ASADES

Avances en Energías Renovables y Medio Ambiente
Vol. 10, 2006. Impreso en la Argentina. ISSN 0329-5184

UN PRIMER ANÁLISIS DE DATOS MULTIVARIADOS DE VARIABLES CLIMÁTICAS EN LA PROVINCIA DE CÓRDOBA

J. Adaro, J. Barral, M. I. Pontin
Grupo de Energía Solar Facultad de ingeniería Universidad Nacional de Río Cuarto
Ruta Nac. 36 Km 601. C. P. 5800. Río Cuarto. Argentina
Telefono: 0358-4676488, mail: aadaro@ing.unrc.edu.ar

RESUMEN: El presente trabajo es el resultado de la aplicación de técnicas multivariadas para el estudio de variables climáticas y tiene como finalidad realizar un análisis estadístico exploratorio, a partir de datos relevados por el Servicio Meteorológico Nacional en la provincia de Córdoba y provincias vecinas. Se utilizaron las técnicas del análisis de agrupamiento y de componentes principales. Se presupuso que distintos grupos de localidades analizadas experimentan comportamientos similares y responden de cierta manera a algunas variables fundamentales o combinación de ellas y se pretende con este estudio adquirir un conocimiento sobre las distintas condiciones climáticas que permita en un futuro utilizar los resultados para la toma de decisiones respecto al desarrollo de proyectos de ingeniería en el área de energía solar.

Palabras clave: análisis multivariado, variables climáticas, agrupamiento, componentes principales, energía solar.

INTRODUCCIÓN

Entre las opciones del análisis exploratorio de datos, se utiliza en esta oportunidad el análisis de agrupamiento dado que entre sus objetivos fundamentales está formar grupos con individuos u objetos similares de acuerdo a una medida de distancia. Existen un número importante de técnicas de agrupamiento clasificadas en dos grande grupos: técnicas jerárquicas y no jerárquicas. Las limitaciones de ambos procedimientos son que los resultados dependen, en primer lugar de la distancia entre objetos elegida y del algoritmo. Entre estás últimas se utilizaron lo métodos del vecino más cercano, más lejano y del centroide (single linkage, complete linkage y centroid) utilizados por Johnson D. E. (2000).

En el análisis de componentes principales se persigue obtener un número menor de variables, o un combinación lineal de las primitivas, que se denominan componentes principal o factor, y cuya posterior interpretación permitirá un análisis más simple del problema a estudiar. El análisis de componentes principales permite describir de un modo sintético, la estructura y las interrelaciones de las variables originales en el fenómeno que se estudia a partir de las componentes obtenidas y que sin lugar a dudas habrá que interpretar.

Se presupone que distintos grupos de localidades analizadas experimentan comportamientos similares y responden de cierta manera a algunas variables fundamentales o combinación de ellas y se pretende con este estudio adquirir un conocimiento sobre esta situación que permita en un futuro utilizar los resultados para la toma de decisiones respecto al desarrollo de proyectos de ingeniería en el área solar térmica y fotovoltaica.

El objetivo central de este trabajo es realizar un agrupamiento de localidades en función de datos de sus variables climáticas y un análisis de las interrelaciones entre dichas variables que permita reducir la dimensionalidad del conjunto de datos e identificar nuevas variables significativas subyacentes.

Otro objetivo que se persigue es la adquisición de experiencia en el manejo de softwares vinculados a esta temática como, los cuales presentan distintas opciones de análisis de datos multivariados, ventajas y desventajas, logrando de ser posible el manejo de una herramienta que pueda ser utilizada en el tratamiento de las variables climáticas en conjunto con otra variables que puedan ser de interés para los diferentes casos de estudios.

Los datos considerados en este trabajo fueron extraídos de una base de datos original adquirida oportunamente del Servicio Meteorológico Nacional, que corresponden a la década 81-90 y consta de la siguiente información:

Datos diarios:

- Cuatro datos de Temperatura, a las 2, 8, 14 y 20 horas
- Cuatro datos de Humedad, a las 2, 8, 14 y 20 horas
- Cuatro datos de Velocidad de Viento, a las 2, 8, 14 y 20 horas
- Cuatro datos de Dirección de Viento, a las 2, 8, 14 y 20 horas
- Un dato de Heliofanía

Estos datos corresponden a nueve localidades de la Provincia de Córdoba y alrededores según se indica en la tabla 1 y se muestran en figura 1.

