

ENFOQUE PROBABILÍSTICO PARA LAS SITUACIONES DE VIENTO EXTREMO

Macarena RODRIGO ⁽¹⁾⁽²⁾, José Antonio LÓPEZ ⁽²⁾

⁽¹⁾ Consorcio de Compensación de Seguros (CCS), Madrid, España

⁽²⁾ Agencia Estatal de Meteorología (AEMET), Madrid, España

ccseguros@aemet.es, jlopezd@aemet.es

RESUMEN

La Tempestad Ciclónica Atípica o TCA es un fenómeno extremo de la naturaleza en el que están incluidos los tornados y los vientos extraordinarios definidos como aquellos que dan lugar a rachas máximas de viento superiores a 120 km/h. El Consorcio de Compensación de Seguros, CCS, compensa los daños producidos por estos fenómenos y la Agencia Estatal de Meteorología, AEMET, es la encargada de elaborar los informes técnicos que concretan la zona geográfica afectada. En este trabajo se presenta una nueva metodología basada en un enfoque probabilístico utilizando la desviación estándar de la estimación puntual de la técnica operativa. Se busca el percentil de la estimación puntual que corrige el error sistemático de la técnica operativa haciendo que el nuevo estimador sea insesgado para las rachas máximas elevadas.

Palabras clave: Racha Máxima, Tempestad Ciclónica Atípica, Riesgo Extraordinario, Métodos de Interpolación, Percentil, Validación.

ABSTRACT

The Atypical Cyclonic Storm, TCA, is an extreme natural phenomenon in which are included tornadoes and extraordinary winds defined as wind gusts exceeding 120 km/h. The Spanish Insurance Compensation Consortium, CCS, provides insurance coverage against these phenomena and the Spanish Meteorological Agency, AEMET, is responsible for developing the technical reports which specify the geographical area affected. This paper presents a new methodology based on a probabilistic approach using the standard deviation of the point estimation of the operative technique. The percentile of the point estimate is used to correct the systematic error of the operative technique such that the new estimator is unbiased for high maximum gusts.

Key words: Maximum Gust, Atypical Cyclonic Storm, Extraordinary Risk, Interpolation Methodology, Percentile, Validation.

1. INTRODUCCIÓN

La técnica desarrollada en AEMET para la delimitación geográfica de las zonas que han superado la racha de viento de 120 km/h (Consorcio de Compensación de Seguros, 2012) aplica una técnica geoestadística, el kriging universal, que se apoya en los datos de

observación y toma como variables externas la elevación del terreno, la distancia al mar y las salidas del modelo HIRLAM de alcance más corto del campo de racha máxima.

El kriging es una técnica de interpolación geoestadística en la que la estimación en un punto cualquiera es una combinación lineal de las observaciones en el resto de puntos debidamente ponderadas (se busca que el estimador sea insesgado y que presente varianza mínima). En concreto, la técnica del kriging universal es la elegida para las estimaciones de las áreas de rachas máximas de viento, y se basa en un modelo de tendencia lineal en el que entran en juego las variables auxiliares citadas y se obtienen, de esta manera, variaciones suaves en la interpolación (Burrough y McDonnell, 1998).

La estructura del semivariograma usada se especifica “sin efecto pepita” de tal forma que el estimador del kriging es exacto: en los puntos de observación el valor estimado coincide con el valor observado. Esto es importante para el CCS puesto que es necesario respetar al máximo los valores observados ya que son los que tienen mayor fuerza legal (García-Legaz y Valero, 2013).

Mediante una evaluación del procedimiento operativo para los vientos más intensos se ha constatado que, además del error aleatorio inevitable en cualquier técnica de estimación en las zonas sin observación, existe otra componente del error de carácter sistemático.

Esta evaluación se ha realizado mediante la validación de 15 situaciones diferentes de TCA ocurridas en diferentes áreas geográficas y con extensiones y repercusión de diversas dimensiones. En primer lugar, se seleccionan una serie de estaciones de observación que no se tendrán en cuenta en el momento de realizar la interpolación geoestadística, y que posteriormente se utilizarán para comparar el resultado de las estimaciones en estas estaciones con su valor observado. El porcentaje de estaciones seleccionadas es de un 20% en principio aleatoriamente, pero aplicando así mismo alguna condición adicional que nos asegure tanto una cobertura espacial como una presencia de estaciones con valores elevados de racha máxima, pues son los de nuestro interés principal.

