# Big Data i intel·ligència artificial: una revolució arqueològica

**Hèctor Orengo** es va incorporar el març de 2018 a l'ICAC com a investigador Ramon y Cajal a l'equip de recerca d'Arqueologia del Paisatge (GIAP), liderat per Josep M. Palet.

Des de llavors, està tirant endavant programes de recerca sobre arqueologia del paisatge i arqueologia computacional. En concret, està desenvolupant **algoritmes d'aprenentatge automàtic** (*machine learning*) per a la detecció automàtica de jaciments.

En col·laboració amb el McDonald Institute for Archaeological Research (University of Cambridge, Regne Unit), aplica aquestes tècniques d'anàlisi en la construcció de la paleohidrografia de la civilització de l'Indus (nord-est de l'Índia), de l'època del Bronze (2600-1900 aC).

## En què consisteixen, aquestes tècniques d'anàlisi?

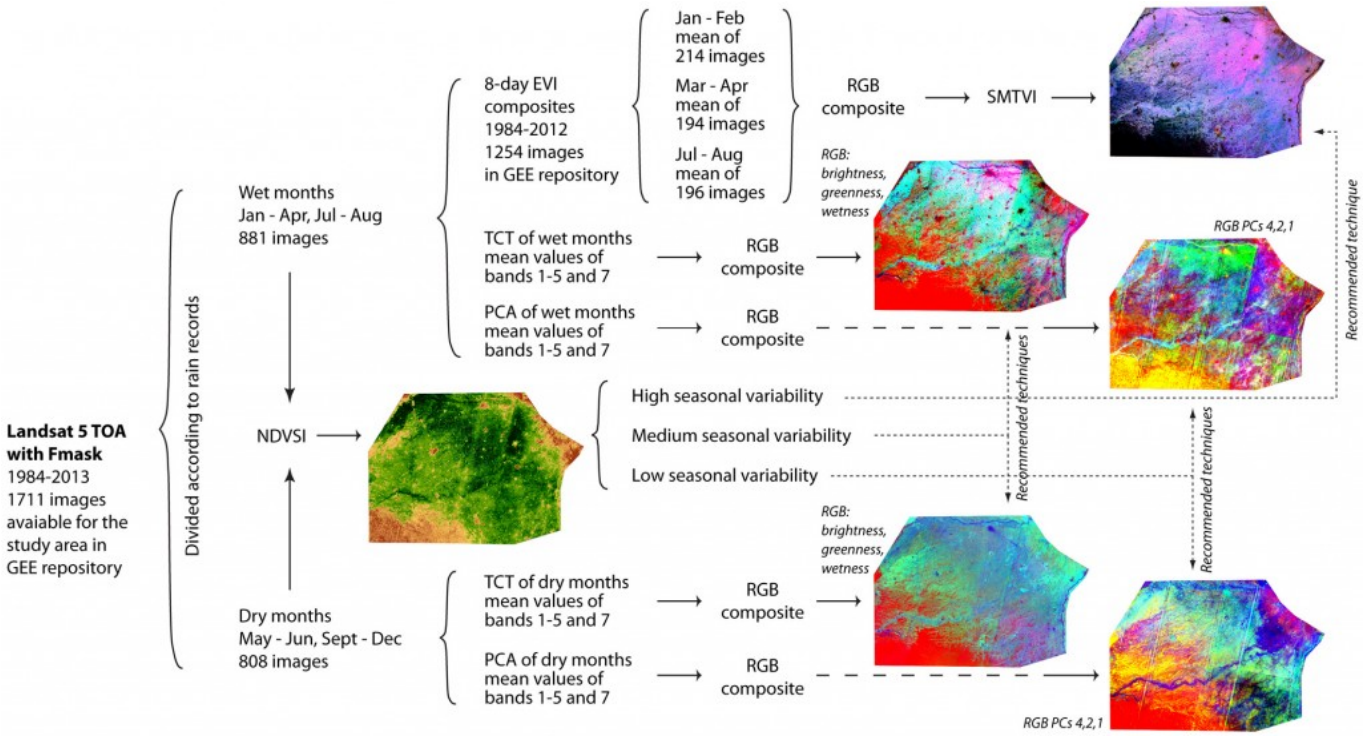
Els algoritmes que hem desenvolupat permeten captar amb imatges d'àrees de terreny enormes (gairebé a escala continental) i transformar les dades, ja siguin imatges satèl·lit o de relleu, i això millora significativament la visibilitat d'elements d'interès arqueològic. Concretament, són dues tècniques que funcionen de manera complementària i es centren en diferents característiques de les estructures antigues que pretenen localitzar.

