

ORIGINAL

TRATAMIENTO Y ESTUDIO DE SERIES DE TEMPERATURA PARA SU APLICACIÓN EN SALUD PÚBLICA. EL CASO DE CASTILLA- LA MANCHA (*)

Isidro J. Mirón Pérez (1), Juan Carlos Montero Rubio (1), Juan José Criado- Álvarez (1), Gonzalo Gutiérrez Ávila (1), Daniel Paredes Beato (2), Sheila Mayoral Arenas (3) y Cristina Linares Gil (4).

(1) Consejería de Sanidad, Junta de Comunidades de Castilla- La Mancha.

(2) Dpto. Física de la Tierra II, Facultad de CC. Físicas Universidad Complutense de Madrid.

(3) Sociedad Castellanomanchega de Medicina Preventiva y Salud Pública.

(4) Asesora de la Fundación General de la Universidad Autónoma de Madrid para el Dpto. de Educación para el Desarrollo Sostenible del Ayuntamiento de Madrid.

(*) Este trabajo se ha financiado con una ayuda a la Sociedad Castellanomanchega de Medicina Preventiva y Salud Pública de la Consejería de Sanidad de la Junta de Comunidades de Castilla- La Mancha (Expediente PI 22/2003).

RESUMEN

Fundamento: Numerosos trabajos relacionan variables atmosféricas con indicadores sanitarios. En regiones extensas, como Castilla- La Mancha, puede ser necesario dividirla en áreas en función de las variables atmosféricas disponibles, eligiendo una estación meteorológica representativa para cada zona. El objetivo de este artículo es analizar los datos diarios de temperaturas de numerosos observatorios de Castilla- La Mancha y su reducción a unas pocas estaciones representativas para ser utilizadas en estudios que relacionen variables atmosféricas con indicadores sanitarios de esta región.

Métodos: Se seleccionaron estaciones meteorológicas de Castilla- La Mancha en función del número de años disponibles y de datos perdidos. Tras rellenar las lagunas de las series elegidas, detectar posibles discontinuidades y homogeneizar las series, los datos diarios de temperaturas se utilizan en análisis de conglomerados jerárquico y factorial mediante componentes principales.

Resultados: El análisis factorial extrae un solo factor utilizando las series de temperaturas máximas, medias o mínimas. En las máximas, ese factor explica el 93,45% de la varianza, con autovalor 39,249. La estación Toledo «Compuesta» tiene coeficientes de correlación en la matriz de componentes principales de 0,987; 0,991 y 0,981 para las series de temperaturas máximas, medias y mínimas respectivamente.

ABSTRACT

Treatment and Temperature Series Study for Use in Public Health. The Case of Castilla-La Mancha, Spain

Background: Numerous articles relate atmospheric variables to health indicators. In large regions, such as Castilla-La Mancha, it may be necessary to divide the region into areas in terms of the atmospheric variables available by selecting a representative weather station for each zone. This article focuses on analyzing the daily temperature data from numerous Castilla La Mancha observatories and reducing the number thereof to a few representative stations for being used in studies relating atmospheric variables to health indicators in this region.

Methods: Castilla-La Mancha weather stations were selected in terms of the number of years available and missing data. After filling in the gaps in the selected series, to detect any possible discontinuities and to homogenize the series, the daily temperature data is used in hierarchical cluster and factorial analyses by principal components.

Results: Factorial analyses extract one single factor by using the maximum, mean or minimum temperature series. For the maximum temperatures, this factor explains 93.45% of the variance, with an eigenvalue of 39.249. The «Compuesta» station in Toledo shows correlation coefficients in the principal components matrix of 0.987, 0.991 and 0.981 respectively for the maximum, mean and minimum temperature series.

Correspondencia:

Juan Carlos Montero Rubio

Sección de Microbiología Clínica y Ambiental

Servicio de Laboratorio

Instituto de Ciencias de la Salud de Castilla- La Mancha

Carretera de Extremadura km 114

45600 Talavera de la Reina (Toledo)

Correo electrónico: jcmonte@jccm.es

Conclusiones: Castilla-La Mancha es una región isoclimática en función de la temperatura y la estación Toledo «Compuesta» la elegida como representativa regional para estudios en salud pública. Los resultados permiten la realización de estudios desagregados en unidades menores como las provincias, con las estaciones de las capitales administrativas como referencia.

Palabras clave: Temperatura ambiental. Mortalidad. Series temporales. Análisis por conglomerados.

Conclusions: Castilla-La Mancha is an isoclimatic region in terms of the temperature, the «Compuesta» station in Toledo being selected as the representative station for the region for public health studies. The results afford the possibility of conducting studies broken down into small units such as the provinces, with the stations in the government capitals as a reference.

Key words: Temperature. Mortality. Clusters analysis. Principal components analysis.

