

RESCATE DE DATOS CLIMÁTICOS EN EL PROYECTO INDECIS

Joan R. COLL¹, Gerard VAN DER SCHRIER², Enric AGUILAR¹, Andrés BISHOP¹

¹ *Centro en Cambio Climático (C3), Universidad Rovira i Virgili, España*

² *Royal Netherlands Meteorological Institute (KNMI), The Netherlands*

joanramon.coll@urv.cat

RESUMEN

En el marco del proyecto europeo INDECIS; “*Integrated approach for the development across Europe of user oriented climate indicators for GFCS high-priority sectors: agriculture, disaster risk reduction, energy, health, water and tourism*” (INDECIS 2017-2020), alrededor de 600K valores están siendo rescatados a escala diaria para dos sub-regiones europeas (Europa Central y la región de los Balcanes) para la temperatura máxima y mínima, la precipitación, las horas de sol y el grosor de la capa de nieve a lo largo del siglo XX. Dichos datos serán incorporados a la base de datos ECA&D, incrementando la cobertura espacio-temporal en Europa para mejorar la calidad y fiabilidad de los productos y servicios climáticos derivados. En primer lugar la base de datos ECA&D fue examinada con el fin de encontrar lagunas de datos en ciertas regiones europeas o series climáticas con un bajo porcentaje de datos para ser candidatas a ser rescatadas. Este ejercicio previo sirvió para focalizar los esfuerzos en la Europa Central y en la zona de los Balcanes. El método de digitalización usado es el “*key as you see*”, digitalizando directamente de la fuente de datos original o secundaria. Los valores introducidos son contrastados con los valores de la fuente original los días 10, 20 y 30 de cada mes para asegurar que se están introduciendo correctamente. Además, los valores promedio y totales mensuales son revisados, garantizando una óptima digitalización. El uso de la metodología de digitalización descrita está proporcionando una transcripción correcta de los datos según los estadísticos obtenidos.

Palabras clave: Proyecto INDECIS, rescate de datos, digitalización, base de datos ECA&D, Europa

ABSTRACT

In the framework of the European co-funded project entitled “*Integrated approach for the development across Europe of user oriented climate indicators for GFCS high-priority sectors: agriculture, disaster risk reduction, energy, health, water and tourism*” (INDECIS 2017-2020), around 600K station observations are being rescued over European sub-regions (mainly Central Europe and Balkans region) for the main climate variables (maximum and minimum temperatures, rainfall, sunshine duration and snow depth) along the 20th century at a daily scale. The rescued dataset will be incorporated into the ECA&D to increase its spatial and temporal data coverage across

Europe. An extensive examination of the ECA&D dataset was conducted first to find data gaps in European regions, or stations with low percentages of daily data as prior candidates for data recovery. The Central Europe and the Balkans were key sub-regions, among others, where data rescue efforts should be focused. Digitizing method is being carried out by using a strict "key as you see" method, meaning that the digitizers type the values provided in the data images, rather than using any coding system. Digitizers carefully cross check their values with the original source values for the 10th, 20th and 30th day of each month. This ensures that no days were skipped or repeated during the digitization process. Monthly totals and summaries are used by the digitizers to calculate these values from their daily transcribed data to check accuracy. The digitizing method is ensuring an accurate data transcription according to the obtained statistics.

Key words: INDECIS Project, data rescue, digitization, ECA&D dataset, Europe

1. INTRODUCCIÓN

El estudio de las dinámicas climáticas observadas en el pasado y actuales, el desarrollo de modelos climáticos que proyectan distintos escenarios climáticos futuros y la elaboración de productos y servicios climáticos, entre otros, precisan de observaciones meteorológicas en formato digital para poder ser usadas por la comunidad científica. Pero la falta de datos todavía se percibe hoy en día en algunos productos climáticos (Brunet and Jones, 2011). Por esa razón los esfuerzos destinados al rescate de datos se está incrementando, especialmente en países en vías de desarrollo y para datos históricos, los cuales están normalmente almacenados en formato papel a riesgo de ser extraviados (WMO, 2016).

