Big Data & artificial intelligence: an archaeological revolution

Hèctor Orengo joined ICAC on March 2018 as a Ramón y Cajal researcher (postdoctoral senior grant from the Spanish government) at the GIAP research group, led by Josep M. Palet.

Since then, he is working on landscape archaeology and computational archaeology research programs. In particular, he is developing machine learning algorithms for automatic sites detection.

In collaboration with the McDonald Institute for Archaeological Research (University of Cambridge, UK), he applies these techniques to the paleohydrography construction of the Indo civilization (northwest India), from the Bronze Age (2600-1900 BC).

What are these analytical techniques?

The algorithms we have developed allow to capture bigger field areas with digital images (almost at a continental scale) and to transform data, either satellite or relief images, and this significantly improves the visibility of elements of archaeological interest. Specifically, they are two techniques that work in a complementary way and focus on different characteristics of the old structures they intend to locate.

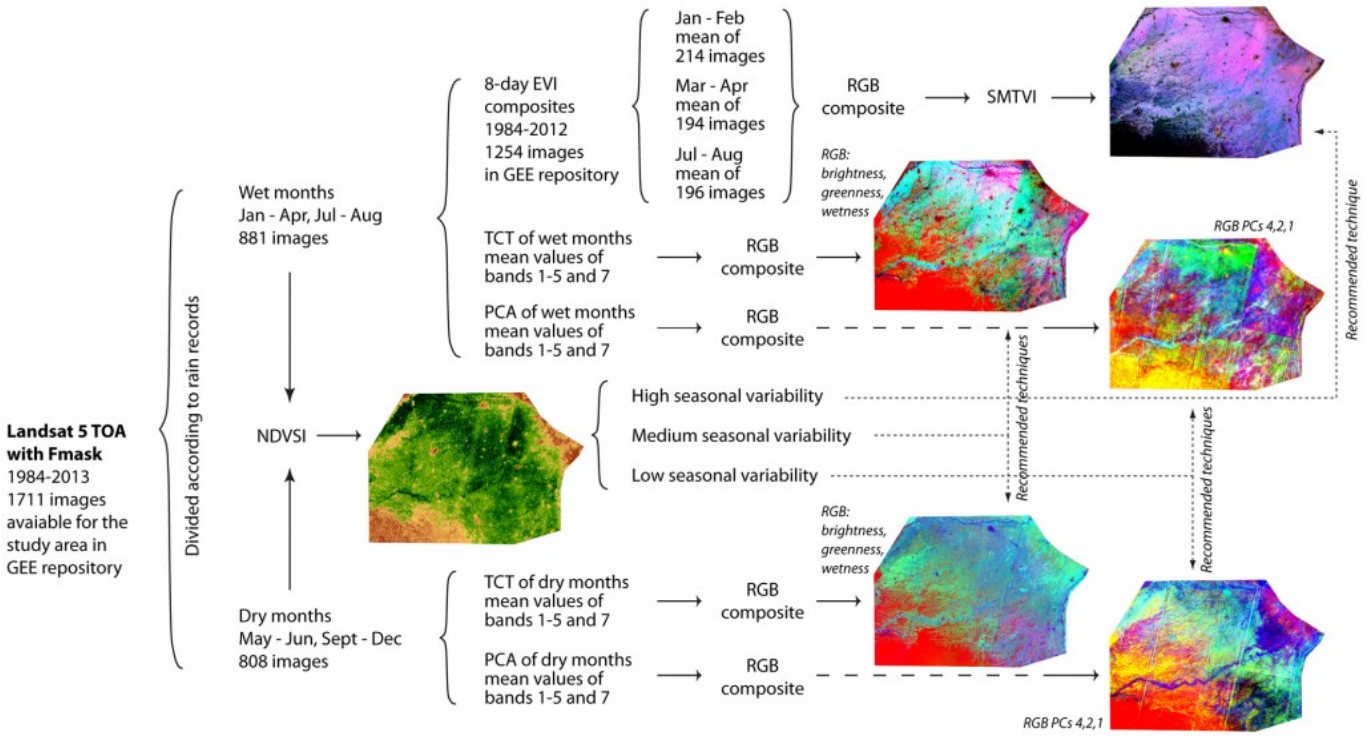
We use JavaScript programming language because we work with Google Earth Engine (not to be confused with Google Earth) and it is the default programming language in this online platform. We also work with other programming languages, like Python or R.