INTRODUCCIÓN

Numerosos estudios relacionan variables atmosféricas con diversos indicadores sanitarios, la mayor parte referidos a ciudades o a zonas geográficas reducidas. La influencia de la temperatura sobre la mortalidad se ha observado en numerosos países¹⁻³, incluyendo España^{4,5,6}, así como el papel de otras variables como la humedad⁷, velocidad del viento⁸ o el efecto conjunto temperatura-viento (wind-chill)⁹ sobre la mortalidad general o causas específicas de mortalidad, e incluso la relación de las horas de luz con la prevalencia de enfermedades psiquiátricas¹⁰ o la estacionalidad de las afecciones afectivas¹¹. La sobremortalidad causada por la ola de calor del verano de 2003^{12,13} puso de actualidad ante la opinión pública este tipo de estudios y sirvió para implementar medidas preventivas ante pronósticos de situaciones meteorológicas similares. Sin embargo, la eficacia de las mismas va a depender de condicionantes que exigen el tratamiento previo de los datos meteorológicos suministrados a fin de cumplir con unos criterios razonables de representatividad. Un problema de este tipo se plantea en el caso de estudios referidos a ámbitos geográficos relativamente extensos u orográficamente diversos, con posibles variaciones termométricas significativas que aconsejan la utilización de varias estaciones meteorológicas para analizar la pertinencia, o no, de dividir la región considerada en áreas, conglomerados o clusters, eligiéndose una estación de referencia para cada uno de los conglomerados resultantes. Por otro lado, para todo estudio climatológico las series de variables meteorológicas a analizar deben abarcar un mínimo de 30 años, según la Organización Meteorológica

Mundial (OMM)¹⁴, lo que complica notablemente el proceso ya que las series de datos diarios suelen contener lagunas (días en los que no se registran datos) más o menos extensas, incluso años, que hay que rellenar con criterios científicos consistentes.

Existen herramientas disponibles ampliamente utilizadas en climatología para salvar todos estos inconvenientes a fin de contar finalmente con datos meteorológicos fidedignos y representativos en la realización de estudios ecológicos en salud pública, especialmente si se han de utilizar para establecer medidas preventivas o alertas fiables a partir, por ejemplo, de cierto nivel de temperatura ambiental.

Por consiguiente, el objetivo de este trabajo es el análisis y tratamiento de los datos de temperatura de un amplio número de estaciones de Castilla-La Mancha para su reducción a una cantidad de estaciones representativas a efectos de su utilización en estudios que relacionen variables atmosféricas con indicadores sanitarios, como la mortalidad, en este ámbito territorial.

MATERIAL Y MÉTODOS

Variables de estudio y estaciones seleccionadas

Los datos diarios de temperaturas máximas, medias y mínimas fueron suministrados por el Instituto Nacional de Meteorología (INM). De 329 estaciones meteorológicas preseleccionadas dentro del territorio de Castilla-La Mancha se escogieron aquellas

que contenían series de mediciones diarias de temperatura entre el 1 de enero de 1974 y el 31 de diciembre de 2003 como mínimo (30 años), pérdidas de datos o lagunas incluidas, cumpliendo estas condiciones 69 observatorios. Las estaciones Guadalajara «Serranillo» y Toledo «Buenavista» son continuación de Guadalajara «Instituto» y Toledo «Lorenzana» respectivamente, si bien mediando un cambio de ubicación en ambas aunque a localizaciones próximas dentro de las respectivas ciudades. La unión de las mismas las denominaremos Guadalajara «Compuesta» y Toledo «Compuesta».

Relleno de lagunas

Puesto que solo 6 estaciones (Albacete «Los Llanos Base Aérea», Ciudad Real «Escuela de Magisterio», Molina de Aragón, Salto de Zorita, Toledo «Lorenzana» y Toledo «Buenavista») no presentan pérdidas en sus datos diarios, se hace necesario completarlos en las series que sí las contienen, de forma que no altere las características propias de las mismas y tratar con la mayor certeza posible de que la variabilidad de las series se deba solamente a factores meteorológicos, sobre todo si se pretende estudiar episodios de valores extremos de temperatura, como es el caso de las olas de calor y de frío, en las que es preciso estimar la temperatura de cada día perdido de la forma más aproximada posible al valor original. Para ello son más adecuados los métodos basados en la interpolación de datos de las estaciones

cercanas^{15,16} que los que se basan en funciones probabilísticas a partir de los datos que se posea de la propia serie¹⁷. Sin embargo, han de establecerse criterios mínimos respecto a las series de temperaturas a seleccionar para no alterar significativamente las características originales de las series que posteriormente se van a rellenar. Así, se incluirán en el estudio las que tengan menos de 40 meses perdidos (considerando como tales aquellos que presenten más de 15 días consecutivos sin valores) y en las que el número de días sin datos sea menor del 6,5% del total, siguiendo el esquema de Eischeid et al.¹⁸. Son 44 las estaciones en principio seleccionadas (13,3% de las 329 originales). Puesto que Guadalajara «Compuesta» y Toledo «Compuesta» son, como se ha señalado, unión de dos estaciones, será finalmente 42 el número de estaciones a considerar.

