

## Perfiles verticales de temperatura y humedad a partir de observaciones IASI: Implicaciones para el futuro MTG-IRS

Niobe Peinado-Galan<sup>(1)</sup>, Xavier Calbet<sup>(2)</sup>

<sup>(1)</sup> AEMET, DT de la C. Valenciana Av Botánico Cavanilles 3, 46010, Valencia, npeinadog@aemet.es

<sup>(2)</sup> AEMET, C/ Leonardo Prieto Castro 8, 28040 Madrid, xcalbeta@aemet.es

**Resumen:** El objetivo principal de este trabajo es caracterizar las capacidades del MTG-IRS para suministrar información de los perfiles de temperatura y humedad atmosférica utilizando IASI-MetOp como *proxy* de MTG-IRS. Para lograr este objetivo, hemos desarrollado un método de recuperación no lineal basado en técnicas de *Machine Learning/Artificial Intelligence* (ML/AI) para estimar los perfiles verticales atmosféricos a partir de medidas de IASI (como *proxy* de MTG-IRS). Se trata de un modelo de *Kernel Ridge Regression* (KRR), el cual nos permite minimizar el conjunto de datos de entrenamiento, seleccionando una cantidad representativa de *Support Vectors* (SV) para el entrenamiento y predicción de los perfiles atmosféricos de temperatura y humedad. Los resultados muestran una mejora significativa en el costo computacional de las recuperaciones de estas variables a partir de sondeadores hiperspectrales, así como en la calidad de las mismas. Esto nos permite, además de recuperar estas variables con gran precisión, determinar los perfiles de inestabilidad atmosférica en cielos cubiertos con una fracción de nubes de hasta el 80%. Los resultados de la aplicación del modelo KRR al futuro MTG-IRS proporcionarán una gran mejora en los análisis meteorológicos para las situaciones meteorológicas analizadas.

**Palabras clave:** MTG, IASI, KRR, *Machine Learning*, inteligencia artificial.

### *Temperature and humidity atmospheric vertical profiles from MTG-IRS data*

**Abstract:** In this work, a non-linear retrieval algorithm is presented for IASI-MetOp (as a proxy of MTG-IRS) to characterize its capabilities in providing information of the atmospheric temperature and humidity profiles. To achieve this goal, we have developed a method based on Machine Learning/Artificial Intelligence techniques for IASI-MetOp (as a proxy of MTG-IRS) to estimate atmospheric vertical profiles. Obtained results show that the Kernel Ridge Regression (KRR) retrieval model developed in this work, is able to retrieve accurate atmospheric temperature and humidity profiles below skies with cloud fraction until 80%. This result is particularly important for providing the atmospheric structure information needed to improve convection forecasts. It is important to note that the principal goal for the KRR model is to enable the real-time production of the physical variables from the satellite hyperspectral resolution radiance measurements. The results from the application of the KRR model to future MTG-IRS will provide a big improvement in the meteorological analyses for analysed weather situations.

**Keywords:** MTG, IASI, KRR, *Machine Learning*, artificial intelligence.

## 1. INTRODUCCIÓN

Durante los últimos 30 años, EUMETSAT ha proporcionado información crítica para contribuir a la predicción operativa y al seguimiento del clima. En el futuro, EUMETSAT añadirá a su flota de satélites el Meteosat de Tercera Generación (MTG), que supondrá una evolución en la monitorización de la atmósfera desde satélites. La serie de satélites comprenderá: cuatro satélites de imágenes (MTG-I) y dos satélites de sondeo (MTG-S).

MTG-S incluye el *Infrared Sounder* (IRS), que se basa en un interferómetro de imágenes con una resolución espectral muy alta, comúnmente conocido como sondas hiperspectrales infrarrojas.

Hoy en día, las observaciones de datos satelitales, y en particular las sondas hiperspectrales infrarrojas, proporcionan una gran cantidad de datos en forma de muestras espectrales. A partir de ellos, se pueden derivar el

estado actual de la atmósfera, la temperatura y los perfiles de vapor de agua. Pero estos datos requieren técnicas de procesamiento que minimicen el alto costo computacional que induce esta gran cantidad de información. Este es un procedimiento complejo debido al proceso físico no lineal que está presente en la conversión de radiancias medidas en el infrarrojo a estados atmosféricos.

Como punto de partida, los sensores hiperspectrales actuales a bordo de los satélites polares proporcionan un buen indicador de los datos MTG-IRS. En este sentido, el instrumento *Infrared Atmospheric Sounding Interferometer* (IASI) a bordo del satélite MetOp, es un buen ejemplo. Actualmente, IASI, entre otras sondas hiperspectrales a bordo de satélites, se utilizan para calcular productos satelitales a partir de los perfiles atmosféricos de temperatura y humedad recuperados. Estos productos se generan con una alta resolución espacial y

temporal para admitir aplicaciones meteorológicas en tiempo real.

