

ANÁLISIS DE CALIDAD DE LOS DATOS DE LAS ESTACIONES DE LA XARXA AGROMETEOROLÒGICA DE CATALUNYA (XAC). APLICACIÓN A LA TEMPERATURA

Antonio GÁZQUEZ PICÓN*, M.^a Carmen LLASAT i BOTIJA**, Juan Carlos PEÑA RABADÁN***, Jaume PALLARÉS i BASSETS**** y Miquel PERPINYÀ i ROMEU*****

* *Departament d'Agricultura, Ramaderia i Pesca. Generalitat de Catalunya.* ***Departamento de Astronomía y Meteorología. Facultad de Física. Universidad de Barcelona.* ****Grupo de Climatología. Universidad de Barcelona.* *****PENTA MSI, S.A.*

RESUMEN

Con el objetivo de realizar un análisis de calidad y una depuración de la base de datos de las estaciones agroclimáticas que componen la *Xarxa Agrometeorològica de Catalunya (XAC)*, se presentan los aspectos metodológicos utilizados y su aplicación a uno de los parámetros observados: la temperatura. El análisis consta de dos niveles: un primer nivel donde se realiza un análisis intrínseco de cada observatorio para detectar datos erróneos y un segundo nivel en donde se comparan los datos entre estaciones a partir de una regionalización tanto climática como agroclimática de Catalunya

Palabras clave: *Xarxa Agrometeorològica de Catalunya (XAC)*; Análisis de Calidad; Temperatura

ABSTRACT

The aim of this paper is to show the methodology followed to do the quality analysis of the data obtained from the agrometeorological stations of the Agrometeorological Network of Catalonia (XAC) and its specific application to the temperature. The analysis has two steps: the first one is constituted by the analysis of the individual data station by station; the second one includes the comparison between the data of the different stations placed in homogeneous zones from an agroclimatic point of view.

Key words: *Xarxa Agrometeorològica de Catalunya (XAC)*; *Quality Analysis*; *Temperature*

1. INTRODUCCIÓN

La Red Agrometeorológica de Catalunya (XAC) está constituida en la actualidad por más de 80 estaciones meteorológicas automáticas repartidas por todo el ámbito rural y forestal de Catalunya

y conectadas por teléfono o radio con el centro de control situado en Barcelona. Se trata de un instrumento auxiliar para el mundo agrario, motivo por el cual se ha desarrollado una amplia serie de aplicaciones informáticas que pueden ser utilizadas sistemáticamente (consejos de riego, prevención de plagas, previsión de heladas, riesgo de propagación de incendios, etc.). La validez de los datos de las estaciones es la piedra angular sobre la que se sustenta la bondad de estas aplicaciones.

En base a la normativa presentada por CIMIS (SNYDER *et al.*, 1985), recopilada en un estudio inédito (LLASAT, 1998), retocada y ampliada sobre la base de la experiencia de 10 años de observaciones horarias de las estaciones de la XAC, así como de las peculiaridades propias de Catalunya, se presenta a continuación un análisis de la calidad de los datos de los observatorios agrometeorológicos de la XAC.

En un control de calidad (LLASAT, 1998), se pueden distinguir cuatro fases:

- 1.- Control de calidad sobre el estado de los sensores, así como sobre los datos enviados por la estación. Para el primer caso, las estaciones se calibran "in situ" una vez al año con sensores de referencia calibrados en el laboratorio. Para el segundo caso, a partir de febrero del año 2000 se hace un seguimiento diario exhaustivo de los datos recogidos, haciendo especial referencia a las temperaturas, humedad, radiación global y neta, dirección y velocidad del viento y, por último, la precipitación. Caber decir que este seguimiento se ha venido realizando desde la instalación de las primeras estaciones en 1987, si bien no de una forma tan sistematizada como la que aquí se presenta.
 - 2.- Control de la calidad sobre los datos derivados de los anteriores ya sean directamente parámetros meteorológicos (ETo, presión de vapor, temperatura húmeda y seca) como de las diferentes aplicaciones que dependen directamente de la XAC: PAC-REG (Programa para la gestión del agua de regadío en Catalunya) y el PAC-INC (Programa para la prevención de los incendios forestales en Catalunya).
 - 3.- Control de la calidad sobre los datos acumulados, es decir, los datos diarios y mensuales, tanto de los parámetros meteorológicos como de las aplicaciones que cuelgan de la XAC.
 - 4.- Control de calidad sobre la base histórica, de importancia capital al ser ésta en donde se apoyan todos los cálculos estadísticos y aplicaciones que requieran el uso de series.
- Al controlar la primera y segunda fase, sistemáticamente a partir de febrero del 2000, nos centraremos en el presente estudio en proponer una metodología útil para la cuarta fase, de la cual se derivarán nuevas metodologías no sólo para la tercera fase, sino para mejorar el control que se hace diariamente sobre la primera y la segunda.

