

Análisis de la calidad de series largas de registros de código de estado del tiempo presente para las estaciones de Cuba

Dra. Lourdes Álvarez Escudero | *lourdes.alvarez@insmet.cu* | Centro de Física de la Atmósfera, Instituto de Meteorología
Lic. Israel Borrajero Montejo | *israel.borrajero@insmet.cu* | Centro de Física de la Atmósfera, Instituto de Meteorología
Lic. Maydes Bárcenas Castro | *maydes.barcenas@insmet.cu* | Centro de Física de la Atmósfera, Instituto de Meteorología

Recibido: abril 3, 2014; aceptado: mayo 16, 2014.

Resumen

El conocimiento sobre el comportamiento de los fenómenos meteorológicos resulta de gran importancia por la incidencia que estos tienen sobre las personas, la economía y el medio ambiente. Para el estudio climático de los fenómenos se necesitan series largas de información lo más completas posibles. El objetivo del presente trabajo es evaluar los criterios de calidad de las series de código de estado de tiempo presente como una variable clasificadora de fenómenos, determinada en 69 estaciones del territorio cubano, y sugerir los períodos y horarios para estudios climáticos posteriores. De las series por estaciones de código de estado de tiempo presente, 98.5 % pueden calificarse como buenas, muy buenas o excelentes con respecto a la calidad de la información contenida; en este sentido, la limitante principal es la falta de registros en los horarios de la noche y la madrugada. Se recomienda usar el período 1989-2010 con miras a estudios climáticos posteriores, excepto para la marcha diaria, en cuyo caso se recomienda el período 2005-2010.

PALABRAS CLAVE: Calidad de los datos meteorológicos, series de observaciones meteorológicas, código de estado de tiempo presente.

Abstract

Knowledge of meteorological phenomena is relevant for their incidence on people, the economy and the environment. In order to carry on climatic studies of different phenomena, long, as complete as possible data series are needed. The purpose of this work is to evaluate quality criteria for current weather state code series, as a phenomena classifying variable at 69 stations on Cuban territory and suggest time lapses and hours for further climatic studies. 98.5 % of current weather state code series in all stations can be considered as good, very good or excellent, with regard to the amount of information contained, being the lack of night time observations the main limiting issue. It is recommended to use the period 1989-2010 for further climatic studies, except for daily courses, for which the recommended period is 2005-2010.

KEYWORDS: Meteorologic data quality, meteorological observations series, current weather state codes.

Introducción

El estudio de los fenómenos meteorológicos es de gran importancia para brindar información útil, tanto a las personas, como a los intereses socioeconómicos.

cos, con vistas a la mitigación de sus efectos adversos y(o) el aprovechamiento de las ventajas que pueda generar.

Para identificar los fenómenos meteorológicos, en las estaciones meteorológicas existen varias formas de registros: el código de estado de tiempo presente, el código de estado de tiempo pasado y el registro de fenómenos, cada uno con sus características específicas. En estudios anteriores (Álvarez *et al.*, 2009) se concluyó que los registros más adecuados para el trabajo con los fenómenos nieblas, neblinas y tormentas son los de código de tiempo presente. Este análisis puede extenderse a los demás fenómenos clasificados mediante el código mencionado.

Al emprender el estudio del comportamiento de cualquier variable climática, es necesario contar con series de registros suficientemente largas en cuanto al número de años, completos y confiables, y lograr que las series construidas sean homogéneas, o sea, que sean representativas de la población que describe el parámetro de análisis; sin embargo, los registros de código de tiempo presente no son siempre correctos ni completos, por lo cual es necesario examinar la calidad de los datos para saber qué estudios pueden emprenderse a partir de estos y cuáles son sus limitaciones.

Para su uso, la Organización Meteorológica Mundial (OMM, 1992) sugiere que los datos meteorológicos deben pasar tres etapas fundamentales de control de calidad: la verificación de errores groseros, la verificación de la coherencia interna y de la coherencia temporal, y la verificación de la coherencia espacio-temporal. Verificar y completar las series de las variables meteorológicas medidas en las estaciones ha sido motivo del trabajo de muchos investigadores (Gandin, 1988; Graybeal *et al.*, 2004b; Feng *et al.*, 2004; Hubbard *et al.*, 2005; You *et al.*, 2008 y Araya, 2011). Las variables más estudiadas son: la temperatura, la precipitación, la humedad y el viento; las verificaciones se han realizado entre variables de superficie y, en algunos casos, con variables de aire superior.

