

Plan de Monitoreo del Aire de la Comunidad: Apéndice E

Iniciativa de Monitoreo Móvil Estatal de California (SMMI)
Sistema de garantía de calidad de productos de datos
basado en mejoras hiperlocales (v2.1)



16 de mayo de 2025

Este documento contiene descripciones de propiedad intelectual, metodologías e invenciones cubiertas por patentes estadounidenses e internacionales, o patentes pendientes que

son propiedad exclusiva de Aclima Inc.



La Iniciativa de Monitoreo Móvil Estatal es parte de Inversiones Climáticas de California, una iniciativa estatal que utiliza miles de millones de dólares de Tope y Comercio para reducir las emisiones de gases de efecto invernadero, fortalecer la economía y mejorar la salud pública y el medio ambiente, particularmente en comunidades desfavorecidas.

Tabla de contenido

1.0 Introducción	4
2.0 Antecedentes y metodología	6
2.1 Detección de eventos de mejora	6
2.2 Metodología de eventos de mejora	8
2.2.1 Derivación de la línea base	8
2.3.2 Identificación de eventos de mejora	9
2.2.3 Procesamiento de eventos de mejora	10
2.2.4 Resumen de parámetros de entrada críticos	10
2.3 Clústeres de mejora	11
3.0 Impacto de la incertidumbre del sensor	12
4.0 Casos de uso generales y limitaciones	15
4.1 Casos de uso	15
4.2 Limitaciones	15
5.0 Citas	16
7.0 Apéndice	17
7.1 Definiciones	17

1.0 Introducción

El mapeo móvil de la contaminación del aire es un método flexible para medir concentraciones de una amplia gama de contaminantes del aire y gases de efecto invernadero en grandes áreas geográficas con alta resolución espacial. Un producto de datos fundamental que Aclima produce a partir del mapeo móvil son los mapas hiperlocales de la concentración típica de contaminación durante un período de tiempo definido que ilustran áreas de mayor o menor concentración a nivel de calle. Este producto de datos de concentración ambiental está diseñado para producir estimaciones de la concentración ambiental promedio a largo plazo que serían comparables a una medición estacionaria promedio realizada en las cercanías. Como tal, el producto de datos de concentración ambiental puede minimizar la influencia de grandes aumentos de concentración de corta duración que se pueden observar al medir cerca de fuentes de emisión. La detección de estas mediciones superiores al promedio en los datos con resolución temporal es una indicación de una fuente de emisión localizada, y el análisis de estas señales puede facilitar más directamente la identificación y caracterización de fuentes a escalas espaciales hiperlocales.

A medida que la plataforma móvil se mueve a través de una columna de emisiones, se puede observar un aumento temporal en la concentración con el tiempo, lo que llamamos evento de mejora. Definimos estos eventos de mejora como una elevación localizada de la concentración de un contaminante dado dentro de la columna, que se distingue medible mente del fondo ambiental (medido en una ubicación cercana en un momento cercano). Estos eventos de mejora individuales representan la detección de un evento de emisión para un contaminante específico en las inmediaciones de la fuente. En conjunto, estos eventos de mejora se compilan y agregan espacialmente en grupos localizados que pueden utilizarse para identificar la ubicación de las fuentes de contaminación atmosférica.

Este tipo de análisis están diseñados para cumplir los siguientes objetivos de seguimiento:

- Caracterizar la distribución espacial hiperlocal de las fuentes de emisiones contaminantes
- Proporcionar información sobre el tipo de fuente que provoca el aumento de contaminantes.
- Evaluar la persistencia de esa fuente contaminante en un lugar determinado
- PAGProporcionar una perspectiva cualitativa a semicuantitativa Indicación de la magnitud de las mejoras observadas en cualquier ubicación determinada, dentro de las limitaciones de la medición del sensor

Dos características del enfoque de monitoreo de área amplia del SMMI son particularmente adecuadas para respaldar estos basados en mejoras objetivos: 1) medición simultánea de múltiples contaminantes y 2) el muestreo multipaso acercarse.

El monitoreo de áreas específicas también es adecuado para apoyar los objetivos mencionados. Es importante destacar que, dado que el monitoreo de áreas específicas se realiza durante períodos más cortos (es decir, aproximadamente dos semanas), los resultados no son suficientes para caracterizar la persistencia a largo plazo de los contaminantes. Sin embargo, los Laboratorios Móviles de los Socios del SMMI pueden realizar una caracterización química más detallada de los eventos de mejora, lo que fortalece la capacidad de identificación de fuentes y facilita una mejor comprensión del riesgo potencial asociado a cada fuente de contaminantes. (Para obtener más detalles sobre los procedimientos de control de calidad de los laboratorios móviles de socios de SMMI, consulte el Apéndice G de CAMP).

La medición simultánea de múltiples contaminantes permite mayor especificidad a la hora de identificar la fuente de la emisión que provocó el aumento. Por ejemplo, la medición del metano (CH_4) y etano (C_2H_6) proporciona simultáneamente los medios para distinguir entre fuentes biogénicas y termogénicas (es decir, fósiles) de metano. El muestreo multipaso añade una dimensión crítica al análisis: la persistencia. Persistencia se define como la consistencia temporal con la que una ubicación específica se ve afectada por una fuente de emisiones específica. Una alta persistencia de los eventos de mejora puede ser importante para aumentar la confianza en la presencia de una fuente específica en una ubicación específica, ya que significa que es improbable que el grupo de mejora sea el resultado de un pico esporádico debido al ruido del sensor (es muy improbable que un pico esporádico ocurra en la misma ubicación en múltiples ocasiones) y que sea improbable que se deba a la detección de una fuente altamente transitoria (es decir, un solo vehículo altamente contaminante que pasa cerca de la plataforma móvil en una ubicación por lo demás libre de contaminación).