2. OBJETIVOS

El objetivo principal que se planteó fue no tan sólo hacer un control sobre la calidad de los datos, sino además depurar en todo lo que fuera posible la base de datos histórica, con el fin de dejarla en unas condiciones óptimas para su consulta por parte del usuario.

Las premisas de partida para dar solución a este objetivo fueron cuatro:

- 1.- La formación de una base de datos a resolución horaria que comprendería desde el inicio del funcionamiento de la estación hasta diciembre de 1999, siempre y cuando ésta tuviese más de dos años de observaciones.
- 2.- Para cada observatorio se eligió otro de referencia con la intención de hacer la comparación de los datos entre estaciones. Esta selección se hizo en función de dos factores:
 - El observatorio elegido como referente, en primer lugar, tendría que ser representativo de la zona agroclimática correspondiente, en base a la Caracterización Agroclimática de Catalunya efectuada por J. Cunillera y M.C. Llasat (las características de cada una de las zonas se pueden consultar en CUNILLERA, 1995; CUNILLERA y LLASAT, 1997; y LLASAT, 1997). En ella se presentaba una clasificación basada en una metodología de regionalización a partir del análisis multivariante de las series de datos meteorológicos (superiores a 30 años) de más de 100 observatorios, de la aplicación de los criterios de Papadakis y de la información extraída a partir de imágenes de satélites heliosíncronos, amén de las características propias del suelo y de la vegetación. Por ella, Catalunya quedaba atomizada en 67 zonas agroclimáticas diferentes (ver Fig. 1).
 - Si no fuera este el caso, se escogería el observatorio en función de:
 - a) Una clasificación climática a partir del Coeficiente de Variación (CV), calculado sobre los datos horarios, *en un primer intento de clasificación térmica de las principales zonas agrarias catalanas a partir de las estaciones de la XAC*, por la que el territorio catalán queda dividido en siete zonas diferentes (ver Fig. 2). De esta manera el país presenta una regionalización en franjas longitudinales dirección NE-SW, prácticamente paralelas entre ellas y organizadas de menor a mayor valor del CV de la costa hacia el interior, diferenciando entre las zonas típicamente mediterráneas con una componente marítima muy importante que modera las temperaturas a lo largo de todo el año (valores de CV por debajo de 35%) hasta zonas típicamente continentales con unas temperaturas muy extremadas tanto en el verano como en el invierno (coeficientes de CV por encima del 60%). La clasificación se ha realizado sobre la base de los siguientes intervalos:
 - 1.- Valores bajos cuando el CV es menor que 40%.
 - 2.- El CV queda comprendido entre el 40 y el 44%, considerándolos como valores moderadamente bajos.
 - 3.- Si se encuentran entre 44 y 48%, se consideraran como valores moderados.
 - 4.- Coeficientes entre el 48 y el 52%, se considerará al CV moderadamente alto.
 - 5.- Los valores entre 52 y 56% se han de interpretar como valores altos.
 - 6.- Si el CV se encuentra entre el 56 y el 60%, se considerará como muy alto.
 - 7.- Por último, si está por encima del 60%, se interpretará como zonas con un componente continental extremo.

El problema que presenta es la escasa longitud que tienen las series (su extensión en el tiempo no va más allá de los 10 años), a pesar que los coeficientes están calculados a partir de las bases horarias. En cambio, la ventaja que presenta la clasificación es que está en relación con la propia red y, por tanto, quedan implícitas las posibles relaciones que puedan haber entre los diferentes observatorios.

b) Como último recurso, se escogería la estación referente en base a la distancia entre observatorios, siendo la distancia mínima la óptima. Es decir, si no se cumplen ninguno de los dos factores anteriores, al menos tendrán que cumplir que la distancia entre ellos no sea muy elevada.