Localidad	Número
Ceres	1
Villa María de Río Seco	2
Chepes	3
Pilar Observatorio	4
Villa Dolores	5
Laboulaye	6
General Pico	7
Marcos Juarez	8
Chamical	9

Tabla 1: Las localidades consideradas en el estudio con su respectivo número de identificación

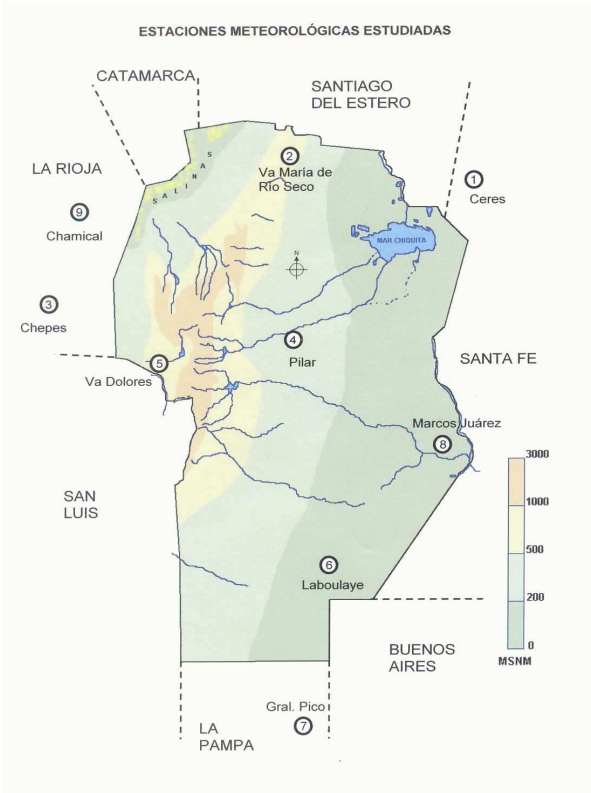

Figura 1: Mapa de la provincia de Córdoba y alrededores con la ubicación de las localidades analizadas.

A partir de esta base de datos se organizaron las matrices para el análisis exploratorio de los mismos (ver Anexo 1). Se tomaron para este análisis la información referida a los meses de enero, abril, julio y octubre para cubrir el comportamiento de los distintos casos durante diferentes épocas correspondiente a las estaciones del año y comparar su variabilidad. Se realizó además un análisis del comportamiento anual, en todos los casos para el año 1981. Para cada caso las variables son TMIN (temperatura mínima en ºC), TMAX (temperatura máxima en ºC), TMED (temperatura media en ºC), HUM (humedad en %), HE (heliofanía en %) y VEL V. (velocidad de viento en Nudos). Se tomaron estas variables entendiendo que ofrecen información representativa de las características meteorológicas de cada localidad.

ANÁLISIS ESTADÍSTICO EXPLORATORIO

En primera instancia sobre cada matriz de datos, correspondientes a cada mes y a los valores anuales, se estandarizaron los datos y se determinó en primer instancia la matriz de distancias utilizando la distancia euclidea.

Para este estudio de agrupamientos se usaron las herramientas provistas por el software comercial utilizado, habida cuenta de la facilidad que presenta para la evaluación del coeficiente cofenético, además de la graficación de los dendogramas y determinación de número de agrupamientos.

A partir de cada matriz de distancias, se realizó un análisis del coeficiente cofenético para los métodos utilizados en el agrupamiento, siendo estos: el vecino más cercano, el vecino más lejano y del centroide según los establece Johnson D. E. (2000).