Objetivos de calidad de datos de Aclima para el uso de eventos de mejora para caracterizar las fuentes de emisión de contaminantes atmosféricos se utilizan los siguientes:

- Encuentre y mapee los puntos de probable origen de la contaminación detectando picos notables en las lecturas de medición que superen claramente los niveles de fondo normales. Más específicamente, esto significa que la medición del pico debe tener una relación señal-ruido de al menos 3.
- Garantizar un alto nivel de confianza en las ubicaciones donde se detectan fuentes de contaminación minimizando la presencia de falsos positivos. Estos falsos positivos se minimizan asegurándose de que se detecten múltiples fuentes de emisiones en la misma ubicación antes de identificarla como una posible fuente de contaminación. Esto se cuantifica generalmente como el número de detecciones por visita a una ubicación específica.
- Monitorear y hacer un seguimiento del rendimiento de cada medición subyacente utilizando los indicadores clave de calidad de datos de ganancia, deriva y límite de detección.

Apoyamos la identificación de la fuente de los siguientes contaminantes:

- CH₄, C₂H₆, BC, PM_{2.5}, NO, CO, TVOC, contaminantes tóxicos del aire

En las siguientes secciones, ofrecemos una explicación detallada de la metodología de procesamiento de datos, descripciones de los procesos de control de calidad que sustentan el diseño de productos de datos, una descripción de los procesos de control de calidad relevantes para la recopilación de datos (que se analizan con más detalle en otra sección), un análisis de las limitaciones y casos de uso de los productos de datos, y detalles sobre cada producto de datos individual, junto con ejemplos de validación para algunos de sus casos de uso previstos. Esta metodología es generalizable y puede aplicarse tanto al monitoreo de áreas amplias como al monitoreo de áreas específicas, a los datos de la plataforma móvil Aclima y a los datos del Laboratorio Móvil Colaborador.

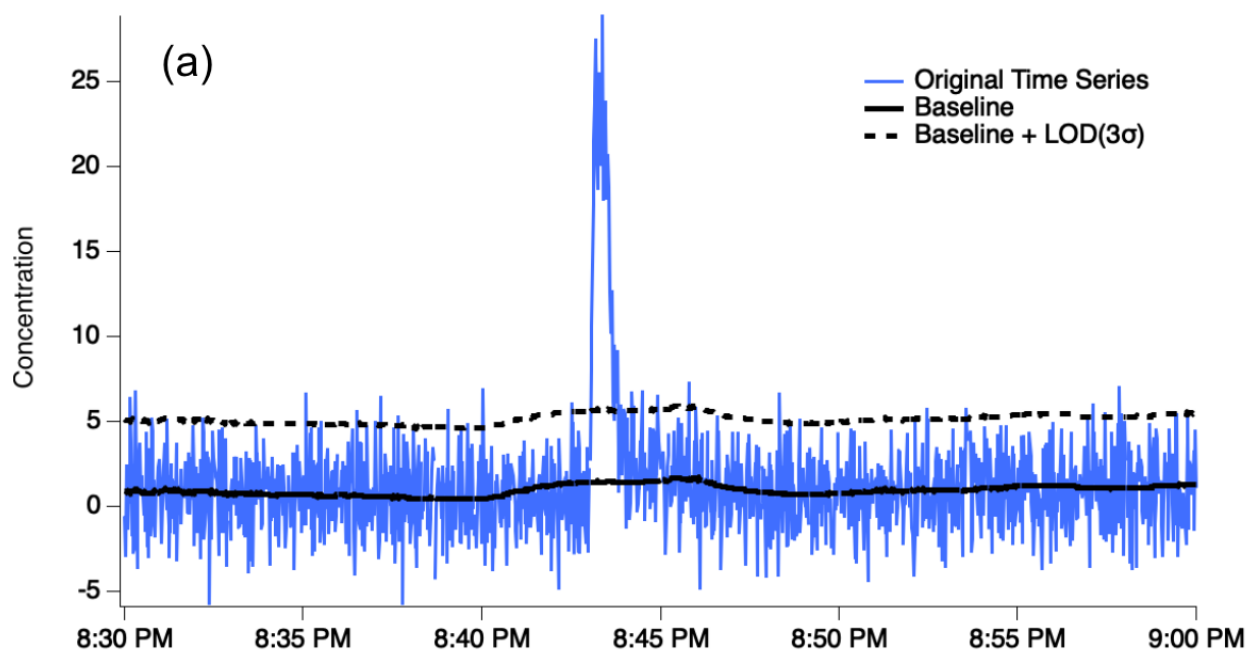
2.0 Antecedentes y metodología

2.1 Detección de eventos de

Las concentraciones de contaminación en un entorno urbano se componen de aire procedente de diferentes regiones afectadas por distintas fuentes. El aire procedente de zonas menos pobladas situadas a barlovento del entorno urbano, denominado fondo regional, se transporta a la región. A medida que las fuentes de contaminación atmosférica de la región se emiten y se mezclan con la atmósfera, se forma un fondo urbano. Este fondo urbano es una mezcla de contaminantes que se emiten directamente o se forman mediante reacciones químicas en la atmósfera. Como resultado, el complejo entorno urbano presenta numerosas fuentes de contaminación que varían en diferentes escalas de tiempo y duración. Fundamentalmente, un evento de mejora que refleja una fuente de emisiones cercana puede distinguirse del fondo ambiental por las diferencias en las escalas de longitud espacio-temporales en la serie temporal del sensor. Las variaciones más lentas en la serie temporal se atribuyen al fondo ambiental, mientras que las más rápidas se atribuyen a mejoras locales. (Wang y otros, 2018, Zimmerman y otros, 2019). En el contexto del monitoreo móvil, las mejoras locales pueden experimentarse en escalas de tiempo aún más rápidas, ya que los vehículos pueden atravesar columnas de emisiones en cuestión de segundos. Este enfoque basado en mejoras se ha demostrado utilizando plataformas móviles para diversas fuentes, incluyendo metano termogénico y biogénico. (Weller y otros, 2018, Moore y otros, 2023).