Para el relleno de lagunas en las series de temperaturas se escogen como máximo 4 estaciones de referencia para cada serie a completar siempre que estén a menos de 300 Km de distancia de la misma y tengan una correlación con ella mayor de 0,35 (nivel de significación de un 95%)¹⁹. El siguiente paso consiste en la descomposición de cada serie original y de sus respectivas series referencia en series mensuales con su posterior normalización (restando a cada dato el promedio total y dividiéndolo por su desviación estándar). De este modo, para cada mes de una serie primitiva con datos perdidos hay un máximo de 4 series de referencia alternativas para completar esos datos. Para ello, se utilizan 5 métodos de interpolación (Tabla 1)^{20,21}

Tabla 1

Métodos de interpolación utilizados para el relleno de lagunas

1. Método de la razón normalizada.
2. Método del inverso de la distancia.
3. Regresión múltiple.
4. Mejor estimador.
5. Media de los anteriores

de manera que se escoge como serie de corrección aquella que, entre todos los métodos de interpolación testados, presente mayor correlación con la serie original en ese mes.

Detección de discontinuidades y homogenización de las series de temperatura

Una vez subsanadas las lagunas que existían en las series originales de temperaturas, es preciso asegurarse de que la variabilidad de estas series de temperaturas rellenas obedece a factores meteorológicos y climatológicos y no a la aplicación de la técnica relatada. Es decir, hay que tratar de detectar posibles inhomogeneidades o discontinuidades que también pueden ser debidas a, por ejemplo, un cambio de ubicación de la estación o del instrumento de medida.

Puesto que no se tiene información completa de la historia de las estaciones, han de utilizarse métodos que no parten de posibles puntos de discontinuidad^{22,23}, basados en la detección de variaciones significativas en la media de las series temporales entre un año y otro, siendo necesario contar con una serie de referencia homogénea que represente la variabilidad climática regional de la serie a testar a fin de poder compararla con ella. Para conseguir esta serie de referencia se parte de las series de las estaciones más próximas.

Por tanto, en primer lugar han de crearse tantas series de medias anuales como series de temperaturas (máxima, mínima y media) y estaciones tenemos. A continuación se seleccionan las cinco estaciones cercanas con mayor coeficiente de correlación respecto a cada serie en estudio²⁴, número suficiente para que una discontinuidad en las series no aparezca como significativa en la serie final^{25,26}.

Una vez realizada la selección de las estaciones queda la construcción de una serie de

referencia para cada serie original de temperatura (tres por estación: máximas, mínimas y medias) a partir del último valor de la serie original, calculando el resto de valores hacia atrás a fin de asegurar la coincidencia en el valor más actual y que no suponga discontinuidad. La interpolación se efectúa mediante el método de la media ponderada por el cuadrado del coeficiente de correlación²⁴ utilizando las 5 series de diferencias seleccionadas. Puede afirmarse así que cada una de las series de referencia ha quedado libre de cualquier inhomogeneidad derivada de las series de las estaciones cercanas utilizadas en el proceso.

La detección de discontinuidades en las series temporales de temperaturas se realiza en dos etapas²⁷. La primera, a partir de un modelo de regresión en dos fases y, posteriormente, la utilización de varios tests estadísticos que permitan asegurar que el punto de discontinuidad identificado no sea debido a factores climáticos. El modelo de regresión en dos fases para series climáticas²⁸ se basa en la creación de series de diferencias entre las series de medias anuales y sus series de referencia. Se ajusta la serie de diferencias mediante una regresión lineal simple y se calcula la suma de los cuadrados de los residuos. Para cada punto de la serie se divide ésta en dos partes, calculándose igualmente la suma del cuadrado de los residuos para cada parte de la serie y sumando los valores de ambas partes. Así, sucesivamente, para todos los puntos de la serie hasta obtener una serie de valores límite. Cada punto de la serie con valor inferior al límite correspondiente será considerado como potencial punto de discontinuidad, siempre que sea estadísticamente significativo. Para evitar la detección de puntos ficticios de discontinuidad³⁵, se realizan, consecutivamente, un test de diferencia de medias entre ambas partes de la serie y el método de multirrespuesta a permutación (MMP)^{29,30}. Los puntos que superen ambas pruebas estadísticas serán considerados discontinuidades.

El siguiente paso en la preparación de las series de temperaturas sería la homogeneización o reconstrucción de las series en las que se han detectado inhomogeneidades. Esta se realiza siempre desde el final de la serie (31 de diciembre de 2003), manteniendo los datos hasta la primera discontinuidad. Desde este punto hasta el principio de la serie (1 de enero de 1974) ha de eliminarse el error que introduce la discontinuidad calculándose el valor de la diferencia de medias entre las dos partes de la serie. Ese valor se sumará o restará, según su signo, a la primera parte de la serie, es decir, desde el inicio de la misma hasta el punto de discontinuidad.

