

Herramienta para la evaluación de la resiliencia en los sectores agrícola y forestal mediante tecnologías Big Data

VM. Fernández-Pacheco^{a1}, E. Antuña Yudego^{b1}, E. Álvarez Álvarez^{a2}, J. Fernández Francos^{a3}, MJ. Suárez López^{a4}, A. Fernández-Villán^{b2}, JL. Carús-Candás^{b3}, V. Wolfs^{c1}, V. Ntegeka^{c2}, P. Willems^{c3}, M. García^{d1}, A. Palacios-Orueta^{e1}, J. Litago^{e2}, L. Juana^{e3}

^a Universidad de Oviedo. Escuela Politécnica de Mieres-C/Gonzalo Gutiérrez Quirós s/n, Mieres, Asturias. E-mail: ^{a1}victorpacheco.ing@gmail.com, ^{a2}edualvarez@uniovi.es, ^{a3}jffrancos@uniovi.es, ^{a4}suarezmaria@uniovi.es.

^b TSK- C/Ada Byron, 220, Gijón. E-mail: ^{b1}elena.antuna@grupotsk.com, ^{b2}alberto.fernandez@grupotsk.com, ^{b3}juanluis.carus@grupotsk.com.

^c Sumaqua- C/Tiensesteenweg 28, Leuven, Bélgica. E-mail: ^{c1}vincent.wolfs@sumaqua.be, ^{c2}victor.ntegeka@sumaqua.be, ^{c3}patrick.willems@sumaqua.be

^d Department of Environmental Engineering, Technical University of Denmark (DTU)- C/2800 Kgs. Lyngby, Denmark. E-mail: ^{d1}mgarc@env.dtu.dk

^e Universidad Politécnica de Madrid- Ciudad Universitaria s/n, 28040 Madrid, España. E-mail: ^{e1}alicia.palacios@upm.es, ^{e2}javier.litago@upm.es, ^{e3}luis.juana@upm.es

Línea temática B | Hidrología, usos y gestión del agua.

RESUMEN

El estrés hídrico constituye un grave problema que se incrementa cada día. En la actualidad, las soluciones tradicionales no son capaces de predecir indicadores de resiliencia útiles para el sector agrícola y forestal. La solución propuesta, bajo el acrónimo FORWARD (operational monitoring and FOREcasting system for Resilience of agricultura and forestry under intensification of the WAtER cycle), pretende investigar, desarrollar y configurar una herramienta de software capaz de identificar y recopilar variables ecohidrológicas desde fuentes de datos libres. Estos datos serán analizados mediante combinación de minería de datos con distintos modelos de análisis para obtener como resultados indicadores de resiliencia y parámetros ecohidrológicos.

Palabras clave | Big Data; Resiliencia hídrica; Evapotranspiración; Sentinel; Modis.

INTRODUCCIÓN

El estrés hídrico constituye un grave problema que se incrementa día tras día debido a la actividad y al desarrollo humano. A esta circunstancia se deben sumar los efectos cada vez más notables del cambio climático, caracterizados por eventos extremos de lluvia y temperaturas (IPCC, 2014). Este escenario provoca un gran impacto negativo en el sector agrícola y forestal, especialmente en aquellas regiones con recursos hídricos muy limitados (Knapp et al., 2008). De esta forma, surge una nueva necesidad de predecir las consecuencias de este escenario para evaluar la resiliencia de los ecosistemas bajo condiciones hidroclimáticas variables. En la actualidad, las soluciones tradicionales propuestas para realizar predicciones no son capaces de afrontar este problema con resultados satisfactorios. El principal motivo de esta limitación se debe buscar en la gran variedad de escalas (tanto espaciales como temporales) asociadas a los procesos hidrológicos (Willems, 2013, Wheeler y Von Braun, 2013, Piao et al., 2010) lo que resulta en una gran complejidad que los métodos clásicos no son capaces de afrontar correctamente. Para superar esta barrera, se propone una nueva metodología de análisis capaz de manejar de forma eficiente las distintas escalas consideradas y obtener unos resultados de predicción fiables que permitan incrementar la resiliencia y eficiencia de uso de los recursos hídricos (Scheffer, 2009).