La Figura 1 ilustra algunos conceptos y definiciones clave utilizados en la descripción de los métodos de procesamiento de datos para detectar eventos de mejora. Si bien las implementaciones individuales de estos métodos varían según los diferentes productos de datos que admiten distintos tipos de fuentes de emisión, esta sección describe el enfoque general, consistente en todas las implementaciones previstas para SMMI. La identificación de un evento de mejora comienza con los datos de la serie temporal de 1 Hz de cada plataforma móvil y generalmente consta de los siguientes pasos de procesamiento: 1) derivar la serie temporal de referencia a partir de la serie temporal original (Figura 1a); 2) restar la serie

temporal de referencia de la serie temporal original para obtener una serie temporal ajustada a la línea base (Figura 1b); 3) encontrar eventos de mejora en la serie temporal ajustada a la línea base que sean estadísticamente significativos; y 4) calcular estadísticas descriptivas sobre el evento de mejora. La Figura 1a muestra la serie temporal original para un contaminante específico (se muestra un evento de mejora de carbono negro, pero el ejemplo pretende ser ilustrativo de cualquier contaminante), la línea de base derivada y la línea de base derivada con un valor constante agregado que corresponde al límite de detección (LOD) del sensor que definimos como tres veces la desviación estándar (σ) de las mediciones del sensor de 1 Hz al medir una concentración relativamente constante (ver detalles en el Sistema de garantía de calidad de medición de la contaminación del aire ambiental móvil) y se denota como 3σ . En este ejemplo se destaca un único evento de mejora de gran magnitud, pero también se pueden observar varios eventos de mejora de menor magnitud ligeramente superiores a la línea discontinua que representa el LOD. La Figura 1b muestra la serie temporal ajustada a la línea base que resulta de restar la línea base de la serie temporal original, donde los valores representan la concentración de mejora. En esta serie temporal transformada, el LOD se representa mediante un único valor, en lugar de un valor que varía con la línea base. En situaciones donde se desea centrarse únicamente en los eventos de mejora más intensos, se puede utilizar un umbral secundario superior al LOD, como se muestra en este ejemplo hipotético. El valor pico de mejora (ilustrado en la Figura 1b) es una métrica típica utilizada para describir cada evento de mejora y se define simplemente como la concentración máxima de la serie temporal ajustada a la línea base durante la ventana del evento de mejora.



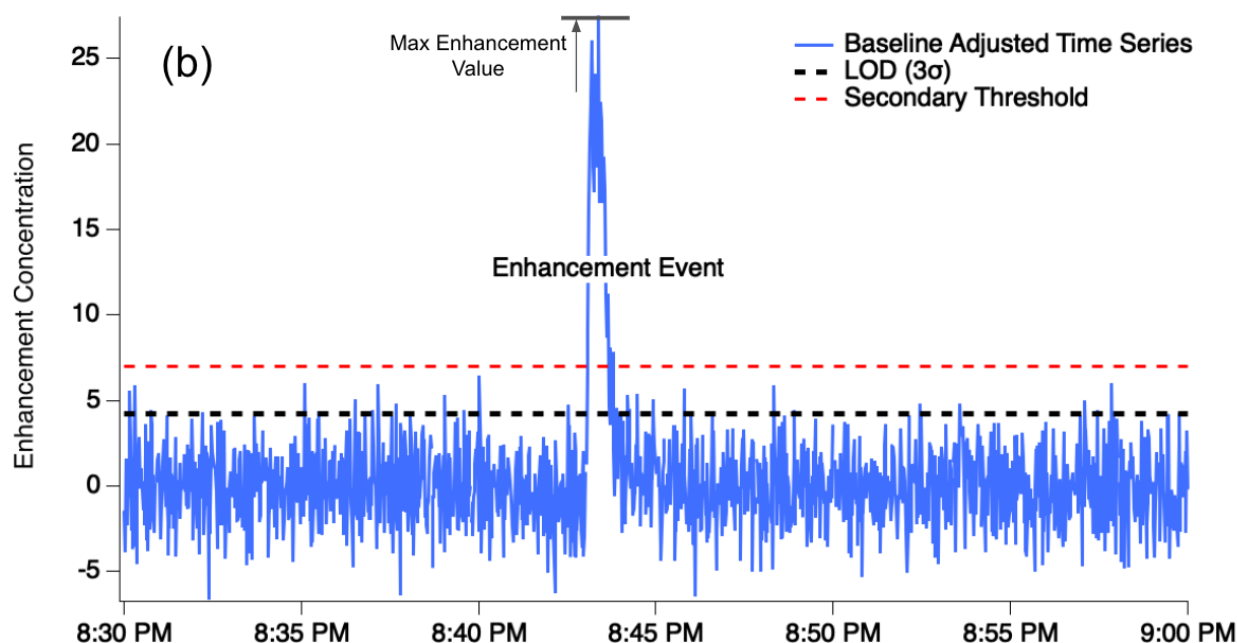


Figura 1: Serie temporal que ilustra un ejemplo de un evento de mejora. Arriba (a) se muestran la serie temporal original, la línea base derivada y la línea base más el LOD de 3σ . Abajo (b) se muestra la serie temporal ajustada a la línea base, calculada restando la línea base de la serie temporal original, con el LOD de 3σ , así como un umbral secundario hipotético que podría utilizarse en casos donde se desee filtrar características de mejora menores que estén solo ligeramente por encima del LOD. Se muestran datos de carbono negro, pero los conceptos se aplican a cualquier contaminante.

Las siguientes secciones proporcionan detalles adicionales sobre cada uno de estos pasos y las diferentes variaciones utilizadas en los productos de datos basados en mejoras de Aclima.

2.2 Metodología de eventos de mejora

2.2.1 Derivación de la línea base

El primer paso en el proceso de procesamiento de datos es derivar la serie temporal de referencia a partir de la serie temporal de 1 Hz. La línea de base de la serie temporal se puede derivar utilizando cualquier método que aisle las características de alta frecuencia de las de baja frecuencia. Además de un enfoque simple de mediana móvil, algunos métodos comúnmente utilizados en la literatura científica atmosférica que aíslan diferentes componentes de frecuencia de los datos de series temporales de contaminantes incluyen el modelo CCGCRV de la NOAA (Thoning y otros, 1989) y el filtrado de Kolmogorov-Zurbenko (Wise y Comrie, 2005). Aclima utiliza principalmente medianas móviles con el fin de derivar la línea de base de la serie temporal original de 1 Hz.