En las dos últimas décadas, muchos proyectos han dirigido esfuerzos al rescate de datos con el fin de mejorar la calidad y longevidad de las series climáticas. Uno de los más recientes es el proyecto europeo titulado "Uncertainties in Ensembles of Regional ReAnalyses" (UERRA 2014-2017). Dicho proyecto permitió rescatar 8.8 millones de observaciones meteorológicas de las variables climáticas esenciales a escala subsidiaria en Europa y en regiones de la cuenca del Mediterráneo para el período 1877-2012 (Ashcroft et al., 2018). La nueva base de datos UERRA de alta calidad está ya incorporada a las bases de datos globales y regionales con el propósito de mejorar las salidas de los modelos de reanálisis regional y estimar las incertidumbres asociadas. Por otro lado, la iniciativa llevada a cabo por la "Atmospheric Circulation Reconstructions over the Earth" (ACRE, Allan et al., 2011) coordina actividades de rescate de datos a escala global mientras que otros proyectos como el "Mediterranean Data Rescue" (MEDARE) y la "Historical Instrumental Climatological Surface Time Series of The Greater Alpine Region" (HISTALP) lo hacen a escala regional (Auer et al., 2007; Brunet et al., 2014a, Brunet et al., 2014b). Otras iniciativas también se están llevando a cabo a nivel nacional, lideradas mayoritariamente por los servicios meteorológicos nacionales, como es el caso de Alemania (Kaspar et al., 2015).

Actualmente, el proyecto europeo INDECIS (*Integrated approach for the development across Europe of user oriented climate indicators for GFCS high-priority sectors: agriculture, disaster risk reduction, energy, health, water and*

tourism), liderado por la Universidad Rovira i Virgili (Tarragona), se centra en el desarrollo de indicadores climáticos orientados a usuarios de sectores económicos prioritarios a escala europea. Para llevarlo a cabo, parte de los esfuerzos han sido destinados al rescate de datos diarios que, potencialmente, alcanzará las 600K nuevas observaciones (de las cuales la mitad ya han sido digitalizadas). El rescate de datos está enfocado a cubrir mejor ciertas zonas de Europa desprovistas de datos continuados o con una densidad de series climáticas más baja. Este es el caso de ciertas regiones de la Europa Central y especialmente en la región de los Balcanes. Las series climáticas candidatas cubren varias décadas a lo largo del siglo XX para las variables climáticas de temperatura máxima y mínima, precipitación, horas de sol y grosor de la capa de nieve. La nueva base de datos digitalizada en INDECIS será finalmente incorporada a la base de datos del European Climate Assessment and Dataset (ECA&D) para mejorar la calidad y fiabilidad de los productos y servicios climáticos derivados.

2. MATERIALES Y MÉTODOS

2.1. Identificación de lagunas en los datos

El propósito principal del rescate de datos que se está llevando a cabo en el marco del proyecto INDECIS es mejorar, por una parte, la cobertura de datos espacial (introducir nuevas series climáticas) y, por otra parte, el relleno de lagunas de datos de series climáticas ya existentes para incrementar la calidad y longevidad de las mismas a escala europea. Las variables de interés se centraron en la temperatura máxima y mínima (TX/TN), precipitación (RR), horas de sol (SS) y grosor de la capa de nieve (SD) a escala diaria.

El primer paso consistió en examinar detenidamente la base de datos ECA&D (<http://eca.knmi.nl/>) para averiguar qué regiones contaban con un porcentaje de datos menor comparado con otras regiones con mayor densidad de estaciones. Los resultados de este primer ejercicio mostraron que las regiones situadas más al este de Europa eran las que presentaban una menor densidad de series climáticas así como también una mayor presencia de lagunas de datos, muy variable entre periodos a lo largo del siglo XX. Concretamente, la región de los Balcanes (Croacia, República de Serbia, Montenegro, Bosnia y Herzegovina y República de Macedonia) era una de las que más falta de datos registraba (Figura 1) mientras que también se identificaron lagunas de datos importantes en ciertas zonas de la Europa Central (especialmente en las Repúblicas Checa y Eslovaca).