División de Castilla-La Mancha en áreas homogéneas respecto a la temperatura:

La potencial división de la Comunidad Castellanomanchega se estudia mediante el análisis de conglomerados jerárquicos y su posterior validación por análisis factorial.

El análisis de conglomerados jerárquico va creando grupos de variables (series de temperaturas) en función de la distancia (medida del grado de similitud) que se encuentre entre unas y otras, partiendo de grupos individuales hasta obtener un único grupo o conglomerado compuesto por todas las variables objeto del análisis. Este proceso permite comprobar la distancia a la que se van fundiendo elementos o grupos en cada etapa del mismo, lo que supone valorar la heterogeneidad de los conglomerados que se van uniendo en el análisis y decidir en qué paso la fusión de elementos la incrementa excesivamente. El análisis factorial mediante componentes principales rotadas es una técnica habitualmente utilizada en climatología^{31,32} basada en que toda la información de las diferentes series se encuentra contenida en su varianza, pudiéndose distinguir áreas o regiones con una variabilidad común, afectadas por las mismas causas³³. Se obtienen para ello los autovalores de una matriz de varianza-covarianza que expresa la varia-

bilidad conjunta de todas las variables, de forma que las primeras componentes explican la mayor parte de la varianza de las series de temperaturas, la cual es directamente proporcional a su autovalor. La solución factorial propiamente dicha se encuentra en la matriz de componentes principales, en la que aparecen los componentes con autovalores superiores a 1, siguiendo el criterio de Kaiser³⁴, y que contiene las correlaciones entre las variables originales (o saturaciones) y cada uno de los factores extraídos. En el caso de extraerse dos ó más componentes, para facilitar la interpretación y estimación de las puntuaciones de los factores, se efectuaría la rotación de la matriz de componentes principales por el método Varimax, que es el más usado en climatología^{35,36}.

RESULTADOS

Las series de referencia utilizadas (4 series de corrección en cada serie a completar) en la interpolación de datos para el relleno de lagunas en las series de temperaturas máximas (Tmax), medias (Tmed) y mínimas (Tmin) presentan unos coeficientes de correlación entre 0,46 y 0,98. El método de interpolación más utilizado es la regresión múltiple. Las series originales con mayor porcentaje de datos perdidos son las correspondientes a Guadalajara «Compuesta», con un 13,2 %. Las lagunas de éstas se completan con datos de series con correlaciones entre 0,78 y 0,93 con las originales, e intervienen como series de corrección con coeficientes de correlación a partir de 0,60.

Se detectan años de discontinuidad en 28, 29, y 31 de las series corregidas de temperaturas máximas, medias y mínimas, respectivamente, por lo que se procedió a su homogeneización. Los dos únicos observatorios en los que consta un cambio de ubicación son Toledo «Compuesta» (en 1982) y Guadalajara «Compuesta» (en 1985), pudiéndose apreciar que las discontinuidades se locali-

Tabla 2

Matriz de componentes principales de temperaturas máximas (Tmax), medias (Tmed) y mínimas (Tmin)

Provincia	Estaciones	Tmax	Tmed	Tmin	
Albacete	Embalse Fuensanta	0,952	0,967	0,939	
	Los Llanos Base Aérea	0,982	0,987	0,973	
	Molinicos	0,951	0,959	0,920	
	Munera	0,973	0,980	0,961	
	Socovos	0,943	0,964	0,944	
Ciudad Real	Abenojar	0,873	0,920	0,894	
	Alcazar de San Juan	0,985	0,991	0,981	
	Almaden Minas	0,967	0,972	0,946	
	Argamasilla de Alba	0,963	0,981	0,970	
	Calzada de Calatrava COOP	0,981	0,977	0,928	
	Ciudad Real Escuela Mag.	0,988	0,990	0,979	
	Cortijo Arriba	0,975	0,984	0,967	
	Fuencillana	0,984	0,987	0,977	
	Pantano Torre Abraham	0,976	0,982	0,957	
	Tomelloso	0,979	0,987	0,981	
	Torre Juan Abad	0,969	0,976	0,945	
	Torrenueva	0,984	0,869	0,968	
	Cuenca	Abia Obispalia	0,963	0,963	0,943
		Arrancacepas	0,958	0,975	0,960
Cañizares Central Vadillo		0,952	0,973	0,843	
Cuenca		0,986	0,990	0,977	
Honrubia		0,978	0,981	0,967	
Motilla del Palancar		0,976	0,981	0,965	
Picazo la Varga		0,977	0,981	0,953	
Guadalajara	Villares Saz	0,961	0,963	0,932	
	Aranzueque	0,988	0,984	0,899	
	Guadalajara Compuesta	0,985	0,976	0,958	
	Molina de Aragón	0,974	0,977	0,845	
	Pantano Vado	0,943	0,966	0,944	
	Salto Zorita	0,980	0,983	0,948	
	Valdelcubo	0,959	0,969	0,917	
Toledo	Belvis de la Jara	0,954	0,956	0,937	
	Cabezamesada	0,921	0,983	0,964	
	Carpio de Tajo	0,881	0,984	0,965	
	Cazalegas	0,971	0,978	0,945	
	Mora	0,983	0,968	0,939	
	Navalucillos	0,974	0,974	0,945	
	Puebla de Almoradiel	0,982	0,990	0,976	
	Santa Olalla	0,982	0,982	0,950	
	Toledo Compuesta	0,987	0,991	0,981	
	Ventas de Retamosa	0,980	0,968	0,941	
Villarejo Montalban	0,968	0,981	0,966		