No obstante, el código de estado de tiempo presente no ha sido de los más favorecidos y, en ocasiones, se ha utilizado como criterio, junto con otros elementos y metadatos, para perfeccionar y completar registros de otras variables (Graybeal *et al.*, 2004a). Dai (2001) plantea que los reportes de tiempo presente presentan a nivel global un contraste evidente en cantidad y calidad para distintas regiones; por ello, se recomienda que los registros mencionados se interpreten con precaución.

El objetivo del presente trabajo es evaluar los criterios de calidad de las series de código de estado de tiempo presente como una variable clasificadora de fenómenos, determinada en 69 estaciones del territorio cubano, y sugerir períodos y horarios para estudios climáticos posteriores.

Materiales y métodos

La base fundamental de información utilizada para el estudio la constituyen los registros de código de estado de tiempo presente de 69 estaciones a lo largo y ancho del territorio cubano. La tabla 1 muestra la identificación de la estación y algunas características de las series utilizadas. Como puede observarse, las series varían en largo, entre 5 años y 41 años, con un largo promedio de 34.4 años, aunque existen importantes faltantes de información, con 12 estaciones por debajo de 70 % de información útil y un promedio general de 84.1 %. Para todos los registros en estudio, los faltantes de información están, en promedio, entre 5 días y 8 días al mes, con un máximo, en febrero, de 7.5 días.

Con respecto a los datos por horas del día, los faltantes con menos de 30 % de la información posible mostraron horarios menos favorecidos a las 01:00 h, 04:00 h (máximo con 170 días promedio al año de 365 posibles) y 22:00 hora local, mientras que para los horarios diurnos, el faltante promedio oscila entre 7 días y 15 días al año.

– **Artículos originales:** Análisis de la calidad de series largas de registros
de código de estado del tiempo presente para las estaciones de Cuba –

