

MANUAL 5

PROTOCOLO DE MANEJO DE DATOS DE LA CALIDAD DEL AIRE



1.INTRODUCCIÓN.	5
2. MANEJO DE DATOS.	7
3. BASES DE DATOS DE CALIDAD DEL AIRE Y CRITERIOS BÁSICOS PARA SU MANEJO.	12
3.1. Formato de Bases de Datos.	12
3.2. Sistemas de Medición Automática.	14
3.2.1. Limpieza y verificación de datos en los SMCA.	14
3.2.1.1. Limpieza de datos.	15
3.2.1.2. Macros para la asignación de banderas durante la limpieza de datos.	19
3.2.2. Verificación de datos.	20
3.2.2.1. Criterios de verificación automática de datos.	21
3.2.2.2. Criterios de verificación manual de datos.	22
3.2.2.3. Herramientas estadísticas útiles para la verificación de datos.	24
3.3. Sistemas de Medición Manual.	25
3.3.1. Banderas sugeridas y su relación con el Control y Aseguramiento de la Calidad.	27
3.4. Características de las bases de datos verificadas.	27
3.4.1. Unidades de medición, claves de parámetros y cifras significativas.	28
4. PROCEDIMIENTO PARA OBTENER INDICADORES DE LA CALIDAD DEL AIRE.	30
4.1. Tipos de datos que se utilizan para el cálculo de distintos indicadores de la calidad del aire.	31
4.2. Nota importante sobre los criterios de compleción de datos.	38
4.2.1. Desempeño anual y desempeño histórico por estación de medición.	38
4.3. Cómo se definen y calculan los indicadores de la calidad del aire.	41

4.3.1. Indicadores relacionados con las normas de calidad del aire.	41
4.3.1.1. Consideraciones para la interpretación de las Normas Oficiales Mexicanas de calidad del aire.	43
4.3.2. Indicadores del comportamiento de los contaminantes a lo largo del tiempo.	47
4.4. Representación gráfica de los indicadores de calidad del aire.	48
4.4.1. Ejemplos de las representaciones gráficas de los indicadores.	50
4.4.1.1. Indicadores de impacto en la salud.	50
4.4.1.2. Indicadores relacionados con el comportamiento del contaminante.	51
5. LOS OBJETIVOS DE CALIDAD DE LOS DATOS (OCD) Y LA EVALUACIÓN DE LA CALIDAD DE LOS DATOS (ECD) EN EL CICLO DE VIDA DE LOS DATOS EN UN PROYECTO.	55
5.1. Objetivo de la calidad de los datos.	56
5.2. Evaluación de la calidad de los datos.	57
5.3. La ECD y su ciclo de vida.	58
5.4. Los cinco pasos en la ECD	59
6. REFERENCIAS.	60
ANEXO I. AJUSTE DE DATOS POR RESULTADOS DE LA VERIFICACIÓN CERO-SPAN.	65
ANEXO II. LÍMITES DE DETECCIÓN DE LAS PRINCIPALES MARCAS DE EQUIPOS DE MEDICIÓN DE LA CALIDAD DEL AIRE EN MÉXICO.	69
ANEXO III. EJEMPLO DE LA OBTENCIÓN DE LOS INDICADORES PERCENTIL 98 Y PROMEDIO ANUAL.	73
ANEXO IV. HERRAMIENTAS ESTADÍSTICAS.	83

1. INTRODUCCIÓN.

Durante los últimos años se ha observado en el país un ritmo constante en el desarrollo de los diferentes sistemas de medición de la calidad del aire, SMCA, principalmente en lo que se refiere al conjunto de equipos de medición y los de adquisición de datos, por lo que es una necesidad fundamental lograr la homologación de las prácticas de medición de calidad del aire, estableciendo esquemas de control y aseguramiento de calidad (CC y AC) que garanticen la veracidad de los datos que generan tal como se establece en el Programa Nacional de Monitoreo Atmosférico (INE-DGCENICA, 2004).

Actualmente la confiabilidad de la información generada por los SMCA queda, en gran parte, supeditada a la calidad del trabajo operativo y al rigor de las instituciones responsables de su manejo. En México son las autoridades ambientales locales las que cuentan con la atribución de administrar, medir y evaluar los niveles de los contaminantes en la atmósfera (DOF, 2006).

En muchos países los métodos de medición están normados con el fin de asegurar que la información generada por las redes de medición sea confiable, comparable con otras redes y de suficiente calidad. En este sentido y con el fin de atender algunas de las deficiencias en el manejo de datos de calidad

del aire que se presentan en algunas redes de medición locales, el Instituto Nacional de Ecología, INE, ha desarrollado este documento como una referencia metodológica que presenta una serie de elementos y criterios que se deben considerar en el manejo de las bases de datos de la calidad del aire, desde la generación del dato en la estación de medición, hasta el cálculo de indicadores y su reporte, para de esta manera asegurar la compatibilidad de los resultados que se generen a nivel nacional.

Con este documento se complementa una serie de seis manuales con los que se pretende asegurar la calidad de los datos generados por las redes de medición de la calidad del aire del país para estar a nivel competitivo con otras regiones y para que la toma de decisiones hacia el mejoramiento de la calidad del aire en beneficio de la salud y el ambiente se lleven a cabo sobre bases sólidas. Los documentos contienen información relevante que cubre todos los aspectos de la medición de la calidad del aire, desde el diseño de una red de medición hasta la validación de los datos y su difusión pública, pasando por la operación, mantenimiento, actividades de aseguramiento y control de calidad y auditorías.

Este documento se divide en cinco capítulos principales, bibliografía y se complementa con tres anexos. El Capítulo 1 da una introducción al Manual. En el Capítulo 2 se exponen, de manera general, los pasos del manejo de datos y la importancia de que dicho manejo se realice de forma adecuada.

El Capítulo 3 tiene por objeto introducir al lector en las primeras etapas del manejo de datos de los sistemas de medición automática: limpieza y verificación. Presenta además, temas como el formato de las bases de datos, definiciones, uso y ejemplos de aplicación de banderas para calificar los datos y también se señalan algunas herramientas estadísticas que se usan por lo común en estas etapas de limpieza y verificación. También se incluye una sección para el manejo de datos de redes manuales y finalmente se describen las características que deberán tener las bases de datos verificadas que provienen de ambos tipos de sistemas.

En el Capítulo 4 se presenta una selección de indicadores de calidad del aire, su definición y los elementos técnicos que se deben considerar en su procesamiento. Asimismo, se propone la forma de interpretar las especificaciones de las Normas Oficiales Mexicanas (NOM) de calidad del aire, publicadas por la Secretaría de Salud, y se dan algunas propuestas de cómo presentar los indicadores de manera gráfica.

El Capítulo 5 describe los objetivos de calidad de datos, su importancia, cómo se realiza su evaluación y su ciclo de vida.

En la serie de anexos, el Anexo I presenta la ecuación del ajuste de datos por resultados de la verificación Cero-Span; en el Anexo II se muestran los límites de detección para las principales marcas de equipos empleados en la medición de la calidad del aire en México; finalmente, el Anexo III contiene una descripción detallada de las herramientas estadísticas que se usan en el manejo de datos.

2. MANEJO DE DATOS.

La medición de la calidad del aire debe responder a los objetivos específicos que establezca el SMCA (ver Manual 2: Sistemas de Medición de la Calidad del Aire), como evaluar el grado de cumplimiento de las normas de calidad del aire y presentar los avances de este cumplimiento al público; para identificar las tendencias a largo plazo de los contaminantes atmosféricos; para desarrollar y evaluar programas y estrategias para el manejo de la calidad del aire; entre otros (US-EPA, 1998; y WHO, 2000). Para cumplir con estos objetivos, las estaciones de medición deben estar ubicadas estratégicamente para cuantificar los niveles de contaminación en sitios representativos que permitan caracterizar adecuadamente la contaminación en una región específica. En este sentido, el diseño de las redes de medición puede orientarse a determinar las concentraciones máximas (valores pico) en el área cubierta por dicha red de medición; a informar al público acerca de los niveles de contaminantes en la atmósfera o a determinar el impacto de una fuente o categoría específica de fuentes (fijas, móviles, lineal y de área) en la calidad del aire (Ver Manual 3, Redes, Estaciones y Equipos de Medición de la Calidad del Aire).

Tal y como se mencionó en el Manual 3, las redes de medición se componen de un número variable de estaciones de medición, don-

de están ubicados los equipos de medición, y de un centro de control donde se almacenan, verifican y validan los datos generados en las estaciones (Figura 1). Para realizar el manejo de datos que generan las estaciones, es necesario homologar los criterios y procedimientos usados en la limpieza, verificación y validación, así como el control y aseguramiento de la calidad de la información, desde que se genera el dato en las estaciones de medición hasta que se reporta en forma de indicadores de la calidad del aire para asegurar que la información sea compatible y comparable entre estaciones de una misma red y entre redes de diferentes localidades.

El éxito de las decisiones que se tomen con estos datos dependerá de su calidad, de la facilidad con que se pueda acceder a ellos, de la forma como están agrupados y de la interpretación que se les dé. Asimismo, para que la toma de decisiones en materia de calidad del aire se fundamente en información veraz sobre la concentración de contaminantes en la atmósfera, es necesario que en los SMCA se cuente con un sistema de calidad que incluya no sólo los planes de aseguramiento y control de calidad (AC y CC), sino también los objetivos de calidad de los datos y su evaluación (OCD y ECD).

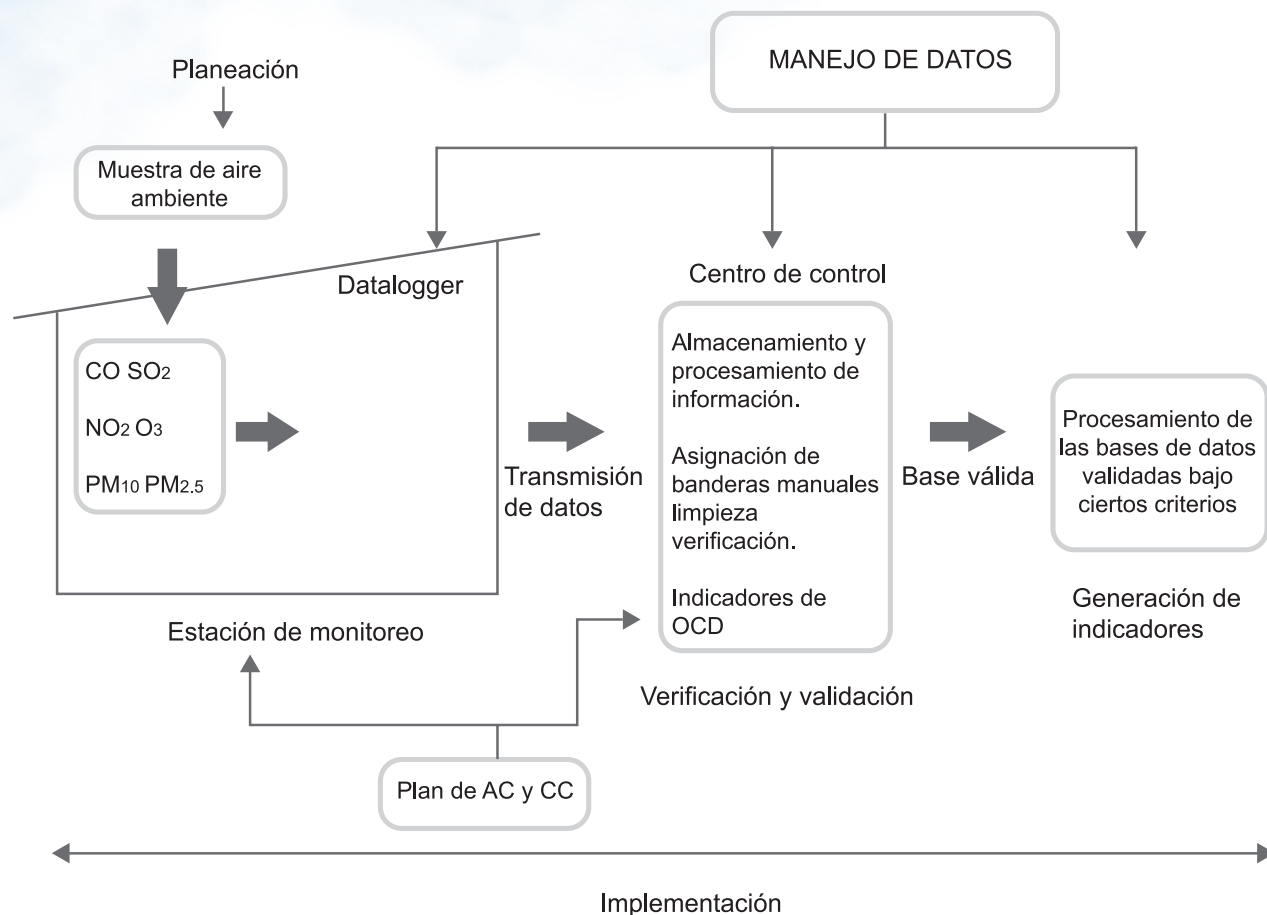


Figura 1. Flujo en el manejo de datos de calidad del aire.

En los siguientes capítulos se describen los procedimientos que deben seguirse para el manejo de datos de la medición de la calidad del aire, que comprende: la limpieza, verificación y validación de los datos, así como la generación de indicadores de calidad del aire. Estos procesos están incluidos en la fase de implementación dentro del ciclo de vida de los datos (Figura 1).

Para que se cumplan los objetivos de medición, la operación de las redes es tan importante como la limpieza, la verificación y la validación de los datos porque todo en conjunto permite aumentar al máximo su integridad.

La limpieza de los datos consiste, entre otras cosas, en la asignación (automática o manual) de banderas para la identificación de valores nulos y fuera de intervalo; la verificación de datos ocurre después de la limpieza inicial y comprende la revisión de los datos limpios y la asignación de banderas en situaciones específicas para lo cual se pueden utilizar herramientas complementarias, como macros en Excel, que automatizan la asignación de banderas a los datos horarios. Los procesos de limpieza y verificación se describen a detalle en el Capítulo 3 de este documento.

De acuerdo con la Agencia de Protección Ambiental de Estados Unidos (US-EPA, por sus siglas en inglés) la verificación de datos se refiere al proceso de evaluar que éstos estén completos, sean correctos y correspondan a los métodos, procedimientos o requisitos establecidos (US-EPA, 2002). Por otra parte, la validación de datos se concibe como un proceso para determinar la calidad analítica de un conjunto de datos de acuerdo con las necesidades particulares (por ejemplo, un plan de aseguramiento de calidad).

Después de que los datos han sido verificados se aplican los diferentes criterios mencionados en el Manual 2 (precisión, sesgo, representatividad y completión, comparabilidad, trazabilidad) para evaluar los objetivos de calidad y validar los datos. La verificación

y validación son pasos secuenciales, por lo general, y con frecuencia se llevan a cabo por diferentes personas. Por ejemplo, la verificación de los datos empieza, en algunas ocasiones, en las estaciones de medición mientras que la validación de datos se lleva a cabo posteriormente en el centro de control de la Red.

Una vez que los datos han sido validados, de acuerdo al objetivo para el cual fueron generados, pueden utilizarse para generar indicadores de calidad del aire. Un indicador se define, de manera general, como un valor que cuantifica y simplifica un fenómeno y que ayuda a comprender condiciones complejas (Web 1). Para el caso de la calidad del aire, un indicador es un medio compacto, una representación numérica de una gran cantidad de datos de calidad del aire medidos en un área dada (CARB, 1989).

De esta manera, los indicadores sobre calidad del aire permiten evaluar la situación de la contaminación atmosférica y comunicar al público la calidad del aire que respira, ambos objetivos de la medición de la calidad del aire (Zuk, et. al., 2007).

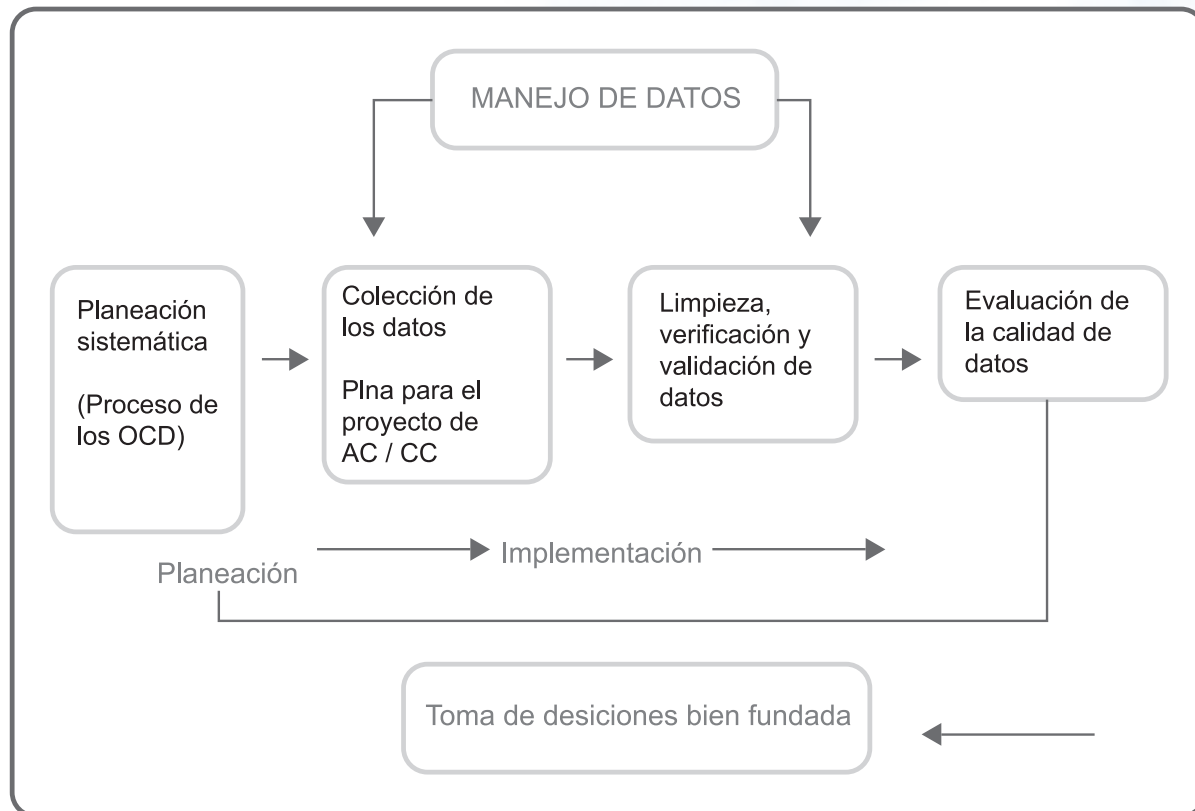
Para generar los indicadores de calidad del aire es necesario definir criterios para el procesamiento de la información registrada en las redes de medición. El propósito es homologar estos procedimientos en el país y redu-

cir la incertidumbre asociada con su cálculo. Los procedimientos y criterios para este propósito se especifican a detalle en el capítulo tres, y son aplicables a los datos generados en cada estación de medición aunque no para determinar las condiciones generales de la ciudad o zona metropolitana, debido a que actualmente sólo el Sistema de Monitoreo Atmosférico de la Ciudad de México reporta la representatividad de sus estaciones. El cálculo de indicadores para representar estrictamente las condiciones de una ciudad o zona metropolitana podrá llevarse a cabo una vez que se caracterice, evalúe y reporte la representatividad de las estaciones de medición, lo que permitiría una mejor y más adecuada interpretación de los indicadores de calidad del aire.

Como ya se mencionó, en México no se cuenta con la estructura para soportar un sistema de calidad en el que se consideren los objetivos de calidad de los datos (OCD) y la evaluación de su calidad (ECD). Por lo tanto, es importante tomar en cuenta estos conceptos y su relevancia dentro del contexto del ciclo de vida de los datos y de la medición de la calidad del aire de acuerdo con el sistema de la Agencia de Protección al Ambiente de los Estados Unidos.

De acuerdo a la US-EPA (US-EPA, 2000c), el ciclo de vida de los datos comprende tres etapas: planeación, implementación y eva-

luación. Durante la fase de planeación los procesos de los objetivos de la calidad de los datos (OCD) se utilizan para definir criterios cuantitativos y cualitativos para determinar cómo, cuándo, dónde y cuántas muestras (mediciones) se van a recolectar para alcanzar el nivel de confiabilidad deseado. En la etapa de implementación se considera la información que se generó en la etapa anterior, el plan de aseguramiento de calidad documentado que contiene los métodos de muestreo, los procedimientos analíticos y un esquema apropiado de aseguramiento y control de calidad (AC y CC) así como la recolección, limpieza, verificación y validación de los datos de acuerdo con las especificaciones de dicho plan. Por último, la evaluación de la calidad de los datos completa el ciclo de vida del proyecto al proveer la evaluación necesaria para determinar si se cumplió con los objetivos planteados en la etapa de planeación (Figura 2).



Fuente: adaptado de USEPA

Figura 2. Etapas del ciclo de vida de los datos en un proyecto.

En este sentido, al proceso de evaluar los datos contra los OCD se le llama evaluación de la calidad de los datos (ECD) y consiste en volver a revisar las rutinas de AC y CC y las de limpieza, verificación y validación de datos para determinar si los OCD se cumplieron y que los datos son adecuados para el uso previsto. En el Capítulo 5 se abordan con más profundidad estos conceptos.

3. BASE DE DATOS DE LA CALIDAD DEL AIRE Y CRITERIOS BÁSICOS PARA SU MANEJO.

En el presente capítulo se muestran las primeras etapas del manejo de datos: la limpieza y la verificación, presentando además, los primeros pasos a seguir para la conformación de una base de datos y estableciendo una propuesta de formato para ésta. Se describen conceptos para calificar los datos y se dan algunos ejemplos de aplicación y también se señalan algunas herramientas estadísticas que se usan por lo general en estas etapas. Posteriormente, se comentan algunos aspectos importantes para el manejo de datos de redes manuales y finalmente, en el último apartado, se describen las características que deberán tener las bases verificadas de datos que provienen de ambos tipos de equipos de medición.

3.1. Formato de Bases de Datos.

Una vez que los equipos de medición, automáticos o manuales, hayan obtenido los datos de calidad del aire y éstos se encuentren disponibles, deberá conformarse una base de datos. En el caso de la medición automática, las señales analógicas generadas por los equipos de medición se envían a un sistema de adquisición de datos, SAD, en donde dichas señales son transformadas a señales digitales. Estas señales digitales son procesadas por el SAD para obtener los valores respectivos de concentración o unidades físicas, en términos de los canales de datos preconfigurados por el operador de dicho sistema, conformando finalmente la base de datos. Ésta es respaldada tanto en la memo-

ria del SAD como en una computadora que por lo general se tiene conectada a dicho SAD. En el caso de los equipos manuales, las variables necesarias para obtener la concentración de partículas se registran en una hoja de cálculo que sirve tanto para vaciar los registros como para realizar los cálculos. Finalmente se recomienda que en otra hoja se registren los resultados finales de todas las estaciones de la red. Esta hoja se considera como la base de datos del muestreo manual.

Las bases de datos de calidad del aire, como cualquier base de datos, deberán tener una estructura definida y ordenada, donde la fecha y la hora determinen el orden. Asimismo, se deben especificar los campos que definen esta base de datos. Éstos pueden ser los parámetros a medir o las estaciones donde se mide cada parámetro con sus respectivos atributos que les sean característicos. Los valores pueden tener decimales o ser números enteros. La estructura de una base de datos cruda estará definida por el software de adquisición de datos, pudiendo ser desde archivos conocidos como valores separados por comas o formato de texto (txt), archivos en formato de base de datos (dbf, data base file), de Excel® (xls) o de tipo dat. De cualquier manera, el administrador o analista de la base de datos podrá definir el formato que a su juicio sea el más conveniente y fácil de usar, modificando el especificado por el software de adquisición de datos. No obstante que la decisión final sobre el formato a usar

la tienen los responsables del manejo de datos, la DGCENICA ha desarrollado programas basados en macros de Excel® que tienen la finalidad de apoyar a las redes del país en las tareas de limpieza y verificación de datos. Para el uso de estos programas se emplea el formato propuesto en el Cuadro 1.