3.- Se considerarán como variables a tratar, la temperatura, la lluvia, la velocidad y dirección del viento, la humedad, la radiación solar y neta y la ETo.

4.- En base al análisis horario de las variables, se pasará a la depuración de la base diaria y mensual.

Dado el carácter pretencioso del trabajo, el objetivo de la comunicación se centrará en el parámetro de más fácil resolución, la temperatura. Su análisis no sólo nos ha de servir para empezar a controlar y depurar la base de datos, sino que, además, se utilizará como experiencia para tratar el resto de parámetros que, en principio, han de presentar un análisis más problemático.

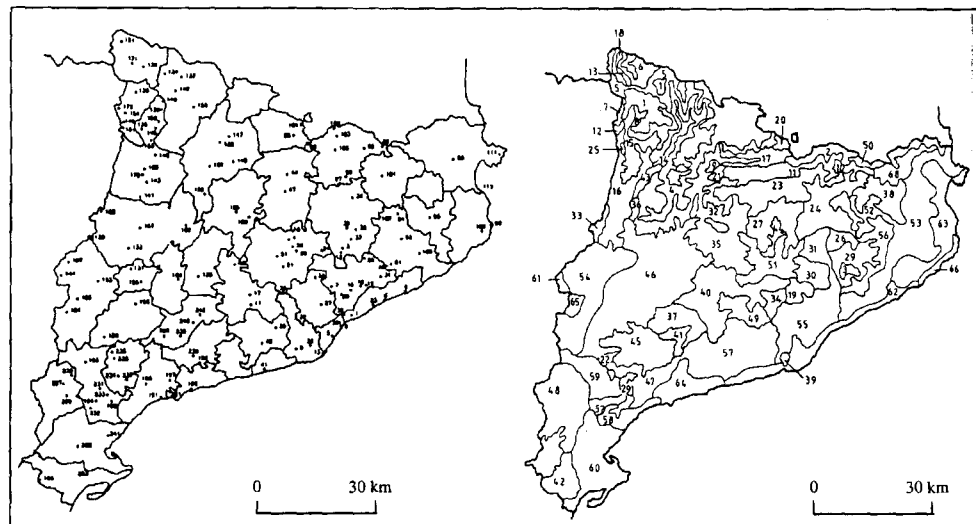


Fig.1: Situación de los observatorios del INM (con más de 30 años de datos) y la regionalización agroclimática del territorio catalán (CUNILLERA, 1995)

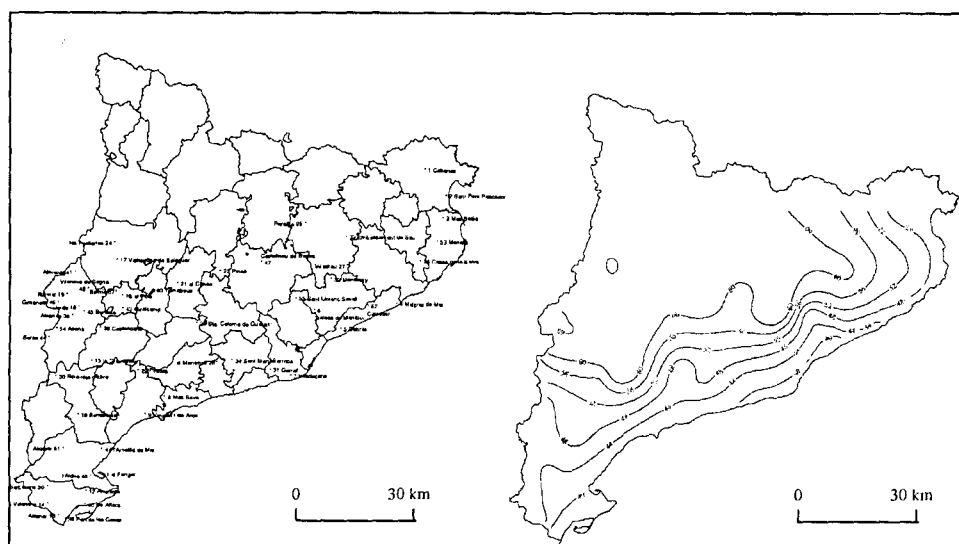


Fig. 2: Situación de los observatorios de la XAC y clasificación climática de las principales zonas agrarias de Catalunya, en base a la regionalización térmica del Coeficiente de Variación

