

Título: Homogenización y pronósticos de Series de Temperaturas Decenales.

**Ricardo Osés Rodríguez.*
Guillermo Saura González.***

*** Centro Meteorológico Provincial. Villa Clara. Cuba**

**Calle Marta Abreu. No. 59. Esquina Juan Bruno Zayas. Santa Clara. Villa Clara.
Cuba**

Palabras Claves: homogeneización, pronóstico.

Resumen:

En el presente trabajo se analiza la homogeneidad de 3 series meteorológicas correspondientes a la Provincia de Villa Clara en el período 1961-1993 que han sufrido cambios de posición. Se determina el momento del cambio y se cuantifica la magnitud del mismo mediante la utilización de los modelos ARIMA para modelar el comportamiento de la serie. Se realizaron análisis de intervención utilizando variables de salto. Se obtienen buenos modelos para describir la serie y se evalúan los resultados. El método utilizado tiene la particularidad de que no utiliza estaciones vecinas por lo que abre un camino para aquellas series largas para las cuales no se cuenta con estaciones vecinas, se evidencia la posibilidad de la utilización de modelos ARIMA estacionales para el pronóstico operativo de temperaturas decenales, además se presentan los principales indicadores de los modelos así como los principales índices para su evaluación.

1.Introducción.

Se han realizado trabajos importantes donde se ha contado con datos adicionales y de archivos meteorológicos [Bradley y otros (1985), y Jones y otros (1985)] donde además se ha contado con información histórica donde se indican los cambios de estación [Goodess y otros (1985)]. En Cuba se han realizado trabajos en este sentido [Cárdenas y otros (1994)] donde se analiza la homogeneidad de la Serie utilizando el Método propuesto por [Conrad y Pollack (1962)], para datos mensuales.

En este trabajo se describe el estudio de la homogeneidad de las series, se discuten las posibles causas de los saltos así como se cuantifica la magnitud de los mismos, se modelan las Series con ayuda de los métodos ARIMA y se utiliza una muestra independiente para evaluar el modelo, además se dan los principales estimadores.

2. Materiales y Métodos.

En el trabajo se utilizaron las series de temperaturas máximas medias decenales de tres estaciones meteorológicas de la provincia de Villa Clara, Cuba de las cuales se cuenta con cierta información sobre la historia de las mismas, más exhaustiva en unas que en otras. Para las series de Yabú (Tx 343) y Caibarién (Tx 348) se utilizó el período 61-93 mientras para Sagua Tx 338 se tomó el período 77-93.

Según los reportes con que se cuenta en los archivos los cambios ocurrieron en las siguientes fechas:

- Tx 343 Febrero 1966, Septiembre 1976 aproximadamente.
- Tx 348 Enero 1966, en la década de los años 70 y en Agosto 1987(Carlos ,1995).
- Tx 338 no se conoce de cambio registrado en los archivos.

En todos los casos de saltos no se contaba con la fecha exacta del cambio, en el caso de Tx 343 se encontraba hasta febrero del 66 en la azotea de un edificio a 87 m sobre el nivel del mar, posterior al 1966 se trasladó hacia un lugar cercano rodeado de césped a 81 m sobre el nivel del mar con gran abundancia de árboles en los alrededores, por último en Septiembre de 1976 se trasladó hacia un lugar más despejado situado a 116.4 m sobre el nivel del mar.

En el caso de Tx 348 se conoce que hasta Febrero de 1966 se encontraba en la azotea en la casa de un aficionado en ese mismo año comienza a trabajar sobre césped cercana a una Secundaria, en la década de los 70 se trasladó hacia la azotea de la Secundaria debido a la influencia de unos árboles que habían crecido alrededor de la misma posteriormente en Agosto de 1987 pasar a un lugar cubierto por césped a 45 m sobre el nivel del mar.

Principales causas de las inhomogeneidades.