	A	B	C	D	E	F
1	FECHA	HORA	CE 1	BCE 1	CE 2	BCE 2
50	03 01 2004	1	3.6	VA	5.3	VA
51	03 01 2004	2	8.7	VA	5.5	VA
52	03 01 2004	3	18.1	VA	5.2	VA
53	03 01 2004	4	10.4	VA	5.4	VA
54	03 01 2004	5	6.1	VA	5.7	VA
55	03 01 2004	6	6.1	VA	6.5	VA
56	03 01 2004	7	3.6	VA	5.9	VA
57	03 01 2004	8	5.1	VA	-0.2	VA
58	03 01 2004	9	11	VA	-0.1	VA
59	03 01 2004	10	34.1	VA	9.6	VA
60	03 01 2004	11	78.4	VA	23	VA
61	03 01 2004	12	70.0	VA	39.2	VA
62	03 01 2004	13	43.4	VA	60	VA
63	03 01 2004	14	39.2	VA	33.2	VA
64	03 01 2004	15	42.9	VA	11.0	VA
65	03 01 2004	16	48.3	VA	27.0	VA
66	03 01 2004	17	44.8	VA	27.9	VA
67	03 01 2004	18	36.1	VA	31.8	VA
68	03 01 2004	19	26.2	VA	29.7	VA
69	03 01 2004	20	13.5	VA	26.6	VA
70	03 01 2004	21	7.4	VA	16.5	VA
71	03 01 2004	22	7.7	VA	8.2	VA
72	03 01 2004	23	4.6	VA	5.2	VA
73	03 01 2004	24	3.8	VA	4.9	VA
74	03 01 2004	1	3.3	VA	5.3	VA
75	03 01 2004	2	11.5	VA	5.2	VA

Cuadro 1. Formato para manejo de datos.

En el ejemplo del Cuadro 1 se sugiere establecer la base de datos en función de los parámetros a medir, por lo que los campos estarán definidos de la siguiente manera:

La primera columna corresponderá a la fecha, la segunda a la hora, la tercera a la clave de la primera estación seguida de una columna para banderas, la quinta a la segunda estación seguida de una columna para banderas, y así sucesivamente. Este archivo será denominado de acuerdo al parámetro medido que se estableció. Todos los registros de las estaciones que se encuentran midiendo este parámetro deberán ser ordenados en columnas consecutivas en el mismo archivo en función de la estación, fecha y hora en que fueron generados. Asimismo, los campos para banderas deberán también conservar aquellas previamente asignadas por el software de adquisición de datos. Finalmente, se recomienda guardar las bases de datos originales de cada etapa del proceso de manejo de datos. Esta primera base de datos se conoce con el nombre de base de datos cruda.

En relación con la propuesta anterior, es importante mencionar que se continúa con la búsqueda de un formato estándar, fácil de usar y de aplicar por todos los SMCA para homologar las bases de datos a nivel nacional.

3.2. Sistemas de Medición Automática.

Las banderas¹ asignadas y los formatos de las bases de datos de los SMCA de cada zona o ciudad, como se señaló, difieren incluso entre estaciones del mismo sistema debido a que se emplean diversas marcas y modelos de sistemas de adquisición de datos. Para fortalecer el análisis, la evaluación y la comparación de los datos a nivel nacional, será necesario homologar los procedimientos de manejo de datos (EMEP, 2003. Web 2). Inicialmente en este apartado se propone una metodología base para calificar los datos resultantes de la medición de calidad del aire.

3.2.1. Limpieza y verificación de datos en los SMCA.

La limpieza y verificación de datos, como ya se mencionó, constituyen la primera etapa del manejo de datos. Ésta representa la valoración de la concordancia entre los procedimientos operativos y los métodos de medición y sirve para identificar y señalar todos aquellos valores que no representen situaciones reales en la medición de la calidad del aire y que fueron producto de errores o de actividades programadas en el sistema de medición o para identificar eventos extraordinarios que afecten directamente los datos generados por la estación de medición.

1 Bandera: Es un código alfa-numérico que califica el estado de cada dato y que define si éste puede ser utilizado para reportes con un mayor o menor nivel de certidumbre. En caso contrario el dato se invalida.

Es importante señalar que la limpieza y la verificación de datos se encuentran estrechamente ligadas y que el buen desempeño de ambas dependerá de una adecuada coordinación y comunicación entre los diferentes equipos de trabajo (operadores y analistas de datos), así como de la existencia y disponibilidad de todos los registros operativos (hojas de calibración, bitácoras, listas de verificación, programas de mantenimiento y cartas de registro.) Esta organización interna del personal y los registros serán de gran utilidad para la comprobación de los eventos, los cuales serán turnados a cada equipo revisor, quien les dará seguimiento, los confirmará y valorará. (Web 3, 2007. US-EPA, 2002).

3.2.1.1. Limpieza de datos.

Es la etapa inicial del manejo de datos. En ésta se aplican una serie de criterios, de forma manual, semiautomática o automática en algunos casos, para diferenciar los datos correctos de los falsos o incorrectos. Es decir, aquellos valores que no corresponden a mediciones de calidad del aire.

En la limpieza de datos se asignan por lo menos cinco banderas que se señalan en el Cuadro 2. Estas banderas se encuentran relacionadas directamente con los sucesos (programados y no programados) que se presentaron durante el periodo de medición (por ejemplo: pruebas de cero y span de los equipos de medición, mantenimiento y calibración, y cortes de energía eléctrica). Estos sucesos deben estar debidamente identificados en la bitácora de operación y en algunos casos también en la bitácora electrónica del sistema de adquisición de datos con la bandera respectiva ligada al dato. En cuanto a la asignación de banderas de limpieza con ayuda del SAD, su aplicación dependerá de dos elementos: 1) las características del SAD (marca y modelo) y 2) la adecuada programación de sus funciones.

Bandera	Significado del dato
VA	Válido
IC	Inválido por calibración
IR	Inválido por rango de operación
VZ	Válido igualado a cero o al límite de detección
ND	No disponible
IF	Inválido por falla del Equipo de Monitoreo

Cuadro 2. Banderas usadas en la limpieza de datos.

A continuación se describen ampliamente estas banderas:

- VA (Dato válido). Aquel dato que se considera correcto o verdadero pero que está sujeto a revisión en la etapa de verificación de datos.
- IC (Dato inválido por calibración). En esta categoría serán incluidos los datos que provengan de las actividades de calibración de los equipos de medición. Asimismo, se incluirán aquellos datos generados como resultado de la ejecución del programa de pruebas de cero/span.
- IR (Dato inválido por rango de operación). En esta categoría se incluyen valores fuera de los límites superior e inferior fijados en el equipo de medición (ver Cuadro 3 y Cuadro 4). Para el caso de los contaminantes, el límite superior se fija de acuerdo a las condiciones de la localidad y podrá diferir en aquellas localidades que se encuentren en zonas críticas de calidad del aire, con alta densidad industrial, en donde los niveles medidos pudieran ser mayores que los límites

superiores recomendados en este documento. Para el caso de partículas suspendidas e hidrocarburos, el Cuadro 3 sugiere dos intervalos de operación que comúnmente se usan en estos equipos de medición.

- VZ (Dato válido igualado al límite de detección o a cero). Esta bandera se aplica a datos con valores negativos los que, excepto en el caso de las mediciones de temperatura, deberán ser igualados a cero o al límite de detección correspondiente para poderlos incluir en la base de datos revisada. El Cuadro 5 (SMA-DF, 2004) señala tanto los contaminantes como los valores límite tolerados para la asignación de esta bandera. El uso de esta bandera es necesario por tres razones: 1) los equipos de medición de gases pueden presentar un corrimiento electrónico que por lo general es negativo y que muchas veces no es recomendable ajustar con una frecuencia mayor a la que establece el programa de control y aseguramiento de calidad; 2) permite cuantificar la cantidad de datos negativos encontrados por cada equipo de medición y sirve como un indicador cuantitativo de la calidad de la medición; y 3) la modificación sugerida para el dato corresponde a menos del 1% del intervalo de operación del equipo de medición por lo que los valores negativos que son resultado de un mal funcionamiento deberán quedar excluidos de este tipo de ajuste.

- IF (Dato inválido por falla en equipo). Esta bandera debe asignarse a aquellos datos de los que exista evidencia en bitácora o en listas de verificación de alarmas en equipos, es decir, aquellos cuyos parámetros de operación (temperaturas, flujos, presiones, voltajes) se encuentran fuera de los valores normales de operación.

- ND (Dato que no está disponible). Se aplica a cualquier dato ausente ya sea por fallas en la transmisión-recepción o por cortes de energía. A este dato se le asigna en la base de datos un número que le dé carácter de nulidad, por ejemplo 9999 (Web 4).

Contaminante	Rango de operación
Ozono (O ₃)	0-500 ppb
Dióxido de azufre (SO ₂)	0-500 ppb
Dióxido de nitrógeno (NO ₂)	0-500 ppb
Óxido nítrico (NO)	0-500 ppb
Óxidos de nitrógeno (NO _x)	0-500 ppb
Monóxido de carbono (CO)	0-50 ppm
Hidrocarburos totales	0-10 / 0-20 ppmc
Metanos	0-10 / 0-20 ppmc
No metanos	0-10 / 0-20 ppmc
Partículas PM ₁₀	0-500 / 0-1000 µg/m ³
Partículas PM _{2.5}	0-500 / 0-1000 µg/m ³

Cuadro 3. Rangos típicos de operación de los equipos automáticos de medición.

Parámetro	Rango de operación
Temperatura ambiental	-50 a 500 C
Humedad relativa	0 – 100 %
Presión Barométrica	500 – 760 mmHg
Precipitación Pluvial	0 – 10 mm
Radiación Solar	0-2000 W/m ²
Radiación UV-A*	0 – 300 mW/m ² 0 – 2000 mW/m ²
Radiación UV-B*	0 – 5 MED/hr 0 - 600 mW/m ²
Velocidad de Viento	0 – 50 m/s
Dirección de Viento	0 – 360°

*Para estos instrumentos se muestran los intervalos de operación de dos de las marcas usadas con mayor frecuencia, Solar Light y Keep & Zonnen.

Cuadro 4. Rangos de operación de los equipos meteorológicos.

Contaminante	Mínimo inferior*
Ozono (O ₃)	-3 ppb
Dióxido de Azufre (SO ₂)	-3 ppb
Dióxido de Nitrógeno (NO ₂)	-3 ppb
Oxido Nítrico (NO)	-3 ppb
Óxidos de Nitrógeno (NO _x)	-6 ppb
Monóxido de Carbono (CO)	-0.4 ppm

Fuente: SMA-DF, 2004.

Cuadro 5. Límite inferior sugerido para valores negativos.

La asignación de las banderas IR, relacionadas con los rangos de operación definidos para cada contaminante o parámetro meteorológico medido, y VZ, relativa a la eliminación o ajuste de valores negativos con base en los criterios definidos por el sistema de control y aseguramiento de la calidad se pueden establecer de forma automática mediante macros de Excel.

3.2.1.2. Macros para la asignación de banderas durante la limpieza de datos

Como resultado del procedimiento de limpieza de datos planteado en este documento la Dirección General del Centro Nacional de Investigación y Capacitación Ambiental (DGCENICA) del INE ha estado trabajando en el desarrollo de macros en Excel® para “limpiar” los datos horarios conforme a los intervalos de operación de los equipos de medición y a los valores negativos con la tolerancia adecuada para cada contaminante. Estas macros asignan automáticamente las banderas relacionadas con cada uno de los casos que se han descrito.

El uso de las macros parte de las siguientes consideraciones:

1. Se deben respetar las banderas que ya están asignadas y en caso necesario asignar aquellas que corresponden a datos fuera

de intervalo de operación, igualar a cero los valores negativos que están dentro del límite de tolerancia, abanderándolos con IR, VZ o ND, según sea el caso.

2. En ningún momento se borran datos. Solamente se señalan con banderas para que posteriormente sean verificados por el personal responsable de las estaciones de medición o usados para alguna estadística sobre el funcionamiento de los equipos de medición.

3. Se usan dos columnas en la base de datos en relación a cada estación de medición. En una se manejan los valores adquiridos por los equipos de medición y en la otra las banderas asignadas a los datos (Ver Cuadro 6).

El programa que contiene las macros de limpieza de datos estará disponible en la página Web del Instituto Nacional de Ecología (www.ine.gob.mx). Estas macros fueron diseñadas para utilizarse en el manejo de datos de las estaciones de medición de calidad del aire de la DGCENICA y deberán adaptarse a las condiciones particulares de los equipos de medición de cada SMCA.

Hora	O3	B1	SO ₂	B2	NO	B3	CO	B4	PM ₁₀	B5
01:00	0.042	VA	-0.004	IR	0.0004	VZ	3.296	VA	9999	ND
02:00	0.044	VA	-0.004	IR	0.0004	VZ	3.266	VA	9999	ND
03:00	0.044	VA	-0.004	IR	0.0004	VZ	3.442	VA	9999	ND
04:00	0.043	VA	0.001	VZ	0.0004	VZ	3.584	VA	9999	ND
05:00	0.044	VA	0.001	VZ	0.0004	VZ	4.266	VA	9999	ND
06:00	0.046	VA	0.001	VZ	0.0004	VZ	4.049	VA	9999	ND
07:00	0.041	VA	0.001	VZ	0.001	DV	3.852	VA	9999	ND
08:00	0.043	VA	-0.004	IR	0.004	DV	3.231	VA	9999	ND

Cuadro 6 Ejemplo de la aplicación de la macro de limpieza de datos.

3.2.2. Verificación de datos.

La verificación se considera como la etapa intermedia del manejo de datos y consiste en una revisión a detalle que se aplica en función de los objetivos y alcances de cada red de medición de la calidad del aire. En esta etapa se pretende, como su nombre lo indica, confirmar la veracidad de datos que parezcan sospechosos o con valores extremos y que hayan sido identificados en la revisión de las bases de datos por medio de procedimientos estadísticos. En esta etapa se incluye la comparación de los datos con algunas fuentes externas de información como serían la correlación espacial y temporal con otros sitios de monitoreo cercanos, los resultados de auditorías de equipos o la confrontación con datos a otras escalas (US-EPA, 1998). Es por lo tanto, una revisión a detalle de los datos para confirmar o mo-

dificar las banderas asignadas en la etapa de limpieza así como para añadir más banderas dependiendo de la experiencia del evaluador y la tendencia de los datos. Las banderas más usadas en la verificación de datos se señalan en el Cuadro 7.

Bandera	Significado del dato
IO	Inválido por operador
VC	Válido calculado
VE	Válido con evento extraordinario
DS	Dato sospechoso

Cuadro 7 Banderas usadas en la verificación de datos.

La verificación de datos se puede realizar en dos partes, una automática con base en los criterios que se describen en el apartado siguiente y la otra manual basada en la experiencia del evaluador.

3.2.2.1. Criterios de verificación automática de datos.

Para la asignación automática de banderas la US-EPA y la California Air Resources Board, CARB, sugieren utilizar los siguientes criterios (US-EPA, 2000a. CARB, 2001):

Para contaminantes:

1. Señala datos, es decir coloca banderas, cuando existen valores constantes por más de tres horas consecutivas de CO, NOx, NO2, NO, O3, SO2, PM10, PM2.5. (Asignar bandera DS).
2. Señala datos cuando la suma de NO y NO2 dividida entre NOx se encuentra fuera del intervalo (0.85, 1.15). (Asignar bandera IO).
3. Señala datos cuando la relación de PM2.5 entre PM10 es mayor a 1.15. (Asignar bandera IO).

Para datos meteorológicos (USEPA APT1, 1989. USEPA, 2000b. Web 5, 2003):

1. Señala datos de radiación (total, UVA, UVB, UVC) cuando ésta es diferente a cero durante la noche. (Asignar bandera IO).
2. Señala datos de velocidad de viento cuando ésta no varía en más de 0.1 m/s en tres horas consecutivas. (Asignar bandera IO).
3. Señala datos de velocidad de viento cuando éstos no varían en más de 0.5 m/s en doce horas consecutivas. (Asignar bandera IO).
4. Señala datos de dirección de viento cuando ésta no varía en más de 1° por más de tres horas consecutivas. (Asignar bandera IO).

5. Señala datos de dirección de viento cuando ésta no varía en más de 10° durante dieciocho horas consecutivas. (Asignar bandera IO).

6. Señala datos de temperatura si ésta cambia en más de 5 °C respecto a la hora previa. (Asignar bandera IO).

7. Señala datos de temperatura si ésta no varía en más de 0.5 °C en doce horas consecutivas. (Asignar bandera IO).

8. Señala datos de presión barométrica cuando ésta tiene cambios de más de 0.75 mm Hg en tres horas consecutivas. (Asignar bandera IO).

3.2.2.2. Criterios de verificación manual de datos

Para la asignación de cada bandera manual en la verificación de datos se usan las siguientes consideraciones:

IO (Dato inválido por operador). Es una bandera que aplica el operador basándose en un procedimiento estadístico en el que se debe revisar la congruencia temporal y espacial de los datos, así como las relaciones entre contaminantes y/o las relaciones de éstos con los datos meteorológicos (USEPA- APT1, 1984). Esta bandera abarca una lista de criterios y consideraciones algunos de los cuales se pueden automatizar (sección 3. 2. 2.1).

Sin embargo, la asignación y el buen uso de esta bandera depende de la experiencia del operador y/o del analista quien deberá saber identificar los patrones típicos de cada contaminante (actividad diurna y nocturna) y estar familiarizado con la operación de los equipos de medición y sus fallas comunes. Incluye el uso de series de tiempo, y la revisión de la estadística básica: máximos, mínimos, promedios, entre otros. (NAPS, 2004. Martínez, A. 1996. UNEP/WHO, GEMS, AIR, 1994). A continuación se muestran algunas consideraciones para la asignación de esta bandera:

Series de Tiempo

Por lo general, el CO, HC y NO aumentan y disminuyen juntos. El NO y el O₃ presentan una relación inversa y no pueden existir juntos a niveles superiores a sus valores límite de calidad del aire.

Temperatura interna de la estación

Este valor debe de mantenerse entre 20 y 30° C (SMA-GDF, 2004) ya que los equipos de medición de gases presentan respuestas erróneas a temperaturas elevadas, sobre todo los de bióxido de azufre y los de óxidos de nitrógeno. En caso de no contar con registros automáticos de la temperatura interna de la estación, para el manejo de datos se podrán utilizar los registros de alarma de

temperatura que presentan los equipos de medición en sus parámetros de operación y éstos deberán anotarse en la bitácora de la estación. El o los datos provenientes de los equipos de medición que presentaron este tipo de alarma deberán ser marcados con la bandera (IO) durante el período de afectación.

Períodos de estabilización de equipos de medición por cortes de energía eléctrica.

Este evento se relaciona con cortes en el suministro de energía eléctrica. El número de datos a los que se asignará esta bandera es variable (desde 3 a 5 minutos hasta un máximo de 60 minutos) y estará en función, tanto de la duración del corte de energía eléctrica, como de las diferentes marcas y modelos de equipos de medición empleados. Las series de tiempo son una herramienta útil para identificar este tipo de eventos. Asimismo, se recomienda marcar con color los cortes de energía identificados en la base de datos a revisar.

VC (Dato válido calculado). Este tipo de bandera se usa en aquellos datos que son obtenidos por medio de interpolación o por el uso de series de tiempo. También es aplicable a aquellos datos que son corregidos de acuerdo con los resultados de las verificaciones de cero y span. Es importante comentar que esta corrección se lleva a cabo después de una evaluación previa y minuciosa del equi-

po de medición que originó los datos ya que aquellos producidos por equipos con antecedentes de mal funcionamiento no podrán ser objeto de este tipo de ajuste y deberán ser invalidados. Esta bandera puede usarse cuando se corrige el valor de dióxido de nitrógeno, siempre y cuando su valor no supere la tolerancia establecida en el sistema de control y aseguramiento de la calidad.

VE (Dato válido con evento extraordinario). Esta bandera se aplica a los datos cuya frecuencia y duración han sido registrados en la bitácora general de la estación o cuando existe alguna otra fuente documental de eventos que afectaron la medición, tales como: incendios, tolveneras o actividades de construcción.

DV (Dato a verificar). Es una bandera temporal que incluye datos que deberán ser confirmados en alguna de las categorías anteriores, según corresponda (verificado o no), con ayuda de las bitácoras de operación y de los criterios y consideraciones mencionados. También se puede asignar esta bandera a datos que se presentan fuera de los límites establecidos en base a tendencias históricas de cada uno de los contaminantes medidos o parámetros meteorológicos (NAPS, 2004). Un ejemplo de aplicación de esta bandera son los niveles de alarma que se han establecido para la aplicación de planes de contingencia.

3.2.2.3. *Herramientas estadísticas útiles para la verificación de datos.*

La US-EPA, en su guía para evaluar la calidad de los datos, describe en el apartado de revisión y verificación de datos las principales herramientas estadísticas que se pueden emplear para la verificación de las bases de datos. Estas herramientas estadísticas se dividen en: estadística descriptiva básica y gráficos (US-EPA, 2000c). A continuación se listan cada una de ellas. Es importante mencionar que algunos de los cálculos y gráficos que se utilizan en la verificación de datos también se utilizan para la generación de indicadores y la validación. El anexo III describe con mayor detalle cada herramienta y da ejemplos de aplicación.

Estadística descriptiva básica.

En algunas investigaciones se obtienen numerosos datos que deben reducirse para lograr una interpretación adecuada. En estas situaciones la estadística descriptiva se utiliza como un valioso instrumento para describir y analizar las características de las observaciones y para encontrar las relaciones que existen con otros conjuntos de datos con los que se comparen. El objetivo de esta actividad es resumir las características cuantitativas del conjunto de datos utilizando estadística descriptiva básica. Algunos indicadores útiles en el análisis incluyen el número de observaciones, las medidas de tendencia central (media, mediana o moda),

medidas de posición relativa (percentiles), medidas de dispersión (rango, varianza, desviación estándar, coeficientes de variación, rango intercuartílico) y medidas de asociación (correlación).

Gráficas.

El propósito es identificar patrones y tendencias en los datos, los cuales podrían pasar inadvertidos al usar únicamente métodos numéricos. Las gráficas pueden usarse para descubrir eventos significativos, identificar problemas potenciales y para sugerir medidas correctivas. Además, algunas representaciones gráficas se utilizan para registrar y guardar datos compactados o para presentar la información de otra manera. Las presentaciones gráficas de los datos incluyen el despliegue individual de los datos, cantidades estadísticas, datos temporales, datos espaciales y dos ó más variables. La persona que este haciendo uso de las gráficas deberá escoger aquellas que muestren las principales características de los datos, debido a que un solo gráfico no proveerá la información completa de los datos. Se sugiere escoger el menor número de gráficas que presenten la información completa de los datos. Si un conjunto tiene un componente espacial o temporal se seleccionará la gráfica que mejor represente ese componente. Si el conjunto de datos comprende más de una variable se deberá tratar cada variable individualmente antes de desarrollar gráficas para varias variables.

3.3. Sistemas de Medición Manual.

Esta sección es aplicable a aquellos sistemas de medición que utilizan el método manual para la determinación de partículas suspendidas (PST, PM₁₀ y PM_{2.5}). Las partículas suspendidas son generalmente un sistema complejo multifase, partículas de baja presión de vapor que son transportadas por el aire y cuyo diámetro aerodinámico va de 0.01 a 100 µm y mayores (US-EPA, 1999).

El aseguramiento de calidad de los datos provenientes de muestreadores de partículas del tipo de alto volumen incluye toda la operación del muestreador e implica dar seguimiento al procedimiento estandarizado que viene descrito de forma general en la NOM-035-ECOL-1993 para la determinación de partículas suspendidas totales, PST. Sin embargo, este procedimiento se ajustará de acuerdo con la marca y modelo del muestreador que se esté usando, por lo cual cada SMCA deberá establecer su propio procedimiento estándar de operación.

En el caso de los muestreadores de PM₁₀ y PM_{2.5} para los que no se tiene norma muchas de las consideraciones que señala la NOM-035 son aplicables siempre y cuando se utilice el método de alto volumen que depende de dos variables: la masa colectada en el filtro y el volumen de aire muestreado. Además del procedimiento operativo se debe contar con un procedimiento de aseguramiento y control de calidad que garan-

tice que la calidad de la medición de estas dos variables (masa y volumen) es la adecuada. El procedimiento operativo referente al mantenimiento y calibración de los muestreadores de partículas de alto volumen se encuentra en el Manual 4: Operación, Mantenimiento y Calibración de los Componentes de una Estación de Medición de la Calidad del Aire. La Figura 3 presenta un diagrama que esquematiza los pasos a seguir en el proceso operativo para asegurar la calidad de los datos en un muestreo de partículas.

El detalle de este procedimiento se describe en la NOM-035-ECOL-1993 y en el CFR

Parte 50 Apéndice B, que señalan el método de determinación de PST, y el CFR Parte 50.