zan para el primero en 1991 (Tmax), 1995 (Tmed) y 1981 (Tmin), y en 1984 (Tmed) y 1985 (Tmin) para el segundo. Solamente las estaciones de Abenójar, Almadén «Minas», Cuenca, Molina de Aragón y Cabezamesada no presentan punto de discontinuidad en ninguno de los 30 años analizados y en ninguna de sus series de temperaturas (Tmax, Tmed y Tmin). De ellas, únicamente las de Molina de Aragón no contienen datos perdidos. En ninguna serie de temperatura se han detectado inhomogeneidades en más de un año.

El análisis de conglomerados jerárquicos no nos indica diferencias claras entre los grupos de series de temperatura que se van formando, aunque mucho más evidente y explicativo resulta el análisis factorial mediante componentes principales, en el que se extrae un solo componente o conglomerado para cada categoría de series de temperaturas, según muestran sus valores de varianza total explicada. Sirva como ejemplo la correspondiente a la categoría de series de temperaturas máximas (Tmax) ya que los resultados son similares para las otras dos categorías (Tmed y Tmin), observándose que sólo un factor explica el 93,450 % de la varianza, con el único autovalor superior a 1 (39,249). El siguiente factor presenta un autovalor de 0,293 y solamente sumaría un 0,697 % al total de la varianza explicada.

La matriz de componentes principales (Tabla 2) exhibe las correlaciones existentes entre las series de temperaturas y el único factor extraído. La menor correlación para Tmax la ofrece la estación de Abenójar (0,873) y la mayor Aranzueque (0,988), siendo Torrenueva (0,869) y Toledo «Compuesta» (0,991, junto con Alcázar de San Juan), respectivamente, para Tmed, y Cañizares Central Vadillo (0,843) y Toledo «Compuesta» (0,981, igual que Alcázar de San Juan y Tomelloso) para la categoría Tmin. De forma conjunta, si se calculara un coeficiente medio de correlación máxima sería la estación Toledo «Compuesta» la que mayor valor alcanzaría (0,986). La otra esta-

ción con cambio de emplazamiento en su secuencia temporal, Guadalajara «Compuesta», tiene coeficientes de correlación con el único componente del modelo de 0,985, 0,976 y 0,958 para Tmax, Tmed y Tmin, respectivamente.

La matriz de componentes principales rotados no se calcula cuando en el análisis se extrae un solo componente, como es el caso.

DISCUSIÓN

El análisis de componentes principales es concluyente en cuanto al número de conglomerados o clústers que extrae el modelo: uno. El autovalor del único factor seleccionado para las series Tmax (39,249), el porcentaje de varianza que explica (93,450 %), los autovalores inferiores a 1 y la insignificante varianza adicional explicada si se consideran más conglomerados dejan poco margen a otra interpretación, teniendo en cuenta que los resultados son similares si el análisis se realiza con las series de temperaturas medias (Tmed) o mínimas (Tmin). Por tanto, no se hallan diferencias significativas entre las 42 estaciones con datos termométricos seleccionadas de Castilla-La Mancha respecto a las series de temperaturas Tmax, Tmed y Tmin. Es decir, la podemos considerar como una Región isoclimática en función de la temperatura, como apunta algún estudio ya realizado con 45 observatorios distribuidos por toda España (3 en Castilla-La Mancha)¹⁹. El hecho de ser todas provincias interiores, sin influencia marítima, puede favorecer el resultado. Según esto, podrían utilizarse los datos de cualquiera de sus estaciones para estudiar la relación de las temperaturas con la morbi-mortalidad o con cualquier otro indicador sanitario regional. Sin embargo, parece lógico que de elegir una sola estación representativa de la Región ésta sería Toledo «Compuesta», ya que es la que presenta un mayor coeficiente de correlación con el factor obtenido si se consideran las tres categorías de temperatura (Tmax,

Tmed y Tmin) de forma conjunta, coincidiendo, además, la condición de Toledo como capital administrativa de Castilla-La Mancha.

Este análisis puede permitirnos realizar estudios desagregados, por ejemplo, por provincias tomando como estación de referencia la de la capital correspondiente, por áreas de salud o por cualquier otra unidad administrativa que se consideren oportuna. También se podría dividir en función de alguna característica de interés: actividad económica principal, rural vs urbana, pirámide poblacional, etc.