Fem servir el llenguatge de programació JavaScript perquè estem treballant amb Google Earth Engine (que no és el mateix que Google Earth) i aquest és el llenguatge predeterminat en la plataforma en línia. També treballem amb altres llenguatges, com Python o R.

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

L'ús dels algoritmes ens ha permès reconstruir més de 20.000 km de paleorius que no són visibles en superfície ni amb imatges satèl·lit sense tractament

Un dels atractius de treballar amb Google Earth Engine és l'immens volum de dades satèl·lit a què dona accés la plataforma: petabytes! (1 petabyte equival a un milió de GB). A més, les dades satèl·lit disponibles cobreixen els darrers 40 anys de programa espacial . A més, ens permet fer el processament d'aquestes dades als seus supercomputadors al núvol. Sense aquest servei que ofereix la plataforma Google Earth Engine, seria molt complicat descarregar la quantitat d'imatges que necessitem per fer les anàlisis, i el seu processament requeriria disposar d'un supercomputador propi.




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

## On les apliqueu?

En aquest moments els nostres algoritmes s'estan aplicant amb excel·lents resultats a moltes problemàtiques arqueològiques, com ara l'estudi de paisatges rituals del Neolític anglès, els tells del primer neolític Europeu a Tessàlia (Grècia) o els paisatges de la colonització grega al Mediterrani, particularment al projecte d'Empúries en el marc del nostre projecte I+D InterArPa (HAR2015-64636-P).

Un dels objectius del meu darrer projecte postdoctoral era la reconstrucció de la xarxa hidrogràfica de l'antiga civilització de l'Indus. Seguint aquesta línia de recerca, l'ús dels algoritmes ens ha permès reconstruir més de 20.000 km de paleorius que no són visibles en superfície ni amb imatges satèl·lit sense tractament.

Ara estem desenvolupant algoritmes de d'aprenentatge automàtic i aprenentatge profund (*deep learning*), molts cops classificats com a intel·ligència artificial

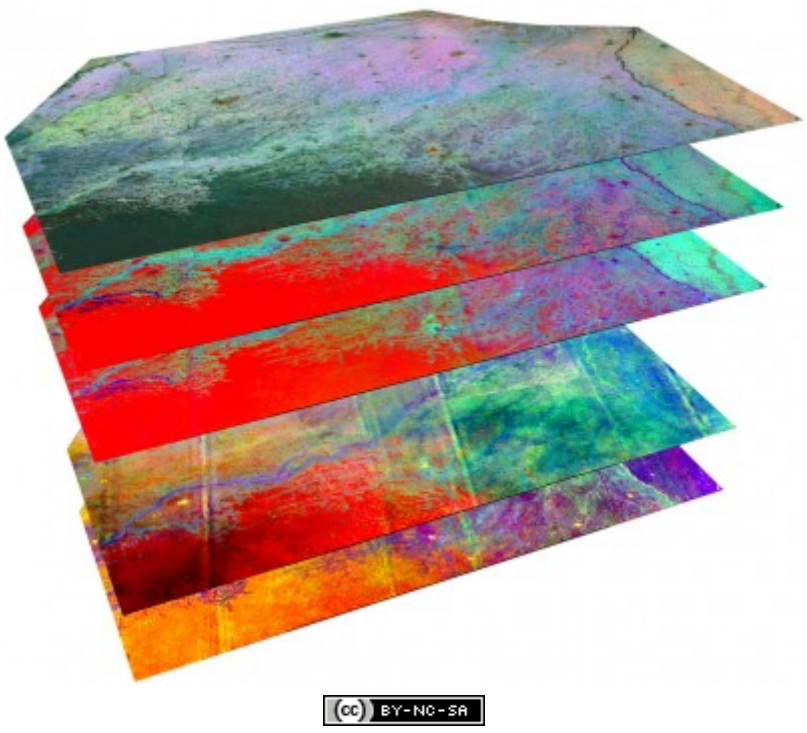
Hem pogut localitzar molts jaciments desconeguts fins aleshores i extraure importants conclusions sobre la forma d'ocupació del territori, l'economia i la eventual desaparició d'aquesta civilització.