Tabla 1. Características de la información de las estaciones en estudio

Número de la estación	Nombre de la estación	Año Inicio	Año Final	Largo de la serie	Datos posibles	Datos Válidos	Información útil (%)
308	La Piedra	2006	2010	5	14 608	14608	100
309	Amistad Cuba Francia	2003	2010	8	23 376	23080	98.7
310	Cabo San Antonio	1970	2010	41	119 800	107190	89.5
312	Santa Lucía	1981	2010	30	87 664	80964	92.4
313	Isabel Rubio	1970	2010	41	119 800	110577	92.3
314	San Juan y Martínez	1970	2010	41	119 800	94389	78.8
315	Pinar del Río	1978	2010	33	96 424	92372	95.8
316	La Palma	1971	2010	40	116 880	101688	87
317	Paso Real de San Diego	1970	2010	41	119 800	114102	95.2
318	Bahía Honda	1977	2010	34	99 352	93614	94.2
319	Valle de Cuajeri	1992	2010	19	55 528	38558	69.4
320	Güira de Melena	1970	2010	41	119 800	93515	78.1
321	La Fe	1970	2010	41	119 800	110242	92
322	Batabanó ¹	1986	2010	24	73 048	56537	77.4
323	Güines	1971	2010	40	116 880	77 320	66.2
324	Punta del Este	1971	2010	40	116 880	90 062	77.1
325	Casablanca	1972	2010	39	113 968	113 894	99.9
326	Santo Domingo	1980	2010	31	90 592	86 635	95.6
327	Unión de Reyes	1977	2010	34	99 352	95 589	96.2
328	Varadero	2003	2010	8	23 376	23 369	100
329	Indio Hatuey	1970	2010	41	119 800	79 041	66
330	Jovellanos	1970	2010	41	119 800	107 999	90.1
331	Jagüey Grande ²	1977	2010	32	99 352	69 099	69.5
332	Colón	1975	2010	36	105 192	100 037	95.1
333	Playa Girón	1970	2010	41	119 800	114 416	95.5
334	Palenque Yateras ³	1993	2010	18	52 600	51 136	97.2
335	Aguada de Pasajeros	1975	2010	36	105 192	76 918	73.1
337	Trinidad	1970	2010	41	119 800	98 025	81.8
338	Sagua la Grande	1970	2010	41	119 800	106 629	89
339	Cayo Coco	1990	2010	21	61 360	60 874	99.2
340	Bainoa	1979	2010	32	93 504	69 538	74.4
341	Jíbaro	1977	2010	34	99 352	89 514	90.1
342	Topes de Collantes	1970	2010	41	119 800	88 216	73.6
343	El Yabu	1977	2010	34	99 352	93 475	94.1
344	Cienfuegos	1975	2010	36	105 192	92 158	87.6
345	Júcaro	1970	2010	41	119 800	84 739	70.7
346	Venezuela	1977	2010	34	99 352	85 827	86.4
347	Camilo Cienfuegos	1977	2010	34	99 352	86 661	87.2
348	Caibarién	1970	2010	41	119 800	107 336	89.6
349	Sancti Spiritus	1970	2010	41	119 800	106 324	88.8
350	Florida	1970	2010	41	119 800	94 589	79
351	Santa Cruz ⁴	1970	2010	38	119 800	84 125	70.2
352	Esmeralda	1970	2010	41	119 800	92 236	77
353	Nuevititas	1970	2010	41	119 800	99 681	83.2
354	Palo Seco	1976	2010	35	102 280	73 071	71.4
355	Camaguey	1970	2010	41	119 800	119 779	100
356	Jamal	1992	2010	19	55 528	47 350	85.3
357	Las Tunas	1975	2010	36	105 192	90 056	85.6
358	Puerto Padre	1975	2010	36	105 192	93 704	89.1
359	Manzanillo	1975	2010	36	105 192	86 954	82.7
360	Cabo Cruz	1970	2010	41	119 800	113 947	95.1
361	Jucarito	1976	2010	35	102 280	66 076	64.6
362	La Jiquima	1975	2010	36	105 192	73 528	69.9
363	Contramaestre	1977	2010	34	99 352	84 209	84.8
364	Santiago de Cuba	1970	2010	41	119 800	102 834	85.8
365	Punta Lucrecia	1970	2010	41	119 800	108 009	90.2
366	Gran Piedra	1970	2010	41	119 800	118 507	98.9
367	Ped. de Holguín ⁵	1993	2010	18	52 600	49 390	93.9
368	Guantánamo	1982	2010	29	84 736	80 743	95.3
369	Maisí	1970	2010	41	119 800	113 645	94.9
370	Guaro	1975	2010	36	105 192	85 405	81.2
371	Pinares de Mayarí	1975	2010	36	105 192	83 931	79.8
372	La Sabana ⁶	1979	1993	15	43 832	26 158	59.7
373	Santiago de las Vegas ⁷	1970	2010	40	119 800	83 535	69.7
374	Tapaste	1976	2010	35	102 280	69 776	68.2
375	Melena del Sur	1974	2010	37	108 112	77 983	72.1
376	Bauta	1974	2010	37	108 112	72 221	66.8
377	Veguita	1977	2010	34	99 352	67 316	67.8
378	Velazco	1976	2010	35	102 280	67 215	65.7

Notas: ¹ A la estación Batabanó (322) le falta el año 1991 completo; ² A la estación Jagüey Grande (331) le faltan 2006 y 2007 completos; ³ La estación Palenque de Yateras (334) tiene dos períodos 1970–1978 y 1993–2008; ⁴ A la estación Santa Cruz (351) le faltan 1983, 1984 y 1985 completos; ⁵ Aunque la estación Pedagógico de Holguín tiene, según el código de la OMM, el número 372, aquí se usará 367 para que no se confunda con La Sabana; ⁶ La estación La Sabana (372) estuvo activa de 1979 a 1993 y su número pasó a la estación Pedagógico de Holguín, que empezó en 1993; ⁷ A la estación Santiago de las Vegas (373) le falta 2007 completo.

Los datos fueron tomados de la base de datos Nueva Thor (Álvarez *et al.*, 2012b) a la cual se agregaron los registros correspondientes al período 2009–2010, obtenidos de la base de datos del Centro del Clima del INSMET, datos estos que fueron sometidos a varias validaciones y adaptaciones con miras a lograr la uniformidad en los registros de la base.

