



**UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO
PROGRAMA DE POSGRADO EN CIENCIAS DE LA TIERRA
CENTRO DE CIENCIAS DE LA ATMÓSFERA**

**“ASIMILACIÓN DE DATOS DE VARIABLES METEOROLÓGICAS Y
QUÍMICAS EN LA MODELACIÓN DE LA CALIDAD DEL AIRE.
CASO DE ESTUDIO: CENTRO DE MÉXICO”**

TESIS

QUE PARA OPTAR POR EL GRADO DE
DOCTOR EN CIENCIAS DE LA TIERRA

PRESENTA:

EVELYN ELISA MARTÍNEZ SABARI

TUTOR:

DR. JOSÉ AGUSTÍN GARCÍA REYNOSO, CENTRO DE CIENCIAS DE LA ATMÓSFERA, UNAM

MIEMBROS DEL COMITÉ TUTOR:

DR. LUIS GERARDO RUIZ SUÁREZ, CENTRO DE CIENCIAS DE LA ATMÓSFERA, UNAM
DR. MICHEL GRUTTER DE LA MORA, CENTRO DE CIENCIAS DE LA ATMÓSFERA, UNAM
DR. ERNESTO DOS SANTOS CAETANO NETO, INSTITUTO DE GEOGRAFÍA, UNAM

CIUDAD UNIVERSITARIA, CD. MX.

JUNIO, 2021



Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

AGRADECIMIENTOS

A mi tutor el Dr. José Agustín García Reynoso por orientarme en esta compleja tarea, brindarme su apoyo en todo momento y confiar en mí.

Al Dr. Arthur Mizzi por su vital ayuda para el desarrollo de esta investigación.

Al Dr. Ernesto Dos Santos Caetano Neto, al Dr. Luis Gerardo Ruiz Suárez y al Dr. Michel Grutter de la Mora, miembros de mi Comité Tutor, por sus evaluaciones semestrales y recomendaciones.

Al Dr. Oscar Peralta Rosales, al Dr. Víctor Almanza Veloz y a la Dra. Graciela Herrera Zamarrón por aceptar ser parte de mi jurado examinador y proceder a la revisión de esta tesis.

Al Consejo Nacional de Ciencia y Tecnología (CONACYT) por el apoyo económico otorgado para realizar mis estudios de doctorado.

Al Laboratorio Nacional de Cómputo de Alto Desempeño (LANCAD) y a la Dirección de Computo y de Tecnologías de Información y Comunicación (DGTIC) de la Universidad Nacional Autónoma de México (UNAM) por el tiempo y recursos computacionales otorgados en la supercomputadora Miztli.

Al Posgrado en Ciencias de la Tierra y al Centro de Ciencias de la Atmósfera por permitirme vivir esta experiencia investigativa y el uso del clúster Ometeotl. A sus profesores y plantel en general.

Al Fondo Especial de Ingresos Extraordinarios y su Beca de Apoyo para Titulación y Obtención de Grado.

Al Programa de Apoyo a los Estudios de Posgrado (PAEP) de la UNAM por el apoyo económico brindado para la participación en eventos científicos.

RESUMEN

En este trabajo se realizó la asimilación de datos meteorológicos y químicos de diferentes fuentes para determinar su influencia en la modelación de la calidad del aire en el Centro de México. Para ello, se adaptó el módulo *WRF Data Assimilation* (WRFDA) al modelo *Weather Research and Forecasting Model with Chemistry* (WRF-Chem) y se implementó el sistema WRF-Chem/DART (*Data Assimilation Research Testbed*) para su uso en el área de estudio, incorporando al módulo DART, la capacidad de asimilar especies químicas procedentes de la Red Automática de Monitoreo Atmosférico (RAMA). Los resultados demostraron que la asimilación de datos meteorológicos con WRFDA sí modifica las condiciones iniciales en el modelo WRF-Chem, el cual presentó un mejor desempeño al iniciarse el proceso a las 12 UTC, empleándose una combinación de datos de archivos PREPBUFR con los satelitales. Se comprobó la influencia de asimilar este tipo de variables sobre las concentraciones de ozono (O_3), insuficiente para obtener resultados superiores en la modelación de la variable química. Dentro de las técnicas de asimilación implementadas en WRFDA, se concluyó que el método *Four-Dimensional Variational* (4DVAR) conduce a un mejor desempeño del modelo, a pesar del alto costo computacional asociado y *Ensemble Four-Dimensional Variational* (En3DVAR) ofrece un balance entre los requerimientos de cómputo y la calidad de los resultados. Las métricas estadísticas indicaron mayor correspondencia entre los valores modelados y observados en el período sin contingencia ambiental. En cuanto a la asimilación de variables químicas con WRF-Chem/DART, las diferencias apreciadas entre los campos resultantes, con y sin asimilación de datos, demostraron que el proceso influye en el comportamiento de las variables consideradas (temperatura, monóxido de carbono y O_3) y, además, conduce a un mejor desempeño del modelo. La influencia de asimilar datos puntuales de la RAMA fue menos notable respecto a los satelitales. Estos últimos presentaron métricas estadísticas cualitativamente superiores. Este estudio contribuye a la creación de nuevas capacidades para la investigación de la contaminación atmosférica en México. Además, los resultados obtenidos aportan una información valiosa para la posible incorporación de la asimilación de datos al pronóstico de la calidad del aire en el país.

ABSTRACT

Meteorological data assimilation is carried out to determine its influence on the air quality modeling in Central Mexico area, as well as the assimilation start time and the correct combination of data sources to assimilate. To do so, the Weather Research and Forecasting Data Assimilation module is adapted to WRF-Chem. All the changes resulting from the assimilation to the initial conditions are observed, in addition, it is obtained that the correspondence between the outputs of the model and the observations starting at 12 UTC and using a combination of PREPBUFR data with satellite data offers better statistical results. Likewise, it is verified that assimilation influences in ozone concentrations, although it is not enough to improve the performance of the model for this chemical variable. Later the *Three-Dimensional Variational* (3DVAR), *Four-Dimensional Variational* (4DVAR) and *Ensemble Three-Dimensional Variational* (En3DVAR) methods are applied and a comparison of their performance is made in each period studied. It can be concluded that despite the high computational cost associated with the 4DVAR technique, its use allows obtaining superior results. If you want to obtain a balance between computational cost and acceptable results then the technique to use is En3DVAR. The calculated scores turned out to be better in the period without environmental contingency, which is associated with a less abrupt behavior of the modeled variables. Finally, the WRF-Chem / DART system is implemented, given its ability to assimilate chemical variables from the *Data Assimilation Research Testbed* module (DART). The capacity to assimilate chemical observations from the Automatic Atmospheric Monitoring Network is developed; in addition, CO data from MOPITT were assimilated, all in conjunction with PREPBUFR meteorological data, for a total of eight assimilation cycles. Through the differences between the resulting fields with and without data assimilation, it is shown how the process itself has a certain influence on each of the variables shown (temperature, CO and O₃). In addition, a better correspondence between the modeled values and those measured in the monitoring stations is observed, for all cases with assimilation compared to those without assimilation. The influence of the assimilation of point data is less notable than that of spatial data, obtaining metrics closer to the desired value in the second case.

ÍNDICE GENERAL

INTRODUCCION.....	1
HIPÓTESIS	5
OBJETIVO GENERAL.....	5
OBJETIVOS ESPECÍFICOS.....	6
CAPÍTULO 1: FUNDAMENTACIÓN TEÓRICA Y ANTECEDENTES	9
1.1 La asimilación de datos en la modelación numérica de la atmósfera	9
1.2 Métodos de asimilación de datos	10
1.2.1 Métodos variacionales.....	11
1.2.2 Métodos basados en el Filtro de Kalman.....	13
1.3 Módulos de asimilación de datos	14
1.3.1 Módulo Weather Research and Forecasting Data Assimilation (WRFDA)	14
1.3.2 Data Assimilation Research Testbed (DART)1.3.1 Módulo Weather Research and Forecasting Data Assimilation (WRFDA)	16
1.3.3 Gridpoint Statistical Interpolation (GSI)	18
1.4 Asimilación de datos de variables meteorológicas y químicas en WRF-Chem.....	19
1.5 Antecedentes.....	21
CAPÍTULO 2: MATERIALES Y MÉTODOS	26
2.1 Modelo WRF-Chem: descripción y configuración general.....	26
2.2 Área de estudio y dominios de modelación.....	29
2.3 Descripción de los experimentos numéricos	32
2.3.1 WRF-Chem/WRFDA	32
2.3.2 WRF-Chem/DART	34

2.4 Configuración de los módulos de asimilación de datos dentro del WRF-Chem	37
2.4.1 Adaptación del módulo WRFDA a WRF-Chem	37
2.4.2 Configuración del módulo DART en el WRF-Chem.....	38
2.5 Metodología para incluir observaciones de especies químicas procedentes de estaciones de la RAMA en el módulo DART	40
2.6 Métricas estadísticas	42
CAPÍTULO 3: RESULTADOS Y DISCUSIÓN	26
3.1 Comportamiento de las condiciones iniciales en la asimilación de datos con el módulo WRFDA	45
3.2 Evaluación de desempeño del modelo WRF-Chem en la asimilación de datos meteorológicos con WRFDA.	57
3.3 Evaluación del desempeño del modelo WRF-Chem en la asimilación de datos con WRFDA para las técnicas 3DVAR, 4DVAR y En3DVAR	65
3.3.1 Caso de estudio con contingencia ambiental	65
3.3.2 Caso de estudio sin contingencia ambiental.....	71
3.4 Rendimiento numérico de los métodos de asimilación de datos utilizados en el módulo WRFDA.....	77
3.5 Resultados de las simulaciones numéricas con el sistema WRF-Chem/DART	81
3.6 Resultados de la modelación con el sistema WRF-Chem/DART con asimilación de datos procedentes de observaciones de la RAMA.....	88
3.7 Evaluación de desempeño del sistema WRF-Chem/DART en la asimilación conjunta de datos meteorológicos y químicos	94
CONCLUSIONES.....	97
RECOMENDACIONES	100
REFERENCIAS BIBLIOGRÁFICAS	101

APÉNDICES.....	121
APÉNDICE A: Métodos iniciales de asimilación de datos	121
APÉNDICE B: Métodos variacionales de asimilación de datos	126
APÉNDICE C: Métodos basados en el Filtro de Kalman.....	129
APÉNDICE D: Datos PREPBUFR y BUFR.....	134
APÉNDICE E: Implementación del sistema WRF-Chem/DART	135
APÉNDICE F: Sensores satelitales utilizados.....	136
APÉNDICE G: Pasos para la ejecución de las técnicas de asimilación de datos 3DVAR, 4DVAR y En3DVAR en el módulo WRFDA.....	139
APÉNDICE H: Principales directorios del sistema WRF-Chem/DART	144
APÉNDICE I: Resultados de las simulaciones numéricas con el sistema WRF-Chem/DART para el caso de prueba	147

ÍNDICE DE FIGURAS

Figura 1. Diagrama de flujo del proceso de asimilación en WRFDA (Adaptado del manual de usuario del módulo)	15
Figura 2. Esquema para WRF-Chem/DART (Adaptado de Mizzi, 2020).....	18
Figura 3. Área de estudio.....	29
Figura 4. Dominios de modelación	31
Figura 5. Procesos para los ciclos inicial y de asimilación en el sistema WRF-Chem/DART	39
Figura 6. Asimilación de datos de la RAMA en DART	41
Figura 7. Estaciones de monitoreo de la RAMA seleccionadas.....	42
Figura 8. Estaciones de la Red del Área Metropolitana de Monterrey empleadas en la investigación (http://aire.nl.gob.mx/).....	43

Figura 9. Campos de primera aproximación (a), de análisis (b, c) y de incrementos de análisis (d, e) para la variable T2 y para los casos PB y (PB+RD), el día 2013-05-01 a las 00 UTC	46
Figura 10. Campos de primera aproximación (a), de análisis (b, c) e de incrementos de análisis (d, e) para la variable T2 y para los casos PB y (PB+RD), el día 2013-05-01 a las 12 UTC	47
Figura 11. Campos de primera aproximación (a), de análisis (b, c) y de incrementos de análisis (d, e) para las variables VV y DD, para los casos PB y (PB+RD), el día 2013-05-01 a las 00 UTC	50
Figura 12. Campos de primera aproximación (a), de análisis (b, c) y de incrementos de análisis (d, e) para las variables VV y DD, para los casos PB y (PB+RD), el día 2013-05-01 a las 12 UTC	52
Figura 13. Evolución temporal de la distribución espacial (a, b, c) y los incrementos de análisis para la variable T2 y para los casos PB (d, e, f) y (PB+RD) (g, h, i) para el tercer, sexto y noveno ciclos de asimilación (comenzando desde las 00 UTC).....	53
Figura 14. Evolución temporal de la distribución espacial (a, b, c) y los incrementos de análisis para la variable VV y para los casos PB (d, e, f) y (PB+RD) (g, h, i) para el tercer, sexto y noveno ciclos de asimilación (comenzando desde las 00 UTC).....	54
Figura 15. Evolución temporal de la distribución espacial (a, b, c) de O3 y las diferencias entre los campos de O3 respecto al campo SA para los casos PB (d, e, f) y (PB+RD) (g, h, i) para el tercer, sexto y noveno ciclos de asimilación (comenzando desde las 00 UTC).....	56
Figura 16. Series temporales de los valores simulados y observados de T2 sobre todas las estaciones (a, d), MER (b, e) y SFE (c, f), para todos los casos de asimilación (00 UTC y 12 UTC).....	58

Figura 17. Series temporales de los valores simulados y observados de VV sobre todas las estaciones (a, d), MER (b, e) y SFE (c, f), para todos los casos de asimilación (00 UTC y 12 UTC)	59
Figura 18. Valores de IOA y RMSE para las variables T2 y VV para cada una de las estaciones seleccionadas y para todos los casos de asimilación (00 UTC y 12 UTC)	60
Figura 19. Series temporales de las concentraciones de O₃, simuladas y observadas, para todas las estaciones (a, d), MER (b, e) y SFE (c, f) para todos los casos de asimilación (00 UTC y 12 UTC).....	61
Figura 20. Valores de IOA y RMSE para las concentraciones de O₃ para cada una de las estaciones seleccionadas y para todos los casos de asimilación (00 UTC y 12 UTC)	62
Figura 21. Series temporales de los valores observados y modelados de T2 para el conjunto de todas las estaciones y las seleccionadas, a partir de las diferentes técnicas de asimilación de datos	66
Figura 22. Series temporales de los valores observados y modelados de O₃ para el conjunto de todas las estaciones y las seleccionadas, a partir de las diferentes técnicas de asimilación de datos	69
Figura 23. Series temporales de los valores observados y modelados de T2 para el conjunto de todas las estaciones y las seleccionadas, a partir de las diferentes técnicas de asimilación de datos	72
Figura 24. Series temporales de los valores observados y modelados de O₃ para el conjunto de todas las estaciones y las seleccionadas, a partir de las diferentes técnicas de asimilación de datos	75
Figura 25. Funciones de costo (a) y gradiente (b) para cada método de asimilación de datos	78
Figura 26. Valores medios de los incrementos de análisis de las variables u (a), v (b) y t (c) para los diferentes niveles verticales calculados por el modelo WRF-Chem	79