Apéndice J (Web 3) y en US-EPA, Quality Assurance Handbook, que señalan el método para la determinación de PM₁₀.

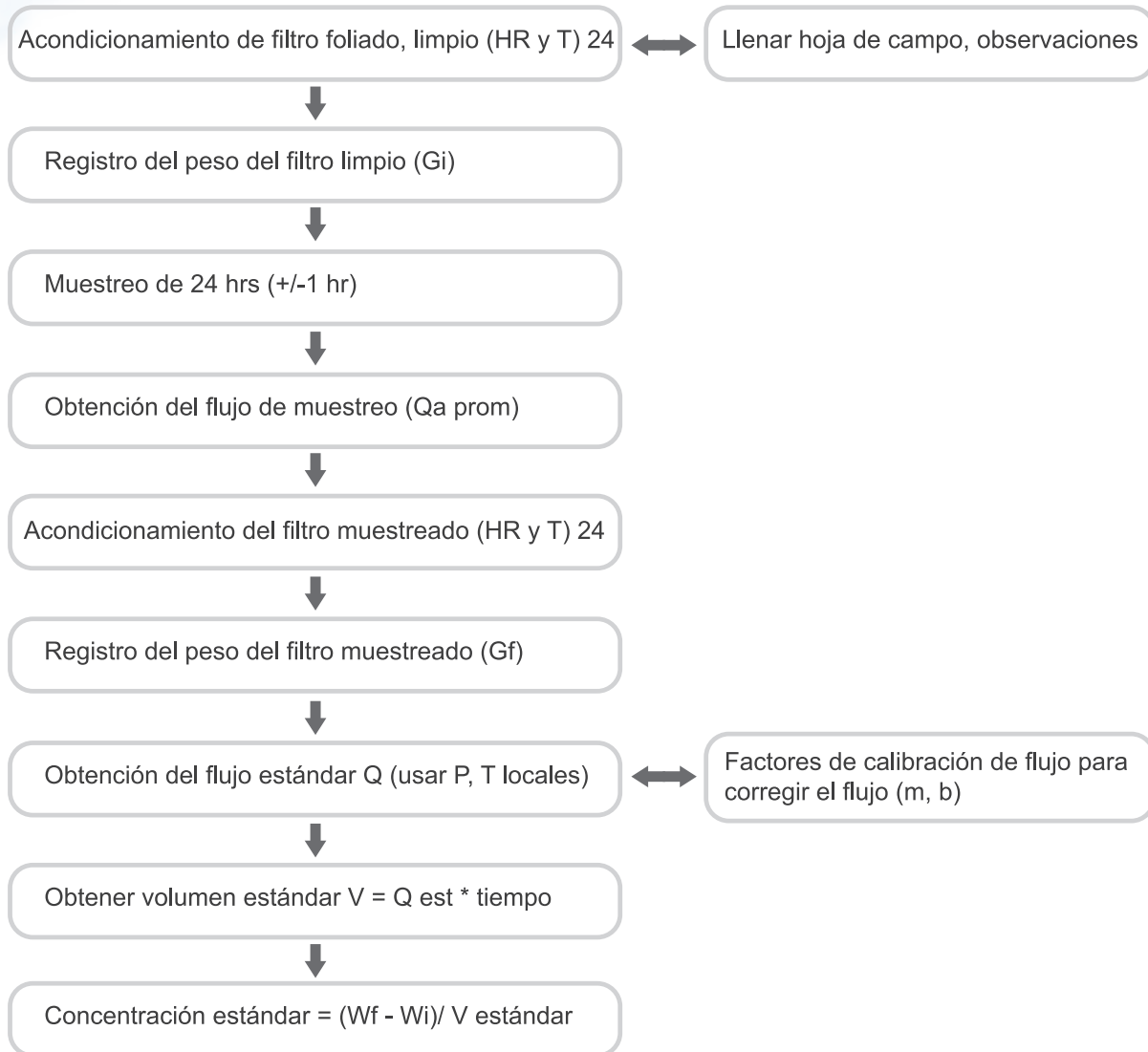


Figura 3. Generación de datos en el muestreo de partículas suspendidas.

3.3.1. Banderas sugeridas y su relación con el Control y Aseguramiento de la Calidad.

Cuando se ha llevado a cabo el procedimiento operativo, el de mantenimiento y los controles de calidad, finalmente se conforma una base de datos en donde se registran los resultados de masa y volumen obtenidos durante el muestreo y se calculan y revisan las concentraciones. Se sugiere agregar a esta base de datos, al igual que a la base de datos de la medición automática, una columna con la bandera que lo califique al lado derecho de cada dato. Las cinco banderas sugeridas a usarse se señalan en el Cuadro 8, que se presenta a continuación:

Bandera	Significado del dato
VA	Dato Válido.
IF	Filtro en mal estado (roto, rasgado, incompleto)
IQ	Flujo de operación menor a 1.1 m ³ /min o inestable
IN	Diferencia de peso negativa. Implica pérdida de peso
IT	Tiempo de muestreo menor a 18 horas

Cuadro 8. Banderas sugeridas para el muestreo manual.

3.4. Características de las bases de datos verificadas.

Una vez que los datos han sido limpiados y verificados y partiendo de que la base de datos tiene el formato sugerido a principio de este capítulo, en los siguientes apartados de esta sección se describen las características que debe tener la base de datos verificada.

3.4.1. Unidades de medición, claves de parámetros y cifras significativas.

En México existe una creciente necesidad de homologar las bases de datos verificadas para poder comparar la información de calidad del aire que se genera a nivel nacional.

De esta forma, las bases de datos verificadas deberán tener establecidos los nombres de cada uno de los parámetros que las integran, así como las unidades de medición y las cifras significativas (US-EPA y la Estrategia Americana de Investigación del Ozono Troposférico, NARSTO, por sus siglas en inglés). La homologación de las variables y atributos de las bases de datos verificadas permitirá que éstas puedan ser utilizadas por todos los usuarios que las requieran.

Atendiendo a las sugerencias anteriores, en el Cuadro 9 se establecen las unidades de medición, las cifras significativas y las claves que se deben usar para cada parámetro medido. Es importante destacar que las cifras significativas han sido definidas con base en la precisión y en el límite de detección de las diferentes marcas de equipos de medición que se emplean en el país. Las características de los datos que provienen del muestreo manual, como cifras significativas, redondeos y unidades de reporte se definen en la NOM-025-SSA1-1993 en su versión del 26 de septiembre de 2005.

A continuación se presenta un ejemplo de la aplicación del número de cifras significativas que se deben usar para el caso del ozono. Como el Cuadro 9 señala, para este parámetro se consideran cuatro cifras significativas, una de ellas entre paréntesis, corresponde a un decimal. De esta forma, se tienen disponibles 4 valores, 3 para enteros y uno para el decimal. Si durante la medición se obtienen por ejemplo 68.698 ppb de ozono, el valor de esta concentración en la base de datos deberá ser redondeado sólo al primer decimal obteniendo 68.7 ppb. Esto se basa en que el equipo de medición tiene un límite de detección de 0.6 ppb. Mayores detalles sobre los límites de detección de otros equipos de medición de calidad del aire pueden ser consultados en el Anexo II.

Asimismo, y para facilitar el manejo de datos, en sus etapas de limpieza, verificación, y validación, se sugiere que el archivo sea guardado manteniendo los registros con una resolución horaria y con una frecuencia mensual, es decir un archivo por mes. Posteriormente, al término de un año, se deberá compilar el archivo con el total de registros obtenidos durante los doce meses.

Parámetro	Fórmula	Unidad de medición	Cifras significativas (decimales)
Ozono	O ₃	ppb	4 (1)
Óxidos de Nitrógeno	NO _x	ppb	4 (1)
Óxido nítrico	NO	ppb	4 (1)
Dióxido de nitrógeno	NO ₂	ppb	4 (1)
Monóxido de carbon	CO	ppm	5 (2)
Dióxido de azufre	SO ₂	ppb	4 (1)
Hidrocarburos totales	HCT	ppm	4 (1)
Partículas suspendidas totales	PST	µg/m ³	4 (0)
Partículas suspendidas con diámetros menores a 10 micras	PM ₁₀	µg/m ³	4 (0)
Partículas suspendidas con diámetros menores a 2.5 micras	PM _{2.5}	µg/m ³	4 (0)
Metano	CH ₄	ppm	4 (1)
Hidrocarburos diferentes al metano	HDM	ppm	4 (1)
Ácido sulfhídrico	H ₂ S	ppm	4 (1)
Velocidad del viento	VV	m/s	3 (1)
Dirección del viento	DV	grados	3 (0)
Temperatura	TMP	°C	3 (1)
Humedad Relativa	HR	%	3 (0)
Radiación Solar	RS	W/m ²	3 (0)
Radiación Ultravioleta A	UVA	mW/m ²	3 (0)
Radiación Ultravioleta B	UVB	MED/hr	3 (0)
Presión barométrica	PB	mb	4 (1)
Precipitación pluvial	PP	mm	2 (1)

Fuente: US-EPA QA/G8, 1999; NARSTO, 2006; NOM-025-SSA1/1993; y Web 6

Cuadro 9. Parámetros y características de la base de datos verificada.

4. PROCEDIMIENTO PARA OBTENER INDICADORES DE LA CALIDAD DEL AIRE.

Como se mencionó, un indicador se define como un valor que cuantifica y simplifica un fenómeno, y que ayuda a entender condiciones complejas (Web 1). En este caso los indicadores de la calidad del aire permiten, entre otras cosas, evaluar el estado de la contaminación atmosférica y comunicar al público cuál es la calidad del aire que respira.

En la Dirección General de Investigación sobre la Contaminación Urbana y Regional (DGICUR) del Instituto Nacional de Ecología (INE) se desarrolló un Sistema de Consulta de Indicadores de Calidad del Aire (SICA).

En este sistema se automatiza el procesamiento de las bases de datos de calidad del aire históricas de algunas ciudades que han entregado sus bases de datos validadas para tal efecto a la DGICUR. El SICA incluye la generación de indicadores de calidad del aire, reduciendo con ello la probabilidad de errores humanos durante su procesamiento. En el sistema el usuario puede realizar consultas de los indicadores de los contaminantes criterio PM₁₀, PM_{2.5}, O₃, CO, SO₂ y NO₂ por ciudad o zona metropolitana (estaciones de monitoreo agregadas), por cada una de las estaciones de monitoreo; para cualquier periodo de tiempo, y por año, mes, semana o día; tanto para las mediciones automáticas como manuales. Entre otras consultas, el

usuario puede solicitar --con diferentes criterios de compleción de datos-- las estadísticas descriptivas básicas de cada parámetro, los promedios hora a hora, los promedios día a día, los promedios mes a mes y frecuencias absolutas y relativas de la dirección y velocidad del viento; así como los indicadores relacionados con las NOM de calidad del aire.

El SICA estará disponible para el público en general a través de la página de Internet de la Dirección de Investigación sobre Calidad del Aire (<http://www.ine.gob.mx/dica>).

Por otra parte, es necesario que los responsables de las redes de monitoreo que no están incluidas en el SICA realicen al cálculo de los diferentes indicadores de la calidad del aire de una manera homogénea y con la menor incertidumbre posible. En este capítulo se dan elementos técnicos para el procesamiento y manejo de los datos registrados en los SMCA con equipos manuales o automáticos. El procesamiento y manejo de datos se lleva a cabo con las bases de datos validadas, esto es, con aquellas bases que cumplieron con el esquema de AC y CC.

4.1. Tipos de datos que se utilizan para el cálculo de distintos indicadores de la calidad del aire.

Los indicadores se calculan con diferentes tipos de datos (promedios de una hora, promedios móviles y concentraciones diarias), de acuerdo con el equipo de medición con el que se generaron y el fenómeno que representan.

El promedio de una hora (también referido como dato horario o concentración horaria) es el promedio de las concentraciones minutas o de algún intervalo de tiempo dentro de la hora. El promedio se calcula sobre el periodo de tiempo anterior a la hora que se está determinando, esto es, si se quiere determinar el promedio de la hora 01:00, éste se obtiene con datos del periodo que comprende de las 00:01 a las 01:00 horas (Figura 4).

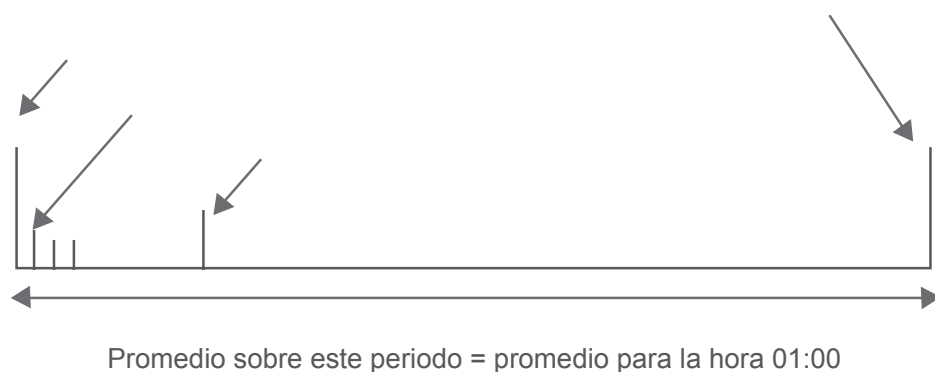


Figura 4. Descripción del promedio de una hora.

Los promedios móviles de 8 horas de CO y O₃ se calculan de las concentraciones horarias, tomando el promedio de la hora seleccionada con las 7 concentraciones registradas en las horas previas. Por ejemplo, para estimar el promedio móvil de 8 horas de las 13:00 horas, se calcula el promedio de las concentraciones registradas para cada hora desde las 06:00 hasta la 13:00 horas (Figura 5).

Las concentraciones diarias, dependiendo del contaminante, se calculan a partir de las concentraciones horarias o de los promedios móviles de 8 horas y representan: el valor máximo diario, el promedio diario o el muestreo de 24 horas.

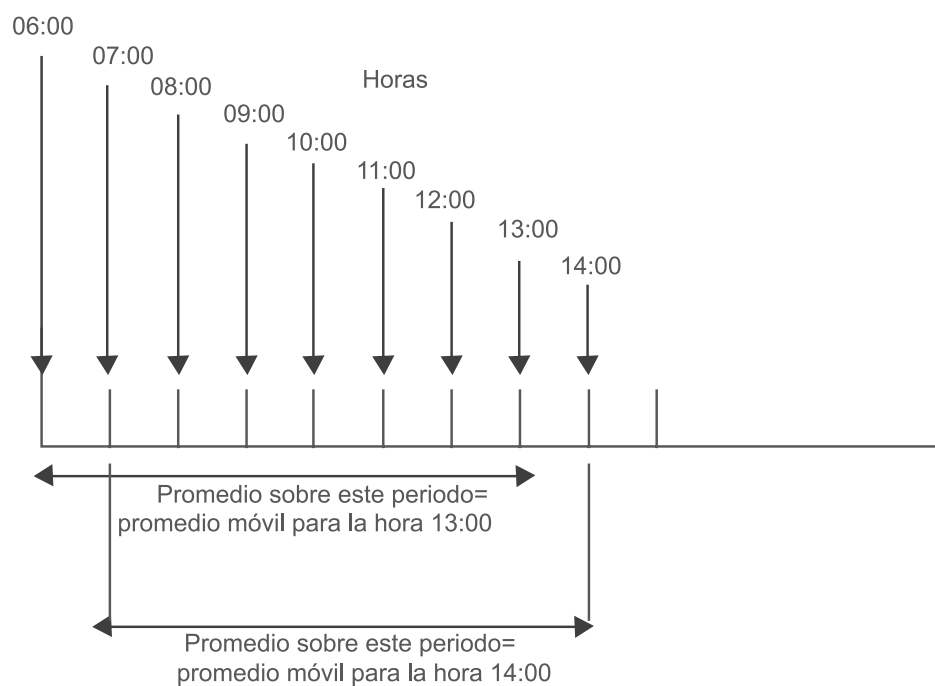


Figura 5. Descripción de los promedios móviles horarios de 8 horas.

Máximo diario (O₃, NO₂, y CO)

Se refiere al valor más alto de los 24 valores horarios o promedios móviles registrados durante el día (Figura 6).

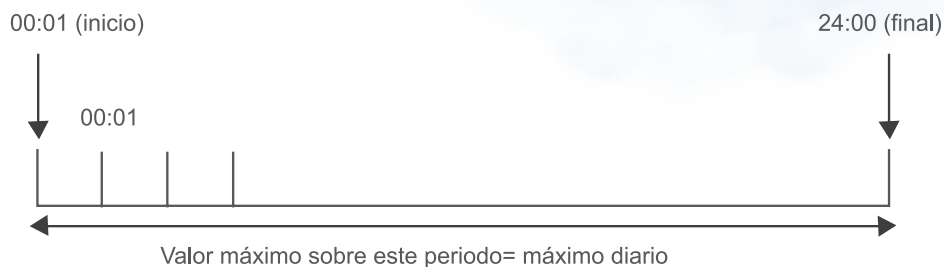


Figura 6. Descripción del máximo diario.

Promedio diario (PM₁₀, PM_{2.5}, y SO₂)

Se refiere al promedio de los 24 valores horarios registrados en el día (Figura 7).

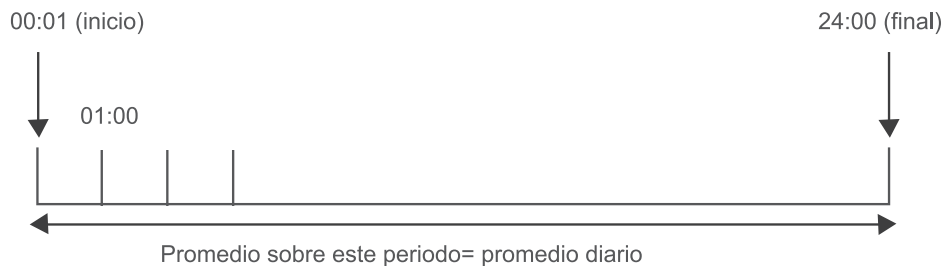


Figura 7. Descripción del promedio de 24 horas.

Muestreo de 24 horas (PM₁₀, PM_{2.5}, y PST),

Se refiere a la medición de la contaminación del aire por medio de la toma de muestras de forma discontinua, usualmente cada seis días. Para obtener los datos descritos en los párrafos anteriores, y que sean considerados representativos del periodo al que se refieren, se requiere contar con una cantidad mínima de información, es decir se deben de cumplir los criterios de completación de datos (ver sección 4.2).

En el Cuadro 10 se describe el tipo de dato dependiendo del equipo de medición, los requerimientos de compleción de datos y los parámetros que se les asocian para obtener los indicadores de la calidad del aire.

Equipo de monitoreo	Tipo de dato	Requerimiento de compleción de datos	Parámetros
Automático	Dato horario	Debe contar con un 75% o más de los registros minutales (45 minutos o más)	PM10, PM2.5, O3, CO, SO2, NO, NOx, NO2, TMP, HR, VV y DV
	Promedio móvil de 8 horas	Debe contar con un 75% o más de los registros horarios (de 6 a 8 datos horarios)	O3 y CO
	Máximo diario	Su base son los datos horarios de un día, y debe contar con un 75% o más de los registros horarios (18 horas o más)	O3, CO, NO, NOx, NO2, TMP, HR y VV
	Máximo diario	Su base son los promedios móviles de 8 horas de un día y debe contar con un 75% o más de los registros horarios (18 horas o más)	O3 y CO
	Promedio diario	Debe contar con un 75% o más de los registros horarios (18 horas o más)	PM10, PM25, SO2, TMP, HR y VV
Manual	Muestreo de 24 horas	Se acepta, si el muestreo duró entre 23-25 horas (NOM-035-ECOL-1993)	PST, PM10, PM2.5 y Pb

Cuadro 10. Tipos de datos para obtener los indicadores de la calidad del aire.

Una vez que se calculan los datos anteriores, es posible generar los indicadores de interés (ver sección 4.3 para el procedimiento de cálculo específico de cada uno). En el Cuadro 11 se muestran una serie de indicadores, para evaluar diferentes aspectos de la calidad del aire. En este caso, se muestran los indicadores de impacto en la salud (relacionados con las NOM de la Secretaría de Salud) y de comportamiento en el tiempo.

Cuadro 11. Indicadores de la calidad del aire.

Aspecto a evaluar	Indicador	Parámetro	Tipo de dato	Equipo de monitoreo
Impacto en la salud	Percentil 98	PST, PM ₁₀ y PM _{2.5}	Concentración diaria	Automático o manual
	Quinto máximo	O ₃	Concentración diaria de los promedios móviles de 8 horas	Automático
	Máximo	O ₃	Dato horario	Automático
	Segundo máximo	COa	Promedio móvil de 8 horas	Automático
		SO ₂	Concentración diaria	
		NO ₂	Dato horario	
	Número y porcentaje de horas arriba del límite de la norma	O ₃ y NO ₂	Dato horario	Automático
	Número y porcentajes de días arriba del límite de la norma	PM ₁₀ , PM _{2.5} y SO ₂	Concentración diaria	Automático
		O ₃ y NO ₂	Concentración diaria de datos horarios	
		CO	Concentración diaria de los promedios móviles de 8 horas	
Promedio anual	SO ₂	Dato horario	Automático	
	PM ₁₀ y PM _{2.5}	Concentración diaria	Automático o manual	
Comportamiento	Máximo y mínimo	PM ₁₀ , PM _{2.5} , O ₃ , CO, SO ₂ , NO, NO _x , NO ₂ , HR, TMP y VV	Dato horario	Automático
		CO y O ₃	Promedio móvil de 8 horas	
		PST, PM ₁₀ , PM _{2.5} , O ₃ , CO, SO ₂ , NO, NO _x y NO ₂	Concentración diaria	Automático o manual
		HR, TMP y VV	Máximo o promedio diario	Automático



Cuadro 11. Indicadores de la calidad del aire.

Comportamiento	Media o promedio	PM10, , PM2.5, O3, CO, SO2, NO, NOx, NO2, HR, TMP y VV	Dato horario	Automático
		CO y O3	Promedio móvil de 8 horas	
		PST, PM10, PM2.5, O3, CO, NO, NOx, NO2 y SO2	Concentración diaria	Automático o manual
		HR, TMP y VV	Máximo o promedio diario	Automático
	Promedio o mediana hora a hora	PM10, PM2.5 , O3, CO, SO2, NO, NOx, NO2, HR, TMP y VV	Dato horario	Automático
		CO y O3	Promedio móvil de 8 horas	
	Promedio o mediana hora a hora	PM10, PM2.5 , O3, CO, SO2, NO, NOx, NO2, HR, TMP y VV	Dato horario	Automático
		CO y O3	Promedio móvil de 8 horas	
	Promedio o mediana día a día	PM10, PM2.5 , O3, CO, SO2, NO, NOx, NO2, HR, TMP y VV	Dato horario	Automático
		CO y O3	Promedio móvil de 8 horas	Automático
		PM10, PM2.5 , O3, CO, SO2, NO, NOx y NO2	Concentración diaria	Automático o manual
		CO y O3	Concentración diaria de los promedios móviles de 8 horas	Automático
		HR, TMP y VV	Máximo o promedio diario	



Comportamiento	Promedio o mediana mes a mes	PM ₁₀ , PM _{2.5} , O ₃ , CO, SO ₂ , NO, NO _x , NO ₂ , HR, TMP y VV	Dato horario	Automático
		CO y O ₃	Promedio móvil de 8 horas	
		PM ₁₀ , PM _{2.5} , O ₃ , CO, SO ₂ , NO, NO _x y NO ₂	Concentración diaria	Automático o manual
		CO y O ₃	Concentración diaria de los promedios móviles de 8 horas	Automático
	Percentiles	HR, TMP y VV	Máximo o promedio diario	Automático
		PM ₁₀ , PM _{2.5} , O ₃ , CO, SO ₂ , NO, NO _x , NO ₂ , HR, TMP y VV	Dato horario	
		CO y O ₃	Promedio móvil de 8 horas	
		PST, PM ₁₀ , PM _{2.5} , O ₃ , CO, SO ₂ , NO, NO _x y NO ₂	Concentración diaria	
Tendencia	Intervalos de concentración	PM ₁₀ y PM _{2.5} , O ₃ , CO, SO ₂ y NO ₂	Dato horario	Automático
		CO y O ₃	Promedio móvil de 8 horas	
		PST, PM ₁₀ , PM _{2.5} , O ₃ , CO, SO ₂ y NO ₂	Concentración diaria	Automático o manual

* Concentración diaria se refiere al dato diario que le corresponde a cada contaminante de acuerdo a la tabla 3.1.

