

INFORME DE PROYECCIONES FUTURAS DEL INDICADOR ODS 6.4.2 – NIVEL DE ESTRÉS HÍDRICO

Resumen ejecutivo

El informe evalúa métodos para predecir el indicador de estrés hídrico (ODS 6.4.2), tomando con caso concreto los valores anuales conocidos de AQUASTAT. Aunque se trata de estimaciones concretas basadas en técnicas de extrapolación, es posible obtener datos de interés para construir una propuesta metodológica sobre una futura plataforma de prospectiva. Se analizan las técnicas matemáticas más adecuadas, obteniendo las siguientes conclusiones:

- 1) Con enfoque univariante, un spline polinómico de flexibilidad media ofrece la mejor precisión ($\text{MAPE} \approx 4\%$), superando a modelos clásicos de serie temporal (Holt-Winters, tendencia lineal, ARIMA) que rondan un 6 % de error.
- 2) El ajuste polinómico global resulta inestable; los cambios irregulares de la serie exigen modelos segmentados.
- 3) Al analizar trece variables adicionales de AQUASTAT se observa fuerte multicolinealidad: tres factores—población rural, peso de la industria en el PIB e índice nacional de precipitaciones—concentran la mayor información explicativa, haciendo los demás índices irrelevantes para el tratamiento matemático (aunque no lo sean para la posible explicabilidad de los procesos relacionados con la gestión del agua).
- 4) Estos resultados avalan la posibilidad de crear una plataforma de prospectiva a partir de los datos de AQUASTAT sobre recursos hídricos, que podría plantearse como un recurso interactivo en el que el usuario pudiera establecer sus propios entornos de interés para la estimación de diferentes factores de importancia, como es el estrés hídrico.

Conclusiones prácticas

1. Para proyecciones nacionales anuales, basta un spline univariado bien parametrizado.
2. Desde el punto de vista matemático, los indicadores complementarios añaden poco valor sin tratar la colinealidad; los tres citados son los más útiles.
3. Se recomienda recopilar (o usar en el caso de que se disponga de ellos) datos más granulares (mensuales o por cuenca) para mejorar la capacidad predictiva y orientar políticas de gestión hídrica.

ÍNDICE

Resumen ejecutivo	2
1. Introducción	4
2. Estudio del estrés hídrico con series temporales.....	6
3. Estudio del estrés hídrico con Splines	8
4. Estudio del estrés hídrico con ajuste de polinomios	9
5. Estudio del estrés hídrico en combinación con otras variables	10
6. Conclusiones	12
Segunda parte: prospectiva conceptual sobre el estrés hídrico.....	13
Matriz de relaciones entre términos de U	15
ANEXO: INFORMACIÓN COMPLEMENTARIA SOBRE LAS PROYECCIONES SEMÁNTICAS.	17

1. Introducción

En este informe se plantean diferentes aproximaciones para la proyección del indicador ODS 6.4.2, que mide el nivel de estrés hídrico, definido como la proporción entre la extracción total de agua dulce y los recursos renovables de agua dulce disponibles en un país, teniendo en cuenta las necesidades de caudales ambientales. En concreto, se ha tomado como caso de estudio España con los datos anuales disponibles en la web de Aquastat.

La serie temporal de la variable *estrés hídrico* puede verse en la Figura 1. Presenta un pico en el año 1986, seguido de un comportamiento relativamente estable con un nivel de estrés hídrico del 50% entre 1990 y 2010, para volver a disminuir hasta 39.83% en 2019, siendo los últimos tres valores disponibles 43.25% (2020, 2021 y 2022).



Figura 1: Representación temporal de la variable Estrés Hídrico en España.

En cuanto a las estrategias planteadas, se dividen en 4 grupos:

1. Estudio mediante series temporales.
2. Estudio mediante Splines.
3. Estudio mediante aproximación de polinomios.
4. Estudio matemático de correlación y dependencia de variables de AQUASTAT para la optimización del modelo.

El criterio para utilizar exactamente estas técnicas, es que cada una de ellas se incluyó porque aporta un ángulo complementario para describir la evolución del indicador de estrés hídrico y, al mismo tiempo, permite contrastar sus limitaciones. El análisis de series temporales (Holt-Winters, tendencia lineal, ARIMA) se eligió como primera aproximación porque parte únicamente del historial del propio indicador y ofrece pronósticos rápidos con supuestos estadísticos bien conocidos. Los splines se incorporaron para captar comportamientos no lineales y perturbaciones locales que los modelos clásicos suelen suavizar en exceso, manteniendo al mismo tiempo una cierta regularidad, además de aportar visualizaciones de fácil interpretación. La aproximación polinómica se probó como línea base matemática sencilla: si un polinomio global describiera bien la serie, su uso

reduciría drásticamente la complejidad computacional, aunque finalmente, como es habitual, se constató su inestabilidad en presencia de cambios bruscos. Por último, el estudio de correlación multivariante permite explorar cuánto valor añadido aportan, para la prospectiva, otras variables conocidas sobre la gestión del agua. Conjuntamente, estos cuatro métodos permiten evaluar la robustez de las proyecciones bajo distintos supuestos y escalas de complejidad.