Existen cuatro factores que han sido identificados como los de mayor incidencia [Mitchell, 1953], [Bradley y Jones, 1985]: cambios en los instrumentos de medición, cambios de lugar, cambios en el tiempo de observación y los métodos para calcular los valores medios y cambios en el entorno alrededor de la estación.

En las grandes escalas los efectos de urbanización no son de considerable magnitud [P.D. Jones, S.C. Raper, R. S. Bradley, H. F. Díaz, P. M. Kelly and T. M. L. Wingley, (1985)], sin embargo en dependencia de las regiones bajo estudio las inhomogeneidades debidas a efectos de urbanización pudieran tener cierta importancia.

Homogenización

Existen dos formas de atacar el problema de la homogeneidad de la serie como por ejemplo.

1. Chequear todos los records exhaustivamente cuando se cuenta con la historia de la estación para que sirva de guía.
2. Utilizar las inconsistencias con respecto a estaciones vecinas como una guía para las inhomogeneidades mayores en la base de datos.

Como la elección depende en última instancia de la aplicación preferimos seguir diferente pasos lógicos que nos permitan obtener un acercamiento a la forma 1 para que en el caso del estudio de una serie particular donde no se cuenta con estaciones vecinas como es nuestro caso sirva de guía. Por lo tanto consideramos los siguientes pasos.

1. Todas las estaciones bajo estudio pueden tener inhomogeneidades.

...Para esto calcularemos las medias anuales con vista a eliminar la afectación de factores de variación decenal y mensual, y utilizando el método de los correlogramas determinaremos si los valores anuales corresponden a un ruido blanco, lo que de ser cierto indicaría la presencia de una serie homogénea (estacionaria).

2. Una vez determinada la no homogeneidad planteamos los valores anuales para por simple inspección determinar en que momento se ha producido el salto (La serie diferenciada anual puede ayudar a determinar el momento del salto.
3. Una vez determinado el momento del salto por simple inspección planteamos los valores decenales para determinar may exactamente el momento del salto.

4. Una vez determinados los puntos de cambio decenales modelamos la serie con la utilización de los modelos ARIMA de la parte más reciente de los datos.
5. Creamos las variables de salto unitario y evaluamos la influencia de los saltos sobre la serie.
6. Utilizando la magnitud del salto homogenizamos la serie y volvemos al paso uno esta vez con la serie corregida.

3.Resultados y Discusión.

Al calcular los correlogramas para las series largas (1961-1993) observamos que para Tx 343 (Fig. 1.) difieren significativamente de un ruido blanco según el estadístico de Box-Ljung [Box, G. E. P. and Jenkins, G. M., (1976)], además los autocorrelogramas presentan espigas lo que nos indica la existencia de inhomogeneidad en la serie por tanto, plotamos los valores anuales.

Fig 1. Correlogramas Tx 343.