## Us heu fixat nous objectius, a partir d'aquesta experiència?

Sí, i tant, ens hem adonat que treballar amb aquestes escales tan grans i amb una multiplicitat de dades elevades comporta un gran esforç d'anàlisi dels elements d'interès arqueològic. És a dir, quan treballes àrees que poden ser més grans que la Península Ibèrica i les restes arqueològiques s'han d'identificar a partir de la combinació de diferents imatges resultants dels algoritmes, resulta que la detecció visual de possibles jaciments pot trigar mesos a configurar-se, tot i que les imatges mostrin clarament elements d'interès.

Per això ara estem desenvolupant algoritmes de d'aprenentatge automàtic i aprenentatge

profund (*deep learning*), molts cops classificats com a intel·ligència artificial, per analitzar aquestes imatges i localitzar de forma automàtica els elements arqueològics que ens interessen.

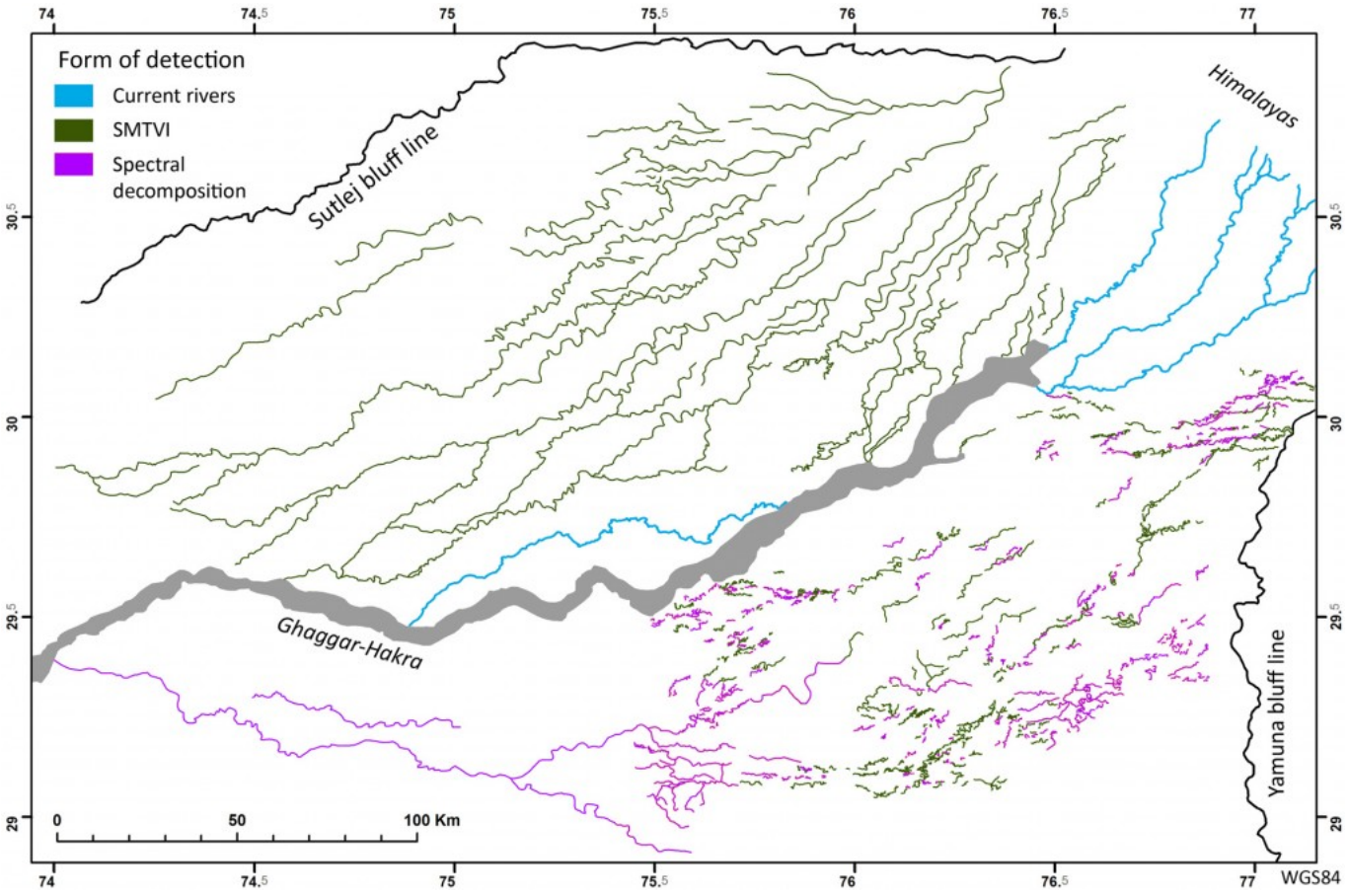



## Quina importància poden tenir, aquests resultats?

Curiosament sempre hi ha problemàtiques per a les que el nostre treball té una gran aplicació de les que no som conscients quan desenvolupem les eines. Pel que fa als paleorius de l'Índia, per exemple, no només tenen una gran importància per a comprendre una de les grans civilitzacions de l'edat del Bronze, sinó que també poden esdevenir un recurs hídric molt important per la seva capacitat actual d'acumulació d'aigua. I això és un factor de gran importància per a l'Índia, on viuen milions de persones i on la producció agrària és una activitat cabdal.

Les nostres anàlisis poden contribuir a les previsions d'evolució dels aqüífers de l'àrea i proporcionar informació de gran valor sobre aspectes com la variabilitat de la productivitat

agrària a l'àrea o la influència del canvi climàtic en l'expansió del desert del Thar, a més de proveir el govern d'eines de gestió cultural.




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

## Veurem cada vegada més arqueòlegs i arqueòlogues davant les pantalles d'ordinador?

Em sembla que la major part dels arqueòlegs ja es passen una part important del seu temps davant de l'ordinador. Però sí, cada cop més l'arqueologia fa servir tècniques computacionals complexes. De fet, camps com l'arqueobotànica o la zooarqueologia estan començant a integrar anàlisis complexes com l'estadística bayesiana o l'anàlisi de xarxes.