Una metodología como la que aquí se propone viene a superar las limitaciones encontradas en estudios que relacionan variables ambientales con la mortalidad y que solamente podían incluir ciudades, con su población situada en un espacio geográfico lo suficientemente pequeño como para suponer a todos los individuos sometidos a iguales condiciones climáticas y un tamaño de población tal que lleve asociado un número de defunciones suficiente para que el análisis estadístico reflejara resultados significativos, sobre todo si se pretendía desagregar según sexo, edad o causas específicas de mortalidad. Proyectos como los del Euro-winter Group³⁷ en Europa o los llevados a cabo por Schwartz^{38,39} o Curriero³ en Estados Unidos de América estudian diferencias regionales dentro de su área geográfica de trabajo analizando exclusivamente ciudades con gran número de habitantes que presenten características socioeconómicas parecidas. En Castilla-La Mancha este tipo de planteamiento impediría el análisis ya que sus poblaciones son pequeñas y dispersas. Sin embargo, existen estudios que estiman la sobremortalidad producida por la temperatura en un país completo, como Holanda¹, o en otra Comunidad Autónoma como Madrid^{7,8}, en los que se utiliza una sola estación meteorológica de referencia sin una base metodológica sólida que la sustente, aunque en el caso de Madrid se justifica la

elección de la estación por ser sus valores los que más se aproximan a la media diaria de todos los observatorios de esa Comunidad, sin entrar en otras consideraciones. No se conoce en ambos estudios ecológicos si efectivamente toda la población está expuesta a análogas condiciones de temperatura aún tratándose de ámbitos geográficos más reducidos que Castilla-La Mancha.

Por tanto, el hecho de que en este trabajo los resultados justifiquen, a nuestro entender, la elección de las estaciones de las capitales de provincia en Castilla-La Mancha como representativas para estudiar la relación de las variables ambientales con la mortalidad u otros indicadores sanitarios provinciales, no obvia que sean necesarios análisis similares en otros ámbitos territoriales (como otras Comunidades Autónomas), especialmente si son extensas, de orografía diversa o incluyen zonas costeras. Más aún si la estación meteorológica de alguna capital que se pretenda utilizar como referencia se encuentra en el interior de una provincia litoral (por ejemplo, Girona o Granada) y se considera como base del estudio la mortalidad provincial. Los resultados podrían verse gravemente distorsionados de tener un peso significativo la población (y por tanto, la mortalidad) de la franja costera, como de hecho suele ocurrir. El resultado podría ser el establecimiento, en su caso, de umbrales de temperatura de alerta a la población por olas de frío o de calor cuanto menos aventurados.

La distribución geográfica de las 42 estaciones seleccionadas abarca prácticamente toda la extensión de la Comunidad Castellano-manchega como puede observarse en la Figura 1, por lo que puede entenderse que representan de una manera fiel toda la gama de variaciones termométricas que pueden darse en esta región. Se decidió incluir Guadalajara «Compuesta» (unión de Guadalajara «Serranillo» y Guadalajara «Instituto») en el estudio a pesar de no cumplir con el criterio de no superar el 6,5 % de datos perdidos

Figura 1

Estaciones meteorológicas seleccionadas



(13,2 %). Las series de temperatura de esta estación son corregidas en sus datos perdidos y participa en el relleno de lagunas de otras series con coeficientes de correlación similares a las de otros observatorios. Su serie de temperaturas máximas (Tmax) no presenta discontinuidad en ningún año pese a constar un cambio de ubicación de la estación en 1985, aunque sí se detecta en Tmed (1984) y Tmin (1985), si bien solo en ésta coinciden el año con discontinuidad y el cambio de localización. La otra estación «Compuesta» (Toledo) no ofrece coincidencia inhomogeneidad- cambio de ubicación en ninguna de sus tres series. Esto puede estar indicándonos la pertinencia de la unión de datos pre y post deslocalización de ambos

observatorios ya que de las seis series implicadas cinco presentan continuidad, según el análisis realizado, precisamente donde cabría esperar discontinuidad. En cualquier caso, allí donde se detecta se subsana mediante la correspondiente homogenización de la serie implicada. Por tanto, no se observa comportamiento diferencial de las series de temperatura de la estación Guadalajara «Compuesta» en los análisis efectuados respecto al resto de estaciones seleccionadas, lo que viene a ser corroborado por los coeficientes de correlación calculados respecto al modelo final extraído en el posterior análisis factorial (Tabla 2). Todo ello justifica la inclusión de dicha estación en el modelo ya que, además, se encuentra en una capi-

tal de provincia y suministra datos de otras variables ambientales muy interesantes para estudios posteriores. Por otra parte, sería poco entendible y poco pedagógico utilizar, llegado el caso, Molina de Aragón o Aranzueque como referencia para establecer, por ejemplo, una temperatura de disparo de la mortalidad en la población de la provincia de Guadalajara por ola de calor. Dado que la finalidad última de estos estudios en salud pública es la implementación de medidas preventivas, conviene hacerlo en términos fácilmente comprensibles por la población afectada.