Para resolver el problema de inversión, los métodos de regresión lineal ofrecen una herramienta simple que proporciona recuperaciones atmosféricas estadísticas. Sin embargo, debido a las dependencias no lineales entre las mediciones de espectros y el vector de estado atmosférico, los métodos no lineales a menudo obtienen mejores resultados que los lineales. Por esta razón, los métodos de regresión no lineal, aunque más complejos desde el punto de vista del algoritmo del modelo y de su entrenamiento, están diseñados para hacer frente a este problema (Aires et al., 2002, Blackwell et al., 2005).

En este trabajo, se presenta un algoritmo de recuperación no lineal para IASI (como un *proxy* de MTG-IRS) para caracterizar sus capacidades para proporcionar información de los perfiles de temperatura y humedad atmosférica. Para lograr este objetivo, hemos desarrollado un método basado en técnicas de *Machine Learning/Inteligencia Artificial (ML/AI)* para IASI (IASI se entenderá como *proxy* de MTG-IRS en el resto del texto) para estimar perfiles verticales atmosféricos. Con el objetivo de diseñar un método de regresión no lineal operativo que compense las limitaciones debidas a: 1) la gran cantidad de datos en forma de muestras espectrales proporcionadas por sondas hiperespectrales y 2) la necesidad de minimizar el alto costo computacional de las técnicas de procesamiento requerido por estos datos. El método desarrollado en este trabajo es un modelo *Kernel Ridge Regression (KRR)* siguiendo los pasos establecidos en (Camps-Valls et al., 2012).

Los resultados obtenidos muestran que el modelo de recuperación de KRR presentado es capaz de recuperar perfiles precisos de temperatura y humedad atmosférica bajo cielos con una fracción de nubes de hasta el 80%. Este resultado es particularmente importante para proporcionar la información de la estructura atmosférica necesaria para mejorar los pronósticos de convección. Es importante tener en cuenta que el objetivo principal del modelo KRR es permitir la producción en tiempo real de las variables físicas a partir de las mediciones de radiancia de resolución hiperespectral del satélite. Los resultados de la aplicación del modelo KRR al futuro MTG-IRS proporcionarán una gran mejora en los análisis meteorológicos para las situaciones meteorológicas analizadas. Estos resultados, junto con otros datos y productos satelitales, son una poderosa herramienta para el NWC SAF para analizar y monitorizar todas las situaciones atmosféricas accesibles para las observaciones de MTG.

## 2. MATERIAL Y MÉTODOS

El método KRR utilizado en este trabajo para recuperar las variables atmosféricas a partir de las mediciones de IASI consiste en un Kernel gaussiano de función de base radial (*Radial Basis Function kernel*), cuya ecuación es:

$$K = \frac{-|X_{test} - X_{train}|^2}{2\sigma^2} \quad (1)$$

Donde  $X_{test}$  es el nuevo vector de datos medido por IASI del cual se quieren obtener las recuperaciones de las variables atmosféricas (en este caso del 15 de julio de 2015 entre las 9:45Z y las 11:27Z).  $X_{train}$  son los vectores de datos medidos por IASI y asimilados por el modelo para su entrenamiento (datos de los días 13 y 14 de julio de 2015). Y  $\sigma$  se define como la distancia al kernel.

Debido a la gran cantidad de datos proporcionados por IASI en forma de radiancias, hemos desarrollado un método que nos permite reducir el costo computacional y mejorar el modelo KRR, reduciendo el tamaño de la muestra introducido como *input*.

Para lograr este objetivo, hemos reducido la dimensionalidad de los datos a través del análisis de componentes principales (PCA, por sus siglas en inglés). Este es un procedimiento aconsejable debido al tamaño del conjunto de datos. A continuación, hemos establecido un método de selección basado en *Support Vectors (SVs)* (Fine, 2001; Camps-Valls et al., 2004; Foody y Mathur, 2004; Melgani y Bruzzone, 2004; Camps-Valls y Bruzzone, 2005; Yang et al., 2006; Durbha et al., 2007; Hastie, 2009; Muñoz-Marí et al., 2009; Arenas-García et al., 2013; Pérez-Suay et al., 2017), el cual nos permite, a partir de las componentes principales de los datos de muestra, seleccionar los datos que mejor representan las condiciones atmosféricas sin redundancia. El algoritmo selecciona los SV del conjunto de datos de entrenamiento basándose en la distancia del kernel. Esta selección de SV permite seleccionar una cantidad limitada de datos, minimizando el costo computacional requerido para las recuperaciones (*retrievals* en inglés) de las variables atmosféricas a partir de las radiancias de IASI (como un *proxy* de MTG-IRS), sin perder información del estado de la atmósfera y por lo tanto sin perder calidad en las recuperaciones de dichas variables atmosféricas.

