

Complutum

ISSN: 1131-6993

<https://dx.doi.org/10.5209/cmpl.72488> EDICIONES
COMPLUTENSE

Valoración de la calidad de los datos arqueológicos a través de la gestión de su vaguedad. Aplicación al estudio del poblamiento tardorromano¹

Leticia Tobalina-Pulido,² César González-Pérez³

Recibido: 16/09/20 / Aceptado: 08/11/20

Resumen. El presente trabajo, que constituye parte de la tesis doctoral de la primera autora, plantea una propuesta de medición de la calidad de los datos arqueológicos a través de la gestión de la vaguedad, concretamente de tres variables: incertidumbre, imprecisión e inexactitud. Si tenemos presente que el dato arqueológico es parcial desde su origen, valorar ese grado de parcialidad debería ser uno de los objetivos desde la obtención de los datos en las intervenciones arqueológicas y durante todo el proceso de investigación, con el objetivo de que nuestros resultados muestren las debilidades y fortalezas de nuestros datos y que los resultados de los análisis posteriores sean lo más honestos posibles con el lector. Aplicaremos la metodología propuesta a un caso de estudio centrado en el análisis de las dinámicas de poblamiento durante la Antigüedad Tardía, valorando la calidad de los datos como fase previa al análisis pero también como parte de los propios cálculos estadísticos.

Palabras clave: Arqueología teórica; gestión de los datos; vaguedad; incertidumbre; inexactitud; imprecisión.

[en] Assessment of Quality of Archaeological Data through Vagueness Management – An Application to the Study of Late Roman Settlements

Abstract. The present work, which is part of the PhD of the first author, proposes a measurement of the quality of archaeological data through the management of vagueness, specifically of three variables: uncertainty, imprecision and inaccuracy. If we bear in mind that archaeological data is partial from its origin, assessing its degree of partiality should be one of the objectives from the time the data is obtained in the archaeological interventions and during the whole research process, with the aim that our results show the weaknesses and strengths of our data and are as honest as possible. We will apply the proposed methodology to a case study focused on the analysis of the dynamics of settlement during Late Antiquity, assessing the quality of the data as a preliminary phase to its statistical analysis.

Keywords: Theoretical archaeology; data management; vagueness; uncertainty; inaccuracy; imprecision

Sumario. 1. La importancia de valorar la calidad de los datos. 1.1. Los datos en Arqueología. 1.1.1. La definición del dato arqueológico. 1.1.2. Los problemas de los datos arqueológicos. 1.2. Diferentes propuestas para el tratamiento de la vaguedad. 2. Enfoque Teórico Propuesto. 2.1. La incertidumbre. 2.2. La imprecisión. 2.3. La inexactitud. 2.4. El error. 3. Criterios de aplicación. 3.1. La incertidumbre de nuestros datos. 3.1.2. La imprecisión de nuestros datos. 3.1.3. La inexactitud de nuestros datos. 4. Aplicación práctica. 4.1. La localización. 4.2. La datación. 4.3. La tipología. 5. Problemas y perspectivas de mejora. Bibliografía

Cómo citar: Tobalina-Pulido, L.; González-Pérez, C. (2020). Valoración de la calidad de los datos arqueológicos a través de la gestión de su vaguedad. Aplicación al estudio del poblamiento tardorromano. *Complutum*, 31 (2): 343-360.

¹ Este artículo ha sido realizado durante una estancia de un mes en septiembre de 2020 en el Incipit CSIC en Santiago de Compostela, bajo la supervisión del Dr. César González-Pérez. La estancia fue financiada gracias a una beca de concurrencia competitiva COST Action SEADDA (Saving European Archaeology from the Digital Dark Age, www.seadda.eu) ECOST-STSM-Request-CA18128-47039.

² Université de Pau et des Pays de l'Adour, E2S UPPA, ITEM, Pau, France
Universidad de Navarra, Facultad de Filosofía y Letras, Departamento de Historia, Historia del Arte y Geografía
leticiatobalina@gmail.com

³ Incipit CSIC
cesar.gonzalez-perez@incipit.csic.es

1. La importancia de valorar la calidad de los datos

El paulatino aumento de la cantidad de los datos producidos – sea la disciplina que sea –, hace necesario considerar su calidad para poder generar conocimiento a partir de su posterior tratamiento. En el caso de los datos arqueológicos, “*évaluer la qualité de l’information est nécessaire afin de pouvoir mesurer notre «légitimité» à formuler certaines hypothèses et interpréter les résultats des analyses effectuées avec ces mêmes données au sujet des structures et des dynamiques spatiales passées*” (Fusco 2016). Si cuantificamos la evidencia disponible en cuanto a calidad de los datos, descubriremos la probabilidad de cometer un error, por ejemplo, al tratar datos poco precisos como si lo fuesen mucho. También podremos considerar cuál es la vaguedad ontológica de los datos (es decir, su imprecisión), su vaguedad epistémica (es decir, su inexactitud) o su usabilidad. Medir todas esas variables conlleva en ocasiones una pérdida de información, más aún cuando tratamos de evaluar la calidad de los datos arqueológicos, es decir, de una ciencia humana, siendo la calidad de los datos altamente compleja y, a menudo, nada fácil de establecer. Es por ello que la elección de las variables a medir será uno de los primeros retos a tener en cuenta.

Uno de los principales problemas al que nos enfrentamos en Arqueología es la variedad y parcialidad de los datos desde su origen. El aumento de las intervenciones arqueológicas a partir de los años ochenta, junto al desarrollo de la prospección como técnica arqueológica, supusieron un gran incremento en la cantidad de informes arqueológicos, acumulándose por decenas en los archivos de la Administración. Paralelamente, el estudio del poblamiento empezó a cobrar fuerza, multiplicándose los intentos de estudios a escala local y regional. En estos análisis, los trabajos de prospección se afianzaron. Sin embargo, ni la acumulación de documentación ni la necesidad de homogeneizar la información producida para analizar los patrones de asentamiento llevaron a una reflexión teórica de la problemática de la calidad de los datos. Una gran cantidad de información que, si queremos explotar de manera honesta y responsable, deberíamos valorar previamente para poder estandarizarla. Curiosamente, esto no es así a pesar del “giro científico” que, según algunos autores (Marila 2017; Sørensen

2016), está realizando la Arqueología, y que, en principio, debería promover la preocupación por la adecuada cuantificación de la vaguedad de los datos.

Si tenemos presente que el dato arqueológico es parcial desde su origen, valorar ese grado de parcialidad debería ser uno de los objetivos desde la obtención de los datos en las intervenciones arqueológicas, con el objetivo de que nuestros resultados muestren las debilidades de nuestros datos. En esa misma línea, P. Diarte y M. Sebastián (2011: 264) señalan que “uno de los principales déficits de la Arqueología es la visión sesgada que ofrece el estudio exhaustivo de un yacimiento frente, en la mayor parte de los casos, al parco conocimiento de la relación de este con sus yacimientos vecinos o no tan cercanos”. Sin embargo, el problema no está en el diferente tratamiento que ha sufrido cada yacimiento sino en la manera en la que gestionamos y analizamos los datos. Trabajar con entidades arqueológicas de prospección y de excavación conjuntamente no debería suponer un problema en el análisis, siempre y cuando se realice una correcta evaluación de la calidad, empleando una serie de criterios para homogeneizar el posterior tratamiento de la información (Fusco 2016). Esta tarea no es sencilla y requiere de un importante proceso de reflexión previo.

Lo ideal, por tanto, en cualquier estudio o investigación arqueológica sería realizar una evaluación de la calidad de los datos durante el proceso de la captura de éstos, es decir, durante la realización de las intervenciones arqueológicas, tanto en la prospección y la excavación como en otro tipo de intervención (fotografía aérea, estudios de materiales arqueológicos, análisis de materiales, etc.). Como bien afirman G. Fusco *et al.* (2014: 2), tenemos que integrar la incertidumbre en todo el proceso de producción del conocimiento, desde la definición del tema de estudio hasta los análisis y la modelización de los fenómenos. Y, además, sería también ideal que existiese un marco conceptual y metodológico que nos permitiese hacer esto de forma simple, clara y sistemática, sin necesidad de construir una aproximación teórica nueva para cada intervención. Todo esto nos permitiría determinar la incertidumbre generada en razón de errores en la toma de los datos, así como de inconsistencias en el momento de registro de los datos (Sánchez Trigueros 2013: 17).

Pese a su importancia, no son muchos los autores a nivel europeo (Bevan *et al.* 2013; De

Runz 2013). y, aún menos, a nivel estatal, que tienen en cuenta esta cuestión en sus trabajos (Parcero Oubiña y González Pérez 2011: 114; Barceló 2008; Sánchez Trigueros 2013). La precisión de la información, del grado de detalle y de la fiabilidad han sido tratados por la bibliografía, pero “este tipo de debates se han extendido mucho menos al campo de la propia gestión de la información” (Parcero-Oubiña y González-García 2016: 482).

Algunos autores consideran que realizar análisis cuantitativos sobre datos arqueológicos no dará resultados fiables o válidos, ya que “[...] es la calidad de nuestros datos la que genera un mayor grado de imprecisión en los resultados” (Baena Preysler y Ríos Mendoza 2006: 25). Esta afirmación, sin embargo, no es del todo correcta, ya que no es la calidad de nuestros datos la que generará la imprecisión, sino que la imprecisión es parte intrínseca y consustancial a los datos. Dicho de otro modo, cada dato nace ya con cierta imprecisión y es nuestro deber realizar una valoración de esta para poder mostrar las carencias y virtudes de nuestros datos. Así, por ejemplo, si tenemos en cuenta la perceptibilidad, es decir, “la probabilidad de que determinados conjuntos de materiales arqueológicos puedan ser descubiertos con una técnica específica” (Ruiz Zapatero y Fernández Martínez 1993), son muchos los datos que no podremos documentar, sea por las características del manto vegetal, por el potencial arqueológico de la zona o por otras cuestiones. Sin embargo, esto no nos impedirá poder realizar un análisis histórico pese a contar con datos parciales, pudiendo valorar el grado de incertidumbre de los datos y, por ende, la calidad de estos.