3. METODOLOGÍA

A continuación se presentará la metodología utilizada para realizar un análisis de calidad exhaustivo y una depuración óptima de la base de datos de las temperaturas de la XAC. Se han de diferenciar tres fases:

- 1.- Preparación de la base de datos a tratar.
- 2.- Control de calidad selección e identificación de los datos malos, diferenciando entre los datos dudosos, los considerados como errores y los que están fuera de orden (datos de control del tipo -6999, 999).
- 3.- Tratamiento estadístico de los resultados obtenidos análisis estadístico de los observatorios a comparar, cotejando ambas muestras sobre la base del análisis de la variabilidad y la comparación de las medias de sus poblaciones teóricas, para pasar al cálculo de los coeficientes de correlación y la construcción de las rectas de regresión con la intención de corregir los datos erróneos.

3.1. Preparación de la base de datos a tratar

El objetivo es hacer una cuantificación del volumen de las estaciones y de los datos a tratar, una regionalización de los observatorios (tanto agroclimática como climática) y, por último, una primera selección de las estaciones de referencia para el posterior análisis estadístico.

Las fases a seguir serían las siguientes:

- 1.- Listado de las estaciones existentes en base a la premisa primera, indicando el periodo de fun-

cionamiento y el número potencial de datos.

2.- Clasificación de éstas en función a la zona agroclimática o climática a la que pertenecen.

3.- Selección del par de estaciones en base a la fase 2 para poder realizar el análisis estadístico correspondiente.

3.2. Control de calidad de los datos

Con la finalidad de hacer una selección de los datos erróneos del total de la base de datos se construyó una aplicación informática que se fundamentaba sobre la siguiente premisa básica:

Definición del intervalo a partir del que se podrá distinguir entre los valores erróneos, los dudosos y los aceptables, siendo necesario tomar el mismo criterio para todas las estaciones.

Normalmente, en el caso de las temperaturas, a partir del valor medio y la desviación tipo de la muestra, se compara si cada valor se encuentra dentro del intervalo comprendido entre ese valor medio más/menos k veces la desviación estándar. En este caso, el efecto local vendría implícito en la propia desviación tipo. Se han de considerar dos casos:

1.- *En el caso de que la distribución sea normal:* se hará el control automáticamente cogiendo como válidas todas las observaciones que se encuentren a una distancia de $\pm 4S$.

2.- *En el caso de que la distribución no sea normal:* Se aplicará la desigualdad de Tchebysheff según la cual entre la media y k veces la desviación estándar hay un mínimo de $100[1-1/k^2]\%$ de las observaciones, independientemente de la distribución. En este sentido, por ejemplo, el intervalo definido por el valor medio y $\pm 4S$ incluye como mínimo el 93,75% de las observaciones. Aquellos valores que estén fuera del intervalo se pueden considerar como dudosos o erróneos.

Los módulos del programa son los siguientes:

a) El programa calcula para cada observatorio el valor medio y la desviación estándar para un total de $24 \times 30 \times 12$ distribuciones, es decir, calculará los dos estadísticos para cada hora de cada día y para cada mes sobre la base de los N años de funcionamiento de cada estación. Con estos datos se construye una matriz con los umbrales de temperaturas superiores y otra matriz para los umbrales de temperaturas inferiores.

b) Se comparan todos los valores horarios del archivo histórico con estas matrices, distinguiendo los posibles errores y los días en que falte registro. Los errores se tipifican de la siguiente forma:

- *O:* Aquellos valores que sean superiores a 100 o inferiores a -100. Se consideraran como valores fuera de rango.

- *E:* Serán valores erróneos aquellos que se encuentren fuera del umbral propuesto, en nuestro caso, aquellos que estén por encima de $\pm 6S$.

- *D:* Se consideraran como valores dudosos los registros horarios que se encuentren entre $\pm 4S$ y $\pm 6S$. Éstos requerirán una comprobación manual.

- *F:* Falta registro.

c) Una vez que se ha hecho correr el programa, éste hará un listado con el número total de casos O , E , D y F , marcando la hora y el día en que se han producido para facilitar su identificación, la comprobación y su posible corrección.