Using algorithms have allowed
us to reconstruct more than
20.000 km of paleorivers that

Contingut extret de la web de l'ICAC (www.icac.cat)

are not visible on the surface
neither with satellite images
without processing

Google Earth Engine is an interesting and powerful tool to work with satellite images data: it gives you access to petabytes! (1 petabyte equals one million GB). And, it covers the last 40 years of the Landsat program. Besides, this online platform enables us data processing on the cloud. Otherwise, it would be very difficult to download the huge amount of images that we need to make the analyses, and data processing would require having our own supercomputer.



Orengo & Petrie (fig. 5) (cc) BY-NC-SA

Where do you apply these methods?

At this time our algorithms are being applied with excellent results to many archaeological problems, such as the study of ritual landscapes of the English Neolithic, the tells of the first European Neolithic in Thessaly (Greece) or the landscapes of Greek colonization in the

Mediterranean, particularly in the Ampurias project within the framework of our project InterArPa (HAR2015-64636-P).

One goal of my last postdoctoral project was the reconstruction of the hydrographic network of the old Indo civilization. Moving forward this research line, using algorithms have allowed us to reconstruct more than 20.000 km of paleorivers that are not visible on the surface neither with satellite images without processing.

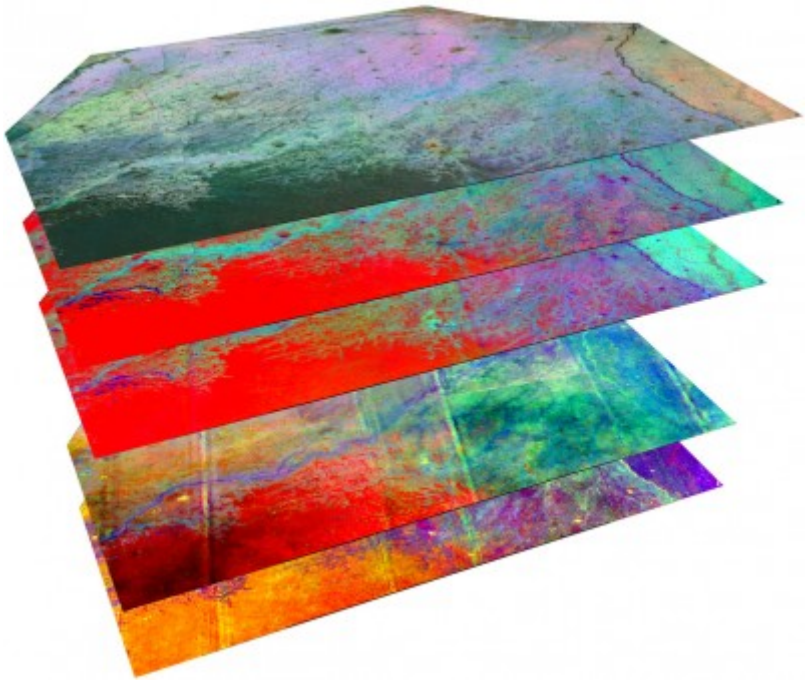
We are now developing machine learning and deep learning algorithms, often considered artificial intelligence

With these field surveying methodologies, we located many archaeological sites previously unknown and to draw relevant conclusions on the way of territory occupation, the economy and the eventual disappearance of this Indo civilization.

Have you set new goals, based on this experience?

Yes, of course. We have realized that working with such massive levels and variety of data carries a great effort to analyse the elements of archaeological interest. That is indeed the case when you work on areas that may be larger than the Iberian Peninsula and the archaeological remains have to be identified from the combination of different images resulting from the algorithms. Visual detection of possible sites can take months to be configured, although the images clearly show elements of interest.