## Podríem dir que hi ha una especialització, dins l'arqueologia?

I tant! Ara hi ha una tendència a dir-li *arqueologia computacional* però també hi ha diversos subcamps força actius que, en certa manera, es solapen amb àmbits com l'arqueologia digital o l'arqueologia virtual. De fet, l'aplicació de noves tecnologies a la investigació arqueològica ja la trobem des dels anys 90. Avui dia hi ha moltes activitats científiques dedicades a aquesta disciplina, com el congrés internacional CAA, Computer Applications and Quantitative Methods in Archaeology, i revistes acadèmiques especialitzades, com la *Internet Archaeology* o, més recentment, la *Journal of Computer Applications in Archaeology*.

D'altra banda, l'arqueologia computacional forma part dels programes formatius de moltes universitats; sobretot al Regne Unit, però també a molts altres països, s'ofereixen mòduls de grau centrats en diversos tipus d'arqueologia computacional, i també existeixen màsters universitaris dedicats específicament a aquesta disciplina.

Aquí a Catalunya tenim investigadors, com en Juan A. Barceló i en César Carreras (tots dos a la UAB), que fa temps que desenvolupen recerca computacional i que han estat pioners a la península Ibèrica en aquesta matèria.



Moment d’un seminari a l’ICAC. Orengo, a la dreta (cc) BY-NC-SA

## Fa poc vau organitzar una activitat formativa, en aquesta línia, oi?

Sí, el GIAP va organitzar a l’ICAC, del 30 d’octubre al 9 de novembre, unes sessions de treball sobre arqueologia computacional avançada, en el marc del projecte InterArPa (finançat pel programa I+D del MINECO – HAR2015-64636-P). Vàrem treballar l’ús de teledetecció, SIG i altres eines geoespacials. Cada participant desenvolupava una recerca pròpia per a la que la intel·ligència artificial o el recursos computacionals complexos resultaven necessaris. El treball en comú va resultar en el desenvolupament de metodologies, algoritmes i processos de gran interès per tots els participants i un parell d’articles resultants d’aquest dies de feina ja estan en preparació.

**El seminari és fruit d'un projecte de recerca que es lidera des de l'ICAC, però també és el resultat dels teus anys de formació prèvia**

Si, a la meva anterior posició com a investigador postdoctoral al McDonald Institute for Archaeological Research de la Universitat de Cambridge estava al càrrec del desenvolupament de noves metodologies computacionals per a l'anàlisi de paisatges antics. Ara, a l'ICAC, voldria seguir amb aquesta línia de treball i per això el meu projecte Ramón y Cajal té un fort enfocament tecnològic.


- 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.