2.2. Localización de las fuentes de datos originales

Una vez localizadas las regiones europeas con una menor cobertura espacio-temporal de datos disponibles, fue necesario obtener fuentes de datos no digitalizados para las zonas identificadas.

Se produjeron algunos contactos con servicios meteorológicos e hidrológicos nacionales en los países de las regiones de interés con la intención de obtener imágenes escaneadas de datos disponibles para la digitalización. Solamente el Servicio Meteorológico de Croacia respondió satisfactoriamente a la petición con el envío de varios archivos pdf con datos escaneados directamente de las fuentes de datos

originales. El escaneo de las fuentes originales requiere tiempo y tiene cierto coste, con lo que el envío de datos está siendo lento y laborioso.

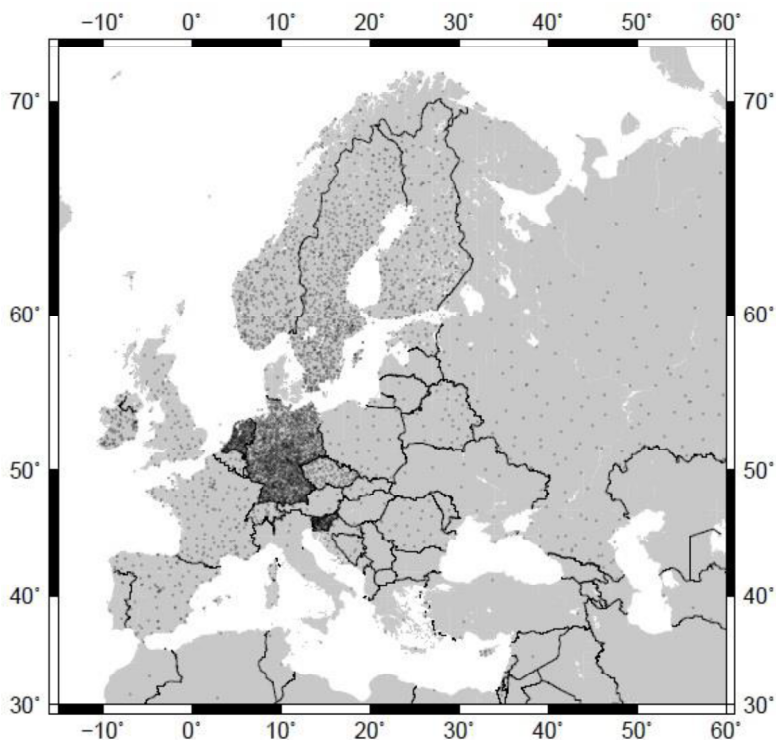


Fig. 1: Densidad espacial de series climáticas en Europa a partir de la base de datos ECA&D

Por otro lado, en la iniciativa “Mediterranean Data Rescue” (MEDARE) y en el proyecto europeo UERRA ya se localizaron varias fuentes de datos online a través del “United States of America’s National Oceanic and Atmospheric Administration/National Climatic Data Center (NOAA/NCDC) Climate Data Modernization Project” (CDMP: http://docs.lib.noaa.gov/rescue/data_rescue_home.html) para las regiones de la Europa del este, los Balcanes y regiones del Mediterráneo (Ashcroft et al., 2018; Brunet et al., 2014^a; Brunet et al., 2014b). Los datos sub-diarios de presión atmosférica, temperatura del aire y dirección y velocidad del viento ya fueron digitalizados durante el proyecto UERRA, pero muchas observaciones quedaron por digitalizar a escala diaria. De este modo, el proyecto INDECIS proporciona una gran oportunidad para rescatar esos valores diarios no digitalizados anteriormente usando las mismas fuentes de datos ya escaneadas.

La mayor parte de fuentes de datos localizadas online a través del CDMP son secundarias. Desafortunadamente las fuentes de datos secundarias son más proclives a padecer errores de transcripción si se compara con las fuentes originales. En algunas

fuentes de datos las observaciones están manuscritas mientras que en otras están mecanografiadas. También cabe destacar que la calidad de las imágenes no siempre es buena y en algunos casos cuesta leer los datos con lo que aumenta la probabilidad de cometer errores de transcripción.

