

Algoritmos de control de calidad para detección de secuencias atípicas de contactos en pluviómetros de balancín

Fecha de recepción: 10/03/2008

Fecha de aceptación: 06/06/2008

José L. Araya¹

Eric J. Alfaro²

Palabras clave

Control de calidad, estación meteorológica automática, programación en grabador de datos, pluviómetro de balancín, secuencia de tiempo entre contactos, obstrucción parcial.

Key words

Quality control, automatic weather stations, Data Logger programming, tipping bucket, inter-tip time sequence, partial blockage.

Resumen

Se prueban diversos Algoritmos de Control de Calidad de Datos (ACCD) para precipitación, con el fin de sugerir metodologías alternativas para detección de posibles valores atípicos generados por pluviómetros de balancín en condiciones operativas. Tres diferentes ACCD fueron probados, uno de ellos para la detección de tiempos entre contactos atípicos y otros dos basados en la detección de

secuencias de tiempos entre contactos. Las diversas pruebas que fueron llevadas a cabo muestran que estas pruebas pueden ser útiles en la detección de tiempos entre contactos atípicos y obstrucciones parciales, detectando situaciones de obstrucción en el 70% de los casos. Los resultados obtenidos en condiciones controladas señalan que es posible utilizar estos algoritmos en tiempo real en forma conjunta, como una forma de aumentar la posibilidad de detección de valores sospechosos.

Abstract

Quality control algorithms are tested in order to present a statistical approach for dealing with atypical values generated by rain gauges under operational conditions. Three different quality control algorithms were tested, one of them for detecting atypical inter-tip times and two of them for outlying suspicious inter-tip time sequences. As part of this research, tests under laboratory-controlled conditions were made to verify their performance.

1. Meteorólogo. Instituto Meteorológico Nacional. Tel. 2222-5616. Correo electrónico: jl araya@imn.ac.cr
2. Profesor e investigador. Escuela de Física, Centro de Investigaciones Geofísicas (CIGEFI) y Centro de Investigación en Ciencias del Mar y Limnología (CIMAR), Universidad de Costa Rica. Tel. 2511-5096. Correo electrónico: ealfaro@cosmos.ucr.ac.cr

The data sets generated show this simple setup is adequate to reproduce situations that induce partial blockages. Around 70% of the induced partial blockages were detected by the quality control algorithms. Apart from that, results show the three tests can be used together to have a higher degree of success in detecting suspicious values.

Los resultados obtenidos en condiciones controladas señalan que es posible utilizar estos algoritmos en tiempo real en forma conjunta, como una forma de aumentar la posibilidad de detección de valores sospechosos

Introducción

El control de calidad de la información meteorológica es de gran importancia. Esto resulta claro dentro del contexto de la automatización de las mediciones meteorológicas, donde el volumen creciente de datos hace muy difícil llevar a cabo el control de calidad de forma manual en todos los niveles. En particular, el objetivo de esta investigación es mostrar algunas metodologías para la revisión de los datos generados por pluviómetros de balancín.

Este artículo describe algunas características básicas del equipo de medición aplicado, así como las fuentes de error comunes. Seguidamente se discuten las pruebas realizadas así como los algoritmos programados. Se finaliza con los resultados obtenidos de estas pruebas.

Descripción del pluviómetro de balancín

Los modelos de pluviómetros de balancín TE 525MM y TE 525 para la medición de la lluvia son los que se han implementado en la red del Instituto Meteorológico Nacional (IMN) desde que se instaló la red de estaciones meteorológicas automáticas. La figura 1 muestra un esquema general del instrumento. La diferencia entre los modelos TE 525MM y TE 525 se encuentra, básicamente, en que el primero presenta un área de captura mayor, lo que le permite obtener muestras instantáneas de datos con mayor resolución.

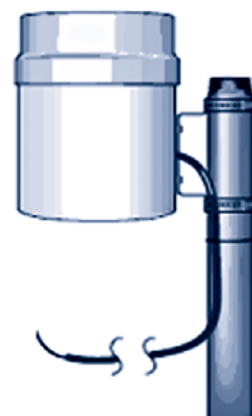


Figura 1. Pluviómetro de balancín modelo TE 525I (Campbell Sci., 2004b).

Este tipo de instrumentos que incluyen un balancín generan un impulso eléctrico digital mediante de la oscilación de un balancín que se encuentra en equilibrio inestable en el interior de la cubeta receptora (Figura 2), conforme las gotitas de lluvia llenan cada uno de los compartimentos de este balancín (Ureña, 1998). Cuando uno de estos compartimentos se llena con un volumen de agua de 0,1mm se genera una señal que es registrada por un canal de pulsos en el grabador de datos (GD) (Campbell Sci., 2004b). El error en la subestimación de lluvia a altas intensidades es dependiente del diseño del instrumento, y está relacionado con el hecho de que el contacto del balancín no es instantáneo, por lo que hay una pérdida de agua durante el movimiento del balancín. De hecho, la duración del contacto determina la inercia del sistema (el momento de inercia relacionada con el balancín) y al momento generado por el agua dentro de la cubeta (Alexandropoulos y Lacombe, 2005).

Existen varios errores de diseños asociados a los sensores de precipitación de depósito basculante o de balancín (Brock y Richardson, 2001), los cuales son:

Magnitud de la velocidad del viento: El flujo de viento tiende a causar deflexión

El error en la subestimación de lluvia a altas intensidades es dependiente del diseño del instrumento, y está relacionado con el hecho de que el contacto del balancín no es instantáneo, por lo que hay una pérdida de agua durante el movimiento del balancín.

en las gotitas de agua que podrían haber precipitado sobre el área de captura del pluviómetro de balancín; esto provoca una subestimación en la cantidad de precipitación que es función del tipo de lluvia, el diámetro de las gotitas, la magnitud de la velocidad del viento en el momento del evento de precipitación y el tipo de protector de viento utilizado, cuando se le ha instalado alguno.

Salpicadura: Cuando gotitas de cierto tamaño golpean la parte superior del instrumento, tiende a darse un efecto de salpicadura hacia afuera del receptáculo, que puede producir que parte del agua precipitada no sea registrada por el sensor.

Obstrucción del embudo: El embudo que habilita el área de captura puede ser obstruido por maleza, hojarasca, animales pequeños, etc., lo cual imposibilita el paso del agua hacia el balancín.

Humedecimiento del embudo y evaporación: Cuando se dan situaciones en las que una lluvia leve precipita, puede darse el caso de que esta agua se mantenga en la superficie del embudo, retardando la hora de inicio del evento, por lo cual se desconocen las horas de inicio y fin de una lluvia tenue (Suárez, 1996). Si la superficie del embudo se encuentra caliente, parte de esta agua podría además perderse por evaporación.

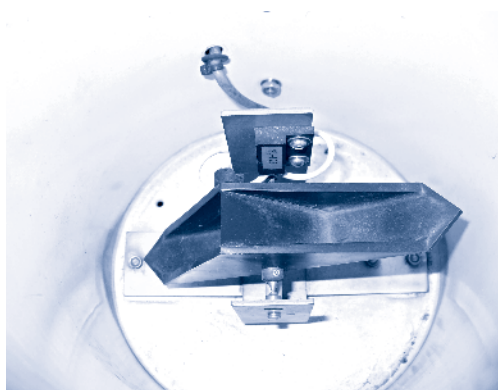


Figura 2. Detalle de las cazoletas para un pluviómetro de balancín TE 525MM.