A continuación, se presentará cada una de ellas detallando sus características y su viabilidad para la predicción del estrés hídrico. Los datos para este indicador deben recopilarse anualmente. Sin embargo, aunque periodos menores pueden enriquecer las extrapolaciones, con los datos que se conocen hasta la actualidad un período de referencia de hasta tres años puede considerarse aceptable.

2. Estudio del estrés hídrico con series temporales

En este apartado se utilizan tres técnicas estadísticas para el pronóstico de datos de series temporales: el método Holt-Winters, un modelo lineal y un modelo ARIMA. Se utilizan los datos hasta 2010 para entrenar los modelos, y se utilizan los últimos años para comprobar la bondad de la predicción. Estas técnicas pueden ser útiles al tomar la información de la estructura de la serie temporal y a partir de ella hacer predicciones del valor del indicador.

En primer lugar, el método Holt-Winters utiliza el suavizado exponencial, por medio de ecuaciones que actualizan el nivel, la tendencia y los componentes estacionales (si los hay) en cada paso de tiempo. Dado que no se observa ningún patrón estacional, se plantea un modelo Holt-Winters aditivo, en el que se tienen en cuenta únicamente el nivel y la tendencia.

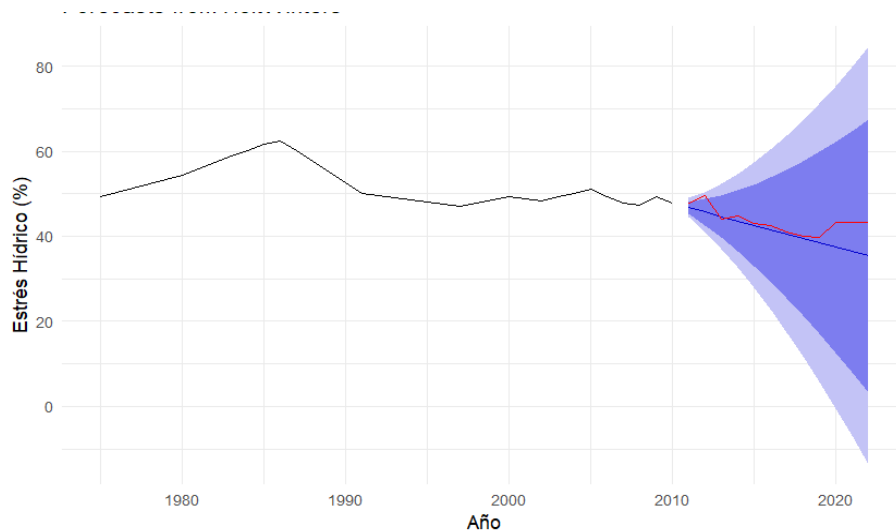


Figura 2: Predicción de la variable Estrés Hídrico en España con Holt-Winters.

Otra opción es ajustar un modelo lineal a la serie temporal, utilizando el mismo enfoque que con el método anterior.

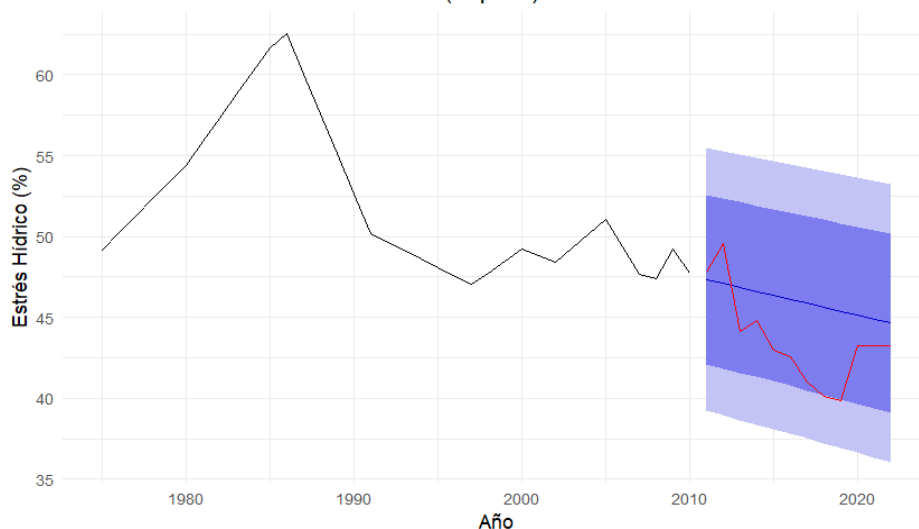


Figura 3: Predicción de la variable Estrés Hídrico en España con un modelo lineal.