1.1. Los datos en Arqueología

1.1.1. La definición del dato arqueológico

Quizás uno de los términos más usados en lo personal y lo profesional, sin importar la disciplina, es el término “dato”. Bancarios, económicos, sociales, en los informativos, en los periódicos, en internet; estamos rodeados de datos, de personas que hablan de datos.

En Arqueología también hacemos referencia a ellos continuamente. Muchas veces no nos paramos a pensar qué son o a qué hacemos referencia con dicho vocablo. En nuestra disciplina, no son muchos los autores (al menos a nivel hispanico) que han reflexionado en tor-

no a este concepto (Sánchez Trigueros 2013), quizás debido al todavía limitado número de trabajos de S.I.G.⁴ o al uso de otros términos como el de “evidencia” 593, “registro arqueológico” (Criado-Boado 1993) o “información arqueológica” (Diarte Blasco and Sebastián López 2011; Parcero-Oubiña and González-García 2016), dejando la palabra dato para las disciplinas de ciencias experimentales.

En una lectura materialista, el dato arqueológico no es más que la materia prima del proceso del conocimiento. Esto podría ser discutible, ya que como bien señala E. Domínguez Berenjeno (2001), el dato es un producto cultural y no puede ser aprehendido de manera séptica. Los datos no se nos presentan, sino que son construidos durante el proceso de investigación.

En Francia, la reflexión teórica es algo más extensa. En los años noventa, en su estudio sobre la formalización de los datos arqueológicos, B. Desachy (1990) los define como “*des vestiges ou ensembles de vestiges matériels, avec ces informations dont ils sont porteurs. Ces vestiges font l’objet d’une analyse, d’un traitement, d’un commentaire, en bref d’une construction intellectuelle, mobilisant un certain “savoir” [...], et débouchant sur une interprétation*”. En esta misma línea, A. Chaillou propone la siguiente definición: “*La donnée est tout élément que l’archéologue trouve dans la fouille [ou la prospection], mobilier, objet, couches, ainsi que les liens entre ces éléments, des observations et des interprétations. [...] pourra aussi être le résultat de l’étude de ces premières données, études des mobiliers...*” (Chaillou 2003, 29). Por su parte, en su diccionario de Arqueología, I. Shaw y R. Jameson lo definen como las “*measurements made on archaeological objects (using the term in the widest possible sense, to include for example artefacts, features, assemblages, sites), as well as counts of such objects*” (Shaw and Jameson 1992), incluyendo dentro de esas medidas tanto las físicas o químicas como las tipológicas. C. Galán Saulnier y J.L. Sánchez Meseguer (2009, 274) distinguen tres tipos: los alfanuméricos, los numéricos y los gráficos, es decir, textos, cifras e imágenes.

⁴ Consideramos conveniente realizar una precisión. Indicamos que el número de trabajos que emplean S.I.G. es limitado porque son muchos los investigadores que lo utilizan para la elaboración de cartografía, pero no para la gestión de los datos espaciales y alfanuméricos, objeto de un S.I.G.

En ese sentido, los datos arqueológicos pueden ser desde un pequeño fragmento cerámico hasta una fase de una ciudad, desde un hallazgo aislado hasta un sitio rural; es decir, presentan diversa naturaleza, procedencia y escala. Por tanto, no son algo homogéneo, sino muy variado, heterogéneo y parcial. Teniendo en cuenta la fuente de origen, podemos distinguir entre los datos “brutos” (aquellos obtenidos directamente en las intervenciones arqueológicas) y los de segunda mano (la literatura que estas producen). A. Chaillou (2003, 30) distingue dos categorías, los primarios y los tratados; así, los primeros serán aquellos que “*proviennent de l’observation directe du site, de la fouille, et de la documentation que l’on peut en tirer*” (2003, 30). Por su parte, los datos tratados son “*la résultante d’études menées sur les données primaires*” (2003, 30).

Por tanto, los datos arqueológicos son la materia prima que extraemos, tanto de informes de intervenciones arqueológicas como de monografías o artículos especializados, y de la que generaremos posteriormente información. Sin embargo, la “capacidad del investigador de construir sus datos por medio de la representación hace que no exista una objetividad completa en el dato” (López Aguilar 2012, 251). Por ello, el dato tiene una naturaleza compleja que lo hace objetivo pero, a su vez, subjetivo; la persona encargada de la toma de datos va a tratar de tomar los datos de una manera objetiva pero su mirada siempre le dará un valor subjetivo.

1.1.2. Los problemas de los datos arqueológicos

A partir de los años 1980, con el crecimiento de la importancia de la prospección en los trabajos arqueológicos, pero también con el impulso de la Arqueología preventiva, el volumen de datos arqueológicos ha ido aumentando progresivamente (García Sanjuán 2003; Trément 2013), hasta tal punto que, como señala F. Trément (2013: 53), “*le site archéologique passe dans les années 1970-80 du statut de « denrée rare » à celui d’« objet surabondant »*”. Este desarrollo y su relación con otras ciencias, técnicas y disciplinas, especialmente las más “científicas” (Barceló 2008; Marila 2017; García *et al.* 2018), ha hecho que la Arqueología sea cada vez más compleja, convirtiéndose en una inagotable fábrica de datos (Galán Saulnier y Sánchez Meseguer 2009: 274). Estos datos son muy diversos y, a me-

nudo, pueden llegar a constituir un problema para los arqueólogos si no se gestionan adecuadamente.

Los datos generados en las intervenciones arqueológicas y por la bibliografía especializada, en general, son fuertemente heterogéneos: están a diferentes escalas, en ocasiones el acceso a la documentación es restringido o parcial (algunos informes son de consulta limitada) y/o presentan diferente naturaleza (geográfica, arqueológica, histórica). De esta manera, hay una gran variedad de escalas de trabajo, desde el estudio aislado de un conjunto de materiales hallados de manera fortuita hasta la excavación bien documentada de otros; y se usan técnicas diferentes de recogida de los datos, siendo todavía mínimos los espacios y yacimientos en los que contamos con una combinación de prospecciones y excavaciones intensivas y en extensión (Tobalina-Pulido 2020).

Esta gran heterogeneidad se traduce a menudo en la necesidad de usar datos de forma conjunta (comparada, interrelacionada, etc.) con niveles de vaguedad muy diferentes. En Arqueología, además, se ha prestado atención “a definir los márgenes de precisión y resolución con los que la información es manejada, buscando [...] incrementar el grado de certeza con el cual podemos comprender los resultados del análisis de la información arqueológica” (Parcero-Oubiña and González-García 2016: 483). En ocasiones, no se trata tanto de buscar la certeza o la precisión de los datos, sino de medir ese grado con el objetivo de mostrar cuál es el grado de calidad de los datos con los que trabajamos. Lo que no se mide no se puede gestionar ni administrar. Pese a esta cuantificación de la calidad de los datos, seguirán siendo inciertos y “los resultados tienen que ser considerados como probabilidades, antes que como una respuesta definitiva” (Barceló 2008, 23).

En la mayoría de los casos contamos con datos de prospecciones, pero son mínimos los yacimientos donde contamos con datos de excavación en extensión. Tenemos escasos datos sobre algunos yacimientos por falta de publicación de los resultados frente a otros que cuentan con memorias inéditas, artículos específicos e incluso monografías; diferencias de la calidad y la cantidad de la información aportada por las memorias administrativas, contando en algunos casos con apenas las coordenadas y en otros con

cientos de páginas detalladas con los materiales recogidos, planos y fotografías; gran heterogeneidad en las fechas de captura de los datos. Esto incluye, por ejemplo, desde descubrimientos en el siglo XIX sin otras menciones posteriores, hasta yacimientos recientemente hallados gracias a la Arqueología de urgencia. Además, las cronologías suelen ser imprecisas, sobre todo para algunos periodos históricos, con resoluciones que van desde medio siglo hasta más de dos siglos.

1.2. Diferentes propuestas para el tratamiento de la vaguedad

A lo largo de las últimas décadas, la gestión de la vaguedad de los datos arqueológicos ha ido cobrando importancia en la investigación; algunos autores la han tratado de aplicar de manera práctica en sus proyectos y otros han reflexionado de manera teórica. Si bien la bibliografía española sobre esta cuestión no es muy abundante, en el ámbito anglosajón y francés, sobre todo a partir de los años noventa, sí se ha llevado a cabo una reflexión interesante. Conviene mencionar aquí el proyecto « *Faire Science avec l'Incertitude* », de carácter interdisciplinar; iniciado en el 2013 en la *Université Nice Sophia Antipolis*, tiene como objetivo abordar la incertidumbre desde las especificidades epistemológicas y metodológicas de las ciencias sociales y humanas (Fusco *et al.* 2014: 2 y ss.).

A nivel teórico, algunos autores han establecido una jerarquización para la incertidumbre. M. Skeels *et al.* (2008) establecen tres niveles jerárquicos y dos transversales, siendo el nivel más bajo el relativo a la precisión de la medida, el medio la completitud de la información y el nivel más alto los datos producidos por inferencias C. De Runz (2008), en su tesis doctoral, plantea que la naturaleza de las imperfecciones de la información se compone de cuatro aspectos: incertidumbre, imprecisión, ambigüedad e incompletitud que, a su vez, se subdividen en dos subfactores. Por su parte, F. Sánchez Trigueros (2013: fig. 2) propone 10 factores a tener en cuenta.