Independientemente del método utilizado, uno de los parámetros de entrada clave es una escala de longitud temporal que determina el límite de tiempo/frecuencia para su inclusión en la línea base. La elección de la escala de longitud temporal influirá en los tipos de

características que se interpretan como mejoras y las que se interpretan como fondo en un producto de datos determinado, y los casos de uso deben tener esto en cuenta. Una vez extraída la línea base de la serie temporal de 1 Hz, se genera la serie temporal ajustada a la línea base (que se muestra en la Figura 1b) restando la línea base de la serie temporal original de 1 Hz. Esta serie temporal ajustada a la línea base se utiliza como datos de entrada para encontrar eventos de mejora..

2.3.2 Identificación de eventos de mejora

Utilizamos dos enfoques generales para identificar eventos de mejora: 1) eventos basados en picos y 2) eventos basados en pases. Cada método tiene ventajas y desventajas, pero la principal diferencia entre ambos es la escala espacial que se representa con mayor eficacia.

Eventos basados en picos Se identifican mediante un algoritmo de detección de picos que los encuentra y caracteriza en la serie temporal ajustada a la línea base. Los picos se caracterizan según su altura (o prominencia) y anchura, su ubicación (latitud y longitud) según el punto de máxima intensidad, y sus horas de inicio y fin. Estas características proporcionan cierta especificidad a la apariencia de un evento de mejora válido y pueden ser ventajosas para identificar fuentes de emisiones con una extensión espacial específica. Por ejemplo, las fugas de gas natural de las tuberías de distribución suelen ser columnas muy discretas espacialmente (aunque no siempre), y resulta útil establecer anchuras de pico mínimas y máximas para reducir la influencia de otras fuentes y determinar con precisión la ubicación donde la intensidad de la columna es máxima. Para identificar la ubicación más probable para la reparación de la fuga.

Eventos basados en pases se identifican en función de una evaluación de una propiedad estadística relevante de un pase de unidad, como por ejemplo si la media, la mediana o el máximo exceden algún umbral. A pase de conducir se define como una agregación de datos de 1 Hz recopilados a medida que un vehículo viaja a través de una unidad espacial definida (por ejemplo, segmento de carretera, hexbin, etc.). A diferencia del enfoque basado en picos, este enfoque es más flexible y tolerante a las características del evento de mejora (es decir, no es sensible a la anchura del pico), sino que permite la detección de un evento que abarca una distancia mayor que la que se detectaría con el enfoque basado en picos. Una desventaja es que asociará un evento de fuente puntual discreto en toda el área cubierta por la unidad espacial; sin embargo, también asociará con mayor precisión los eventos de mejora con mayor distribución espacial a lo largo de toda la trayectoria. Se detectó, lo cual es deseable para capturar la extensión espacial de las emisiones de fuentes de áreas grandes, como ciertos vertederos o granjas.

El factor clave que impulsa la decisión La elección del enfoque más adecuado para productos de datos específicos depende de si las fuentes esperadas son fuentes discretas localizadas fijas en el espacio, fuentes de área mayor o fuentes móviles. Para fuentes discretas localizadas, utilizamos mejoras basadas en picos, y para fuentes de área mayor o móviles, mejoras basadas

en pases. Existe un alto grado de intercambiabilidad entre ambos enfoques y, si bien las mejoras basadas en picos pueden favorecer eventos más localizados, seguirán capturando fuentes de área mayor, y viceversa para las mejoras basadas en pases. Comprender las diferencias entre ambos enfoques es importante para interpretar los resultados y comprender sus limitaciones.

2.2.3 Procesamiento de eventos de mejora

La serie temporal ajustada a la línea base se utiliza como conjunto de datos de entrada para procesar los eventos de mejora mediante cualquiera de estos dos métodos. Los eventos de mejora se derivan de cada vehículo por separado y posteriormente se combinan en un único conjunto de datos que es una compilación de todos los eventos de mejora. Cada evento de mejora contiene metadatos adicionales necesarios para el seguimiento y el posterior procesamiento de datos, incluyendo: un identificador único del evento, un identificador del vehículo, marcas de tiempo de inicio y fin, ubicación (latitud, longitud y, cuando corresponda, un identificador de unidad espacial), velocidad del vehículo y dirección del vehículo. También se calculan y rastrean métricas descriptivas sobre los eventos de mejora. Las métricas clave para las mejoras basadas en picos son la concentración y la anchura de los picos (tanto en unidades de tiempo como en unidades espaciales). Para las mejoras basadas en pases, la métrica clave es la concentración de mejora media de los pases. Finalmente, cada evento de mejora detectado para el contaminante primario está asociado con información sobre los contaminantes detectados simultáneamente en forma de un campo booleano (verdadero/falso) que indica si una mejora particular del contaminante se detecta simultáneamente o no, así como métricas cuantitativas que describen la magnitud de estas mejoras simultáneas (si están presentes).

2.2.4 Resumen de parámetros de entrada críticos

El conjunto de datos de eventos de mejora resultante contiene toda la información relevante necesaria para identificar los eventos que cumplen los requisitos para su inclusión en el análisis de datos pertinente. (No todos los parámetros se utilizan en todos los análisis). A continuación, se presenta un resumen de los parámetros de entrada relevantes:

- **Límite de detección del sensor (LOD)** Este es el umbral mínimo para la detección de eventos de mejora.
- **Umbral secundario para la magnitud de la mejora** En algunos casos es deseable utilizar un umbral secundario para definir eventos calificativos que pueden ser más altos que el LOD del sensor.
- **Tamaño de la ventana temporal de referencia u otros parámetros** En este momento, la mayoría de los productos de datos de Aclima utilizan una mediana móvil simple que requiere un único parámetro de entrada que define la longitud de la ventana.
- **Parámetros de entrada de búsqueda de picos** Para las mejoras basadas en picos, se requieren una serie de parámetros de entrada que definen el tamaño de la ventana de búsqueda, la prominencia mínima (diferencia entre los puntos más altos y más bajos en la serie temporal), el ancho mínimo y la distancia mínima entre picos.