Pruebas de algoritmos de precipitación

Los algoritmos descritos a continuación no fueron instalados en ninguna de las estaciones meteorológicas automáticas (EMA) del IMN, utilizadas en esta investigación, debido a que la naturaleza de los errores que detectan es tan especial que la única manera de tener plena certeza de su correcto funcionamiento era someterlos primero a diversas pruebas en el laboratorio, ya que son algoritmos cuyas premisas se basan en suposiciones muy específicas sobre las secuencias de tiempos entre contactos, susceptibles de ser simuladas y de ser mejor entendidas en el laboratorio antes de hacerlos operativos en la red. Es importante tener en cuenta que el objetivo de las pruebas de los algoritmos para pluviómetros de balancín que serán descritas es determinar si los algoritmos programados en el GD eran consistentes.

Los materiales utilizados en estas pruebas fueron:

- Programa para prueba de algoritmos.
- Recipiente calibrador conteniendo 16 onzas (473mL) de agua.
- Nivel.
- Pluviómetro de balancín TE 525 MM. Este mismo aparato fue utilizado durante toda la prueba.
- GD modelo CR10X
- Batería de 12V-7 A.
- Regulador de voltaje.
- Interfase SC 32 y cables de conexión.
- Teclado portátil para visualización de datos CR10KD.

Además, se tomaron las siguientes precauciones:

- Nivelación apropiada del instrumento.
- Limpieza previa del balancín y embudo, con el fin de remover

cualquier material ajeno al sistema que pudiera influir en la prueba, tal como polvo, grasa, etc.

- Prueba básica de calibración, la cual se realizó antes de iniciar la prueba de los tres algoritmos de lluvia. Se trató de un método simple de calibración estática, en el cual el instrumento es nivelado y ajustado, de modo que un volumen específico de agua induce cierto número de contactos en el balancín (Humphrey *et al.*, 1997).

Detección de valores atípicos en el registro de lluvia

Las siguientes pruebas han sido utilizadas para la detección de errores en registros de datos con un solo pluviómetro de balancín. Upton y Rahimi (2003) han sugerido varias pruebas para comprobación del funcionamiento de pluviómetros de balancín, considerándolos ya sea como instrumentos aislados o como parte de una red. Los algoritmos programados con los que se efectuaron estas pruebas pueden ser encontrados en Araya (2007). Estas pruebas son útiles cuando se tienen varios detectores en una región relativamente pequeña. Para efectos de esta investigación, se pretende analizar aquellas pruebas que permitan abanderar valores sospechosos por medio de varios algoritmos de programación.

Detección de una frecuencia alta de contactos

Descripción

Esta prueba fue desarrollada con el fin de distinguir intervalos entre contactos atípicos. Cuando el balancín genera un contacto, puede producirse un rebote que, en caso de no estar el sistema perfectamente ensamblado, generará un segundo contacto erróneo. Para detectar esto, se supone que la razón de precipitación que se presenta en cierto instante se asemejará a la del próximo instante (Upton y Rahimi, 2003).

Sea $t_0, t_1, t_3, \dots, t_n$ los tiempos en los cuales un pluviómetro de balancín reporta un contacto, se define el periodo entre contactos como

$$\tau_k = t_k - t_{k-1} \quad (1)$$

Si se supone que la razón de lluvia varía poco, entonces;

$$\tau_k = t_k \approx \tau_{k-1} \quad (2)$$

Para efectos de los trópicos, el periodo entre contactos puede variar dependiendo de la intensidad de la lluvia, lo cual puede resolverse considerando la razón de intervalos entre contactos adyacentes en lugar de la diferencia. De allí que conviene definir un parámetro λ_k de la siguiente manera:

$$\lambda_k = \left| \ln \left(\frac{\tau_k}{\tau_{k-1}} \right) \right| \quad (3)$$

Cualquier contacto se rechaza si se cumple la relación

$$\lambda_k > \delta, \quad (4)$$

donde δ es una constante que se debe determinar.

La figura 3 muestra la distribución de λ_k para diferentes periodos entre contactos. El eje de las abscisas muestra el tiempo entre contactos sospechosos, mientras que el eje de las ordenadas muestra el valor de λ_k correspondiente. Cada una de las curvas corresponde a una secuencia de tiempos entre contactos uniformes que van desde 1 segundo hasta 300 segundos.

De la figura 3 se puede observar como el parámetro λ_k es un buen indicador del tiempo entre contactos consecutivos, ya que permite saber si se presenta un tiempo entre contactos anormalmente corto. Por ejemplo, supóngase que en un aguacero el sensor de precipitación registra un contacto cada 300 s (recuérdese que se asume que los tiempos entre contactos son similares) y que después de uno de estos contactos se registra otro contacto 2 s después. En

Cuando el balancín genera un contacto, puede producirse un rebote que, en caso de no estar el sistema perfectamente ensamblado, generará un segundo contacto erróneo

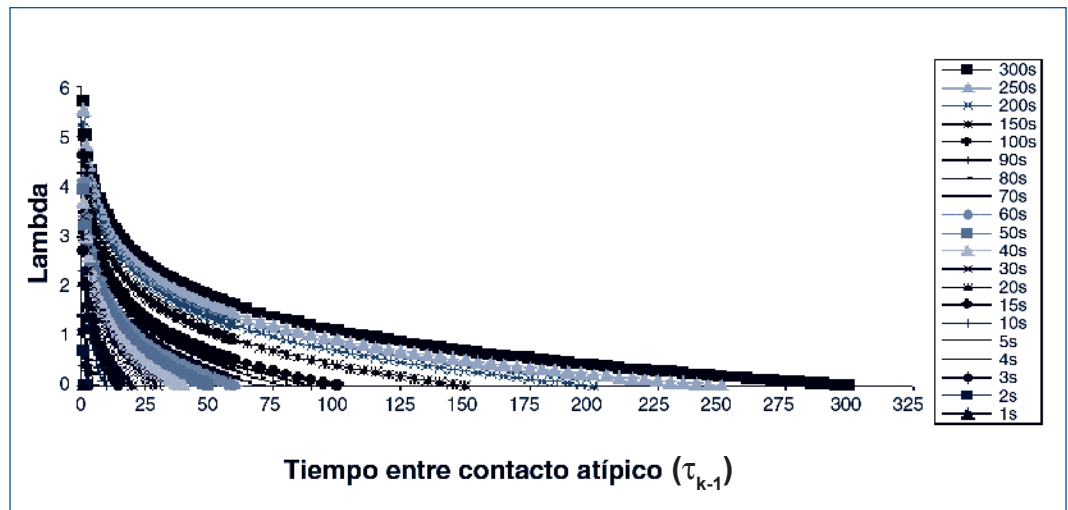


Figura 3. Variación de lambda para tiempos entre contactos atípicos suponiendo secuencias de tiempos entre contactos uniformes (La leyenda muestra los valores τ_k).

este caso, el valor calculado para λ_k es 5,01. El algoritmo abanderaría el tiempo atípico entre contactos que es menor o igual a dos segundos en una secuencia uniforme de 300 s. Cuanto menor sea el tiempo entre contactos, menor debe ser el tiempo atípico entre contactos para que sea detectado. En realidad, la elección de $\lambda_k = 5$ sugerida por Upton y Rahimi (2003) permite abanderar un tiempo entre contactos muy corto cuando este represente el 0,673% o menos del tiempo entre contactos de mayor duración anterior o posterior a este.