La metodología para determinar la calidad de la información está dirigida a analizar el completamiento de las series de registros y el sesgo que la falta de datos podría introducir en su distribución. Para ello, se utilizó la metodología desarrollada por Álvarez y colaboradores (2009), adaptada con variaciones pequeñas, según lo planteado por Álvarez y colaboradores (2012a), basada en la determinación del largo de la serie, la información utilizable, y los sesgos mensual y horario. Los cuatro indicadores de la calidad de las series que se evaluarán estarán dados por el largo de la serie (ILargo), el porcentaje de información útil (IU), el sesgo mensual (Sm) y el sesgo horario (Sh); y la calidad general de la serie (ICALI) se define por el promedio de los cuatro índices mencionados.

Resultados y discusión

De acuerdo con la metodología referenciada en el epígrafe anterior, puede resumirse la calidad de las series por estaciones en cuanto a registros de código de estado de tiempo presente (Tabla 3).

En la tabla 3 se aprecia que, con respecto al largo de las series, la mayoría califica de excelente o muy buena, y quedan solo diez series con calificativo de buena o inferior, por lo general, estaciones que cerraron años atrás o de creación reciente. Con respecto a la información útil, once series fueron calificadas como regulares y una como mala a causa de, fundamentalmente, registros perdidos o no medidos en meses específicos o en horarios de la noche y la madrugada. El sesgo mensual no constituye un pro-

blema, puesto que todas las estaciones muestran calificadores entre excelente y muy bueno. El sesgo horario es el calificador que acumula más series con resultados regulares o malos (64 %), dado que en los años anteriores a 2000, muchas estaciones tenían regímenes de observación que solo abarcaban el horario diurno. Esta ausencia de información influye en índices tales como el de la información útil y debe tenerse en cuenta para obtener conclusiones acerca de la ocurrencia de los fenómenos en estudio, por lo cual se hace necesario determinar las horas de máxima ocurrencia de cada uno con vistas a evaluar la pertinencia de utilizar o no los datos.

Según los índices analizados, y dándoles a todos la misma importancia, las series por estaciones se califican de la manera siguiente: diez excelentes, 42 muy buenas, 16 buenas y solo una regular, lo cual arroja que 98.5 % de las estaciones calificaron sus series como buenas o con categorías superiores.

En general, teniendo en cuenta las limitaciones impuestas por la ausencia de la información, las series analizadas pueden utilizarse para describir particularidades de cada fenómeno en estudio; sin embargo, para la marcha interanual, las series deben poseer la información más completa posible de año en año, aunque ello no implique toda la información registrada.

Si se analizan, de conjunto, todos los registros de tiempo presente de la base Nueva Thor, teniendo en cuenta que las primeras estaciones comenzaron a reportar en 1970 y que al cierre del estudio existen 69 estaciones, puede obtenerse el porcentaje de completamiento de la información para diferentes períodos, al dividir los datos válidos entre los posibles (Fig. 1). Se observa que a medida que se acorta el período hacia el final de la serie, mayor resulta el porcentaje de información válida; de este modo, en el período 2005–2010 se alcanza el máximo, con 98 %.

Si se escogiese un período de estudio basado en el completamiento de la información, sería recomendable utilizar el período 1989–2010, puesto que serían

– **Artículos originales:** Análisis de la calidad de series largas de registros
de código de estado del tiempo presente para las estaciones de Cuba –

Tabla 3. Calificadores por índices para cada serie en estudio de código de estado de tiempo presente por estaciones