La limitación que supone la utilización de datos exclusivamente termométricos en este análisis de conglomerados viene impuesta por la información suministrada desde el INM: solamente 6 observatorios de Castilla-La Mancha disponen de datos sobre otras variables meteorológicas o ambientales como horas de luz, humedad o presión atmosférica (las 5 capitales de provincia más Molina de Aragón, en Guadalajara).

En resumen, podemos afirmar que la metodología utilizada nos permite corregir de forma coherente las pérdidas de datos que las series temporales originales puedan presentar y, partiendo de un gran número de estaciones meteorológicas y al menos 30 años de valores diarios de temperaturas, se puede realizar una selección de series representativas de las temperaturas diarias registradas en una región extensa como Castilla-La Mancha, cumpliendo criterios climatológicos internacionalmente aceptados. De esta manera, podemos estimar de una manera fiable la exposición que la población de Castilla-La Mancha tiene a la temperatura del aire.

En resumen, podemos afirmar que la metodología utilizada nos permite corregir de forma coherente las pérdidas de datos que las series temporales originales puedan presentar y, partiendo de un gran número de estaciones meteorológicas y al menos 30

años de valores diarios de temperaturas, se puede realizar una selección de series representativas de las temperaturas diarias registradas en una región extensa como Castilla-La Mancha, cumpliendo criterios climatológicos internacionalmente aceptados. De esta manera, podemos estimar de una manera fiable la exposición que la población de Castilla-La Mancha tiene a la temperatura del aire.

Glosario de términos:

- Análisis de conglomerados jerárquicos: agrupación progresiva, paso a paso, de elementos (variables en nuestro caso) en conglomerados indivisibles cada vez más grandes, en función de su similitud, hasta formar uno solo.
- Análisis factorial: técnica de reducción de datos para encontrar grupos homogéneos de variables a partir de un conjunto numeroso de ellas.
- Análisis de componentes principales: método de extracción del análisis factorial que muestra la variabilidad conjunta de todas las variables, con expresión del número óptimo de factores o conglomerados que explica la mayor parte de la varianza total.
- Autovalores: expresan en un análisis de componentes principales la cantidad de varianza explicada por cada factor. Cada autovalor tiene asociado un porcentaje de varianza explicada.
- Matriz de componentes principales: contiene las correlaciones entre las variables originales y cada uno de los factores extraídos o componentes principales.
- Matriz de componentes principales rotada: opción para facilitar la interpretación de los resultados, siempre y cuando la solución incluya un mínimo de dos factores extraídos.

BIBLIOGRAFÍA

1. Kunst EA, Looman CWN, Mackenbach JP. Outdoor air temperature and mortality in the Netherlands: A time-series analysis. *Am J Epidemiol* 1993; 137:331-41.
2. Wen-Harn P, Lung-An L, Ming-Jan T. Temperature extremes and mortality from coronary heart disease and cerebral infarction in elderly Chinese. *Lancet* 1995; 345:353-5.
3. Curreiro FC, Heiner KS, Samet JM, Zeger SL, Strug L, Patz JA. Temperature and mortality in 11 cities of the Eastern of the United States. *Am J Epidemiol* 2002; 155:80-7.
4. Ballester F, Corella D, Pérez-Hoyos S, Sáez M, Hervás A. Mortality as a function of temperature. A study in Valencia, Spain 1991-1993. *Int J Epidemiol* 1997; 26:551-61.
5. Saez M, Sunyer J, Castellsagué J, Murillo C, Antó M. Relationship between weather temperature and mortality: A time series analysis approach in Barcelona. *Int J Epidemiol* 1995; 24:576-82.
6. Sierra Ortiz, M, Díaz Jiménez J, Montero Rubio JC, Alberdi Odriozola JC, Mirón Pérez IJ. Mortalidad diaria en la Comunidad de Madrid durante el periodo 1986-1991 para el grupo de edad de 45 a 64 años: su relación con la temperatura del aire. *Rev Esp Salud Pública* 1997; 71:149-60.
7. Alberdi JC, Díaz J, Montero JC, Miron I. Daily mortality in Madrid Community 1986-1992: Relationship with meteorological variables. *Eur J Epidemiol* 1998; 14:571-8.
8. Montero JC, Mirón IJ, Díaz J, Alberdi JC. Influencia de variables atmosféricas sobre la mortalidad por enfermedades respiratorias y cardiovasculares en los mayores de 65 años en la Comunidad de Madrid. *Gac Sanit* 1997; 11:164-170.
9. Gill JS, Davis P, Gill SK, Beever DG. Wind-chill and the seasonal variation of cerebrovascular disease. *J Epidemiol Community Health* 1992; 46:261-5.
10. Lambert GW, Reid C, Kaye DM, Jennings GL, Esler MD. Effect of sunlight and season on serotonin turnover in the brain. *Lancet* 2002; 360:1840-2.
11. Montejo Pernas J. Estacionalidad de los trastornos afectivos. *Monografías de Psiquiatría* 2000; Vol XII, N°1.
12. García-Herrera R, Díaz J, Trigo RM, Hernández E, Dessai S. Extreme summer temperatures in Iberia: Health impacts and associated synoptic conditions. *Ann Geophy* 2005; 23:239-51.
13. Martínez F, Simón-Soria F, López-Abente G. Valoración del impacto de la ola de calor del verano de 2003 sobre la mortalidad. *Gac Sanit* 2004; 18:250-8.
14. 49 Reglamento Técnico de la OMM. Documentos Fundamentales n° 2; Vol. I-Generalidades, 1988.
15. DeGaetano AT, Eggleston KL, Knapp WW. A Method to Estimate Missing Daily Maximum and Minimum Temperature Observations. *J Appl Meteorol* 1995; 34:371-80.
16. Schneider T. Analysis of Incomplete Climate Data: Estimation of Mean Values and Covariance Matrices and Imputation of Missing Values. *J Climate* 2001; 14:853-71.
17. García Herrera R, Gallego D, Hernández E. Influence of the North Atlantic Oscillation on the Canary Islands precipitation. *J Climate* 2001; 14:3889-3903.
18. Eischeid JK, Pasteris PA, Diaz HF, Plantico MS, Lott NJ. Creating a Serially Complete, National Daily Time Series of Temperature and Precipitation for the Western United States. *J Appl Meteorol* 2000; 39:1580-91.
19. Prieto L, García Herrera R, Díaz J, Hernández E, Del Teso T. Minimum extreme temperatures over Peninsular Spain. *Global Planetary Change* 2004; 44:59-71.
20. Eischeid JK, Baker CB, Karl TR, Diaz HF. The quality control of longterm climatological data using objective data analysis. *J Appl Meteorol* 1995; 34:2787-95.
21. Young KC. A three-way model for interpolating monthly precipitation values. *Mon Weather Rev* 1992; 120:2561-9.
22. Vicent, L. A. A technique for the identification of inhomogeneities in Canadian temperature series. *J Climate* 1998; 11:1094-1104.
23. Rhoades DA, Salinger MJ. Adjustment of temperature and rainfall records for site changes. *Int J Climatol* 1993; 13:899-913.
24. Peterson TC, Easterling DR. Creation of homogeneous composite climatological reference series. *Int J Climatol* 1994; 14:671-9.