2.3. Método de digitalización de datos

Antes de empezar directamente con la digitalización de los datos es necesario conocer en profundidad las fuentes de datos para familiarizarse con el formato, la estructura de las hojas de observaciones, el idioma o las anotaciones adicionales que aportan una valiosa información como metadatos. Esta inspección previa de las fuentes de datos puede evitar errores importantes derivados, por ejemplo, de la ausencia de algunas hojas de datos para ciertos meses, valores perdidos o variables inexistentes entre otros. Una vez las fuentes de datos han sido inspeccionadas exhaustivamente, puede empezar el proceso de digitalización. El método de digitalización usado en este estudio es el de un estricto “key as you see”, cosa que significa que el digitalizador transcribe los valores tal y como están escritos en la fuente de datos, sin usar ningún sistema de códigos, siguiendo las recomendaciones de la Organización Meteorológica Mundial (WMO, 2016).

El presupuesto del proyecto no ha permitido utilizar otros métodos de digitalización más complejos, como por ejemplo la digitalización por duplicado o triplicado en el que dos o más digitalizadores transcriben los mismos datos para detectar errores de digitalización al contrastarlos (Brönnimann et al., 2006). Tampoco ha sido posible usar técnicas avanzadas como el reconocimiento óptico de caracteres (OCR) o el reconocimiento de voz debido al elevado coste asociado y a que gran parte de las fuentes de datos están manuscritas, con lo que un OCR no puede leerlas correctamente.

De este modo, se ha optado por la digitalización manual para realizar este estudio. El digitalizador es el encargado de preparar una hoja de cálculo para introducir los datos con el formato año, mes, día y el valor de cada variable usando media pantalla del ordenador para insertar los datos en la hoja de cálculo y la otra media para mostrar la fuente de datos (Figura 2). Él mismo es quien contrasta los valores digitalizados con la fuente de datos los días 10, 20 y 30 de cada mes para asegurarse que el proceso de digitalización transcurre correctamente (sin valores repetidos o valores obviados). Las lagunas con valores perdidos y las posibles variaciones registradas en las fuentes de datos quedan también registrados en una hoja de metadatos siguiendo la recomendaciones de Aguilar et al., (2003). Además, cuando la fuente de datos incorpora medias o totales mensuales, el digitalizador se encarga de calcularlos a partir de los datos transcritos para comprobar su veracidad. En los casos que esos totales mensuales no estén disponibles en las fuentes de datos, un efectivo control de calidad preliminar es efectuado a partir de la detección de valores fuera de rango para cada variable transcrita con el fin de detectar valores aberrantes fruto de un error de digitalización. Los errores de digitalización encontrados se documentan en una plantilla específica (Figura 3) mientras que las correcciones se efectúan usando una copia de la serie original para preservar la trazabilidad del proceso.