Según se desprende del análisis de la ecuación 3, la elección de λ_k también depende del intervalo de ejecución mínimo de detección del GD; el algoritmo solo detectará aquellos valores dentro de la resolución en la que el GD es capaz de detectar los pulsos del sensor de lluvia. En el caso de los GDs Campbell CR10X toleran un intervalo de ejecución de mínimo de 0,015625s. Cuanto menor sea el intervalo de ejecución del programa, mayor capacidad tendrá el algoritmo de detectar un tiempo entre contactos que difiere, sustancialmente, del anterior. Para

la presente investigación el intervalo de interrogación utilizado fue 1 s.

Datos utilizados y metodología

La base de datos del IMN no cuenta con datos que permitan la validación de este algoritmo, ya que nunca se ha medido el tiempo entre contactos generado por el GD. Es por ello que para todos los algoritmos de lluvia aquí descritos, se desarrollaron pruebas para generar los datos en el laboratorio por medio de contactos inducidos, en condiciones muy idealizadas, como se explicó anteriormente. En el cuadro 1 se muestra el número total de pruebas que se realizaron. Dos tipos de pruebas fueron realizadas para corroborar la consistencia del algoritmo: Las primeras fueron realizadas mediante la generación de contactos con razón de cambio en la precipitación constante, utilizando una bomba de precisión para tal fin. En esta fase se puso a prueba la consistencia del algoritmo de medir tiempos atípicos entre contactos usando la ecuación (3).

La segunda fase fue llevar a cabo pruebas manualmente: la metodología seguida fue generar contactos con tiempos

cronometrados, de modo que la frecuencia de contactos fuera aproximadamente constante, e inmediatamente se inducía manualmente un contacto cuyo tiempo entre el último contacto difería notablemente del anterior, de modo que hubiera certeza de que el valor de λ_k sobrepasaría el valor umbral definido y de esta forma probar que el algoritmo trabajaba correctamente. El ACCD se probó con dos valores diferentes para λ_k . Además, nótese que en los datos de las pruebas manuales realizadas con la bomba no habría habido diferencia entre la usar $\lambda_k = 3$ ó $\lambda_k = 5$, ya que la diferencia del orden de magnitud de un tiempo entre contactos “normal” y uno atípico es muy grande, de modo que se habría dado detección de la misma forma, independientemente de estos valores utilizados.

Resultados

La figura 4 muestra los resultados para uno de los experimentos realizados con el fin de probar este algoritmo. Nótese en

los ejes de las abscisas que la experiencia mostrada se inicia a las 17:38 h y se extiende hasta las 18:05 h. Durante este periodo, se almacenaba la información cada segundo, por lo que fue necesario llevar a cabo un filtrado de la información en el tiempo, con el fin de mostrar los detalles importantes en la figura 4. La parte a) indica la magnitud del tiempo entre contactos registrados; nótese que la secuencia de tiempo entre contactos en general se mantiene entre 125 s y 178 s. Las disminuciones abruptas en la línea de la curva son debidas a la reducción sustancial del tiempo del último contacto con respecto al anterior; estos son los contactos atípicos artificialmente inducidos para que fueran detectados por el algoritmo. La parte b) de la misma figura muestra el contador de los valores, donde λ_k excede el límite máximo impuesto. Nótese que por cada tiempo entre contactos anormalmente corto en la parte a) de la figura 4 el contador se incrementa en dos unidades en la parte b), indicando que este

Cuadro 1. Pruebas de laboratorio realizadas para el algoritmo de detección de contactos atípicos en pluviómetros de balancín

N.º pruebas	Fecha	Hora inicio	Hora final	Rango para lambda	Tiempos entre contactos (s)
1	02/02/2006	17:23	17:48	5	35
2	02/02/2006	16:35	16:56		45
3	02/02/2006	16:02	16:21		50
4	02/02/2006	15:41	15:47		55
5	08/02/2006	15:52	16:04		65
6	08/02/2006	16:12	16:24		75
7	08/02/2006	16:31	16:43		85
8	08/02/2006	16:51	17:02		95
9	13/03/2006	13:57	15:52		300
10	11/05/2006	15:44	16:26	3	150-200
11	11/05/2006	16:31	17:36		231-286
12	11/05/2006	17:38	18:05		125-178
13	12/05/2006	15:10	16:51		366-438
14	12/05/2006	16:59	18:56		422-805

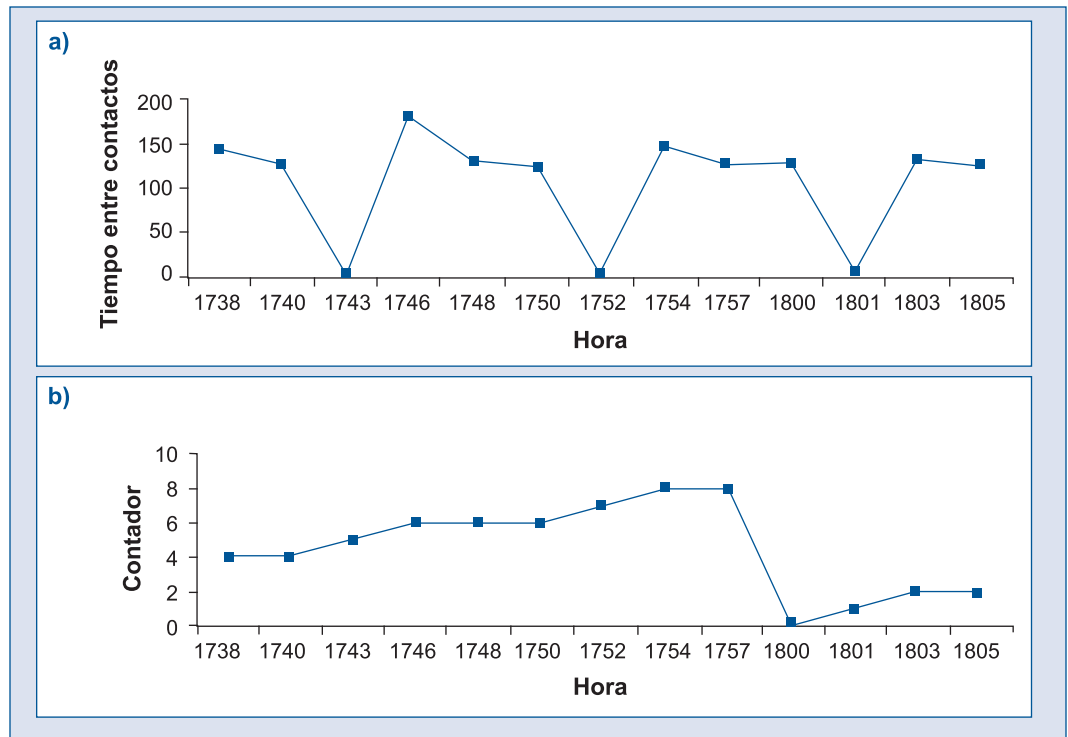


Figura 4. a) Secuencia de tiempos entre contactos aproximadamente uniformes con tiempos entre contactos inducidos anormalmente cortos. b) Variación del contador de tiempos entre contactos atípicos (Prueba 12 del Cuadro 1).

almacena información en cuanto a los tres picos abruptos en la secuencias de tiempos entre contactos mostrada en dicha parte a). Obsérvese también que el contador para valores fuera de rango se inicia exactamente al comienzo de cada hora, por lo que el contador siempre comienza asumiendo un valor nulo.