A la hora de gestionar la incertidumbre a nivel práctico, es la teoría de los conjuntos difusos (*fuzzy sets*) la que más ha sido aplicada en Arqueología. Se trata de una lógica

aplicable a conceptos que no tienen límites claros, introduciendo un grado de vaguedad en las cosas que califica, asignando para ello valores en el intervalo de 0 a 1. Un ejemplo fácil de comprender es el siguiente. Una persona que mide 180 cm, ¿en qué grado podemos considerar que es alta? La lógica difusa nos permite realizar una distribución de posibilidades y poder afirmar que es “poco mediano” y “más bien alto”, ya que no hay un valor cuantitativo que defina los términos “alto” o “bajo”.

En Arqueología, si bien no vamos a enumerar todos los trabajos realizados, podemos destacar algunos. J. Barceló planteó, ya en los años noventa para su proyecto PYGMALION, la aplicación de la lógica difusa para el estudio de la cerámica fenicia (Barceló 1996: 158 y ss.). S. Hermon y F. Niccolucci (2002: 227–29) también aplican la teoría de la lógica difusa al estudio de cincuenta piezas líticas procedentes de Israel y su clasificación. Su planteamiento consiste en la asignación de un coeficiente difuso. Las piezas se hicieron analizar por cinco expertos que debían asignar un número (de 0 a 1) en función de la seguridad que tenían al atribuir una u otra clasificación a la pieza. El resultado es una tabla que permite calcular el coeficiente R considerando la seguridad que los expertos tienen al asignar la clasificación y contar así con una referencia sobre la incertidumbre del estudio. C. de Runz y E. Desjardin se centran en las vías romanas, considerando las variables de localización, datación y orientación en su estudio (2011). Su análisis, al igual que el anterior, se basa en la aplicación de la teoría de los conjuntos difusos. En esta misma línea, J. Fusco (2016) también parte de dicha teoría para el estudio del poblamiento en Siria y Líbano, partiendo de los datos del proyecto *PaléoSyr/PaléoLib*.

También son interesantes las aportaciones realizadas por el proyecto *ArchaeDyn* (Favory *et al.* 2008), donde la gestión de la incertidumbre, centrada sobre todo en la cuestión cronológica (Fig. 1), fue uno de los pilares de la investigación. *ArchaeDyn* hace tratamiento de la calidad de los datos cronológicos a través de la gestión de la incertidumbre y la precisión. Por ejemplo, para la cronología, optan por cifras que indiquen la incertidumbre de los años mientras que, en el caso de la precisión, se limitan a “preciso” o “impreciso”, como podemos ver en la tabla siguiente.

Fecha de inicio (TPQ)	5000
Incertidumbre de esta fecha de inicio (+- 200) si es necesario, sino poner ()	200
Fecha de final (TAQ)	4001
Incertidumbre de esta fecha de final (+- 50) si es necesario, sino poner ()	50
Precisión de la datación: preciso o impreciso (dos valores posibles)	Preciso
Método de datación (tipo-cronológico, C14, dendrocronología...)	Tipo-cronológico

Fig. 1. Ejemplo de aplicación de los criterios cronológicos establecidos para el proyecto ArchaeDyn mediante un ejemplo (Traducido según la tabla de Favory y Nuninger 2008: 55)

Otro de los métodos más empleados es el basado en procedimientos estadísticos bayesianos en el que las observaciones se emplean para inferir la probabilidad de que una hipótesis pueda ser cierta. Uno de los que ha aplicado este método es K.L. Wescott. Para este autor, en una zona prospectada en la que conocemos un alto número de yacimientos, la incertidumbre del resultado es menor. Lo mismo ocurriría para una zona de potencial bajo que ha sido prospectada y en la que no se localizaron yacimientos. En su estudio, el grado de calidad de la información arqueológica inicial es la que condiciona la precisión del resultado (Mehrer y Wescott 2005: 56–61). Finalmente, también se ha empleado el método aorístico, un enfoque que permite cuantificar las incertidumbres temporales e incorporarlas en los posteriores análisis. Partiendo de un conjunto de puntos y sus distribuciones de probabilidad temporal, tendremos un número determinado de patrones espacio-temporales posibles. Podemos generar diferentes posibilidades de patrones espaciales basándonos en las probabilidades temporales y obtener así una distribución con los resultados más y menos probables (Bevan *et al.* 2013).

2. Enfoque Teórico Propuesto

El enfoque teórico que proponemos se basa en el concepto de *vaguedad*, referida al hecho de que los datos no son “limpios” y perfectos,

sino que a menudo no describen el mundo de una forma clara y precisa. La vaguedad se manifiesta a través de diferentes aspectos, como se detalla a continuación.

- Por incertidumbre nos referimos a las dudas que podemos tener sobre la correspondencia entre un dato y aquello que pretende representar. Por ejemplo, si alguien nos dice que cree que Alejandro Magno tuvo cuatro hijos pero no está seguro, está diciendo que no sabe con certeza si esto es así o no. Pero si nos dicen que Alejandro Magno tuvo un padre y una madre, estaremos seguros de que es cierto; no hay incertidumbre. La incertidumbre suele venir dada por la vaguedad epistémica, por conocimiento imperfecto del mundo y/o por el error de las medidas.
- Por imprecisión nos referimos al grado de detalle o cantidad de información que proporciona un dato. Por ejemplo, si decimos que el Everest tiene una altitud de entre 8000 y 9000 metros estamos siendo bastante imprecisos; en cambio, si decimos que tiene una altura de 8848 metros estamos siendo mucho más precisos, porque aportamos más información. A menudo, la imprecisión viene dada por la vaguedad ontológica inherente a las cosas, es decir, aquella que depende exclusivamente de la naturaleza de las cosas que estudiamos, sin que influya nuestra

percepción o nuestra forma de representarlas. Por ejemplo, una montaña posee límites imprecisos porque no comienza ni finaliza en un lugar concreto, sino que existe un cambio gradual (como nos muestra la Paradoja de Sorites).

- Por inexactitud nos referimos a la diferencia entre un dato y el valor verdadero de aquello que trata de representar; es decir, un dato es más exacto cuanto más se acerca al valor verdadero de aquello que describe. Por ejemplo, decir que el Everest tiene una altitud de aproximadamente 8000 metros es menos exacto que decir que tiene una altitud de 8848 metros.
- Por error nos referimos a la variación que se produce al generar un dato de la misma manera y sobre la misma cosa de forma repetida. Por ejemplo, si medimos la altitud del Everest utilizando cierta técnica repetidas veces, podríamos obtener medidas como 8841, 8850, 8846, etc., debido a la imperfección de nuestros instrumentos de medida y protocolos de trabajo. Cuanta más variación exista, más error hay y, por lo tanto, mayor incertidumbre sobre el verdadero valor de la altitud del Everest.

Aunque el concepto de error es fundamental para una comprensión conjunta de nuestro tratamiento de la vaguedad, no lo hemos aplicado en el caso de estudio recogido en este artículo, por lo que no lo trataremos con detalle. A continuación, se describen y exploran el resto de los conceptos de forma más detallada.

2.1. La incertidumbre

La incertidumbre viene a menudo dada por la “vaguedad epistémica”, uno de los dos tipos de vaguedad existentes junto con la ontológica (ver más abajo), presentes en los diversos aspectos del mundo (González-Pérez 2018: 175–76). La incertidumbre se refiere a situaciones en las que nuestro conocimiento sobre algo es poco claro o incompleto (González-Pérez 2018: 177). Por ejemplo, podemos decir que “la villa Liédena presenta una importante ampliación en el siglo IV d.C.” Al decir esto, mencionamos toda la centuria porque no sabemos exactamente cuándo tuvo lugar esta ampliación, aunque sospechamos que este fenómeno se produjo en un lapso mucho más breve que un siglo que, debido a la falta de datos, no podemos concretar.

En Arqueología, como ya hemos indicado anteriormente, “*la plupart du temps [los datos son] fragmentaires et difficiles à dater*” (Fusco 2016: 21). Además, el continuo incremento de los datos, principalmente con el desarrollo de las bases de datos y del impulso de la Arqueología preventiva en las últimas tres décadas, la heterogeneidad y diversidad de los datos es todavía más evidente. La cantidad de datos que acumulamos, pero sobre todo la forma en la que cada grupo de investigación, cada administración o cada servicio arqueológico los registra, ha conllevado un incremento de la importancia de tratar la incertidumbre de los datos arqueológicos de una forma adecuada.

Este incremento de los datos, que algunos han llegado a calificar de *big data* (Gattiglia 2015), ha supuesto que los datos arqueológicos en la actualidad tengan unas características diferentes a los de hace unas décadas. Algunos autores han realizado una comparación entre los datos “tradicionales” y los datos denominados “masivos” o *big data*. Esto ha llevado a una gran disparidad entre los datos de la Administración, de la bibliografía, de los informes arqueológicos, y de otras fuentes. Así, por ejemplo, los datos de la Administración, principalmente de los inventarios arqueológicos (que podríamos considerar como “datos masivos”), distan bastante de los datos extraídos de trabajos de investigación y artículos científicos (“datos tradicionales”), en los que, de manera general, la información está más estructurada, detallada y explicada.