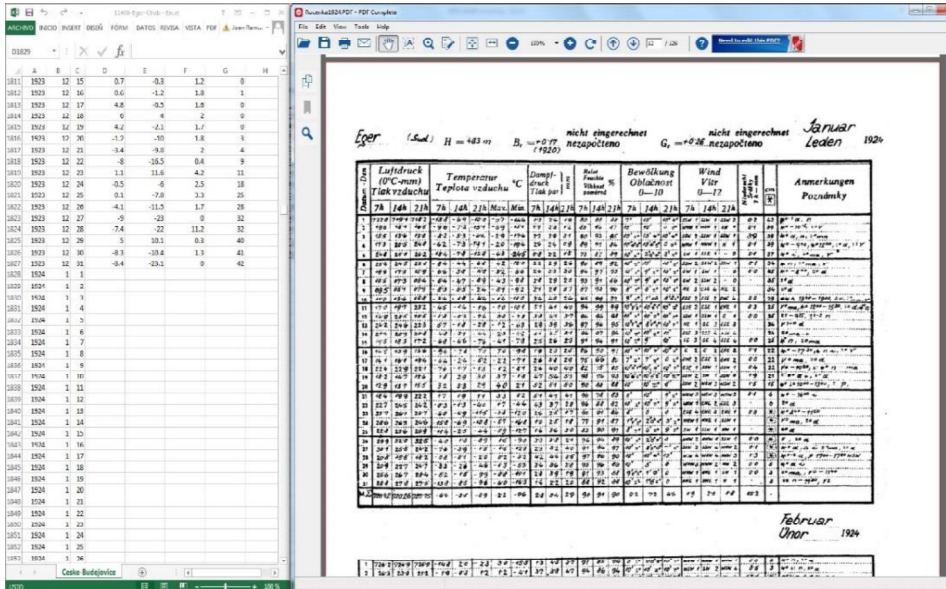


Fig. 2: Proceso de digitalización empleado en el proyecto INDECIS

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	country	station name	code	Year	Month	Day	Element	Original value	Replacement value	Detection test	Type of error	Procedure	Comments	
2	Croatia	Brodanci	5080	1931	2	27	RR	0	0,2	Visual checking	Transcription	corrected	Difficulties to be read	
3	Croatia	Brodanci	5080	1931	2	28	RR	0	4,6	Visual checking	Transcription	corrected	Difficulties to be read	
4	Croatia	Brodanci	5080	1931	10	21	RR	4	4,9	Visual checking	Transcription	corrected	Difficulties to be read	
5	Croatia	Brodanci	5080	1931	10	27	RR	3	3,9	Visual checking	Transcription	corrected	Difficulties to be read	
6	Croatia	Brodanci	5080	1931	12	10	RR	1	0,9	Visual checking	Transcription	corrected	Difficulties to be read	
7	Croatia	Brodanci	5080	1933	1	25	RR	3,4	3,7	Visual checking	Transcription	corrected	Difficulties to be read	
8	Croatia	Brodanci	5080	1933	2	23	RR	5,9	5,1	Visual checking	Transcription	corrected	Difficulties to be read	
9	Croatia	Brodanci	5080	1933	3	7	RR	0,2	0,6	Visual checking	Transcription	corrected	Difficulties to be read	
10	Croatia	Brodanci	5080	1933	6	29	RR	2,5	2,8	Visual checking	Transcription	corrected	Difficulties to be read	
11	Croatia	Brodanci	5080	1933	8	2	RR	1,5	1,6	Visual checking	Transcription	corrected	Difficulties to be read	
12	Croatia	Brodanci	5080	1933	10	14	RR	0,9	6,9	Visual checking	Transcription	corrected	Difficulties to be read	
13	Croatia	Brodanci	5080	1934	6	12	RR	4	4,9	Visual checking	Transcription	corrected	Difficulties to be read	
14	Croatia	Brodanci	5080	1936	5	27	RR	2,2	12,2	Visual checking	Source error	corrected	Typing error	
15	Croatia	Brodanci	5080	1936	7	4	RR	37,3	37,1	Visual checking	Transcription	corrected	Difficulties to be read	
16	Croatia	Brodanci	5080	1936	10	2	RR	10,8	10,3	Visual checking	Transcription	corrected	Difficulties to be read	
17	Croatia	Brodanci	5080	1937	2	18	RR	0,8	0,07	Visual checking	Transcription	corrected	Difficulties to be read	
18	Croatia	Brodanci	5080	1938	11	1	RR	3	3,3	Visual checking	Transcription	corrected	Difficulties to be read	
19	Croatia	Brodanci	5080	1939	6	28	RR	19,8	19,3	Visual checking	Transcription	corrected	Difficulties to be read	
20	Croatia	Brodanci	5080	1939	9	22	RR	2,6	9,2	Visual checking	Transcription	corrected	Typing error	
21	Croatia	Brodanci	5080	1939	10	28	RR	2,5	2,8	Visual checking	Transcription	corrected	Difficulties to be read	
22	Croatia	Brodanci	5080	1940	2	1	RR	2,8	2,9	Visual checking	Transcription	corrected	Difficulties to be read	