Por último, también puede plantearse un modelo ARIMA sin componentes estacionales.

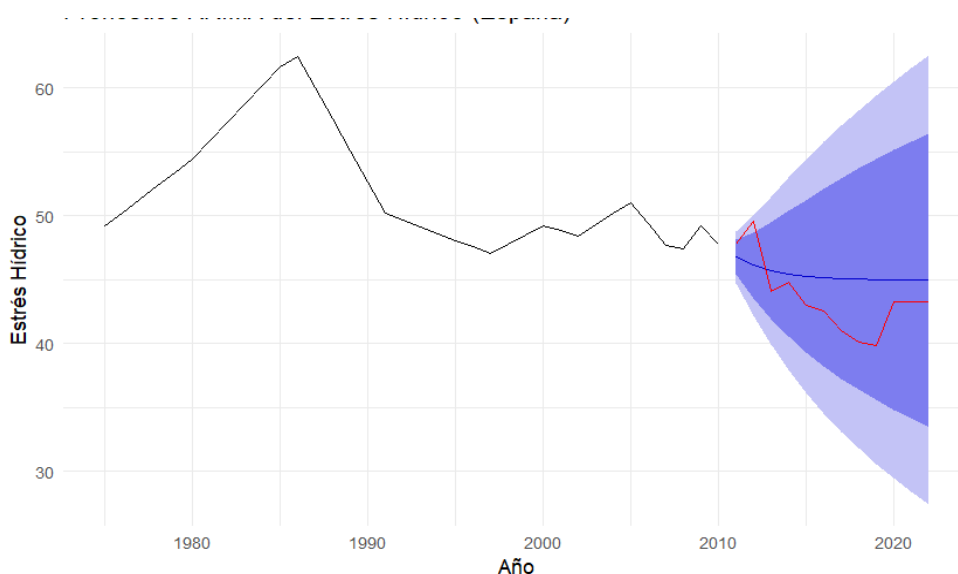


Figura 4: Predicción de la variable Estrés Hídrico en España con un modelo ARIMA.

Los modelos obtienen las siguientes medidas de error sobre el conjunto de validación. Aunque no pueden considerarse estimaciones precisas por los valores de los errores relativos, proporcionan una guía que podría ser de interés para una posible plataforma de prospectiva.

	Holt-Winters	Lineal	ARIMA
MAPE	5.89 %	6.95 %	6.05 %
MAE	2.585	2.942	2.573
RMSE	3.694	3.348	2.944

3. Estudio del estrés hídrico con Splines

Por otro lado, se ha planteado una aproximación de la serie temporal mediante Splines, ajustando diferentes partes de la serie temporal a polinomios con diferentes grados de libertad. Esto significa que a mayor grado de libertad mayor es la obligación de ajustarse a la variable, lo que hace que sea beneficioso buscar un equilibrio entre un buen ajuste y una buena generalización. Este enfoque permite modelar la variable a tramos, de forma que la predicción parte principalmente del último tramo de la serie temporal, pero tomando también información de los tramos anteriores.

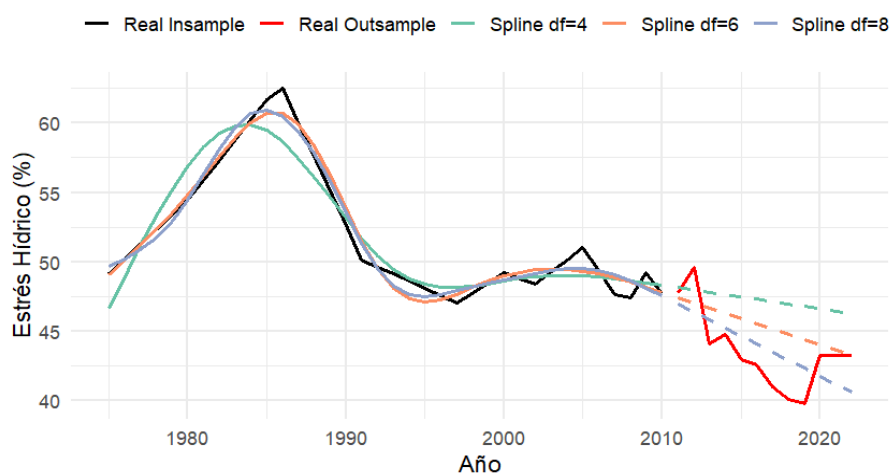


Figura 5: Ajuste de la variable Estrés Hídrico en España con Splines.