* El segundo máximo se obtiene sin traslape de información. Esto es, que el valor de los promedios móviles de ocho horas que se calcule como segundo máximo no incluya concentraciones horarias con las que se calculó el promedio móvil de ocho horas que originó el primer máximo.

La lista de indicadores podría ser más extensa; sin embargo, los que se muestran son suficientes para evaluar los diferentes aspectos de la calidad del aire. En el SIMAT de la ZMVM se obtienen otros indicadores más que evalúan tendencias: prueba de tendencia, prueba de homogeneidad y porcentaje de cambio; estos últimos indicadores no se describen en este documento.

Los indicadores se pueden reportar para representar las condiciones de una ciudad o de una zona metropolitana, pero debido a que en nuestro país son muy pocas las redes de medición que han evaluado la representatividad de sus estaciones se recomienda que el reporte de indicadores se haga para cada estación de medición, en tanto no se caracterice, evalúe y reporte su representatividad (Zuk et al., 2007).

4.2. Nota importante sobre los criterios de completación de datos.

Los equipos automáticos arrojan de manera continua registros, minuto a minuto, pero debido a la capacidad de almacenamiento que demanda el alto volumen de datos, por lo general las bases reportadas contienen promedios horarios. En cambio, los equipos manuales arrojan datos diarios (de tres a seis por mes).

Con el propósito de que los indicadores de calidad del aire sean confiables se debe revi-

sar que la información que se utilice para su generación cuente con un mínimo de datos validados. Esto es, que se cumpla con los criterios de completación de datos.

El cumplimiento de la completación de datos se debe de revisar para cada contaminante o para cada variable meteorológica en cada una de las estaciones de medición, para identificar aquellas que cuenten con la información necesaria en una hora, en el día, en un mes, en un trimestre y en algunas ocasiones, en el año o durante un periodo de años (ver Cuadro 10). La revisión de la completación debe efectuarse sobre la base de datos validada, una vez que se han identificado los datos válidos y eliminado los registros incongruentes o aislados.

Actualmente, en México sólo se aplica el criterio de completación para validar los datos y generar indicadores de calidad del aire.

4.2.1. Desempeño anual y desempeño histórico por estación de medición.

Para generar indicadores que sean representativos del comportamiento de un contaminante o de una variable meteorológica a lo largo de uno o varios años se recomienda evaluar el desempeño anual y el histórico de las estaciones de medición. El desempeño anual se define como el porcentaje de datos

válidos de cada estación de medición en un año y el desempeño histórico como el porcentaje de años que una estación de medición opera con un buen desempeño anual.

El SIMAT de la ZMVM califica el desempeño anual de una estación de medición como bueno cuando registra el 75% o más de datos válidos; regular, cuando registra el 50% o más de los datos sin llegar al 75%; y malo cuando registra menos del 50% de datos válidos (Cuadro 12).

Las estaciones de medición que son etiquetadas con un desempeño anual bueno tienen información completa para la generación de indicadores, mientras que las estaciones de medición con una clasificación menor requieren mejoras para que en el futuro registren mayor cantidad de datos válidos.

Equipo de monitoreo	Tipo de dato	Datos horarios válidos		
		Malo (< 50%)	Regular (>= 50% y < 75%)	Bueno (>= al 75%)
Automático	Horario	0 a 4391 (año bisiesto)	4392 a 6587 (año bisiesto)	6588 a 8784 (año bisiesto)
		0 a 4379 (año no bisiesto)	4380 a 6569 (año no bisiesto)	6570 a 8760 (año no bisiesto)
Manual	Muestreos de 24 horas	0 a 30	31 a 45	46 a 61

Cuadro 12. Categorías de desempeño anual.

Asimismo, el SIMAT califica el desempeño histórico de una estación de medición en tres categorías: bueno cuando el porcentaje es del 75% o más de los años; regular cuando es del 50% o más sin llegar al 75% de los años; y malo cuando es menor del 50% de los años (Cuadro 13).

Número de años de operación	Número de años con desempeño anual bueno (1)	Porcentaje de años con desempeño anual bueno	Desempeño histórico		
			Malo	Regular	Bueno
N	m	$\frac{100 \times m}{N}$	< 50%	>= 50% y < 75%	>= 75%

(1) No puede ser mayor a los años de operación.

Cuadro 13. Desempeño histórico de las estaciones de medición.

Por ejemplo, para calificar el desempeño histórico de tres estaciones de medición en un periodo de seis años, y suponiendo que se sabe que la primera estación tiene cinco años con desempeño anual bueno, la segunda cuatro y la tercera sólo dos; el porcentaje de desempeño anual bueno de las estaciones es 83% ($5/6 \times 100$), 67% ($4/6 \times 100$), y 33% ($2/6 \times 100$) respectivamente. Esto quiere decir que el desempeño histórico de cada una de las estaciones es bueno, regular y malo, respectivamente. Por lo anterior, se obtiene que la estación de medición que registró un desempeño histórico bueno, cuenta con información suficiente a lo largo de los años.

En la práctica, se recomienda que la información para generar indicadores de calidad del aire cumpla por lo menos con los criterios de completación de una hora, de un día, de un mes y de un trimestre (al menos el 75% de los minutos, horas o días respectivamente –Cuadro 10) y si es posible que se verifique el desempeño anual e histórico. Se recomienda que para cada una de las estaciones de medición se dé a conocer con qué cantidad de datos se obtienen los indicadores para que los usuarios les den un uso y una interpretación adecuados. En este sentido sólo el SIMAT de la ZMVM genera indicadores de las estaciones de medición que cumplen con un desempeño anual e histórico bueno.

4.3. Cómo se definen y calculan los indicadores de la calidad del aire.

4.3.1. Indicadores relacionados con las normas de calidad del aire

Como se mencionó en el Manual 2, uno de los objetivos de la medición de calidad del aire es evaluar el cumplimiento de las normas de calidad del aire. En general, las normas de calidad del aire se establecen con base en los resultados de estudios toxicológicos y epidemiológicos que evalúan la relación entre la exposición a un contaminante y sus efectos en la salud. Cuando existe una concentración umbral, es decir, un nivel abajo del cual no se identifican impactos en la

salud, se toma esa concentración como el límite permisible. Sin embargo, para algunos contaminantes, como las partículas suspendidas, aún no se han detectado umbrales para algunos efectos en la salud. De hecho, se han encontrado impactos en la salud aun a niveles muy bajos (Pope et al., 2002). Por lo tanto, el establecimiento de normas de calidad del aire para ciertos contaminantes representa la aceptación de un grado de riesgo para la población. En muchos casos, las normas de calidad del aire son conservadoras, con el fin de proteger la salud de las poblaciones más susceptibles, como los niños, los ancianos y los enfermos (WHO, 2000).

En México, la Secretaría de Salud es la responsable de evaluar la evidencia de los impactos de la contaminación atmosférica en la salud y de establecer los límites permisibles de concentración de los contaminantes. Las normas oficiales mexicanas (NOM) de calidad del aire para partículas (PST, PM10 y PM2.5), el O₃, el CO, el SO₂, el NO₂ y el Pb se muestran en el Cuadro 14. En este Cuadro se presentan las concentraciones, el tiempo para el promedio y la frecuencia máxima aceptable de los contaminantes normados, así como el número de la norma correspondiente y su fecha de publicación.

Contaminante	Valores límite			Normas oficiales mexicanas
	Exposición aguda		Exposición crónica	
	Concentración y tiempo promedio	Frecuencia máxima aceptable	Concentración y tiempo promedio	
Partículas suspendidas totales (PST)	210 µg/m ³ (24 horas)	2% de mediciones al año	-	Modificación a la NOM-025-SSA1-1993 (DOF, 2005)
Partículas menores de 10 micrómetros (PM ₁₀)	120 µg/m ³ (24 horas)	2% de mediciones al año	50 µg/m ³ (promedio aritmético anual)	
Partículas menores de 2.5 micrómetros (PM _{2.5})	65 µg/m ³ (24 horas)	2% de mediciones al año	15 µg/m ³ (promedio aritmético anual)	
Ozono (O ₃)	110 ppb (1 hora) (216 µg/m ³)	No se permite	-	Modificación a la NOM-020-SSA1-1993 (DOF, 2002)
	80 ppb (8 horas)	4 veces al año	-	
Monóxido de carbono (CO)	11 ppm (8 horas) (12 595 µg/m ³)	1 vez al año	-	NOM-021-SSA1-1993 (DOF, 1994 ^a)
Bióxido de azufre (SO ₂)	130 ppb (24 horas) (341 µg/m ³)	1 vez al año	30 ppb (promedio aritmético anual)	NOM-022-SSA1-1993* (DOF, 1994b)
Bióxido de nitrógeno (NO ₂)	210 ppb (1 hora) (395 µg/m ³)	1 vez al año	-	NOM-023-SSA1-1993(DOF, 1994c)

* NOM en proceso de modificación.

Cuadro 14. Normas Oficiales Mexicanas de Calidad del Aire.

4.3.1.1. Consideraciones para la interpretación de las Normas Oficiales Mexicanas de calidad del aire.

Los límites permisibles especificados en las normas consideran para su cálculo y para la evaluación del cumplimiento la concentración, el tiempo para el promedio y el número de veces que se permite que se exceda dicho valor en un determinado periodo de tiempo (frecuencia máxima aceptable 1 vez al año, 4 veces al año, entre otros). Sin embargo, existen diferencias en la interpretación de estos términos, además de que las especificaciones para el manejo de datos sólo se incluyen en las normas que han sido modificadas después de 1994. Por lo anterior a continuación se proponen algunas guías para la interpretación de las NOM y sus especificaciones.

Lapso de tiempo considerado para el reporte de los promedios horarios en las bases de datos.

En las NOM no se especifica de manera exacta cómo reportar el promedio horario. Por esta razón, en algunas redes de medición los datos horarios van de 1:00 a 24:00 horas y en otras de 0:00 a las 23:00 horas. Esto incide en la interpretación de la información y por lo tanto en que la comparación de información entre redes pueda no ser correcta. Por ejemplo, si se quisiera comparar información de estaciones de dos redes de medición que manejan diferentes formatos puede suceder que la información registrada

a la misma hora no corresponda a los mismos intervalos de tiempo. Es decir, en un caso puede tratarse del tiempo transcurrido en la hora anterior y en el otro, el tiempo transcurrido en la hora que le sigue. El estándar para datos meteorológicos definido por la Organización Meteorológica Mundial es el tiempo transcurrido en la hora anterior y está basado en la necesidad de predecir sobre una situación que ya sucedió de manera real (MENZ, 2009). Con base en este estándar se sugiere que a partir de la publicación de este documento, la notación para reportar los promedios horarios en las bases de datos sea: de 1:00 a 24:00 horas, donde el 1:00 represente el promedio horario que va de las 0:01 horas hasta la 1:00; el 2:00, de la 1:01 a las 2:00; y así sucesivamente hasta las 24 horas (Figura 4).

Intervalo de tiempo para el SO₂

En la norma de SO₂ no se especifica a qué hora empieza y termina el intervalo de tiempo de veinticuatro horas. Se sugiere que éstas correspondan al promedio de las concentraciones horarias de las 24 hrs. del día, como se indicó en la Figura 7.

Criterios de compleción de datos y redondeo de cifras decimales significativas.

Sólo en las modificaciones de las normas de O₃ y de partículas se especifica un criterio de compleción de datos de al menos el 75% para el cálculo de valores de acuerdo

al tiempo para realizar el promedio de las concentraciones y también el procedimiento para realizar el redondeo de las cifras decimales. Para homologar el criterio de completación de datos se sugiere que en todas las NOM se requiera de al menos el 75% de datos necesarios para el cálculo. Asimismo, se recomienda que se aplique la siguiente regla para el redondeo a un determinado número de cifras significativas: si la cifra siguiente a la definida como significativa para cada contaminante es menor o igual a cinco, el valor de la cifra significativa no se incrementa y si es mayor se incrementa al número inmediato superior.

Frecuencia máxima aceptable en un año.

Ozono. En el numeral 4.1 de la NOM se especifica que “la concentración de ozono, como contaminante atmosférico debe ser menor o igual a 110 ppb, promedio horario, para no ser rebasado una vez al año, calculado como se especifica en el numeral 4.3.1 de esta Norma”. Se recomienda interpretar este texto como que el límite no se debe rebasar. Esto es, si el límite se rebasa una vez, se considerará que no se cumple con la norma horaria.

CO, SO₂ y NO₂. Para estos contaminantes, en su respectiva NOM se indica que la frecuencia máxima aceptable es “una vez en un año”. Esto se interpretará como que el límite

permisible se puede rebasar una vez para el tiempo promedio especificado en un año. Si para el tiempo especificado, el promedio se rebasa más de una vez, la respectiva norma no se cumple. Por ejemplo, para el SO₂, si la concentración del segundo máximo de los promedios de 24 horas en un año excede el límite de 130 ppb se considera que no se cumple con la norma diaria para este contaminante debido a que en la norma se especifica que sólo se puede exceder ese valor una vez al año.

Partículas PM₁₀, PM_{2.5} y PST.

La frecuencia máxima aceptable con que se permite que se rebasen los límites especificados en la modificación de la NOM-025-SSA1-1993 es del 2% de las concentraciones diarias en un año para el límite de 24 horas. Es decir, si el percentil 98 de las concentraciones diarias rebasa el límite, no se cumple con la norma.

4.3.1.2. Definición y cálculo de los indicadores

La evaluación de la calidad del aire tiene como referencia obligada los límites máximos permisibles establecidos en las NOM de calidad del aire. Así se puede determinar para cada estación de medición si se cumple con las NOM y de manera complementaria

calcular el número de días del año en los que los niveles de un contaminante se encuentran por arriba de los límites establecidos en la NOM correspondiente. A continuación se describe de manera breve en qué consiste cada uno de los indicadores relacionados con impactos en la salud especificados en el Cuadro 11.

Percentil 98.

Valor por debajo del cual se acumula el 98% de los registros dejando arriba de este valor el 2% del total. Este sirve para evaluar el valor límite permisible para la concentración de partículas PST, PM₁₀ y PM_{2.5}. La modificación a la NOM-025-SSA1-1993 indica que para cada trimestre del año se requerirá de 75% o más de las concentraciones diarias.

Si en un trimestre no se cumple con esta condición la información de ese trimestre no se considerará para el cálculo del indicador.

Para la validación del año es necesario contar con al menos tres trimestres que cumplan con el número de muestras especificado. En caso contrario no podrá evaluarse el cumplimiento de la norma para ese año.

El cálculo del percentil 98 se realiza de acuerdo con el siguiente algoritmo: se multiplica el número de concentraciones diarias (n) por

0.98; del resultado obtenido se toma la parte entera, i (por ejemplo: si n=298 entonces $298 \times 0.98 = 292.04$ y así la parte entera es $292 = i$). A la parte entera se le suma 1 (i +1).

Por último se ordenan las concentraciones diarias de menor a mayor. El Percentil 98 es el dato que está en el lugar i +1 de los datos ordenados (DOF, 2005).

Quinto máximo.

Se obtiene como el quinto valor más grande registrado en un conjunto de datos. El cumplimiento de la norma de ocho horas para O₃ se obtiene a partir del cálculo del quinto máximo anual de las concentraciones diarias de los promedios móviles de ocho horas, el cual debe ser igual o menor a 0.08 ppm. La modificación a la NOM-020-SSA1-1993 indica que para cada año se requerirá de al menos 75% de concentraciones diarias provenientes de los promedios móviles de ocho horas. En el caso del cálculo de la concentración diaria se debe considerar que si no se cumple con la compleción de datos de al menos el 75% de los promedios móviles de ocho horas durante el día, pero el valor máximo rebasa 80 ppb, éste debe registrarse como la concentración máxima del día (DOF, 2002).

Máximo.

Se obtiene como el máximo valor registrado en un conjunto de datos. Ilustra el comportamiento de eventos extraordinarios de contaminación. Este indicador determina el cumplimiento o incumpliendo del límite horario especificado en la NOM de protección a la salud de O₃ (DOF, 2002) y se calcula a partir de las concentraciones horarias.

Segundo máximo.

Se obtiene como el segundo valor máximo registrado en un conjunto de datos. Indica que determina el cumplimiento o no cumplimiento de los límites especificados en las NOM de protección a la salud en relación con el tiempo especificado para el promedio de la concentración de una hora para NO₂, de 8 horas para CO y de 24 horas para SO₂ y la frecuencia de tolerancia de una vez al año (DOF, 1994a; DOF, 1994b; DOF, 1994c).

Promedio o Media.

Valor medio que resulta de dividir la suma de todos los valores entre el número de éstos. Es un indicador que determina el cumplimiento de los límites anuales de partículas PM₁₀, PM_{2.5} y de SO₂ (DOF, 2005; DOF, 1993). Para las partículas PM₁₀ y PM_{2.5} la norma anual se evalúa con el promedio aritmético de las concentraciones diarias y para el SO₂ con el promedio aritmético de las concentraciones horarias.

Número de horas arriba del límite.

Se obtiene como el conteo de horas por arriba del valor límite establecido en las NOM de calidad del aire. Indica el impacto que tiene la contaminación por un contaminante específico en una región, es un indicador complementario a los de la evaluación del cumplimiento de límites de las NOM de calidad del aire. Es aplicable únicamente para los contaminantes que cuentan con norma horaria (O₃ y NO₂). La información que genera este indicador es equivalente a la que proporciona el indicador “porcentaje de horas arriba del límite” y se obtiene dividiendo el número de horas por arriba del valor límite establecido en las NOM de protección a la salud entre el total de horas con concentraciones válidas del período que se va a analizar.

Número de días arriba del límite.

Se obtiene como el conteo de días por arriba del valor límite establecido en las NOM de calidad del aire. Este valor indica el impacto que tiene la contaminación por un contaminante específico en una región dada, es un indicador complementario a los de la evaluación del cumplimiento de límites de las NOM de calidad del aire. La información que genera este indicador es equivalente a la que proporciona el indicador “porcentaje de días arriba del límite” y se obtiene como la división del número de días por arriba del valor límite establecido en las NOM de protección a la salud entre el total de días con concen-

traciones válidas del período a analizar. El cálculo del número de días o el porcentaje de días en que se rebasa el valor de los límites establecidos en la norma se debe hacer con las concentraciones diarias y con un criterio de completación para el cálculo de la concentración diaria de al menos un 75% de los datos. Sin embargo, dada la escasa cantidad de datos de los que se dispone en algunas redes de medición, se sugiere en un principio contar con al menos el 50% de los datos necesarios para el cálculo de la concentración diaria e ir aumentando cada vez más este porcentaje hasta alcanzar el 75%.

En el Anexo III se presentan dos ejemplos del cálculo para obtener los indicadores percentil 98 y promedio anual para las partículas PM₁₀, de acuerdo al manejo de datos especificados en la modificación de la NOM-025-SSA1-1993 de calidad del aire.

4.3.2. Indicadores del comportamiento de los contaminantes a lo largo del tiempo.

La utilidad de los indicadores del comportamiento del contaminante es que permiten analizar sus variaciones durante el día (hora a hora), durante la semana (día a día), durante el año (mes a mes) y su distribución en un periodo de tiempo. A continuación se describe en qué consiste cada indicador relacionado con el comportamiento del contaminante especificados en el Cuadro 11.

Promedio o mediana hora a hora.

Muestra el cambio de las concentraciones de un contaminante a lo largo de un día promedio mostrando los valores típicos de cada hora. Se obtiene como el promedio (la mediana) de los registros horarios de un periodo de tiempo para cada hora del día. Permite asociar la concentración de los contaminantes con la intensidad de las actividades antropogénicas o patrones meteorológicos.

Promedio o mediana día a día.

Muestra el cambio de las concentraciones de un contaminante a lo largo de una semana promedio mostrando los valores típicos de cada día. Se obtiene como el promedio (la mediana) de los registros horarios o diarios de un periodo de tiempo de cada uno de los días de la semana. Permite asociar la concentración de los contaminantes con la intensidad de las actividades antropogénicas o patrones meteorológicos.

Promedio o mediana mes a mes.

Muestra el cambio de las concentraciones de un contaminante a lo largo de un año promedio mostrando los valores típicos de cada uno de los meses del año. Se obtiene como el promedio (la mediana) de los registros horarios o diarios de cada mes de un periodo de tiempo. Permite asociar la concentración de los contaminantes con la intensidad de las actividades antropogénicas o patrones meteorológicos.

Percentil.

Señala el valor del parámetro que acumula cierto porcentaje del total de datos. Después de ordenar un conjunto de datos (el cual puede ser de un mes, un año o un periodo de años) por su magnitud (de menor a mayor) se obtiene el porcentaje de datos menor o igual a cada valor. Los percentiles permiten caracterizar el comportamiento de un contaminante en un período de tiempo y posiblemente el impacto de los programas de mejoramiento ambiental, así como la influencia de fenómenos temporales o regionales. En particular, los percentiles 98, 90, 75, 50, 25 y 10 se emplean como indicadores de calidad del aire. El rango intercuartil se obtiene como la diferencia entre el percentil 75 y el percentil 25, caracteriza la dispersión de los valores usuales. El uso de este percentil evita la influencia del 25% de los valores altos y del 25% de los valores bajos de un conjunto de datos, representa de manera adecuada el comportamiento típico de un conjunto de datos de calidad del aire.

El máximo y el promedio que se describieron en la sección anterior también se usan para determinar el comportamiento de los contaminantes.

4.4. Representación gráfica de los indicadores de calidad del aire.

Una de las formas de presentar los indicadores de calidad del aire es a través de gráficas.

Con éstas se comunican de manera sencilla los resultados obtenidos. A continuación se muestran algunos tipos de gráficas en las que se ilustran algunos indicadores que se describieron en la sección anterior:

Gráficas en el tiempo.

Con este tipo de gráficas es posible visualizar el comportamiento (o patrón) del fenómeno de interés a lo largo del tiempo.

Gráficas de dos escalas.

La característica principal de esta gráfica es que se pueden visualizar dos indicadores con diferentes unidades de medición.

Gráficas de caja.

Con este tipo de gráficas es posible determinar visualmente un resumen muy completo de la distribución de los datos debido a que proporciona las estadísticas de orden de la variable de interés además de varios indicadores como el percentil 25, 75, el intervalo intercuartil, el promedio, la mediana y el máximo (Figura 8).

En los tres tipos de gráficas descritas es posible ilustrar indicadores de impacto a la salud, de comportamiento y de tendencia.

Mapas.

La distribución espacial de los contaminantes en las ciudades se puede ilustrar mediante mapas de contorno en los que la impresión visual de las concentraciones se conjunta con los límites geográficos de la zona a la que pertenecen. Con este tipo de representación es posible ilustrar principalmente indicadores de comportamiento.

Se recomienda que se incluya en la representación gráfica de los indicadores las unidades del contaminante, el criterio o criterios de compleción con que se generaron, el periodo de tiempo de la información que se está representando, la fuente de la información y un título auto explicativo de los indicadores.