25. Easterling DR, Peterson TC. Techniques for detecting and adjusting for artificial discontinuities in climatological time series: a review. Fifth International Meeting on Statistical Climatology, June 22-26, Toronto, Ontario, pp. J28-J32, 1992.
26. Easterling DR, Peterson TC. A new method for detecting undocumented discontinuities in climatological time series. *Int J Climatol* 1995; 15:369-77.
27. Solow AR. Testing for climate change: An application of the two-phase regression model. *J Climate Appl Meteorol* 1987; 26:1401-5.
28. Lund R, Reeves J. Detection of Undocumented Change-points: A Revision of the Two-Phase Regression Model. *J Climate* 2002; 15:2547-54.
29. Mielke PW, Barry KJ, Brier, GW. Application of Multi-Response Permutation Procedures for Examining Seasonal Changes in Monthly Mean Sea-Level Pressure Patterns. *Mon Weather Rev* 1981; 109:120-6.
30. Karl TR, Williams CN. An Approach to Adjusting Climatological Time Series for Discontinuous Inhomogeneities. *J Climate Appl Meteorol* 1987; 26:1744-63.
31. White D, Richman M, Tarnal B. Climate Regionalization and Rotation of Principal Component Analysis. *Int J Climatol* 1991; 11:1-25.
32. Fernández-Mills G, Lana X, Serra C. Catalonian precipitation patterns: Principal component analysis and automated regionalization. *Theor Appl Climatol* 1994; 49:201-212.
33. Ribera P. Patrones de variabilidad climática en el Mediterráneo. Tesis doctoral. 173p., Universidad Complutense. Madrid, 1999.
34. Kaiser HF. The Varimax criterion for analytic rotation in factor analysis. *Psychometrika* 1958; 23:187-200.
35. Richman MB. Review article, rotation of principal components. *Int J Climatology* 1986; 6:293-335.
36. Stone RC. Weather types at Brisbane, Queensland: an example of the use of principal components and cluster analysis. *Int J Climatol* 1989; 9:3-32.
37. The Eurowinter Group. Cold Exposure and winter from ischaemic heart disease, cerebrovascular disease, respiratory disease, and all causes in warm and cold regions of Europe. *Lancet* 1997; 349:1341-6.
38. O'Neill MS, Zanobetti A, Schwartz J. Modifiers of the Temperature and Mortality Association in Seven US Cities. *Am J Epidemiol* 2003; 157:1074-82.
39. Schwartz J. How Sensitive Is the Association between Ozone and Daily Deaths to Control for Temperature?. *Am J Respir Crit Care Med* 2005; 171:627-31.