Figura 27. Valores medios de RMSE de las variables u (a), v (b) y t (c) para los diferentes niveles verticales calculados por el modelo WRF-Chem.....	80
Figura 28. Distribución espacial de CO (media del ensamble en el primer paso de pronóstico) para el dominio Centro de México con asimilación de datos químicos de MOPITT (a), sin asimilación (b) e incrementos de observación (c).....	82
Figura 29. Distribución espacial de O₃ (media del ensamble en el primer paso de pronóstico) para el dominio Centro de México con asimilación de datos químicos de MOPITT (a), sin asimilación (b) e incrementos de observación (c).....	83
Figura 30. Distribución espacial de T2 (media del ensamble en el primer paso de pronóstico) para el dominio Centro de México con asimilación de datos químicos de MOPITT (a), sin asimilación (b) e incrementos de observación (c).....	84
Figura 31: Campos de diferencias entre las condiciones iniciales antes de después de la asimilación de datos para las variables CO (a) y O₃ (b).....	85
Figura 32. Distribución espacial de los perfiles verticales de MOPITT CO para el momento en que se inicia el primer ciclo de asimilación (20140725 a las 00 UTC), mapa completo (a) y acercamiento al dominio de modelación (b)	86
Figura 33. Incrementos de análisis para la variable T2 en el segundo, cuarto, sexto y octavo ciclos de asimilación con el sistema WRF-Chem/DART	89
Figura 34. Incrementos de análisis para la variable CO en el segundo, cuarto, sexto y octavo ciclos de asimilación con el sistema WRF-Chem/DART	90
Figura 35. Incrementos de análisis para la variable O₃ en el segundo, cuarto, sexto y octavo ciclos de asimilación con el sistema WRF-Chem/DART	91
Figura 36. Incrementos de análisis para la variable SO₂ en el segundo, cuarto, sexto y octavo ciclos de asimilación con el sistema WRF-Chem/DART	92
Figura 37. Series temporales de los valores simulados y observados de T2 (a), SO₂ (b), O₃ (c) y CO (d) para todas las estaciones de la RAMA disponibles	94
Figura 38. Valores de IOA para las variables T2 (a), SO₂ (b), O₃ (c) y CO (d) para cada una de las estaciones de la RAMA seleccionadas	95

Figura G1. Ventanas de tiempo para 3DVAR (a) y 4DVAR (b)	140
Figura H1. Directorio FRAPPE_DIR o WORK_DIR.....	144
Figura H2. Directorio real_PANDA_RELEASE_TEST	145
Figura H3. Directorio PANDA_REAL_TIME_DATA	146
Figura I1. Distribución espacial de CO (media del ensamble en el primer paso de pronóstico) para el dominio de Prueba con asimilación de datos químicos de MOPITT (a), sin asimilación (b) e incrementos de observación (c).....	147
Figura I2. Distribución espacial de O₃ (media del ensamble en el primer paso de pronóstico) para el dominio de Prueba con asimilación de datos químicos de MOPITT (a), sin asimilación (b) e incrementos de observación (c).....	148
Figura I3. Distribución espacial de T2 (media del ensamble en el primer paso de pronóstico) para el dominio de Prueba con asimilación de datos químicos de MOPITT (a), sin asimilación (b) e incrementos de observación (c).....	149
Figura I4. Campos de diferencias entre las condiciones iniciales antes de después de la asimilación de datos para las variables CO (a) y O₃ (b)	150

ÍNDICE DE TABLAS

Tabla I. Emisiones de contaminantes criterio en México (INEM-2016 en Mg/año) (https://www.gob.mx/semarnat/documentos/documentos-del-inventario-nacional-de-emisiones)	2
Tabla II. Características de las herramientas de asimilación de datos meteorológicos disponibles para el modelo WRF-Chem	20
Tabla III. Características de las herramientas de asimilación de datos meteorológicos disponibles para el modelo WRF-Chem.....	21
Tabla IV. Parametrizaciones físicas y químicas empleadas en la configuración del modelo WRF-Chem	28

Tabla V. Casos de estudio determinados para la asimilación de datos con WRFDA	33
Tabla VI. Descripción de los casos de estudio considerados en la asimilación de datos con WRFDA	35
Tabla VII. Valores de IOA, BIAS, RMSE y ρ para T2, VV and concentraciones de O₃ para los casos SA, PB y (PB+RD) (00 UTC y 12 UTC)	64
Tabla VIII. Valores de IOA, BIAS, RMSE y ρ para la variable T2 y para las técnicas de asimilación empleadas: SA, 3DVAR, 4DVAR y En3DVAR	67
Tabla IX. Valores de IOA, BIAS, RMSE y ρ para la variable O₃ y para las técnicas de asimilación empleadas: SA, 3DVAR, 4DVAR y En3DVAR	70
Tabla X. Valores de IOA, BIAS, RMSE y ρ para la variable T2 y para las técnicas de asimilación empleadas: SA, 3DVAR, 4DVAR y En3DVAR.....	73
Tabla XI. Valores de IOA, BIAS, RMSE y ρ para la variable O₃ y para las técnicas de asimilación empleadas: SA, 3DVAR, 4DVAR y En3DVAR	76
Tabla XII. Costo computacional asociado a las técnicas de asimilación de datos utilizadas para un solo ciclo.....	81
Tabla XIII. Valores de IOA, BIAS, RMSE y ρ para la variable CO en simulaciones numéricas con asimilación de datos MOPITT CO y sin asimilación para estaciones de monitoreo en Monterrey (24 al 25 de julio de 2014)	87
Tabla XIV. Valores medios de IOA, BIAS, RMSE y ρ calculados considerando todas las estaciones de la RAMA disponibles y para las variables T2, SO₂, O₃ y CO	96

INTRODUCCIÓN

El problema de la contaminación atmosférica es uno de los más importantes hoy en día a nivel mundial, principalmente por las afectaciones que provoca en la salud humana. Según la Organización Mundial de la Salud (OMS), en 2012 murieron 7 millones de personas por causas relacionadas con la contaminación ambiental, 3.7 millones de ellas de forma prematura, asociadas con enfermedades respiratorias y cardíacas, alergias, accidentes cerebrovasculares, cáncer de pulmón y neumopatías crónicas y agudas como el asma. Los niños y adultos mayores fueron los principales afectados (OMS, 2014).

Una buena calidad del aire, se puede definir, como el conjunto de concentraciones de sustancias presentes en la atmósfera en un periodo de tiempo dado que no causan daño a la salud, al bienestar de la población, al equilibrio ecológico o a los materiales con valor económico (PROAIRE, 2012). Por otra parte, la contaminación atmosférica, se refiere a la situación en la que la concentración de dichas sustancias son los suficientemente altas sobre los niveles normales en el ambiente, produciendo efectos nocivos en los seres humanos, animales, vegetación o materiales (Seinfeld y Pandis, 2006).

A escala global, se emiten cada año a la atmósfera grandes cantidades de partículas y gases nocivos que afectan la calidad del aire. Los contaminantes primarios son aquellos que se encuentran en la atmósfera tal cual fueron emitidos, entre los que se encuentran gases como el óxido de azufre (SO_2), el óxido de nitrógeno (NO_x), el amoníaco (NH_3), el metano (CH_4), el monóxido de carbono (CO) y los clorofluorocarbonos (CFC). Además, partículas sólidas o líquidas que son denominadas, en general, como material particulado (PM). Estas se dividen, a su vez, según el diámetro de la partícula en menor a $10\ \mu\text{m}$ (PM_{10}) y menor a $2.5\ \mu\text{m}$ ($\text{PM}_{2.5}$). De igual modo, existen los llamados contaminantes secundarios como el ozono (O_3), los cuales se forman en la atmósfera por reacciones químicas de los primarios, generalmente, en presencia de radiación solar. Por su parte, para evaluar la calidad del aire, se consideran los contaminantes criterio siguientes: O_3 , NO , SO_2 , CO , $\text{PM}_{2.5}$, PM_{10} y plomo (Rodríguez, 2016).

México no es una excepción a esta problemática. Existen serios problemas de calidad del aire en las principales zonas metropolitanas del país, fundamentalmente en la Zona Metropolitana del Valle de México (ZMVM), la más más conocida y documentada, aunque también son de interés los casos de Guadalajara, Monterrey, Puebla, Valle de Toluca, entre otros. Según el Inventario Nacional de Emisiones de 2016 (INEM-2016), se emiten en el país cantidades importantes de contaminantes, tanto gases como partículas, incluyendo los compuestos orgánicos volátiles (COV) que contribuyen a la formación de O₃ (Tabla I). Debido a lo anterior, la gestión de la calidad del aire en México se ha convertido en un tema de alta prioridad (INECC, 2018)

Tabla I. Emisiones de contaminantes criterio en México (INEM-2016 en Mg/año)

(<https://www.gob.mx/semarnat/documentos/documentos-del-inventario-nacional-de-emisiones>)

Tipo de Fuente	PM₁₀	PM_{2.5}	SO₂	CO	NO_x	COV	NH₃
Áreas	603,763	391,647	17,640	2,727,762	130,341	1,591,541	843,906
Fijas	193,527	128,063	1,314,310	288,916	439,065	101,035	3,028
Móviles	92,455	82,843	28,989	5,225,820	1,337,571	519,009	9,077
Naturales	0	0	0	0	1,152,964	9,789,402	0

Una de las formas principales en la actualidad para estimar la calidad del aire en una región es la modelación numérica. Constituye una herramienta físico-matemática que permite simular condiciones reales de transporte y dispersión de contaminantes en la atmósfera, como resultado de la interacción entre las condiciones meteorológicas y las relacionadas con las fuentes de emisión (Pu y Kalnay, 2019). Esto se realiza a través de modelos de calidad del aire, muy utilizados para identificar las contribuciones de las distintas fuentes de emisión, para asistir en el diseño de estrategias encaminadas a la reducción de sustancias nocivas en el aire, para definir la ubicación y la resolución de los sistemas de medición y también para resolver sus deficiencias en términos de resolución espacial y temporal. Una vez validado y adaptado a las condiciones del área donde se va a aplicar un modelo de calidad del aire, este constituye una opción representativa, menos costosa y más rápida que las mediciones, aunque ambas deben complementarse para obtener mejores resultados (Turtós, 2013).

La dinámica atmosférica se rige por leyes físicas expresadas a través de ecuaciones matemáticas que los modelos numéricos resuelven como un problema de condiciones iniciales integradas en el tiempo (Kalnay, 2003), lo cual constituye una tarea compleja y que depende de la precisión de un gran número de parámetros externos. A pesar de los avances en tecnologías de cómputo, las técnicas de recolección de datos iniciales, la selección de parametrizaciones y la resolución de los dominios, entre otros, las evaluaciones de desempeño de los modelos de calidad del aire continúan mostrando errores en sus soluciones numéricas. Entre los factores más importantes que influyen en dichos errores, se encuentran el conocimiento aún insuficiente sobre patrones y tasas de emisión de contaminantes, así como de las condiciones iniciales y de frontera de las variables meteorológicas y de las especies químicas (Russel y Dennis, 2000).

Desde las primeras experiencias en modelación numérica y especialmente en modelos de calidad del aire, se hizo evidente que una de las principales limitantes era el alto margen de error en las condiciones iniciales y de frontera, determinado principalmente por la escasez de datos y el empleo de técnicas inapropiadas para la construcción de las mallas de modelación (Charney, 1951). En el caso de los modelos de área limitada (Mesinger, 1997), los más usados actualmente en contaminación atmosférica, estos emplean los resultados de modelos globales o regionales como datos de entrada principales, los cuales, en muchas ocasiones, no son representativos del área de estudio o no tienen la resolución necesaria, introduciendo mayor incertidumbre en la simulación numérica. Por ello, se desarrollan continuamente diferentes herramientas para refinar estos datos de entrada, según las observaciones reales.

La asimilación de datos es una poderosa herramienta en este sentido. Constituye el proceso mediante el que se incluyen datos de observaciones en un modelo, manteniendo la coherencia en el tiempo de las leyes que gobiernan la evolución y las propiedades físicas de la atmosfera (Bouttier y Courtier, 1999). Existen disímiles ejemplos de su uso en la predicción numérica del tiempo, los cuales demuestran sus ventajas. Incluyen la asimilación de observaciones en superficie, radiosondeos, mediciones indirectas de radares meteorológicos, así como la asimilación de datos obtenidos por complejas técnicas de teledetección mediante satélites.

Los primeros pasos en asimilación de datos se dieron con los inicios de la predicción numérica del tiempo, cuando Richardson (1922) y luego Charney et al. (1950), realizaron interpolaciones manuales de observaciones disponibles para inicializar una malla regular de puntos, que luego sería digitalizada. Sin embargo, su uso en química atmosférica es mucho más reciente, pues los modelos de calidad del aire comenzaron a utilizarse, con mayor frecuencia, a partir de mediados de la década de 1990s (Bocquet et al., 2015).

Actualmente, el desarrollo de esta temática está enfocado en los modelos acoplados química-meteorología, debido a que estos ofrecen la posibilidad de asimilar ambos tipos de datos, de forma independiente o interrelacionados, para lo cual se han desarrollado numerosas técnicas (Zhang, 2008; Baklanov et al., 2014; Bocquet et al., 2015). En este sentido, muchas son las investigaciones realizadas a nivel mundial sobre la asimilación de datos empleando el modelo WRF-Chem. No obstante, en el caso de México, aún son muy escasas, destacándose los trabajos de Bei et al. (2008), Bei et al. (2010), Bei et al. (2012) y Bei et al. (2014).

Teniendo en cuenta lo anterior y dada la importancia que tiene para la modelación numérica el establecer condiciones iniciales y de frontera representativas del área de estudio, en el presente trabajo se realiza la asimilación de datos meteorológicos y químicos de diferentes fuentes, tanto de observaciones en superficie como de radiancia de satélites, para inicializar el modelo acoplado WRF-Chem, con el objetivo de mejorar el proceso de modelación de la calidad del aire en el Centro de México.

Primeramente, se adapta el módulo WRFDA, disponible en WRF, para su uso en WRF-Chem, con el propósito de obtener una primera aproximación a la influencia que tiene la asimilación de datos meteorológicos sobre el comportamiento de las variables químicas simuladas por el modelo, utilizando para ello la técnica *Three-Dimensional Variational* (3DVAR) disponible en WRFDA. Se selecciona un período de estudio comprendido entre el 1 y el 4 de mayo de 2013 (2013-05-01 al 2013-05-04), en el cual se activó el Programa de Contingencia Ambiental y Atmosférico (PCAA) por O₃ en la Ciudad de México. Se definen seis casos de estudio, combinando dos elementos: las fuentes de datos a asimilar y el horario en que inician las simulaciones numéricas.

Luego de realizar los experimentos numéricos, se hace un análisis descriptivo de los resultados obtenidos durante el proceso de asimilación y se evalúa el desempeño del modelo WRF-Chem para las variables químicas y meteorológicas simuladas, utilizando las métricas estadísticas que se describen. Además, se determina el mejor horario para iniciar la asimilación de datos meteorológicos y la fuente más apropiada.

Partiendo de los resultados anteriores, se realiza una evaluación comparativa entre las técnicas de asimilación de datos 3DVAR y *Four-Dimensional Variational* (4DVAR) y una versión híbrida que combina 3DVAR con *Ensemble Kalman Filter* (EnKF), todas disponibles en el módulo WRFDA. Lo anterior, con el objetivo de seleccionar el método que permita un mejor desempeño del modelo WRF-Chem en dos períodos de tiempo con condiciones ambientales diferentes: el primero de ellos coincide con el descrito previamente, mientras que el segundo fue situado entre el 4 y el 7 de mayo de 2013 (2013-05-04 al 2013-05-07), en el cual no se produce activación del PCAA.