- **Criterios de multicontaminantes** Los parámetros anteriores también deben definirse para cualquier especie secundaria relevante que se utilice en los criterios de identificación del tipo de fuente o su categorización. Además, deben especificarse los criterios para utilizar mejoras de múltiples contaminantes para definir un tipo de fuente específico.
- **Filtro de velocidad del vehículo** Se puede definir una velocidad mínima y/o máxima del vehículo para garantizar que todas las mejoras se muestran de una manera relativamente consistente.
- **Ancho del pico** En algunos casos, se especifican los criterios para un rango de ancho de pico aceptable (tanto como duración como distancia) para garantizar que el sensor u otros artefactos no afecten el producto de datos.

2.3 Clústeres de mejora

Los eventos de mejora individuales proporcionan información sobre las fuentes de emisión presentes en una ubicación específica y en un momento específico. Para aumentar la confianza en que estos eventos de mejora están vinculados a una fuente persistente de emisiones en esa ubicación específica, los eventos de mejora detectados a lo largo del tiempo en la misma ubicación deben agruparse y caracterizarse como un todo. Estas agrupaciones de eventos de mejora se denominarán clústeres de mejora. Para lograr esta agrupación de manera efectiva, se requieren algunas suposiciones con respecto a la distancia a la que dos eventos de mejora diferentes se consideran de la misma fuente. Hay dos enfoques generales: 1) usar agregación espacial simple a una resolución fija o 2) usar un algoritmo de agrupamiento basado en la densidad más complejo (es decir, DBSCAN) que puede resultar en agrupaciones de mejora de tamaño variable dependiendo de la extensión espacial de las detecciones de eventos de mejora que parecen estar asociadas con la misma fuente.

La ventaja del enfoque basado en la densidad para definir grupos de mejora es que las fuentes que están distribuidas espacialmente de manera más amplia pueden caracterizarse con mayor precisión mediante un único grupo de eventos de mejora, manteniendo al mismo tiempo la precisión en la Ubicación de clústeres con eventos que se detectan cerca. Este enfoque funciona bien con productos de datos que representan fuentes de emisiones discretas con una superposición espacial limitada. El enfoque basado en la densidad no es eficaz en casos donde los eventos de mejora son ubicuos espacialmente, como los TVOCs estos escenarios, regiones enteras pueden terminar agrupándolos en un solo grupo si las entradas del algoritmo de detección de mejora y agrupamiento no están finamente ajustadas a cada región específica, lo cual es un enfoque que no funciona bien cuando se intenta aplicar un modelo consistente en múltiples áreas de recolección únicas (por ejemplo, urbanas, rurales, suburbanas, etc.).

El uso de una agregación espacial de resolución fija para definir clústeres de mejora es una solución sencilla para escenarios donde el algoritmo basado en la densidad no funciona correctamente. La desventaja es que las suposiciones sobre las escalas espaciales relevantes

para las zonas de impacto de las fuentes de emisión son fijas. En algunos casos, múltiples fuentes hiperlocales distintas pueden agruparse en un solo clúster, mientras que las fuentes de área más grande pueden repartido en múltiples clústeres.

Estos grupos de mejoras se describen utilizando un conjunto de métricas descriptivas, que varían ligeramente según los diferentes productos de datos, pero generalmente se agrupan en dos categorías diferentes: métricas de magnitud y métricas de persistencia. Métricas de magnitud describir la agregación de concentraciones de eventos de mejora. Métricas de persistencia Describe la tasa de detección de un evento de mejora. Para las métricas de magnitud, una métrica común es el valor máximo de mejora, que corresponde a la mayor concentración de picos de mejora detectada dentro de un clúster de mejora. Para las métricas de persistencia, algunas métricas comunes son el conteo de picos (o número de eventos de mejora), la relación pico-paso (número de picos detectados por cada paso de un vehículo cerca de la ubicación del clúster) y los días únicos (número de días únicos en los que se detectó un evento de mejora). Para los eventos de mejora basados en el paso, se aplica el mismo conjunto de métricas de persistencia, pero se utiliza una nomenclatura diferente: en lugar de conteo de picos y relación pico-paso, se utilizan conteo de aciertos y tasa de aciertos, respectivamente. Los detalles sobre las métricas específicas se proporcionan en la descripción de cada producto de datos.

La siguientes un resumen de los parámetros de entrada críticos necesarios para definir las propiedades de los clústeres de mejora:

- **Tipo y tamaño de agregación espacial** Especifique la resolución basada en densidad o fija, junto con los parámetros relevantes para el tipo especificado. La agregación basada en densidad requiere varios parámetros, en particular la distancia máxima entre eventos. Para la resolución fija, es necesario especificar el tipo y el tamaño de la geometría (es decir, nivel de celda hexbin, longitud del segmento de carretera, etc.).
- **Número mínimo de eventos o días distintos** En algunos casos, se requiere identificar un número mínimo de eventos de mejora para generar un clúster válido. Además, los eventos deben detectarse en un número mínimo de días distintos. Esto proporciona un umbral mínimo de persistencia para el seguimiento de un evento y reduce el número de clústeres debido a eventos esporádicos o artefactos de medición, donde es deseable tener una baja tasa de falsos positivos.

3.0 Impacto de la incertidumbre del sensor

El documento de Aclima sobre Garantía de Calidad en la Medición de la Contaminación del Aire Ambiental analiza el rendimiento de los sensores, las fuentes de incertidumbre y los procesos de control de calidad en los dispositivos móviles. Los productos de datos basados en mejoras no son tan sensibles a ciertas fuentes de incertidumbre como los productos de datos de concentración ambiental. La Tabla 1 resume las fuentes típicas de incertidumbre de los sensores y su respectiva influencia en la calidad de los productos de datos basados en mejoras.

