



Cuestionario de la segunda parte del ejercicio

Especialidad: A6 D3. Modelado y análisis de datos.


Por favor, lea detenidamente antes de comenzar:

- **NO** abra el **CUESTIONARIO** ni empiece el examen hasta que se le indique.
- Para realizar este primer ejercicio se hace entrega de dos documentos:
 1. Cuadernillo con los **casos prácticos prácticos**, sobre las materias del programa de esta convocatoria.
 2. **Hoja de respuestas** donde se consignará la respuesta correcta a cada pregunta.
- Al finalizar la prueba se hará entrega de la hoja de respuestas.
- Sólo se calificará las respuestas desarrolladas en la **HOJA DE RESPUESTAS**
- Una vez abierto el cuestionario, compruebe que consta de todas las páginas y preguntas y que sea legible. En caso contrario solicite uno nuevo al personal del aula.
- Verifique que el número de la solapa donde se recogen sus **datos personales coincide con el número de la hoja** de examen donde se consignan las respuestas.
- El examen se realizará con bolígrafo azul o negro. Si no dispone de uno, solicítelo al Tribunal.
- El cuestionario consta de **2 casos** propuestos
- La persona candidata deberá **ELEGIR UNO de esos dos escenarios, haciéndolo constar en la hoja de respuesta** y, basándose en la afirmación aportada por el tribunal, construir justificadamente un caso específico y plantear las formas de abordar la situación, proponiendo vías de soluciones o mejoras e intervenciones a llevar a cabo, todo debidamente argumentado
- El **enunciado** del caso se entregará **en INGLÉS**. La **contestación** al mismo se desarrollará **en CASTELLANO**.
- Numera las hojas de respuesta en orden de lectura e indique los datos personales solicitadas en la misma.
- Se podrán pedir hojas en blanco para utilizar como borrador, pero estas **NO** serán calificadas.
- **NO Separe** ninguna de las copias de la **HOJA DE RESPUESTAS**. Una vez finalizado, el personal del aula le indicará los pasos a seguir.
- **Dispone de 120 minutos**, máximo, para realizar este ejercicio.



Proceso selectivo por el sistema de acceso libre para ingreso en la Escala de Tecnólogos de los Organismos Públicos de Investigación, convocado por resolución de 22 de diciembre de 2025 (BOE N°314 30 de diciembre) – OEP 2023-2024-2025
Primer Ejercicio

Fecha:
10/05/2026
Página: 2 de 4

	Proceso selectivo por el sistema de acceso libre para ingreso en la Escala de Tecnólogos de los Organismos Públicos de Investigación, convocado por resolución de 22 de diciembre de 2025 (BOE N°314 30 de diciembre) – OEP 2023-2024-2025 Primer Ejercicio	Fecha: 10/05/2026 Página: 3 de 4
---	--	--

Estudio de CASO NUMERO 1

A research institute has a remote sensing unit which has accumulated over 10 years of satellite imagery over a network of research stations across the Iberian Peninsula, which includes products of weekly vegetation greenness indices (spatial resolution: 250m), Land Surface Temperature (spatial resolution: 1km), soil moisture content (spatial resolution: 1 km) and monthly burned area (500m), along with hourly meteorological in-situ sensor data such as air temperature, precipitation, relative humidity and solar radiation. The data are stored in heterogeneous formats (NetCDF, CSV, GeoTIFF) and with different time steps and spatial resolutions, along with being stored across different local servers with no unified access system. The research group wants to develop a predictive model to estimate wildfire risk at regional scale (1 km resolution) with a 72-hour forecast horizon, to be integrated into an operational early warning system for the entire Iberian Peninsula. The system must be capable of generating updated risk maps within 2 hours of receiving new meteorological observations. However, the project faces three critical problems:

1. The dataset contains significant gaps (up to 15% missing values in some sensors), inconsistent temporal resolution across sources, and no documented metadata standard. The satellite optical products (vegetation index, land surface temperature) are additionally affected by cloud cover, resulting in irregular temporal gaps that vary by season and region.
2. The research group has no established protocol for model validation or for quantifying prediction uncertainty.
3. The resulting datasets and model outputs are intended to be shared with other European research institutions under the Horizon Europe Open Data mandate, but no data management plan exists.

Based on this scenario, develop a structured technical proposal addressing the following aspects:

- A. The **data preprocessing** and integration strategy, including treatment of missing values, harmonisation of formats and implementation of a metadata standard (5 points)
- B. The **design and justification of a machine learning modelling approach** suitable for spatiotemporal prediction, including model selection, training strategy and uncertainty quantification (7 points)
- C. The **validation framework to evaluate model performance**, specifying metrics and validation strategy appropriate for spatiotemporal data. (5 points)
- D. The **data governance and open data strategy**, including the design of a Data Management Plan, choice of repository, licensing and compliance with FAIR principles. (5 points)
- E. Any **infrastructure or tooling recommendations** to make the system operational and sustainable within a public research institution. (3 points)



Estudio de CASO NUMERO 2

A research institute is conducting a long-term environmental monitoring study of air quality across 50 urban and peri-urban stations in Spain. The study generates multivariate time series of pollutant concentrations (NO_2 , $\text{PM}_{2.5}$, O_3) sampled at hourly intervals over 8 years. The institute has been tasked with developing a statistical modelling framework to:

1. Characterise the spatiotemporal variability of pollutant concentrations and identify anomalous episodes.
2. Quantify the uncertainty of the model estimates to support evidence-based policy recommendations to regional authorities.
3. Integrate the results into a reproducible, automated analysis pipeline that can be rerun as new data arrives.

However, the project faces several challenges:

1. The time series show non-stationarity, seasonal patterns and frequent outliers caused by sensor malfunctions, with no established quality control protocol.
2. Previous analyses were performed manually in spreadsheets with no version control or documentation, making results non-reproducible.
3. The institute is under pressure to publish the datasets and methodology in open access, but has no established workflow for this.

Based on this scenario, develop a structured technical proposal addressing the following aspects:

- A. The **quality control and preprocessing pipeline** for multivariate environmental time series, including outlier detection, treatment of non-stationarity and sensor fault identification. (5 points)
- B. The **statistical modelling approach** to characterise spatiotemporal variability and quantify **uncertainty**, justifying the choice of method and its assumptions. (7 points)
- C. The design of a **reproducible analysis pipeline**, specifying tools, version control strategy and documentation standards. (5 points)
- D. The **open data and dissemination strategy**, including dataset documentation (metadata, data paper), repository selection, licensing and alignment with the institute's Open Access mandate under Spanish Law 14/2011. (5 points)
- E. Recommendations for **institutionalising the workflow** so that it can be maintained and updated by future staff with minimal dependency on individual researchers. (3 points)