Se encuentra la matriz de distancia y luego se realiza el agrupamiento utilizando el método del vecino más cercano, repitiendo este procedimiento para los métodos del vecino más lejano y del centroide, y se calcula en todos los casos el correspondiente coeficiente cofenético, y se opta por el método que da el mayor valor para dicho coeficiente. Se procede a continuación a hacer el dendograma correspondiente para ser utilizado a posterior en el agrupamiento de las localidades. Se muestra en la figura 2 el dendograma para el mes de enero de 1981, utilizando el modelo de agrupamiento del vecino más lejano, ya que el coeficiente cofenético fue el mayor respecto a los otros modelos con un valor de equivalente a 0,6357.

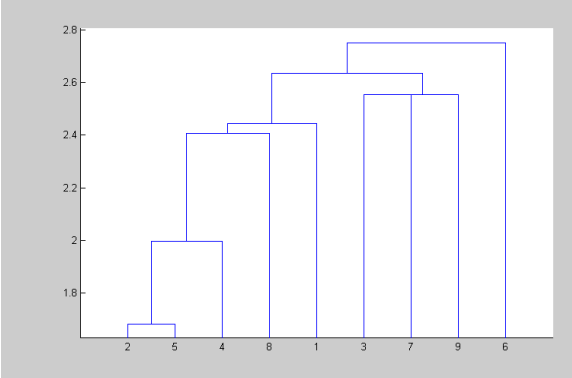

Figura 2: Dendograma para enero de 1981.

Para el análisis de componentes principales se utiliza como software específico aquel con la ventaja respecto a las posibilidades de establecer de ante mano las opciones de su obtención, ya que contiene un procedimiento que puede analizar las componentes principales sobre datos estandarizados.

Componente	Autovalores iniciales(a)		
	Total	% de la varianza	% acumulado
1	116,624	73,995	73,995
2	33,313	21,136	95,131
3	3,955	2,509	97,640
4	2,733	1,734	99,375
5	,773	,490	99,865
6	,213	,135	100,000

Tabla 2: Varianza total explicada por método de extracción: Análisis de Componentes principales.

La figura 3 muestra el gráfico de sedimentación, que sirve para determinar el número óptimo de factores, y para obtener una idea de la representación gráfica del tamaño de los autovalores. Un autovalor indica la cantidad de varianza explicada por una componente principal. Tanto en la Tabla 2 de varianza explicada como en el gráfico de sedimentación ofrece los autovalores ordenados de mayor a menor. De acuerdo a Johnson *et. al.(*1982) el primer autovalor es el mayor de los posibles, el segundo autovalor es el segundo mayor, y así sucesivamente, si un auto valor se aproxima a cero, significa que el factor correspondiente a ese autovalor es incapaz de explicar una cantidad relevante de la varianza total.

Se tomaron los dendogramas de los métodos con mayor coeficiente cofenético para cada mes y para el caso anual y a partir de estos se agruparon las localidades conforme a los pasos del agrupamiento. Esta situación se muestra en los mapas de que se presentan en las figuras 4, 5 y 6.

Al analizar los mapas se observa en términos generales el agrupamiento en primera instancia de las localidades de tras las sierras por un lado, de las localidades de la pampa húmeda por otro y como un tercer grupo las de la zona más central del país. Si bien esta distribución no es exacta en todos los mapas, si se muestra como una tendencia clara.

Al realizar el análisis de componentes principales mostrados en la Tabla 3, se observa claramente en la primer componente un fuerte influencia de la variable heliofanía por sobre las otras, y con una incidencia mucho más débil la humedad relativa. En todos los casos la segunda componente está influenciada fundamentalmente por la humedad y la tercera componente generalmente por algunas de las temperaturas. Es de destacar que en el caso más desfavorable con dos componentes se explica el 90 % de la varianza, y con tres componentes más del 95 % de la varianza.

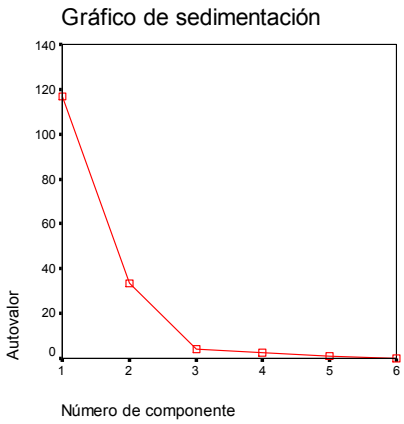

Figura 3: Gráfico de sedimentación

	Componente		
	1	2	3
TMIN	-,099	,112	,617
TMAX	-,290	,640	,090
TMED	-,538	,509	,386
HUM	4,104	-5,226	-,147
HE	9,954	2,154	,195
VEL	,570	,827	-1,832

Tabla 3: Matriz de componentes principales.