La incertidumbre, es decir, la falta de certeza de algo “emerge de representaciones imperfectas del mundo y su análisis implica la toma de conciencia de que pueden existir contradicciones entre los datos y el fenómeno real representado por ellos” (Sánchez Trigueros 2013: 13). La incertidumbre es una propiedad de los datos, ya que se refiere a la relación entre nuestras formas de representar (nuestro conocimiento del mundo, en definitiva) y aquello representado. Pero no sólo eso, pues la incertidumbre no es sólo inherente a los datos, por ejemplo, las formas de datación o las variaciones en una prospección, sino que también “*aux choix que l’on effectue à partir de ces données, et à l’interprétation que l’on en fait*” (Fusco 2016: 43).

De esta manera, un dato es de mejor calidad cuanto menos incierto sea, pues “*la gestion de l’imperfection (l’incertitude est une forme d’imperfection) des données fait partie de la*

gestion de la qualité” (De Runz 2008: 11). A la hora de medir la incertidumbre, hay muchos criterios que han sido tenidos en cuenta e incluso grados dentro ella (Fusco 2016; Gavignet *et al.* 2016).

El análisis del grado de incertidumbre de los datos debería ser una cuestión a tratar en cualquier estudio arqueológico. Son datos inciertos aquellos que no nos dan seguridad, confianza o certeza sobre su capacidad de representar de forma adecuada aquello que pretenden representar. La incertidumbre “está relacionada con nuestro conocimiento sobre las cosas, de modo que varía subjetivamente dependiendo de quién observe o describa algo” (González-Pérez 2018: 178); esto implica que diferentes datos generados por diferentes personas sobre la misma cosa pueden presentar diferentes grados de incertidumbre. Por ejemplo, en nuestro caso, un dato incierto sería una referencia a un asentamiento tardío en el que la cronología viene dada por el tipo de cerámica hallada en superficie durante una prospección y del que no conocemos información sobre el material hallado (ni láminas, ni dibujos, por ejemplo). Otro tipo de dato incierto sería un hallazgo aislado en los años 1920 y del que no contamos con otras referencias posteriores.

2.2. La imprecisión

Por otro lado, está la imprecisión. Algo es impreciso si “cualquier afirmación sobre [ello] sigue siendo verdadera incluso cuando varía la calidad expresada para dicha característica” (González-Pérez 2018: 176). Es decir, algo impreciso no posee un valor concreto, sino que se manifiesta de forma gradual o “borrosa”. La imprecisión surge habitualmente de la vaguedad ontológica o naturaleza gradual de las cosas; por ejemplo, una montaña no posee límites claros y definidos, sino que existe una transición suave entre su cumbre y el valle anejo. No es posible, por tanto, especificar de forma precisa dónde empieza o acaba la montaña, o qué superficie ocupa.

La imprecisión también puede venir dada por las limitaciones de nuestros instrumentos de medida o de los sistemas de trabajo. Un ejemplo habitual es el de la localización de los yacimientos. Diríamos que un yacimiento está localizado con alta precisión, por ejemplo, si contamos para él con sus coordenadas expresadas en UTM con más de dos decimales, mientras que diríamos que posee una precisión

baja si sólo contamos con las coordenadas sin decimales. De este modo, la imprecisión tiene que ver con la falta de detalle en un dato.

Finalmente, la imprecisión puede surgir también de forma intencional, cuando la “inyectamos” a propósito en un dato para mitigar su incertidumbre. Por ejemplo, podríamos decir que “la villa de Arellano se abandona entre finales del IV y principios del V d.C.”. En este caso, como no sabemos con certeza cuándo esta villa fue abandonada, inyectamos imprecisión de forma intencionada para mitigar la incertidumbre de los datos y conseguir que las probabilidades de que nuestra afirmación sea verdadera aumenten (González-Pérez 2018: 178).

2.3. La inexactitud

La exactitud mide la diferencia entre un dato y el valor real que tratamos de representar⁵. Así, en ocasiones se atribuye inexactitud a un atributo para que sea menos incierto, como ya señalamos anteriormente con el ejemplo de Arellano. Aplicado a la tipología, esta inyección de inexactitud se manifiesta, por ejemplo, cuando hablamos de “una estructura agrícola-residencial”. No tenemos clara la tipología de la estructura por lo que la denominamos de manera más genérica (inyectando inexactitud), para aproximarnos más a la realidad. De esta manera, aunque el valor sea más inexacto reducimos la vaguedad epistémica. Al igual que para la precisión, una entidad puede ser muy precisa pero inexacta, o ser a la vez inexacta e imprecisa.

La imprecisión y la inexactitud son aspectos independientes y pueden combinarse de todos los modos posibles. Un dato puede ser preciso y exacto, o impreciso e inexacto, o preciso y a la vez inexacto, o al revés. Por ejemplo, como propone F. Sánchez Trigueros en artículo sobre la calidad de los datos arqueológicos (2013: 11), si consideramos un fondo de cabaña desenterrado durante una excavación arqueológica, al referirnos a la tipología de esta estructu-

⁵ El diccionario Larousse define la exactitud como “Caractère de ce qui est conforme à la vérité, à la réalité” mientras que el D.R.A.E. la define como el grado en el que algo se asemeja a otra entidad que es tomada como modelo. El V.I.M. (Vocabulario Internacional de Metrología. Conceptos fundamentales y generales, y términos asociados) la define como “proximidad entre un valor medido y un valor verdadero de un mensurando” (Centro Español de Metrología 2012, 31).

ra, podemos decir que se trata de una cabaña, con lo cual estaríamos siendo muy exactos (el dato representa fielmente el mundo) y además bastante precisos (al decir “cabaña” aportamos bastante información sobre el mundo). Sin embargo, también podríamos decir que se trata de una construcción; en este caso, estaríamos siendo igual de exactos que antes (la cabaña encontrada es una construcción, al fin y al cabo), pero mucho más imprecisos, porque “construcción” aporta menos información. De la misma forma, también podríamos decir que se trata de un horno para pan, lo cual es muy preciso (aporta mucha información) pero muy poco exacto (no representa nada bien al mundo).

2.4. El error

El error corresponde a la variación que se da cuando tomamos medidas u observamos el mundo de un modo u otro, para generar datos de forma repetida. Es habitual, por ejemplo, que distintas dataciones de C14 para la misma muestra arrojen resultados ligeramente distintos, debido a la imperfección de los instrumentos de medida o de nuestros protocolos de trabajo. El error ha sido definido como el “grado de variabilidad de una medida tomada repetidamente sobre una entidad o del resultado que se deriva de ésta” (Sánchez Trigueros 2013: 10). Aunque tanto el error como la imprecisión a menudo tienen su fuente en las limitaciones de los instrumentos de medida, constituyen fenómenos distintos; por ejemplo, un instrumento que proporcione medidas altamente repetibles pero con pocos decimales produciría muy poco error pero gran imprecisión, mientras que un instrumento que arrojase medidas muy variables pero con muchos decimales estaría produciendo mucho error pero gran precisión.

Aquí debemos llamar la atención sobre una fuente de confusión habitual. El V.I.M. (Vocabulario Internacional de Metrología. Conceptos fundamentales y generales, y términos asociados) define “precisión” como la “proximidad entre las indicaciones o los valores medidos obtenidos en mediciones repetidas de un mismo objeto, o de objetos similares, bajo condiciones especificadas” (Centro Español de Metrología 2012: 31), pero esta definición corresponde a nuestro concepto de error, y no a nuestra noción de precisión. Es importante resaltar que nuestro concepto de precisión es

independiente de la medida (o de la observación) y se basa, sin embargo, en la vaguedad ontológica o en la inyección intencional para mitigar la incertidumbre, como hemos explicado anteriormente.

Un error mayor suele conducir a una mayor incertidumbre. Por ejemplo, si repetidas dataciones de una misma muestra arrojan siempre el mismo resultado, estaremos bastante seguros (baja incertidumbre) de que el resultado es una buena representación del verdadero valor. Pero si estas dataciones varían mucho entre ellas, existirá una alta incertidumbre sobre cuál es el valor verdadero. En el caso de la Arqueología, su aplicación práctica resulta complicada; no podemos conocer el grado de error en las mediciones salvo en el caso de contar con análisis de C14 o similares. En nuestra propuesta, por tanto, no tendremos en cuenta esta variable, al ser esta una cuestión medible en ciencias naturales, pero de difícil evaluación y aplicación en humanidades.

3. Criterios de aplicación

Partiendo de la reflexión teórica anterior, hemos aplicado este marco conceptual al estudio de las dinámicas espaciales y temporales de poblamiento durante la Antigüedad tardía, en el espacio comprendido entre el Ebro medio y la cadena pirenaica. El *corpus* se compone de 350 entidades arqueológicas que, a su vez, se subdividen en hallazgos aislados y sitios arqueológicos. La cronología de los sitios se extiende entre los siglos III d.C. y VII d.C. Los datos recopilados proceden de los archivos de la Administración (informes, estudios de material, inventarios y cartas arqueológicas) y de publicaciones (artículos especializados, monografías, tesis doctorales y actas de congresos).

Aplicamos los criterios anteriormente indicados a tres variables arqueológicas: la localización, la datación y la tipología. Hemos establecido los siguientes parámetros para cada criterio y cada variable.

3.1. La incertidumbre de nuestros datos

Para medir la incertidumbre nos basamos en el origen de los datos, es decir, su fuente. Debemos resaltar que la determinación de la incertidumbre, siendo esta un fenómeno fuertemente subjetivo, es un proceso basado en la confianza y en la percepción del riesgo, por lo

que los criterios que se adopten pueden variar según el contexto y las personas implicadas, sin que esto constituya un defecto del sistema. Dado que trabajamos con fuentes de segunda mano (informes, bibliografía, tesis doctorales, etc.), era necesario valorar la certidumbre/incertidumbre de estas. Para ello, partiendo de la base documental de fuentes empleadas, establecimos una serie de criterios considerando las características del conjunto de datos.