3.3. Depuración de la base de datos

El término depuración se ha de entender como sinónimo de comprobación y la posible estimación del dato considerado como erróneo. La finalidad de este último paso del análisis será, en consecuencia, comprobar y, en su caso, hacer una estimación de aquellos datos identificados como poco consistentes. La solución metodológica se ha de buscar en la construcción de un modelo de regresión lineal para cada observatorio sobre la base de un análisis estadístico previo basado en la variabilidad y en la semejanza de las medias poblacionales. Éste constará de tres partes:

a) Comprobar la semejanza entre la variabilidad del par de estaciones. Decidir si las varianzas son aproximadamente iguales con la intención de aplicar un test sobre las medias de las poblaciones en base a la suposición de igualdad entre varianzas. Para confirmar esta decisión, se puede seguir el criterio siguiente:

1.- Se calcula el estadístico experimental (F_{exp}) como el cociente de las varianzas muestrales corregidas de ambas muestras:

$$F_{exp} = \frac{n_1 S_2^2(n_2 - 1)}{n_2 S_1^2(n_1 - 1)}$$

2.- Se determina en la tabla de la distribución de Fisher-Snedecor (valor F), los valores que dejan a la derecha respectivamente áreas $1-(\alpha/2)$ y $\alpha/2$, cogiendo $\alpha=0,05$ y con unos grados de libertad de n_1-1 y n_2-1 .

3.- Se hace el contraste de hipótesis:

- La Hipótesis Nula (H_0): las varianzas son iguales

- La Hipótesis Alternativa (H_1): existen diferencias significativas entre las varianzas,

aceptando la H_1 si el estadístico experimental se encuentra entre los dos valores de F críticos estimados.

b) En base al resultado de la primera fase del análisis, se hace una comparación entre las medias de las muestras, para ver si hay diferencias significativas entre las poblaciones. Se trabajará de la forma siguiente:

1.- Supondremos que la variable estudiada X , en nuestro caso la temperatura, cumple que:

$X \approx N(\mu_1, \sigma_1)$ en la población 1

$X \approx N(\mu_2, \sigma_2)$ en la población 2

Es decir, que nuestras poblaciones son normales y que, además, las desviaciones tipo son semejantes.

2.- Se plantean las hipótesis de contraste:

- La Hipótesis Nula (H_0): $\mu_1 = \mu_2$.

- La Hipótesis Alternativa (H_1): $\mu_1 \neq \mu_2$.

3.- Se calcula el estadístico experimental:

$$t_{\text{exp}} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{n_1 S_1^2 + n_2 S_2^2}{n_1 + n_2 - 2} \left(\frac{n_1 + n_2}{n_1 n_2} \right)}}$$

4.- El criterio de decisión se basa en buscar en la tabla de la distribución t-Student para n_1+n_2-2 grados de libertad y un nivel de confianza del 95%, el valor $t_{\alpha/2}$, aceptando la H_1 si:

$$|t_{\text{exp}}| > t_{\alpha/2}$$

c) Si se acepta que tanto la variabilidad como el valor central de ambos observatorios son semejantes, se construye el modelo de regresión lineal, considerándolo como aceptable si el coeficiente de correlación de Pearson es superior a 0,9, al aceptar la poca variabilidad de las muestras, además de pertenecer a zonas tanto agroclimática como climáticamente parecidas. Si se acepta el modelo se construyen tres tablas que informarán sobre los siguientes parámetros:

- La primera tabla define los estadísticos de la regresión: coeficiente de correlación, coeficiente de determinación, coeficiente de determinación corregido, error típico y el número total de observaciones.
- La segunda tabla hace un análisis de la varianza que nos informa si es válido el modelo de regresión calculado.
- En la última tabla es donde aparecen los coeficientes del modelo, tanto la pendiente como la intercepción, al igual que los intervalos de confianza, para poder construir la recta de regresión. Esta recta será la que se utilizará para estimar los datos defectuosos y rellenar los vacíos horarios de menos de 24 h.

4. RESULTADOS

4.1. Preparación de la base de datos a tratar

De las 85 estaciones agrometeorológicas que en enero del año 2001 funcionaban en la XAC, 55 observatorios (un 65%) cumplían la primera premisa del apartado de objetivos, de los que el más antiguo en funcionamiento es el de Monte-Julia con 13 años de datos horarios, siendo Sant Llorenç de Savall el más moderno con casi tres años de registros. El total potencial de datos a analizar es de 2774912. En la Tabla 1 aparece el listado de las estaciones junto con la comarca en la que se encuentran; la fecha de inicio de toma de registros; el total de datos para cada observatorio; y, por último, la zona agroclimática y la zona climática a la que pertenecen.