En el mapa anual se agrupan por un lado Chepes y Villa Dolores, y luego con Chamical. Estas localidades se encuentran detrás de las Sierra de los Comechingones, mientras que las otras localidades forman un grupo bien determinado y diferente de las localidades anteriores. Esta situación se corresponde con el mapa del promedio anual de radiación global diaria presentado por Grossi Gallegos (1997), siendo ésto razonable dado que en el análisis de componentes principales la variable fundamental en el primer componente es la heliofanía, que se comporta linealmente con el promedio de la radiación global, pero de ninguna manera significa que ésta sea una variable dependiente.

Al observar el mapa de enero, el agrupamiento en primera instancia de Ceres con Laboulaye separado de Marcos Juárez, Pilar, Villa Dolores y Villa María de Río Seco, es congruente con la línea de separación entre la región de 6 y 6,5 KWh/m^2 en el mapa de Grossi Gallegos (1997). Se ve un desplazamiento de dicha línea respecto al agrupamiento de Chepes, General Pico y Chamical.

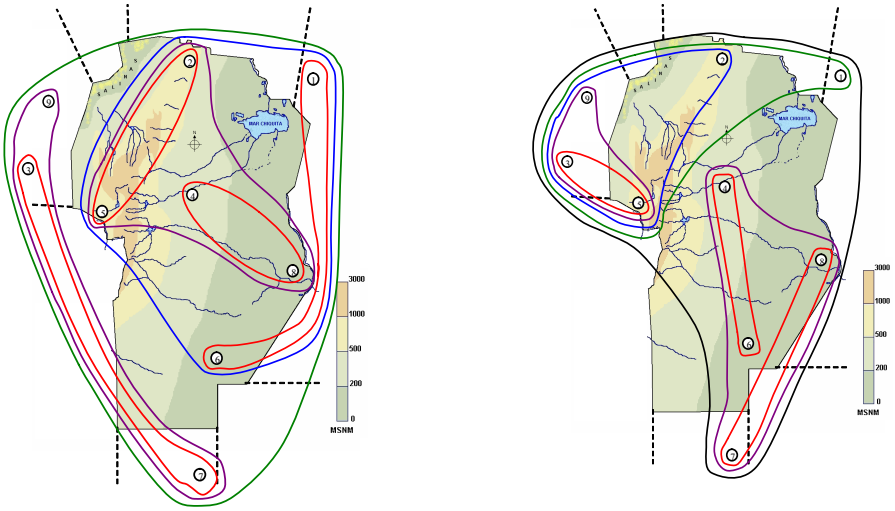

Figura 4: Agrupamiento para el mes de enero de 1981 por el método del vecino más alejado y para el mes de abril de 1981 por el método del centroide.

En abril solo hay consistencia en el agrupamiento en primera instancia de General Pico con Marcos Juárez, respecto a la línea de separación entre la región de 3,5 y 3 KWh/m^2 en el mapa de Grossi Gallegos (1997). En el resto de los agrupamientos no se visualizan coincidencias.

En el mapa de julio se ve una clara consistencia en los dos últimos agrupamientos respecto a la línea que separa las regiones de 2 y 2,5 KWh/m^2 en el mapa de Grossi Gallegos (1997).

En el mapa de octubre en los dos últimos agrupamientos hay congruencia respecto a la línea que separa las regiones de 5 y 5,5 KWh/m^2 en el mapa de Grossi, a excepción de la localidad de General Pico.