Autocorrelaciones: TX343

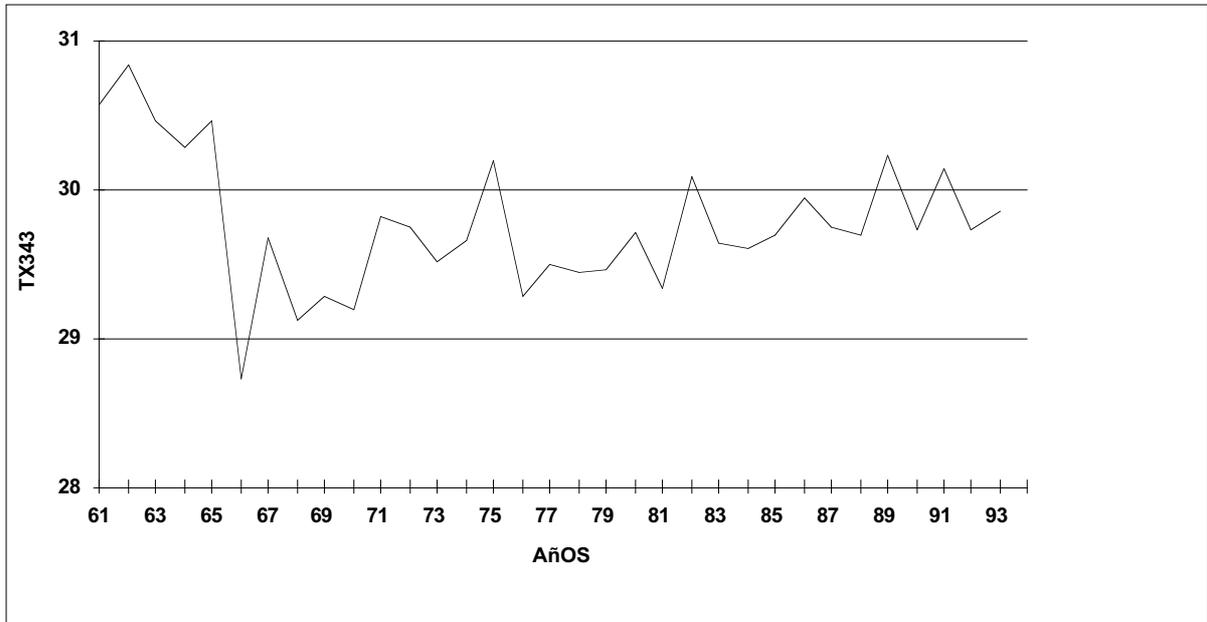
Auto- Stand.

Lag	Corr.	Err.	-1	-.75	-.5	-.25	0	.25	.5	.75	1	Box-Ljung	Prob.
1	.339	.166	4.145	.042
2	.374	.164	9.354	.009
3	.133	.161	10.037	.018
4	-.046	.158	10.120	.038
5	-.288	.156	13.543	.019
6	-.218	.153	15.571	.016
7	-.097	.150	15.992	.025
8	-.097	.147	16.430	.037
9	-.074	.144	16.692	.054
10	.092	.141	17.119	.072
11	.057	.138	17.289	.100
12	-.041	.135	17.380	.136
13	-.035	.132	17.449	.180
14	-.093	.128	17.971	.208
15	-.111	.125	18.755	.225
16	-.204	.121	21.595	.157

Plot Symbols: Autocorrelations * Two Standard Error Limits .

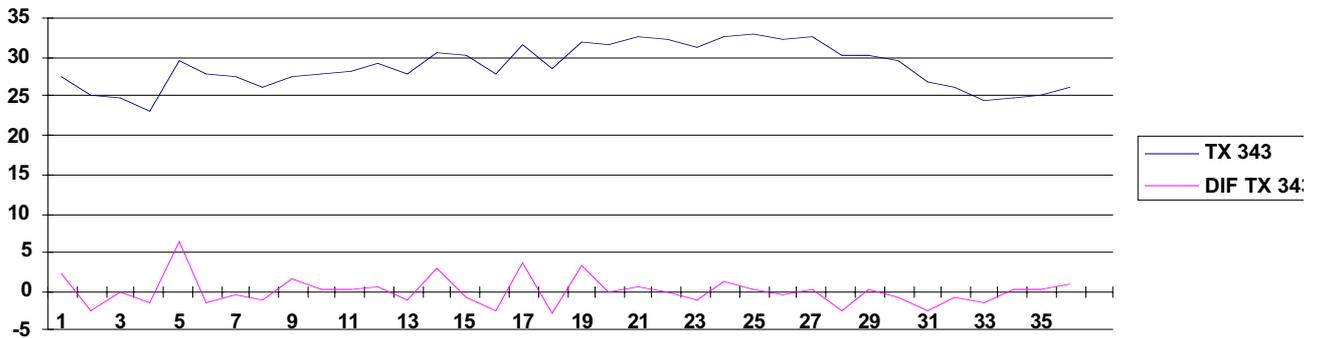
En la Fig. 2 puede observarse claramente la existencia de dos saltos uno en 1966 y otro en 1976.