Posteriormente, se realiza la asimilación combinada de datos meteorológicos y químicos con el sistema WRF-Chem/DART, con el objetivo de definir un criterio sobre su influencia en las condiciones iniciales que se obtienen y a su vez, en el desempeño del modelo WRF-Chem, en la modelación de la calidad del aire en el Centro de México. Además, se implementa en el módulo DART la capacidad de asimilar datos de especies químicas procedentes de la Red Automática de Monitoreo Atmosférico (RAMA), estableciéndose una metodología para ello.

En sentido general, la novedad científica de esta investigación radica en dos elementos principales. Por una parte, la adaptación realizada al módulo de asimilación de datos WRFDA, haciendo posible su funcionamiento en conjunto con el modelo WRF-Chem para la simulación de la calidad del aire, lo cual nunca antes ha sido publicado. Por otra parte, la validación del funcionamiento del sistema WRF-Chem/DART para la asimilación conjunta de datos químicos y meteorológicos en un área de estudio dentro de México, lo cual tiene gran importancia para la simulación numérica de contaminantes atmosféricos.

Otro aporte significativo consiste en la inclusión de las observaciones de la RAMA como una nueva fuente de datos a asimilar en el sistema WRF-Chem/DART. Esto permitió estudiar el comportamiento del WRF-Chem ante la asimilación de datos procedentes de diferentes orígenes en el área de estudio, es decir, conocer cómo varía el desempeño del modelo al ejecutarse con condiciones iniciales que sean generadas a partir de asimilar observaciones en superficie, además de las satelitales. La metodología desarrollada, no sólo es aplicable al área del Centro de México, sino que su uso puede extenderse a la asimilación de datos en todo el país.

Por último, con el diseño y funcionamiento de los sistemas WRF-Chem/WRFDA y WRF-Chem/DART para el Centro de México, se puede contar con herramientas útiles para estudios de calidad del aire en México, sentando las bases para la inclusión de la asimilación de datos en el pronóstico de calidad del aire en el país. Por este motivo, en la redacción de la tesis se tiene en cuenta la descripción detallada de los procedimientos realizados, los experimentos que se diseñaron, así como la metodología desarrollada en la investigación, permitiendo así que todos estos elementos puedan ser aplicados en trabajos futuros

HIPÓTESIS:

- La asimilación de datos de variables meteorológicas y químicas, procedentes de diferentes fuentes, para inicializar el modelo WRF-Chem, contribuye a mejorar su desempeño y obtener resultados más cercanos a las condiciones reales, por lo que, en consecuencia, se contaría con una herramienta eficaz para la modelación de la calidad del aire en el Centro de México.

OBJETIVO GENERAL

- Asimilar observaciones de variables meteorológicas y químicas procedentes de diferentes fuentes para la modelación de la calidad de aire con WRF-Chem en el Centro de México.

OBJETIVOS ESPECÍFICOS:

- 1- Asimilar datos de variables meteorológicas procedentes de observaciones PREPBUFR y de radiancia de satélites con el módulo WRFDA, para mejorar el campo de inicialización del modelo WRF-Chem.
- 2- Evaluar el desempeño del modelo WRF-Chem respecto a las observaciones disponibles sobre el área de estudio, a partir de los resultados obtenidos en las simulaciones numéricas para cada una de las fuentes de datos empleadas y horarios de inicio de la asimilación.
- 3- Evaluar comparativamente el uso de las diferentes técnicas de asimilación de datos disponibles en el módulo WRFDA para generar condiciones iniciales que puedan mejorar el desempeño del modelo WRF-Chem, utilizando recursos computacionales razonables.
- 4- Implementar el módulo de asimilación de datos DART para su uso en el área del Centro de México de manera eficiente, teniendo en cuenta la capacidad de cómputo disponible.
- 5- Asimilar datos de variables químicas procedentes de satélites y observaciones en superficie con el módulo de asimilación DART.
- 6- Evaluar el desempeño del modelo WRF-Chem utilizando las condiciones iniciales generadas a partir de la asimilación de datos con el módulo DART

En el Capítulo 1 se presentan los métodos y módulos de asimilación de datos, tanto meteorológicos como químicos, más utilizados en las ciencias atmosféricas y descritos en la literatura internacional, señalando sus ventajas y desventajas, así como las principales fuentes de datos a asimilar y su origen. Además, se hace una revisión biobibliográfica acerca del estado del arte del tema que se aborda en la presente investigación, tanto a nivel global como para el caso de México.

En el Capítulo 2 se describen los principales materiales y métodos utilizados en el desarrollo de esta tesis, así como las metodologías empleadas en la implementación de los módulos WRFDA y DART, para la asimilación de datos en el área de interés. Además, se presentan los casos de estudio definidos para la validación del funcionamiento de ambos módulos y sus características generales de configuración.

En el Capítulo 3 se presentan los resultados obtenidos para cada uno de los casos de estudio considerados. Para el módulo WRFDA, se describe el comportamiento de los incrementos de análisis y se determina la fuente de datos y el horario de inicio del proceso de asimilación que permiten mejorar el desempeño del modelo WRF-Chem, a partir de la evaluación estadística realizada. Además, se muestran las diferencias en el comportamiento numérico de los tres métodos de asimilación presentes en la versión de WRFDA que se utiliza. Por otro lado, en lo que respecta al módulo DART, se describen los resultados de la asimilación de especies químicas procedentes de estaciones de la RAMA y se evalúa el comportamiento del sistema WRF-Chem/DART en cuanto a la aplicación de la metodología propuesta.

Finalmente, se presentan las Conclusiones obtenidas con base en los resultados de esta investigación y que responden a los objetivos trazados, así como las Recomendaciones necesarias para trabajos futuros. Luego, se muestran las Referencias Bibliográficas consultadas para el desarrollo de la tesis y debidamente señaladas a lo largo del texto, además de los Apéndices, que sirven de complemento a los resultados que se describen.

CAPÍTULO 1: FUNDAMENTACIÓN TEÓRICA Y ANTECEDENTES

1.1 La asimilación de datos en la modelación numérica de la atmósfera

La modelación numérica es actualmente una de las vías más efectivas y utilizadas para la estimación y evaluación de la calidad del aire. El comportamiento de la atmósfera es gobernado por leyes físicas que pueden expresarse por ecuaciones matemáticas que consideran campos como la temperatura, la dirección y fuerza del viento, la humedad y otros de origen químico, como las concentraciones de fondo de los principales contaminantes atmosféricos.

Si estas ecuaciones pueden resolverse, se tiene una percepción del estado de la atmósfera y las mismas varían según el modelo, pero todas parten del sistema de ecuaciones primitivas, compuesto por: las ecuaciones de movimiento de Navier-Stokes, devenidas de la Segunda Ley de Newton; la ecuación de la termodinámica; la ecuación de continuidad; otra ecuación análoga a la anterior pero aplicada al vapor de agua y la ecuación del gas ideal (Skamarock et al. 2008).

En el caso particular de los modelos de química atmosférica, las ecuaciones agregadas son de tipo conservativo y de esta forma, se aplica un algoritmo matemático a cada fenómeno que experimentan los contaminantes, sumando sus contribuciones en una ecuación de balance denominada advección-difusión, que para n contaminantes i , se puede representar en las tres direcciones j , como señalan Seinfeld y Pandis (2006):

$$\frac{\partial (j c_i)}{\partial x_j} = u \frac{\partial c_i}{\partial t} + D_i \frac{\partial^2 c_i}{\partial x_j \partial x_j} + R_i(c_1, c_2, \dots, c_n, T) + S(\vec{X}, t); i = 1, 2, \dots, n; j = 1, 2, 3$$

El primer término representa la variación de la concentración del contaminante i en el tiempo; el segundo, la entrada neta del contaminante asociada al flujo atmosférico de velocidad u (advección) en las diferentes direcciones j ; el tercero, la variación de la concentración del contaminante debido a la difusión turbulenta; el cuarto, la velocidad de desaparición del contaminante por reacción química o deposición (sumidero) y el quinto, es el término fuente, el cual se refiere a la incorporación del contaminante al entorno considerado, proveniente de focos emisores propios.

La resolución del sistema de ecuaciones físico-matemáticas que compone los modelos, es un problema de condiciones iniciales que son integradas en el tiempo según las leyes que rigen la física atmosférica (Kalnay, 2003). La obtención de resultados más cercanos a la realidad depende de cuán realista sea el modelo matemático y del grado de exactitud de las condiciones iniciales. Este último aspecto es un problema de gran importancia y complejidad, que se ha convertido por sí mismo en una ciencia (Daley, 1991). La disponibilidad de datos y el uso de mejores técnicas para su asimilación son factores de gran peso en la exactitud de los modelos numéricos que se emplean.

La asimilación de datos puede definirse como el proceso mediante el cual se incluyen datos en un modelo, manteniendo la coherencia de las leyes que rigen la evolución en el tiempo y las propiedades físicas de la atmósfera (Bouttier y Courtier, 1999). Por su parte, se define el análisis objetivo, como el método mediante el cual se crea una malla regular de valores (o cualquier otra representación espectral) con los campos de las variables del modelo, a partir de valores observados distribuidos aleatoriamente en el espacio. En la construcción de los campos puede estar presente una aproximación inicial que se combinará con los valores observados. Un buen análisis objetivo debe ser representativo de los datos observados, y mantener la suavidad del campo analizado.

Se denomina campos de primera aproximación a los datos de entrada principales del modelo numérico que usualmente son las salidas de pronóstico de otro modelo. En el proceso de asimilación de datos, estos campos se combinan con datos de observaciones mediante un análisis objetivo de los mismos.

1.2 Métodos de asimilación de datos

Los primeros trabajos de asimilación de datos a nivel mundial, fueron realizados por Richardson (1922) y Charney et al. (1950), los cuales consumían un tiempo muy elevado y se hizo necesario automatizar el proceso (Charney, 1951). Además, la escasez de datos existentes provocó que los campos de las variables mostraran una distribución espacial y temporal poco uniforme, por lo que fue imprescindible el uso de información adicional (aproximación inicial o *background*) para la construcción de las condiciones iniciales de los modelos (Bergthorsson y Döös, 1955).

Debido a lo anterior, se produce un desarrollo paulatino de varios métodos iniciales de asimilación de datos, pudiéndose mencionar: los de aproximaciones sucesivas, los de *nudging* y el de interpolación óptima, los cuales se describen en el Apéndice A.

Estos métodos iniciales son muy simples y no consideran restricciones dinámicas como el balance geostrófico o hidrostático, por lo que se generan ondas de gravedad inercial en los primeros pasos de integración. Dado que los movimientos atmosféricos no son lineales, una consideración de balance geostrófico lineal no es suficiente para asegurar condiciones iniciales balanceadas. En este caso, el método normal de inicialización de modo no lineal, se puede utilizar como una alternativa, separando las componentes de alta y baja frecuencias de los datos de entrada iniciales, mediante funciones de la estructura vertical y horizontal de la atmósfera inicial.

1.2.1 Métodos variacionales

Los métodos variacionales de asimilación de datos son los más populares dentro de la modelación numérica del tiempo pues permiten eliminar el paso de inicialización. Estos incluyen: 3DVAR (Sasaki, 1970) y 4DVAR (Bouttier y Rabier, 1997) y pueden representarse a través de una matriz de ecuaciones para múltiples nodos de malla y variables, cuya expresión está dada por:

$$J(x) = \frac{1}{2}(x - x_b)^T B^{-1}(x - x_b) + \frac{1}{2}(x - x_o)^T R^{-1}(x - x_o)$$

donde $J(x)$ es el jacobiano de x (vector de $n \times m$ valores, siendo n el número de puntos y m el número de variables dependientes); B y R son las matrices que contienen la información referente a los errores estadísticos del *background* y las observaciones, respectivamente; $(x - x_b)$, término conocido como incrementos de análisis, constituye la diferencia entre el campo de análisis x (obtenido después de la asimilación de datos) y el campo de primera aproximación x_b (condición inicial obtenida, en esta tesis, de campos procedentes del modelo NARR); mientras que x_o representa las observaciones, las cuales son usualmente transformadas en una matriz consistente con el sistema a través de un operador lineal.

De esta forma, a partir de una variable de control, se obtiene el mínimo de $J(x)$, que se denomina función de costo, buscando que el campo de análisis se acerque lo más posible al *background*. Si son incluidas observaciones anteriores o posteriores al tiempo de análisis, entonces la técnica se considera tetradimensional (4DVAR), de lo contrario, tridimensional (3DVAR). Más detalles matemáticos sobre algunas variantes de estos métodos, se muestran en el Apéndice B.

En el caso de la técnica 3DVAR, no existe una selección de datos, es decir, todos los disponibles se utilizan para la asimilación de forma simultánea, evitando inestabilidades en los límites de los campos de las diferentes observaciones. La matriz de los errores de covarianza del *background* se define con un enfoque global y más general respecto a las aproximaciones locales que se utilizan en los métodos iniciales.

En 3DVAR es posible añadir restricciones a la función de costo para controlar el ruido, por ejemplo, sin aumentar los requerimientos computacionales de la minimización, así como incorporar relaciones no lineales entre las variables observadas y las del modelo.

El método 4DVAR tiene de ventaja sobre 3DVAR que las diferencias entre el modelo y las observaciones se calculan en el mismo horario en que se tomó la observación y, además, con una mayor resolución temporal. En 4DVAR se busca la evolución de la predicción y se ajustan, de forma óptima, las observaciones disponibles. Es una técnica mucho más compleja que 3DVAR pues utiliza un modelo de predicción lineal y su adjunto para resolver el problema de la minimización de la función de costo de manera eficiente. Las matrices de los errores de covarianza referidas a las observaciones (R) y al análisis (B), se estiman con mayor precisión.

La dinámica y la física del modelo de predicción es una parte integral de la técnica 4DVAR, de forma que las observaciones se utilizan de un modo más consistente desde el punto de vista meteorológico y con un cuidadoso control de calidad. Además, combina observaciones de diferentes tiempos en la ventana atmosférica de asimilación, reduciendo los errores del análisis. Sin embargo, 4DVAR es una opción más compleja que 3DVAR pues necesita un modelo de predicción linealizado y su adjunto para resolver el problema de la minimización de la función coste eficientemente

1.2.2 Métodos basados en el Filtro de Kalman

Los métodos de asimilación de datos basados en el Filtro de Kalman (KF), constituyen una alternativa a las técnicas variacionales anteriormente presentadas. Estos funcionan a través de un mecanismo de predicción y corrección, con el cual se obtiene una nueva estimación del estado de la variable asimilada (vector estado), a partir de su estimación anterior. Para ello, se adiciona un término de corrección que incorpora la información proporcionada por las nuevas mediciones, de forma que el error de la predicción se minimiza estadísticamente.

KF se considera un procedimiento matemático secuencial, pues recalcula la solución cada vez que una nueva medición está disponible, sin necesidad de utilizar los datos antiguos nuevamente. De esta forma, se obtienen estimaciones lineales insesgadas de mínima varianza para el estado de un sistema a partir de datos con ruido. Se basa en dos expresiones principales: la ecuación de estado y la ecuación de medición, las cuales contienen, respectivamente, la información referida al pronóstico y a las observaciones a asimilar. La primera ecuación describe la evolución del estado del sistema a través del tiempo y la segunda relaciona al estado con los datos.

Existen diferentes enfoques teóricos para comprender la génesis de cada una de las variantes asociadas al método KF, los cuales son desarrollados ampliamente en Mitchell (2001), Anderson (2001) y Anderson (2003). Una descripción sobre las variantes más utilizadas dentro de los módulos de asimilación de datos, se muestra en el Apéndice C.

Los métodos de asimilación basados en KF son sencillos de diseñar y codificar, no necesitan de un modelo lineal adjunto para su ejecución y tienen la capacidad de calcular los errores del *background* y el análisis al igual que 4DVAR, pero son actualizados a partir de las perturbaciones que se le van realizando a las condiciones iniciales generadas. Todas las observaciones a asimilar son incluidas, considerando su propiedad de localización (Petrie, 2008) Además, son menos costosos que 4DVAR, dependiendo sobre todo de cuantos miembros se usen para generar el ensamble.