Para estos valores elevados de racha máxima se ha obtenido un sesgo negativo en las estimaciones con un valor medio de -13 km/h, esto es, el valor estimado infraestima el observado. Este sesgo negativo cabe explicarlo por el hecho de que el kriging produce un estimador insesgado globalmente pero sobre un rango particular de observaciones acerca el estimador a la media. Como se seleccionó un rango de valores de observación más alto que la media es de esperar un sesgo negativo (Rodrigo y López, 2013).

2. METODOLOGÍA

Para mejorar las estimaciones de racha máxima elevada con el procedimiento operativo se ha desarrollado un método basado en un enfoque probabilístico. Este enfoque ya está de hecho presente de forma implícita en la base de la actual técnica operativa, puesto que los valores estimados en un punto se interpretan como el valor medio o esperado de la función de densidad de probabilidad de la racha máxima de viento estimada en ese punto. Para tener un conocimiento más detallado de la distribución de probabilidad no basta con conocer las medidas de tendencia central sino también es necesario conocer las medidas de dispersión (Cressie, 1991). Por tanto, si se conociera la desviación típica de la distribución del viento en

cada punto se podría dar una estimación basada, no sólo como ahora en la media estimada, sino también en cualquier percentil que se desee.

La herramienta utilizada hasta el momento para obtener las estimaciones de racha máxima mediante el método operativo es el Sistema de Información Geográfica SAGA (System for Automated Geoscientific Analyses). Mediante este software no es posible calcular las medidas de dispersión puesto que sólo realiza el cálculo para el valor esperado, medida de centralidad. Con objeto de obtener la desviación típica de la estimación ha sido necesario programar la técnica operativa, que usa un krigeado universal, en el lenguaje y entorno de programación R que permite extraer tanto las medidas centrales como las de dispersión (R Core Team, 2013).

2. 1. Validaciones

Las validaciones realizadas incluyen un total de 15 TCA diferentes que engloban fenómenos ocurridos en diferentes zonas y con extensiones varias, al igual que ocurre en la selección realizada al realizar la evaluación del procedimiento operativo. Por tanto, son estas mismas situaciones las que se han utilizado en este estudio. Dentro de estas, es necesario destacar aquellas que presentan mayor extensión puesto que son las de mayor impacto económico y en las que están incluidas: Floora (13 al 15 de enero de 2010), Xynthia (27 al 28 de febrero de 2010), TCA 15 al 17 de diciembre de 2011 y Gong (18 al 20 de enero de 2013).

Para estimar de forma precisa el error sistemático de la técnica operativa se han realizado validaciones aleatorias múltiples para este conjunto de situaciones de viento fuerte. La selección de estaciones de validación que permite evaluar los sesgos obtenidos se aplica mediante un programa en R que tiene en cuenta las características que presenta la TCA, (Dalgaard, 2008).

Para las situaciones que presentan mayor extensión, detalladas anteriormente, se ejecutan un total de 50 validaciones por cada TCA. Dado el número elevado de observaciones de racha fuerte (esto es, por encima de 80 km/h) en este tipo de TCA, en cada validación se incluyen 10 de estas observaciones. Además, cada validación se completa con un número de estaciones de observación con valores de racha máxima por debajo de 80 km/h. Este número de estaciones que completa cada validación depende de la proporción existente entre valores altos y bajos de racha máxima en los datos totales, los que incluyen todas las estaciones de observación. En ambos casos las estaciones se seleccionan de manera aleatoria y sin reemplazo. Asimismo se tiene en cuenta la densidad local de estaciones imponiendo que no se repita ninguna provincia dentro de una misma validación. Además se impone la condición de que cada observación por encima de 80 km/h entre al menos una vez en alguna validación, y otra condición que asegura que este tipo de observación no se concentre para que no se perturbe demasiado el campo estimado al suprimirlas.