Fig. 2. Marcha Anual de la Temperatura (TX 343).



Una vez determinado el momento del salto en la serie anual planteamos los valores decenales para determinar el momento exacto del salto. En la Fig. 3 se muestra que para 1966 ocurre un cambio brusco en la quinta decena, para asegurarnos planteamos la serie diferenciada, llegando a la conclusión que ha ocurrido un salto de 6,2 °C.

Fig. 3: Marcha de la Temperatura (TX 343) y de la Serie Diferenciada (DIF TX 343) del año 66 .



Con la parte más reciente (76-93) se modelaron las series y tras diferentes intentos con otros modelos alternativos ARIMA se obtuvieron las siguientes ecuaciones.

Analysis of Variance TX 338:

	DF	Adj. Sum of Squares	Residual Variance
Residuals	537	806.07004	1.3103220

Variables in the Model:

	B	SEB	T-RATIO	APPROX. PROB.
AR1	.20791320	.04000643	5.196994	.00000030
AR2	.14648593	.03969837	3.689973	.00024740
SMA1	.9430963	.08080262	11.671606	a .00000000

Analysis of Variance TX 343:

	DF	Adj. Sum of Squares	Residual Variance
Residuals	1149	1956.7109	1.5625626

Variables in the Model:

	B	SEB	T-RATIO	APPROX. PROB.
AR1	.24970869	.02805126	8.901871	.00000000
AR2	.14563083	.02804650	5.192478	.00000026
SMA1	.97286338	.04401957	22.100703	.00000000

Analysis of Variance TX 348:

	DF	Adj. Sum of Squares	Residual Variance
Residuals	540	620.43433	.98360296

Variables in the Model:

	B	SEB	T-RATIO	APPROX. PROB.
AR1	.19144172	.03939633	4.8593792	.00000159
AR2	.15645850	.03942595	3.9684140	.00008237
SMA1	.89153921	.24413827	3.6517798	.00028603
SMA2	.20501393	.04874490	4.205854	.00003062

Para la Tx 338 y Tx 343 se obtuvieron modelos regresivos de segundo orden (AR1, AR2) combinados con un término de media móvil de primer orden estacional diferenciado (SMA1) donde todos los términos son significativos, para el caso de Tx 348 el modelo obtenido difiere un poco de los anteriores ya que presenta un término de media móvil de segundo orden donde también los parámetros son significativos

Una vez con estos modelos creamos las variables salto unitario STEP que toman los siguientes valores:

STEP = 0. Fuera del período de salto.

STEP = 1. Dentro del período de salto.

En el caso de Tx 343, la variable STEP3431 toma el valor de 1 para el período 61 hasta 66 (Decena 5) y toma el valor 0 para el resto de la serie, en el caso del segundo salto la variable STEP3432 toma el valor de 1 para el período 66 (Decena 6) hasta 76 (Decena 31), y toma el valor de 0 para el resto, es de destacar que el segundo salto se produjo en Noviembre y no en

Septiembre como consta en los archivos, esto puede ser debido a que cuando se cambiaba una estación de lugar se realizaban mediciones conjuntas durante algún tiempo, lo que pudo influir en que para la nueva localización el asentamiento de los datos se hiciera en meses posteriores (dos meses) a la fecha que inicialmente se debía cambiar la estación.