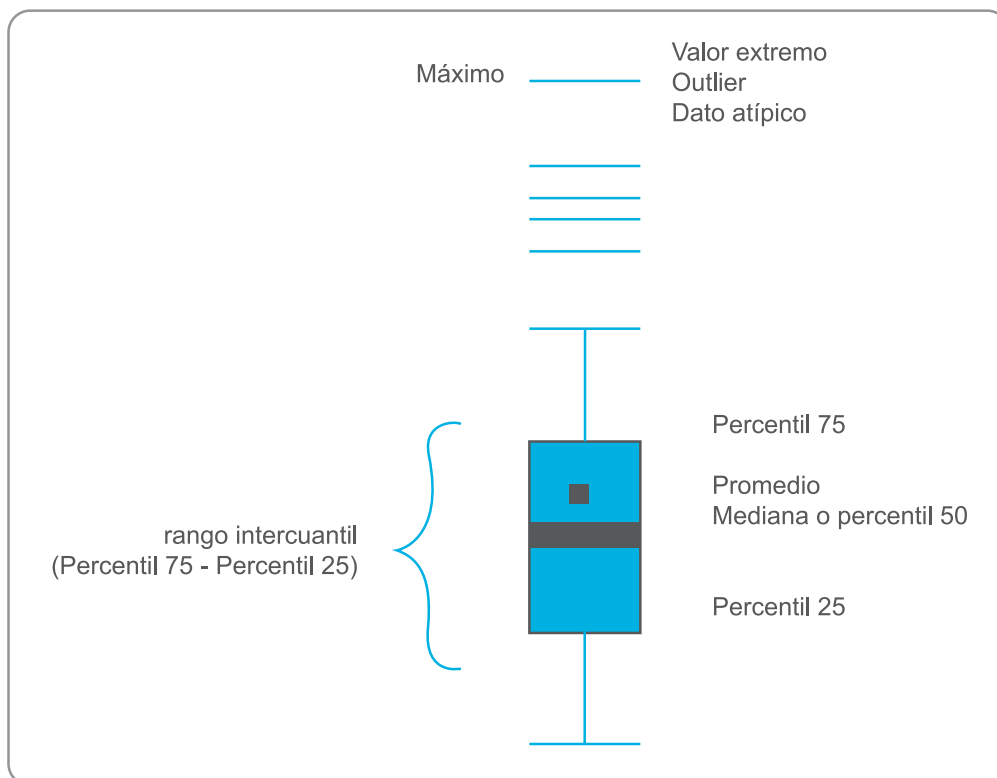


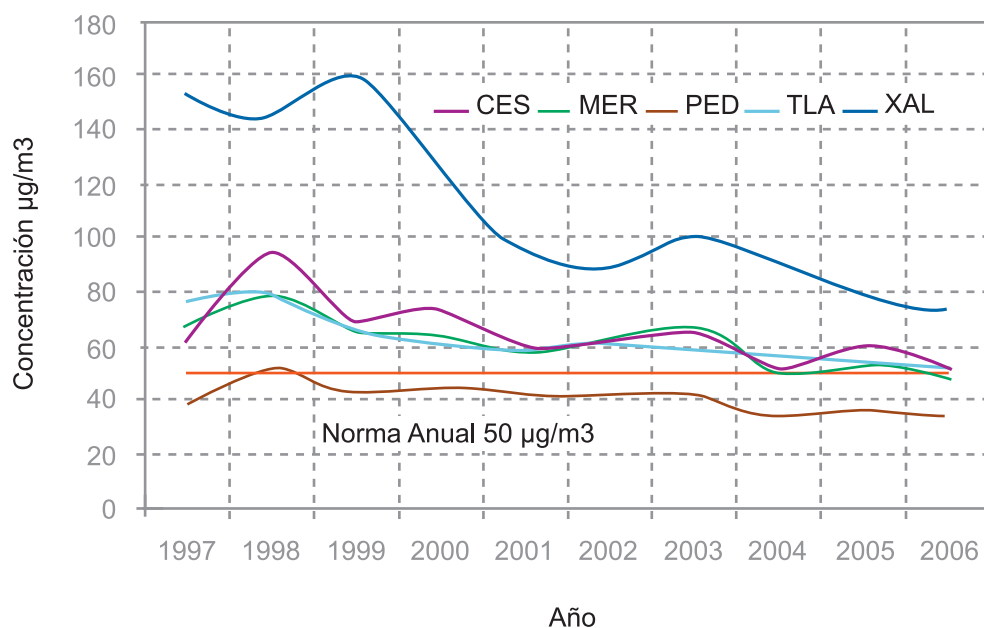
Figura 8. Gráfica de caja.

4.4.1. Ejemplos de las representaciones gráficas de los indicadores.

4.4.1.1. Indicadores de impacto en la salud.

En la Figura 9 se presenta como ejemplo una gráfica en el tiempo que ilustra el promedio anual de los muestreos manuales de PM_{10} en la ZMVM para el periodo de 1997 a 2006.

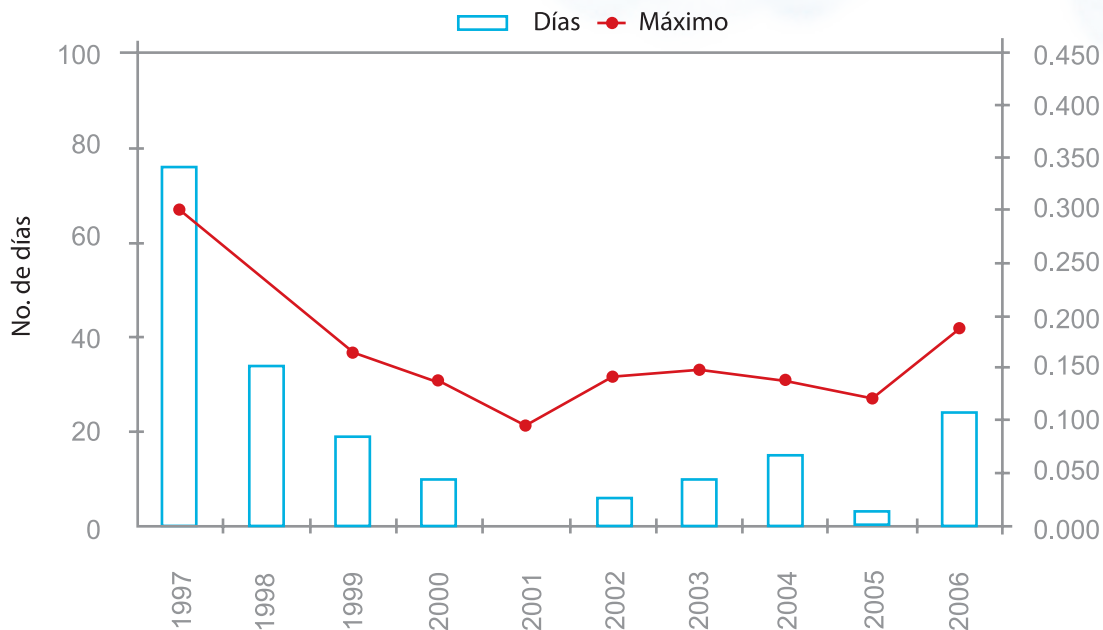
En la Figura 10 se presenta una gráfica con dos escalas relativa a los días en que se excede la norma horaria (0.11 ppm) y la concentración máxima de O_3 para 1997-2006 en la ZMG en relación a la NOM-020-SSA1-1993.



* Xalostoc (XAL), Tlalnepantla (TLA), Merced (MER), Cerro de la Estrella (CES) y Pedregal (PED) (equipo manual).
Elaboración propia de los autores.

Compleción de datos: al menos 3 trimestres con por lo menos el 75% de los muestreos manuales válidos al año.
Fuente de los datos: Red Manual del Sistema de Monitoreo Atmosférico de la Ciudad México.

Figura 9. Promedio anual de PM_{10} en 5 estaciones* de la ZMVM (1997-2006).



Elaboración propia de los autores.

Compleción de datos para las concentraciones diarias: al menos el 50% de los datos horarios.

Fuente de los datos: Red Automática de Monitoreo Atmosférico de la ZMG.

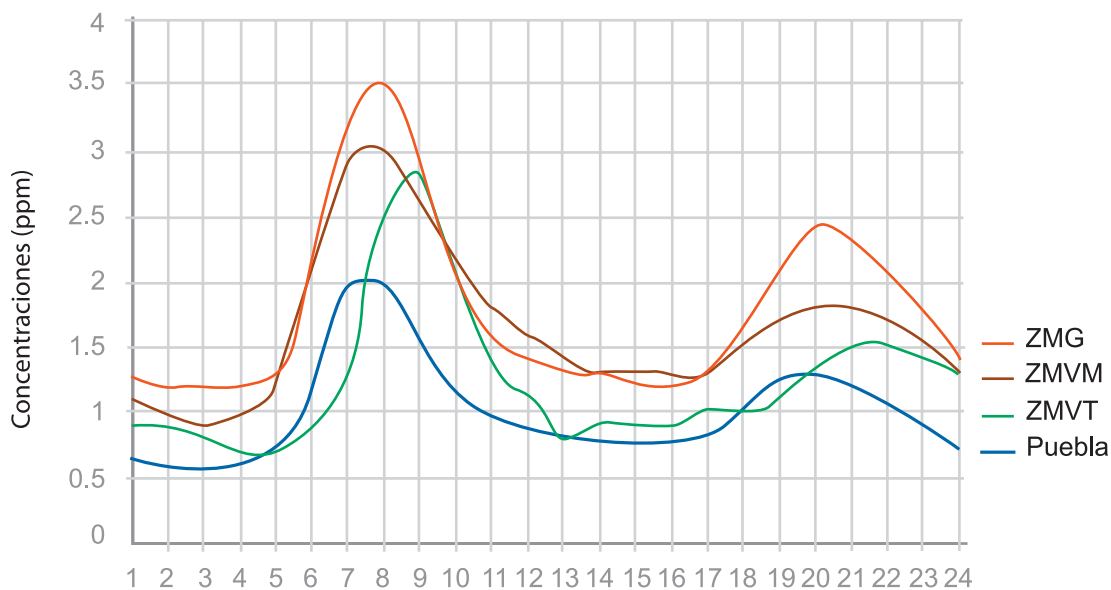
Figura 10. O₃ en la estación Tlaquepaque de la ZMG: días en que se excede la norma horaria (0.11 ppm) y concentración máxima (1997-2006)

4.4.1.2. Indicadores relacionados con el comportamiento del contaminante.

A continuación se ilustra con una gráfica en el tiempo (Figura 11) el comportamiento horario (mediana hora a hora) de las concentraciones de CO en tres zonas metropolitanas y una ciudad (INE-DGICUR, 2007).

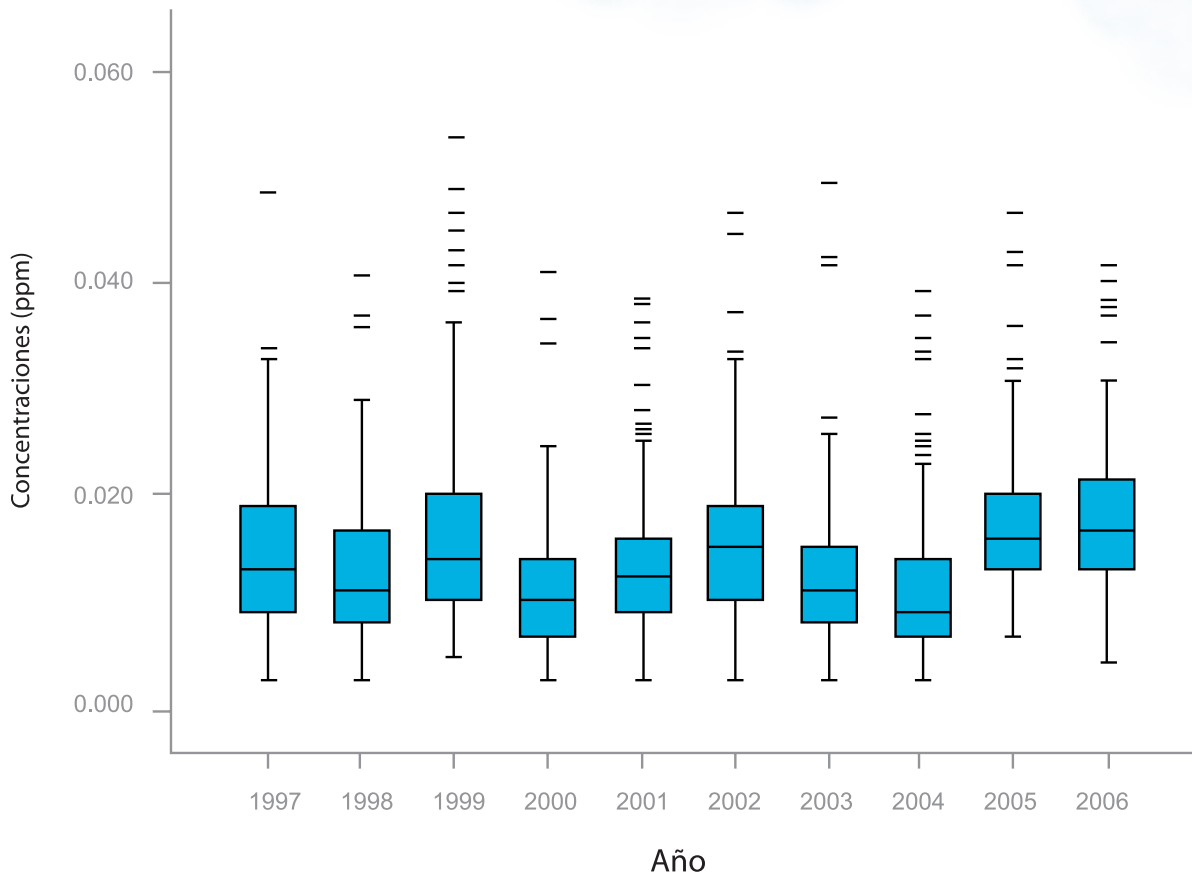
En la Figura 12, empleando una gráfica de cajas se muestra el comportamiento y tendencia de los promedios de 24 horas de SO₂ en la ZMVT en la estación de medición San Mateo de 1997 a 2006.

Por último, en la Figura 13 se presenta la distribución espacial del NO₂ en la ZMG para el año 2005.



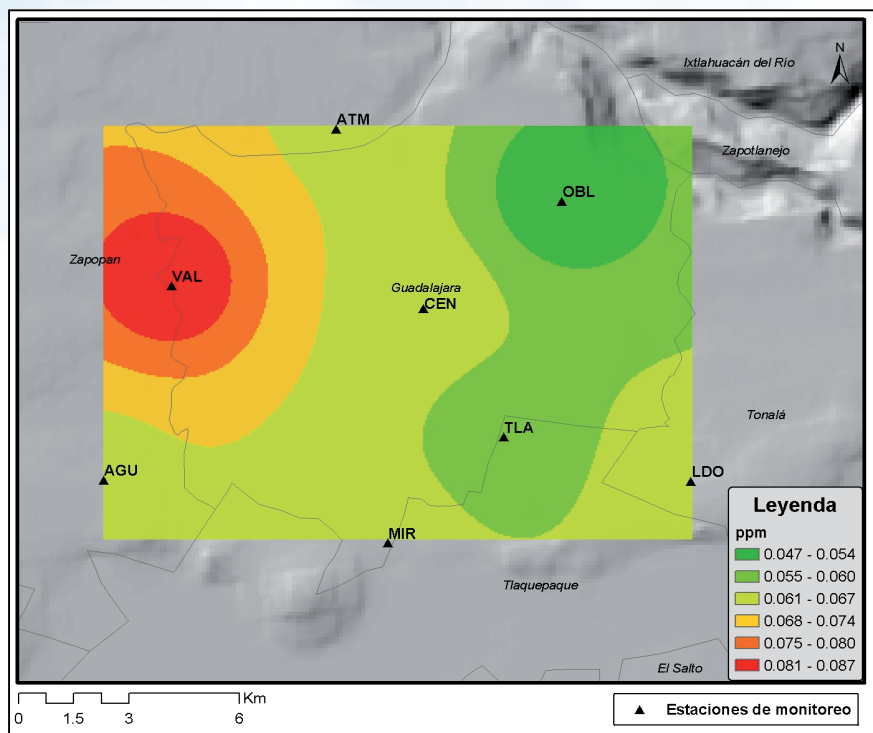
Fuente: Tercer Almanaque de datos y tendencias de la calidad del aire en nueve ciudades mexicanas (Zuk et al., 2007).

Figura 11. Comportamiento horario del CO en 4 ciudades mexicanas (1997-2005).



Elaboración propia de los autores con datos: Red Automática de Monitoreo Atmosférico de la ZMVT
 Compleción de datos para las concentraciones diarias: al menos el 50% de los datos horarios.

Figura 12. Tendencia de las concentraciones diarias de S02 en la estación de medición San Mateo de la ZMVT (1997-2006).



Fuente: Tercer Almanaque de datos y tendencias de la calidad del aire en nueve ciudades mexicanas (Zuk et al., 2007).

Figura 13. Distribución espacial del NO₂ en la ZMG (2005).

5. LOS OBJETIVOS DE LA CALIDAD DE LOS DATOS (OCD) Y LA EVALUACIÓN DE LA CALIDAD DE LOS DATOS (ECD) EN EL CICLO DE VIDA DE LOS DATOS DE UN PROYECTO.

Los objetivos de calidad de los datos, OCD, se mencionaron ya en el Manual 2, donde se presentaron los aspectos que deben ser observados en su establecimiento. En este Capítulo se describen con mayor profundidad sus conceptos. Asimismo, se presenta cómo se realiza su evaluación dentro de su ciclo de vida. Los OCD, dentro del sistema de calidad de la US-EPA y dentro del contexto de medición de la calidad del aire, se refieren a aseveraciones cuantitativas y cualitativas derivadas del proceso de los OCD. Este proceso consiste en una serie de pasos lógicos que deben de llevar a cabo los responsables de las redes de medición de la calidad del aire, y se inicia desde que se planea y se aplica a la toma de decisiones. Los pasos incluyen la definición de los siguientes aspectos: el problema que se va a resolver (por ejemplo, detectar el cumplimiento de las normas de calidad del aire, identificar tendencias a largo plazo de los contaminantes atmosféricos, proveer de bases de datos para investigación, entre otros), la decisión a tomar, la información requerida para tomar la decisión, los límites del estudio, la regla de decisión y los límites de incertidumbre aceptables. En suma, se requiere optimizar el diseño del estudio (US-EPA, 2006).

Los diferentes objetivos de medición pueden requerir de diferentes OCD lo que puede vol-

ver complejo su desarrollo. Para solventar esta situación se propone establecer los OCD para el objetivo que necesita una calidad de datos más rigurosa de manera que la calidad de los datos para los demás objetivos también se cumpla (US-EPA, 1998). En la US-EPA los OCD se han establecido, en primera instancia, para asegurar que las decisiones que se tomen en relación al cumplimiento de los estándares (National Ambient Air Quality Standards, NAAQS) se encuentren dentro de un grado de certidumbre específico.

Una vez que se han especificado los OCD se debe valorar y controlar la calidad de los datos para asegurar que se mantienen dentro de los criterios de aceptación establecidos. La evaluación de los OCD se diseña para controlar varias fases del proceso de medición (muestreo, preparación y análisis) y para asegurar que la incertidumbre total se encuentre dentro del intervalo prescrito en los OCD. La evaluación de los OCD se define con los siguientes criterios de calidad: precisión, sesgo, representatividad, posibilidad de detección, completación y grado de comparabilidad.

La evaluación de la calidad de los datos (ECD) se define como una valoración científica y estadística de un conjunto de datos para determinar si éstos son adecuados para

usarse de acuerdo a un propósito específico en cuanto a su calidad, cantidad y tipo. La evaluación de la calidad de los datos en la US-EPA se basa en una premisa fundamental: la calidad de los datos, como concepto, es significativa solamente cuando está relacionada al uso para el cual fueron generados los datos. Por ello, es fundamental conocer en qué contexto se van a utilizar para establecer un criterio relevante a fin de determinar la conveniencia de utilizarlos.

El proceso de la ECD se desarrolla para los casos donde se han establecido OCD formales. Sin embargo, en relación con los OCD hay una técnica propuesta por la EPA para la aceptación de la calidad de los datos si es que una red no tiene OCD formales (esto es, que fue establecida antes del desarrollo del proceso de los OCD). (US-EPAc, 2000).

Con lo anterior, se pretende mostrar la importancia de que los OCD y la ECD sean desarrollados en las redes de medición de la calidad del aire. El alcance de este documento es dar a conocer que existen estos procedimientos y que aunados a una adecuada implantación para la colección de datos, las actividades del AC y CC y el manejo adecuado de datos dentro de su ciclo de vida, respaldan la certeza de que las decisiones que se tomen con base en dicha información están bien fundamentadas y soportadas. A continuación se tratarán estos temas con mayor detalle.

5.1. Objetivo de la calidad de los datos

Los objetivos de calidad de los datos (OCD) son criterios que clarifican los objetivos del estudio, definen los tipos apropiados de adquisición de datos y especifican los niveles tolerables de errores de decisión potencial. Inicialmente, los OCD se utilizaron para describir los indicadores de calidad de los datos, tales como, los parámetros de precisión, exactitud, representatividad, comparabilidad y compleción. Actualmente, estos procedimientos son una herramienta de planeación para el manejo de los errores de decisión.

El propósito general del procedimiento de OCD es asegurar que la generación de los datos de calidad del aire se efectúe tan eficazmente como sea posible. El OCD es un procedimiento de planeación sistemática para generar datos suficientes que garanticen el uso para el que fueron diseñados y a su vez, mejore la efectividad de planeación, la eficacia del diseño y el resguardo de los resultados.

El procedimiento de OCD genera datos apropiados por su tipo, calidad y cantidad; facilita el método óptimo de toma de decisión entre alternativas racionales, provee especificaciones de los niveles tolerables de errores de decisión potencial y promueve la documentación adecuada de resultados de planeación sistemática.

La finalidad del procedimiento de OCD es descifrar detalladamente las formas en un proceso lógico para obtener el tipo correcto de datos. Además, facilita la administración de las acciones y el soporte de los procesos. Entre los pasos del proceso de establecimiento de los OCD se incluyen:

- Establecimiento del Problema
- Identificación de Decisiones
- Identificación de insumos para las decisiones
- Definición de los límites del estudio
- Desarrollo de reglas de decisión
- Especificación de límites de tolerancia para los errores de decisión
- Optimización del diseño para la obtención de datos.

5.2. Evaluación de la calidad de los datos

La evaluación de la calidad de los datos (ECD), de acuerdo a la US-EPA, se define como una valoración científica y estadística de un conjunto de datos para determinar si éstos son adecuados para ser utilizados, de acuerdo a un propósito específico, en cuanto a su calidad, cantidad y tipo. En México no se realiza este tipo de evaluación para los datos de calidad del aire, el cual es una parte fundamental para garantizar los resultados.

En este capítulo se describe como aplicar algunas herramientas gráficas y estadísticas para la evaluación de un conjunto de datos.

La evaluación de la calidad de los datos es construida en la EPA en una premisa fundamental: la calidad de los datos, como un concepto, es significativo solamente cuando está relacionado al uso para el cual fueron generados, por lo que es importante conocer en qué contexto un conjunto de datos va a ser usado para establecer un criterio relevante y así juzgar si estos son adecuados o no. Con la evaluación de la calidad de los datos la EPA plantea que se pueden contestar dos preguntas fundamentales:

1. Dada la calidad del conjunto de datos, ¿puede ser tomada una decisión con respecto de ellos, con la confianza deseada?
2. ¿Qué tan bien se puede esperar que el diseño de muestreo se haya efectuado sobre un extenso rango de posibles resultados? esto es, si se usa la misma estrategia en el diseño de muestreo para un estudio similar, ¿se podría esperar que los datos soporten el mismo uso para el cual fueron generados con el nivel de confianza deseado, en particular, si las mediciones resultaron más altas o más bajas que en el estudio original?

La primera pregunta está orientada a las necesidades inmediatas del usuario de los datos y permite al tomador de decisiones, con base en la información que están proporcionando los datos, tomar una elección informada. Por ejemplo: si los datos proveen fuerte evidencia a favor de una acción sobre otra, entonces, el tomador de decisiones puede

tomar la decisión sabiendo que estará soportada por datos correctos; por el contrario: si los datos no muestran una fuerte evidencia a favor de una alternativa, entonces los datos alertaran al tomador de decisiones de esta incertidumbre. (Tal como coleccionar más o diferentes datos antes de tomar una decisión o proceder con la decisión a pesar de una relativamente alta pero aceptable, probabilidad de una conclusión errónea).

La segunda pregunta está orientada a las necesidades potenciales futuras del usuario de datos. Por ejemplo: si el investigador decide usar cierto diseño de muestreo en una localización diferente de donde el diseño fue primeramente utilizado, deberá determinar qué tan bien puede esperarse que el diseño se lleve a cabo, dado que los resultados y las condiciones del medio ambiente del muestreo serán diferentes de aquellas del original.

Lo anterior, puesto que las condiciones del medio ambiente varían de un lugar a otro y de una fecha a otra, la conveniencia de un método de muestreo debe ser evaluada sobre un extenso rango de posibles resultados y condiciones.

5.3. La ECD y su ciclo de vida

El ciclo de vida de los datos comprende tres etapas, planeación, implementación y evaluación. Durante la fase de planeación, los

procesos de los objetivos de la calidad de los datos (o algún otro procedimiento de planeación sistemático) son usados para definir criterios cuantitativos y cualitativos para determinar cómo, cuándo, dónde y cuántas muestras (mediciones) se van a coleccionar para alcanzar el nivel de confianza deseado.

Esta información junto con los métodos de muestreo, procedimientos analíticos y procedimientos apropiados de aseguramiento y control de calidad son documentados en el plan de aseguramiento de calidad. Entonces los datos son coleccionados de acuerdo a las especificaciones del plan mencionado. La evaluación de la calidad de los datos completa el ciclo de vida del proyecto al proveer la evaluación necesaria para determinar si se cumplió con los objetivos. Durante la fase de evaluación, los datos son validados y verificados para asegurar que los protocolos de muestreo y análisis especificados en el proyecto de aseguramiento de calidad se hayan seguido y que los sistemas de medición se hayan llevado a cabo de acuerdo con los criterios especificados en el plan de aseguramiento de calidad de los datos. La evaluación de la calidad de los datos procede entonces a usar el conjunto de datos validados para determinar si la calidad de los datos es satisfactoria.