En cambio, para las TCA de menor extensión, el procedimiento puede ser más exhaustivo. En este caso, en cada validación se elimina una sola estación con racha máxima observada superior a 80 km/h del mismo modo, selección aleatoria y sin reemplazo. De este manera el número de validaciones para cada una de estas situaciones se corresponde exactamente con el número de estaciones de observación que presentan rachas máximas elevadas (en nuestro caso oscila entre un mínimo de 69 y un máximo de 165).

3. EL MÉTODO PROBABILÍSTICO

Esta técnica utiliza el procedimiento operativo programado en R con objeto de obtener no sólo las estimaciones propias de la variable de interés, la racha máxima, sino una estimación de la desviación estándar. Para ello aplica la función “krige” del paquete de R “gstat” que genera tanto las medidas centrales como las de dispersión (Bivand *et al.*, 2008). El semivariograma se ajusta con las funciones de este mismo paquete.

El semivariograma, similar a la covarianza sin centrar, es la función que caracteriza la estructura de momentos de segundo orden mediante un solo argumento (espacial). Para la técnica geoestadística considerada se impone la estacionariedad, es decir, invarianza de los momentos de segundo orden frente a desplazamientos. Se trabaja con los residuos y se estima la matriz de covarianzas (o semivariograma) trabajando con estos residuos. Por otro lado, en el semivariograma se utiliza no sólo la variable de racha máxima, tal y como permite SAGA, sino que se incluyen todas las variables externas.

La técnica probabilística se basa en la descripción de la distribución de probabilidad mediante los parámetros estimados de dicha distribución, valor medio y desviación estándar, buscando los percentiles que corrijan el sesgo negativo obtenido en las estimaciones correspondientes a las observaciones con velocidades de racha máxima superiores a 80 km/h.

Las validaciones efectuadas inicialmente para la técnica operativa utilizando R y para los valores de racha máxima elevadas muestran un sesgo sistemático negativo. Este sesgo varía según el tipo de TCA, para las más extensas se presenta un sesgo menor.

Por otro lado, se observa cómo a medida que tomamos percentiles mayores se produce una clara mejora en el sesgo, para el caso de las rachas máximas elevadas, hasta que llega a hacerse nulo. Como cabría esperar, la anulación del sesgo depende del tipo de TCA, de manera que las de mayor extensión se anulan a percentiles menores y viceversa.

3. 1 Calibración del método

Fecha TCA	Percentil	Probabilidad	Extensión TCA
20100113a15	0,42	66%	229.338
20100227a28	0,43	67%	310.622
20111023a27	0,92	82%	60.705
20111112a14	0,86	80%	29.578
20111215a17	0,62	73%	148.913
20120105a08	0,80	79%	30.716
20120202a05	0,85	80%	80.420
20120206a08	0,78	78%	70.906
20120415a17	1,37	91%	31.316
20120423a26	1,10	86%	55.365
20121027a28	0,75	77%	31.165
20121124a25	0,84	80%	31.380
20121213a16	0,93	82%	81.492
20130118a20	0,47	68%	201.691
20130123a24	0,87	81%	71.162

Tabla 1: RESULTADOS DE LAS VALIDACIONES PARA CADA TCA

En la Tabla 1 se presentan los resultados, para cada una de las TCA, de las validaciones descritas anteriormente. La fecha de la TCA viene dada en el formato año, mes y período de días que comprende. En la siguiente columna se detallan los percentiles para los que el sesgo se anula teniendo en cuenta los valores elevados de racha máxima superior a 80 km/h, así como la probabilidad a la que corresponden. Por último, para dar una idea general de la extensión de la TCA se han contado los puntos de grid (1km x 1km) que presentan valores de racha máxima superiores a 80 km/h en las estimaciones realizadas con el método operativo utilizando R, que se reflejan en la última columna.

Los anteriores resultados se representan en la Fig. 1, en la que se muestra la dependencia entre los percentiles obtenidos para cada una de las TCA y su extensión. Para facilidad de interpretación las TCA grandes y normales se diferencian por el color. Como se observa, aunque las dos poblaciones por extensión de TCA se separan bien por los percentiles, no hay una forma simple de dependencia general del percentil con la extensión (especialmente para las TCA de extensión menor). Por ello se ha optado por definir dos percentiles correctores, uno para cada tipo de TCA según extensión.