1.3 Módulos de asimilación de datos

Con el objetivo principal de resumir el trabajo realizado por diferentes investigadores, se han desarrollado algunos módulos de asimilación de datos, los cuales tienen la ventaja de reunir, en un solo conjunto de instrucciones, la ejecución de distintos métodos de asimilación. Pueden formar parte de un modelo numérico en sí, como el caso del WRFDA, o ser independientes y adaptables al sistema numérico en cuestión, como ocurre con el DART o el *Gridpoint Statistical Interpolation* (GSI).

1.3.1 Módulo Weather Research and Forecasting Data Assimilation (WRFDA)

WRFDA es un módulo implementado en el modelo WRF, el cual permite la asimilación de datos meteorológicos mediante diferentes técnicas. En la versión 3.8, utilizada en la presente investigación, están disponibles: 3DVAR (Barker et al. 2004), 4DVAR (Huang et al., 2009) y una variante híbrida que combina las anteriores con ensambles (Wang et al., 2008a; Wang et al., 2008b).

Estas técnicas de asimilación variacional se aplican mediante un proceso iterativo de minimización de una función de costo conocida (en este caso, el método iterativo implementado en el módulo WRFDA es el Gradiente Conjugado), por lo que no es necesario el paso de inicialización. Uno de sus beneficios principales consiste en que las restricciones dinámicas, tales como el balance geostrófico e hidrostático, son incluidas cuando la función de costo es minimizada.

El WRFDA asimila observaciones en formato *American Standard Code for Information Interchange* (ASCII), aunque cuenta con una herramienta, llamada *Observation Processor* (OBSPROC), capaz de convertir otros formatos de datos a ASCII. También asimila datos de radiancia de satélite en formato *Binary Universal Form* (BUFR), el cual es muy empleado en aquellos de origen meteorológico, así como datos de archivos *Prepared BUFR data* (PREPBUFR) de distintas fuentes (Barker et al., 2012). Más detalles sobre los datos BUFR y PREPBUFR se muestran en el Apéndice D.

El diagrama de flujo que explica el proceso de asimilación se presenta en la Figura 1, en la cual x^b es el background y x^{lbc} son las condiciones de frontera de x^b ; ambos campos se construyen con el Sistema de Preprocesamiento o *Preprocessing System* (WPS) del modelo WRF y el ejecutable encargado de generar dichas condiciones (`real.exe`); x^a es el análisis; x^f es el pronóstico obtenido con WRF; y^o son las observaciones; B_0 es una matriz que representa los errores estadísticos del background; mientras que R es la matriz de covarianza de los errores referidos a las observaciones.

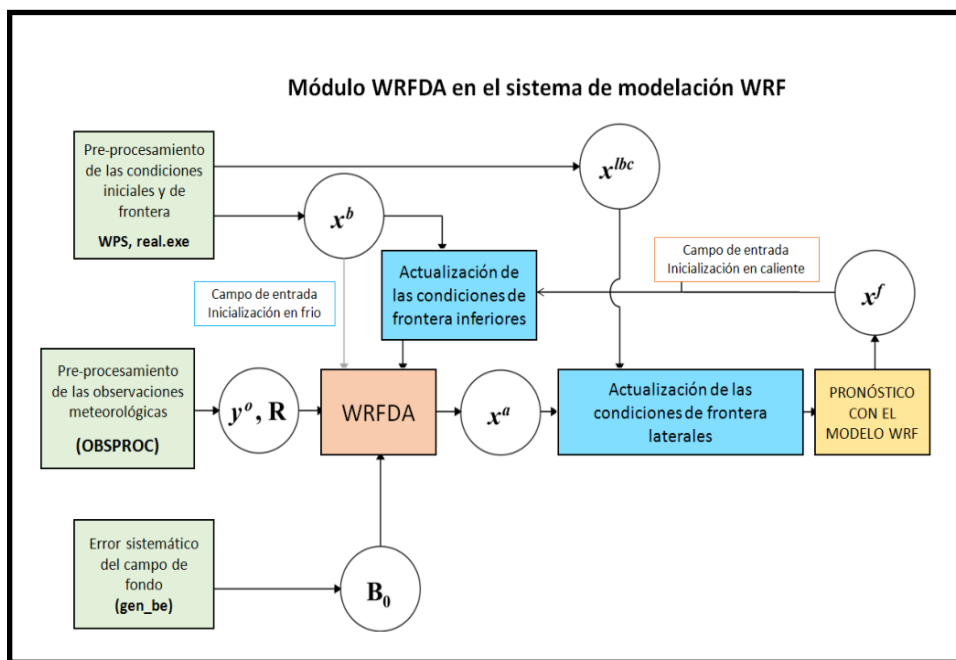


Figura 1. Diagrama de flujo del proceso de asimilación en WRFDA (Adaptado del manual de usuario del módulo)

En la figura, se puede apreciar, que el módulo WRFDA permite la asimilación de tipo *cold start* (inicialización en frío) y *warm start* (inicialización en caliente), es decir, obtener la condición inicial x^b utilizando, respectivamente, *WPS/real.exe* o una salida del modelo propiamente. Más detalles del módulo en su manual de usuario (https://www2.mmm.ucar.edu/wrf/users/docs/user_guide_V3.9/users_guide_chap6.htm).

1.3.2 Data Assimilation Research Testbed (DART)

DART es un entorno de software desarrollado por la *Data Assimilation Research Section* (DAReS) del *National Center for Atmospheric Research* (NCAR), que permite la exploración de una variedad de algoritmos de asimilación de datos con diferentes fuentes observacionales y modelos numéricos. Emplea un enfoque de programación modular para aplicar el método de asimilación EnKF, lo cual facilita que los modelos subyacentes converjan a un estado que es más consistente cuando se mezclan con la información de las observaciones. Estos modelos pueden intercambiarse dentro y fuera del módulo, al igual que los diferentes algoritmos correspondientes a EnKF, el cual, de forma general, requiere ejecutar varias instancias de un modelo para generar un conjunto de estados. Un operador directo, apropiado para el tipo de observación a asimilar, se aplica a cada uno de los estados, para generar la estimación de dicha observación hacia el modelo.

Otra ventaja de este módulo radica en que sus algoritmos están diseñados de modo que permiten la incorporación de nuevos modelos y diferentes tipos de observaciones empleando la codificación mínima de un pequeño conjunto de rutinas de interfaz, sin modificar el código existente. Además, se pueden crear operadores directos para nuevos tipos de observación, de una manera que es casi independiente del modelo de pronóstico, considerando que muchos de los operadores estándar están disponibles "fuera de la caja" y funcionarán sin codificación adicional.

El modelo WRF-Chem es uno de los que ya tiene añadido el DART (ver detalles de su implementación en el Apéndice E). El WRF-Chem/DART es un sistema regional multivariante de asimilación de datos meteorológicos y químicos, de forma independiente o simultánea, desarrollado por NCAR, pero aún no disponible a la comunidad científica internacional, en su capacidad de asimilar datos químicos (Mizzi et al., 2016). Dentro de los métodos EnKF existentes, WRF-Chem/DART utiliza el EAKF. Detalles del algoritmo e implementación de EAKF en DART se documentan en Anderson (2001), Anderson y Collins (2007) y Anderson et al. (2009). Además, en la Sección 2.4.2 de esta tesis, se aborda la temática relacionada con WRF-Chem/DART y su implementación.

El sistema WRF-Chem/DART suele ejecutarse mediante múltiples ciclos de pronóstico y análisis. Este procedimiento es esencial para obtener los beneficios de la técnica utilizada por el módulo, pero mucho más complejo y costoso, desde el punto de vista computacional, que ejecutar un módulo de asimilación del tipo WRFDA y GSI, cuya base son las técnicas variacionales. La covarianza de los errores de medición se estima a partir del ensamble creado por los pronósticos y, además, las desviaciones de cada miembro del conjunto, sobre la media, suelen ser representativas de los errores estadísticos del pronóstico a corto plazo, en cada instante (Mizzi et al., 2016).

Por todo lo anterior, el ensamble sí refleja los efectos de las observaciones recientes y la evolución de los errores de pronóstico bajo la dinámica del modelo, lo cual constituye una de las deficiencias de otros métodos existentes. Para el pronóstico, unos 3-4 días de asimilación, en ciclos de 6 horas, suelen ser suficientes (Mizzi et al., 2016).

En función de la configuración utilizada, en la asimilación de datos meteorológicos las variables perturbadas son: temperatura, humedad, componente zonal y meridional del viento y presión superficial. La covarianza de los errores estadísticos se estima mediante el método NMC (*National Meteorological Center*), en el cual se calculan las diferencias entre dos salidas de pronóstico del GFS para dos períodos de 24 y 48h respectivamente (Parrish y Derber, 1992). La condición de frontera para el tiempo de análisis es ajustada para que coincida con el análisis de DART.

Para las emisiones se utiliza un esquema de actualización, o sea, se incluyen en el vector de estado del ensamble mediante una estimación como parámetros evolutivos por hora, enfoque que proporciona información sobre las emisiones más un promedio para un período de tiempo específico. Un desafío en este sentido lo constituye la ausencia de un modelo de pronóstico para evolucionar las variables de emisión hacia adelante en el tiempo. Una alternativa consiste en actualizar los factores de escala de emisión en cada ciclo de asimilación, en lugar de tratar las emisiones, en cada hora, en un lugar específico de la malla, como proporcionan los inventarios de emisiones (Mizzi et al., 2016).

Según el flujo de procesos en WRF-Chem/DART (Figura 2), para realizar la asimilación se incluyen condiciones iniciales y de frontera meteorológicas y químicas y emisiones.

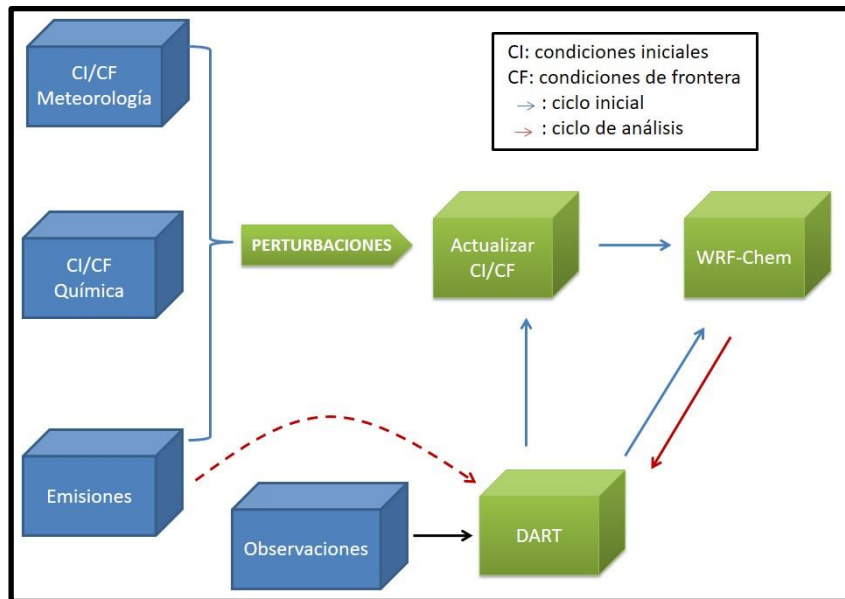


Figura 2. Esquema para WRF-Chem/DART (Adaptado de Mizzi, 2020)

Dentro de las fuentes de datos que pueden asimilarse con WRFChem/DART están: archivos PREPBUFR de NCEP, salidas del WRFDA en formato OBSPROC y mediciones de CO del sensor *Measurement of Pollution in the Troposphere* (MOPITT), entre otras. Sin embargo, en la bibliografía solo se reporta el trabajo con PREPBUFR (Mizzi et al., 2016; Mizzi et al., 2018 y Ma et al., 2019a).

1.3.3 Gridpoint Statistical Interpolation (GSI)

El GSI es un sistema de asimilación de datos variacional, flexible, avanzado y eficiente para varias plataformas de computación paralelas, de dominio público y disponible gratuitamente para su uso. Desarrollado de forma continua por el *Developmental Testbed Centre* (DTC), GSI se centra en aplicaciones regionales de predicción numérica del tiempo junto con los modelos WRF y GFS. Puede emplearse como un sistema de asimilación de datos híbridos conjunto-variacional, combinado con un sistema de conjunto, como ocurre en el Sistema Global de Asimilación de Datos (GDAS) de NCEP.

Junto con los modelos de predicción y adjuntos, GSI puede convertirse en un sistema variacional de cuatro dimensiones, ofreciendo capacidades para calcular la sensibilidad de observación en estudios de impacto operacional de datos. Los operadores de observación en GSI pueden utilizarse en un sistema EnKF u otros de análisis de datos, transformando variables modelo a otras observadas en el espacio observacional. (Ver el manual de usuario en: https://dtcenter.org/sites/default/files/community-code/gsi/docs/usersguide/GSIUserGuide_v3.5.pdf)

1.4 Asimilación de datos de variables meteorológicas y químicas en WRF-Chem

Actualmente, los principales módulos para la asimilación de datos son WRFDA, DART y GSI. En cuanto a la asimilación de datos meteorológicos, WRFDA y GSI se han utilizado ampliamente como parte de los modelos de pronóstico del tiempo, tanto para el trabajo operativo como para la investigación. Sin embargo, para el caso de los modelos de calidad del aire, la asimilación de datos puramente meteorológicos es un tema muy poco abordado en la literatura científica (Martínez y García, 2020).

En cuanto a la asimilación de datos químicos, el módulo más avanzado es el DART, aunque existen dificultades o retos en la modelación numérica de estas variables que limitan su desarrollo. Las observaciones de especies químicas, realizadas mayormente en superficie y con fines regulatorios, suelen estar muy dispersas mientras que las de la atmósfera superior, obtenidas principalmente por satélite de forma indirecta, presentan mucho volumen para una baja densidad de área, requiriendo de mayor recurso computacional para su procesamiento. La interacción de la meteorología con las observaciones químicas en las variables de estado, durante el proceso de asimilación, también puede afectar el comportamiento de las variables químicas.

De igual modo, aunque se conoce la existencia de una gran sensibilidad en los resultados de los pronósticos numéricos a la calidad de las emisiones, no se tiene una medida real de su incertidumbre y ni de los errores de covarianza asociados. Un elemento adicional que debe considerarse es el conjunto de transformaciones químicas que ocurren en la atmósfera en escalas de tiempo variables, desde horas hasta meses.

Los módulos WRFDA, DART y GSI pueden ser utilizados para la asimilación de datos de variables meteorológicas para inicializar el modelo WRF-Chem, con características propias que los diferencian entre sí (Tabla II). El WRFDA tiene implementado un mayor número de métodos de asimilación, incluyendo técnicas variacionales y ensambles. Además, este es el módulo que forma parte, precisamente, del modelo WRF.

Tabla II. Características de las herramientas de asimilación de datos meteorológicos disponibles para el modelo WRF-Chem

Sistema	WRF-Chem/WRFDA	WRF-Chem/GSI	WRF-Chem/DART
Origen de los datos	Observaciones en superficie, satélites, radares y GPS; sistemas de datos BUFR y PREPBUFR	Observaciones en superficie, satélites, radares y GPS; sistemas de datos BUFR y PREPBUFR	Sistema de datos PREPBUFR.
Formatos de los datos	ASCII, BUFR y PREPBUFR	BUFR y PREPBUFR	PREPBUFR y obsequence_file
Métodos de asimilación	3DVAR, 4DVAR, EnFK, híbridos	3DVAR, 4DVAR, híbridos	EAKF

En el caso de la asimilación de datos químicos para WRF-Chem, los módulos WRFDA y GSI no tienen incorporados, propiamente, dicha funcionalidad. Sin embargo, en algunos estudios reportados, se han desarrollado opciones para permitirlo, describiéndose además las características de los mismos (Tabla III).