En cuanto al error asociado a cada uno de los ajustes, tomando las mismas medidas de error que en el apartado anterior:

Grados de libertad	4	6	8
MAPE	9.368 %	3.954%	4.403%
MAE	5.429	2.298	2.800
RMSE	4.526	1.951	2.104

4. Estudio del estrés hídrico con ajuste de polinomios

Como última aproximación basada en la distribución de una única variable, se plantea un ajuste de polinomios de distinto grado. Sin embargo, como se observa en la Figura 6, esta aproximación no obtiene muy buenos resultados al tratarse de una serie temporal con patrones muy irregulares.

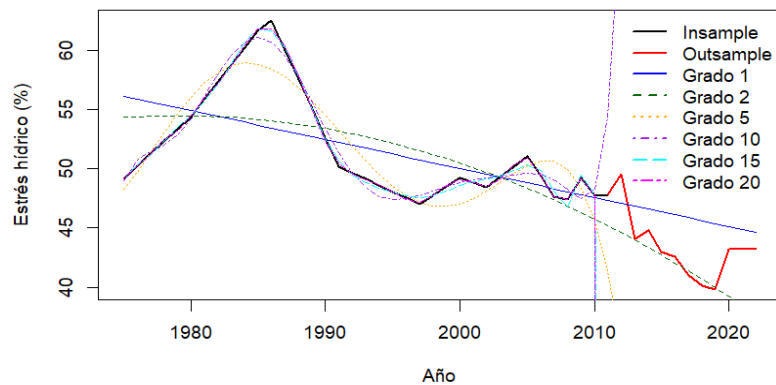


Figura 6: Ajuste de la variable Estrés Hídrico en España con polinomios.

5. Estudio del estrés hídrico en combinación con otras variables

Por último, se ha planteado el estudio de una selección de variables de Aquastat para comprobar su utilidad como variables predictoras del Estrés Hídrico. En concreto, tomando en cuenta características de las variables como la completitud y no formar parte del cálculo del Estrés Hídrico, se han seleccionado las siguientes variables:

1. Year
2. Agriculture, value added (%GDP)
3. GDP per capita
4. Industry, value added to GDP
5. Services, value added to GDP
6. % of total country area cultivated
7. Cultivated area (arable land + permanent crops)
8. Population density
9. Rural population
10. Urban population
11. Total harvested irrigated crop area (full control irrigation)
12. Total dam capacity
13. National Rainfall Index (NRI)
14. SDG 6.4.2. Water Stress

La correlación de cada una de las combinaciones de las variables aparece representada en la Figura 7.

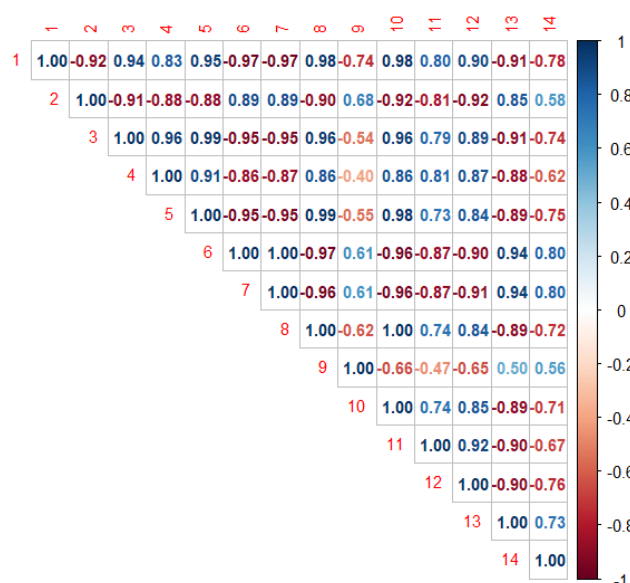


Figura 7: Matriz de correlaciones de las variables seleccionadas.

Puesto que muchas de las variables cuentan con correlaciones muy fuertes, se ha planteado un análisis de componentes principales para ponderar la importancia de cada una de las variables en el conjunto, representado en la Figura 8. De este análisis se

concluye que las variables con mayor importancia para recoger la varianza de los datos son: “Rural population”, “Industry, value added to GDP” y “National Rainfall Index (NRI)”.