Esto es consistente con el hecho de que el algoritmo fue programado para que el contador dé información en resolución horaria acerca de cuántas veces se presenta un tiempo anormal entre contactos. Nótese, además, que, dada una serie de tiempos entre contactos aproximadamente iguales, un contacto atípico que difiera, sustancialmente, del resto será contado dos veces, suponiendo tiempos entre contactos similares entre sí antes y después de este. Es importante señalar en este punto que lo único que se inicia es el valor del contador; a nivel interno, el algoritmo sigue

ejecutándose. Igualmente, sucede con los otros dos algoritmos de precipitación que han sido considerados.

Detección de pluviómetros de balancín parcialmente bloqueados usando el algoritmo de secuencias crecientes de tiempos entre contactos

Descripción

Para el caso de un pluviómetro de balancín que se encuentra parcialmente bloqueado, ya sea por hojarasca, insectos o desperdicios de aves que se posicionan sobre el embudo, se puede suponer que el agua tiende a acumularse en lugar de caer inmediatamente hacia el balancín, dependiendo de la magnitud de la obstrucción y de la cantidad de agua que se acumule en el embudo. Es así como las gotas de agua que caen hacia el balancín lo harán a una razón de cambio que

tiende a disminuir conforme disminuye el volumen de agua acumulada, suponiendo que ha dejado de llover. Partiendo de esta suposición, se puede afirmar que un pluviómetro de balancín podría estar bloqueado si el tiempo entre contactos aumenta (Upton y Rahimi, 2003). Para el almacenamiento de la información con respecto a tiempos entre contactos crecientes o decrecientes, se define un índice de cambio δ_k , que clasifica los tiempos entre contactos de la siguiente manera:

$$\delta_k = 1, \tau_k > \tau_{k-1}, \quad (5)$$

$$\delta_k = 0, \tau_k = \tau_{k-1}, \quad (6)$$

$$\delta_k = -1, \tau_k < \tau_{k-1}, \quad (7)$$

donde τ_k es el tiempo entre contactos. Los tiempos entre contactos se separan en dos clases, excluyendo aquellos casos en los que no hubo variación en los intervalos entre contactos; la fracción “p” de ellos que cumple $\tau_k > \tau_{k-1}$ y la fracción “q” que cumple que $\tau_k < \tau_{k-1}$.

Es posible calcular la probabilidad de que exista una fracción p de intervalos crecientes en la serie de tiempos entre contactos usando la distribución binomial (Spiegel y Stephens, 2002). Si p es la probabilidad de obtener $\delta_k=1$ para un tiempo entre contactos y “q” la probabilidad de obtener $\delta_k = -1$, entonces la probabilidad de que se genere un número de X tiempos entre contactos crecientes en N tiempos entre contactos consecutivos vendrá dada por:

$$p(X) = \frac{N!}{X!(N-X)!} p^X q^{N-X} . \quad (8)$$

A partir de este razonamiento, se calcula la probabilidad de obtener 12 tiempos entre contactos crecientes en una muestra consecutiva de 12 tiempos entre contactos. En las condiciones analizadas por Upton y Rahimi (2003), y después de remover datos erróneos, se tuvo un registro de tiempos entre contactos de 200 000 valores de espacios consecutivos, donde un 51%

cumplía que $\tau_k > \tau_{k-1}$, contra un 48% en que $\tau_k < \tau_{k-1}$, de lo cual se deduce que la probabilidad p es aproximadamente 0,51 para esta muestra de datos. De modo que suponiendo que el número de tiempos entre contactos crecientes es aproximadamente igual al número de tiempos entre contactos decrecientes para una muestra suficientemente larga, y suponiendo que la ubicación de estos tiempos en la serie fuera al azar, se tiene que la probabilidad de encontrar 12 o más tiempos entre contactos crecientes es $0,51^{12} = 0,003$. Nótese que la elección de una secuencia de 12 tiempos entre contactos consecutivos crecientes fue hecha por la siguiente razón: para una secuencia con este número tiempos entre contactos la probabilidad de una secuencia consecutiva con $\delta_k=1$ es muy baja, de allí que un pluviómetro de balancín podría estar bloqueado si una secuencia de 12 o más valores positivos de δ_k es encontrada.

Datos utilizados y metodología

La prueba de este ACCD se llevó a cabo utilizando diversos volúmenes de agua, con el fin de detectar si se generaban secuencias crecientes de contactos. La obstrucción en el embudo era inducida utilizando pasta selladora, un material moldeable utilizado en la red de estaciones para el sellado de orificios en las cajas de las EMA, el cual, después de varias pruebas con otros materiales, resultó ser el más adecuado para producir los orificios que permitieran generar las secuencias buscadas, ya que posibilita regular fácilmente el tamaño del orificio utilizando agujas de diversos diámetros, además de su gran capacidad adhesiva. La figura 5 muestra el sistema utilizado para la simulación de este tipo de obstrucciones, donde puede observarse la colocación del pluviómetro de balancín dentro de un contenedor de agua con el fin de confinar el agua usada para las pruebas. En la parte superior del diagrama de lo que representa el pluviómetro de balancín puede observarse la representación del embudo en la cual se colocaba una

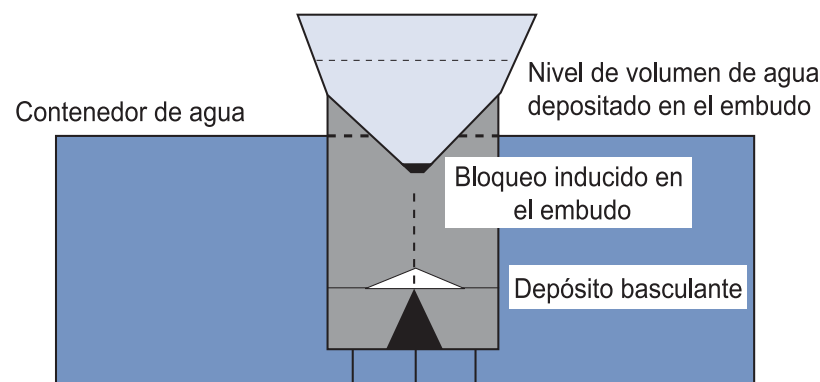


Figura 5. Simulación de la obstrucción en pluviómetro de balancín sin precipitación en el momento de la obstrucción.