- **Incertidumbre de la localización:** consideramos que las coordenadas geográficas extraídas de los inventarios e informes de la administración presentan un nivel de incertidumbre bajo. Las empresas que los han realizado son casi siempre las mismas y, por tanto, entendemos que en la toma de datos emplean siempre los mismos criterios. Por otra parte, se ha establecido el año 2005 como momento de transición en la incertidumbre al ser necesario tomar una fecha de referencia. En dicho año se publicó la Ley Foral 14/2005, de 22 de noviembre, del Patrimonio Cultural de Navarra, momento a partir del cual se homogeneizó la entrega de memorias arqueológicas (en Aragón la ley se publicó en 1997 y 1999, por lo que tomamos como referencia la de Navarra, de fecha más reciente). Además, se ha constatado que los informes arqueológicos son, en general, más completos a partir de los años 2004-2005 en ambas comunidades autónomas. Por tanto, los criterios quedan establecidos de la siguiente manera:

1. **Incertidumbre baja:** la localización del elemento ha sido extraída del inventario arqueológico del servicio de Arqueología regional de la comunidad autónoma en la que se localiza.
2. **Incertidumbre media:** la localización del elemento ha sido extraída de informes arqueológicos o de la bibliografía posterior al año 2005.
3. **Incertidumbre alta:** la localización ha sido extraída de informes y/o bibliografía anterior al año 2005 sin contar con otra referencia a la misma en otro documento.

- **Incertidumbre de la datación:** consideramos la fuente bibliográfica de la que extraemos el dato cronológico. En este caso

fue algo más sencillo atribuir los grados de incertidumbre dado que, en general, hay un consenso bastante alto entre la comunidad científica en los grados de certidumbre de las dataciones. Cuando contamos con una datación C14 para el abandono de una fase, valoramos que tiene un grado de certidumbre alto. Dado que no siempre contamos con dataciones de este tipo, hemos incluido también en este grado aquellas obtenidas a partir de estudios de materiales por criterios crono-tipológicos realizados en los últimos 10 años⁶. El resto de los grados de certidumbre han quedado establecidos a partir de la fecha y del tipo de estudio de materiales realizados, así como por la presencia o no de láminas y dibujos en los estudios –eso nos da un grado de certeza de la fuente más elevado–.

1. **Incertidumbre baja:** la datación viene dada por dataciones de C14 o criterios crono-tipológicos de materiales en estudios realizados en los últimos 10 años, con láminas de los materiales y/o varios estudios sobre el mismo.
 2. **Incertidumbre media:** la datación viene dada por criterios crono-tipológicos mediante estudios realizados entre 1985 y 2005, y disponemos de láminas con los dibujos de los materiales.
 3. **Incertidumbre alta:** la datación viene dada por la tipología de los materiales y/o los estudios de material son anteriores a 1985, no disponemos de láminas y/o solo hay una referencia bibliográfica sin detalle.
- Incertidumbre de la tipología del sitio

Por último, para la tipología partimos de los criterios establecidos en la tipología.

1. **Incertidumbre baja:** contamos con elementos arqueológicos claros para la atribución tipológica.
2. **Incertidumbre media:** contamos con algunos elementos para la atribución tipológica, pero son dudosos.

⁶ El estudio de materiales para nuestro periodo, sobre todo, en lo que concierne a la cerámica, ha mejorado considerablemente en la última década, pudiendo tener dataciones bastante fiables y precisas (no en todos los casos).

3. Incertidumbre alta: solo contamos con una tipología en función de la tipología actual del lugar.

3.1.2. La imprecisión de nuestros datos

- Imprecisión de la localización
 - 1. Imprecisión baja:** las coordenadas UTM del sitio tienen dos decimales o más.
 - 2. Imprecisión media:** las coordenadas UTM del sitio tienen un decimal.
 - 3. Imprecisión alta:** las coordenadas UTM del sitio no tienen decimales.
- Imprecisión de la datación
 - 1. Imprecisión baja:** cuando podemos dar una datación con un detalle de un máximo de 50 años, sea cual sea la metodología por la que se ha obtenido la misma.
 - 2. Imprecisión media:** cuando podemos dar una datación con un detalle de entre 50 y 100 años, sea cual sea la metodología por la que se ha obtenido la misma.
 - 3. Imprecisión alta:** cuando podemos dar una datación con un detalle de más de 100 años, un siglo, sea cual sea la metodología por la que se ha obtenido la misma.
- Imprecisión de la tipología del sitio

En este caso, la precisión va a venir determinada por el tesoro que hemos establecido para la tipología de sitios que, al estar jerarquizado, nos permite establecer niveles de detalle en la definición del tipo de ocupación de cada sitio.

- 1. Imprecisión baja:** cuando el sitio tiene una categoría de nivel 3 del tesoro de “tipo de ocupación” (A21: *villa* de pequeñas dimensiones; A21: *villa* de grandes dimensiones etc.).
- 2. Imprecisión media:** cuando el sitio tiene una categoría de nivel 2 del tesoro “tipo de ocupación” (A1: sitio rural indeterminado, A2: *villa*, etc.).
- 3. Imprecisión alta:** cuando el sitio tiene una categoría de primer nivel del tesoro “tipo de ocupación” (A: Rural, etc.)

Esta propuesta pretende simplificar la forma de tratar la calidad de los datos arqueológicos.

3.1.3. La inexactitud de nuestros datos

La inexactitud solo la aplicamos a la localización porque, al ser expresada de forma numérica, es fácil evaluar la diferencia entre un dato y la realidad. No podemos saber hasta qué punto nuestros datos cronológicos y de ocupación se acercan a la realidad del momento.

- Inexactitud de la localización
 - 1. Inexactitud baja:** contamos con las coordenadas geográficas exactas de localización del elemento, y esto es verificable empíricamente.
 - 2. Inexactitud media:** contamos con la localización del elemento dentro de una parcela, sin tener las coordenadas exactas. Se pueden calcular las coordenadas a partir del centroide de la parcela como aproximación a la localización exacta.
 - 3. Inexactitud alta:** contamos con la localización del elemento dentro de un municipio, sin mayores referencias a la posible ubicación exacta de este. Como en el caso anterior, es posible calcular las coordenadas aproximadas a partir del centroide del municipio.

4. Aplicación práctica

Como venimos apuntando a lo largo de las páginas anteriores, medir la calidad de los datos a través del tratamiento de la vaguedad es uno de los objetivos principales de nuestro trabajo (Cf. Tobalina Pulido 2020, para una explicación detallada de la metodología completa propuesta). Por ello, a continuación, mediremos cuál es la calidad de nuestros datos en cuanto a su localización, su datación y su tipología. Esto nos va a permitir conocer cuál es la calidad global de nuestro conjunto de datos, pudiendo tener en cuenta esto a la hora de llevar a cabo otros análisis como los de proximidad, donde contar con unos datos de buena calidad en la localidad resulta importante. Así, al considerar la calidad, podremos valorar mejor los resultados de nuestros análisis.

4.1. La localización

Nos centramos, en primer lugar, en la localización de las entidades arqueológicas. En un primer momento, optamos por realizar un análisis de agrupamiento por similitud de los datos, estableciendo automáticamente en *ArcMap* tres grupos. Sin embargo, el programa no permite prefijar muchos parámetros, lo que hace que el resultado sea bastante impreciso. Por ello, optamos por hacer un análisis ponderando las cifras de exactitud, certidumbre y precisión, para no dejar esto al azar estadístico del programa informático. Creemos que la exactitud prima sobre las otras dos variables, por lo que la ponderamos con un 40%, mientras que la fiabilidad lo haría con un 40% y la precisión con un 20% (Fig.2). Estos porcentajes responden al peso que damos a cada variable en nuestra investigación, en función de nuestras prioridades. Aunque es cierto que el propio arqueólogo que genera y procesa los datos participa en la generación de la vaguedad asociada, no toda esta depende de él. Primeramente, la vaguedad ontológica es completamente independiente de las personas, por definición. En segundo lugar, mucha de la vaguedad epistémica viene dada por las características de los instrumentos que manejamos (como el error típico de un receptor de GPS, por ejemplo) y de las técnicas y tecnologías que aplicamos, y no de cada individuo. Es cierto que cada individuo, con sus sesgos y preferencias personales, tiñe los datos que genera de un tono particular, pero no determina completamente la vaguedad en sus datos. De este modo, creemos que se puede hablar de “objetividad”, no en términos absolutos, sino en el sentido de algo compartido y comprensible por varios individuos con experiencias y formación similares.

De esta manera, en nuestro caso, los valores más significativos son la exactitud y la certidumbre. Nos interesa especialmente que los datos sean exactos y certeros, mientras que la precisión es una variable, en nuestro caso, secundaria. Teniendo esto en cuenta, damos a las dos primeras un mayor peso en el cálculo. Aunque la precisión es importante, no nos resulta tan relevante como las otras dos. La fórmula que hemos establecido es la siguiente, donde E se corresponde con la “Exactitud”, C con “Certidumbre” y P con

“Precisión”, siendo C^{lo} el índice de la calidad de los datos de localización.

$$C^{lo} = (E*0,40) + (C*0,40) + (P*0,20)$$

Fig. 2. Fórmula aplicada para la ponderación de la calidad de la localización

De esta manera, para un sitio con los valores siguientes: E=3, C=2, P=2, el valor de C^{lo} será de 2,4 (resultado de la operación: $(3*0,40) + (2*0,40) + (2*0,20)$). De esta manera, los valores más próximos a 3 serán aquellos con una peor calidad, mientras que los más próximos a 1 serán los mejores, siendo los valores iguales a 1 los que presentan una mejor calidad (Fig. 3).