Estación	ILARGO	IU	Sm	Sh	ICALI	Estación	ILARGO	IU	Sm	Sh	ICALI
308	M	E	E	E	MB	345	E	B	E	M	MB
309	M	E	MB	E	MB	346	E	MB	E	R	MB
310	E	MB	E	R	MB	347	E	MB	E	R	MB
312	E	E	MB	R	MB	348	E	MB	E	M	MB
313	E	E	E	R	MB	349	E	MB	MB	R	MB
314	E	B	E	M	MB	350	E	B	E	M	MB
315	E	E	E	MB	E	351	E	B	MB	M	B
316	E	MB	E	R	MB	352	E	B	E	M	MB
317	E	E	MB	B	MB	353	E	B	E	M	MB
318	E	E	MB	B	MB	354	E	B	E	M	MB
319	B	R	MB	M	B	355	E	E	E	E	E
320	E	B	E	M	MB	356	B	MB	MB	E	MB
321	E	E	E	B	E	357	E	MB	MB	M	MB
322	B	B	E	M	B	358	E	MB	E	M	MB
323	E	R	E	M	B	359	E	B	E	M	MB
324	E	B	MB	M	B	360	E	E	MB	B	MB
325	E	E	E	E	E	361	E	R	E	M	B
326	E	E	MB	MB	E	362	E	R	MB	M	B
327	E	E	MB	MB	E	363	E	B	MB	R	MB
328	M	E	E	E	MB	364	E	MB	E	R	MB
329	E	R	E	M	B	365	E	E	MB	B	MB
330	E	E	E	B	E	366	E	E	E	E	E
331	E	R	E	M	B	367	B	E	MB	E	MB
332	E	E	E	MB	E	368	MB	E	MB	MB	MB
333	E	E	MB	B	MB	369	E	E	MB	MB	E
334	B	E	MB	E	MB	370	E	B	MB	M	B
335	E	B	E	M	MB	371	E	B	E	M	MB
337	E	B	E	M	MB	372	B	M	MB	M	R
338	E	MB	MB	B	MB	373	E	R	E	M	B
339	B	E	MB	E	MB	374	E	R	E	M	B
340	E	B	E	M	MB	375	E	B	E	M	MB
341	E	E	E	M	MB	376	E	R	E	M	B
342	E	B	MB	M	B	377	E	R	E	M	B
343	E	E	MB	B	MB	378	E	R	E	M	B
344	E	MB	E	M	MB						

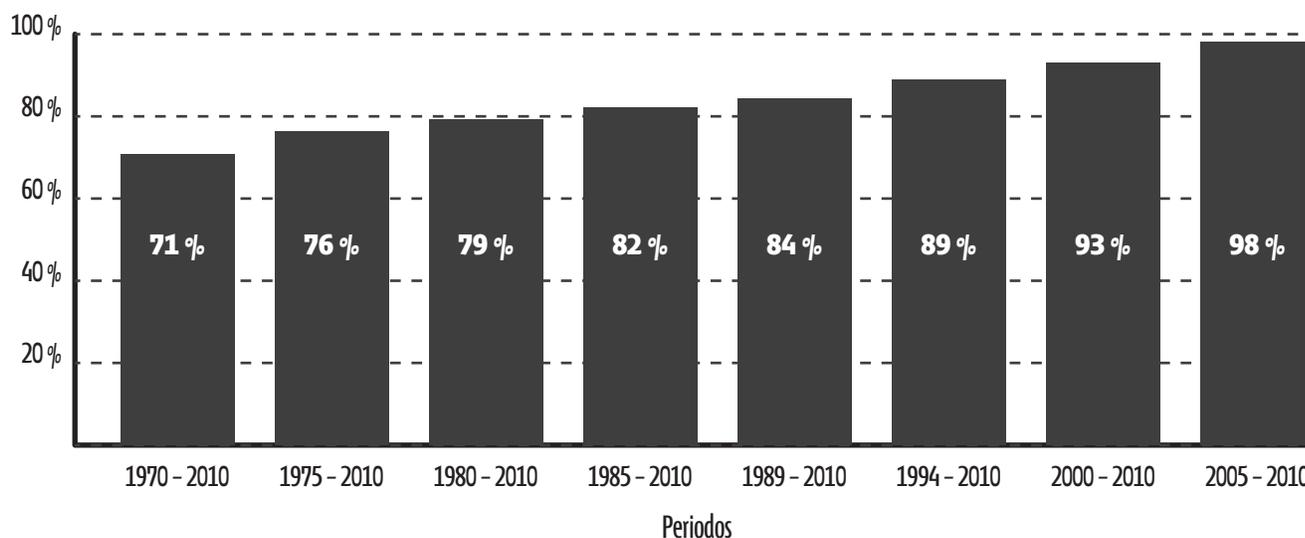


Fig. 1 Porcentaje de observaciones válidas con respecto a las posibles para diferentes períodos de la información contenida en la base de datos Nueva Thor.