En total se realizaron 9 pruebas en las cuales solo había agua contenida en el embudo del pluviómetro de balancín (Figura 5), esto, con el fin de probar las premisas del ACCD para una secuencia de tiempos entre contactos crecientes

cantidad previamente definida de agua que precipitaba lentamente hacia el balancín debido a la obstrucción inducida en el orificio de salida de agua del embudo.

En total se realizaron 9 pruebas en las cuales solo había agua contenida en el embudo del pluviómetro de balancín (Figura 5), esto, con el fin de probar las premisas del ACCD para una secuencia de tiempos entre contactos crecientes. Para este otro ACCD se realizaron 11 pruebas, en las cuales hubo cierto volumen de agua precipitando (al mismo tiempo que había también cierto volumen de agua confinado en el embudo). Se experimentó con diferentes volúmenes de agua en el embudo; asimismo, el tamaño del orificio fue variable, ya que se regulaba el tamaño de este al inicio de cada una de las pruebas para inducir tiempos iniciales entre contactos de diferente duración. De nuevo, estas pruebas son muy elementales y pretendían inducir obstrucciones en el embudo del sensor de tal forma que se diera la condición de tiempos entre contactos crecientes, ya que solo interesaba demostrar que el algoritmo es capaz de detectar la condición de bloqueo si se dan las condiciones apropiadas. Nótese, además, que en ocasiones los ACCD no están en condiciones de detectar la

situación. De las experiencias llevadas a cabo, se observó que hay algunos errores que pueden afectar la capacidad de estos ACCD de detectar bloqueos parciales, a saber:

- *Volumen de agua acumulado:* en el embudo podría no acumularse el volumen de agua suficiente para que se detecte una secuencia sospechosa. Por ejemplo, debería esperarse que se tengan al menos 1,2 mm acumulados en el embudo para que el ACCD para una secuencia entre contactos crecientes fuera capaz de detectar el bloqueo parcial; en caso contrario, el bloqueo pasará inadvertido.
- *Magnitud inestable del bloqueo:* se observó que las propiedades físicas de la pasta selladora podían variar durante la experiencia, debido al contacto directo con el agua, haciéndola más blanda conforme más tiempo permaneciera en contacto directo con el agua. Esto podría afectar la constancia en el diámetro del orificio inducido sobre esta masa, lo cual tendría su efecto en la magnitud de los tiempos entre contactos.
- *Nivelación adecuada del pluviómetro de balancín:* si el instrumento no está bien nivelado, la duración del tiempo entre contactos puede resultar afectada.
- *Diseño del embudo del pluviómetro de balancín TE525MM:* El embudo de este instrumento no es una superficie cónica perfecta, sino que presenta un estrechamiento abrupto en la superficie interna, por lo que la variación de la profundidad del volumen cónico de agua con su radio no es constante durante todo el proceso de evacuación. Esto afecta las secuencias crecientes de tiempos entre contactos.
- *Adecuada calibración del pluviómetro de balancín:* aunque para esta investigación se llevó a cabo una

calibración previa del sensor, una mala calibración tendría su efecto en la magnitud de los tiempos entre contactos, ya que la adecuada calibración depende de que ambas cazoletas se mantengan al mismo nivel y que sean capaces de contener el mismo volumen de agua.

- *Cazoletas limpias*: Cuando material extraño, tal como grasa o tierra se adhiere sobre las cazoletas, el momento de inercia del balancín podría variar hasta el punto de afectar la calibración y con ello la magnitud de los tiempos entre contactos.

Es importante que este tipo de factores se tomen en consideración ya que en condiciones operativas pueden tener su efecto en la capacidad de detección de estos ACCD para bloqueos parciales.

Resultados

La figura 6 muestra el resultado de una de las pruebas llevadas a cabo para comprobar la consistencia del algoritmo. Nótese que de nuevo el eje temporal da la hora en horas y minutos (se inició a las 8:16 h y finalizó a las 10:56 h). En este caso, también se filtró el archivo de datos con muestreo a un segundo, con el fin de mostrar solamente los detalles de interés. La parte a) de la figura 6 muestra la variación del índice de cambio conforme se generaba la secuencia de contactos.

Obsérvese en las secuencias que el índice de cambio en general asume un valor alternante entre -1 y 1 , hasta que en algún momento comienza a darse una persistencia de 1 , lo cual denota una secuencia de tiempos entre contactos que crece en el tiempo. La parte b) muestra cómo el tiempo entre contactos va aumentando progresivamente conforme transcurre el tiempo; el último tiempo entre contactos decreciente se da aproximadamente a las 9:03 h, y a partir de esa hora hay una tendencia creciente, mostrando que efectivamente la convención asumida para el índice de

cambio está señalando una secuencia en la que el tiempo entre contactos aumenta con el tiempo. Finalmente, la parte c) muestra como el Indicador de Alerta (IA) es activado exactamente cuando el algoritmo detecta una secuencia de doce tiempos entre contactos crecientes, generándose una constante “-10”, en contraste con la condición nula de no alerta. Obsérvese que cuando se comparan la parte b) con la c) el IA, se activa en el mismo momento en el que se han generado la secuencia de 12 tiempos entre contactos crecientes.

Detección de pluviómetros de balancín parcialmente bloqueados, usando el algoritmo de sumas acumulativas

Descripción

Este algoritmo de sumas acumulativas se propone con el fin de tratar aquellos casos en los que, dada una tendencia creciente en la magnitud de los tiempos entre contactos en un bloqueo parcial en un pluviómetro de balancín, se presenta una disminución en la magnitud de los tiempos entre contactos. Esto puede ocurrir, por ejemplo, en aquellos casos en los que se tiene una obstrucción parcial y al mismo tiempo algo de lluvia comienza a precipitar sobre el embudo. La ventaja de este ACCD es que permite tolerar cierto número de disminuciones en los tiempos entre contactos (lo cual no permite el ACCD para secuencias de tiempos entre contactos crecientes).

Esto puede hacerse usando un gráfico de control de sumas acumulativas, el cual es método de control estadístico que ha probado ser de gran utilidad en la industria, ya que permite detectar pequeñas desviaciones con respecto a la media (Montgomery, 2005). En el contexto del Proceso de Control de Calidad de Datos en meteorología, los gráficos de control de sumas acumulativas pueden ayudar a evaluar sesgo e histéresis con respecto a mediciones independientes (como, por ejemplo, las efectuadas con instrumentos