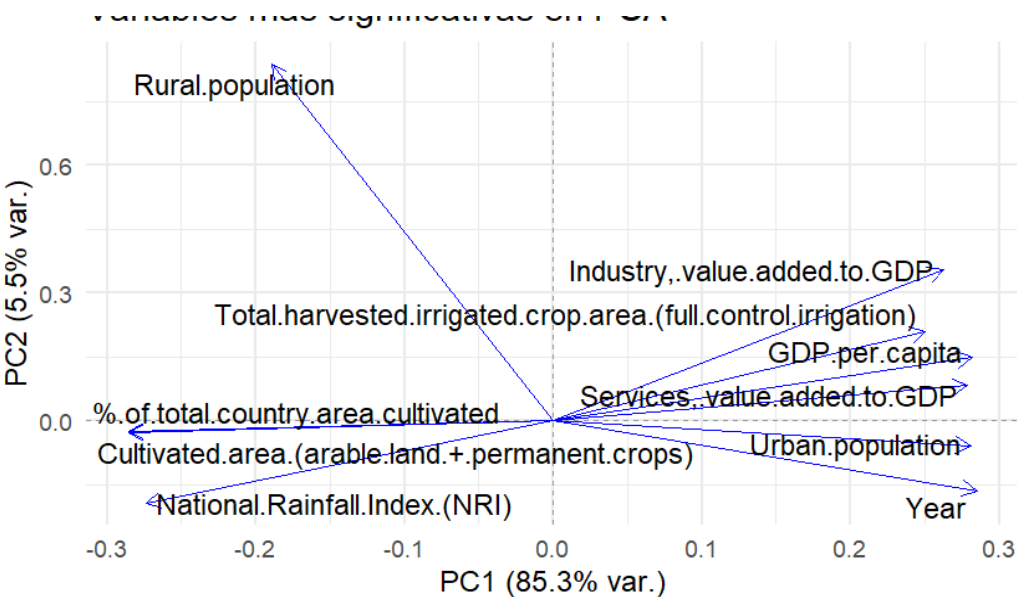


Figura 8: Análisis de Componentes Principales para la selección de variables.

6. Conclusiones

Este informe ha recogido diferentes enfoques para la predicción del Estrés Hídrico, tomando los últimos datos como conjunto de validación para poder obtener una medida para comparar entre los modelos. En primer lugar, la aplicación de técnicas para series temporales obtiene resultados razonables en los tres modelos planteados, aunque peores que un modelo de Splines con una buena elección del grado de libertad (en este caso seis). Por otro lado, enfoques como el ajuste por polinomios de diferentes grados no obtiene buenos resultados por la irregularidad de la serie temporal. Por último, del estudio Estrés Hídrico junto a otras variables el resultado más destacable es la gran correlación entre las variables, obteniendo por Análisis de Componentes Principales que aquellas de más utilidad para recoger la variabilidad de los datos son “Rural population”, “Industry, value added to GDP” y “National Rainfall Index (NRI)”.

Cabe destacar que estas técnicas pueden proporcionar mejores resultados con datos más desagregados (como podrían ser datos a nivel de cuenca, o datos mensuales en lugar de anuales), o con series temporales más largas que permitan detectar más patrones y tendencias en los datos.

Los datos disponibles harían posible la construcción de una plataforma interactiva en la que los usuarios pudieran calcular sus propias proyecciones, proporcionando un control del error y otros parámetros de fiabilidad de las extrapolaciones.

Segunda parte: prospectiva conceptual sobre el estrés hídrico

En esta sección presentamos un ejemplo de aplicación de nuestra metodología de prospectiva, sobre problemas que afectan al estrés hídrico. Nuestro método determina los factores que pueden afectar a este asunto, que se codifican en forma de palabras o expresiones cortas para definir lo que se denomina un universo U, y que depende del punto de vista del análisis que se quiera aplicar. Sobre él se calculan lo que se denominan proyecciones semánticas, que cuantifican de formas diferentes (hay muchas proyecciones semánticas) hasta qué punto el estrés hídrico comparte contexto semántico con los factores que se han considerado relevantes al plantear el estudio.

Las proyecciones semánticas son números reales entre 0 y 1, y pueden interpretarse como la probabilidad de que un término del universo U comparta un cierto entorno de conocimiento (bases de datos científicos, colecciones de artículos especializados, informes de instituciones relevantes, o simplemente, los documentos indexados por Google u otro buscador), con la (o las) expresión(es) lingüística(s) que sintetiza(n) el concepto a estudiar, en este caso el estrés hídrico.

Los estudios de prospectiva basados en proyecciones semánticas buscan cuantificar cómo inciden distintos factores —previamente identificados como pertinentes— sobre un tema específico; en nuestro caso, el estrés hídrico en la agricultura. El método asigna un peso a cada factor, de modo que revela cuáles ejercen mayor influencia y en qué sentido orientan la discusión sobre el tema específico. Con la premisa de que estos factores anticipan las direcciones de la investigación, las políticas y las soluciones tecnológicas, la proyección semántica permite priorizar los más determinantes y prever hacia dónde se encaminarán las iniciativas futuras relacionadas con el tema, que es en este ejemplo el estrés hídrico.