22 años de información, hacia la parte final de la serie y 84 % de los registros posibles. Para estudios relativos a la marcha diaria, el período analizado debe reducirse a las series entre 2005 y 2010, con solo seis años de información, pero con 98 % de información real con respecto a la información posible.

Conclusiones y recomendaciones

De las series por estaciones de código de estado de tiempo presente, 98.5 % puede calificarse como buenas, muy buenas o excelentes con respecto a la calidad de la información contenida.

El calificador sesgo horario acumula 64 % de las series con resultados regulares o malos dada la falta de registros en horarios de la noche y la madrugada durante varios años consecutivos.

Para estudios climáticos como la representatividad de los fenómenos, la marcha anual y la distribución espacial, se recomienda trabajar con los datos del período 1989-2010.

Para los estudios de marcha diaria, se recomienda el período 2005-2010, por la ausencia de registros en la noche y la madrugada.

Referencias

- Álvarez, L, Borrajero, I, Álvarez, R. y Rojas, Y, 2009: 'Estudio preliminar del fenómeno niebla a partir del registro de fenómenos y su relación con la variable código de estado del tiempo presente', *Memorias del V Congreso Cubano de Meteorología*, Ciudad Habana, del 30 de noviembre al 4 de diciembre de 2009 (publicación electrónica), ISBN 978-959-7167-20-4, <F:\Data\Trabajos completos.pdf>, pp. 1456-1466.
- Álvarez, L, Borrajero, I, Álvarez, R, Aenlle, L, Rivero, I, Iraola, C, Rojas, Y. y Hernández, M, 2012a: 'Estudio de la marcha interanual de la frecuencia de ocurrencia de observaciones con tormenta para el territorio cubano', *Revista de Climatología*, vol. 12, pp. 1-21, <<http://webs.ono.com/reclim/reclim12a.pdf>>.
- Álvarez, L, Borrajero, I, Álvarez, R, Aenlle, L. y Barceñas, M, 2012b: 'Actualización de la distribución espacial de las tormentas eléctricas en Cuba', *Revista Cubana de Meteorología*, vol. 18, no. 1, pp. 83-99.
- ARAYA, J. L, 2011: 'Resultados de un control de calidad de datos de temperatura superficial', *Tecnología en Marcha*, vol. 24, no. 1, pp. 33-49.
- DAI, A, 2001: 'Global Precipitation and Thunderstorm

- Frequencies. Part I: Seasonal and Interannual Variations', *Journal of Climate*, vol. 14, no. 6, pp. 1092-1111.
- FENG, S, HU, Q. AND QIAN, W, 2004: 'Quality Control of Daily Meteorological Data in China, 1951-2000: A new dataset', *International Journal of Climatology*, vol. 24, pp. 853-870.
- GANDIN, L. S, 1988: 'Complex quality control of meteorological data', *MWR*, vol. 116, pp. 1137-1156.
- GRAYBEAL, D. Y, DEGAETANO, A. T. Y EGGLESTON, K. L, 2004a: 'Complex Quality Assurance of Historical Hourly Surface Airways Meteorological Data', *J. Atmos. Oceanic Technol*, vol. 21, pp. 1156-1169.
- GRAYBEAL, D. Y, DEGAETANO, A. T. AND EGGLESTON, K. L, 2004b: 'Improved Quality Assurance for Historical Hourly Temperature and Humidity: Development and Application to Environmental Analysis', *Journal of Applied Meteorology*, vol. 43, pp. 1722-1735.
- HUBBARD, K. G, GODDARD, S, SORENSEN, W. D, WELLS, N. AND OSUGI, T. T, 2005: 'Performance of Quality Assurance Procedures for an Applied Climate Information System', *J. Atmos. Oceanic Technol*, vol. 22, pp. 105-112.
- OMM, 1992: *Manual del Sistema Mundial de Proceso de Datos y de Predicción*, OMM, no. 485, vol. II: Aspectos regionales, Ginebra, Suiza.
- YOU, J, HUBBARD, K. G. AND GODDARD, S, 2008: 'Comparison of methods for spatially estimating station temperatures in a quality control system', *International Journal of Climatology*, vol. 28, no. 6, pp. 777-787.