Fig. 3: Ejemplo de plantilla para documentar el control de calidad de la digitalización

3. RESULTADOS

A lo largo del proyecto INDECIS está previsto rescatar un total de 600K observaciones diarias (300K ya han sido rescatados) sobre temperatura máxima y mínima (en °C), precipitación (en mm), horas de sol (en horas) y grosor de la capa de nieve (en cm) en la Europa Central y la región de los Balcanes con períodos muy variables a lo largo del siglo XX. La Figura 4 muestra la distribución espacial de las

25 series climáticas que están siendo rescatadas (12 series ya lo han sido) en 7 países europeos; 11 en la República Checa, 5 en la República Eslovaca, 3 en la República de Serbia, dos en Bosnia y Herzegovina, dos más en la República de Macedonia, una en Croacia y otra en Montenegro. Los períodos a digitalizar son muy variables a lo largo del siglo XX cubriendo el período 1949-2012 para las series climáticas ubicadas en la región de los Balcanes y el período 1917-1968 para las series localizadas en la Europa Central.



Fig. 4: Distribución espacial de las series climáticas a rescatar

La Figura 5 muestra la cantidad de valores digitalizados (en millares) hasta el momento (300K observaciones) para cada serie climática y para cada variable. La cantidad total de valores digitalizados depende básicamente de la extensión del período digitalizado y del número de variables disponibles para ser rescatadas. Las series climáticas situadas en la región de los Balcanes son las que han acumulado una mayor cantidad de valores digitalizados, por ejemplo; Podgorica (Montenegro), Brodanci (Croacia), Skopje (República de Macedonia) y Zlatibor (República de Serbia). Por otro lado, las variables con más observaciones digitalizadas son la precipitación (RR) y el grosor de la capa de nieve (SD), que representan un 52% de los datos rescatados. Las observaciones de temperatura máxima (TX) y temperatura mínima (TN) rescatadas representan el 34% de los valores digitalizados mientras que las observaciones relacionadas con las horas de sol (SS) representan tan solo el 14% de las observaciones. La diferencia en la cantidad de observaciones digitalizadas depende básicamente de la disponibilidad o no de dichas variables en las fuentes de datos.

Finalmente, el control de calidad de la digitalización consistente en contrastar las medias y totales mensuales proporcionados por la fuente de datos (en la mayoría de casos) con las medias y totales mensuales calculados a partir de los datos rescatados ha demostrado que los errores cometidos durante el proceso de digitalización representan solamente el 0,6% del total de datos digitalizados. Gran parte de los errores se deben a una mala lectura del valor debido a la baja calidad de la fuente de

datos en algunas hojas, mientras que, en menor medida, existen otros errores de transcripción al introducir mal un valor (ej. insertar 104,5°C en vez de 10,5°C).

Al finalizar el proceso de digitalización y el control de calidad de la digitalización, las 25 nuevas series climáticas (600K valores) serán incorporadas a la base de datos ECA&D para mejorar la cobertura espacial y temporal en las regiones de la Europa Central y los Balcanes. Naturalmente, los nuevos datos deberán someterse a nuevos controles de calidad más exhaustivos para encontrar errores no sistemáticos así como a varias pruebas de homogeneidad para asegurar una alta calidad de los mismos y ponerlos a disposición de la comunidad científica.