Tabla III. Características de las herramientas de asimilación de datos meteorológicos disponibles para el modelo WRF-Chem

Sistema	WRF-Chem/WRFDA	WRF-Chem/GSI	WRF-Chem/DART
Variables	Observaciones en superficie y aire superior de O ₃ , CO y CH ₄	Observaciones en superficie de O ₃ y partículas; AOD de MODIS	Observaciones en superficie; perfiles de CO de MOPITT e IASI; columna total de NO ₂ de OMI; AOD de MODIS; columna total o parcial de O ₃ de IASI.
Métodos de asimilación	3DVAR y 4DVAR	3DVAR	EAKF

1.5 Antecedentes

El uso de la asimilación de datos en la química de la atmósfera es muy reciente, partiendo del hecho que los modelos numéricos de química atmosférica, se utilizan de forma más frecuente para el pronóstico de calidad del aire, desde mediados de la década de 1990.

Elbern y Schmidt (2001) es uno de los estudios pioneros de reanálisis y confección de mapas de concentración de contaminantes. Analizaron el estado químico de un episodio de contaminación atmosférica por O₃, empleando el modelo EURAD CTM, con asimilación de datos de observaciones en superficie y radiosondeos. Sus resultados mostraron que la optimización de condiciones iniciales ofrecía una mejora significativa para el pronóstico de O₃ de 6 a 12 h. Conclusiones similares se obtuvieron en otras investigaciones (Wu et al., 2008; Tombette et al., 2009; Wang et al., 2011; Curier et al., 2012), mientras que Chai et al. (2006) lo hicieron para NO, NO₂ y peroxiacetilnitrato (PAN), empleando el modelo STEM-2K1. De igual forma, la 2008/50 Dirección de Calidad del Aire de la Unión Europea sugirió el uso de modelos en combinación con mediciones, con el objetivo de "proporcionar información adecuada sobre la distribución espacial de la calidad del aire" (Borrego et al., 2015; OJEU, 2008).

Enfoques de asimilación de datos se han utilizado también para combinar las mediciones y los resultados de modelos en el contexto de las evaluaciones de la calidad del aire (Candiani et al., 2013), mejorar inventarios de emisiones (Mijling y van der A, 2012; Koohkan et al., 2013) y establecer de forma óptima redes de monitoreo (Rayner, 2004; Wu et al., 2010; Wu y Bocquet, 2011; Lauvaux et. al, 2012). Además, en la modelación inversa para mejorar o identificar errores en las tasas de emisión (Elbern et al., 2007; Vira y Sofiev, 2012; Yumimoto et al., 2012), para las condiciones de frontera (Roustan y Bocquet, 2006), en el ajuste de los parámetros del modelo (Barbu et al., 2009; Bocquet, 2012) y para asimilación de datos de variables químicas procedentes de satélite (Wang et al., 2011), entre otros. (Bocquet et. al, 2015)

En cuanto a la asimilación de datos en modelos acoplados, este es un tema mucho más reciente aún y de menor desarrollo. Sin embargo, su empleo ha venido creciendo en los últimos años, debido a que estos modelos ofrecen la posibilidad de asimilar datos meteorológicos y químicos de forma simultánea. Para ello, emplean una gran variedad de técnicas como las ensambles, las cuales han demostrado una alta convergencia a la solución real, a pesar de ser bien complejas y con un alto costo computacional (Zhang, 2008; Baklanov et al., 2014; Bocquet et al. 2015).

En este sentido, varias son las investigaciones que emplean el modelo WRF-Chem. Pagowski et al. (2010) utilizaron GSI y 3DVAR para asimilar observaciones en superficie de O₃ y PM_{2.5} en Estados Unidos. Liu et al. (2011) asimilaron AOD de MODIS para simular un episodio de tormenta de polvo ocurrido en Asia en 2010, mientras que Chen et al. (2014) utilizaron un enfoque similar para mejorar las simulaciones de PM_{2.5} y carbón orgánico en superficie durante un evento de incendio por quema de biomasa en Estados Unidos. Pagowski y Grell (2012) compararon las técnicas 3DVAR y EnFK durante la asimilación de concentraciones en superficie de PM_{2.5}.

Schwartz et al. (2012) también emplearon GSI y 3DVAR para asimilar AOD de MODIS y observaciones en superficie de PM_{2.5} en Estados Unidos, mientras que Schwartz et al. (2014) asimilaron los mismos datos, pero utilizando *Ensemble Square Root Filter* (EnSRF) y un método híbrido de ensamble con 3DVAR. Jiang et al. (2013) también usaron GSI y 3DVAR para asimilar observaciones en superficie de PM₁₀ en China.

Peng et al. (2017) asimilaron concentraciones en superficie de $PM_{2.5}$ con EnKF, mientras que Peng et al. (2018) emplearon EnKF para asimilar observaciones en superficie de varias especies químicas, ambos para mejorar el pronóstico de calidad del aire en China.

Los trabajos mencionados realizan asimilación de datos solo de variables químicas. La meteorología juega un papel crucial en la modelación de la calidad del aire, al definir el entorno físico y dinámico de la química atmosférica. Tiene una fuerte influencia en la emisión y transformación de especies químicas, aerosoles o partículas en la atmósfera. Además, las reacciones químicas que intervienen en los procesos de formación de especies secundarias están afectadas directamente por factores meteorológicos como la radiación solar, la humedad relativa, la temperatura, el viento y la cobertura nubosa.

Por ello, la incertidumbre en los datos meteorológicos utilizados en la simulación de la calidad del aire, puede afectar negativamente a los resultados del modelo. En este sentido, a medida que continúa el rápido crecimiento de la tecnología informática y aumenta la disponibilidad de instrumentos de teledetección más modernos y avanzados, es posible la recopilación de un mayor volumen de datos meteorológicos para ser asimilados, de manera independiente o simultánea con los químicos, para contribuir a mejorar el desempeño de los modelos de calidad del aire (Seaman 2000; Seaman 2003).

En los últimos años, se han llevado a cabo diferentes investigaciones conducidas por NCAR, para desarrollar la capacidad del módulo WRFDA de asimilar datos de variables químicas y meteorológicas simultáneamente. Guerrette y Henze (2015) desarrollaron un nuevo módulo basado en WRFDA y utilizando la técnica 4DVAR, el cual fue probado en un estudio de sensibilidad para un evento de incendio por quema de biomasa ocurrido en California, Estados Unidos. El principal problema relacionado con esta herramienta radica en que aún no se encuentra incluida en la versión de WRFDA que está disponible para la comunidad científica internacional (Comunicación personal con los autores).

Eltahan y Alahmadi (2019) mostraron el impacto de asimilar datos meteorológicos con WRFDA y los algoritmos 3DVAR y 4DVAR, en la simulación numérica de AOD durante una tormenta de polvo ocurrida en Egipto. Encontraron que la asimilación de observaciones de velocidad del viento mejoró el desempeño del modelo.

De forma paralela, se han desarrollado otras investigaciones en las cuales se utiliza el módulo de asimilación de datos DART, introducido al modelo WRF-Chem por Mizzi et al. (2016), en conjunto con la asimilación de *compact phase space retrievals (CPSRs)* que se derivan de los perfiles de CO obtenidos de MOPITTT e IASI. Liu et al. (2017) realizaron asimilación de datos de toda la columna de NO₂ procedente de OMI, con el objetivo de investigar su potencial para el ajuste de inventarios de emisiones.

Mizzi et al. (2017) aplicaron por primera vez el sistema WRF-Chem/DART al pronóstico de calidad de aire durante la campaña FRAPPE de NCAR. Esta temática se continúa trabajando de forma paralela en colaboración con el Instituto de Meteorología Max Plank en Alemania (Dr. Brasseur y Dr. Bouarar), la Universidad Nanjung en China (Dr. Wang y Dr. Ma) y el Grupo de Fisicoquímica de la Atmósfera del Centro de Ciencias de la Atmósfera de la Universidad Nacional Autónoma de México (autores de la presente Tesis de Investigación Doctoral). Algunos resultados de este trabajo conjunto se presentan en Ma et al. (2019a) y Ma et al. (2019b).

Mizzi et al. (2018) demostraron que la asimilación de datos de CO de MOPITT con WRF-Chem/DART, permite obtener un mejor desempeño del modelo respecto al empleo de observaciones convencionales, tanto para análisis como para pronóstico.

En el caso de México, son muy pocos los estudios reportados en la literatura científica sobre asimilación de datos. Lobato (2003) realizó de forma experimental la asimilación de datos meteorológicos de la Red Estaciones Meteorológicas Automáticas del Servicio Meteorológico Nacional (EMA), para inicializar el modelo mesoescalar NCAR/Penn State (MM5). Para ello, empleó el módulo LITTLER, incluido entre los subprogramas de dicho modelo. Los resultados obtenidos muestran que los pronósticos para 24 y 48 horas son más cercanos a las observaciones al agregarse información que permite refinar las condiciones iniciales, capturando fenómenos de escala menor a la sinóptica.

Así mismo, Zitácuaro y Caetano (2004) realizaron simulaciones del campo de viento para el Valle de México, también con el modelo MM5, asimilando datos meteorológicos de la Red Automática de Monitoreo Atmosférico de la Ciudad de México (RAMA) y EMA. Los resultados mostraron una mejor representación de velocidad y dirección del viento, considerando las características impuestas por las montañas alrededor del valle y se obtuvieron los patrones de viento a escala local con una alta resolución.

Específicamente, en el tema de la química atmosférica, Bei et al. (2008) usaron 3DVAR para asimilar los campos meteorológicos de viento, humedad y temperatura, con el objetivo de mejorar las simulaciones numéricas de O₃ en Ciudad de México. Estas fueron realizadas con un sistema de modelación compuesto por el MM5 y el *Air Quality Model with extensión* (CAMx). Los mejores resultados se obtuvieron en períodos matutinos, sobre todo en los picos de O₃ y la posición de la pluma.

Basados en la investigación anterior, Bei et al. (2010) asimilaron datos meteorológicos durante la campaña MCMA-2006/MILAGRO (Molina et al., 2010) con similares objetivos, pero usando un pronóstico ensamble obtenido de condiciones iniciales meteorológicas generadas con la técnica WRF-3DVAR (Barker et al., 2004). Posteriormente, siguiendo la misma idea, Bei et al. (2012) investigaron la incertidumbre en la simulación numérica de aerosoles orgánicos secundarios en la Zona Metropolitana del Valle de México (ZMVM), utilizando el modelo WRF-Chem y simulaciones ensamble. Además, Bei et al. (2014) evaluaron los impactos del uso de EnKF en la simulación de la calidad del aire en la región fronteriza entre México y California, Estados Unidos, durante la campaña Cal-Mex de 2010 y lo compararon con las técnicas de ensamble y *nudging*.

CAPÍTULO 2: MATERIALES Y MÉTODOS

2.1 Modelo WRF-Chem: descripción y configuración general

El *Weather Research and Forecasting Model with Chemistry* (WRF-Chem, Grell et al. 2005; Fast et al. 2006) es uno de los sistemas de modelación numérica de la atmósfera más utilizados a nivel mundial. Sus aplicaciones en distintos campos permiten explicar episodios pasados, evaluar los potenciales efectos de las estrategias de reducción de emisiones o realizar pronósticos de calidad del aire a escala regional y global. En esta investigación, se utiliza la versión 3.8.1 para realizar los experimentos numéricos.

WRF-Chem es desarrollado por *National Oceanic and Atmospheric Administration Earth System Research Laboratory* (NOAA/ESRL) y NCAR. Tiene estructura modular que permite considerar variedad de procesos fisicoquímicos, como son: transporte, emisión, deposición, mezcla, transformación química, interacción con los aerosoles, procesos de fotólisis, transferencia radiativa, entre otros, de manera simultánea con la meteorología, es decir, ambos componentes se encuentran totalmente acoplados y utilizan el mismo esquema de transporte, igual malla numérica, las misma resolución espacial y temporal e idénticos esquemas físicos. WRF-Chem posee opciones para la discretización espacial, la difusión, el anidamiento, las condiciones iniciales y de frontera, las parametrizaciones físicas, los esquemas de uso de suelos y vegetación, las emisiones biogénicas y antropogénicas, las emisiones volcánicas, los mecanismos químicos en la fase gaseosa, entre otras (Peckham et al., 2017).

Los módulos más importantes del modelo son: el Sistema de Preprocesamiento o *Preprocessing System* (WPS), el módulo de asimilación de datos o *Data Assimilation* (WRFDA), el módulo que resuelve numéricamente las ecuaciones que modelan la atmósfera (ARW-solver), además del módulo de Postprocesamiento y Visualización.

El módulo WPS tiene como objetivo principal definir los dominios de modelación e interpolar a estos, los datos orográficos y de uso de suelo, así como los meteorológicos procedentes de otros modelos. Incluye la capacidad de anidamiento y posee una interfaz de usuario para incorporar datos estáticos, entre otros.

Los archivos que se generan en WPS junto con el inventario de emisiones son los datos de entrada para el módulo ARW-solver, en el cual se acopla la química atmosférica y se aplican los mecanismos de reacción para simular la formación de contaminantes y su transporte. Las salidas que se obtienen son, precisamente, las concentraciones de los contaminantes. El módulo final se refiere al post procesamiento de las salidas y cuenta con una variedad de visualizadores en dos y tres dimensiones.

Existen otras herramientas de análisis estadístico complementarias, las cuales no forman parte del WRF-Chem, pero son muy utilizadas, como el *Model Evaluation Tools* (MET), desarrollado por el *Development Testbed Center* (DTC). El MET permite comparar las salidas del modelo con observaciones de estaciones de monitoreo, empleando para ello dos procesos: *Point_stat_analysis*, en el cual se extraen por interpolación bilineal los datos del modelo con base en la ubicación de las observaciones y *Stat_analysis*, donde se calculan las métricas estadísticas (García y Mora, 2017).

En la Tabla IV, se muestran las principales parametrizaciones físicas y químicas que fueron seleccionadas para configurar el modelo WRF-Chem, con una breve descripción general de las mismas. Estas se utilizaron en todos los casos de estudio desarrollados en la investigación. Se activó un mecanismo que posee el modelo (*aer_ra_feedback* = 1) para considerar las interacciones aerosoles-radiación con un efecto de tipo directo de las emisiones en la meteorología. Las emisiones de COV generan aerosoles orgánicos secundarios, los cuales influyen en la radiación solar que llega a la superficie, modificando la temperatura superficial y otras variables meteorológicas calculadas por el modelo. Además, se incluye una opción (*cu_rad_feedback* = .true) para que los esquemas de onda corta y fotólisis consideren los efectos de nube que no hayan sido resueltos en la simulación (Forkel et al. 2012).