Fig.3. Escala cromométrica de calidad

Una buena parte de las entidades arqueológicas se localiza en la parte baja del gráfico, es decir, contamos con datos de calidad media-baja, pero la mayoría se localizan en una parte media del gráfico, entre 1,2 y 1,9 puntos, lo que les atribuye una calidad media-aceptable (Fig.4). Tan solo una treintena de entidades arqueológicas se localiza en la parte más baja del diagrama, en color verde, siendo los datos de localización con un mejor índice de calidad. Al menos una cuarta parte de los sitios se sitúa en la parte superior del gráfico, es decir, tienen una mala calidad. Si representamos el resultado cartográficamente, las entidades arqueológicas con una mejor calidad se centran en la zona navarra, mientras que las de peor calidad, se localizan entre Navarra y la actual Cinco Villas, principalmente. Esto puede ser debido a que muchos de los hallazgos realizados en esta área de la zona de estudio fueron llevados a cabo por eruditos locales durante el siglo XIX e inicios del XX (Altadill 1923), no habiéndose realizado estudios posteriores y contando, por tanto, con datos bastante parciales y poco precisos en algunos casos.

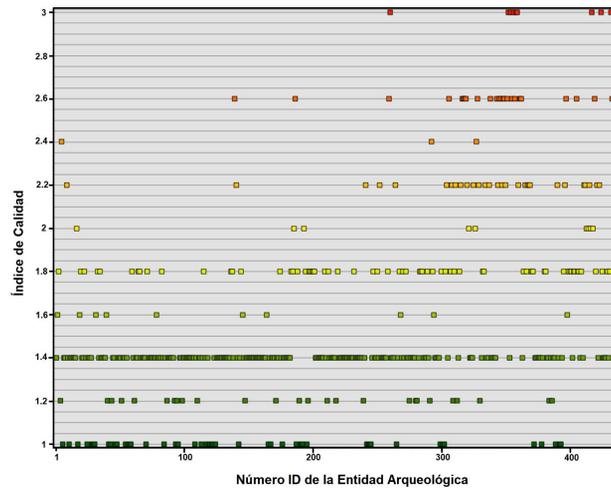


Fig. 4. Diagrama de calidad de los datos de la localización de las entidades arqueológicas (de verde a rojo en función de la calidad)

Una mayoría de entidades arqueológicas presenta una calidad media en su localización, cubriendo todas las áreas de la zona de estudio, lo que nos indica que, de manera general, contamos con datos de una calidad aceptable. Sin embargo, si discernimos entre los hallazgos aislados y los sitios arqueológicos, se observa una diferencia interesante entre la calidad de los datos de localización de ambos. Los hallazgos aislados se sitúan en la parte superior del gráfico, con índices por encima de 1,2 en todos los casos y con una mayoría superando el 1,6, lo que les da a los hallazgos aislados una calidad de localización bastante inferior que a los sitios (Fig. 5). Esto no resulta extraño; nor-

malmente los hallazgos han sido realizados en el marco de trabajos agrícolas y se desconoce su procedencia exacta, siendo además muchas veces hallazgos realizados en el siglo XIX o inicios del siglo XX, cuya certidumbre puede ser, en ocasiones, dudosa, a juzgar por los datos que tenemos. Por el hecho de ser un hallazgo del siglo XIX, no tiene por qué ser un dato de mala calidad, sino que, de manera general, los datos que tenemos de dichas cronologías para nuestra zona son deficientes en cuanto a las indicaciones de la localización de los hallazgos y la exactitud/precisión de los mismos, lo que hace que sean datos de una calidad inferior a otros posteriores.

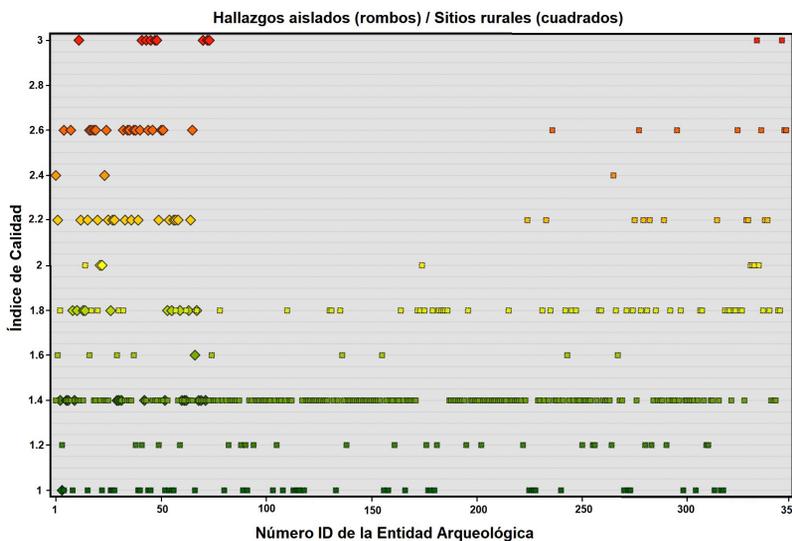


Fig. 5. Diagrama de dispersión para mostrar la calidad de las Entidades Arqueológicas. Rombo: Hallazgos aislados; Cuadrados: sitios (de verde a rojo, de calidad más alta a más baja)

Si profundizamos algo más en el análisis y observamos cuál es el grado de certidumbre, precisión y exactitud de los datos, podremos ver más en detalle las características de nuestro *corpus*. Para ello, centrándonos en los sitios calificados como rurales en nuestra base de datos y que conforman el núcleo de estudio de nuestra tesis, podemos apreciar cómo 326 presentan un grado de certeza alto, 15 medio y 10 bajo. Esto es debido a que los datos extraídos de la Administración son certeros, habiendo extraído una gran parte de la información de informes e inventarios arqueológicos. De ahí que el resultado sea elevado en este caso. Si valoramos la exactitud, la situación difiere de este primer resultado: 246 sitios tienen un alto grado de exactitud, es decir, un 70,08%; 93 un grado medio (26,49%) y 12 bajo (3,4%). Esto nos indica que una mayoría de los sitios ha sido localizada al menos en la par-

cela o tenemos las coordenadas exactas del yacimiento. Finalmente, si valoramos la precisión de los datos de localización, la situación es prácticamente la inversa. Así, 68 sitios tienen una precisión alta (un 19,37%), 42 media (11,96%) y 241 baja (68,66%).

Por tanto, esto nos indica que, aunque la calidad sea aceptable, la precisión con la que fue tomado el dato no es elevada, contando en un alto porcentaje de sitios con coordenadas sin decimales y siendo solo un 19,37% los sitios en los que se han tomado coordenadas UTM con dos o más decimales.

Si analizamos de manera detallada el índice de calidad de localización del conjunto de sitios rurales, un 13,10% tiene una muy buena calidad, es decir, un índice 1, un 60,68% aceptable (índice 1,1-1,4), un 19,94% media (1,6-2) y un 6,26% mala (índice 2,1-3), es decir, 259 una calidad aceptable, 70 media y 22 baja, respectivamente (Fig.6).

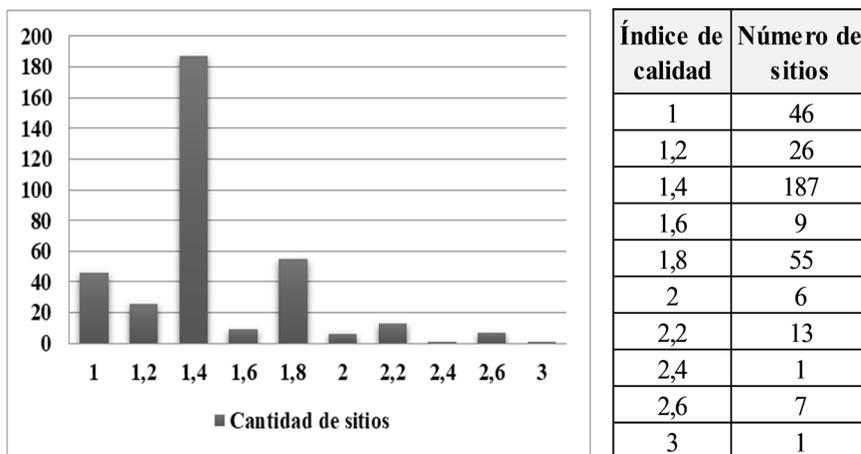


Fig. 6. Cantidad de sitios rurales por índice de calidad, expresado en gráfico de barras a la izquierda y en valores absolutos a la derecha

4.2. La datación

A la hora de calcular el índice de calidad de la datación (C^{da}), hemos aplicado la fórmula siguiente para ponderar los datos de precisión y certidumbre. En el caso de la datación solo tenemos en cuenta dos variables, la precisión y la certidumbre, por lo que a la hora de ponderar los datos hemos optado por hacerlo al 50% (Fig.7).

$$C^{da} = (C*0,50) + (P*0,50)$$

Fig. 7. Fórmula aplicada para ponderar los datos de calidad de la datación

El resultado es un valor del 1 al 3 donde 1 supone que el dato tiene una mayor calidad y 3 menor. Esto nos permite saber cuál es el grado de calidad de la cronología de las fases de nuestros sitios.

Los sitios mejor datados (verde y amarillo) se ubican en la zona central de Navarra y en el área de Labillosa (Huesca), lo que supone un 47,09% de los sitios rurales. Un 27,33% tiene una calidad media, localizándose principalmente en la zona de Navarra media. Finalmente, un 25,58% tiene una datación de mala calidad (Fig. 8); la mayoría de los sitios con una mala datación se centra en la parte norte de la provincia de Zaragoza y

en el sur Huesca, espacios poco trabajados arqueológicamente.