4.2. Control de calidad de los datos

En la Tabla 1 aparecen los resultados después del análisis del total de datos a partir de la aplicación informática. El total de errores detectados ha sido de 122880 datos (alrededor del 4%), sien-

Tabla 1: LISTADO DE OBSERVATORIOS UTILIZADOS

ESTACIÓN	COMARCA	INICIO	TOTAL DATOS	ZONA AGRO	ZONA CLIM	ERR. DETEC	ERR. CORR.	% ERR.	% ERR. CORR.
CABANES	ALT EMPORDÀ	11/6/91	75024	63	3	59	59	0,1	100,0
S. PERE PESCADOR	ALT EMPORDÀ	1/5/89	93480	63	3	2081	222	2,2	10,7
S. MARTÍ SARROCA	ALT PENEDEÈS	24/5/97	22848	57	4	883	519	3,9	58,8
CASTELLNOU BAGES	BAGES	22/1/98	22435	51	6	850	130	3,8	15,3
VINYOLS I ELS ARCS	BAIX CAMP	17/5/89	87877	58	1	5672	246	6,5	4,3
PRADES	BAIX CAMP	12/12/95	35544	45	7	28	28	0,1	100,0
L'ALDEA	BAIX EBRE	4/6/91	75192	60	2	11479	268	15,3	2,3
ALDOVER	BAIX EBRE	1/1/95	43824	48	2	556	177	1,3	31,8
EL FANGAR-DELTEBRE	BAIX EBRE	30/4/92	67248	60	1	1039	391	1,5	37,6
L'AMETLLA DE MAR	BAIX EBRE	29/8/95	38040	60	1	907	91	2,4	10,0
MAS BADIA - LA TALLADA	BAIX EMPORDÀ	1/5/89	93480	63	4	3263	449	3,5	13,8
MONELLS	BAIX EMPORDÀ	4/3/98	16008	53	5	14	14	0,1	100,0
GARRAF-BEGUES	BAIX LLOBREGAT	29/2/96	33648	57	2	560	80	1,7	14,3
VILADECANS	BAIX LLOBREGAT	29/4/93	58512	62	1	367	31	0,6	8,4
EL MONTMELL	BAIX PENEDEÈS	13/7/95	39168	57	3	22	22	0,1	100,0
S.C. QUERALT	CONCA DE BARBERA	19/11/96	27288	40	6	63	63	0,2	100,0
CASTELLDANS	LES GARRIGÜES	16/9/97	20064	46	7	1697	353	8,5	20,8
LA GRANADELLA	LES GARRIGÜES	30/11/92	69432	45	6	4561	241	6,6	5,3
CASSÀ DE LA SELVA	GIRONÈS	8/3/93	60432	53	5	3527	186	5,8	5,3
CABRILS	MARESME	27/11/95	35880	62	1	164	92	0,5	56,1
EL CORREDOR - DOSRIUS	MARESME	9/8/96	29736	53	3	737	65	2,5	8,8
MALGRAT DE MAR	MARESME	10/1/90	87648	62	2	1107	267	1,3	24,1
ALCANAR	MONTSIÀ	6/5/91	75864	60	1	1738	250	2,3	14,4
AMPOSTA	MONTSIÀ	6/2/92	69264	60	1	109	109	0,2	100,0
ELS ALFACS - S.C. RAPITA	MONTSIÀ	6/7/94	48096	60	1	2005	181	4,2	9,0
MAS DE BARBERANS	MONTSIÀ	17/11/92	62376	42	2	9605	149	15,4	1,6
PORT DE LES CASES	MONTSIÀ	30/8/93	55560	60	1	9293	101	16,7	1,1
ELS VALENTINS - ULLDEC.	MONTSIÀ	15/11/90	79992	42	3	2075	203	2,6	9,8
LES AVELLANES	NOGUERA	7/7/95	39312	16	6	10	10	0,0	100,0
VALLFOGONA BALAGUER	NOGUERA	17/10/90	80688	54	7	10637	101	13,2	0,9
TERMENS - BELLVÍS	NOGUERA	5/1/97	26160	46	7	7129	193	27,3	2,7
MONTE-JULIA - BELVER	OSCA	19/3/88	103296		6	2087	162	2,0	7,8
EMBASSAMENT SAU	OSONA	20/5/97	22920	26	7	86	86	0,4	100,0
PERAFITA	OSONA	3/7/95	39384	24	7	27	27	0,1	100,0
VILADRAU	OSONA	3/10/95	37176	28	6	18	18	0,0	100,0
MIRALCAMP	PLÀ D'URGELL	15/3/97	24504	46	7	5168	152	21,1	2,9
EL POAL	PLÀ D'URGELL	14/9/89	90240	46	7	1024	256	1,1	25,0
BENISSANET	RIBERA D'EBRE	14/7/93	56664	67	4	24	24	0,0	100,0
RIBA-ROJA D'EBRE	RIBERA D'EBRE	18/11/96	27312	59	3	48	48	0,2	100,0
EL CANÓS - PLANS DE SIÓ	SEGARRA	20/10/88	98088	46	6	5162	194	5,3	3,8
AITONA	SEGRIÀ	12/3/98	15816	65	7	220	28	1,4	12,7
ALMENAR	SEGRIÀ	5/3/97	24768	54	7	4128	312	16,7	7,6
ALCARRÀS	SEGRIÀ	17/9/97	20040	65	7	549	189	2,7	34,4
BORDETA - LLEIDA	SEGRIÀ	16/9/97	20064	54	7	2253	213	11,2	9,5
LLEIDA	SEGRIÀ	30/3/90	85536	54	6	2849	233	3,3	8,2
GIMENELLS	SEGRIÀ	10/12/96	26784	61	6	718	46	2,7	6,4
RAÏMAT - LLEIDA	SEGRIÀ	31/8/88	99360	54	6	1178	98	1,2	8,3
SERÓS	SEGRIÀ	4/1/98	17448	54	5	2904	380	16,6	13,1
VILANOVA DE SEGRIÀ	SEGRIÀ	16/9/97	20088	54	7	523	163	2,6	31,2
PINÓS	SOLSONÈS	3/7/95	39432	35	7	63	39	0,2	61,9
MAS BOVÈ - CONSTANTÍ	TARRAGONÈS	29/5/90	84096	64	2	2313	595	2,8	25,7
TORNABOUS	URGELL	23/6/97	22128	46	7	4367	311	19,7	7,1
SANT LLORENÇ SAVALL	VALLÈS OCCIDENTAL	24/3/98	15552	19	7	23	23	0,1	100,0
MONTSENY-TAGAMANENT	VALLÈS ORIENTAL	4/1/96	34992	55	5	301	85	0,9	28,2
CALDES DE MONTBUI	VALLÈS ORIENTAL	23/2/90	79104	55	5	4610	499	5,8	10,8
		Total	2774912			122880	9472	4,4	7,7