Dado que se trata de un ejemplo que solo pretende mostrar la metodología, omitimos los detalles de la definición de universo:

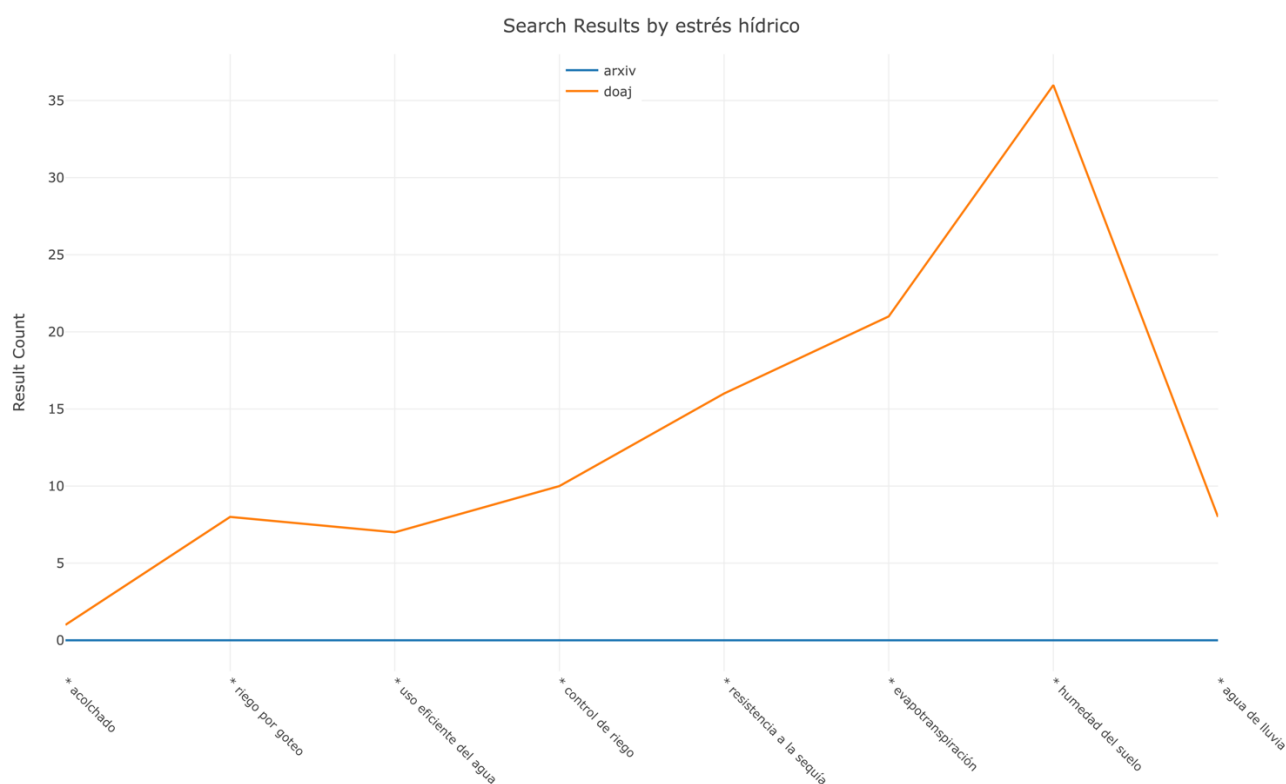
U = {riego por goteo, uso eficiente del agua, humedad del suelo, agua de lluvia, control de riego, acolchado, resistencia a la sequía, evapotranspiración}.

Como término de referencia usaremos simplemente “**estrés hídrico**”.

Utilizaremos solamente la proyección semántica facilitada por la base de datos DOAJ. Se trata de un directorio que reúne revistas académicas de acceso abierto que cumplen ciertos criterios de calidad. Registra metadatos de las revistas (título, ISSN, política de acceso, revisión por pares, etc.) y, en muchos casos, los metadatos de cada artículo. Los resultados de la búsqueda pueden verse en la siguiente tabla.

Búsqueda	doaj
riego por goteo	180
evapotranspiración	375
estrés hídrico AND acolchado	1
estrés hídrico AND riego por goteo	8
estrés hídrico AND uso eficiente del agua	7
estrés hídrico	362
agua de lluvia	431
acolchado	37
resistencia a la sequía	61
uso eficiente del agua	230
estrés hídrico AND control de riego	10
estrés hídrico AND resistencia a la sequía	16
estrés hídrico AND evapotranspiración	21
humedad del suelo	954
control de riego	203
estrés hídrico AND humedad del suelo	36
estrés hídrico AND agua de lluvia	8

La siguiente gráfica representa estos datos conjuntamente.