That is why we are now developing machine learning and deep learning algorithms, often considered artificial intelligence, to analyse these images and automatically locate the archaeological elements we are interested in.

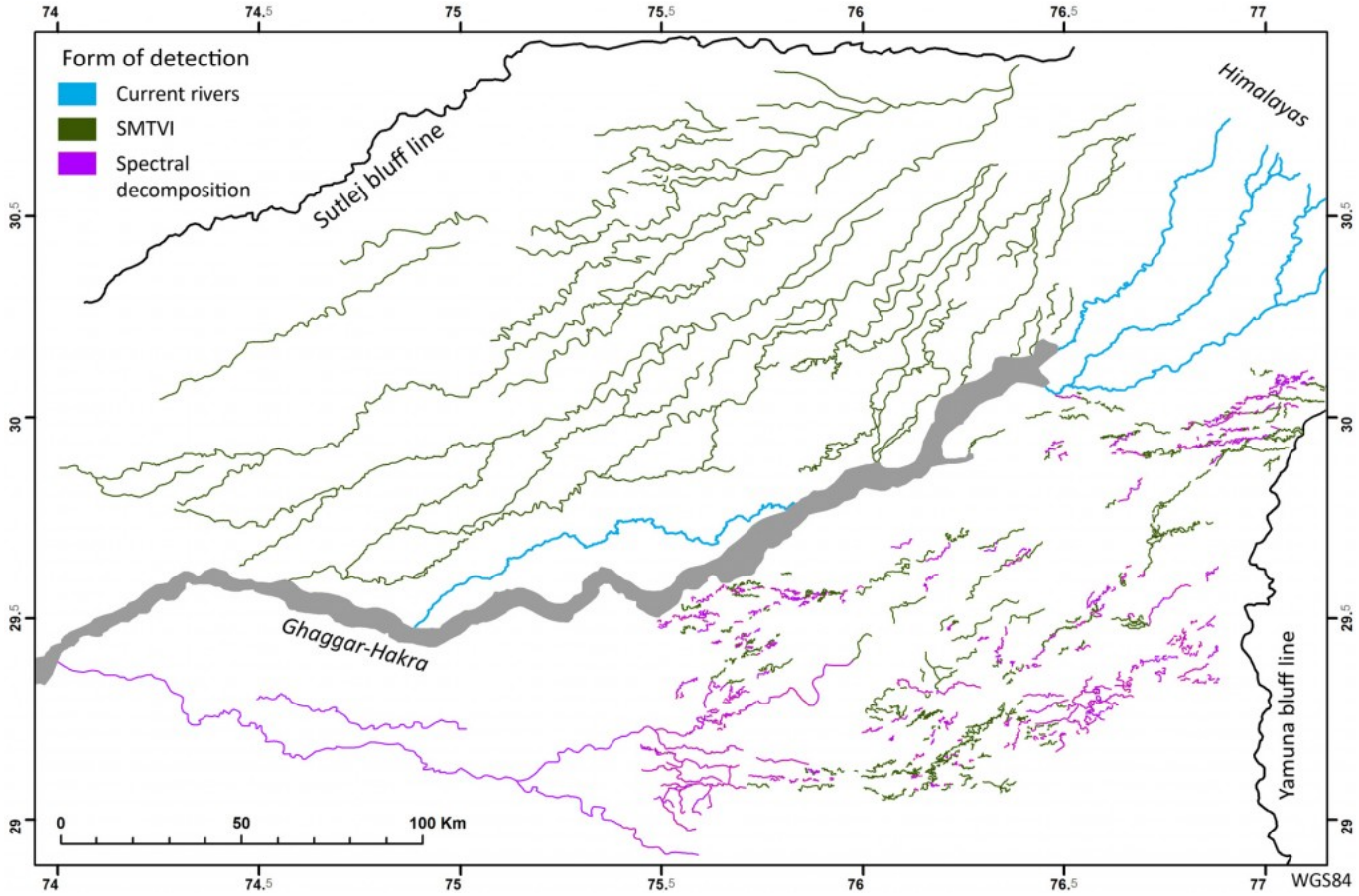


(cc) BY-NC-SA

How important can these results be?