5.4. Los cinco pasos en la ECD

La evaluación de la calidad de los datos incluye cinco pasos que empiezan con la revisión de la documentación de la planeación y finalizan con una respuesta a la pregunta planteada durante la fase de planeación del estudio.

1. Revisión de los objetivos de la Calidad de los datos (OCD) y su diseño de muestreo. Revisar los OCD que se tomaron para asegurar que son aplicables todavía. Si los OCD no han sido desarrollados, especificar los OCD antes de la evaluación de los datos (definir las hipótesis estadísticas y especificar límites tolerables de confianza en la decisión de los errores; para la estimación de problemas, define un intervalo de confianza extenso); revisar el diseño de muestreo y la documentación de la forma como se coleccionaron los datos para la consistencia con los OCD.

2. Revisión preliminar de los datos. Revisar los reportes del aseguramiento de los datos, calcular estadísticas básicas y generar gráficas de los datos. Utilizar esta información para aprender acerca de la estructura de los datos e identificar patrones, relaciones o anomalías potenciales.

3. Seleccionar la estadística de prueba. Seleccionar el procedimiento más apropiado para resumir y analizar los datos, basado en la revisión de los objetivos de la calidad de los datos, el diseño de muestreo y la revisión

de los datos preliminares. Identificar las suposiciones claves bajo las cuales los procedimientos estadísticos serán válidos.

4. Verificar las suposiciones de la Prueba Estadística. Evaluar si se mantienen bajo las suposiciones clave o si las desviaciones son aceptables, dada la información actual y otras informaciones acerca del estudio.

5. Tomar conclusiones de los datos. Llevar a cabo los cálculos requeridos para la prueba estadística y documentar las inferencias tomadas como resultado de estos cálculos. Si el diseño va a ser usado de nuevo, se debe de evaluar.

Estos cinco pasos son presentados en una secuencia lineal, pero la evaluación de la calidad de los datos es de una naturaleza muy iterativa. Por ejemplo: si la revisión preliminar de los datos revela patrones o anomalías en el conjunto de datos que son inconsistentes con los Objetivos de la Calidad de los Datos, entonces algunos aspectos de la planeación del estudio pueden ser reconsiderados. Asimismo, si las suposiciones de las pruebas estadísticas no son soportadas por los datos, entonces los pasos previos de la Evaluación de la Calidad de los Datos debe ser revisada nuevamente. La fortaleza de la Evaluación de la Calidad de los Datos radica en que está diseñada para suscitar un entendimiento de qué tan bien los datos satisfacen el objetivo para el cual fueron generados de una manera lógica y eficiente.

5. REFERENCIAS

- CARB, 1989. Draft Criteria for evaluating air quality-related indicators. California Air Resources Board. Estados Unidos, 33 pp.
- DOF 1994a. Norma Oficial Mexicana NOM-021-SSA1-1993. Diario Oficial de la Federación del 23 de diciembre de 1994.
- DOF 1994b. Norma Oficial Mexicana NOM-022-SSA1-1993. Diario Oficial de la Federación del 23 de diciembre de 1994.
- DOF 1994c. Norma Oficial Mexicana NOM-023-SSA1-1993. Diario Oficial de la Federación del 23 de diciembre de 1994.
- DOF 1994c. Norma Oficial Mexicana NOM-023-SSA1-1993. Diario Oficial de la Federación del 23 de diciembre de 1994.
- DOF 2002. Modificación a la Norma Oficial Mexicana NOM-020-SSA1-1993. Diario Oficial de la Federación del 30 de octubre de 2002.
- DOF 2005. Modificación a la Norma Oficial Mexicana NOM-025-SSA1-1993. Diario Oficial de la Federación 26 de septiembre de 2005.
- DOF. 2006. Ley General del Equilibrio Ecológico y la Protección al Ambiente. Título cuarto, capítulo II, artículo 112, fracción VI.
- EMEP, 2003. Manual for sampling and analyses. Cooperative programme for monitoring and evaluation of long-range transmission of air pollutants in Europe. Noruega, 303 pp.
- INE-DGCENICA. 2004. Programa Nacional de Monitoreo Atmosférico. INE, México, D.F. 53 pp.
- INE-SEMARNAP, 2000. Gestión de la calidad del aire en México. INE-SEMARNAP, México. 184 pp.

- KNODERER, Charley A., Et Al. "Data collected by a surface meteorological station at Autberry, California during Fall 2000/Winter 2001 for the California Regional PM10/PM2.5 air quality study. California Air Resources Board, USA, 2001.
- NAPS, 2004. National Air Pollution Surveillance Network. Quality Assurance and Quality Control Guidelines, Canada, 44 pp.
- MARTÍNEZ, Bolivar Ana P. y Romieu Isabel. Introducción al monitoreo atmosférico. México, 1996, 238 pp.
- MENZ, 2009, Ministry for the Environment of New Zealand. Good-Practice for Air Quality Monitoring and Data Management. New Zealand, 111 pp.
- NARSTO, 2006. (México, Estados Unidos, Canada), Guidance to data exchange, 2006. 6pp.
- POPE, C.A., R.T. Burnett, M.J. Thun, E.E. Calle, D. Krewski, K. Ito y G.D. Thurston. 2002. Lung cancer, cardiopulmonary mortality, and long-term exposure to fine particulate air pollution. Journal of the American Medical Association 287(9): 1132-1141.
- SMA-GDF, 2004. Adquisición y validación de datos de contaminantes atmosféricos y parámetros meteorológicos, Procedimiento: SA/ADDA/GG/01(00). México, 4 pp.
- UNEP/WHO, GEMS/AIR. Methodology Review Handbooks, United Nations Environmental Programs, Nairobi, 1994.Vols. I-IV.
- US-EPA, 1983. Test Method for the determination of suspended particulates in the atmosphere (High-Volume Method), USA, 1983, 44pp.
- US-EPA, 1989. APT-1 Course 470 Quality Assurance for Air Pollution Monitoring Systems. USA, 125 pp.

5. REFERENCIAS

- US-EPA, 1998. Quality Assurance Handbook for air pollution measurements systems. Vols. 1-5.
- US-EPA, 1999. Compendium of Methods for Inorganic Air Pollutants Method IO-2.1. USA, 78 pp.
- US-EPA, 1999. Compendium of Methods for the Determination of Inorganic Compounds in Ambient Air Compendium Method IO-2.1 Sampling of ambient air for Total Suspended Particulate Matter (SPM) and PM10 using High Volume (HV) sampler, EPA/625/R-96/010a. Estados Unidos, 78 pp.
- US-EPA, 2000a. Guidance on Technical Audits and Related Assessments for Environmental Data Operations, EPA/600/R-99/080. USA, 101 pp.
- US-EPA, 2000b. Meteorological Monitoring Guidance for Regulatory Modeling Applications. USA, 107 pp.
- US-EPA. 2000c. Guidance for Data Quality Assessment, Practical Methods for Data Analysis. EPA/600/R-96/084. USA, 219 pp.
- US-EPA. 2002. Guidance on Environmental, Data Verification and Data Validation. EPA/240/R-02/004. USA. 96 pp.
- US-EPA. 2006. Guidance on Systematic Planning Using the Data Quality Objectives Process. EPA/240/B-06/001. USA, 121 pp.
- WHO, 2000. Use of the Guidelines in Protecting Public Health. En: Air Quality Guidelines for Europe. WHO Regional Publications, Copenhagen. pp. 41-55.
- ZUK, M.y M.G. Tzintzun-Cervantes, L. Rojas-Bracho. 2007. Tercer Almanaque de datos y tendencias de la calidad del aire en nueve ciudades mexicanas. INE-SEMARNAT, México. 114 pp.

PÁGINAS WEB

- Web 1. Measurement and Assessment [en línea] [fecha de consulta: octubre de 2006]. Disponible en: <http://www.iisd.org/measure/>
- Web 2. List of flags used in the EMEP data base [en línea][fecha de consulta: junio de 2003].Disponible en: <http://www.nilu.no/projects/ccc/submission.html>
- Web 3. Code Federal of Regulations Part 40, Título 50. [en línea] [fecha de consulta: mayo de 2007]. Disponible en: <http://ecfr.gpoaccess.gov/>
- Web 4. Data Management Handbook (ORNL/CDIAC-112/R2), en <http://cdiac.esd.ornl.gov/programs/NARSTO/> prepared by NARSTO Quality Systems Science Center) Oak Ridge, USA, 1999.
- Web 5. Met Monitoring Guide [en línea][fecha de consulta: marzo de 2003].Disponible en: http://www.webmet.com/met_monitoring/toc.html
- Web 6. Universidad de Sonora, 2006. Cifras significativas y redondeo [en línea][fecha de consulta: agosto de 2006].Disponible en: http://www.tochtli.fisica.uson.mx/fluidos%20y%20calor/cifras_significativas_y_redondeo.htm

ANEXO 1

AJUSTE DE DATOS POR RESULTADOS DE LA VERIFICACIÓN CERO-SPAN

Como parte de un esquema de aseguramiento y control de calidad, de una estación o de una red automática de medición, se recomienda (ver Manual 3) disponer de un sistema de calibración dinámica *in situ* que permita hacer la verificación del cero y del *span* (semanal) en la respuesta de los equipos de medición de gases así como hacer sus calibraciones cuando así se requieran. Lo anterior se hace para verificar con frecuencia que las concentraciones de cada contaminante gaseoso que se están midiendo son reales o si presentan alguna desviación. Los valores aceptados de desviación serán como máximo de: +/- 10% para el *span* y 5 ppb para el cero. Estos valores pueden ser corregidos siempre y cuando se usen valores de cero y *span* en un periodo corto de no más de una o dos semanas. El ajuste de la serie de datos para el periodo en que se cuente con los valores de respuesta del cero y *span* se hace aplicando la siguiente ecuación de la recta:

$$\text{Dato corregido} = (\text{dato crudo} - Az) \times \frac{(Es - Ez)}{(As - Az)}$$

Az= Respuesta actual del cero

As= Respuesta actual del *span*

Es=Respuesta esperada del cero

Ez= Respuesta esperada del *span*

En esta ecuación, por lo general se espera que la respuesta del cero sea cero, de forma que la ecuación anterior se reduce a:

$$\text{Dato corregido} = (\text{dato crudo} - Az) \times \frac{(Es)}{(As - Az)}$$

Este ajuste de datos no deberá aplicarse a los provenientes de equipos de medición que muestran evidencia de un mal funcionamiento. Para este caso, los datos tendrán que ser descartados del análisis, y mencionar el malfuncionamiento en la bitácora del equipo.

ANEXO 2

LÍMITES DE DETECCIÓN DE LAS PRINCIPALES MARCAS DE EQUIPOS DE MEDICIÓN DE LA CALIDAD DEL AIRE EN MÉXICO

Contaminante	Límite de detección equipos Marca API-Teledyne	Límite de detección equipos Marca TEI
Ozono (O ₃)	< 0.6 ppb RMS	1 ppb
Bióxido de azufre (SO ₂)	0.4 ppb RMS	1.0 ppb RMS (1 min. promedio) 0.5 ppb RMS (5 min. promedio)
Bióxido de nitrógeno (NO ₂)	0.4 ppb RMS	0.4 ppb (1 min. promedio)
Óxido nítrico (NO)	0.4 ppb RMS	0.4 ppb (1 min. promedio)
Óxidos de nitrógeno (NO _x)	0.8 ppb RMS	0.8 ppb (1 min. promedio)
Monóxido de carbono (CO)	< 0.050 ppm	0.04 ppm (0.5 min. promedio)
Hidrocarburos totales	0.05 ppmC	----
Metano	0.05 ppmC	20 ppb metano
	0.05 ppmC	50 ppb NMHC como propano
Partículas PM ₁₀	± 5 µg/m ³ promedio 10 min. ± 1.5 µg/m ³ promedio 1 hora.	< 1 µg/m ³ (24 horas)
Partículas PM _{2.5}	± 5 µg/m ³ promedio 10 } min. ± 1.5 µg/m ³ promedio 1 hora.	< 5 µg/m ³ (1 hora) < 1 µg/m ³ (24 horas)

RMS (Runs Minimum Samples): Concentración calculada sobre 25 muestras corridas un mínimo de 7 veces. Definición por la USEPA.



ANEXO 3

EJEMPLO DE LA OBTENCIÓN DE LOS INDICADORES PERCENTIL 98 Y PROMEDIO ANUAL

A continuación se ilustran dos ejemplos para obtener los indicadores percentil 98 y promedio anual para las partículas PM₁₀, de acuerdo al manejo de datos especificados en la modificación de la NOM-025-SSA1-1993 de calidad del aire.

Percentil 98 y promedio anual

El siguiente conjunto de datos corresponde a las concentraciones de PM₁₀ de los muestreos realizados cada seis días durante el año 2005 en cuatro estaciones de medición en una ciudad.

Fecha	Est 1	Est 2	Est 3	Est 4	Fecha	Est 1	Est 2	Est 3	Est 4
04/01/2005	46.98	24.94		86.28	27/06/2005	19.46	92.53	38.59	59.99
10/01/2005	39.94	33.77		193.55	03/07/2005			82.27	
16/01/2005	10.90	18.20	62.28	90.85	09/07/2005		27.62	72.81	81.96
22/01/2005	7.52	5.40	32.89	46.40	15/07/2005		26.15		126.70
28/01/2005	19.26	16.88	47.06	72.93	21/07/2005	21.26	21.41	34.22	114.78
03/02/2005	5.45	29.66	24.26	50.57	27/07/2005	8.92	14.45	19.66	33.96
09/02/2005	19.22	18.13	43.37	82.65	02/08/2005	30.76	29.10	45.14	78.26
15/02/2005	70.90	70.91	95.08	233.08	08/08/2005	17.15	19.96	32.89	61.28
21/02/2005	35.73	38.42	0.74	137.52	14/08/2005	9.06	12.43	16.02	20.67
27/02/2005	8.82	11.11	8.79	18.72	20/08/2005	10.97	13.08	21.76	29.53
05/03/2005	17.08	18.84	36.54	129.32	26/08/2005	45.31	41.43	81.54	138.49
11/03/2005	82.19	87.12	95.08	162.99	01/09/2005	30.76	29.10	45.14	78.26
17/03/2005	54.32	39.02	65.96	146.66	07/09/2005	17.15	19.96	32.89	61.28
23/03/2005		60.78	120.87	235.06	13/09/2005	9.06	12.43	16.02	20.67
29/03/2005		30.10	238.15	184.23	19/09/2005			55.76	58.58
04/04/2005	89.68	88.47	426.32	275.04	25/09/2005	34.05	33.72	75.28	112.23
10/04/2005	82.76	73.61	95.08	168.93	01/10/2005	31.37	11.05	110.66	197.00
16/04/2005	24.25		113.50	136.51	07/10/2005	9.70		26.83	61.28
22/04/2005	35.73	28.31	72.77	114.61	13/10/2005	9.60	13.03	131.05	27.23
28/04/2005	22.48	27.69	128.24	128.06	19/10/2005	19.46	19.96		99.35
04/05/2005	6.83	21.57	49.38	67.89	31/10/2005	22.93	26.06		100.12
10/05/2005	55.19	54.69	151.83	95.97	06/11/2005	41.82	62.63	105.25	125.16
16/05/2005	19.30	22.24	32.34	46.40	12/11/2005	27.79	16.46	47.32	
22/05/2005	45.86	85.05	152.16	125.11	18/11/2005	27.44		116.21	135.74
28/05/2005	63.43	21.05	20.39	38.89	24/11/2005				244.39
03/06/2005	15.98	15.83	80.81	119.95	06/12/2005	52.28		104.59	167.14
09/06/2005	22.37	14.40	59.70	94.20	12/12/2005	68.31	70.43	170.77	191.23
15/06/2005	27.79	27.62	47.51	79.00	18/12/2005	56.98	90.60	167.61	102.79
21/06/2005	27.79	27.62	42.96	88.87	24/12/2005				244.39

Cuadro III.1. Concentraciones de PM₁₀ (µg/m³) reportados, 2005.

Paso 1

De acuerdo con el numeral 5.4.1.de la Modificación de la NOM-025-SSA1-1993, “en cada sitio de monitoreo, la concentración promedio de 24 horas de PST, PM₁₀ y PM_{2.5} se reportará en µg/m³, sin cifras decimales. Si se cuenta con valores de una o más cifras decimales, el valor será redondeado. Si el primer decimal es menor o igual a 5, el valor entero no se incrementa, si es mayor se incrementa al inmediato superior”.

Del Cuadro III.1 se observa que las concentraciones se reportaron con cifras decimales, por lo que el primer paso consiste en redondear las concentraciones a valores enteros de acuerdo con la regla antes descrita. Así los valores de PM₁₀ a considerar para los siguientes pasos son los que se presentan en la Cuadro III.2.

Fecha	Est 1	Est 2	Est 3	Est 4	Fecha	Est 1	Est 2	Est 3	Est 4
04/01/2005	47	25		86	27/06/2005	19	92	38	60
10/01/2005	40	34		193	03/07/2005			82	
16/01/2005	11	18	62	91	09/07/2005		28	73	82
22/01/2005	7	5	33	46	15/07/2005		26		127
28/01/2005	19	17	47	73	21/07/2005	21	21	34	115
03/02/2005	5	30	24	50	27/07/2005	9	14	20	34
09/02/2005	19	18	43	83	02/08/2005	31	29	45	78
15/02/2005	71	71	95	233	08/08/2005	17	20	33	61
21/02/2005	36	38	1	137	14/08/2005	9	12	16	21
27/02/2005	9	11	9	19	20/08/2005	11	13	22	29
05/03/2005	17	19	36	129	26/08/2005	45	41	81	138
11/03/2005	82	87	95	163	01/09/2005	31	29	45	78
17/03/2005	54	39	66	147	07/09/2005	17	20	33	61
23/03/2005		61	121	235	13/09/2005	9	12	16	21
29/03/2005		30	238	184	19/09/2005			56	58
04/04/2005	90	88	426	275	25/09/2005	34	34	75	112
10/04/2005	83	74	95	169	01/10/2005	31	11	111	197
16/04/2005	24		113	136	07/10/2005	10		27	61
22/04/2005	36	28	73	115	13/10/2005	10	13	131	27
28/04/2005	22	28	128	128	19/10/2005	19	20		99
04/05/2005	7	21	49	68	31/10/2005	23	26		100
10/05/2005	55	55	152	96	06/11/2005	42	63	105	125
16/05/2005	19	22	32	46	12/11/2005	28	16	47	
22/05/2005	46	85	152	125	18/11/2005	27		116	136
28/05/2005	63	21	20	39	24/11/2005				244
03/06/2005	16	16	81	120	06/12/2005	52		104	167
09/06/2005	22	14	60	94	12/12/2005	68	70	171	191
15/06/2005	28	28	47	79	18/12/2005	57	90	168	103
21/06/2005	28	28	43	89	24/12/2005				244

Cuadro III.2. Concentraciones de PM₁₀ (µg/m³) sin cifras decimales, 2005

Una fórmula sencilla en Excel para redondear de acuerdo a las especificaciones de la NOM de partículas es usar la siguiente función lógica:

$fx = SI(celda="","",SI(celda-ENTERO(celda)>=0.6,ENTERO(celda)+1,ENTERO(celda)))$

Paso 2

El paso 2 consiste, de acuerdo al numeral 5.4.2. de la modificación de la NOM-025-SSA1-1993, en determinar la cantidad de datos necesaria para la evaluación de la observancia de la Norma que dice: “Para poder verificar la observancia de esta Norma se requerirá de un mínimo de datos en un año.

Este mínimo se evalúa a partir de la cantidad de muestras de 24 horas válidas obtenidas en cada uno de los cuatro trimestres del año (ver Cuadro III.3). Para cada trimestre se requerirá un mínimo de 75% de muestras válidas. Dato que con los sitios donde el monitoreo no se realice diariamente, se tomará como base el número de muestreos calendarizados para dicho periodo. Si la cantidad de muestras es menor se invalidará el trimestre correspondiente. Para la validación del año es necesario contar con al menos tres trimestres válidos que cumplan con el número de muestras válidas ya especificado, en caso contrario no podrá evaluarse el cumplimiento de la norma para ese año”.

Trimestre	Meses
1	Enero, febrero, marzo
2	Abril, mayo, junio
3	Julio, agosto, septiembre
4	Octubre, noviembre, diciembre

Cuadro III.3. Criterios de Calidad de los Datos de Monitoreo.

Así, el número de muestreos con los que se cuenta para poder hacer la evaluación de la norma de 24 horas y anual de PM₁₀ de los datos del Cuadro III.4 se presentan a continuación.

Trimestre	Total de muestras esperadas en 2005	Al menos el 75% de las muestras	Muestras validas por estación de medición y cumplimiento del criterio de compleción de datos							
			Est 1		Est 2		Est 3		Est 4	
			Muestras	Cumple	Muestras	Cumple	Muestras	Cumple	Muestras	Cumple
Ene-Mar	15	12	13	Sí	15	Sí	13	Sí	15	Sí
Abr-Jun	15	12	15	Sí	14	Sí	15	Sí	15	Sí
Jul-Sep	15	12	11	No	13	Sí	14	Sí	14	Sí
Oct-Dic	15	12	11	No	8	No	9	No	12	Sí

Cuadro III.4. Cumplimiento del criterio de compleción de datos por trimestre.

De los resultados de la tabla anterior y de las especificaciones del numeral 5.4.2 se desprende que no es posible evaluar el cumplimiento de los límites de las PM₁₀ en la estación 1, debido a que en el tercero y cuarto trimestres no se cumplió con los criterios de compleción de datos de al menos el 75% de muestras válidas. Las estaciones 2 y 3 cumplieron con el criterio de compleción en tres de los cuatro trimestres, por lo que si es posible evaluar el cumplimiento de los límites de las PM₁₀, por último la estación 4 cumplió en los cuatro trimestres con el criterio de compleción de datos. Es importante aclarar, que para las estaciones 2 y 3, los muestreos del trimestre en el que no se cumplió con el criterio de compleción de datos, no debe considerarse en los cálculos posteriores.

Paso 3

El paso 3 consiste en realizar los cálculos de acuerdo al numeral 5.4.5 siempre y cuando se cumpla con las especificaciones del numeral 5.4.2

Percentil 98

Los datos que se van a considerar para el calculo del percentil 98 corresponden para la estación 2 y 3 a los muestreos realizados en los tres primeros trimestres del año y para la estación 4 a todos los trimestres. Así, y siguiendo el algoritmo descrito en la sección 4.3.1.2, en el Cuadro III.5 se presenta el lugar (i+1), que corresponde a la concentración del percentil 98.

Estación	Total de muestras válidas (n)	AN*0.98	Parte entera, i	i + 1
Est 2	42 (tres trimestres)	41.16	41	42
Est 3	42 (tres trimestres)	41.16	41	42
Est 4	56 (cuatro trimestres)	54.88	54	55

Cuadro III.5. Desglose del lugar $i+1$ que corresponde al percentil 98

La parte entera es muy fácil de obtener a partir de la siguiente función de Excel:

$$fx = \text{ENTERO}(\text{celda}).$$

Cabe destacar que el lugar $i+1$ de las estaciones 2 y 3 es 42, que es el lugar que corresponde a la máxima concentración medida de PM_{10} entre los meses de enero a septiembre.

Por último, en el Cuadro III.6 se muestran, para la estación 4, las concentraciones diarias de los cuatro trimestres de 2005 en una serie ascendente ($x_1, x_2, x_3, \dots, x_n$) en la que cada valor es igual o mayor que el valor anterior ($x_n \geq x_{n-1}$).