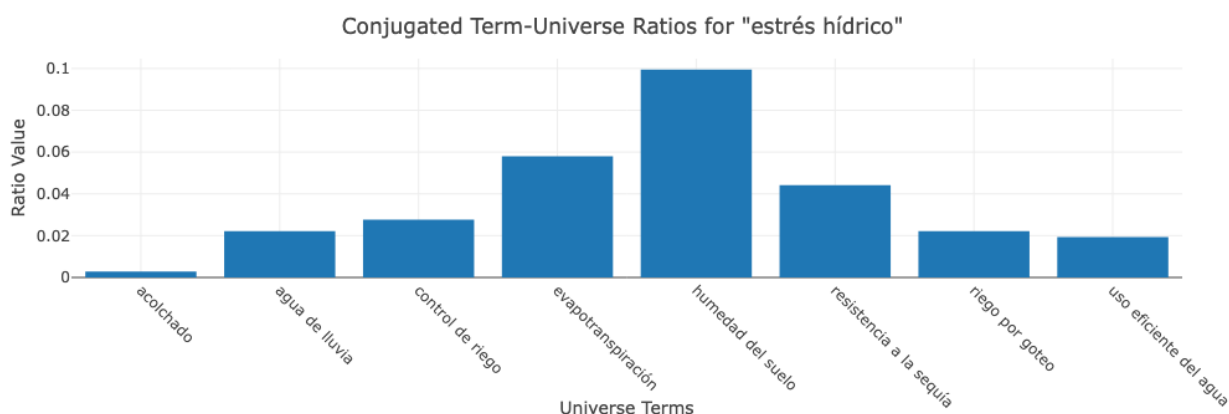
La búsqueda de relaciones dos a dos de los términos de U da las siguientes coincidencias: en cada posición se escribe el número de documentos en los que aparecen los dos términos implicados. La matriz muestra que ciertos términos comparten contexto

semántico en artículos de forma significativa, si comparamos con los datos aportados en la tabla anterior.

Matriz de relaciones entre términos de U

Término	riego por goteo	uso eficiente del agua	humedad del suelo	agua de lluvia	control de riego	acolchado	resistencia a la sequía	evapotranspiración
riego por goteo	0	13	16	3	15	3	0	19
uso eficiente del agua	13	0	11	11	6	1	0	21
humedad del suelo	16	11	0	32	21	1	2	64
agua de lluvia	3	11	32	0	3	1	0	20
control de riego	15	6	21	3	0	0	0	9
acolchado	3	1	1	1	0	0	0	2
resistencia a la sequía	0	0	2	0	0	0	0	0
evapotranspiración	19	21	64	20	9	2	0	0

Las proyecciones semánticas vienen representadas en el siguiente histograma. Esta es la gráfica más importante para la interpretación de los resultados. La proyección semántica da la proporción de apariciones compartidas entre el término y cada elemento del universo, dividido por las apariciones de ese elemento del universo.



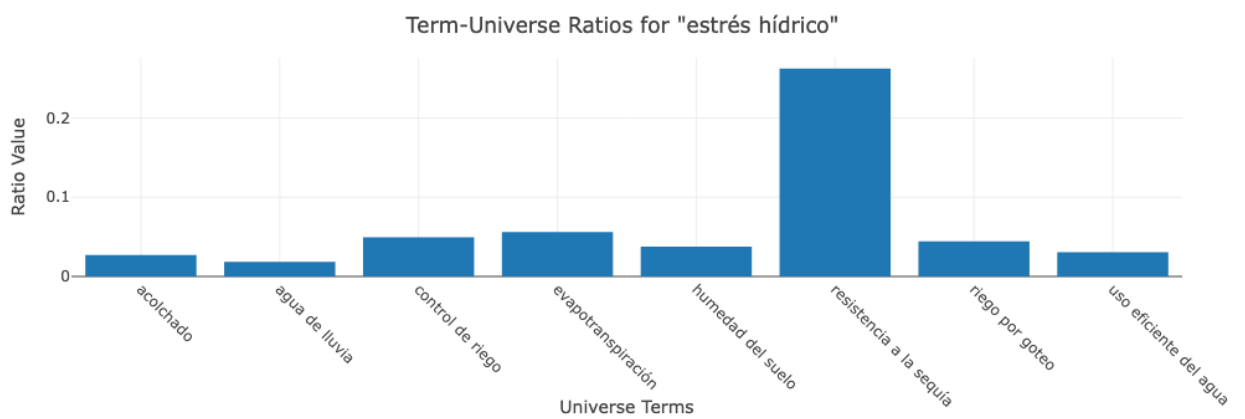
El resultado del análisis es el siguiente:

- 1) Tomando como referencia los artículos indexados en DOAJ, tenemos que la cuantificación de las relaciones no es demasiado alta: todos los valores dan menos de un 10% de coincidencia (índice menor de 0.1). Sin embargo, el resultado es significativo, que normalmente se consideran valores no irrelevantes los que puntúen con un valor mayor del 2%.