Tabla IV. Parametrizaciones físicas y químicas empleadas en la configuración del modelo WRF-Chem

Proceso	Referencia	Descripción
Capa límite planetaria	Yonsei University YSU (Hong y Dudhia, 2003)	Es un esquema no local de primer orden que representa los flujos verticales pequeños como resultado del transporte turbulento en toda la columna atmosférica
Capa superficial	MM5 Similarity (Zhang y Anthes, 1982)	Emplea funciones de estabilidad para calcular los coeficientes de intercambio superficial para calor, humedad y momento
Microfísica	WRF Single-Moment 5-Class WSM5 (Hong et al., 2004)	Es un esquema de masa que pronostica la razón de mezcla de vapor de agua, lluvia, nube, hielo y nieve y permite la formación de nubes en fase mixta
Uso de suelos	Noah Land Surface Modelling LSM (Chen y Dudhia, 2001)	Es un modelo de humedad y temperatura del suelo de cuatro capas que calcula las reservas y flujos de humedad
Radiación de onda larga	Rapid Radiative Transfer Model RRTM (Mlawer et al., 1997)	Calcula los flujos y las tasas de calentamiento de manera eficiente y precisa, con un enfoque K-correlacionado
Radiación de onda corta	MM5 (Dudhia, 1989)	Es un esquema simple de banda ancha y descendente de flujo solar el cual emplea tablas de búsqueda para las propiedades de la nube
Cúmulos	Kain-Fritsch (Kain y Fritsch, 1993)	Representa los procesos convectivos asumiendo que el flujo de masa en la base de la nube está determinado por la cantidad de energía potencial convectiva disponible en el medio ambiente que debe eliminarse

2.2 Área de estudio y dominios de modelación

La región conocida como Centro de México es el área de estudio de esta investigación e incluye los siguientes estados de la megalópolis de la zona central mexicana: Ciudad de México, Estado de México, Morelos, Hidalgo, Tlaxcala y Puebla (Figura 3).

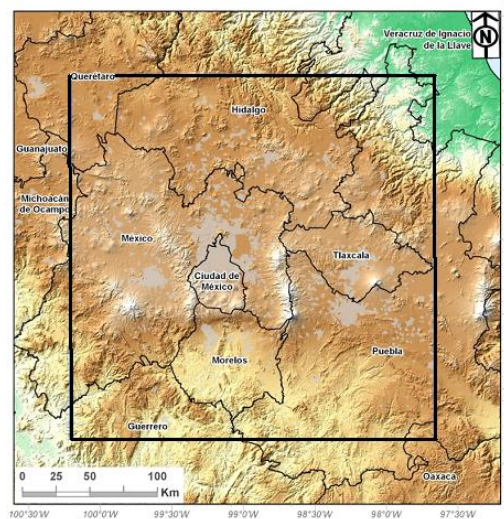


Figura 3. Área de estudio

En las últimas décadas, el crecimiento poblacional y el desarrollo económico experimentado por la megalópolis, considerada la región urbana más poblada del país, de Latinoamérica y entre las más grande del mundo, ha ocasionado gran impacto en la calidad del aire del país. La pérdida de las zonas agrícolas dada la expansión urbana, las emisiones de contaminantes antropogénicos como los solventes de pinturas y los gases de combustión en los automóviles, entre otros, han contribuido a aumentar, de forma significativa, la contaminación atmosférica en la región (CAME, 2018)

La Zona Metropolitana del Valle de México (ZMVM) constituye el núcleo central del área de estudio y la región de mayor interés para la presente investigación, debido a que se considera la más contaminada del país. Se localiza en el interior del Valle de México y está rodeada por elevaciones que inducen la acumulación de los contaminantes sobre ella, provocando serias afectaciones a la salud humana. Su centro geográfico se ubica en $19^{\circ} 30'$ de latitud Norte y $99^{\circ} 02'$ de longitud Oeste (SEMARNAT, 2016).

La fuente que más contribuye a la contaminación ambiental en la ZMVM es el transporte, emitiendo la mayor parte de los precursores de O₃ y partículas a la atmósfera. Por su parte, la actividad volcánica y los incendios forestales, también incorporan sustancias contaminantes de origen natural al aire, que pueden ser transportadas por la acción del viento (INECC, 2018).

Las propias condiciones geográficas y climatológicas de la ZMVM contribuyen a sus problemas ambientales. La menor cantidad de oxígeno producto de su altitud de más de 2,240 m sobre el nivel del mar, las características de su orografía y la elevada incidencia de la radiación solar, que favorece la ocurrencia de reacciones químicas de formación de O₃ en la atmósfera, son elementos a considerar. El patrón de vientos predominante agudiza las inversiones térmicas, principalmente en los meses invernales de diciembre a febrero y, además, transporta los contaminantes hacia diferentes regiones de la ZMVM y otras zonas colindantes, provocando afectaciones a la calidad del aire (INECC, 2018).

Diferentes estudios han demostrado también la interacción que existe entre las cuencas del Centro de México, así como los patrones de viento y ventilación que se presentan cuando se tiene altos niveles de O₃ (García et al., 2009; Ruiz et al., 2010; Salcedo et al., 2012). Se han identificado los sitios donde ocurre el intercambio y algunas de las condiciones meteorológicas que los favorece (Doran y Zhong, 2000; Jazcilevich et al., 2003; de Foy et al., 2008)

En el caso de la ZMVM, esta exporta parcelas de aire hacia el norte (Estado de México e Hidalgo) entre enero y marzo, hacia el estado de Morelos en mayo y de julio a diciembre, hacia el este del Estado de México (Valle de Toluca) de septiembre a diciembre, mientras que en mayo permanecen dentro de la ZMVM y áreas aledañas. A su vez, la ZMVM recibe parcelas de aire desde las cuencas de Toluca (septiembre a diciembre), Puebla-Tlaxcala (junio a septiembre) y Cuernavaca-Cuautla (INECC, 2016). Particularmente, se ha podido evaluar el impacto que tienen las emisiones del volcán Popocatepetl en la calidad del aire de la ZMVM, así como el transporte de la pluma de contaminantes proveniente de la región de Tula, Hidalgo (de Foy et al., 2009; Rivera et al., 2009; Sosa et al., 2013; Almanza et al., 2012, García et al., 2014), entre otros.

Por todo lo anterior, el desarrollo continuo de herramientas para estudiar la calidad del aire en el Centro de México y específicamente en la ZMVM, es una tarea priorizada en el país. Especial interés existe en la implementación y optimización de los modelos de química atmosférica, con el objetivo de que sus salidas numéricas tengan una mayor correspondencia con las observaciones de las estaciones disponibles. La presente investigación, precisamente, brinda un aporte científico a este propósito

Para realizar las simulaciones numéricas con WRF-Chem en el área de estudio, se construyeron dos dominios de modelación (Figura 5), partiendo del criterio de la disponibilidad y la calidad existentes en los datos de las emisiones de los contaminantes en el inventario utilizado, así como en las estaciones de monitoreo atmosférico. Ambos dominios se anidaron de modo *two-way nesting* y se definieron como sigue: uno mayor (D-1) donde se ubicó una malla de 90 x 90 celdas con resolución de 9 km y otro menor (D-2) que cubre la mayor parte del Centro de México, en el cual se empleó una malla de 88 x 88 celdas con resolución de 3 km. Verticalmente se consideran 29 niveles.

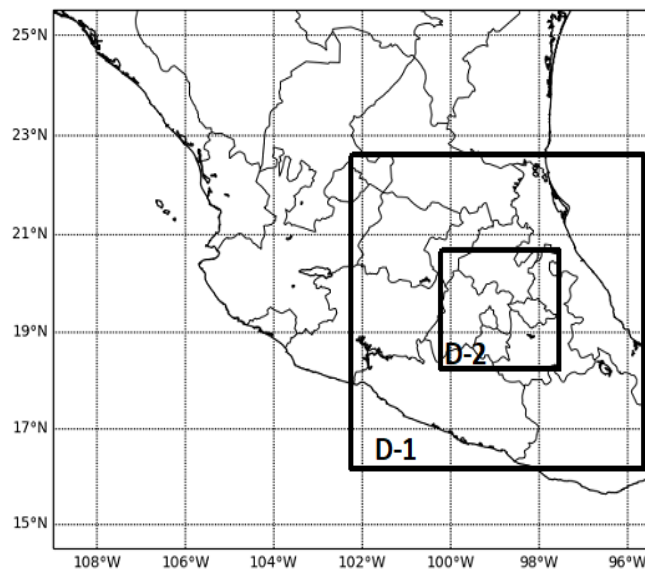


Figura 4. Dominios de modelación

2.3 Descripción de los experimentos numéricos

2.3.1 WRF-Chem/WRFDA

Para llevar a cabo el proceso de asimilación de datos meteorológicos con el módulo WRFDA, se seleccionó un período de estudio comprendido entre el 1 y el 4 de mayo de 2013 (2013-05-01 al 2013-05-04), en el cual se activó el Programa de Contingencia Ambiental y Atmosférico (PCAA). El día jueves, 2 de mayo (2013-05-02), a las 17 horas, se inició un episodio de mala calidad del aire por O₃ en la Ciudad de México, extendiéndose hasta el siguiente día 3 (2013-05-03), a las 20 horas, siempre en tiempo local. En las estaciones de monitoreo ubicadas al sudeste de la zona, se reportaron valores anómalos de concentraciones de O₃, alcanzándose un máximo del Índice Metropolitano de Calidad del Aire (IMECA) de 157 puntos.

Desde el punto de vista climático, en la ZMVM, el mes de mayo es el que presenta los registros más altos de temperatura en todo el año. En el periodo de estudio, los valores reportados por las estaciones meteorológicas, tanto mínimos como máximos, fueron incluso superiores al promedio, debido a la circulación de un sistema de alta presión ubicado en el Océano Atlántico, el cual provocó, además, la ocurrencia de un marasmo barométrico a causa de sus gradientes muy débiles. Los vientos se mantuvieron en calma o con bajas velocidades en los días considerados, lo cual no favoreció a la dispersión de los contaminantes (SMN, 2013).

Para realizar los experimentos numéricos, se determinaron seis casos de estudio, teniendo en cuenta la combinación de dos elementos: la fuente de datos a asimilar y el horario de inicio de las simulaciones numéricas. Por una parte, se definieron tres opciones: sin asimilación (SA), con asimilación de archivos de tipo PREPBUFR (PB) y con asimilación de datos PB más radiancia de satélites en formato BUFR (PB+RD); mientras que, por otro lado, fueron considerados dos horarios de inicio: 1 de mayo (2013-05-01) a las 00 UTC (19 horas tiempo local) y 1 de mayo (2013-05-01) a las 12 UTC (7 horas tiempo local), debido a que son horarios sinópticos y hay un mayor volumen de datos disponibles. De acuerdo a lo anterior, los casos de estudio se definieron como se muestra en la Tabla V.

Tabla V. Casos de estudio determinados para la asimilación de datos con WRFDA

Elementos	00 UTC	12 UTC
SA	00 UTC_SA	12 UTC_SA
PB	00 UTC_PB	12 UTC_PB
(PB+RD)	00 UTC_(PB+RD)	12 UTC_(PB+RD)

Los datos procedentes del *North American Regional Reanalysis* (NARR), desarrollado por el *National Centers for Environmental Prediction* (NCEP), fueron empleados para generar las condiciones iniciales y de frontera de tipo meteorológico, necesarias para las simulaciones numéricas con WRF-Chem. NARR es una base de datos atmosféricos y de hidrología de la superficie terrestre, dinámicamente consistente, de alta resolución y alta frecuencia. Su resolución temporal es de 3 h y espacial de 32 km, con 29 niveles de presión, cubriendo la región de América del Norte (Mesinger et al., 2006).

Por su parte, se utilizaron las condiciones iniciales y de frontera químicas que están predeterminadas y codificadas en el modelo WRF-Chem, las cuales no varían en el tiempo y están basadas en un perfil vertical idealizado para el hemisferio norte y latitudes medias, con un ambiente limpio, elaborado por el *NOAA Aeronomy Lab Regional Oxidant Model* (NALROM) (Peckham et al., 2017).

Las emisiones antropogénicas se obtuvieron del Inventario Nacional de Emisiones de México para el año 2008 (INEM-2008) elaborado por la Secretaría de Medio Ambiente y Recursos Naturales (SEMARNAT). El INEM-2008 cuenta con la clasificación de las emisiones de todas las fuentes fijas, móviles y de área, en sus diferentes categorías e incluye las emisiones anuales por municipio de los contaminantes atmosféricos SO₂, CO, NO, NO₂, PM₁₀, PM_{2.5}, compuestos orgánicos volátiles (VOCs), amoníaco (NH₃), entre otros (García et al., 2018).

Para tener una base de datos de emisiones útil para la modelación con WRF-Chem a partir del INEM-2008, se le aplicó la metodología correspondiente al modelo de distribución espacial, temporal y de especiación química (DiETE). En este caso, se desagrega el inventario, teniendo en cuenta que la distribución espacial de las emisiones en cada celda del dominio de modelación, depende de los datos de población, la cobertura de suelo y la ubicación de industrias (García et al., 2018).

El mecanismo empleado para la química en fase gaseosa fue el *Regional Acid Deposition Model versión 2* (RADM2, Stockwell et al. 1990), el cual determina la distribución de las emisiones de las especies químicas en función de la reactividad con el radical hidroxilo.

La técnica de asimilación de datos utilizada en el sistema WRF-Chem/WRFDA, para todos los casos definidos, es 3DVAR.

Posteriormente, a partir de los resultados obtenidos, se definieron otros casos de estudio, pero esta vez comenzando la asimilación de datos siempre a las 12 UTC y para la fuente de datos (PB+RD), utilizando los métodos de asimilación disponibles en la versión 3.8.1 de WRFDA, empleada en esta investigación: 3DVAR, 4DVAR y En3DVAR.

Se mantuvo el mismo período de tiempo seleccionado anteriormente (1-4 de mayo de 2013, es decir, del 2013-05-01 al 2013-05-04) y se añadió otro, comprendido entre el 4 y 7 de mayo de 2013 (del 2013-05-04 al 2013-05-07), en el cual no hubo contingencia ambiental en la ZMVM. El objetivo principal de estos experimentos fue observar la influencia de las diferentes técnicas mencionadas en el proceso de asimilación de datos, así como en la modelación con WRF-Chem en situaciones reales con condiciones de química atmosférica diferentes.

2.3.2 WRF-Chem/DART

Para llevar a cabo la asimilación de datos con el sistema WRF-Chem/DART, se definieron tres casos de estudio: Prueba, Prueba-México y Centro de México, con las características que se muestran en la Tabla VI.

Tabla VI. Descripción de los casos de estudio considerados en la asimilación de datos con WRFDA

	Caso de Prueba	Caso de Prueba-México	Caso Centro de México
Área de estudio	China	México	Centro de México
Dominios	2	2	2
Centro	25° Lat N / 105° Lon E	19° Lat N / -98° Lon O	19° Lat N / -98° Lon O
Resolución espacial	D-1: 60km (150x112) D-1: 20km (148x157)	D-1: 60km (150x112) D-1: 20km (148x157)	D-1: 9km (90x90) D-1: 3km (88x88)
Resolución temporal	D-1: 240s D-2: 120s	D-1: 240s D-2: 120s	D-1: 30s D-2: 10s
Emisiones antropogénicas	GEI-EPA	GEI-EPA	INEM-2008
Datos químicos asimilados	MOPITT (CO)	MOPITT (CO)	MOPITT (CO) + RAMA (CO, O ₃ , NO ₂ y SO ₂)
Datos meteorológicos asimilados	PREPBUFR	PREPBUFR	PREPBUFR
Ciclos de asimilación	1	1	8

El caso de Prueba (Mizzi, 2018 y Mizzi, 2020), se diseñó en los experimentos conjuntos desarrollados entre investigadores de NCAR y China. Se recomienda su ejecución, en primera instancia, para comprobar la estabilidad numérica del sistema WRF-Chem/DART una vez instalado y validar que los recursos computacionales destinados al mismo, sean suficientes para su correcto funcionamiento. Por su parte, el caso México, permitió validar que los cambios realizados al *script* principal de sistema WRF-Chem/DART eran consistentes para un dominio que cubriera a todos el país.