Este método, ha demostrado ser una técnica prometedora en el pronóstico meteorológico así como en cualquier otra aplicación en la que haya grandes bases de datos.

En este trabajo hemos utilizado como variable independiente de entrada al modelo KRR, datos medidos por IASI, los cuales consisten en: Radiancias, Ángulo Vertical Zenital Solar/Satélite, latitud y presión superficial en cada píxel.

Por otro lado, las variables atmosféricas recuperadas con el modelo KRR a partir de los datos de IASI son: Perfiles verticales de temperatura y humedad atmosféricos en 90 niveles de presión, Temperatura del aire a 2 m de la superficie (SAT, *Surface Air Temperature*), Temperatura al nivel del suelo (SKT, *Skin Temperature*) y Humedad del aire a 2 m de la superficie (SdT, *Surface Dew Point*). Así cada *Field of View (FOV)* de IASI, o medida, tiene asociadas 5 variables atmosféricas de salida del modelo KRR. Los perfiles de Temperatura y Humedad han sido divididas en 90 niveles de presión en toda la atmósfera, pero en este trabajo únicamente se muestran los resultados obtenidos desde el nivel 1000 hPa hasta los 100 hPa, ya que es el rango atmosférico de interés para las aplicaciones meteorológicas.

Para comprobar la calidad de las recuperaciones de IASI procesadas con el modelo KRR, hemos comparado dichas recuperaciones con los análisis del ECMWF. En

primer lugar el modelo KRR se ha entrenado con los dos días previos al día que se ha utilizado para testear los resultados de las recuperaciones de las variables atmosféricas de IASI. Así, para entrenar el modelo KRR, se han utilizado, como variables independientes, los datos medidos por IASI de todas las órbitas a nivel global durante los días 13 y 14 de julio de 2015. Y como variable dependiente, de salida, los análisis de ECMWF de los días 13 y 14 de julio de 2015. A continuación se han calculado, utilizando el modelo entrenado, las recuperaciones de IASI para el día 15 de julio de 2015 medidas entre las 9:45Z y las 11:27Z. Finalmente, para validar la precisión de las recuperaciones de IASI, las hemos comparado estadísticamente con los análisis del ECMWF del día 15 de julio de 2015 a las 12Z. Así, se ha hecho una comparación estadística calculando el BIAS (Ecuación 2) y el *Root Mean Square Error* (RMSE) (Ecuación 3) entre las dos bases de datos.

$$BIAS = Y_{test} - Y_{pred} \quad (2)$$

$$RMSE = \sqrt{(Y_{pred} - Y_{test})^2} \quad (3)$$

Donde  $Y_{pred}$  son las 5 variables atmosféricas recuperadas a partir de los datos de IASI el día 15 de julio de 2015 entre las 9:45Z y las 11:27Z.  $Y_{test}$  son las variables atmosféricas de los análisis del ECMWF a las 12Z, utilizados para validar dichas recuperaciones.

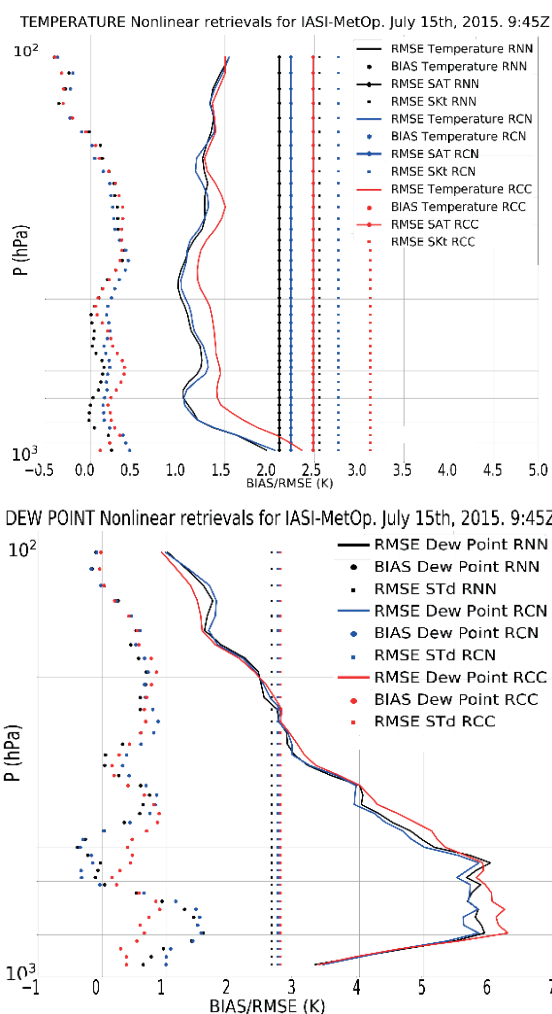
Los resultados muestran una gran similitud entre las recuperaciones de IASI y los datos del ECMWF.