Además, si valoramos solo la exactitud de los datos, un 45,93% tiene una mala exactitud, mientras que algo más de la mitad

(54,07%) buena. Esto no resulta llamativo, ya que un gran número de sitios tiene una cronología muy amplia, muchas veces de más de tres siglos, lo que hace que los datos pierdan calidad.

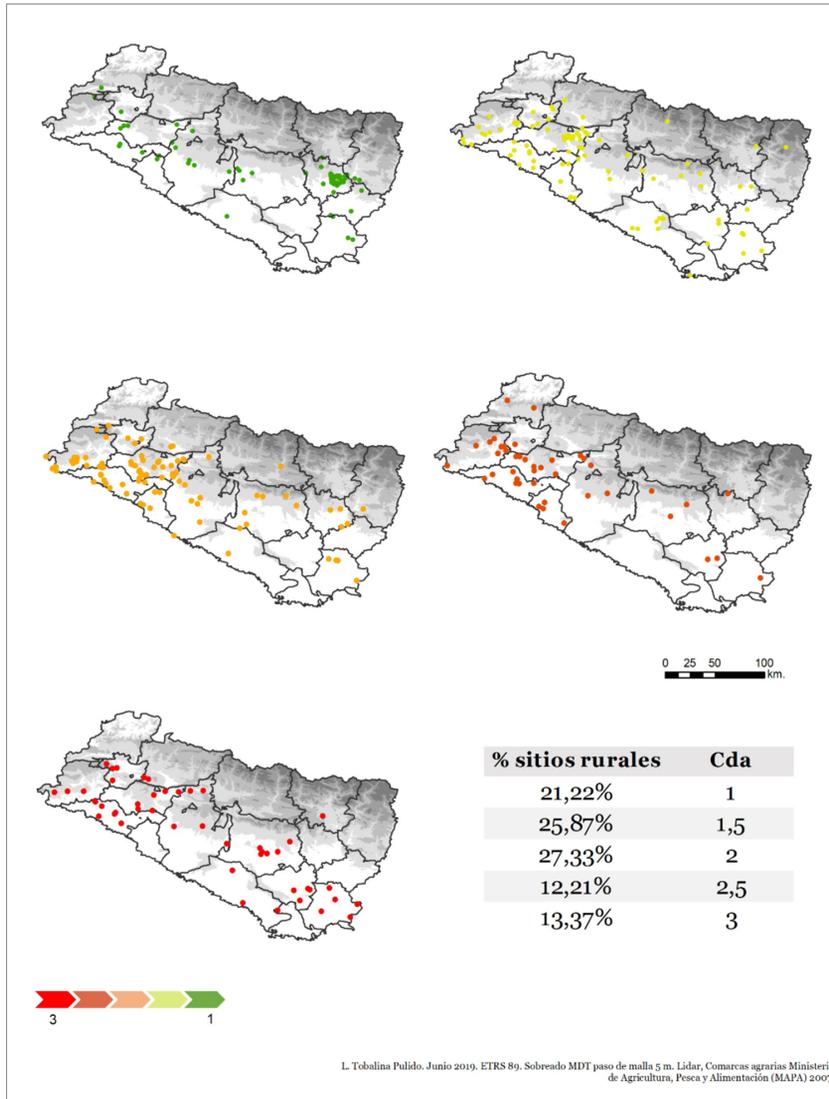


Fig. 8. Calidad de datación de los sitios rurales

4.3. La tipología

En el caso de la tipología (C^{tip}), hemos aplicado la misma fórmula que para la datación, es decir, ponderar las variables al 50% (Fig. 9).

$$C^{tip} = (C*0,50) + (P*0,50)$$

Fig. 9. Fórmula aplicada para ponderar los datos de calidad de la tipología

La calidad de la tipología es muy similar a la de datación, con índices por encima de 1,6 en al menos el 50% de los sitios (Fig. 10).

Finalmente, a modo de ejemplo, nos gustaría presentar una prueba en la que presentamos los sitios con una calidad de tipología y de datación buena (1-1,5). El resultado aplicado a nuestros datos es desalentador. Considerando solo la calidad de la datación, las cifras se mantienen en torno al 50% de sitios (salvo para la séptima centuria) (Fig. 11). Sin embargo, al añadir la calidad de la tipología, el porcentaje se reduce

drásticamente para todo el periodo. Es decir, si solo tuviésemos en cuenta en los análisis los sitios de buena calidad global, estos serían no más de cincuenta. Evidentemente, datos distintos arrojarían resultados distintos. Y, en cualquier

caso, es bastante probable que la calidad de los datos arqueológicos de la Antigüedad Tardía sea en general bastante mala, debido a los escasos estudios de materiales y a la falta de una tipología de materiales consensuada.

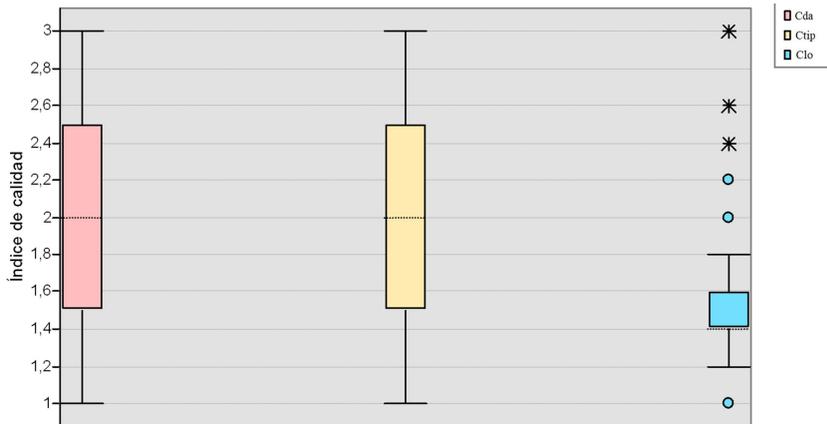


Fig. 10. C^{da}, C^{tip} y C^{lo} de los sitios rurales

	Cda = 1-1,5		Cda y Ctip = 1-1,5	
	Núm. Sitios	% (del total)	Num. Sitios	% (del total)
200-299 d.C.	143	49,83%	48	16,72%
300-399 d.C.	132	46,81%	52	18,44%
400-499 d.C.	69	55,65%	31	25,00%
500-599 d.C.	26	40,63%	12	18,75%
600-699 d.C.	12	29,27%	6	14,63%

Fig. 11. Cuadro que recoge el número de sitios y el porcentaje de estos por siglos considerando la calidad de la tipología y de la datación

Estos criterios y su aplicación en los diferentes análisis nos permiten mostrar al lector cuáles son las carencias de los datos, ofreciendo unos resultados más honestos.

5. Problemas y perspectivas de mejora

Hemos realizado una completa valoración de los datos a través de la gestión de la vaguedad. Esto nos ha permitido estimar, al menos parcialmente, la calidad de los datos, siendo una posible solución a la valoración de los datos arqueológicos. Se trata, por tanto, de un buen punto de partida que necesita afinarse y perfeccionarse pero que nos permite ofrecer al lector unos resultados más honestos. No buscamos reducir la incertidumbre o la imprecisión, pues estas están presentes en los datos, sino medirla, para que nuestros análisis

permitan obtener resultados que tengan en cuenta la calidad de los datos.

Esta estandarización de los datos conlleva una pérdida de información; para poder trabajar a un mismo nivel, tenemos que renunciar necesariamente a ciertas cuestiones como manejar varias escalas de trabajo (en nuestro caso hemos usado la macro) o contar con toda la información detallada de la estratigrafía de cada sitio (hemos tenido que simplificar las fases y la evolución de los enclaves). Por otro lado, hemos evaluado la calidad de los datos arqueológicos, pero no hemos realizado este proceso para los geográficos. Un reto futuro es poder gestionar la vaguedad de ambos en conjunto. Otro aspecto que tenemos que mejorar es el tratamiento de los duplicados en el sistema de registro, ya que trabajar con una cantidad tan elevada de datos suele evitar la repetición de cierta información; una mejor iden-

tificación automatizada de los mismos permitiría reducir estos errores.

Nuestro trabajo ha permitido crear un punto de partida para desarrollar una metodología que permita valorar de manera objetiva una documentación parcial, imperfecta y heterogénea a través de la gestión de la vaguedad. Como perspectivas futuras nos planteamos varias. Uno de los retos es afinar los criterios de los índices de calidad de los datos en todas las variables, pudiendo utilizar el método en el estudio de materiales o directamente al trabajo de campo (prin-

cialmente prospecciones), donde la validación de la propuesta mediante su aplicación a otros proyectos es una de las prioridades. Otro reto es aplicar un método similar a la recogida de datos en prospección, cuestión sobre la que actualmente estamos trabajando y de la que tendremos resultados próximamente.

Por tanto, podemos decir que la creación de un sistema universal y autónomo de tratamiento de la vaguedad para gestionar la calidad es una de las ambiciones investigadoras que nos planteamos en un futuro próximo.