En la nueva solución propuesta, bajo el acrónimo FORWARD (operational monitoring and FOREcasting system for Resilience of agricultura and forestry under intensification of the WAtER cycle), se pretende investigar, desarrollar y configurar una herramienta de software capaz de identificar y recopilar variables ecohidrológicas (evapotranspiración, Índice de Precipitación Estandarizada, Fracción de Evaporación...) desde fuentes de datos libres y analizarlas mediante la combinación de

minería de datos con distintos modelos de análisis para obtener como resultados indicadores de resiliencia (productividad y rendimiento del terreno) y de parámetros hidrológicos clave como la evapotranspiración, así como una predicción climática (precipitación, temperatura, humedad...). Una de las características primordiales de esta herramienta consiste en la implementación de la misma en una plataforma Big Data a la que se incorporará la información de las diversas fuentes (Bharathi et al., 2016, Walli et al., 2013 Moser et al., 2014, Karantzalos et al., 2015, Ma et al., 2015). Este factor se presenta como una gran innovación desde el punto de vista tecnológico ya que la tecnología Big Data es capaz de almacenar y manejar grandes volúmenes de información de forma eficiente, en este caso al servicio de la hidrología, lo que constituye una actuación sin precedentes.

Objetivos

El objetivo principal de la investigación que se presenta, es diseñar y desarrollar un sistema inteligente capaz de identificar y obtener variables ecohidrológicas; permitiendo su análisis y por tanto generando como resultados indicadores de resiliencia y parámetros hidrológicos. Para obtener dicho objetivo principal se identifican una serie de objetivos técnicos basados en el desarrollo de técnicas avanzadas de análisis de datos:

1. Desarrollar e implementar una arquitectura Big Data a medida y adaptable capaz de manejar múltiples fuentes de datos con una alta heterogeneidad de variables en cada una de ellas (imágenes, datos en crudo, agregados, indicadores existentes), técnicas de minería de datos, así como modelos que operen a diferentes escalas (local, regional, nacional y continental).
2. Proporcionar: (a) modelos de previsión mejorados tanto a diferentes escalas temporales (diarios, mensuales o estacionales) como a diferentes escalas espaciales; (b) monitorización de las variables e indicadores más relevantes en los campos de la agricultura y los sistemas forestales mediante la combinación de diferentes tipos de modelos (*data-mining*, *data-driven* y *process-based*) en una arquitectura Big Data.
3. Estudiar y entender la resiliencia del recurso hídrico en ecosistemas forestales y de agricultura de las regiones que presenten carencias de agua y eventos extremos, especialmente en el caso de sequías, y en el contexto del cambio climático. Se pretenden estudiar los sitios más vulnerables a este problema mediante la combinación de todas las fuentes de datos disponibles y técnicas avanzadas de minería de datos.

DESCRIPCIÓN DEL SISTEMA

El sistema de generación de indicadores de resiliencia y ecohidrológicos permitirá servir información a diferentes escalas (tanto espaciales como temporales) aportando información útil en el sector agrícola y forestal. El sistema, tal y como muestra la Figura 1, estará constituido por un módulo de adquisición de datos, una plataforma de procesamiento y una capa de explotación que permita el manejo y visualización de la información adquirida y procesada.

A través de la interacción entre los tres módulos principales del sistema se llevará a cabo la recogida de datos de todas las fuentes seleccionadas, su transformación a indicadores de resiliencia y parámetros ecohidrológicos mediante técnicas avanzadas de análisis y manejo de datos y su explotación a través de gráficos e imágenes GIS que permitan ayudar a la decisión dentro del ámbito agroforestal.



Figura 1 | Diagrama del sistema propuesto.

Módulo de adquisición de datos

El módulo de adquisición de datos es el encargado de captar los datos de las diversas fuentes, llevar a cabo el pre-procesamiento que sea necesario e incorporar esa información al sistema. La Figura 2 muestra los principales bloques que compondrán este módulo y que se describen en detalle a continuación.



Figura 2 | Bloques funcionales del módulo de adquisición de datos.

a) Fuentes

El sistema se alimentará a través de diversas fuentes, todas ellas abiertas. Esto incluirá Observaciones Terrestres (EO, por sus siglas en inglés), redes de calibración, datos de medios sociales no estructurados, bases de datos globales existentes y proyecciones ambientales. Algunas de las fuentes de información más importantes con las que se contará son:

- ECMWF (Centro Europeo de Previsiones Meteorológicas a Plazo Medio, en español) del que se podrán recopilar datos como la temperatura del aire, la radiación solar, la precipitación o la humedad relativa (Doman et al., 2014).
- MODIS (Moderate-Resolution Imaging Spectroradiometer) del que se obtendrán parámetros como el NDVI (Índice de Vegetación de Diferencia Normalizada), FAPAR (Fracción de la Radiación Fotosintéticamente Activa Absorbida), LAI (Índice de Área Foliar), emisividad, LST (Temperatura de la Superficie de la Tierra) y su tiempo o el albedo (porcentaje de radiación que cualquier superficie reflecta respecto a la radiación que incide sobre la misma) (Modis, 2017).
- SENTINEL 2 un satélite más reciente que MODIS que permite obtener parámetros similares con mayor resolución espacial pero no temporal. Los datos de SENTINEL2 se emplearán para validar los modelos (Sentinel, 2017).
- Eddy Covariance flux tower es un sistema que permite medir el dióxido de carbono, el vapor de agua (evapotranspiración), el flujo de energía e incluso mediciones adicionales biológicas y meteorológicas en campo. Sus datos serán utilizados para validar y regionalizar los modelos (Flux-net, 2017).

b) Procesador

El procesador será el encargado de conseguir el correcto funcionamiento de todos los bloques que componen el módulo de adquisición de datos. Este componente permitirá centralizar los datos provenientes de las diferentes fuentes contando con al menos las siguientes capacidades desde el punto de vista del control del sistema:

- Capacidad de configuración y administración avanzada.
- Soporte multiprotocolo para obtener los datos de las diferentes fuentes.
- Medidas para la tolerancia a fallos como por ejemplo mecanismos de autodiagnóstico.
- Recursos de computación suficientes para cumplir con los requisitos de frecuencia de muestreo y de volumen de manejo de datos.
- Capacidades de pre-procesamiento de los datos provenientes de las diferentes fuentes con objeto de extraer información de más alto nivel antes de proceder a su envío a la plataforma de monitorización. Este requisito es especialmente importante en el caso de variables con altas frecuencias de muestreo.

Módulo de procesamiento de datos

El módulo de procesamiento de datos alberga toda la lógica necesaria para poder transformar la información capturada a través de las diversas fuentes de información (MODIS, SENTINEL, ECMWF, etc.) en información valiosa para el usuario. Las tecnologías empleadas en el diseño de este módulo cumplirán requisitos de escalabilidad, alta tolerancia a fallos y redundancia. Teniendo en cuenta estas características, se utilizarán tecnologías especializadas en el tratamiento masivo de datos (Big Data) y que puedan garantizar las condiciones de tiempo real que son exigidas.

La Figura 3 muestra los principales bloques funcionales que componen este módulo y que se describirán a continuación.

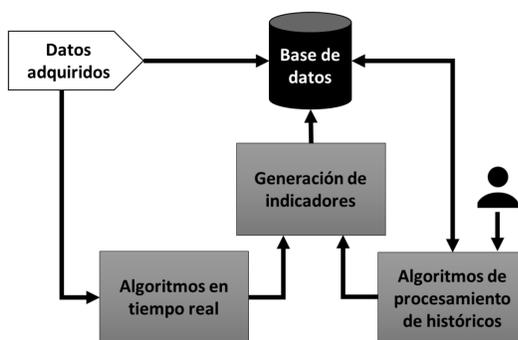


Figura 3 | Bloques funcionales del módulo de procesamiento de datos.

a) Base de datos

Este módulo estará preparado para garantizar un alto rendimiento en la escritura, de tal manera que se pueda realizar el almacenamiento de forma fiable de toda la información que sea generada en el sistema. Esta base de datos albergará la información adquirida de todas las fuentes del sistema, la información que se genere durante el procesamiento y la información relativa a las salidas de la plataforma para su consulta. Por esta razón, todos los bloques que componen este módulo de procesamiento se alimentan de ella.

Se prevé el uso de tecnologías de almacenamiento distribuido para lo que se utilizará una base de datos NoSQL que asegure la integridad de la información y permita consultas cruzadas de datos con una velocidad de respuesta adecuada. Se prestará máxima atención a la escalabilidad del sistema.

b) Algoritmos en tiempo real

A medida que los datos provenientes de las fuentes sean recibidos en el módulo de procesamiento, serán estudiados y procesados para alimentar a los modelos teóricos que simulan la dinámica de la vegetación. Para ello se prevé el uso de Modelos Dinámicos de Vegetación Mundial (DGVM) y del Simulador General de Ecosistemas de Lund-Postdam-Jena (LPJ-GUESS) (Huxman et al, 2004).