Posteriormente introducimos estas variables al modelo, obteniéndose los siguientes resultados

Analysis of Variance TX 343:

	DF	Adj. Sum of Squares	Residual Variance
Residuals	1147	1898.2682	1.4998099

Variables in the Model:

	B	SEB	T-RATIO	APPROX. PROB.
AR1	.21864825	.02801420	7.804909	.00000000
AR2	.11166232	.02801796	3.985384	.00007186
SMA1	.98841376	.09644705	10.248253	.00000000
STEP3431	.71772689	.15265025	4.701773	.00000296
STEP3432	-.28576176	.11818243	-2.417972	.01576227

Analysis of Variance TX 348:

	DF	Adj. Sum of Squares	Residual Variance
Residuals	540	620.43433	.98360296

Variables in the Model:

	B	SEB	T-RATIO	APPROX. PROB.
AR1	.17436092	.04126499	4.2253960	.00002817
AR2	.17100286	.04135381	4.1353818	.00004132
SMA1	.56657933	.05287781	10.714878	.00000000
SMA2	.01740202	.04118542	1.4284779	.00073223
STEP3481	.60701718	.13194020	4.600699	.00000476
STEP3482	.39606785	.10137578	3.906928	.00009919

Para Tx 343 ambos saltos fueron significativos como se puede constatar. La serie 61 hasta 66 (Decena 5) esta 0.72 oC may caliente que la serie may reciente 76 (Decena 31) hasta 93 .La serie 66 (Decena 6) hasta 76 (Decena 31) estaba 0.28 oC por debajo de la más reciente, es interesante destacar como este ultimo cambio esta relacionado con la diferencia de altura de las series que ge de 30 m aproximadamente ,si calculáramos la variación de temperatura debida a la altura podemos obtenemos aproximadamente 0.33 oC lo que no difiere significativamente de la magnitud del salto(0.28 oC). Si analizamos que la estación en la primera serie 61 -66(5) se encontraba en una azotea pudiéramos intentar el cálculo del efecto de la construcción restando 0.28 oC(que evidentemente son debidos al cambio de altura 30 m, considerando que ambas

series tienen similares las condiciones de emplazamiento, salvo quizás una diferencia de altura de 6 m) de los 0.72 oC (que son el resultado de factores del cambio de posición por altura y por efecto de la contracción que se encuentra en la azotea) esto nos daría aproximadamente 0,44 oC.

Para el caso de la Tx 348 el primer salto 61 - 66(5) representa que la estación estaba ubicada en un lugar 0.6 oC mas caliente que la actual 76 (31) -93 , respecto al segundo salto 66 (6) hasta 76 (31) la magnitud fué de 0.4 oC más caliente que la serie actual, en este caso no existe una certeza de la exactitud de las fechas y los lugares de los cambios en los archivos.

Evaluación de los Pronósticos

Una vez establecida la magnitud del cambio se realizaron los pronósticos en una muestra independiente de 36 casos de 1993 obteniéndose los siguientes indicadores.

Muestra independiente 1993.

		TX 348	TX 343	TX 338
Mean Error	Use	- 0.003	0.008	0.035
	Predict	0.085	0.111	0.103
Mean Abs Error	Use	0.805	0.997	0.936
	Predict	0.754	0.939	0.895
Mean Pct Error	Use	-0.121	-0.135	-0.020
	Predict	0.216	0.248	0.197
SSE	Use	635.950	1994.960	840.550
	Predict	34.230	54.800	48.220
MSE	Use	1.173	1.696	1.565
	Predict	0.954	1.522	1.340
RMS	Use	1.083	1.302	1.251
	Predict	0.985	1.234	1.157
Durbin -Watson	Use	2.010	2.060	2.031
	Predict	1.519	1.150	1.125
Correlación	Predict	0.900	0.830	0.870

De todos los estadígrafos el criterio may fuerte de comparación es el MSE (Mean Square Error), además no se observan diferencias significativas en la parte de uso y de predicción por lo que podemos decir que el modelo trabaja correctamente, los mejores resultados se obtienen para Tx 348 . La interpretación de los primeros es obvia , el se calculan utilizando error medio debe ser cercano a cero. Los errores en porciento se calculan utilizando como denominador 1

os valores observados de la serie y luego se promedian incluyendo signos(Mean Pct Error) y en valor absoluto(Mean Abs Pct Error).,SSE denota la suma de cuadrados de los errores, MSE es la media de MSE y RMS es la raíz cuadrada de MSE, por último el estadístico de Durbin Watson[Durbin, J. And Watson, G. S. (1951), nos indica la no existencia de autocorrelación en la fase de uso y cierta correlación positiva en la fase de predicción positiva debida quizás a la heterocedasticidad estacional. En la muestra independiente se calculó el skill obteniéndose valores de 0.83 ,0.87 y 0.90 % lo que nos indica la posibilidad de la utilización de estos modelos para el pronóstico de temperaturas medias decenales.