Tabla 1 Resumen de las incertidumbres de medición y sus impactos en los productos de datos basados en mejoras

Fuente de incertidumbre	Sensibilidad del producto de datos de mejora
Precisión de desplazamiento/deriva	Muy bajo
Ganancia de precisión/deriva	Bajo
Ruido aleatorio de alta frecuencia	Alto
Picos anómalos, fallos, cambios rápidos de desplazamiento, etc.	Medio
Especies interferentes	Potencialmente alto

La mayoría de los sensores de calidad del aire presentan diferentes componentes de incertidumbre que varían en distintas escalas temporales. Normalmente, los sensores presentan ruido aleatorio de alta frecuencia, así como variabilidad de baja frecuencia (a menudo denominada "deriva"). El ruido de alta frecuencia suele ser ruido gaussiano ("blanco"), fácil de caracterizar y abordar mediante métodos estadísticos estándar. Por ejemplo, cualquier evento de mejora puede cuantificarse mediante su relación señal-ruido, donde la mejora máxima es la señal y la precisión 1σ es el ruido. Si el perfil de ruido es gaussiano, una mejora con una relación señal-ruido de 3 o superior (es decir, una señal superior a la precisión 3σ por encima de la línea base de medición) tiene una probabilidad del 99,7 % de ser un evento significativo y no un error de medición aleatorio. (Para más información sobre la precisión del conjunto de sensores de Aclima, consulte la Sección 5.5 del Sistema de Garantía de Calidad para la Medición Móvil de la Contaminación del Aire Ambiental).

Los errores de sensor de baja frecuencia, o deriva, pueden ser más difíciles de predecir, varían en diferentes escalas de tiempo y su magnitud. La deriva puede compensarse hasta cierto punto al recalibrar los sensores, pero en algunos casos puede llegar a ser significativa entre calibraciones. Sin embargo, si se elige con cuidado la escala temporal del fondo ambiental, este error de baja frecuencia puede agruparse con la línea base y, por lo tanto, ignorarse en gran medida. Este es un aspecto especialmente ventajoso de estos productos de datos en el contexto de flotas a gran escala desplegadas remotamente, donde no resulta práctico realizar comprobaciones de calidad de alta frecuencia del desplazamiento y la ganancia del sensor en la medida en que normalmente ocurriría en los sitios de monitoreo regulatorios importante comprender las características temporales de la incertidumbre del sensor y definir la escala de tiempo para identificar la línea de base adecuadamente o garantizar que el límite inferior de la magnitud de la mejora se establezca lo suficientemente alto como para descartar la deriva como fuente de cualquier mejora. Aclima utiliza actualmente escalas de longitud muy cortas (de minutos o menos) para definir las líneas base en todos los productos de datos basados en mejoras, lo que resulta en que la deriva sea una fuente insignificante de incertidumbre en la

mayoría de los casos. Se asume que la incertidumbre del sensor a escalas de tiempo más rápidas es puramente gaussiana y se caracteriza por el LOD del sensor con un nivel de confianza específico (p. ej., 2σ , 3σ , etc.), que se utiliza como umbral mínimo para identificar eventos de mejora.

La descripción anterior es válida si la deriva del sensor se encuentra principalmente en su desplazamiento, es decir, si se trata de un valor fijo independientemente de la concentración medida. En este escenario, la magnitud de la mejora no se verá afectada, ya que la deriva es equivalente para el valor de referencia y el valor máximo del evento de mejora; por lo tanto, la diferencia permanece inalterada. Sin embargo, si la deriva se encuentra en la ganancia del sensor (es decir, si el error aumenta con la concentración), la precisión de la magnitud de los eventos de mejora detectados se verá afectada de forma equivalente. Tanto la incertidumbre de la ganancia como la del desplazamiento (y la deriva a lo largo del tiempo) se caracterizan en el proceso de calibración posterior a la implementación de Aclima, como se describe en el documento del Sistema de Garantía de Calidad de la Plataforma Móvil de Medición de la Contaminación del Aire Ambiental. La Tabla 2 muestra la incertidumbre típica de la ganancia para las mediciones relevantes utilizadas en los productos de datos de Aclima basados en la mejora. Estas incertidumbres en la ganancia se derivan de las comprobaciones de calibración relativas a sus calibraciones anteriores, lo que representa derivas a lo largo de 3 a 12 meses. Estas incertidumbres se propagan directamente a la incertidumbre de la medición de eventos de mejora individuales, pero en general estas incertidumbres son muy pequeñas en comparación con la variabilidad en el comportamiento de una columna de emisiones. Por ejemplo, Von Fischer et al. (2017) y Weller et al. (2019) muestran una variabilidad significativa en las concentraciones observadas y las tasas de fuga estimadas a partir del muestreo repetido de la misma fuga. Se prevé que la variabilidad y la incertidumbre para las fuentes de emisiones reales en zonas urbanas complejas sean aún mayores. Por lo tanto, se espera que la incertidumbre de la ganancia del sensor sea insignificante en este contexto.

Tabla 2 Incertidumbres típicas de ganancia del sensor para la plataforma móvil Aclima, derivadas de comprobaciones de calibración

Contaminante	Incertidumbre de ganancia típica
Metano	1%
Etano	1%
TVOC	32% ¹
Carbono negro	3%
NO	22%
CO	6%

PM2.5	10%
-------	-----

¹La sensibilidad del sensor TVOC es muy variable para diferentes clases de compuestos orgánicos.