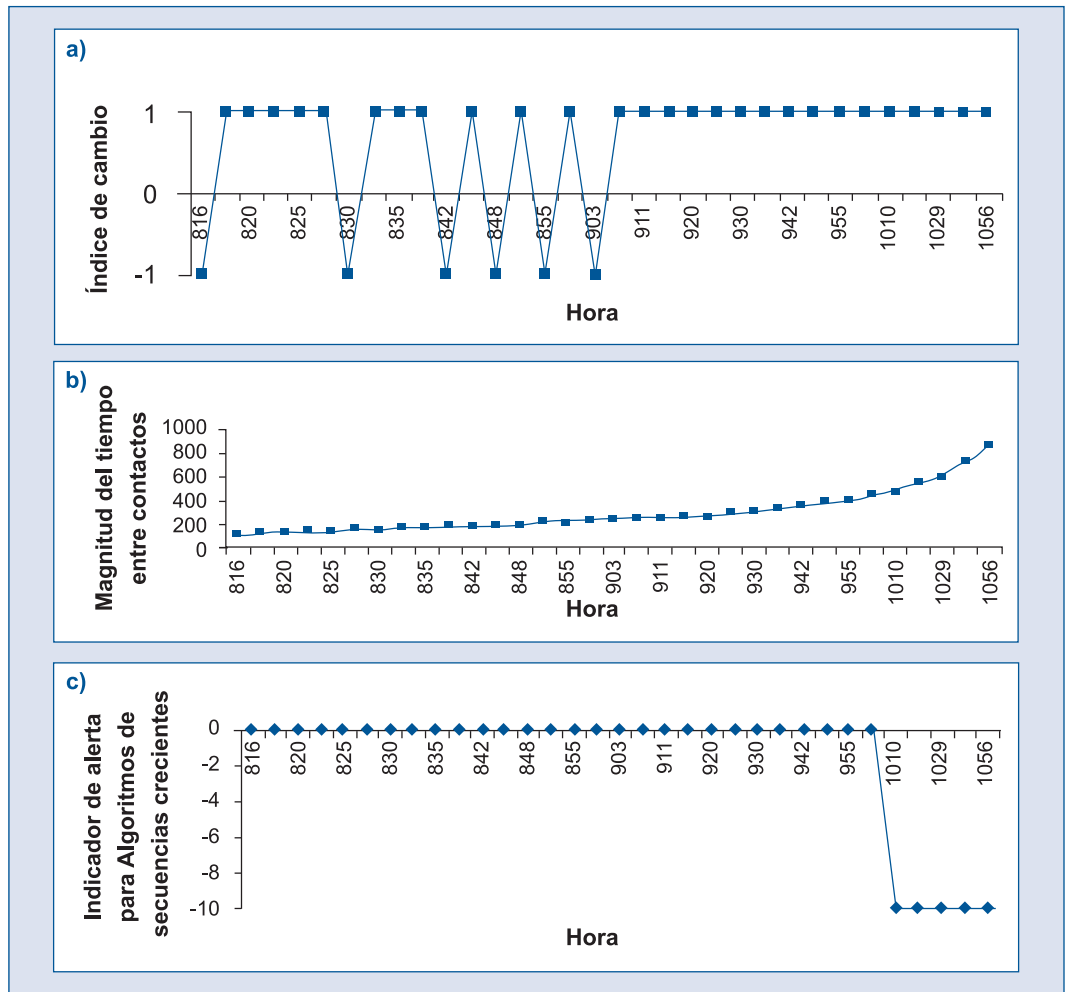


Figura 6. a) Variación del índice de cambio durante la prueba. b) Variación de la magnitud del tiempo entre contactos. c) Variación del indicador de alerta para el algoritmo de secuencia de tiempo entre contactos crecientes.

portátiles) o comparaciones entre sensores redundantes instalados en el emplazamiento o bien los ubicados en estaciones cercanas (Meek y Hatfield, 1994).

En la aplicación de este concepto para la detección de posibles bloqueos parciales se requiere la misma codificación de los tiempos entre contactos, usando el parámetro δ_k tratado en el ACCD para secuencias crecientes de tiempos entre contactos. Se hace uso del estadístico C_k dado por la expresión :

$$C_k = \max_k \left(0, \sum \delta_i \right) \quad (9)$$

donde la sumatoria se efectúa para $i=1,2,\dots,k$, siendo C_k la suma acumulativa después de la k -ésima observación. Obsérvese que se asume que la media objetivo deseada para los valores de δ_k es cero; es decir, dada una secuencia de tiempos entre contactos suficientemente larga, entonces la media de estos δ_k se supone que es cercana a cero. Nótese también que en los casos en que $\sum \delta_k < 0$, entonces se supone el valor $C_k = 0$. El conocimiento de este estadístico hace posible el cálculo de la razón r_k dada por:

$$r_k = \frac{C_k}{N_k} \quad (10)$$

la cual es un índice que señala como se acumula información acerca de la ocurrencia de tiempos entre contactos crecientes. N_k es el número de espacios desde la última vez que se registró un valor de cero para C_k .

De esta manera, Upton y Rahimi (2003) sugieren que “un pluviómetro de balancín podría estar bloqueado si el estadístico C_k excede 12 y además $r_k \geq 0.5$ ”. El algoritmo además inicia en cero los valores de C_k y N_k si encuentra un intervalo mayor a dos horas (véase Montgomery, 2005, para más detalles en cuanto a procesos de visualización de medias usando el algoritmo de sumas acumulativas).

Datos utilizados y metodología

Los datos fueron generados en el laboratorio por medio de pruebas muy sencillas, cuyo único objetivo era lograr generar las condiciones para las cuales el algoritmo fuera capaz de detectar un bloqueo en el embudo. La figura 7 muestra el arreglo experimental usado para modelar las condiciones de bloqueo para este algoritmo. De nuevo se tomaron las medidas básicas de instalación, tales como nivelación del instrumento, limpieza de este, etc., mencionadas también para

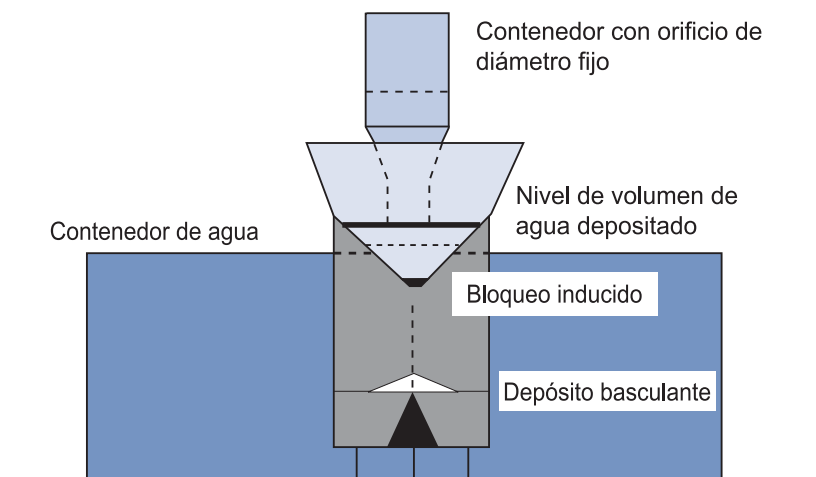


Figura 7. Simulación de la obstrucción en pluviómetro de balancín con precipitación.

el algoritmo de secuencias crecientes de tiempos entre contactos. Para simular el efecto de la lluvia en caso de obstrucción, se utilizó el recipiente calibrador con el orificio regulado, de modo que en todas las pruebas realizadas se dio la situación de que había un bloqueo inducido en el embudo con cierto volumen de agua acumulada, al mismo tiempo que había flujo de agua desde el recipiente calibrador hacia el embudo.