El modelo ha sido testado en FOVs con cielos despajados (recuperaciones sin nubes, RNN), y cielos cubiertos por una fracción de nubes de hasta el 80% (recuperaciones con nubes, RCC). Además se han calculado recuperaciones en cielos sin nubes utilizando los parámetros de normalización de las recuperaciones con fracción de nubes hasta el 80% (RCN).

### 3. RESULTADOS Y DISCUSIÓN

A continuación se muestran los resultados obtenidos en la comparación estadísticas de las 5 variables atmosféricas de interés entre: las recuperaciones de IASI utilizando el modelo KRR entre las 9:45Z y las 11:27Z y los análisis del ECMWF a las 12Z para el 15 de julio de 2015.

Como podemos observar en la Figura 1, los perfiles de RMSE (Ecuación 1) y BIAS (Ecuación 2) calculados para los perfiles verticales de temperatura y humedad del día 15 de julio de 2015 a las 9:45Z, muestran una buena concordancia entre las variables recuperadas a partir de los datos de IASI con el modelo KRR y los análisis del ECMWF utilizados como referencia. Para los RNN, calculados en cielos despejados, la comparación de los perfiles de temperatura entre ambas series de datos, desde los 1000 hPa a los 100 hPa, muestran una RMSE media de 1,24K, mientras que los perfiles de humedad (Temperatura de Rocío), para esa misma capa de la atmósfera muestran un RMSE de 3,87K. Por otro lado, la RMSE media para la SKT es de 2,56, para la SAT es de 2,56 y para la STd es de 2,63K.



**Figura 1.** Comparación entre las recuperaciones de IASI, de las variables atmosféricas de interés, utilizando el modelo KRR a partir de datos del 15 de julio de 2015 entre las 9:45Z y las 11:27Z y los análisis del ECMWF a las 12Z. RMSE (línea continua) y BIAS (línea discontinua) obtenidos para el perfil vertical atmosférico de Temperatura, SAT y SKt (figura de arriba) y para el perfil vertical atmosférico de Humedad y STd (figura de abajo). En rojo se muestra la comparación entre los análisis del ECMWF y RCC. En negro se muestra la comparación entre los análisis del ECMWF y RNN. En azul se muestra la comparación entre los análisis del ECMWF y RCN.

En la Tabla 1 se muestran además las estadísticas para los RCN (recuperaciones en cielos despajados utilizando los parámetros de normalización obtenidos para las recuperaciones en escenas con hasta un 80% de fracción de nubes) y RCC (recuperaciones calculadas en escenas con una fracción de nubes de hasta el 80%).

En todos los casos, las recuperaciones de IASI de las variables atmosféricas de interés estimadas con el modelo KRR muestran gran precisión al compararlas con los análisis del ECMWF, incluso en escenas con gran fracción de nubes, lo que supone un gran avance a la hora de conseguir información sobre la estructura de la atmósfera con sondadores hiperspectrales a bordo de satélites, incluso con cielos nubosos.

**Tabla 1.** RMSE media y BIAS medio entre las recuperaciones de IASI medidos entre las 9:45Z y las 11:27Z, y los análisis del ECMWF a las 12Z para cielos despejados y nubosos el 15 de julio de 2015.