Por último una vez homogeneizada la serie calculamos nuevamente los correlogramas (Fig. 4 a,b) mostrando esta vez la existencia de un ruido blanco lo que nos indica que son homogéneas.

Fig. 4. a: Correlogramas Tx 343 Homogenizada.

Autocorrelations: TX343 HOMOGENIZADA

Auto- Stand.

Lag	Corr.	Err.	-1	-.75	-.5	-.25	0	.25	.5	.75	1	Box-Ljung	Prob.
1	-.022	.166				*						.017	.896
2	.216	.164				****						1.757	.415
3	.008	.161				*						1.760	.624
4	-.010	.158				*						1.764	.779
5	-.236	.156				*****						4.068	.540
6	-.330	.153				*.*****						8.724	.190
7	-.055	.150				*						8.858	.263
8	-.131	.147				***						9.652	.290
9	-.169	.144				***						11.024	.274
10	-.028	.141				*						11.062	.353
11	.130	.138				***						11.951	.367
12	.000	.135				*						11.951	.450
13	.122	.132				**						12.809	.463
14	.061	.128				*						13.036	.524
15	.159	.125				***						14.655	.477
16	-.054	.121				*						14.855	.535

Plot Symbols: Autocorrelations * Two Standard Error Limits .

Fig. 4. b: Correlogramas Tx 343 Homogenizada.

Autocorrelations: TX348 Homogenizada

Auto- Stand.

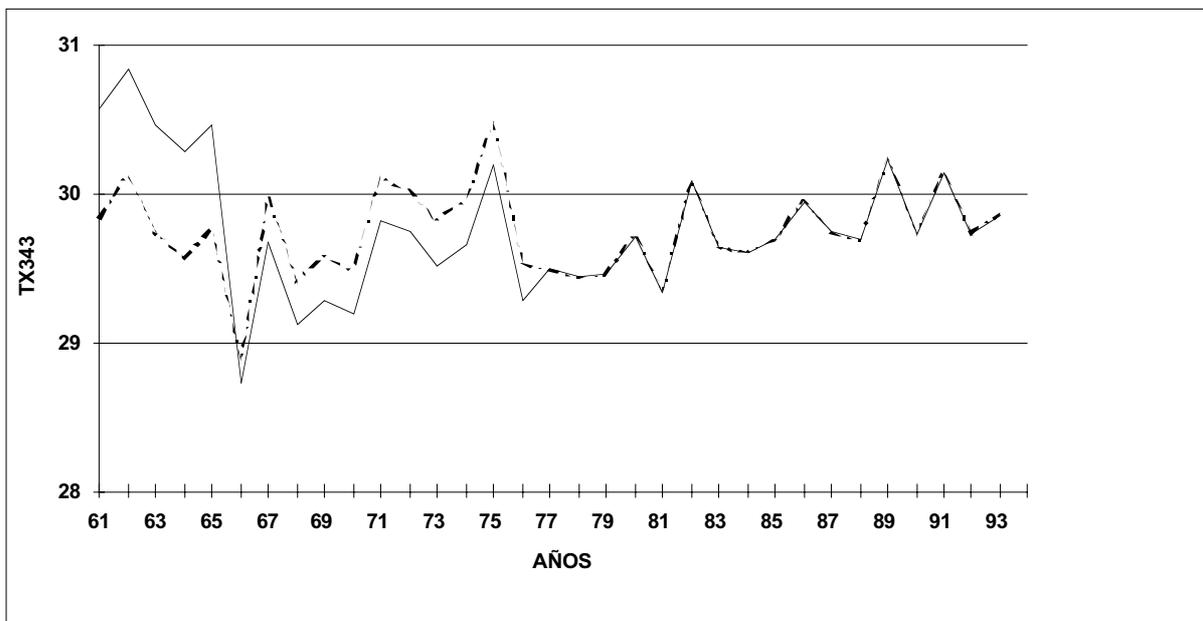
Lag Corr. Err. -1 -.75 -.5 -.25 0 .25 .5 .75 1 Box-Ljung Prob.