Fecha	Serie ordenada de valores	Fecha	Concentración ordenada
27/02/2005	X1 = 19	31/10/2005	X29 = 100
14/08/2005	X2 = 21	18/12/2005	X30 = 103
13/09/2005	X3 = 21	25/09/2005	X31 = 112
13/10/2005	X4 = 27	22/04/2005	X32 = 115
20/08/2005	X5 = 29	21/07/2005	X33 = 115
27/07/2005	X6 = 34	03/06/2005	X34 = 120
28/05/2005	X7 = 39	22/05/2005	X35 = 125
22/01/2005	X8 = 46	06/11/2005	X36 = 125
16/05/2005	X9 = 46	15/07/2005	X37 = 127
03/02/2005	X10 = 50	28/04/2005	X38 = 128
19/09/2005	X11 = 58	05/03/2005	X39 = 129
27/06/2005	X12 = 60	16/04/2005	X40 = 136
08/08/2005	X13 = 61	18/11/2005	X41 = 136
07/09/2005	X14 = 61	21/02/2005	X42 = 137
07/10/2005	X15 = 61	26/08/2005	X43 = 138
04/05/2005	X16 = 68	17/03/2005	X44 = 147
28/01/2005	X17 = 73	11/03/2005	X45 = 163
02/08/2005	X18 = 78	06/12/2005	X46 = 167
01/09/2005	X19 = 78	10/04/2005	X47 = 169
15/06/2005	X20 = 79	29/03/2005	X48 = 184
09/07/2005	X21 = 82	12/12/2005	X49 = 191
09/02/2005	X22 = 83	10/01/2005	X50 = 193
04/01/2005	X23 = 86	01/10/2005	X51 = 197
21/06/2005	X24 = 89	15/02/2005	X52 = 233
16/01/2005	X25 = 91	23/03/2005	X53 = 235
09/06/2005	X26 = 94	24/11/2005	X54 = 244
10/05/2005	X27 = 96	24/12/2005	X55 = 244
19/10/2005	X28 = 99	04/04/2005	X56 = 275

Cuadro III.6. Concentraciones ordenadas de PM₁₀ (µg/m³), 2005.

El percentil 98, P0.98, es el valor de concentración con índice $i + 1$ en la serie ordenada de valores. En el caso de la estación 4 el percentil 98 es $X_{55} = 244 \mu/m^3$, para la estación 2 es $X_{42} = 92 \mu/m^3$ y para la estación 3 $X_{42} = 426 \mu/m^3$.

Una forma muy sencilla de obtener el percentil 98 en Excel sin necesidad de ordenar las concentraciones diarias es utilizar la siguiente función:

$$f_x = K.ESIMO.MENOR(\text{rango}, i+1).$$

De acuerdo con el numeral 5.4.3. para la determinación del cumplimiento del valor normado de 24 horas para PM_{10} , un sitio cumple con la norma de PM_{10} para el promedio de 24 horas cuando el valor del percentil 98 es menor o igual a $120 \mu g/m^3$. De los resultados obtenidos se concluye que en el año 2005 la estación 2 cumple con el valor normado de 24 horas, por el contrario las estaciones 3 y 4 no cumplen con dicho valor.

Es importante mencionar que los algoritmos que se utilizan en los paquetes estadísticos y Excel para calcular el percentil 98 no siguen el algoritmo descrito anteriormente, por lo que no se recomienda su uso para obtener el percentil 98 y evaluar el cumplimiento del valor de 24 horas de la norma de partículas.

Promedio anual

Al igual que para el percentil 98, los datos que se van a considerar para determinar el cumplimiento del valor anual de las PM_{10} son los de aquellas estaciones que cumplieron con al menos tres trimestres con el criterio de completación de datos en 2005.

De acuerdo al numeral 5.4.5.2., el promedio anual se calcula a partir de los promedios trimestrales. En el Cuadro III.7 se muestran los

promedios para cada trimestre de las tres estaciones de medición que cumplieron con el criterio de compleción de datos, así como el promedio anual.

La determinación del cumplimiento del valor normado del promedio anual para PM₁₀ se especifica en el numeral 5.4.4., e indica que un sitio cumple con la norma anual de PM₁₀ cuando el promedio anual de los valores diarios, calculado como se indica en la sección 5.4.5.2., es menor o igual a 50 µg/m³. De los resultados obtenidos se concluye que en el año 2005 la estación 2 cumple con el valor normado de 24 horas; por el contrario, las estaciones 3 y 4 no cumplen con dicho valor (ver Cuadro III.7).

Trimestre	Est 1	Est 2	Est 3
Ene-Mar	33.5	66.9	124.6
Abr-Jun	42.9	100.6	109.3
Jul-Sep	23.0	45.1	72.5
Oct-Dic	ND	ND	141.2
Promedio anual	33.1	70.9	111.9

ND: Trimestre que no cumple con el criterio de compleción de datos del al menos el 75% de muestreos

Cuadro III.7. Promedios trimestrales de PM₁₀ (µg/m³), 2005.

ANEXO 4

HERRAMIENTAS ESTADÍSTICAS

La estadística descriptiva es utilizada como una herramienta para describir y analizar las características de un conjunto de datos, así como las relaciones que existen con las características de otros conjuntos de datos con los que se compare. En general, se usa para organizar y resumir una cantidad grande de datos, auxiliándose de representaciones numéricas (estadísticas descriptivas básicas) y gráficas, lo que ayuda a una mejor y adecuada interpretación de los datos. En relación con los datos de calidad del aire, su uso comprende desde la limpieza de datos hasta la generación de indicadores.

A continuación se describen algunas de las herramientas estadísticas que se usan en las diferentes etapas del manejo de datos de calidad del aire.

Estadísticas descriptivas básicas:

Medidas de tendencia central

Las medidas de tendencia central proporcionan información acerca de los valores centrales de la variable a estudiar, también se conocen como medidas de posición o localización. Los valores medios dan una idea acerca del comportamiento de la variable.

Las tres medidas de tendencia central más comunes son: **la media o promedio, la mediana y la moda.**

La medida más comúnmente utilizada como el centro de un conjunto de datos es la **media**, la cual se denota como \bar{x} y generalmente se calcula como un promedio aritmético de los datos (ver siguiente ecuación).

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} \text{ para todos los valores de la variable } (i = 1, 2, \dots, n)$$

Entre las ventajas de la media se encuentran las siguientes: es fácil de entender y de calcular, siempre existe, es única y refleja todos los valores del conjunto de datos. La desventaja es que cuando en el conjunto de datos hay valores extremos y datos censurados, su valor se ve afectado por éstos y en algunos casos puede distorsionarse tanto que no represente la tendencia central.

La **mediana** de un conjunto de datos (\tilde{x}) es el valor de la variable que cae en la mitad de los datos ordenados de manera ascendente (el 50% de las mediciones se encuentran a la izquierda y el otro 50% a la derecha). A la mediana también se le conoce como percentil 50. Entre las ventajas de la mediana se encuentran las siguientes: es única, es fácil de calcular, no se ve afectada por valores extremos, se puede usar cuando los datos son cualitativos y donde las categorías tienen un

orden. Para calcular la mediana de un conjunto de datos se ordenan los datos del más pequeño al más grande y se denotan por $X(1)$, $X(2)$, ... , $X(n)$; donde, $X(1)$ es el dato con el valor más pequeño, $X(2)$ es el dato con el segundo valor más pequeño, y así sucesivamente hasta $X(n)$, que es el dato con el valor más grande; usando los datos ordenados la mediana se define de la siguiente manera:

$$\bar{X} = \begin{cases} x_{\left(\frac{n+1}{2}\right)} & \text{si } n \text{ es impar} \\ \frac{x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n}{2}+1\right)}}{2} & \text{si } n \text{ es par} \end{cases}$$

La tercera medida de tendencia central de los datos es **la moda**. La moda se define como el valor que ocurre con más frecuencia. Este valor no siempre existe y puede no ser único, además de que es menos usado que las otras dos medidas. La moda es muy útil cuando se tienen datos cualitativos.

El uso de las medidas de tendencia central en datos de calidad del aire se ilustra con el siguiente ejemplo: los datos horarios de PM_{10} ($\mu\text{g}/\text{m}^3$) generados en una estación de medición el día 25 de enero de 2002 se muestran en la Cuadro IV.1.

Hora	Concentración (µg/m ³)	Hora	Concentración (µg/m ³)
1	88.1	13	37.9
2	144.8	14	43.0
3	135.2	15	53.3
4	118.9	16	57.6
5	116.0	17	44.1
6	106.2	18	69.6
7	93.6	19	88.3
8	129.6	20	93.6
9	180.0	22	88.9
10	189.0	23	50.3
11	116.1	24	35.8
12	70.5	25	28.5

Cuadro IV.1. Datos horarios de PM₁₀.

La media o promedio de PM₁₀ para el día 25 de enero de 2002 está dada (o) por:

$$\bar{x} = \frac{88.1 + 144.8 + 135.2 + 118.9 + \dots + 93.6 + 88.9 + 50.3 + 35.8 + 28.5}{24} = 90.7875 \mu\text{g}/\text{m}^3$$

Esto puede ser considerado como el dato representativo del día. La mediana de PM₁₀ para el mismo día es 88.6 μg/m³ que se obtuvo al ordenar los 24 datos (Cuadro IV.2) y aplicando la ecuación:

$$\bar{x} = \frac{X_{\left(\frac{24}{2}\right)} + X_{\left(\frac{24}{2}+1\right)}}{2} = \frac{X_{(12)} + X_{(13)}}{2} = \frac{88.3 + 88.9}{2} = 88.6 \mu\text{g} / \text{m}^3$$

Concentración ordenada	Concentración ordenada
X(1) = 28.5	X(13) = 88.9
X(2) = 35.8	X(14) = 93.6
X(3) = 37.9	X(15) = 93.6
X(4) = 43.0	X(16) = 106.2
X(5) = 44.1	X(17) = 116.0
X(6) = 50.3	X(18) = 116.1
X(7) = 53.3	X(19) = 118.9
X(8) = 57.6	X(20) = 129.6
X(9) = 69.6	X(21) = 135.2
X(10) = 70.5	X(22) = 144.8
X(11) = 88.1	X(23) = 180.0
X(12) = 88.3	X(24) = 189.0

Cuadro IV.2. Datos horarios de PM₁₀ ordenados.

Moda de la muestra. Al calcular el número de veces que cada valor ocurre se tiene que: $93.6 \mu\text{g}/\text{m}^3$ aparece dos veces, así que este valor es la moda del conjunto de datos. Resumiendo, las medidas de tendencia central media, mediana y moda del conjunto de mediciones de PM_{10} para el día 25 de enero de 2002 están dadas por $90.7875 \mu\text{g}/\text{m}^3$, $88.6 \mu\text{g}/\text{m}^3$ y $93.6 \mu\text{g}/\text{m}^3$ respectivamente, lo que indica de inmediato que la distribución de los datos no es simétrica debido a que las tres medidas de tendencia central son diferentes.

Medidas de posición relativa (percentiles)

En ocasiones, al analizar un conjunto de datos surge el interés de conocer la posición relativa de una o varias observaciones en relación a todas las observaciones. Los percentiles son una medida de la posición relativa que es muy útil en el resumen de un conjunto de los datos. Un percentil es el valor del dato que es más grande que o igual a un porcentaje dado de los valores de los datos.

Entre los percentiles más utilizados se encuentran los cuartiles (percentiles 25, 50 y 75). De estos, el percentil 50 o mediana ya se definió en párrafos anteriores; y los percentiles 25 y 75 son utilizados para estimar la dispersión de un conjunto de datos. En los datos de calidad del aire son muy importantes también los percentiles 90, 95, 98 y 99.

Por ejemplo: el percentil 98 es el valor normado de los promedios de 24 horas de las PM_{10} y las $\text{PM}_{2.5}$.

El cálculo del percentil p se realiza de acuerdo con el siguiente algoritmo: sean X_1, X_2, \dots, X_n el conjunto de datos, se listan los datos de manera ascendente y se denotan por $X(1), X(2), \dots, X(n)$ -de la misma manera como se describió para el cálculo de la mediana-, se multiplica el número de datos n por $p/100$, se divide el resultado en la parte entera y en la parte de la fracción, es decir, $n \cdot p/100 = j + g$, donde j es la parte entera y g es la parte de la fracción. Entonces el percentil p , se calcula de acuerdo a la siguiente ecuación :

$$\text{percentil } p = \begin{cases} \frac{X_{(j)} + X_{(j+1)}}{2} & \text{si } g = 0 \\ X_{(j+1)} & \text{si } g \neq 0 \end{cases}$$

Así, para el conjunto de datos del ejemplo anterior, el percentil 95 se calcula de la siguiente manera: $p = 95$, $p/100 = 95/100 = 0.95$ y $n \cdot p/100 = 24 \cdot 0.95 = 22.8 = 22 + .8$, de donde $j = 22$ y $g = .8$; como g es diferente de cero, el percentil 95 = $X(j+1) = X(22+1) = X(23) = 180 \mu\text{g}/\text{m}^3$ (ver Cuadro III.2). Así, el percentil 95 del conjunto de datos de PM_{10} para el día 25 de enero de 2002 es $180 \mu\text{g}/\text{m}^3$.

La ecuación de la mediana o percentil 50 presentada dentro de las medidas de tendencia

central es un caso particular de la ecuación 3. Al calcular el percentil 50 con la ecuación 3 se tiene que $p/100=50/100=0.50$ y $n \cdot p/100 = 24 \cdot 0.50 = 12$, de donde $j=12$ y $g=0$; por lo que el percentil 50 = $[X(j) + X(j+1)]/2 = [X(12) + X(13)]/2 = [88.3 + 88.9]/2 = 88.6 \mu\text{g}/\text{m}^3$.

Medidas de dispersión.

Las medidas de tendencia central son más útiles si están acompañadas por información que indique cómo los datos se dispersan alrededor de ellas. Las medidas de dispersión más usadas de un conjunto de datos son **el rango, la varianza, la desviación estándar, el coeficiente de variación y el rango intercuartilico**.

La medida de dispersión más fácil de calcular de un conjunto de datos es **el rango (R)**. Se define como la diferencia entre el mayor y el menor valor de los datos. Para conjuntos de datos con pocos valores, el rango es fácil de interpretar y representa adecuadamente la dispersión de los datos; por el contrario, para conjuntos grandes el rango no aporta mucha información dado que para su cálculo se consideran valores extremos. La ecuación del rango está dada por:

$$R = \text{Dato mayor} - \text{dato menor}$$

La varianza de un conjunto de datos representa qué tanto se alejan en promedio las observaciones de un valor fijo (alguna medida de tendencia central, usualmente la media aritmética). Se define como la suma de los cuadrados de las desviaciones respecto a la media (se llama desviación respecto a la media a la diferencia entre cada valor de la variable y la media) dividida entre el número de datos menos 1. La varianza de la muestra es afectada por valores extremos y por datos censurados. Esta medida hereda las unidades de medición de los datos al cuadrado. Su expresión matemática está dada por:

$$s^2 = \frac{\sum_{i=1}^n (x_j - \bar{x})^2}{n - 1} = \frac{\sum_{i=1}^n x_i^2 - \frac{1}{n} (\sum_{i=1}^n x_i)^2}{n - 1}$$

La desviación estandar de un conjunto de datos se define como la raíz cuadrada de la varianza y conserva las mismas unidades de medición de los datos originales, esto es:

$$s = \sqrt{\frac{\sum_{i=1}^n (x_j - \bar{x})^2}{n - 1}} = \sqrt{\frac{\sum_{i=1}^n x_i^2 - \frac{1}{n} (\sum_{i=1}^n x_i)^2}{n - 1}}$$

El coeficiente de variación es una medida relativa que no depende de las unidades originales de los datos y permite comparar la variabilidad de varios conjuntos de datos diferentes. Mientras más grande, más variables son los datos. Se define como el núme-

ro de veces que cabe la media en la desviación estándar. Es decir, qué tan grande es la desviación estándar en relación a la media.

$$CV = \frac{s}{\bar{x}}$$

El coeficiente de variación es usado frecuentemente en aplicaciones ambientales porque la variabilidad (expresada como desviación estándar) es frecuentemente proporcional a la media.

Cuando en un conjunto de datos se tienen valores extremos el **rango intercuartilico** es una medida de dispersión más representativa que la desviación estándar. Se define como la diferencia del percentil 75 con respecto al percentil 25 (ecuación 8).

$$\text{Rango intercuartilico} = \text{Percentil 75} - \text{Percentil 25}$$

Las medidas de dispersión del conjunto de datos de PM₁₀ para el día 25 de enero de 2002 son las siguientes:

$$\text{Rango: } R = \text{Dato mayor} - \text{dato menor} = 189 - 28.5 = 160.5 \mu\text{g}/\text{m}^3,$$

Varianza:

$$s^2 = \frac{([88.1]^2 + [144.8]^2 + \dots + [35.8]^2 + [28.5]^2) - \frac{(2178.9)^2}{24}}{24-1} = 1978.61 (\mu\text{g}/\text{m}^3)^2$$

$$\text{Desviación estándar: } s = \sqrt{1978.61} = 44.48 \mu\text{g} / \text{m}^3$$

$$\text{Coeficiente de variación: } CV = \frac{s}{\bar{x}} = \frac{44.48 \mu\text{g} / \text{m}^3}{90.7875 \mu\text{g} / \text{m}^3} = 0.48995$$

Rango intercuartilico = Percentil 75 – Percentil 25 = $117.5 \mu\text{g}/\text{m}^3 - 51.8 \mu\text{g}/\text{m}^3 = 65.7 \mu\text{g}/\text{m}^3$ donde:

Percentil 75, $p/100=75/100=0.75$ y $N \cdot p/100=(24)(.75)=18$ de donde $j=18$ y $g=0$; $y(75) = [X(j) + X(j+1)]/2 = [X(18) + X(19)]/2 = [116.1 + 118.9]/2 = 117.5 \mu\text{g}/\text{m}^3$.

Percentil 25, $p/100=25/100=0.25$ y $N \cdot p/100=(24)(.25)=6.0$, de donde $j=6$ y $g=0$; $y(25) = [X(j) + X(j+1)]/2 = [X(6) + X(7)]/2 = [50.3 + 53.3]/2 = 51.8 \mu\text{g}/\text{m}^3$.

Medidas de asociación.

Usualmente, se mide más de una variable para cada conjunto de datos y puede haber interés en conocer la relación o el nivel de asociación entre dos o más de esas variables. Una de las medidas de asociación más comunes es **el coeficiente de correlación**. El coeficiente de correlación mide el grado de asociación lineal entre dos variables y no implica causa y efecto. Se puede decir que la correlación entre dos variables es alta y que es fuerte, pero no se puede decir que una variable cause el incremento o decremento en la otra sin evidencia y un fuerte control estadístico. Los coeficientes de correlación más usados son el de Pearson y de Spearman basados en rangos.

Coficiente de Correlación de Pearson.

El coeficiente de correlación de Pearson mide la relación lineal entre dos variables y toma valores entre -1 y 1. Se considera que existe una asociación fuerte si el valor es cercano a 1 o -1 y débil si es casi cero. Si el valor de la correlación lineal es negativo se dirá que la asociación es inversa. Si es positivo se dirá que la asociación es directa.

Cuando el valor de la correlación es cercano a cero las variables no presentan asociación o bien presentan algún tipo de asociación no lineal, como es el caso de la asociación cuadrática, la cúbica o la logarítmica. Entre las propiedades del coeficiente de correlación se encuentra que no se ve afectado por cambios en la localización de los datos (sumar o restar o multiplicar una constante a todos los valores de las mediciones X y/o Y). La correlación de Pearson puede ser sensible a la presencia de uno o más valores extremos en especial cuando el tamaño del conjunto de datos es pequeño. Tales valores pueden resultar en una correlación alta lo que sugeriría una fuerte tendencia lineal cuando sólo una moderada tendencia está presente.

Por ejemplo: cuando un solo par de valores (X,Y) tiene valores muy altos para ambas variables mientras el resto de las variables no están correlacionadas. También sucede al revés, cuando la presencia de valores extremos pueden enmascarar bajas correlaciones entre valores de X y Y, o cuando todos los pares (X,Y) excepto uno o dos tienden a agruparse alrededor de una línea recta y los otros puntos tienen un valor muy grande en X pero moderado o pequeño en Y (o viceversa). Dada la influencia de los valores extremos se recomienda utilizar junto con el coeficiente de correlación una gráfica de dispersión. La expresión matemática para calcular el coeficiente de correlación de Pearson está dada por la siguiente ecuación:

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} = \frac{\sum_{i=1}^n x_i y_i - \frac{\sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n}}{\sqrt{\left[\sum_{i=1}^n x_i^2 - \frac{(\sum_{i=1}^n x_i)^2}{n} \right] \left[\sum_{i=1}^n y_i^2 - \frac{(\sum_{i=1}^n y_i)^2}{n} \right]}}$$

con \bar{x} , \bar{y} las medias aritméticas

Coeficiente de correlación de Spearman basado en rangos.

Una alternativa al coeficiente de correlación de Pearson es el coeficiente de correlación de Spearman basado en rangos. Esto es, reemplazando para el cálculo del coeficiente de correlación cada valor de X y Y por su rango, es decir, asignándole, 1 al valor más pequeño, 2 a el segundo valor más pequeño, y así sucesivamente; y cuando los valores se repiten asignado el promedio de los rangos que les co-

responderían a cada uno de ellos. El par de rangos para X y Y son los que se utilizan para el cálculo del coeficiente de Spearman que se obtiene utilizando la misma fórmula usada para el coeficiente de Pearson. El coeficiente de correlación de Spearman no se ve alterado por transformaciones no lineales crecientes de las X's o de las Y's. El coeficiente de correlación de Spearman es menos sensible a los valores extremos que el coeficiente de Pearson por lo que se convierte en una alternativa o en un complemento del coeficiente de correlación de Pearson.

El uso de la medidas de asociación a datos de calidad del aire se ilustra mediante el siguiente ejemplo: los datos horarios de ozono (O₃, ppb) y temperatura (°C) generados en una estación de medición el día 21 de abril de 2001 se muestran en la Cuadro IV.3.

Hora	X = Concentración de ozono (ppb)	Y = Temperatura (°C)	Hora	X = Concentración de ozono (ppb)	Y = Temperatura (°C)
1	33	22.6	13	60	31.5
2	35	21.8	14	56	33.0
3	25	20.4	15	53	34.0
4	22	20.2	16	49	34.3
5	25	18.9	17	50	34.4
6	31	20.6	18	47	33.7
7	10	18.2	19	44	32.5
8	8	18.6	20	39	30.7
9	13	22.4	21	35	27.8
10	24	24.8	22	32	25.5
11	34	27.4	23	35	24.3
12	42	29.2	24	37	23.4

Cuadro IV.3 Concentración de ozono y temperatura.

Aplicando a los datos anteriores la ecuación para calcular el coeficiente de correlación de Pearson se tiene que $r = 0.87$. Como r es cercano a 1 y positivo se puede inferir que existe una asociación lineal directa entre las dos variables. En relación con el coeficiente de correlación de Spearman basado en rangos se obtiene un valor muy similar al que se obtiene con el de Pearson, los rangos de las mediciones de ozono y temperatura se muestran en el Cuadro IV.4.

Hora	Concentración de ozono	X = Rangos de la concentración de ozono	Temperatura	Y = Rangos de la temperatura
1	1	10	22.6	9
2	2	13	21.8	7
3	3	6.5	20.4	5
4	4	4	20.2	4
5	5	6.5	18.9	3
6	6	8	20.6	6
7	7	2	18.2	1
8	8	1	18.6	2
9	9	3	22.4	8
10	10	5	24.8	12
11	11	11	27.4	14
12	12	17	29.2	16
13	13	24	31.5	18
14	14	23	33.0	20
15	15	22	34.0	22
16	16	20	34.3	23
17	17	21	34.4	24
18	18	19	33.7	21
19	19	18	32.5	19
20	20	16	30.7	17
21	21	13	27.8	15
22	22	9	25.5	13
23	23	13	24.3	11
24	24	15	23.4	10

Cuadro IV.4 Rangos de la concentración de ozono y temperatura.

Aplicando la fórmula de correlación a los rangos de la concentración de ozono y los rangos de la temperatura se tiene que el coeficiente de correlación de Spearman es de 0.887, del cual se infiere que existe una asociación lineal directa entre las dos variables, lo anterior no es de extrañar, debido a que es bien sabido que las concentraciones de ozono aumentan con la presencia de luz solar.

Gráficas

El propósito de esta herramienta es identificar patrones y tendencias de los datos que podrían pasar inadvertidos al usar únicamente estadísticas descriptivas. En la estadística las gráficas se usan para comunicar el comportamiento de un conjunto de datos con claridad, precisión y eficiencia. Al desplegar una gráfica se quiere mostrar los datos, evitar distorsiones de lo que los datos dicen, presentarlos en un espacio pequeño (compactados), hacer coherente un conjunto de datos grande y hacer que los datos revelen niveles de detalle a simple vista. Al revisar las gráficas se puede determinar si los datos son razonables y si hay valores extremadamente altos o valores extremadamente bajos. También se puede determinar qué tipo de distribución tienen, si su distribución es simétrica o si es bimodal y si muestran alguna tendencia obvia.