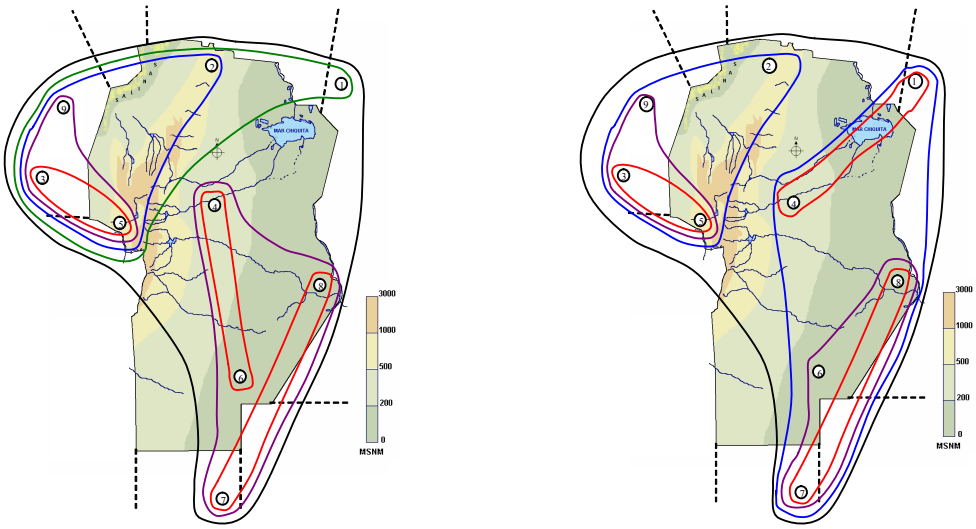

Figura 5: Agrupamiento para el mes de julio de 1981 por el método del vecino más alejado y para el mes de octubre de 1981 por el método del vecino más alejado.

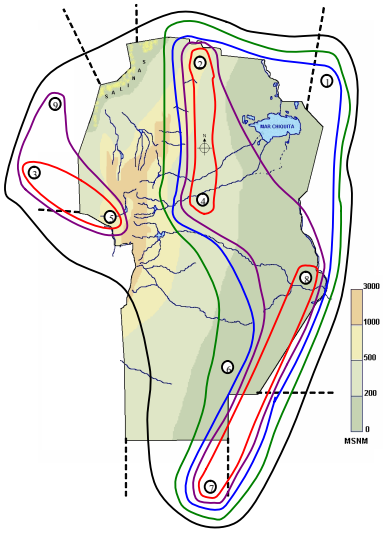

Figura 6: Agrupamiento anual para 1981 por el método del centroide.

CONCLUSIONES

El hecho de haberse tomado la información de sólo un año y de algunos meses, podría explicar la imprecisión de los resultados respecto a la comparación de las cartas de radiación ya existente, lo cual no permitiría en principio aseverar una conclusión definitiva sobre el estudio. Se puede tomar a este estudio como un punto de partida para posteriores análisis de datos, incluyendo más años y más meses, así como otras variables. Por ejemplo en lugar de tomar las temperaturas mínimas y máximas anuales o de cada mes, se podría tomas los valores medio de las mínimas y de las máximas, lo que evitaría trabajar con valores extremos que seguramente se comportarían como datos outliers.

REFERENCIAS

Grossi Gallegos H. (1997) Tesis Doctoral " Evaluación a nivel de superfície de la radiación solar global en la República Argentina. División Física – Departamento Ciencias Básicas. Universidad Nacional de Luján.

Johnson D. E. (2000). Métodos multivariados aplicados al análisis de datos, 1ª edición, pp. 566. Internacional Thomson Editores.

Johnson R. A. y Wichern D. W. (1982) Applied multivariate statistical analysis, 3ª edicción, pp 642. Prentice Hall, New Jersey.

ABSTRACT: This work presents the results obtained through the application of multivariate techniques in the study of climatic variables. One of the objectives is the development of a exploratory statistical analysis taking data from the National Meteorological Service on the Cordoba province and other neighboring provinces in Argentina. Techniques of Cluster Analysis and Principal Components were used. It was assumed that different sets of the analyzed localities experiment similar behaviors and respond in certain way to some fundamental variables or combination of variables. It is expected from this study to get a better knowledge of the climatic conditions, in order to use the information to take decisions in the future about the development of engineering solar projects in this region.

Keywords: multivariate analysis, climatic variables, clustering, principal components, solar energy.