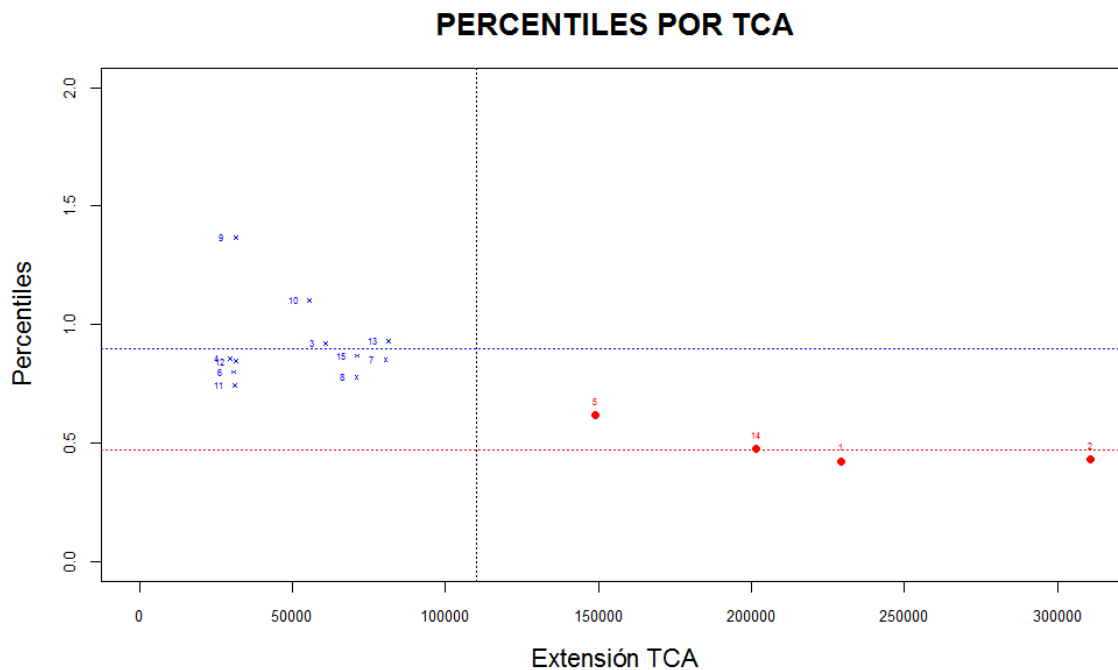


Fig. 1: Dependencia entre el percentil y la extensión para cada una de las TCA

Para las TCA más pequeñas, representadas mediante cruces en azul y que están dispuestas como una nube de puntos, el percentil que se aplica es de 0,9 (línea horizontal discontinua azul), resultado de la media de los percentiles ponderada con las extensiones para esta población de TCA. El criterio para definir esta población de TCA es que su extensión (tal como se definió anteriormente en términos de puntos de rejilla) sea inferior a 120.000 (línea vertical discontinua negra).

Por otro lado, para las TCA de mayor extensión (es decir, con extensión por encima de 120.000), representadas en el gráfico mediante puntos en rojo, el percentil que resulta de la media ponderada es 0,47. De esta forma se define completamente la forma de aplicar el método probabilístico para una TCA cualquiera. Estos resultados sumarios se reflejan en la Tabla 2.

Tipo TCA	Extensión	Percentil	Probabilidad
Pequeña	< 120.000	0,90	82%
Grande	≥ 120.000	0,47	68%

Tabla 2: NORMA DE APLICACIÓN MÉTODO PROBABILÍSTICO

3.2 Comparación del método probabilístico con el operativo

Una vez calibrado el método probabilístico se procedió a valorar las diferencias que presenta con relación al método operativo, utilizando las mismas 15 TCA. Las variables utilizadas para resumir estas diferencias son:

- a) extensión de rachas máximas superiores a 120 km/h (como antes en términos de número de puntos de rejilla)
- b) número de municipios afectados según la condición anterior
- c) número de habitantes de los municipios resultantes.

Hay que aclarar que en la condición b) anterior un municipio se considera afectado si algún punto de rejilla afectado pertenece al municipio. Como puede apreciarse, la primera variable tiene un estricto carácter fisiográfico, mientras que las otras dos permiten aproximar el impacto económico de las TCA. La decisión de tomar estas dos variables se debe a que el CCS aplica la ampliación de las zonas a municipios para cubrir los daños ocasionados según su reglamento.