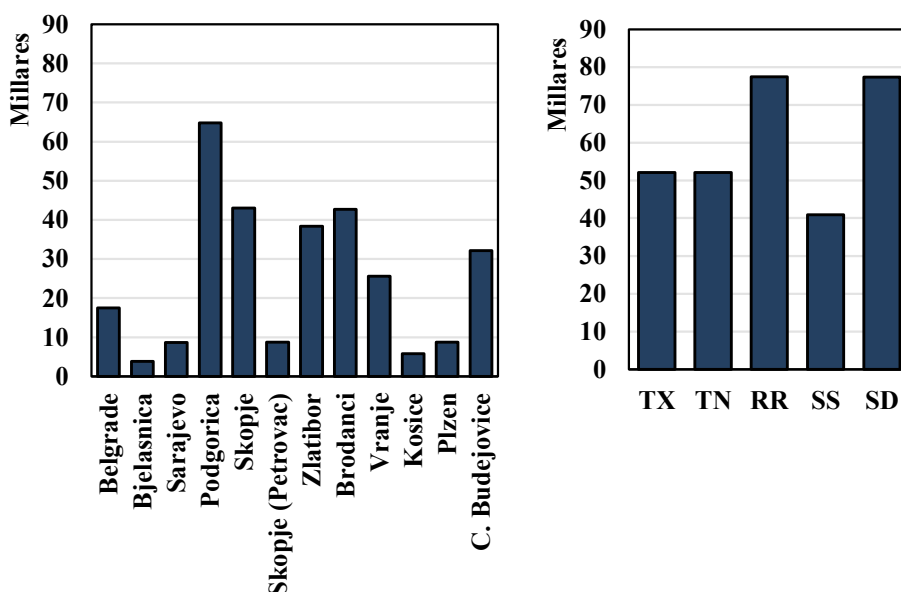


Fig. 5: Cantidad de valores digitalizados (en millares) por estaciones y variables

4. DISCUSIÓN Y CONCLUSIONES

En el marco del proyecto INDECIS, una parte de los recursos han sido destinados al rescate de datos diarios a escala europea con el fin de mejorar la calidad y fiabilidad de los productos y servicios climáticos disponibles. El presente estudio describe detalladamente todo el proceso llevado a cabo; desde la identificación de lagunas en los datos y la obtención de las fuentes de datos no digitalizados, hasta el propio proceso de digitalización y la documentación minuciosa de datos y metadatos, incluyendo también las correcciones derivadas de los errores cometidos durante la transcripción de los datos.

El proceso de identificación de lagunas en los datos, la obtención de las fuentes de datos para ser rescatados y la preparación de dichas fuentes para la digitalización de las observaciones no es tarea fácil, ya que consume mucho tiempo (Brönninmann, 2006). Concretamente, el proceso de digitalización de datos de forma manual requiere

muchas horas de dedicación y capital humano, por eso es crucial diseñar e implementar un método de digitalización efectivo y fiable para que la base de datos final resulte de alta calidad sin un alto coste.

Existen varias recomendaciones disponibles para guiar a los expertos relacionados con proyectos de rescate de datos. Por ejemplo, la guía para digitalizar manualmente datos climáticos proporcionada por Brönnimann (2006) describe el uso de tecnologías basadas en el reconocimiento óptico de caracteres (OCR) o el reconocimiento de voz para agilizar el proceso de digitalización. Pero el estudio demostró que la digitalización manual resultó el método más eficiente en términos de agilidad, errores cometidos y tiempo de procesamiento de datos posterior a la digitalización. La Organización Meteorológica Mundial apoya esta premisa (WMO, 2016) recomendando que el uso de OCRs solo es efectivo en ciertas fuentes de datos, pero el ojo humano todavía es más eficaz transcribiendo fuentes de datos manuscritas. Hoy en día, el método de digitalización más efectivo recae en la transcripción de datos por duplicado o triplicado empleando plantillas que encajen perfectamente con el formato de la fuente de datos original (WMO, 2016). Sin embargo, el coste final también se incrementa notablemente con lo que muchos proyectos no pueden hacer frente a dicho gasto extra. La simple digitalización manual con un control de calidad de la digitalización durante y al final del proceso resulta el equilibrio más efectivo entre coste y calidad final de los datos rescatados sabiendo que todavía existen algunos problemas a resolver (Ashcroft et al., 2018).

Usando el método de la digitalización manual sumado a un control de calidad de la digitalización, un total de 25 series climáticas, que potencialmente sumarán 600K observaciones diarias cuando concluya el estudio, están siendo rescatadas en 7 países de la Europa Central y la región de los Balcanes a lo largo del siglo XX. Las variables climáticas de interés son la temperatura máxima y mínima, la precipitación, las horas de sol y el grosor de la capa de nieve.