Por otro lado, dada la importancia de la evapotranspiración, se presenta especial atención a la estimación de la misma a partir de los datos de Observaciones Terrestres (EO) obtenidos de ECMWF y MODIS aplicando el modelo un modelo de evapotranspiración (García et al, 2013). Además, se realizarán enfoques de Eficiencia de Uso (Light Use Efficiency approach) mediante una adaptación de ese modelo de evapotranspiración.

c) Algoritmos de procesamiento de históricos

El procesamiento de históricos se basa en el análisis y procesamiento de toda la información almacenada en la base de datos del sistema. Para ello, se utilizarán principalmente técnicas de aprendizaje supervisado y diferentes modelos matemáticos. El procesamiento histórico alimentará el módulo de generación de indicadores adaptándolos a las condiciones particulares de cada fuente.

Las técnicas de aprendizaje supervisado y procesamiento estadístico se emplearán para deducir una función a partir de los datos disponibles. Se tendrán en cuenta los indicadores generados por los algoritmos en tiempo real y se emplearán los datos históricos para validar el modelo. El objetivo será crear una función capaz de predecir el valor correspondiente a cualquier objeto de entrada válida después de disponer de suficientes datos de ejemplo. Para ello, se evaluará el rendimiento de diferentes técnicas de aprendizaje supervisado (Redes Neuronales, o Máquinas de Soporte Vectorial, entre otras) y diferentes modelos matemáticos, principalmente estadísticos, donde se aplicarán los principales test y procedimientos para investigación de observaciones bien conocidas (correlación, análisis de regresión, etc.).

d) Generación de indicadores

El módulo de generación de indicadores creará indicadores de resiliencia y de parámetros hidrológicos clave, así como información climática convirtiéndose en una herramienta operativa de apoyo a la decisión. Este módulo se alimenta de la salida tanto del módulo de algoritmos en tiempo real como del módulo de algoritmos de procesamiento históricos.

Su objetivo es transformar la salida del módulo de algoritmos en tiempo real en información operativa para los usuarios. Estos indicadores asistirán en la toma de decisiones de las instituciones (comunidad de regantes, organismos públicos...) o empresas que requieran información. Así, los indicadores deberán ser eficientes en tiempo de procesamiento y fiables desde el punto de vista de sus resultados y significativos. Los outputs del sistema serán por tanto indicadores ecohidrológicos y variables que reflejen la disponibilidad de agua y el estado de los cultivos y bosques, incluyendo, entre otros, el Índice de Precipitación Estandarizada (SPI), la Productividad Primaria Neta (NPP), la Fracción de Evaporación (EF), la Eficiencia de Uso de Agua (WUE) (Dolman et al., 2014) y características del suelo y de caudales.

e) Módulo de explotación de datos

Toda la información generada y procesada será explotada a través de una interfaz de usuario que permitirá la supervisión general del sistema. El módulo de explotación de datos se basará en herramientas Web con distintas funciones, como la posibilidad de generar gráficos y estadísticas interactivas, obtener información con vista sobre mapa o enviar datos a los usuarios adaptados a distintos dispositivos: PCs, teléfonos inteligentes, etc.

PRUEBAS DEL SISTEMA

Ubicación seleccionada

El emplazamiento definido para la realización de las pruebas del sistema es la Península Ibérica, de donde se dispone de datos de fuentes libres, existiendo además la posibilidad de realizar pruebas que permitan la validación de estos modelos. Además, se emplearán escenarios de cambio climático derivados de las bases de datos de CMIP5 y EURO-CORDEX (Willems et al., 2012). Además, un desarrollo futuro permitiría que la metodología empleada fuese aplicable a otras regiones del continente europeo.

Resultados esperados

Los resultados esperados tras el desarrollo e implementación del sistema construido en FORWARD son fundamentalmente:

- Contribución al estado del arte y mejora de la comprensión sobre los procesos implicados en la resiliencia de los sistemas agrícolas y forestales, así como la generación de nuevos indicadores del proceso.
- Desarrollo de nuevas técnicas analíticas y de minería de datos y herramientas para la gestión eficiente del recurso hídrico y apoyo a la toma de decisiones basadas en la tecnología Big Data.
- Contribución a la creación de planes públicos de actuación en colaboración con gobiernos en la lucha contra el cambio climático y protección de los recursos naturales, así como la concienciación y formación social.