Lag	Corr.	Err.	-1	-.75	-.5	-.25	0	.25	.5	.75	1	Box-Ljung	Prob.
1	-.284	.166		*****								2.908	.088
2	-.074	.164		*								3.111	.211
3	.018	.161		*								3.123	.373
4	-.068	.158		*								3.305	.508
5	-.028	.156		*								3.337	.648
6	.003	.153		*								3.337	.766
7	-.017	.150		*								3.350	.851
8	.152	.147			***							4.421	.817
9	-.214	.144		****								6.631	.676
10	-.065	.141		*								6.842	.740
11	.331	.138			*****	*						12.601	.320
12	-.137	.135			***							13.627	.325
13	.041	.132			*							13.723	.394
14	.099	.128			**							14.323	.426
15	-.225	.125		****								17.569	.286
16	-.058	.121		*								17.800	.336

Plot Symbols: Autocorrelations * Two Standard Error Limits .

En la Fig. 5 se observa un ejemplo de la homogenización de la serie TX 343.

Fig. 5: Serie Tx 343 ____ . Homogenizada TX 343 - - - -.



4. Conclusiones

Se demuestra la validez del empleo de los métodos ARIMA para la homogeneización de series de observación meteorológica, así como para el pronóstico de las mismas.

Se evidenció la coincidencia en la fecha de los cambios reflejados en los registros de los archivos históricos y las determinadas por la metodología.

Agradecimientos

Quisiera agradecer la ayuda del Dr. Ricardo Grau por sus sugerencias y al Departamento de Meteorología Aplicada del Centro Meteorológico Provincial, por la información histórica brindada.

Referencias

- Box, G. E. P. and Jenkins, G. M., 1976: Time Series Analysis Forecasting and Control, Holden- Day, San Francisco.
- Bradley et al, 1985: Data bases for detecting CO₂ - induced climatic change. U. S. Department of Energy State of the Art Report, Detecting the Climatic Effects of Increasing Carbon Dioxide, U. S. Dept. Of Energy Carbon Dioxide Research Division, Washington, D.C.(in press).
- Cárdenas et al ,1994: Pronósticos Mensuales de totales de lluvia y temperaturas extremas en Cuba, Elementos de la Variabilidad Climática. Grupo Nacional de Pronóstico a Largo Plazo, INSMET Cuba.
- Conrad, V., and Pollack, L. D. ,1962: Methods in Climatology , Harvard University Press, 459 pp.
- Durbin, J. And Watson, G. S.,1951: Testing for Serial correlation in least squares regression, II, Biometrika 38.
- Mitchell. J. M. Jr., 1953: On the causes of instrumentally observed secular temperature trends. J. Meteor., 10, 244-261.
- P. D. Jones, S. C. B. Raper, R. S. Bradley, H, F. Díaz , P. M. Kelly and T. M. L. Wingley, 1985: Northern Hemisphere Surface Air Temperature Variations: 1851-1984, American Meteorological Society 1986.
- P. M. Kelly. P. D. Jones, H. F. Díaz and C. Goodess,1985: A climatic data bank for the Northern Hemisphere land areas, 1851- 1980. DoE Tech Rep No. TRO17, U. S. Dept of Energy, Carbon Dioxide Research Division, Washington, D.C.,335 pp.