La incorporación de dichas series climáticas a la base de datos ECA&D ayudará a rellenar las lagunas de datos identificadas previamente en estas regiones europeas. La nueva base de datos recuperada deberá someterse a más controles de calidad rigurosos para encontrar errores no sistemáticos así como a varias pruebas de homogeneidad para asegurar una alta calidad de los mismos y poder usarlos para la creación de nuevos productos y servicios climáticos.

AGRADECIMIENTOS

El proyecto INDECIS es parte de ERA4CS, un ERA-NET iniciado por JPI Climate, financiado por FORMAS (SE), DLR (DE), BMFWF (AT), IFD (DK), MINECO (ES), ANR (FR) y co-financiado por la Unión Europea (Grant 690462).

REFERENCIAS

- Aguilar, E., Auer, I., Brunet, M., Peterson, T. C. and Wieringa, J. (2003). Guidelines on Climate Metadata and Homogenization, World Meteorological Organisation, (1186), 55.
- Allan, R., Brohan, P., Compo, G. P., Stone, R., Luterbacher, J., Brönnimann, S., Allan, R., Brohan, P., Compo, G. P., Stone, R., Luterbacher, J. and Brönnimann, S.

- (2011). The International Atmospheric Circulation Reconstructions over the Earth (ACRE) Initiative, *Bulletin of American Meteorological Society*, 92(11), 1421–1425, doi:10.1175/2011BAMS3218.1.
- Ashcroft L., Coll J.R., Gilabert A., Domonkos P., Aguilar E., Sigró J., Castellà M., Uden P., Harris I., Jones P., Brunet M. (2018). A rescued dataset of sub-daily meteorological observations for Europe and the Mediterranean region, 1877–2012. *Earth System Science Data* (under revision).
- Auer, I., Böhm, R., Jurkovic, A., Lipa, W., Orlik, A., Potzmann, R., Schöner, W., Ungersböck, M., Matulla, C., Briffa, K., Jones, P., Efthymiadis, D., Brunetti, M., Nanni, T., Maugeri, M., Mercalli, L., Mestre, O., Moisselin, J.-M., Begert, M., Müller-Westermeier, G., Kveton, V., Bochnicek, O., Stastny, P., Lapin, M., Szalai, S., Szentimrey, T., Cegnar, T., Dolinar, M., Gajic-Capka, M., Zaninovic, K., Majstorovic, Z. and Nieplova, E. (2007). HISTALP—historical instrumental climatological surface time series of the Greater Alpine Region, *International Journal of Climatology*, 27(1), 17–46, doi:10.1002/joc.1377.
- Brönnimann, S., Annis, J., Dann, W., Ewen, T., Grant, A. N., Griesser, T., Krähenmann, S., Mohr, C., Scherer, M. and Vogler, C. (2006). A guide for digitising manuscript climate data, *Climate of the Past*, 2(3), 191–207, doi:10.5194/cpd-2-191-2006.
- Brunet, M. and Jones, P. (2011). Data rescue initiatives: bringing historical climate data into the 21st century, *Climate Research*, 47(1), 29–40, doi:10.3354/cr00960.
- Brunet, M., Gilabert, A., Jones, P. and Efthymiadis, D. (2014a). A historical surface climate dataset from station observations in Mediterranean North Africa and Middle East areas. *Geoscience Data Journal*, 1(2), 121–128, doi:10.1002/gdj3.12.
- Brunet, M., Jones, P. D., Jourdain, S., Efthymiadis, D., Kerrouche, M. and Boroneant, C. (2014b). Data sources for rescuing the rich heritage of Mediterranean historical surface climate data, *Geoscience Data Journal*, 1(1), 61–73, doi:10.1002/gdj3.4.
- Kaspar, F., Tinz, B., Mächel, H. and Gates, L. (2015). Data rescue of national and international meteorological observations at Deutscher Wetterdienst, *Advances on Science Research*, 12, 57–61, doi:10.5194/asr-12-57-2015.
- World Meteorological Organization (2016). *Guidelines on Best Practices for Climate Data Rescue 2016*.