Una fuente adicional de incertidumbre se relaciona con artefactos anómalos del sensor que ocurren aleatoriamente, como picos de ruido, cambios rápidos de compensación, fallos, etc. Estos no son comunes, pero pueden ocurrir ocasionalmente, especialmente cuando un sensor falla. Como se documenta en el documento Sistema de Garantía de Calidad para la Medición de la Contaminación del Aire Ambiental Móvil, muchos de estos tipos de eventos se detectan y marcan automáticamente como parte del proceso de procesamiento de datos. Sin embargo, también detectamos algunos errores manualmente durante nuestro proceso de revisión de datos, que se marcan por omisión posteriormente, antes de la verificación final de los datos. En ciertos casos, dependiendo de la naturaleza del artefacto, estos eventos pueden registrarse como una mejora individual eventos Mientras el mapeo esté en curso, el equipo de revisión de datos eventualmente identificará y eliminará estos eventos. Sin embargo, dado que estos productos de datos pueden provocar una intervención antes de que los datos estén completamente verificados, Es importante considerar su impacto. Como se discutió en la introducción, hay dos características de estos productos de datos que pueden minimizar el impacto de estos artefactos: 1) muestreo multipaso y 2) múltiples entradas de contaminantes.

El aprovechamiento de múltiples pases en estos productos de datos genera confianza en un conjunto de eventos de mejora, partiendo del supuesto de que sería muy improbable que uno de estos artefactos se repita en la misma ubicación. En este sentido, las métricas de persistencia pueden ser un indicador sólido de confianza en un tipo de fuente específico en una ubicación específica. La inclusión de múltiples entradas de contaminantes como criterio para identificar eventos de mejora que cumplan los requisitos también reduce la probabilidad de detectar falsos positivos debido al sensor. artefactos (picos, fallos, etc.) porque es poco probable que dos sensores que funcionan con dos principios de detección diferentes muestren artefactos aleatorios al mismo tiempo. Cuanto más específicos sean los criterios para identificar una mejora, menos probable será que un artefacto resulte en un falso positivo. Por ejemplo, incluir un rango aceptable de proporciones entre dos contaminantes será más selectivo que el criterio de que dos contaminantes deben experimentar mejoras simultáneamente.

4.0 Casos de uso general y limitaciones

4.1 Casos de uso

En general, estos productos de datos identifican lugares donde hay fuentes específicas de emisiones y ofrecen información sobre las formas más eficaces de intervenir o mitigar esas fuentes. Hay dos casos generales de uso de productos de datos:

- 1) Identificar las ubicaciones de fuentes de emisiones específicas donde una intervención directa puede conducir a una reducción o al fin de la fuente de emisiones individual (por ejemplo, detección de fugas de gas natural para señalarlas para su reparación); los productos de datos pueden proporcionar además un medio para clasificar la gravedad o la consistencia de la fuente de emisiones a fin de priorizar las intervenciones.
- 2) Analice la distribución espacial de los diferentes tipos de fuentes indicados por los productos de datos para comprender mejor qué actividades generales podrían correlacionarse con una mayor densidad de grupos de mejora (por ejemplo, vertederos y metano biogénico; TVOC y estaciones de servicio, etc.).

Como hemos descrito, los productos de datos basados en mejoras son mínimamente sensibles a la desviación del sensor y otros artefactos anómalos. Por lo tanto, estos productos pueden aportar valor a los usuarios en las primeras etapas de la recopilación de datos en una región determinada, y se puede obtener información útil de los datos sin esperar a que pasen a la etapa de datos verificados (véase el documento Sistema de Garantía de Calidad para la Medición de la Contaminación del Aire Ambiental de Mobile Aclima para más detalles sobre las etapas de los datos). Esto es especialmente cierto en el caso de uso 1 descrito anteriormente: identificar fuentes de emisiones individuales que puedan surgir durante la recopilación de datos. Para el caso de uso 2, Es necesario incorporar en el análisis una garantía de calidad adicional Por el usuario final para garantizar un muestreo relativamente equitativo y completo para comparar áreas con alta y baja densidad de conglomerados de mejora. Por lo tanto, el caso de uso 2 podría no ser tan efectivo hasta que se complete la recopilación. Sin embargo, una ventaja importante de los productos de datos de mejora de Aclima es que existe una menor probabilidad de detectar falsos positivos, incluso cuando los datos se encuentran en la etapa preliminar.

4.2 Limitaciones

Cada implementación específica de productos de datos basados en mejoras tiene sus propias limitaciones que deben tenerse en cuenta al interpretar los resultados:

- **No es ideal para capturar fuentes de emisiones esporádicas.** Debido a la naturaleza de la recopilación de datos móviles, es posible que no se detecte una fuente de emisión altamente esporádica; las fuentes de emisión más persistentes o continuas tienen una probabilidad mucho mayor de ser detectadas.
- **Las variaciones en la dirección del viento dificultan la identificación precisa de la ubicación de la fuente.** Si bien los datos del viento se pueden utilizar para mejorar la identificación de la ubicación, aún pueden existir desafíos para identificar con precisión las ubicaciones de las fuentes y la ubicación donde se detectan las mejoras puede no reflejar la ubicación de la fuente responsable. Por lo tanto, se debe tener cuidado y considerar cuidadosamente antes de asignar características de mejora observadas a fuentes específicas, especialmente para fuentes que no se encuentran inmediatamente en la carretera o al costado de la carretera.

- **Las tasas de emisión pueden no ajustarse directamente con la concentración de mejora** Las propiedades cuantitativas de los picos y grupos se pueden utilizar para priorizar ciertas fuentes de emisiones para su mitigación o investigación adicional, pero estas métricas deben interpretarse solo como cualitativas o, En el mejor de los casos, Es semicuantitativo y no permite inferir directamente las tasas de emisión. En algunos casos, el sensor también puede ser una fuente importante de incertidumbre al calcular las concentraciones presentes durante un evento de mejora (p. ej., TVOC). Sin embargo, incluso en casos donde las concentraciones reales se miden con precisión, la dinámica de dispersión de la pluma es una fuente importante de incertidumbre al estimar las tasas de emisión a partir de las concentraciones.
- **Sensibilidad a la ventana de tiempo para la definición de la línea base.** Es posible que no se pueda capturar con precisión la extensión espacial de las fuentes de emisiones individuales en los casos en que el tiempo que lleva atravesar la columna es del mismo orden o más largo que la escala de longitud temporal de la derivación de la línea de base. Se espera que las diferentes fuentes de interés tengan diferentes escalas espacio-temporales. Por ejemplo, una fuente puntual discreta ubicada a nivel de calle, como un autobús parado, puede tener una columna de emisiones asociada de unos pocos metros de ancho que se puede atravesar en segundos, mientras que una fuente de área extensa, como un vertedero o una granja, puede tener una columna de emisiones distribuida a lo largo de cientos o incluso miles de metros, cuya penetración en una plataforma móvil tardaría más. Esto implica la necesidad de seleccionar adecuadamente la escala temporal utilizada para aislar los componentes de mayor frecuencia de la serie temporal.