Se indica la comarca donde están instalados, la fecha de inicio, el total potencial de datos, la zona agroclimática y climática a la que pertenecen, el total de errores detectados y corregidos (de tipo o, e, d y f), el tanto por ciento en relación con el total potencial de datos y el tanto por ciento de errores corregidos

do la estación de Termens, situada en la comarca de La Noguera, la que presenta un porcentaje más alto (con 27,3% del total de sus datos), y Benissanet, situada en la Ribera de l'Ebre, la que menos errores presenta (tan sólo 24 datos de los 56.664 posibles). Se han podido corregir completamente 13 estaciones (un 24% de los observatorios analizados), siendo Vallfogona de Balaguer (La Noguera) la estación donde menos errores se han podido corregir (un 0,9% del total de los datos dudosos encontrados en el observatorio). Los errores corregidos suponen un 8% respecto al total de datos malos detectados.

4.3.- Depuración de la base de datos

Dado el objetivo de esta parte del estudio, el análisis de la consistencia del par de observatorios escogidos a partir de la zona agroclimática, climática o la mínima distancia entre ellos, para la corrección y estimación de los datos poco fiables, se ha basado, como se indicaba en el apartado metodológico, en tres fases: análisis de la variabilidad comparando las varianzas de ambas poblaciones; a partir de este primer resultado, se comparan las medias sobre la base de la hipótesis de igualdad de varianzas para pasar, por último, a construir el modelo de regresión correspondiente.