El caso Centro de México se definió, en cuanto a sus principales características de configuración, con dos objetivos fundamentales: incluirle al módulo DART una fuente de asimilación puntual perteneciente a México, en este caso las observaciones de las estaciones de la RAMA y, además, aumentar el número de ciclos de asimilación con respecto al caso de Prueba. Esto permitiría que el sistema WRF-Chem/DART quedara a punto y disponible para su empleo con fines investigativos en el área de interés de la presente investigación, tanto para simulaciones numéricas de diagnóstico como de pronóstico. Por lo anterior, este caso constituye el más importante a considerar en los experimentos diseñados.

Para todos los casos de estudio, el proceso de asimilación comienza el día 24 de julio de 2014 (2014-07-24) a las 18 UTC, con ciclos de 6 horas cada uno. Es preciso indicar que, aunque siempre se consideran dos dominios de modelación (D-1 y D-2), la asimilación de datos se realiza sólo en el de menor resolución espacial (D-1).

Las condiciones iniciales y de frontera meteorológicas se obtuvieron a partir de los datos del modelo *Global Forecast System* (GFS), desarrollado por NCEP, el cual brinda información sobre diferentes variables con una resolución espacial de $1^\circ \times 1^\circ$ y temporal de 6 h (RDA, 2010). Las de tipo químico, fueron generadas con la herramienta MOZBC (Archer et al., 2014), teniendo como base las salidas del *Model for Ozone and Related chemical Tracers* (MOZART, Emmons et al., 2010), modelo químico de transporte global que obtiene la concentración de fondo de 41 especies oxidantes en la atmósfera.

Las emisiones biogénicas se obtienen del *Model of Emissions of Gases and Aerosols from Nature* (MEGAN2, Guenther et al., 1994; Guenther et al., 2006) y las de incendios forestales, del inventario de NCAR (*Global FINN*, Wiedinmyer et al., 2010). En cuanto a las emisiones antropogénicas se utiliza, según el caso, el Inventario de Gases de Efecto Invernadero (GEI) de la *Environmental Protection Agency* (EPA) o el INEM-2008.

2.4 Configuración de los módulos de asimilación de datos dentro del WRF-Chem

En esta sección se muestran los pasos realizados para la configuración, implementación (en algunos casos) y puesta a punto, de forma general, de los dos módulos de asimilación de datos utilizados en la presente investigación: WRFDA y DART.

2.4.1 Adaptación del módulo WRFDA a WRF-Chem

En el proceso de adaptación del WRFDA al WRF-Chem, fue necesario desactivar la parte química del modelo, durante el proceso de compilación del módulo. Una vez que se obtuvo el ejecutable para WRFDA, se reinstala el WRF-Chem en un directorio subyacente a WRFDA, volviendo a activarse la química en el mismo. De esta forma, se pueden generar condiciones iniciales de tipo meteorológico con WRFDA e integrarlas en el tiempo con WRF-Chem.

Para la ejecución del módulo WRFDA, se requieren tres archivos de entrada principales. Sus especificaciones, para la presente investigación, son las siguientes:

- 1- Campo de primera aproximación: archivo `wrfinput` tomado de la salida de `WRF/real.exe` para cada ciclo de asimilación.
- 2- Observaciones: archivo de datos PREPBUFR (estaciones de superficie, boyas, sondeos atmosféricos, observaciones de GPS tales como vapor de agua y tráfico aéreo) y datos BUFR de radiancia de satélites (10 sensores de satélites polares, incluyendo: seis sensores AMSU-A pertenecientes a los satélites NOAA 15-16-18-19, EOS-Aqua y METOP-2; tres sensores MHS pertenecientes a los satélites NOAA 18-19 y METOP-2 y un sensor AIRS perteneciente al satélite EOS-Aqua), ambos proporcionados por NCEP. Las características de los sensores utilizados, se describen en el Apéndice F.
- 3- Un archivo con la matriz de errores de covarianza generada por el background: proporcionado por WRFDA usando la opción CV3 (Parrish y Derber, 1992).

Primeramente, es necesario renombrar los archivos BUFR según la convención de nombres establecida en el manual de usuario del módulo, separarlos según el instrumento del que proceden y colocarlos en el directorio de trabajo del mismo. Luego, se hicieron cambios estructurales en el archivo `namelist.input` del módulo WRFDA, tales como: definir el formato BUFR como el tipo de datos de entrada y añadir información detallada sobre los sensores de satélite utilizados, entre otros. Además, se creó un script para realizar los links a los datos BUFR, PREPBUFR y a los coeficientes necesarios para la ejecución del *Community Radiative Transfer Model* (CRTM). Finalmente, el proceso de asimilación de datos se realizó con un intervalo de 6 horas.

Detalles sobre los pasos realizados para la ejecución de las técnicas 3DVAR, 4DVAR y En3DVAR en el módulo WRFDA, se muestran en el Apéndice G.

2.4.2 Configuración del módulo DART en el WRF-Chem

Para configurar el módulo DART en el modelo WRF-Chem, los códigos necesarios y sus actualizaciones fueron obtenidos en colaboración con el investigador del NCAR, Dr. Arthur Mizzi, debido a que estos no se encuentran abiertamente disponibles a la comunidad científica. Previamente, con el objetivo de garantizar el funcionamiento del módulo, se instalaron una serie de programas, cuyas versiones se listan a continuación:

- WRF: 3.9.1
- WPS: 3.9.1
- WRFDA: 3.9.1
- WRF-Chem: 3.6.1
- DART: manhattan / wrf-chem.r11690 / wrf-chem.r13172

En el caso del DART, se emplearon tres versiones diferentes durante la investigación, debido a la continua incorporación de funcionalidades. Los directorios principales del sistema WRF-Chem/DART y sus características aparecen en el Apéndice H.

Primeramente, se hicieron modificaciones a los *scripts* que permiten ejecutar el WRF-Chem/DART, según los requerimientos de cada clúster utilizado, la supercomputadora Miztli (<http://www.lancad.mx/?p=59>) y Ometeotl (Oropeza et. al, 2020).

En el caso del script principal, llamado `real_time_PANDA_RETR_RELEASE_TEST.ksh`, se establecieron diferentes parámetros, tales como: la fecha y la hora de inicio y fin del ciclo inicial (o de pronóstico), las versiones a utilizar de cada elemento en el sistema WRF-Chem/DART y las rutas a cada subdirectorío, el tiempo y la fecha de los distintos ciclos de asimilación (o de análisis), los requerimientos computacionales, entre otros elementos importantes para el correcto funcionamiento del sistema. Luego, se ejecutan de manera secuencial, cada uno de los pasos de los ciclos inicial y de asimilación en el sistema WRF-Chem/DART, como se observa en la Figura 5.

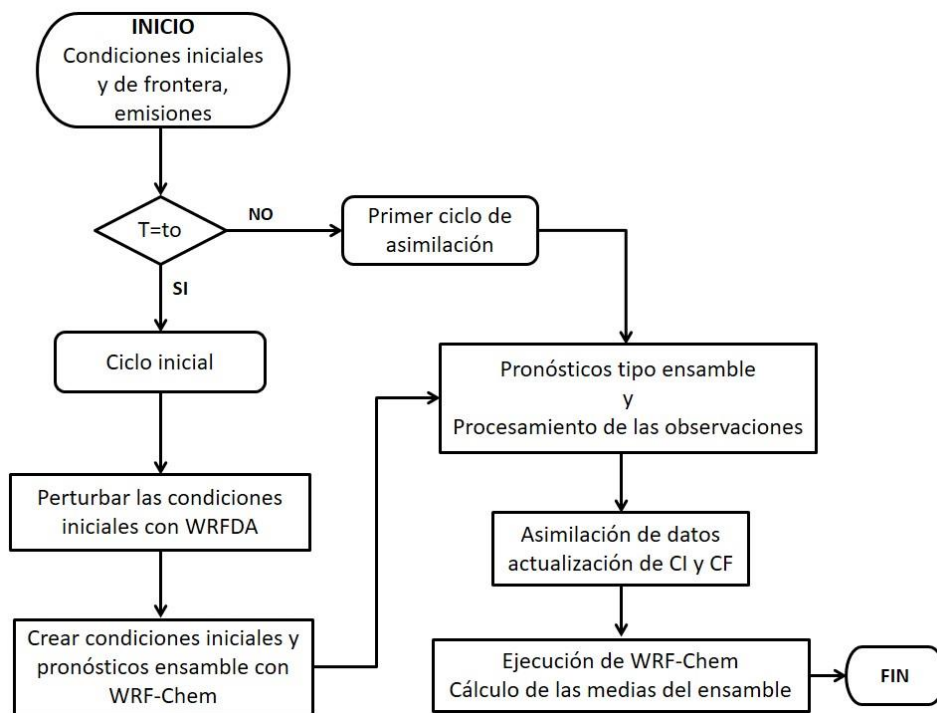


Figura 5. Procesos para los ciclos inicial y de asimilación en el sistema WRF-Chem/DART

Durante el ciclo inicial, aunque se considera el primer ciclo del proceso, no se realiza asimilación de datos propiamente, sino que se crean condiciones iniciales con WRFDA y un pronóstico inicial con WRF-Chem, ambos de tipo ensamble. Este pronóstico inicial se utiliza como *background* en el siguiente ciclo.

Posteriormente, en el segundo ciclo del proceso, se comienza a realizar la asimilación de datos químicos con DART, paso denominado como primer ciclo de asimilación. Se genera un ensamble inicial derivado de los resultados obtenidos en el ciclo anterior, con el objetivo de crear las condiciones de frontera laterales, que serán actualizadas en cada etapa del pronóstico. Además, las observaciones a asimilar se transforman al formato en que DART las lee, utilizando la herramienta IDL para el caso de las observaciones de CO procedentes de MOPITT. Luego, se realiza la asimilación de datos del ensamble, tanto meteorológicos como químicos, volviendo a actualizarse las condiciones de frontera laterales a partir de los cambios provocados por la asimilación.

Una vez obtenidas las nuevas condiciones iniciales y de frontera se ejecuta nuevamente el WRF-Chem en cada miembro del ensamble. Luego, se calculan las medias de los ensambles para cada horario del ciclo, las cuales constituyen las salidas finales y el principal resultado de esta etapa, completándose así los dos primeros ciclos del proceso.

Si se desea añadir más ciclos de asimilación al sistema, el procedimiento se realiza de forma similar a lo descrito anteriormente, aunque teniendo en cuenta que el proceso en DART funciona por pares de ciclos, es decir, para cada nuevo ciclo incorporado no se pueden utilizar las condiciones iniciales obtenidas en el ciclo inicial anterior.

2.5 Metodología para incluir observaciones de especies químicas procedentes de estaciones de la RAMA en el módulo DART

En la Figura 6 se muestra la metodología desarrollada en la presente investigación para incluir en el módulo DART, la capacidad de asimilar datos químicos procedentes de observaciones de estaciones de monitoreo atmosférico de la RAMA, la cual se encuentra actualmente en funcionamiento. Es importante aclarar que, si bien el módulo DART permite la asimilación de datos de especies químicas procedentes de AIRNOW, deben adaptarse las funcionalidades existentes e incorporarles la posibilidad de asimilar datos de la RAMA.

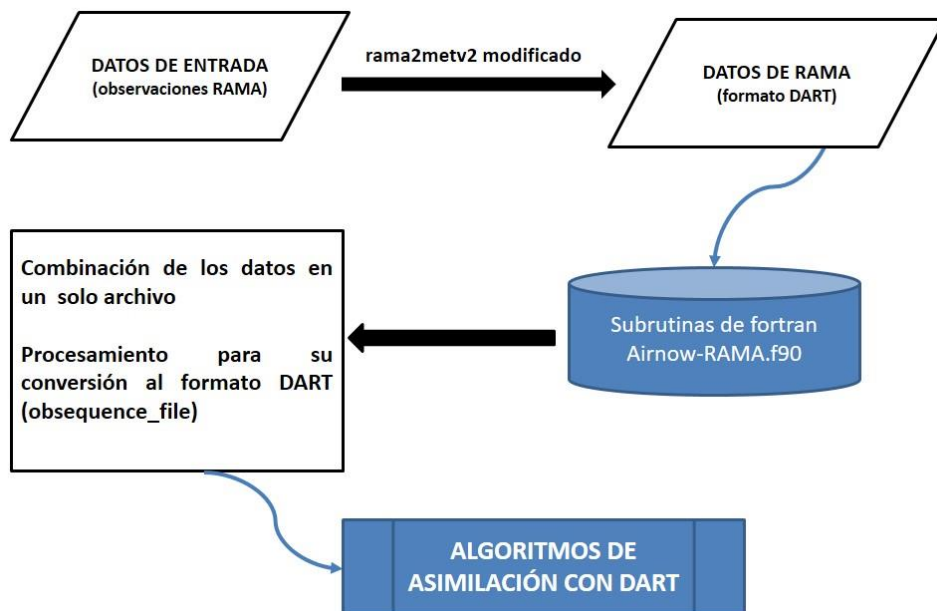


Figura 6. Asimilación de datos de la RAMA en DART

En primer lugar, se convirtieron los datos de la RAMA a un formato legible por los *scripts* de DART. En este caso, se utilizó un formato intermedio entre el que poseen los datos originales y el que presenta AIRNOW, con el objetivo de hacer menos cambios en las subrutinas de lectura de datos de DART. Además, se modificó el *script* `rama2metv2`, el cual fue desarrollado en el Centro de Ciencias de la Atmósfera de la UNAM (<https://github.com/JoseAgustin/ram2metv2>) y se emplea, originalmente, para convertir datos de la RAMA al formato en que la herramienta MET los lee. También se modificaron los *scripts* de DART encargados de la lectura de los datos procedentes de AIRNOW.

Luego, se recompiló nuevamente todo el módulo DART, para que las funcionalidades añadidas a cada subrutina, de cada contaminante asimilado, quedaran listas para su uso, a través de nuevos ejecutables incorporados al módulo de asimilación. Por último, se le agregó al *script* principal de DART, la capacidad de asimilar cada una de las fuentes añadidas, garantizando que el modo en que se mezclan estas observaciones en un archivo común no excluyera a ninguna de las observaciones posibles a asimilar y que, en el proceso de asimilación de datos, propiamente, fueran incluidas.

2.6 Métricas estadísticas

Se realizó un análisis estadístico en cada caso propuesto, a través de una comparación, punto a punto, de los resultados de las simulaciones de temperatura a 2 metros (T2), velocidad (VV) y dirección (DD) del viento a 10 metros y O₃, respecto a las mediciones de siete estaciones de la Red Automática de Monitoreo Atmosférico (RAMA) en la ZMVM (Figura 7), principal área de interés en la presente investigación, dentro de la región del Centro de México. Específicamente, se consideran: Hospital General (HGM), Merced (MER), Pedregal (PED), Santa Fe (SFE), Tlalnepantla (TLA), Tlalpan (TPN) y Xalostoc (XAL), estaciones recomendadas por LTMCE2 (2017) para evaluar el comportamiento de O₃ en la ZMVM y que, además, sus datos estaban disponibles para el período de modelación seleccionado. Las herramientas utilizadas para estos propósitos incluyen el *Unified Post Processor* versión 3.1 (UPPv3.1) y el *Model Evaluation Tools* (METv5.0).