En la Tabla 3 se muestra un resumen de este análisis en el que se calculan estas diferencias teniendo en cuenta todas las situaciones. En cuanto al área que delimita las rachas máximas mayores o iguales a 120 km/h el aumento es del 83%. Al extender estas zonas a municipios el aumento es algo menor, 75%, pero todavía un valor considerablemente alto. Por último, para la variable más representativa de los costes que podría ocasionar una situación de TCA, la población, se obtiene un aumento del 62% del método probabilístico respecto al operativo.

Área	Número municipios	Población municipios
83%	75%	62%

Tabla 3: COMPARATIVA MÉTODOS PROBABILÍSTICO vs. OPERATIVO (% DE AUMENTO DEL MÉTODO PROBABILÍSTICO SOBRE EL OPERATIVO)

A continuación, en las Fig. 2 y Fig. 3, se muestran los mapas resultantes de aplicar el método probabilístico y compararlo con el método operativo. En los ejemplos se muestra una de las TCA de mayor extensión, aplicando el percentil 0,47 en el método probabilístico, y otra TCA pequeña, aplicando el percentil 0,9.

En ambos casos los mapas corresponden a la extensión de municipios que se obtiene por ambos métodos. En rojo están marcados los municipios que presentan en alguno de sus puntos unas estimaciones de racha máxima superiores a 120 km/h mediante el procedimiento operativo implementado en R, y en azul los municipios que, además de los anteriores, se obtienen aplicando el método probabilístico.

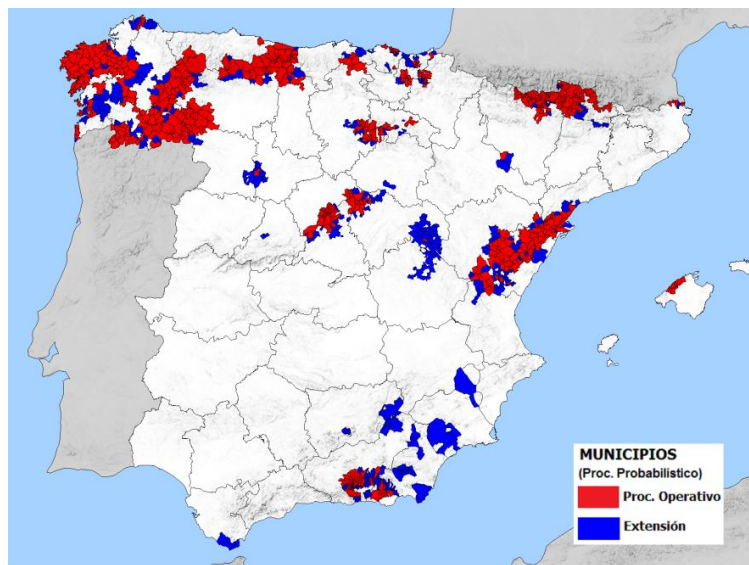


Fig. 2: Resultados de la aplicación de los métodos probabilístico y operativo a TCA grande: 13 al 15 de enero de 2010 (Floora).

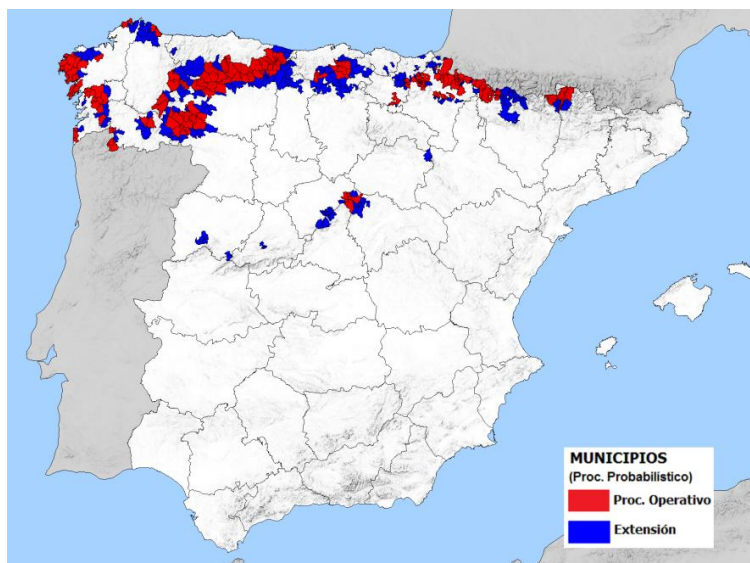


Fig. 3: Resultados de la aplicación de los métodos probabilístico y operativo a TCA pequeña: 23 al 27 de noviembre de 2011.