- 1.- Análisis de las varianzas: no se han observado diferencias significativas a un nivel de significación del 95% en los pares de observatorios analizados, a pesar que la coincidencia entre éstos en el ámbito de zonas agroclimáticas o climáticas es tan sólo del 40% para el primer caso y del 22% para el segundo, si bien, en el caso de la regionalización climática, si tomamos zonas contiguas este porcentaje se eleva hasta el 80%.
- 2.- Dado el nivel de exigencia del test de comparación de medias (el contraste se basaba directamente si las medias son o no son iguales para una $\alpha=0,05$), tan solo una estación, Mas de Barberans, ha pasado el test con éxito. Por esta razón, se ha hecho una clasificación a partir de la variabilidad intrínseca del estadístico experimental (t_{exp}) en la que se distinguen cinco tipos diferentes. El nivel de comparación entre el par de observatorios se ha de considerar como muy bueno en 1 caso, bueno en 26 casos, normal para 14 casos, malo en 13 casos y muy malo en 1 caso. De esta manera tenemos un 75% de los observatorios comparados como aceptables, siendo la estación de Prades los que peores resultados presenta.
- 3.- Como resultado de estas dos fases junto con el valor alto de datos analizados (en todo los casos se han analizado más de 15.000 registros), se han de considerar los modelos de regresión obtenidos como óptimos, variando los coeficientes de correlación entre 0,938 y 0,995. Por tanto, en ninguno de los casos se ha rechazado el modelo como poco consistente.

5. CONCLUSIONES

De la aplicación de la metodología expuesta se pueden sacar las siguientes conclusiones:

- 1.- Se ha realizado el primer intento de clasificación térmica de las principales zonas agrarias catalanas a partir de los observatorios que conforman la XAC sobre la base de sus datos horarios. De esta manera Catalunya quedaría dividida en siete zonas de disposición longitudinal y de dirección paralela a la línea de costa que variarían desde las zonas con características mediterráneas con una clara influencia marítima hasta aquellas áreas con unos caracteres climáticos

de tipo continental.

- 2.- El “estado de salud” de la base de datos que construye la XAC desde un punto de vista térmico se ha de considerar, en líneas generales, como excelente, con tan sólo un 4,4% de errores detectados sobre el total de datos, si bien hay algunos observatorios que pueden tener casi un tercio de los datos de sus series con errores. Los fallos más comunes son por falta de registro (93% del total de errores), es decir, por el no funcionamiento de la estación. Por error del sensor (ha dado un dato erróneo) tan sólo se puede hablar de un 2% de los errores detectados, mientras que datos que el programa ha identificado como dudosos se sitúan alrededor del 1%, de los que la mayor parte de ellos se han dado como válidos.
- 3.- La metodología aplicada a la depuración de los datos ha dado unos resultados óptimos, con unos coeficientes de correlación muy elevados, si bien, a pesar de todo, sólo se han podido llegar a corregir un 8% del total de errores detectados, al ser el resto debidos fundamentalmente no por error del sensor sino por un fallo en el funcionamiento general de la estación, es decir, por falta de datos.

6.- REFERENCIAS

CUNILLERA, J. (1995): “L’ús de la teledetecció i l’anàlisi multivariant a la caracterització agroclimàtica del territori. Aplicació a Catalunya”. Tesis doctoral, Universidad de Barcelona. CUNILLERA, J; LLASAT, M.C. (1997): “Application of Remote Sensing and Multivariate Analysis to the Agroclimatic Characterization of a Mediterranean Region”. *Theoretical and Applied Climatology*, 57, 197-208.

LLASAT, M.C. (1997): “Meteorologia agrícola i forestal a Catalunya”. Secció d’Estadística i Informació Agrària, Gabinet Tècnic, Departament d’Agricultura, Ramaderia i Pesca, Barcelona.

LLASAT, M.C. (1998): “Control de qualitat de les dades de les estacions agrometeorològiques del Departament d’Agricultura, Ramaderia i Pesca”. Informe interno.

SNYDER, R.L., LANINI, B.J., SHAW, D.A. y PRUITT, W.O. (1987): “Using reference evapotranspiration (ET_o) and crop coefficients to estimate crop evapotranspiration (E_c) for agronomic crops, grasses and vegetable crops”. Cooperative Extension, University of California.