Resultados

La figura 8 muestra uno de los casos en los cuales el algoritmo detectó obstrucción en el pluviómetro de balancín. Nótese la escala de tiempo en horas y minutos en el eje de las abscisas. En la parte a) se presenta la magnitud del tiempo entre contactos. Obsérvese cómo el tiempo entre contactos tiende al aumento con el tiempo, aunque con ocasionales disminuciones. La parte b) muestra el comportamiento del índice de cambio; se evidencia que no necesariamente se debe tener una secuencia positiva para el índice de cambio para que el algoritmo detecte un bloqueo. Al respecto, la figura 8c demuestra como crece el valor de la suma acumulativa desde la última vez que esta fue nula; la pendiente de la figura 8d corresponde al valor que asume la razón r_k . Obsérvese que para el contacto número veinticuatro (denotado por N_k en el eje de la abscisa) el valor de la suma acumulativa llega a 13, de modo que la razón r_k en ese punto es igual o superior a 0,5, lo cual demuestra que, efectivamente, el algoritmo considera el valor de esta razón después de que el parámetro de sumas acumulativas alcanza su valor umbral. Finalmente, la parte d) denota el valor del IA y la hora en la que la obstrucción fue detectada.

Caso real de obstrucción detectada en la red de estaciones meteorológicas

Un caso comprobado de obstrucción parcial en un pluviómetro de balancín se presentó

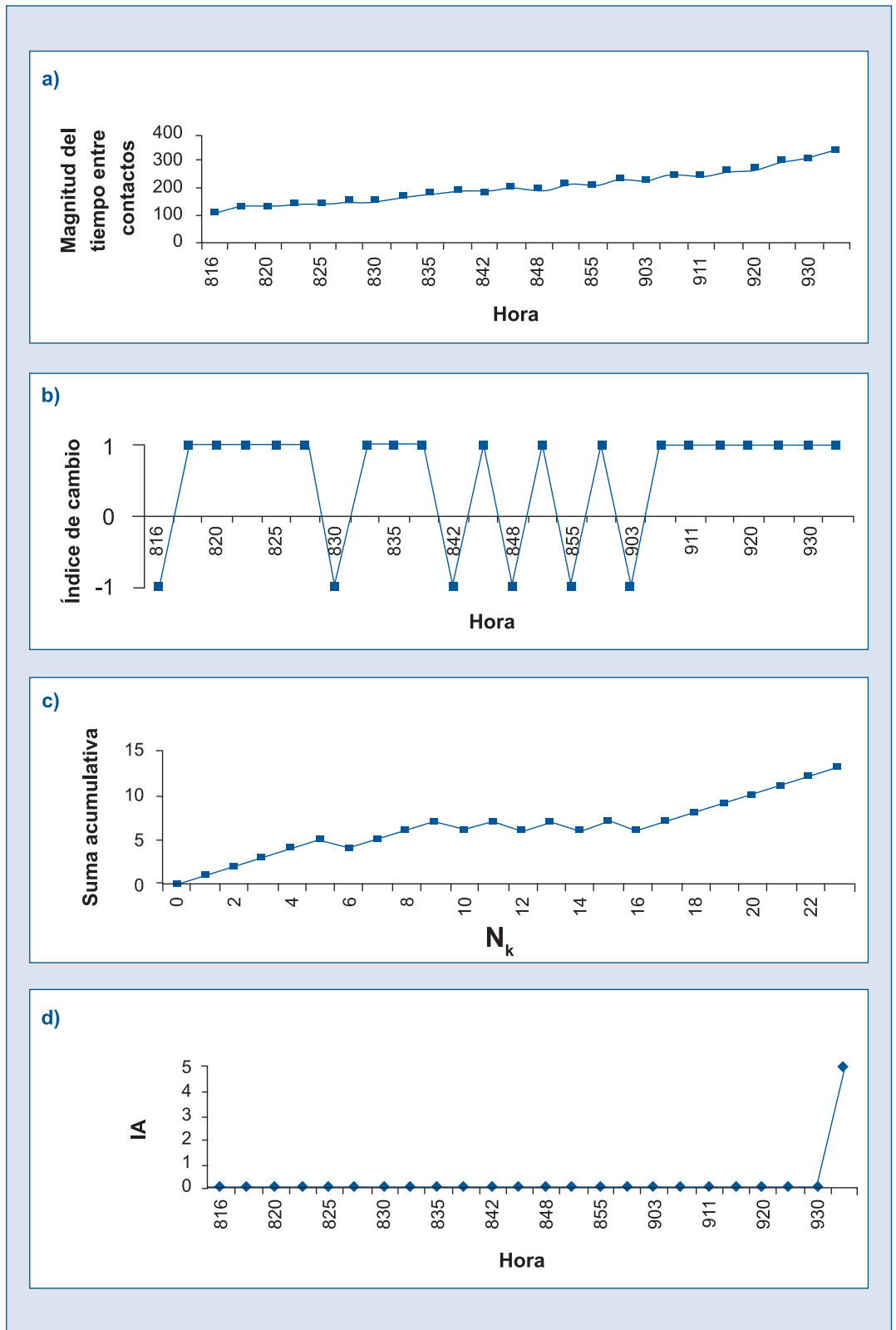


Figura 8. a) Variación de los tiempos entre contactos. b) Variación del índice de cambio. c) Incremento del valor de sumas acumulativas. d) Variación del IA.

en el Aeropuerto Internacional “Juan Santamaría, Costa Rica” (latitud: 9° 59’, longitud 84°13’) entre el 15 y el 18 de marzo, en el cual la estación meteorológica reportaba acumulados de precipitación en horas en las que era claro que no hubo lluvia. Es necesario recalcar que este caso mostrado no fue detectado por ninguno de los algoritmos anteriormente explicados, ya que no fueron probados en condiciones operativas; sin embargo, se recogió suficiente evidencia como para probar que esta serie de tiempo atípica efectivamente fue generada por una obstrucción parcial en el pluviómetro de balancín, como, por ejemplo, el bloqueo del orificio del embudo (el cual fue removido por los técnicos de mantenimiento de la red) y la generación de contactos no existentes en condiciones de cielos despejados que fueron observados por el personal de meteorología sinóptica del IMN. Curiosamente, esta situación en las que el personal de meteorología sinóptica percibía registros extraños se presentó en dos ocasiones, los días 17 y 18 de marzo. Una vez revisado el pluviómetro en el lugar se comprobó que efectivamente estaba obstruido. La figura 9 es interesante, debido a que muestra cómo el tiempo promedio entre contactos aumentaba en el tiempo, a la vez que se registraban acumulados de precipitación

que se daban en horas consecutivas y que disminuían en el tiempo; esto, debido a la disminución del volumen de agua contenido en el embudo. Es en casos como estos en que los ACCD tratados en este estudio resultan de utilidad. Este ejemplo mostrado también refuerza la importancia de la futura implementación a nivel operativo de pruebas de control de calidad como las descritas anteriormente, de modo que el control de calidad pueda realizarse de forma objetiva mediante pruebas que simulen, en la medida de lo posible, la revisión que un ser humano haría de los datos. Esto cobra mayor trascendencia en el contexto de la automatización y conforme más grande sea la red de estaciones meteorológicas que se tenga.