Gráficas de frecuencia - Histograma

Dos de los métodos más usados para resumir la distribución de los datos son las gráficas de frecuencia y los histogramas. Tanto las gráficas de frecuencia como el histograma usan el mis-

mo principio básico para desplegar los datos: dividir el total de los datos en intervalos contando el número de datos que caen dentro de cada intervalo y desplegando los datos como la altura o área dentro de una gráfica de barras. Hay ligeras diferencias entre el histograma y la gráfica de frecuencia.

En la gráfica de frecuencia la altura relativa de las barras representa la densidad relativa de los datos. En un histograma el área dentro de la barra representa la densidad relativa de los datos. La diferencia entre las dos gráficas se hace notable cuando se utilizan intervalos de diferente tamaño y por ende diferentes tamaños de barras.

El histograma y la gráfica de frecuencia proveen un medio de evaluar la simetría y variabilidad de los datos. Si los datos son simétricos entonces la estructura de estas gráficas deberá ser simétrica alrededor de un punto central (la media, por ejemplo). Este tipo de gráficas indican además si los datos están sesgados y la dirección del sesgo (figura IV.1).

Para dibujar una gráfica de frecuencia y un histograma se parte de un Cuadro de frecuencias o distribución de frecuencias en la cual se organizan y distribuyen los valores de una variable. La distribución de frecuencias consiste en formar cuadros resumen de los valores de la variable, los que pueden agruparse o no. El primer paso es determinar el número de intervalos (no hay una regla para determinarlo) y agrupar los valores de la variable en cada intervalo, los cuales se denominan "intervalos de clase". Al número de valores posibles que caen en cada

clase se le llama “frecuencia absoluta de la clase (Ni)”. Los intervalos de clase no necesariamente son de igual tamaño. Al punto medio entre los límites de cada intervalo de clase se le denomina “marca de clase”. Por lo general es útil presentar las frecuencias absolutas en términos relativos calculando la proporción que del total de valores posibles corresponde a cada valor distinto de la variable o a cada intervalo de clase. Dichas frecuencias reciben el nombre de “frecuencias relativas (fi)”.

A continuación se presenta a través de un ejemplo como generar una gráfica de frecuencia y un histograma considerando el conjunto de datos 17.7, 17.45, 22.8, 35.5, 28.6, 17.2, 19.1, <4, 7.2, <4, 15.2, 14.7, 14.9, 10.9, 12.4, 12.4, 11.6, 14.7, 10.2, 5.2, 16.5 y 8.9 que corresponden a concentraciones de monóxido de carbono (ppm).

Los valores de los datos se encuentran alrededor de 0 y 40 ppm. Se pueden establecer intervalos iguales de tamaño 5 ppm, esto es, 0-5ppm, 5-10ppm, etc. y como convención se considerará que el valor que se incluirá en cada intervalo corresponderá al extremo izquierdo. Esto es, en el caso de un valor de 5 ppm, este será considerado en el intervalo 5-10 ppm en lugar del 0-5 ppm.

El intervalo de clase (columna 1) y el número de observaciones (columna 2) que corresponden a cada intervalo se muestran en la Cuadro IV.5. La frecuencia relativa de cada intervalo se muestra en la columna 3 y en la columna 4 se muestra el porcentaje de observaciones por ppm (esto es, como la unidad común para los datos es 1 ppm y en cada intervalo hay 5 unidades comunes (el tamaño del intervalo), para obtener el porcentaje de observaciones por ppm se debe de dividir la frecuencia relativa entre 5).

La gráfica de frecuencia se construye con la columna 1 y la columna 2 de la Cuadro IV.5, mientras que, el histograma se construye con la columna 1 y la columna 4. El eje horizontal para ambas va de 0 a 40 ppm y el eje vertical para la gráfica de frecuencias es de 0-10 y el eje vertical para el histograma es de 0-10% (de acuerdo a los valores que tomaron la columna 2 y la columna 4 como se muestra en la Figura IV.1).

Intervalo	Número de observaciones (frecuencia absoluta)	Porcentaje de observaciones (frecuencia relativa)	Porcentaje de observaciones por ppm
0-5 ppm	2	9.10	1.8
5-10 ppm	3	13.60	2.7
10-15 ppm	8	36.36	7.3
15-20 ppm	6	27.27	5.5
20-25 ppm	1	4.55	0.9
25-30 ppm	1	4.55	0.9
30-35 ppm	0	0.00	0.0
35-40 ppm	1	4.55	0.9

Cuadro IV.5 Resumen de las frecuencias de las concentraciones de monóxido de carbono.

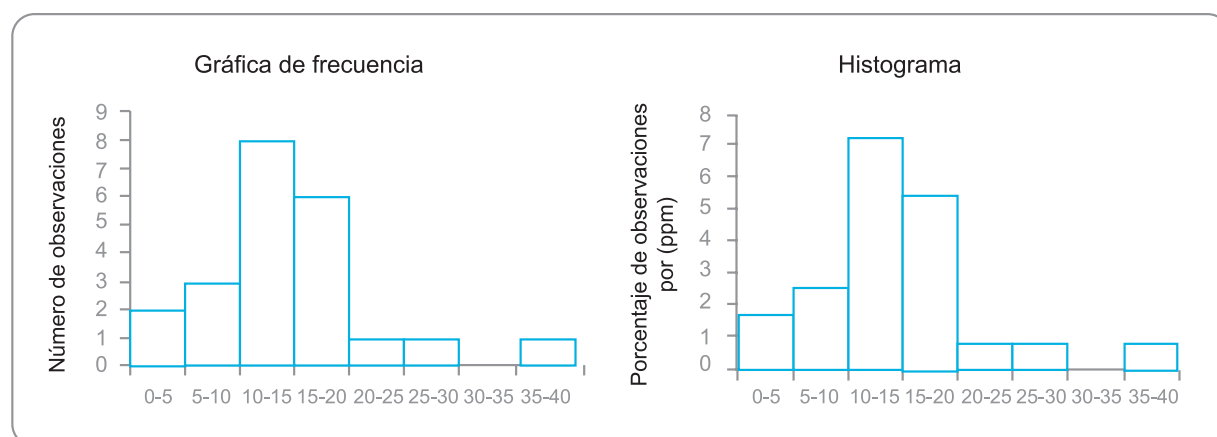


Figura IV.1. Gráfica de frecuencia e histograma de las concentraciones de monóxido de carbono.

En este ejemplo, la gráfica de frecuencia y el histograma se aprecian muy similares porque son iguales los tamaños de los intervalos de clase. En la gráfica de frecuencia la altura de las barras es la que muestra la densidad de los datos, mientras que en el histograma el área de las barras es la que muestra la densidad de los datos.

Gráficas de caja.

Una gráfica de caja, también conocida como de caja y bigotes, es un diagrama esquemático útil para visualizar las estadísticas descriptivas de los datos. La gráfica de caja es útil en situaciones en las cuales no es necesario describir todos los detalles de una distribución.

Una gráfica de caja se encuentra conformada por una caja central dividida por una línea y dos líneas que se extienden fuera de la caja y que se llaman bigotes (figura 8 y la figura IV.2). La longitud de la caja central indica la dispersión de los datos (el 50% de los datos) mientras que la longitud de los bigotes muestra qué tan extendidas están las distribuciones de las colas. El ancho de la caja no tiene un significado particular. La caja puede ser bastante estrecha sin afectar su impacto visual. La mediana de la muestra es la línea horizontal que atraviesa la caja y la media se denota usualmente con el signo de “+” o “x”.

Los valores extremos en la gráfica, ya sean pequeños o grandes, se denotan por un “*” o por “_”. Una gráfica de caja y bigotes puede ser utilizada para evaluar si los datos tienen una distribución simétrica. Si la distribución es simétrica la apariencia de la gráfica será la siguiente: la caja se verá dividida en dos partes iguales por la mediana, la cual coincidirá con la media. La longitud de los bigotes será igual y la cantidad de valores extremos estarán distribuidos de igual manera en cada uno de los lados de la gráfica. La gráfica de caja y bigotes se construye de la siguiente manera:

- Se determina la escala vertical de la gráfica basado en los valores máximo y mínimo del conjunto de los datos y se selecciona un ancho para la caja teniendo en mente que el ancho es solamente una herramienta de visualización. Se le asigna W al ancho de la caja y entonces para la escala horizontal los rangos irán de $-1/2W$ a $1/2 W$.
- Se calcula para el conjunto de datos el percentil 75, el percentil 25, el rango intercuantílico, la media y la mediana.
- Se dibuja una caja a través de los puntos $((-1/2W, \text{percentil } 75), (-1/2W, \text{percentil } 25), (1/2W, \text{percentil } 25), (1/2W, \text{percentil } 75))$ y se dibuja una línea entre los puntos $((1/2W, \text{mediana}), (-1/2W, \text{mediana}))$ y se marca la media con el símbolo “+”.

- Se calcula el tope superior del bigote encontrando [el valor X más grande] < [percentil 75 + 1.5 (rango intercuartilico)] y se dibuja una línea que una los puntos ((0, percentil 75), (0, X)).

- Se calcula el tope inferior del bigote encontrando [el valor Y más pequeño] > [percentil 25 - 1.5 (rango intercuartilico)] y se dibuja una línea que una los puntos ((0, percentil 25), (0, Y)).

- Para todos los puntos $X^* > X$ y $Y^* < Y$, coloque un “*” o “-“ en el punto (0, X^*) y en el punto (0, Y^*).

Al aplicar los pasos anteriores al conjunto de datos 4.0, 6.1, 9.8, 10.7, 10.8, 11.5, 11.6, 12.4, 14.6, 14.7, 14.7, 16.5, 17, 17.5, 20.6, 20.8, 25.7, 25.9, 26.5, 32.0 y 35.5 que corresponden a mediciones de monóxido de carbono (ppm) se tiene:

El rango de los datos van de 4.0 a 35.5 ppm, el cual va a ser el rango del eje vertical. Arbitrariamente se utilizara un ancho de caja de 4 para el eje horizontal.

Aplicando las ecuaciones de la primera sección del anexo: la media del conjunto de datos es 16.87, la mediana es 14.70, el percentil 75 es 20.8, el percentil 25 es 11.5 y el rango intercuartilico es 9.3.

Por lo tanto, la caja se va a dibujar con los puntos (-2,20.8), (-2,11.5), (2,11.5), (2,20.8). La línea que indica la mediana va de (-2,14.7) a (2,14.7) y el punto (0,16.87) se marcará con el signo “+”.

El tope superior esta dado por: percentil 75 + 1.5(rango intercuartilico) = 34.75. El valor más cerrado a este número, pero menor que este es 32.0. Por lo tanto la línea del bigote superior se dibuja de (0,20.8) a (0,32.0).

El tope inferior esta dado por: percentil 25 - 1.5(rango intercuartilico) = -2.45. El valor más cercano a este número, pero más grande que este es 4.0. Por lo tanto la línea del bigote inferior se dibuja de (0,4) a (0,11.5).

Sólo hay un dato con un valor más grande que 32.0, el cual es el 35.5. Por lo que el punto (0,35.5) se marcará con un asterisco “*”. No hay datos con valores menores que 4.0.

Por último en la Figura IV.2 se ilustra la gráfica de caja para las 21 concentraciones de monóxido de carbono.

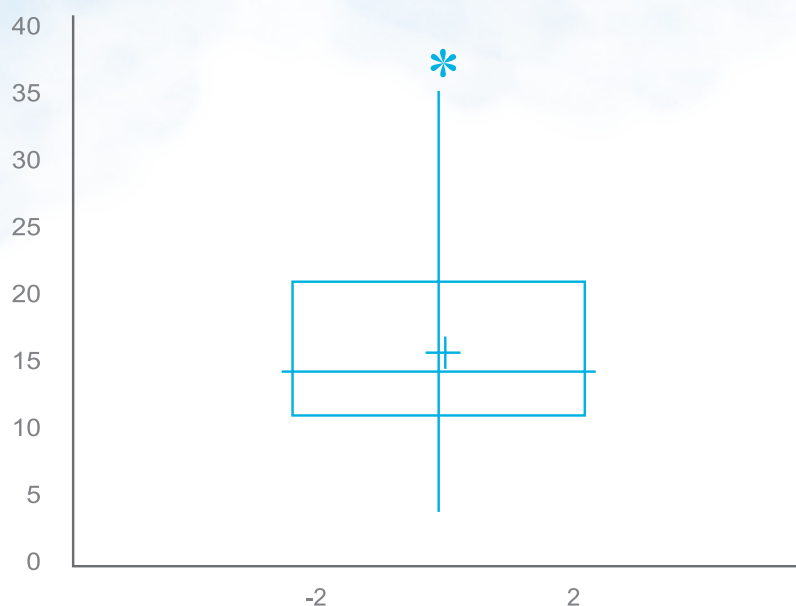


Figura IV.2. Gráfica de caja de las concentraciones de monóxido de carbono.

Gráficas de los rangos de los datos.

La gráfica de los rangos de los datos es una representación útil que es fácil de construir e interpretar y que no hace suposiciones acerca del modelo para los datos. La gráfica de los rangos de los datos despliega cada uno de los valores, por lo que se considera como una representación gráfica de todos los datos en lugar de un resumen, este tipo de gráficas se usa para determinar la densidad de los datos y el sesgo.

La gráfica de los rangos de los datos se genera de la siguiente manera: sea $X_1, X_2, \dots,$

X_n el conjunto de datos y $X(1), X(2), \dots, X(n)$ para $i = 1 \dots n$, los datos listados en orden ascendente, como se describió para la mediana en la primera sección de este anexo. Para generar la gráfica del rango de los datos, se grafican los valores ordenados a lo largo del eje horizontal de los pares $(i, X(i))$. Así, para el conjunto de datos de 22 mediciones de monóxido de carbono (ppm): 17.7, 17.4, 22.8, 35.5, 28.6, 17.2, 19.1, 4.9, 7.2, 4.0, 15.2, 14.7, 14.9, 10.9, 12.4, 12.4, 11.6, 14.7, 10.2, 5.2, 16.5 y 8.9, los rangos de los datos se presentan en la Cuadro IV.6 y la gráfica de los rangos de los datos se ilustra en la Figura IV.3.

I	X(i)	I	X(i)
1	4	12	14.7
2	4.9	13	14.9
3	5.2	14	15.2
4	7.2	15	16.5
5	8.9	16	17.2
6	10.2	17	17.4
7	10.9	18	17.7
8	11.6	19	19.1
9	12.4	20	22.8
10	12.4	21	28.6
11	14.7	22	35.5

Tabla III.6 Rangos de las concentraciones de CO

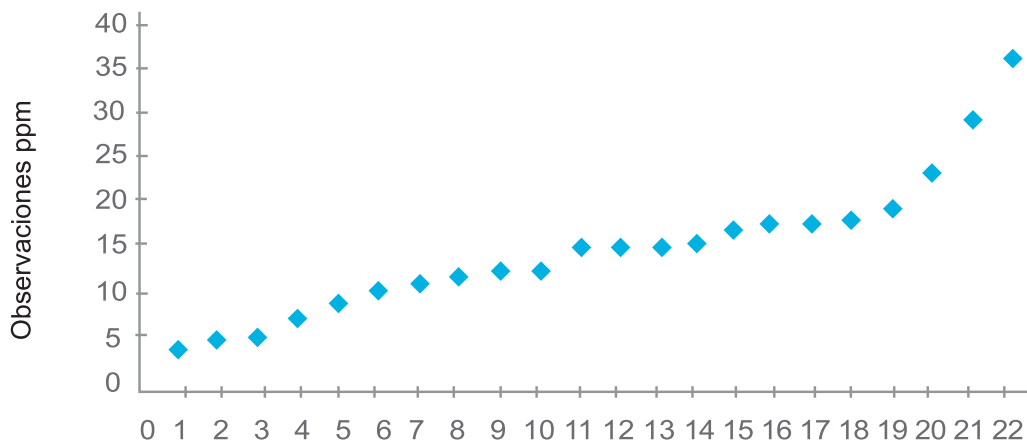


Figura IV.3. Gráfica de los rangos de las concentraciones de CO

Gráficas de los cuartiles.

Una gráfica de cuartiles es una representación del rango de los datos contra la fracción de puntos de datos que excede. Una gráfica de cuartiles puede ser utilizada para leer la información de los cuartiles, tales como la media, los cuartiles y el rango intercuartilico. Es fácil de construir e interpretar, la gráfica de cuartiles presenta cada dato, por lo que es una representación gráfica de toda la información y no un resumen de ésta. De acuerdo a la forma de la gráfica, ésta da una idea de cómo están distribuidos los datos, su densidad y si están sesgados.

Una gráfica de los cuartiles de los datos se genera de la siguiente forma: sea X_1, X_2, \dots, X_n el conjunto de datos y sea $x(i), X(2), \dots, X(n)$ para $i = 1 \dots n$, los datos listados en orden ascendente. Calcule las fracciones $f_i = (i-0.5)/n$, para $i = 1, \dots, n$; grafique los pares $(f_i, X(i))$ con líneas que unan los puntos consecutivos. Se puede adicionar a la gráfica líneas verticales que indiquen el rango intercuartilico y la mediana. Así, para las 22 concentraciones de monóxido de carbono (ppm): 17.7, 17.4, 22.8, 35.5, 28.6, 17.2, 19.1, 4.9, 7.2, 4.0, 15.2, 14.7, 14.9, 10.9, 12.4, 12.4, 11.6, 14.7, 10.2, 5.2, 16.5 y 8.9, las fracciones f_i y los datos ordenados de manera ascendente se muestran en la Cuadro IV.7 y la gráfica de los cuartiles de los datos en la Figura IV.4.

i	X(i)	fi	c	X(i)	fi
1	4	0.02	12	14.7	0.52
2	4.9	0.07	13	14.9	0.57
3	5.2	0.11	14	15.2	0.61
4	7.2	0.16	15	16.5	0.66
5	8.9	0.20	16	17.2	0.70
6	10.2	0.25	17	17.4	0.75
7	10.9	0.30	18	17.7	0.80
8	11.6	0.34	19	19.1	0.84
9	12.4	0.39	20	22.8	0.89
10	12.4	0.43	21	28.6	0.93
11	14.7	0.48	22	35.5	0.98

Cuadro IV.7 Fracciones y concentraciones de CO ordenadas

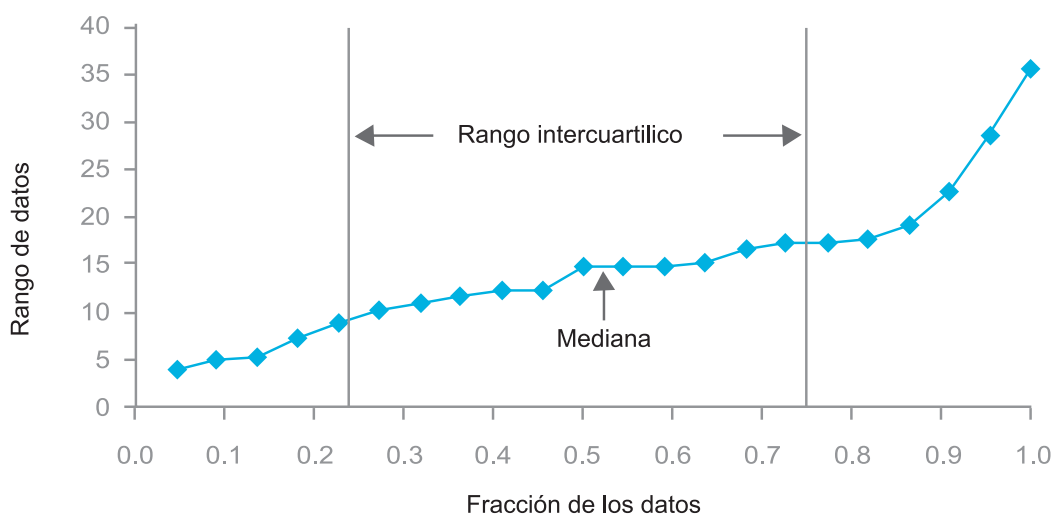


Figura IV.4. Gráfica de los cuartiles de CO

Gráficas de dispersión.

Para grupos de datos de observaciones pareadas donde se miden dos o más variables, una gráfica de dispersión es una de las herramientas más poderosas para analizar la relación entre dos variables. Las gráficas de dispersión son fáciles de construir. Este tipo de gráficas muestra la relación que hay entre dos variables y los datos extremos. En una gráfica de dispersión, si las variables están altamente correlacionadas los pares de datos se agrupan de forma compacta a lo largo de una línea. Adicionalmente, en una gráfica de este tipo se pueden observar muy fácilmente patrones no lineales y agrupaciones de puntos. La gráfica se obtiene al graficar los pares de puntos de las variables X e Y. La gráfica de dispersión de los datos del ejemplo de las medidas de asociación se presenta en la Figura IV.5.

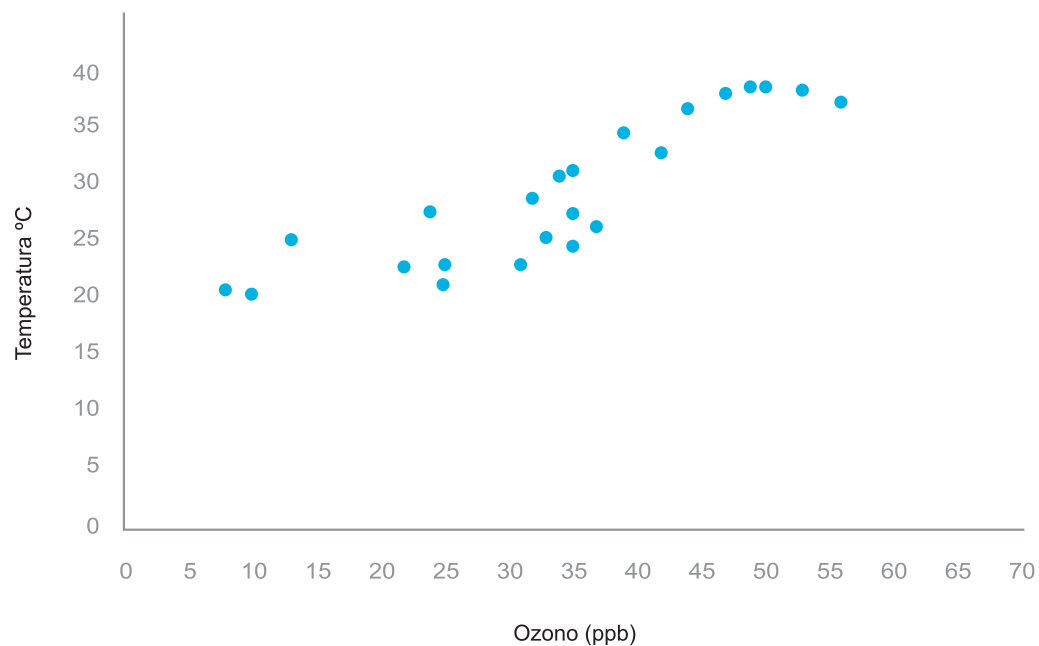


Figura IV.5. Gráfica de dispersión de las concentraciones de ozono vs temperatura

Gráficas en el tiempo.

Es una de las gráficas más simples de generar y ofrece una gran cantidad de información. Esta gráfica hace fácil de identificar las tendencias de corta y larga escala respecto al tiempo, muestra fluctuaciones en periodos pequeños. Por ejemplo, los niveles de monóxido de carbono a lo largo del día comúnmente se incrementan en la hora pico (horas de intenso tráfico vehicular), este comportamiento se repite día con día (Figura 11 de la sección 4.4.1.2). Tendencias a gran escala como fluctuaciones estacionales aparecen como incrementos o caídas en la gráfica. Por ejemplo, los niveles de ozono tienden a ser mayores en el verano que en el invierno. Una gráfica de tiempo puede también mostrar tendencias de dirección e incrementos de variabilidad en el tiempo. Los datos extremos pueden ser fácilmente identificados usando una gráfica en el tiempo.

Gráficas de datos espaciales.

Mapas. La distribución espacial de los contaminantes en las ciudades se puede ilustrar mediante gráficas de contorno en las que la impresión visual de las concentraciones se conjunta con los límites geográficos de la zona a la que pertenecen (ver Figura 13 de la sección 4.4.1.2). Las gráficas de contorno se usan para mostrar todas las tendencias espaciales de los datos interpolando los valores entre los lugares de muestreo. La mayor parte de los procedimientos dependen de la densidad de cobertura de la malla (mayores densidades, darán mayor información). Una gráfica de contorno ofrece una de las mejores imágenes de las características espaciales de los datos.