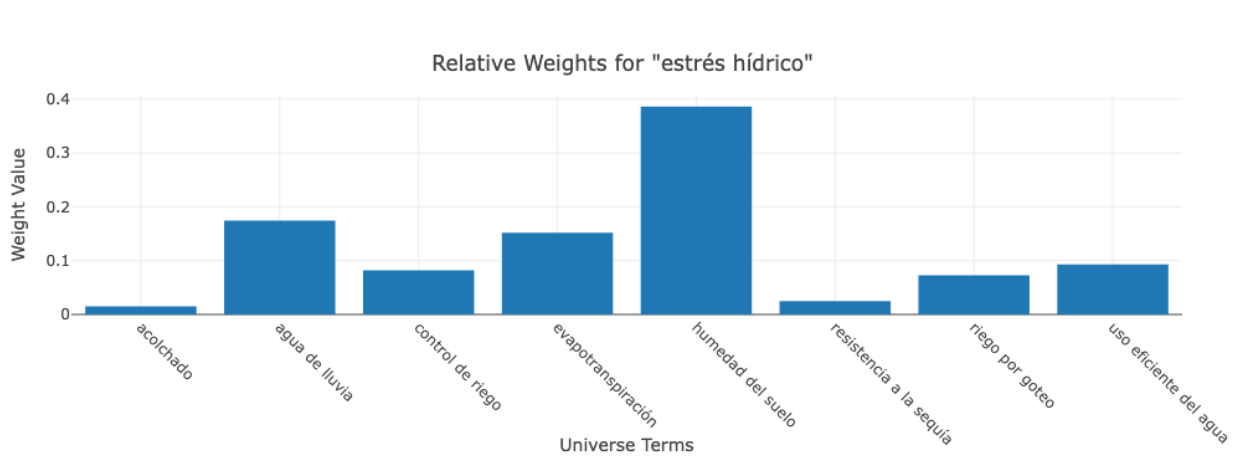
- 2) El **factor más relevante asociado al estrés hídrico** resulta ser la **humedad del suelo**, que resulta el índice más importante en relación al estrés hídrico. En una investigación más profunda, se contrastaría esta información con la obtenida a partir de otras proyecciones semánticas, o por otros medios.
- 3) Significativamente aparecen también la **evapotranspiración y a resistencia a la sequía**, con relaciones próximas al 5% con el estrés hídrico.
- 4) Los otros factores no son tampoco despreciables, pero puntúan en torno al 2%, a excepción de acolchado, que debe ser eliminado de la lista de factores posibles. Aunque es posible que la falta de representatividad se deba a una mala codificación del concepto.

ANEXO: INFORMACIÓN COMPLEMENTARIA SOBRE LAS PROYECCIONES SEMÁNTICAS.

El siguiente histograma aporta información visual sobre la proporción de apariciones compartidas entre el término y el elemento del universo, dividido por las apariciones del elemento del universo.

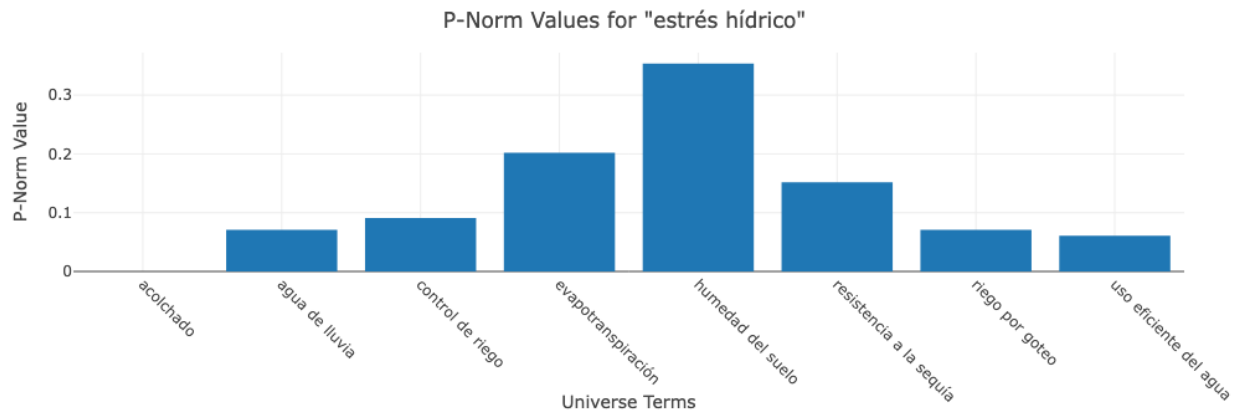


Los siguientes pesos dan información del peso relativo de los elementos de U, en cuanto a su peso según la aparición en documentos científicos indexados en DOAJ.



Por último, la siguiente gráfica indica la relación de significado del término estrés hídrico con cada uno de los elementos de U, usando un word embedding. Valores próximos a 1 indicarían que la coincidencia entre términos observada puede deberse a relaciones de

sinonimia, en lugar de ser asociaciones significativas entre conceptos diferentes. Así, un alto valor de este índice juega en contra de la proyección semántica correspondiente.



Agradecimientos

Nos gustaría agradecer la financiación de la Generalitat Valenciana (España) a través de la subvención PROMETEO 2024 CIPROM/2023/32.