Conclusiones

Después de la programación y prueba preliminar de los ACCD para precipitación previamente analizados, y a la luz de las pruebas de laboratorio realizadas, nótese que estos tres algoritmos son altamente dependientes del conocimiento del tiempo entre contactos consecutivos, por lo que en caso de ser implementados en tiempo real en la red de estaciones del IMN, deberá incluirse un algoritmo para el cálculo de los tiempos entre contactos, ya sea en el

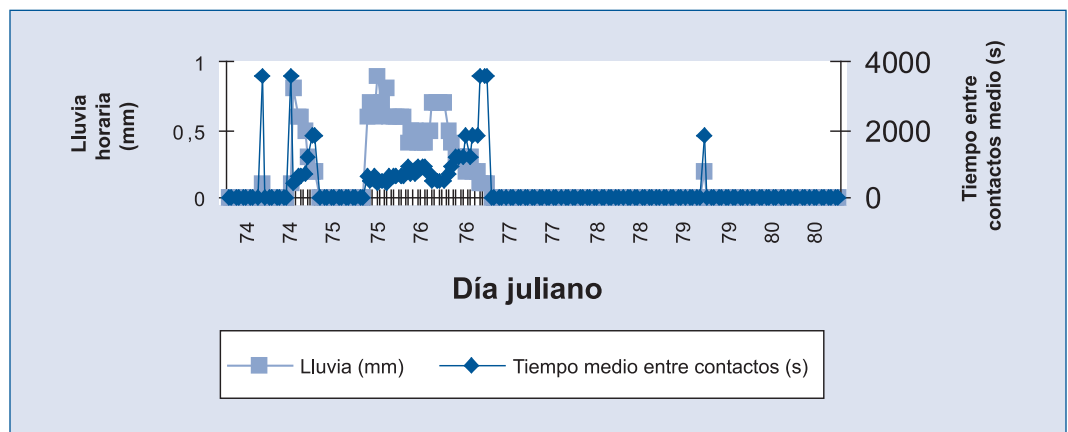


Figura 9. Total horario de lluvia y tiempos medios de tiempos entre contactos por hora en el Aeropuerto Internacional Juan Santamaría. 15 al 17 de marzo del 2007.

emplazamiento (en la programación del GD) o tras el envío de los datos a la OPD, luego se procede a aplicar los ACCD tratados en este capítulo en el nivel de control de calidad que mejor convenga.

Analizando el contexto de la investigación en la que estos ACCD fueron aplicados por Upton y Rahimi (2003) en una red de EMA en la que tenían instalados varios pluviómetros de balancín distribuidos al noroeste de Inglaterra, se encontraron fuentes de error como las siguientes:

- Posibles rebotes en el balancín.
- Bloqueos parciales por deposición de hielo sobre los embudos.

En el contexto de Costa Rica, es posible que en redes de estaciones que usan pluviómetros de balancín se den errores como los comentados aquí. En el caso de los bloqueos parciales, estos se han presentado en la red de estaciones meteorológicas del IMN, solo que estos bloqueos parciales, (o totales) no han tenido que ver con solidificación de agua bloqueando parcialmente el orificio del embudo (Fernández, 1984).

Los ACCD para posibles rebotes y bloqueos parciales identificaron secuencias de tiempos entre contactos sospechosos. En las pruebas de laboratorio realizadas en esta investigación se concuerda con Upton y Rahimi (2003), en que estos ACCD no reclaman infalibilidad en el análisis que efectúan, por lo que siempre deberán ser valorados con un adecuado criterio profesional en caso de detección. Puede suceder que el bloqueo sea detectado por uno u otro algoritmo, por ambos a la vez, o bien, no detectado por ninguno, eso dependerá de las condiciones de obstrucción que se tengan, ya que una vez que el pluviómetro de balancín es colocado en el emplazamiento, las dos condiciones para las cuales se concibieron estos algoritmos pueden darse de forma natural, por lo que es conveniente tener ambos algoritmos para aumentar la posibilidad de

detección en caso de bloqueo parcial. Los resultados de las pruebas de laboratorio mostraron que ambos algoritmos por separado alertaron alrededor del 55% de las obstrucciones. Usados en conjunto, ambos detectaron el 70% de los casos de bloqueo inducido, por lo que la capacidad de detección es mayor usando ambos algoritmos, de allí que es recomendable su uso en conjunto para mejorar la eficiencia en la detección de secuencias sospechosas de tiempos entre contactos (Araya, 2007).

Se espera que la capacidad de detección de contactos atípicos usando estos algoritmos mejore si se tienen varios instrumentos en el área; de hecho, la utilidad de estos ACCD radica, precisamente, en el hecho de que fueron concebidos para la detección de contactos atípicos en tiempo real en el contexto de una red de estaciones debidamente constituida, quedando su implementación a nivel operativo, dentro de una red de estaciones meteorológicas para estudios futuros.

Agradecimientos

Una gratitud especial al Instituto Meteorológico Nacional, por facilitar el equipo y las instalaciones para la realización de los programas y la producción de los datos utilizados en esta investigación; sin su apoyo este trabajo no habría sido posible. Gracias a los señores Ing. Rodolfo Dávila, Víctor Hugo Gómez, por su invaluable ayuda cuando requerí asesoramiento en materia de cómputo. Un agradecimiento al Sr. Marcos Rojas, del Instituto Costarricense de Electricidad, por su explicación de la metodología de calibración de pluviómetros de balancín usada en esta institución. Por último, al proyecto ED-1977-UCR.

Se espera que la capacidad de detección de contactos atípicos usando estos algoritmos mejore si se tienen varios instrumentos en el área....

Bibliografía

- Alexandropoulos, C. y Lacombe, M., 2005. WMO Laboratory Intercomparisons of Rain Intensity Gauges. Meteo-France, Trappes, Francia. Guidelines on Quality Control Procedures for Data from Automatic Weather Stations. World Meteorological Organization, Report 82, 348-355
- Araya, J. L., 2007: Algoritmos de control de calidad de datos en estaciones meteorológicas. automáticas. Tesis de Licenciatura, Escuela de Física, Universidad de Costa Rica. 172 pp.
- Campbell Sci., 2004b: TE 525 Tipping Bucket Rain Gage. Instruction Manual. Campbell Scientific, Inc. Utah. 8 pp.
- Fernández, W., 1984: Comments on Meteorological and Climatological Observations on Coco Island. Revista Geofísica, N.º 20, 9-19.
- Humpfrey, M.D., Istok, J.D., Lee, J.Y., Hevesi, J.A. y Flint, A.L., 1997: A new Method for Automated Dynamic Calibration of Tipping-Bucket Rain Gauges. *J. Atmos. Oceanic Technol.*, 14,1513-1519.
- Meek, D. W., y Hatfield, 1994: Data Quality checking for single station meteorological databases. *Agric. For. Meteor.*, 69, 85-109.
- Montgomery, D.C., 2005: Introduction to Statistical Quality Control. 5th ed. J. Wiley & Sons, 759 pp.
- Spiegel, M.R.y Stephens, L.J., 2002: Estadística. 3.ª edición. Mc Graw Hill, 541 pp.
- Upton, G. J. G. y Rahimi, A. R., 2003: On-line detection of errors in tipping-bucket raingauges. *J. Hidrol.*, 278, 197-212.
- Ureña, F.,1998: Estudio y análisis de los datos generados por una estación meteorológica automática. Tesis de grado de Licenciatura. Escuela de Física, Universidad de Costa Rica. 106 pp.