5.0 Citas

Moore, D.P., Li, N. P., Wendt, L. P., Castañeda, S. R., Falinski, M. M., Zhu, J., Song, C., Ren, Z. J. y Zondlo, M. A. (2023) Subestimación de las emisiones de metano en todo el sector del tratamiento de aguas residuales de los Estados Unidos, *Medio ambiente. Ciencia y tecnología*. 57 (10), 4082-4090, doi: 10.1021/acs.est.2c05373

Thoning, K. W., Tans, P. P., y Komhyr, W. D. (1989) Dióxido de carbono atmosférico en el Observatorio de Mauna Loa: 2. Análisis de los datos GMCC de la NOAA, 1974-1985, *J. Geophys. Res.*, 94(D6), 8549-8565, doi:10.1029/JD094iD06p08549.

por Fischer, J.C., Cooley, D., Chamberlain, S., Gaylord, A., Griebenow, C.J., Hamburg, S.P., Salo, J., Schumacher, R., Theobald, D. y Ham, J. (2017) Identificación rápida, basada en vehículos, de la ubicación y magnitud de fugas en gasoductos urbanos. *Ciencia y tecnología ambiental* 51 (7), 4091-4099, doi: 10.1021/acs.est.6b06095

Wang, J.M., Jeong, C-H., Hilker, N., Shairsingh, K.K., Healy, R.M., Sofowote, U., Debosz, J., Su, Y., McGaughey, M., Doerksen, G., Munoz, T., White, L., Herod, D. y Evans, G.J. (2018) Mediciones de contaminantes del aire cerca de la carretera: Contabilización de la variabilidad entre sitios mediante factores de emisión.

Medio ambiente. Ciencia y tecnología. 52, 16, 9495–9504, doi:10.1021/acs.est.8b01914

Weller, Z.D., Roscioli, J. R., Daube, W. C., Lamb, B. K., Ferrara, T. W., Brewer, P. E. y von Fischer, J. C. (2018) Estudios de metano basados en vehículos para detectar fugas de gas natural y estimar su tamaño: validación e incertidumbre *Medio ambiente. Ciencia y tecnología.* 52 (20), 11922-11930, doi: 10.1021/acs.est.8b03135

Wise, E.K. y Comrie, A.C. (2005) Ampliación del filtro Kolmogorov-Zurbenko: aplicación al ozono, las partículas y las tendencias meteorológicas, *Journal of the Air & Waste Management Association*, 55:8, 1208-1216, doi 10.1080/10473289.2005.10464718

Zimmerman, N., Li, H.Z., Ellis, A., Hauryliuk, A., Robinson E.S., Gu, P., Shah, R.U., Ye, Q., Snell, L., Subramanian, R., Robinson, A.L., Apte, J.S. y Presto, A.A. (2020). Mejora de las correlaciones entre el uso del suelo y las concentraciones de contaminantes atmosféricos mediante análisis wavelet: Perspectivas de una red de sensores de bajo coste. *Calidad del aire en aerosol. Res.* 20: 314-328. doi:10.4209/aaqr.2019.03.0124

7.0 Apéndice

7.1 Definiciones

Recuento de días distintos: La cantidad de días únicos en los que se detectó un pico o evento de mejora en una ubicación específica.

Pase de manejo: se define como una agregación de datos de 1 Hz recopilados mientras un vehículo viaja a través de una unidad espacial definida (segmento de carretera, hexbin, etc.)

Clústeres de mejora: Agrupaciones espaciales de eventos de mejora individuales detectados en diferentes momentos, pero dentro de la misma ubicación.

Eventos de mejora: Elevación localizada en la concentración de un contaminante dado dentro de una columna de emisión que se puede distinguir de manera medible del fondo ambiental.

Límite de detección (LOD): Se define como tres veces la desviación estándar (σ) de las mediciones de 1 Hz del sensor cuando se mide una concentración relativamente constante (ver detalles en Sistema de garantía de calidad de medición de la contaminación del aire ambiental móvil) y se denota como 3σ .

Métricas de magnitud: Una clase de métricas que describe la magnitud de un evento de mejora específico (como concentración) o la suma de la magnitud de múltiples picos detectados dentro del mismo grupo de mejora. La mejora de picos es un ejemplo de métrica de magnitud de evento de mejora. La mejora máxima es un ejemplo de métrica de magnitud de grupo de mejora agregada (es decir, la mejora máxima de picos en todos los picos asociados a un grupo determinado).

Evento de mejora basado en pases: Eventos de mejora que se identifican en función de una evaluación de una propiedad estadística relevante de un pase de unidad, como por ejemplo si la media, la mediana o el máximo superan algún umbral.

Evento de mejora basado en picos: Eventos de mejora que son identificados utilizando un algoritmo de detección de picos que encuentra y caracteriza picos en la serie de tiempo ajustada de referencia.

Relación pico-paso: Describe el porcentaje de veces en que se detecta un evento de mejora de un tipo particular para todos los pases de la unidad espacial.

Métricas de persistencia: Una clase de métricas que describe la consistencia temporal con la que una ubicación específica se ve afectada por una fuente de emisiones específica. La relación pico-paso es un ejemplo de métrica de persistencia.

Antecedentes regionales: Aire de zonas menos pobladas a barlovento del entorno urbano.

Antecedentes urbanos: A medida que las fuentes de contaminación del aire en toda la región se emiten y se mezclan con la atmósfera, se forma un fondo urbano, que es una mezcla de contaminantes que varía lentamente y que se emiten directamente o se forman por reacciones químicas en la atmósfera.