RMSE y BIAS medios para la comparación entre los análisis del ECMWF y RNN

	T (K)	Td (K)	SAT (K)	SKt (K)	STd (K)
RMSE	1.24	3.87	2.11	2.56	2.63
BIAS	0.09	0.47	-	-	-

RMSE y BIAS medios para la comparación entre los análisis del ECMWF y RCN

	T (K)	Td (K)	SAT (K)	SKt (K)	STd (K)
RMSE	1.26	3.83	2.24	2.77	2.73
BIAS	0.16	0.55	-	-	-

RMSE y BIAS medios para la comparación entre los análisis del ECMWF y RCC

	T (K)	Td (K)	SAT (K)	SKt (K)	STd (K)
RMSE	1.46	3.96	2.49	3.13	2.78
BIAS	0.17	0.49	-	-	-

#### 4. CONCLUSIONES

En este trabajo, hemos desarrollado un modelo basado en la técnica KRR para obtener el estado de la atmósfera utilizando únicamente medidas del sondeador hiperspectral IASI basadas en: radiancias, ángulo cenital del satélite y latitudes. Como se puede observar en la Figura 1 y en la Tabla 1, los resultados obtenidos muestran una alta concordancia con los análisis del ECMWF, utilizados en este trabajo para validar las recuperaciones de las variables atmosféricas a partir de las medidas de IASI. Mostrando calidad y precisión para las 5 variables atmosféricas de interés, incluso con cielos con alta nubosidad (cubiertos hasta en un 80% de nubes), lo que proporciona información muy valiosa sobre la estructura de la atmósfera incluso en situaciones convectivas. Este resultado nos muestra el alto potencial del modelo KRR desarrollado en este trabajo y la útil herramienta que suponen estas nuevas recuperaciones de IASI.

Este modelo KRR, junto con otros datos y productos satelitales, proporcionan una potente herramienta al NWC SAF para analizar y monitorizar la atmósfera en situaciones de gran interés para la predicción a corto plazo.

#### 5. BIBLIOGRAFÍA

Aires, F., Rossow, W.B., Scott, N., Chedin, A. 2002. Remote sensing from the infrared atmospheric sounding interferometer instrument 1. Compression, de-noising, and first guess retrieval algorithms. *Journal of Geophysical Research Atmospheres*, 107(D22), 4619.

Arenas-García, J., Petersen, K.B., Camps-Valls, G., Hansen, L.K. 2013. Kernel multivariate analysis framework for supervised subspace learning: a tutorial on linear and kernel multivariate methods. *IEEE Signal Processing Magazine*, 30(4), 16-29.

Blackwell, W.J. 2005. A neural-network technique for the retrieval of atmospheric temperature and moisture profiles from high spectral resolution sounding data. *IEEE Transactions on Geoscience and Remote Sensing*, 43(11), 2535-2546.

Camps-Valls, G., Gómez-Chova, L., Calpe, J., Soria, E., Martín, J.D., Alonso, L., Moreno, J. 2004. Robust support vector method for hyperspectral data classification and knowledge discovery. *IEEE Transactions on Geoscience and Remote Sensing*, 42(7), 1530-1542.

Camps-Valls, G., Bruzzone, L. 2005. Kernel-based methods for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 43(6), 1351-1362.

Camps-Valls, G., Muñoz, J., Gómez-Chova, L., Guanter, L., Calbet, X. 2012. Nonlinear statistical retrieval of atmospheric profiles from metop-IASI and MTG-IRS infrared sounding data. *IEEE Transactions on Geoscience and Remote Sensing*, 50(5), 1759-1769.

Durbha, S., King, R., Younan, N. 2007. Support vector machines regression for retrieval of leaf area index from multiangle imaging spectroradiometer. *Remote Sensing of Environment*, 107(1-2), 348-361.

Fine, S., Scheinberg, K. 2001. Efficient SVM training using low-rank kernel representations. *Journal of Machine Learning Research*, 2, 243-264.

Foody, G.M., Mathur, J. 2004. A relative evaluation of multiclass image classification by support vector machines. *IEEE Transactions on Geoscience and Remote Sensing*, 42(6), 1335-1343.

Hastie, T., Tibshirani, R., Friedman, J.H. 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Second, Springer-Verlag, New York, USA.

Melgani, F., Bruzzone, L. 2004. Classification of Hyperspectral Remote Sensing Images with Support Vector Machines. *IEEE Transactions on Geoscience and Remote Sensing*, 42, 1778-1790.

Muñoz-Marí, J., Plaza, A., Gualtieri, J., Camps-Valls, G. 2009. Parallel programming and applications in grid, P2P and networking systems. In: *Xhafa, F. (Ed.), Parallel Implementation of SVM in Earth Observation Applications*. IOS Press, UK.

Pérez-Suay, A., Amorós-López, J., Gómez-Chova, L., Laparra, V., Muñoz-Marí, J., Camps-Valls, G. 2017. Randomized kernels for large scale Earth observation applications. *Remote Sensing of Environment*, 202, 54-63.

Yang, F., White, M., Michaelis, A., Ichii, K., Hashimoto, H., Votava, P., Zhu, A.X., Nemani, R. 2006. Prediction of continental-scale evapotranspiration by combining MODIS and AmeriFlux data through support vector machine. *IEEE Transactions on Geoscience and Remote Sensing*, 44, 3452-3461