AGRADECIMIENTOS

Los autores desean agradecer a la UE y los siguientes organismos: Centro para el Desarrollo Tecnológico Industrial (CDTI), “Innovation Fund Denmark” (IFD) y “Flanders Innovation & Entrepreneurship” (VLAIO) por su financiación, en el marco del consorcio internacional colaborativo FORWARD bajo la convocatoria WaterWorks2015-ERA-NET Cofund. Esta convocatoria es una parte integral de las Actividades Conjuntas de 2016 desarrollado por “Water JPI”. El consorcio está formado por TSK, DTU (Danmarks Tekniske Universitet) y SUMAQUA con la colaboración de la UPM (Universidad Politécnica de Madrid) y la Universidad de Oviedo.

REFERENCIAS

- Bharathi, T., Ramya, R., & Priya, N. M. (2016). Real-Time Big Data Analytical Architecture for Remote Sensing Application. *Networking and Communication Engineering*, 8(7), 269-270.
- Dolman, A.J., Miralles, D.G., & Jeu, R.A. (2014). Fifty years since Monteith's 1965 seminal paper: the emergence of global ecohydrology. *Ecohydrology*, 8(3), 897-902.
- Flux-net (2017). <https://fluxnet.ornl.gov/>. Consultada: 30-07-2017.
- García, M., Sandholt, I., Ceccato, P., Ridler, M., Mogin, E., Kergoat, L., Morrilas, L., Timouk, F., Fensholt, R., & Domingo, F. (2013). Actual evapotranspiration in drylands derived from in-situ and satellite data: *Assessing biophysical constraints*. *Remote Sensing of Environment* 131, 103-118.
- Huxman, T. E., Smith, M. D., Fay, P. A., Knapp, A. K., Shaw, M. R., Loik, M. E., & Pockman, W. T. (2004). Convergence across biomes to a common rain-use efficiency. *Nature*, 429(6992), 651-654.

- IPCC. 2014. Impacts, Adaptation, and Vulnerability. Part A: Global and Sectoral Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. *USA Cambridge University Press*, 1, 32.
- Karantzalos, K., Bliziotis, D., & Karmas, A. (2015). A scalable geospatial web service for near real-time, high-resolution land cover mapping. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 8(10), 4665-4674
- Knapp A.K., Beier C., Briske D.D., Classen A.T., Luo Y., Reichstein M., Smith M.D., Smith S.D., Bell J.E., Fay P.A., Heisler J.L., Leavitt S.W., Sherry R., Smith B. & Weng E. (2008). Consequences of More Extreme Precipitation Regimes for Terrestrial Ecosystems. *Bioscience*, 58, 811-821.
- Ma, Y., Wang, L., Liu, P., & Ranjan, R. (2015). Towards building a data-intensive index for big data computing – A case study of Remote Sensing data processing. *Information Sciences*, 319, 171-188.
- Modis (2017). <https://modis.gsfc.nasa.gov/>. Consultada: 30-07-2017.
- Moser, L., Voigt, S., Schoepfer, E., & Palmer, S. (2014). Multitemporal wetland monitoring in Sub-Saharan West-Africa using medium resolution optical satellite data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 7(8), 3402-3415.
- Piao, S., et al. (2010). The impacts of climate change on water resources and agriculture in China. *Nature* 467, 43-51.
- Scheffer, M. (2009). *Critical transitions in nature and society*. Princeton University.
- Sentinel (2017). http://www.esa.int/Our_Activities/Observing_the_Earth/Copernicus/Sentinel-2. Consultada: 30-07-2017
- Walli, A., Tøttrup, C., Naeimi, V., Bauer-Gottwein, P., Bila, M., Mufeti, P., & Koetz, B. (2013, September). TIGER-NET—Enabling an earth observation capacity for integrated water resource management in Africa. *In ESA Living Planet Symp.*, Edinburgh, UK.
- Wheeler, T., & Von Braun, J. (2013). Climate change impacts on global food security. *Science*, 341 (6145), 508-513.
- Willems, P., Olsson, J., Arnbjerg-Nielsen, K., Beecham, S., Pathirana, A., Bülow Gregersen, Madsen, H., & Nguyen, V. (2012). Impacts of climate change on rainfall extremes and urban drainage systems. *London – New York: IWA Publishing*
- Willems, P. (2013). Multidecadal oscillatory behaviour of rainfall extremes in Europe. *Climatic Change*, 120(4), 931-944.