancia distribution were found for all three factors.

Firstly, in terms of elevation, there was a relatively intuitive soft upper bound, with estancias rarely existing above 5,000 meters. Perhaps more surprising is that estancias were also rarely located below 3,000 meters, and most were above 4,000 meters. Figure 4 shows the density distributions of estancias at different altitudes compared to the elevation distribution of the whole survey area.

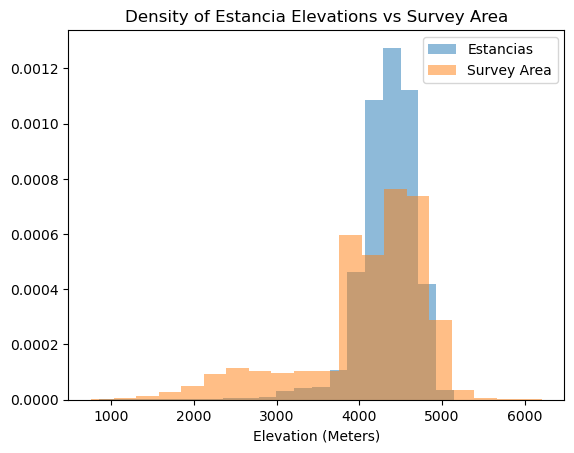
****

Figure 4

The upper bound can be easily explained by the fact that the highest altitudes are inhospitable terrain that don’t host much foliage. The explanation behind the lack of low altitude estancias is a more interesting question. A first hypothesis is that lower altitudes are more suitable to agricultural communities, meaning that those areas are both less likely to be available to ranchers, as well as estancias that are built at those low altitudes are more likely to be developed over later once abandoned and thus are not captured by a modern-day survey.

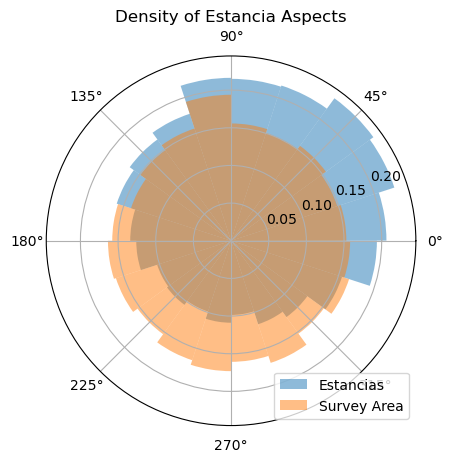


Figure 5

The next factor to be discussed is aspect, or the direction that the land is facing. Although in flatter terrains this may not seem to affect much at all, in the Andean setting it closely related to how much water flows over a surface, as well as the amount of the primarily easterly wind a piece of land receives. Figure 5 shows the distribution of estancias at different aspects, with a clear leaning towards northerly aspects (meaning north is the downhill direction).

This relationship is the most enigmatic one uncovered during the analysis, and the underlying phenomenon remains unclear. A possible first inquiry to follow is hydrological analysis. Water flows south towards the Pacific in the survey region, so south-facing surfaces may get an amount of water flowing over them that either makes them unattractive to ranchers, or more quickly erodes traces of estancias on south-facing aspects. In the latter case, the GeoPACHA dataset may have an important and unexamined deviation from the historical distribution of estancias.

The final facet of the terrain analysis was looking at the absolute grade or steepness of estancias. The results, shown in figure 6, are similar to those achieved in the elevation analysis. There is an intuitive explanation for the avoidance of high grades, as they are harder to navigate. The explanation for the lack of estancias at low grades is likely tied to the lack of estancias at low elevations described earlier in this section

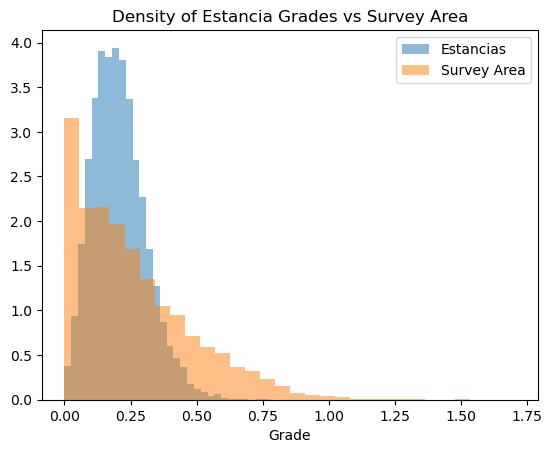


Figure 6

**Next Steps**

Not only are there several natural threads of inquiry left untouched by this analysis, but several new questions and motivations that have arisen from our new findings. Further analysis can help contextualize archaeological findings in the Andean region and deepen the historical understanding of the people that inhabited it.

* Incorporating climate and weather data could be a valuable piece to understanding how different locations wear over time, and the distributions of archaeological sites that may no longer leave any trace.
* The relationship between terrain, vegetation, and climate variables deserves further attention. Although the first look into these relationships did not immediately show interesting results, there is certainly room for more complex relations that require a more involved analysis to uncover.
* Not much attention was paid to the regional differences across the survey area in the examined variables in this analysis. Given the diverse nature of sites in terms of both purpose and history, there is a solid basis to think estancias located many hundreds of kilometers apart would be patterned differently.
* An analysis of the location of estancias relative to other estancias and nearby agricultural settlements can provide important insight into the role of estancias in the unique Andean economy

**References**

Didan, K.. *MODIS/Terra Vegetation Indices 16-Day L3 Global 250m SIN Grid V061. 2021*, distributed by NASA EOSDIS Land Processes DAAC, https://doi.org/10.5067/MODIS/MOD13Q1.061. Accessed 2023-03-31.

European Space Agency, Sinergise (2021). *Copernicus Global Digital Elevation Model*. Distributed by OpenTopography. https://doi.org/10.5069/G9028PQB. Accessed: 2023-03-31

Microsoft Open Source, Matt McFarland, Rob Emanuele, Dan Morris, & Tom Augspurger. (2022). *microsoft/PlanetaryComputer* <https://doi.org/10.5281/zenodo.7261897>

U.S. Geological Survey. *Landsat Enhanced Vegetation Index*. <https://www.usgs.gov/landsat-missions/landsat-enhanced-vegetation-index> Accessed 2023-03-31