4. CONCLUSIONES

La técnica probabilística para corregir el sesgo de las rachas fuertes que se ha presentado tiene una base teórica sencilla. Utiliza la misma técnica geoestadística que la empleada en la técnica operativa, el krigeado universal, utilizando la desviación típica estimada por este para pasar a un percentil en la estimación. Una consideración importante es que, dado que el estimador del krigeado universal con semivariograma sin efecto pepita es exacto, esto es, en los puntos con observación coincide con la observación, y en esos puntos la desviación típica es nula, el estimador de la técnica probabilística sigue siendo exacto. De esta manera, se respetan al máximo los valores observados, condición necesaria para el CCS ya que estos son los que tienen mayor fuerza legal.

Por otro lado, se garantiza la compatibilidad con la técnica actual en el sentido de que las zonas delimitadas con el procedimiento propuesto engloban a las zonas delimitadas con el método operativo.

Además, esta técnica probabilística permite regular fácilmente la extensión de las áreas estimadas de racha máxima para valores altos. En este estudio se han calculado los percentiles más adecuados para aplicar según las características de la TCA, sin embargo, podrían obtenerse diversos escenarios, asociados a diversos percentiles, con gran facilidad para situaciones más complejas, presentando la ventaja, como acabamos de señalar, de que los percentiles mayores siempre incluyen las áreas dadas por los percentiles menores.

Para hallar el percentil corrector se ha estimado el sesgo de la técnica operativa para rachas máximas elevadas (superiores a 80 km/h) mediante múltiples validaciones aleatorias para cada una de varias situaciones de TCA de diferentes características. En las validaciones se ha tenido en cuenta la importante heterogeneidad en la densidad espacial de los puntos de observación. Se ha encontrado que el percentil corrector depende claramente de la magnitud de la TCA y se ha establecido una norma general de aplicación: para las TCA pequeñas el percentil corrector es de un 0,9 y para las TCA grandes se fija en 0,47.

Por último, hay que señalar que esta técnica se ha programado en el lenguaje y entorno de programación R, que permite una total automatización.

5. AGRADECIMIENTOS

Este trabajo surge como una colaboración entre la Agencia Estatal de Meteorología y el Consorcio de Compensación de Seguros, los cuales, junto a AT Sistemas, patrocinan el proyecto. Los autores desean agradecer el apoyo de muchas personas que pertenecen a estas organizaciones y que hacen posible estas investigaciones.

6. REFERENCIAS

Bivand, R.S.; Pebesma, E.J. y Gómez-Rubio, V. (2008). *Applied Spatial Data Analysis with R*. Springer.

Burrough, P.A. y McDonnell, R.A. (1998). *Principles of Geographical Information Systems*. Oxford University Press.

Consortio de Compensación de Seguros (2012). *Recopilación Legislativa. Febrero 2012*. Ministerio de Economía y Competitividad.

Cressie, N. (1991). *Statistics for spatial data*. A Wiley – Interscience Publication.

Dalgaard, P. (2008). *Introductory Statistics with R*. Springer.

López, J.A. y Rodrigo, M. (2013). Desarrollo de una técnica para la delimitación de zonas de racha de viento máxima alta en situaciones de viento extremos. En: García-Legaz, C. y Valero, F. (Eds.). *Fenómenos meteorológicos adversos en España*, AMV Ediciones, pp. 81-96.

Rodrigo, M. y López, J.A. (2013). “Mejora del procedimiento operativo de AEMET para la estimación de las áreas con rachas máximas de viento”. *Tethys*, 10, pp. 35-44.

R Core Team (2013). R Foundation for Statistical Computing: <http://www.R-project.org>.