ANEXO 1

Matrices de Datos utilizadas en el trabajo

Matriz de datos para Enero Año 1981

	TMIN	TMAX	TMED	HUM	HE	VEL V.
Ceres	16,50	35,50	24,64	76,56	83,26	6,23
V.M. de Río Seco	16,70	34,10	24,72	76,32	68,70	2,49
Chepes	13,89	34,80	25,45	59,99	68,70	6,38
Pilar	14,40	31,30	23,10	75,16	73,16	3,73
V. Dolores	15,30	33,00	24,25	70,22	67,29	3,05
Laboulaye	15,40	33,00	23,43	72,58	81,58	9,51
Gra. Pico	16,40	36,40	24,46	64,68	64,68	5,83
M. Juarez	13,40	33,40	22,97	78,35	61,13	6,47
Chamical	16,90	34,30	25,97	63,29	49,81	5,11

Matriz de datos para Abril Año 1981

	TMIN	TMAX	TMED	HUM	HE	VEL V.
Ceres	11,20	34,70	20,08	79,69	55,50	5,84
V.M. de Río Seco	7,20	30,70	18,80	86,19	46,54	1,16
Chepes	8,00	30,00	19,14	71,98	46,54	4,02
Pilar	5,90	28,30	17,65	81,63	52,33	2,30
V. Dolores	8,60	31,00	18,71	76,99	54,63	3,51
Laboulaye	7,40	27,20	17,24	81,96	53,33	8,70
Gra. Pico	7,60	29,80	17,00	85,34	38,90	3,76
M. Juarez	6,80	30,20	17,43	85,13	42,57	3,97
Chamical	9,30	31,20	19,65	75,78	28,53	3,68

Matriz de datos para Julio Año 1981

	TMIN	TMAX	TMED	HUM	HE	VEL V.
Ceres	-0,50	27,00	11,38	82,29	56,29	7,54
V.M. de Río Seco	-6,50	25,80	11,52	69,95	36,77	2,49
Chepes	-2,60	23,00	11,07	65,39	36,77	6,04
Pilar	-3,50	22,30	10,26	69,97	68,00	5,12
V. Dolores	-3,60	25,20	11,33	61,19	29,03	6,24
Laboulaye	-4,80	22,60	9,61	74,29	55,71	11,33
Gra. Pico	-4,00	20,80	9,51	74,90	35,55	4,83
M. Juarez	-2,40	22,50	9,93	82,20	39,39	7,61
Chamical	-1,50	24,60	11,42	48,73	46,97	3,89

Matriz de datos para Octubre Año 1981

	TMIN	TMAX	TMED	HUM	HE	VEL V.
Ceres	6,90	37,90	19,31	59,77	90,55	7,98
V.M. de Río Seco	10,00	37,10	21,22	55,59	65,97	3,51
Chepes	6,40	36,40	21,67	47,32	65,97	5,66
Pilar	9,30	33,00	19,40	53,70	91,42	5,51
V. Dolores	7,40	36,00	21,03	41,72	57,13	5,95
Laboulaye	7,60	33,00	18,23	57,41	63,61	11,96
Gra. Pico	6,80	33,20	18,51	57,72	62,97	6,40
M. Juarez	4,80	33,80	18,18	64,97	46,87	7,33
Chamical	8,90	36,50	22,69	39,26	49,26	7,10

Matriz de datos Anual Año 1981

	TMIN	TMAX	TMED	HUM	HE	VEL V.
Ceres	-1,30	37,90	19,27	74,94	71,48	7,05
V.M. de Río Seco	-6,50	37,10	19,53	70,72	59,42	2,67
Chepes	-2,60	39,60	20,03	60,98	59,42	5,13
Pilar	-5,00	34,10	18,01	68,62	74,68	4,35
V. Dolores	-3,60	37,60	19,61	60,73	56,04	4,35
Laboulaye	-4,80	35,80	17,49	68,61	70,96	10,62
Gra. Pico	-4,00	36,60	17,39	68,82	51,95	5,54
M. Juarez	-2,40	35,00	17,44	76,78	49,37	6,53
Chamical	-1,50	40,70	20,55	57,47	41,44	4,77