Bibliografía

- Altadill, J. (1923): *De Re Geographico-Histórica. Vías y vestigios romanos en Navarra*. Imprenta de la Diputación de Guipúzcoa, San Sebastián. http://www.navarra.es/appsext/bnd/GN_Ficheros_PDF_Binadi.aspx?Fichero=BGN00FApCad11_1400000000000000410.pdf
- Baena Preysler, J.; Ríos Mendoza, P. (2006): Realidad y Abstracción: Límites de la integración de datos en los SIG. *La aplicación de los SIG en la Arqueología del paisaje*: 15–27.
- Barceló, J.A. (1996): Heuristic Classification and Fuzzy Sets. New Tools for Archaeological Typologies. *Analecta Praehistorica Leidensia*, 28: 155–164. https://proceedings.caaconference.org/files/1995/19_Barcelo_CAA_1995.pdf
- Barceló, M. (2008): *Una historia de la informática*. UOC, Barcelona.
- Bevan, A., J. Conolly, C. Henning, A. Johnston, A. Quercia, L. Spencer y J. Vroom (2013): Measuring chronological uncertainty in intensive survey finds: A case study from Antikythera, Greece. *Archaeometry*, 55: 312–328. <https://doi.org/10.1111/j.1475-4754.2012.00674.x>. <https://onlinelibrary.wiley.com/doi/full/10.1111/j.1475-4754.2012.00674.x>
- Centro Español de Metrología (2012): *VIM Vocabulario Internacional de Metrología Conceptos Fundamentales y Generales, y Términos Asociados*. Centro Español de Metrología, Madrid. <https://www.cem.es/sites/default/files/vim-cem-2012web.pdf>
- Chaillou, A. (2003): Nature, statut et traitements informatisés des données en archéologie : les enjeux des systèmes d'informations archéologiques. Tesis doctoral defendida en la Université Lumière Lyon 2. <https://tel.archives-ouvertes.fr/tel-00137986/file/TomeI.pdf>
- Criado-Boado, F. (1993): Visibilidad e interpretación del registro arqueológico, *Trabajos de Prehistoria*, 50: 39-56. <https://doi.org/10.3989/tp.1993.v50.i0.488>
- De Runz, C. (2008): *Imperfection, temps et espace : Modélisation, Analyse et Visualisation dans un SIG Archéologique*. Tesis defendida en la Université de Reims Champagne-Ardenne.
- . 2013. Analyse de données spatiotemporelles imparfaites dans un SIG Archéologique. *Revue du Comité Français de Cartographie (CFC)*, 215: 51–58. <http://www.lecfc.fr/new/articles/215-article-9.pdf>
- De Runz, C. y Desjardin, É. (2011): Exploration de données temporelles floues: Application en archéologie. *Revue des nouvelles technologies de l'Information (RNTI), Numéro Spécial Fouille de Données Complexes*, 21: 249–266. <https://hal.archives-ouvertes.fr/hal-00669222/document>
- Desachy, B. (1990): Le traitement formalisé des données archéologiques, *Brises*, 15-2: 21-36. https://www.researchgate.net/publication/307173542_Le_traitement_formalise_des_donnees_en_archeologie
- Diarte Blasco, P. y Sebastián López, M. (2011): Gestión de La Información Arqueológica y Realidad Virtual: Visarq 1.0. *Archeologia e Calcolatori*, 22: 261–282. http://www.archcalc.cnr.it/indice/PDF22/AC_22_Diarte_Blasco_Sebastian_Lopez.pdf
- Domínguez Berenjeno, E. (2001): Arqueología y territorio: de la interpretación arqueológica al dato histórico, *SPAL: Revista de prehistoria y Arqueología de la Universidad de Sevilla*, 10: 109-122. <http://dx.doi.org/10.12795/spal.2001.i10.05>
- Favory, F.; Nuninger, L.; Bertonecello, F.; Fovet, É.; Estelle Gauthier, Nicolas Poirier, Murielle George-Leroy, et al. 2008. “ArchaeDyn. Dynamique Spatiale Du Peuplement et Ressources Naturelles : Vers Une Analyse Intégrée Dans Le Long Terme, de La Préhistoire Au Moyen Âge., ACI Espaces et Territoires (Contrat ET28, 2005-2007),” 75. <https://halshs.archives-ouvertes.fr/halshs-00483141>.

- Fusco, G.; Bertonecello, F.; Candau, J.; Emsellem, K.; Huet, T.; Longhi, C.; Poinat, S.; Primon, J.L. y Rinaudo, C. (2014): *Faire science avec l'incertitude : réflexions sur la production des connaissances en Sciences Humaines et Sociales. Incertitude et connaissances en SHS : Production, diffusion, transfert*. Informe de proyecto. <https://hal.archives-ouvertes.fr/halshs-01166287/>
- Fusco, J. (2016): *Analyse des dynamiques spatio-temporelles des systèmes de peuplement dans un contexte d'incertitude : Application à l'archéologie spatiale*. Tesis doctoral defendida en la Université de Nice Sophia Antipolis. <https://tel.archives-ouvertes.fr/tel-01341554>
- Galán Saulnier, C. y Sánchez Meseguer, J.L. (2009): Tratamiento y gestión de la información arqueológica: ¿un problema o un reto? *Espacio, Tiempo y Forma. Serie I, Nueva Época. Prehistoria y Arqueología*, 2: 269–98. <http://e-spacio.uned.es/fez/eserv.php?pid=bibliuned:ETFSerieI-2009-2140&dsID=Documento.pdf>
- García, A.; Orengo, H.; Conesa, F.; Green, A. y Petrie, C. (2018): Remote sensing and historical morphodynamics of alluvial plains. The 1909 indus flood and the city of Dera Ghazi Khan (Province of Punjab, Pakistan), *Geosciences*, 9 (1): 21. <https://doi.org/10.3390/geosciences9010021>. <https://www.mdpi.com/2076-3263/9/1/21>
- Gattiglia, G. (2015): Think big about data: archaeology and the Big Data challenge. *Archäologische Informationen*, 38: 113–124. <https://doi.org/https://doi.org/10.11588/ai.2015.1.26155>. <https://journals.ub.uni-heidelberg.de/index.php/arch-inf/article/download/26155/19871/>
- Gavignet, E.; Leclercq, E.; Cullot, N.; Savonnet, M.; Gavignet, E. y Leclercq, E. (2016). Raisonner en logique modale sur l'incertitude liée aux données : application en archéologie. *Revue Internationale de Géomatique*, 26: 467–490. <https://doi.org/10.3166/ri.2016.00007>. <https://hal.archives-ouvertes.fr/hal-01467036>
- Gonzalez-Perez, C. (2018). *Modelado de información para Arqueología y antropología: Principios de ingeniería de software para patrimonio cultural*. Createspace Independent Publishing Platform.
- López Aguilar, F. (2012): Arqueología y complejidad. Sobre la naturaleza del dato arqueológico, *Temas de Arqueología Mexicana*, 17: 237-264. https://www.academia.edu/5974784/Arqueolog%C3%ADa_y_complejidad_Sobre_la_naturaleza_del_dato_arqueol%C3%B3gico
- Marila, M. (2017): Vagueness and Archaeological Interpretation: A Sensuous Approach to Archaeological Knowledge Formation through Finds Analysis. *Norwegian Archaeological Review*, 50: 66–88. <https://doi.org/10.1080/00293652.2017.1325393>. <https://www.tandfonline.com/doi/full/10.1080/00293652.2017.1325393>
- Mehrer, M. W. y Wescott, K.L. eds. (2005): *GIS and Archaeological Site Location Modeling*. CRC Press, Abingdon-on-Thames. <https://doi.org/10.1201/9780203563359>.
- Niccolucci, F. (2002): Estimating Subjectivity of Typologists and Typological Classification with Fuzzy Logic. *Archeologia e Calcolatori*, 13: 217-232. <http://www.archcalc.cnr.it/indice/PDF13/16Niccolucci.pdf>
- Parcero Oubiña, C. y González Pérez, C. (2011): El Dato Geográfico En La Gestión Del Patrimonio Arqueológico. *Revista PH*, 77: 113. <https://doi.org/10.33349/2011.77.3114>. <http://www.iaph.es/revistaph/index.php/revistaph/article/view/3114/3114>
- Ruiz Zapatero, G. y Fernández Martínez, V.M. (1993): Prospección de Superficie, Técnicas de Muestreo y Recogida de Información. *Inventarios y Cartas Arqueológicas*. Junta de Castilla y León, Valladolid: 87-98.
- Sánchez Trigueros, F. (2013): Reflexiones sobre la calidad y la incertidumbre de los datos arqueológicos. *Ligustinus*, 1: 9–19. <http://arqarqt.revistas.csic.es/index.php/arqarqt/article/view/154/166>
- Shaw, I. y Jameson, R. eds. (1992): *A dictionary of Archaeology*. Blackwell Publishing, Oxford.
- Skeels, M. Bongshin Lee, G. S. y Robertson, G. (2008): Revealing Uncertainty for Information Visualization. *Proceedings of the Working Conference on Advanced Visual Interfaces*. ACM Press, New York: 376. <https://doi.org/10.1145/1385569.1385637>. <https://dl.acm.org/doi/10.1145/1385569.1385637>
- Sørensen, T.F. (2016): In Praise of Vagueness: Uncertainty, Ambiguity and Archaeological Methodology. *Journal of Archaeological Method and Theory*, 23: 741–763. <https://doi.org/10.1007/s10816-015-9257-8>. <https://link.springer.com/article/10.1007/s10816-015-9257-8>
- Tobalina-Pulido, L. (2020): Methodological Proposal for the Study of Temporal and Spatial Dynamics during the Late Period between the Middle Ebro and the Pyrenees. *Information*, 11 (7): 366. <https://doi.org/10.3390/info11070366>. <https://www.mdpi.com/2078-2489/11/7/366>
- Trément, F. (2013): *Les Arvernes et Leurs Voisins Du Massif Central à l'époque Romaine, Tome 1*. Clermont-Ferrand: Société des Amis des Universités de Clermont-Ferrand.