There are always problems for which our work has a great application but, curiously, we are not aware of them when we develop the analytical tools. Related to Indian paleorivers, for instance, they not only have great relevance to understanding one of the great civilizations of the Bronze Age, but they can also become a very important water resource due to their current capacity of water accumulation. And that can be a very important issue in India, with millions population and where agricultural production is a major activity.

Our analyses can contribute to the forecasts of the evolution of aquifers in the area and provide valuable information on aspects such as the variability of agricultural productivity or the influence of climate change on the expansion of the Thar desert, in addition to providing cultural management tools to the government.



Orengo & Petrie (fig. 6) (cc) BY-NC-SA

Will we see more and more archaeologists in front of the computer screens?

Well, I think that most archaeologists already spend an important part of their time in front of the computer. But yes, archaeology increasingly uses complex computational techniques. In fact, fields such as archaeobotany or zooarchaeology are beginning to integrate complex analyses such as Bayesian statistics or network analysis.

Can we say that there is specialization, within archaeology?

Sure! There is a currently a trend to talk about computational archaeology but there are different subfields pretty active which, somehow, overlap with fields like digital archaeology or virtual archaeology. In fact, the application of new technologies in archaeological

research turn already up in the 90s. Today there are many scientific activities devoted to this field, such as the international conference CAA, Computer Applications and Quantitative Methods in Archaeology, and specialized academic journals, such as *Internet Archaeology* or, more recently, *Journal of Computer Applications in Archaeology*.

On the other hand, computational archaeology is part of the curricula of many universities that, especially in the United Kingdom but also in many other countries, offer undergraduate programs focused on different types of computational archaeology. There are also master programs on this specific discipline.

In Catalonia, we have researchers, such as Juan A. Barceló and César Carreras (both, UAB), who have been developing computational research for a long time and who have been leaders in the Iberian Peninsula in this field.



In the workshop at the ICAC. Orengo at right (cc) BY-NC-SA

You recently organized a workshop in the ICAC, in this line, right?

Yes, the GIAP organized at the ICAC, from October 30 to November 9, a workshop on advanced computational archaeology, within the framework of the InterArPa project (funded by the MINECO R & D program – HAR2015-64636-P). We worked with the use of remote sensing, GIS and other geospatial tools. Each participant developed their own research for which artificial intelligence or complex computational resources were necessary. The joint work resulted in methodologies and algorithms development, and a couple of papers based on this experience are already in preparation.

This workshop is an outcome of a research project led by the ICAC, but is also the result of your many years of experience

Yes, in my previous position as a postdoctoral researcher at the McDonald Institute for Archaeological Research at the University of Cambridge I was in charge of the development of new computational methodologies for the analysis of ancient landscapes. Now, at the ICAC, I want to continue with this line of research and that is why my Ramón y Cajal project has a strong technological approach.

- Garcia, Orengo et al. "Mapping Archaeology While Mapping an Empire: Using Historical Maps to Reconstruct Ancient Settlement Landscapes in Modern India and Pakistan". *Geosciences* 2019, *9*(1), 11. Open Access.
- Petrie, Orengo et al. "Remote Sensing and Historical Morphodynamics of Alluvial Plains. The 1909 Indus Flood and the City of Dera Ghazi Khan (Province of Punjab, Pakistan)". *Geosciences* 2019, *9*(1), 21. Open Access.
- Orengo & Petrie. "Multi-Scale Relief Model (MSRM): a new algorithm for the visualisation of subtle topographic change of variable size in digital elevation models". *Earth Surface Processes and Landforms* **2018**, 43(6), 1361-1369. Open Access.
- Orengo & Petrie. "Large-Scale, Multi-Temporal Remote Sensing of Palaeo-River Networks: A Case Study from Northwest India and its Implications for the Indus Civilisation". *Remote Sensing* 2017, *9*(7), 735. Open Access.