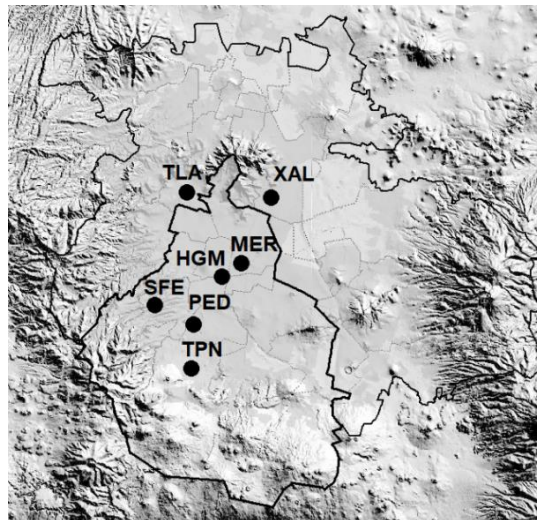


Figura 7. Estaciones de monitoreo de la RAMA seleccionadas

Para analizar la influencia de la asimilación de datos químicos procedentes de MOPITT en el desempeño del modelo WRF-Chem (Sección 3.5), se tomaron las 10 estaciones de monitoreo de la Red del Área Metropolitana de Monterrey (Figura 8), debido a que la distribución espacial de los perfiles verticales de MOPITT, para el momento en que se inicia el primer ciclo de asimilación (2014-07-25 a las 00 UTC), no cubría el área de interés y, por tanto, las estaciones de la RAMA no pueden emplearse para el análisis.

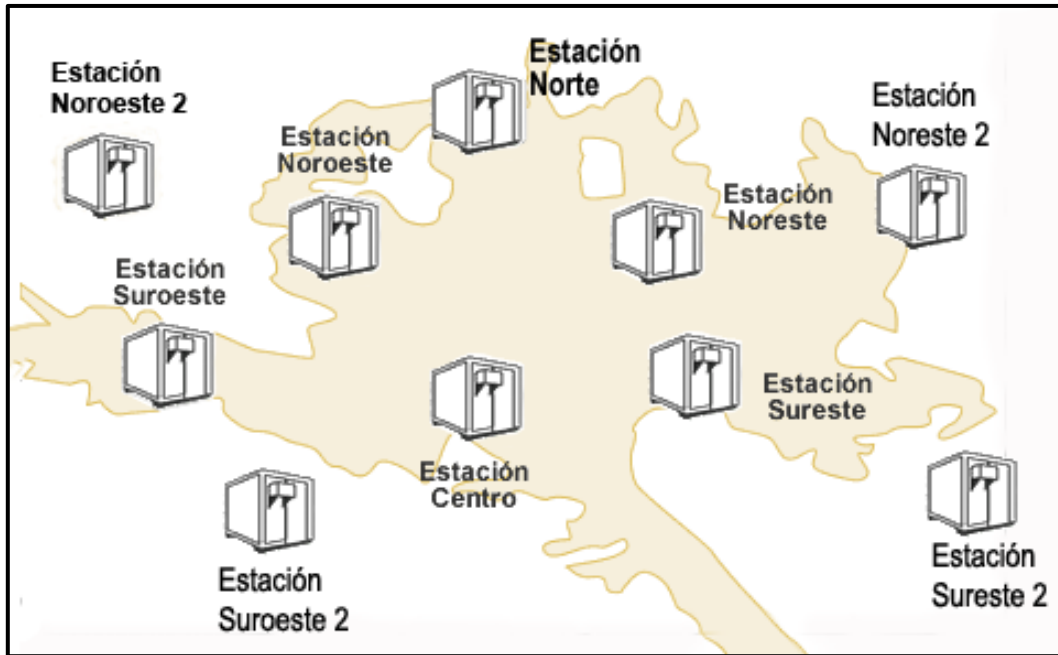


Figura 8. Estaciones de la Red del Área Metropolitana de Monterrey empleadas en la investigación (<http://aire.nl.gob.mx/>)

Las métricas estadísticas utilizadas para evaluar el desempeño del modelo son:

- Índice de Concordancia de Willmott (IOA, Willmott, 1981; Willmott, 1982): considerado actualmente como el principal indicador y evalúa el desempeño del modelo en una escala de 0 a 1, donde 1 representa la correspondencia ideal entre las simulaciones y las observaciones, calculándose:

$$IOA = 1 - \frac{\sum_{i=1}^N (p_i - o_i)^2}{\sum_{i=1}^N (|p_i - \bar{o}| + |o_i - \bar{o}|)^2} \quad (2)$$

- Raíz del error cuadrático medio (RMSE, Fox, 1981): describe la diferencia promedio entre las simulaciones y las observaciones a través de un valor siempre positivo, de forma que un valor más próximo a 0 implica un mejor desempeño del modelo y puede definirse como:

$$RMSE = \left[\frac{1}{N} \sum_{i=1}^N (p_i - o_i)^2 \right]^{1/2} \quad (3)$$

- Coeficiente de Correlación de Pearson (ρ , Pielke, 1984): describe el grado de correspondencia y correlación que existe entre los valores simulados y los observados y se calcula:

$$\rho = \frac{\frac{1}{N} \sum_{i=1}^N (o_i - \bar{o})(p_i - \bar{p})}{\sigma(o)\sigma(p)} \quad (4)$$

- Sesgo Normalizado (BIAS, Pielke 1984): indica si existe sobreestimación (BIAS>0) o subestimación (BIAS<0) por parte de las simulaciones del modelo con respecto a las observaciones y se define:

$$BIAS = \frac{1}{N} \sum_{i=1}^N \frac{(p_i - o_i)}{o_i} * 100 \quad (5)$$

En las ecuaciones anteriores: p_i es el valor simulado; o_i es el valor observado al mismo tiempo i ; \bar{o} es el promedio de todos los valores observados y \bar{p} el de todos los simulados; N es el número de observaciones por estación y σ es la desviación estándar.

CAPÍTULO 3: RESULTADOS Y DISCUSIÓN

3.1 Comportamiento de las condiciones iniciales en la asimilación de datos con el módulo WRFDA

Primeramente, se calcularon los incrementos de análisis, los cuales proporcionan una medida de la variación de cada variable una vez realizado el proceso de minimización. Los resultados se presentan para el dominio D-2 (ver Figura 4) y para las variables VV y T2, debido a que estas constituyen las dos de más influencia sobre el estado de la calidad del aire en el área de interés, dentro de los campos meteorológicos que son perturbados durante la asimilación de datos. La intensa y constante radiación solar que se registra en la ZMVM a lo largo de todo el año, contribuye a la formación de O₃, principal contaminante en la región. Además, las frecuentes inversiones térmicas que se producen, favorecen la acumulación de contaminantes, mientras que el flujo del viento, determina su transporte y dispersión (PROAIRE, 2012)

Las Figuras 9 y 10 muestran los campos de primera aproximación (a), análisis (b, c) y los incrementos de análisis (d, e) para T2, obtenidos el día 2013-05-01 para los casos de asimilación definidos en la Tabla V). En los campos de *background* se observa un ambiente generalmente cálido en todo el dominio para ambos horarios, con valores de T2 superiores a los promedios del mes de mayo en el área, acorde con lo que se reporta en SMN (2013). Los campos de análisis muestran que, ambas opciones de asimilación, sí modifican las condiciones iniciales para T2 en todo el dominio y tienen influencia sobre la variable. Los campos de incremento indican, de forma general, que se recomienda disminuir los valores de T2 en gran parte del área de estudio en un orden de $O(10)$, en correspondencia con su escala característica.

Las correcciones más extremas en los incrementos de análisis se observan, en su mayoría, a las 00UTC, superando los -12°C (Figura 9a), aunque también, en menor medida, a las 12 UTC. Esto sugiere que el campo de primera aproximación sobreestima los valores de T2 en esas áreas del dominio, las cuales se ubican, principalmente, en zonas con relieve extremo, ya sean valles o montañas.

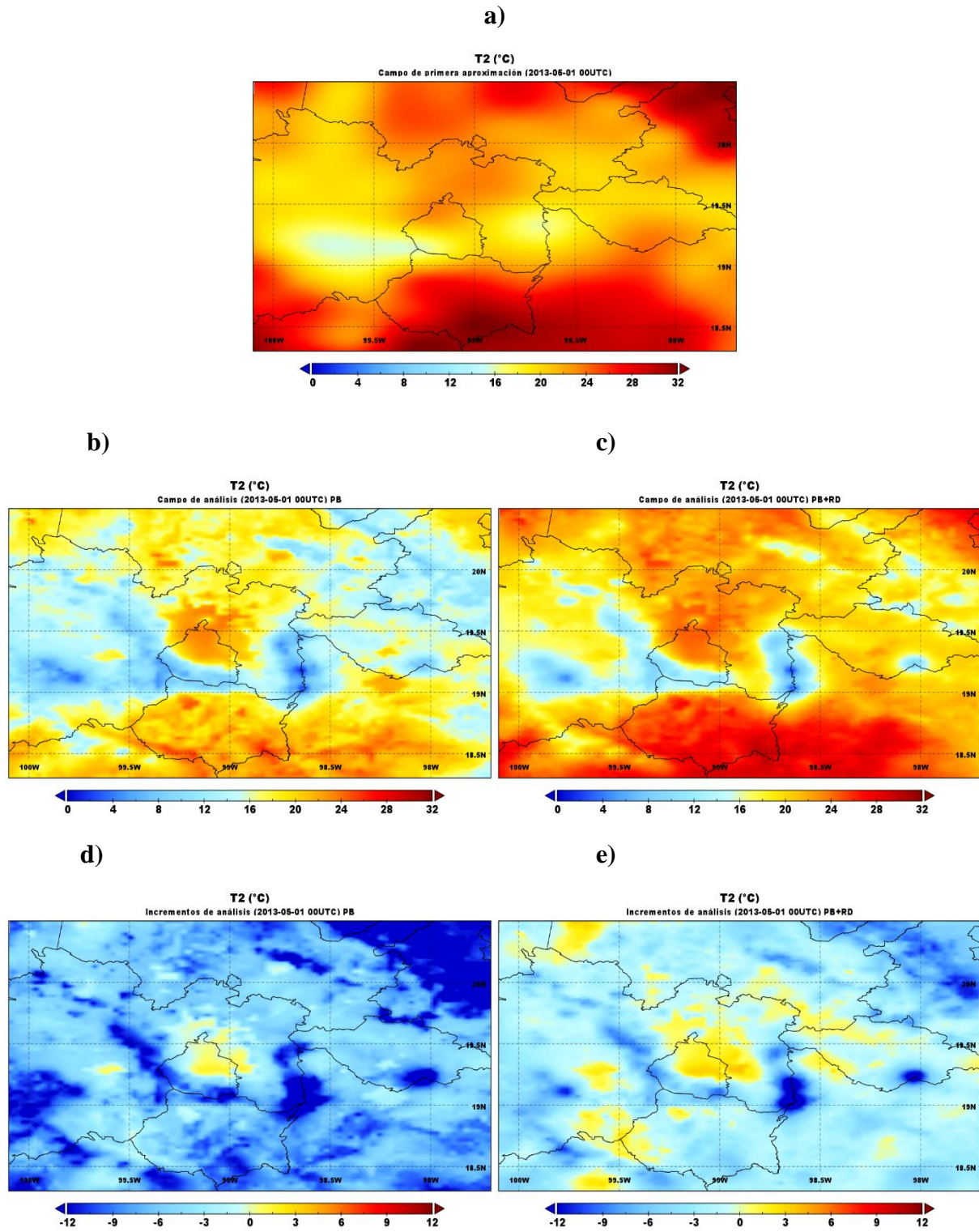


Figura 9. Campos de primera aproximación (a), de análisis (b, c) y de incrementos de análisis (d, e) para la variable T2 y para los casos PB y (PB+RD), el día 2013-05-01 a las 00 UTC

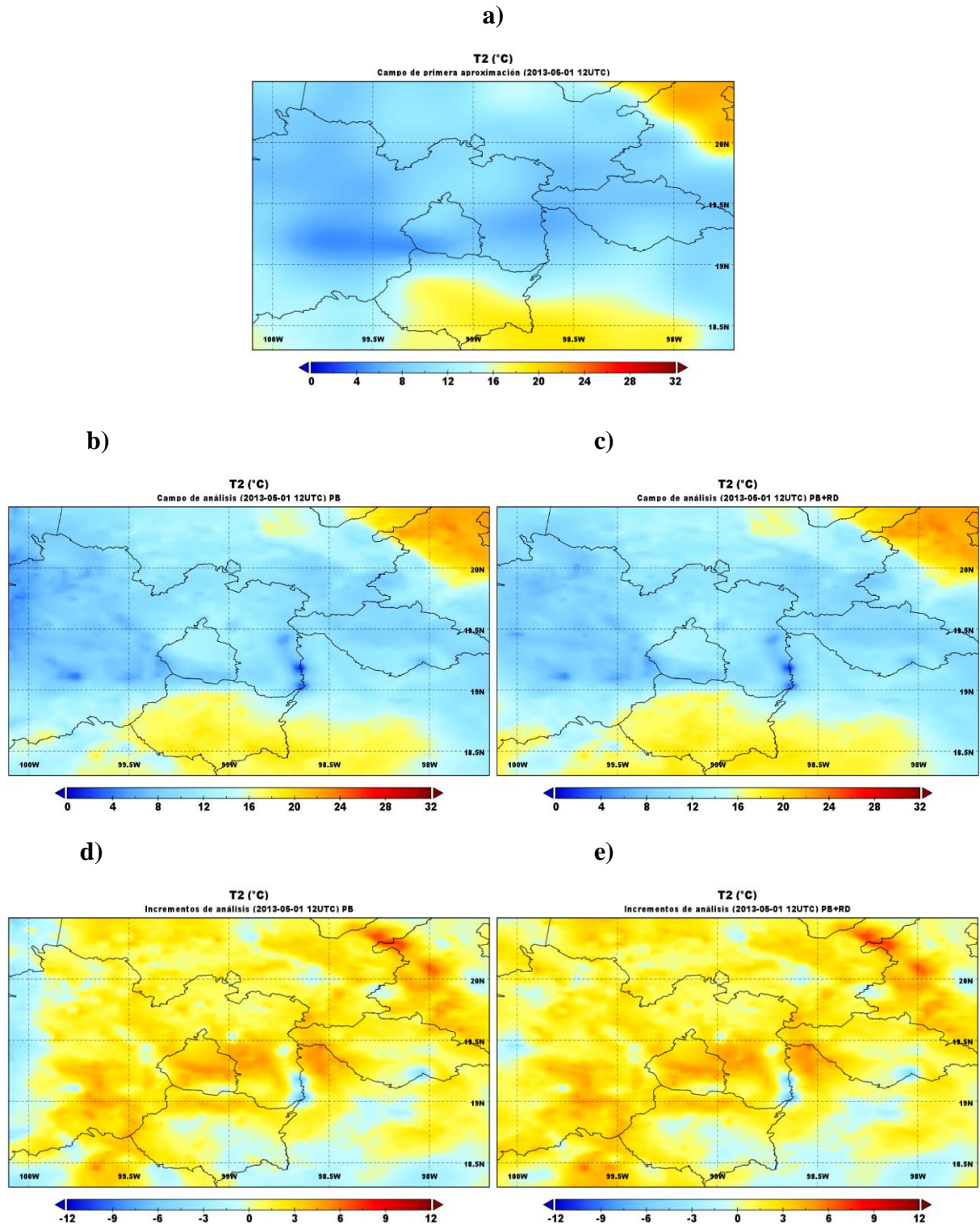


Figura 10. Campos de primera aproximación (a), de análisis (b, c) e de incrementos de análisis (d, e) para la variable T2 y para los casos PB y (PB+RD), el día 2013-05-01 a las 12 UTC

El análisis anterior sugiere que las condiciones iniciales utilizadas originalmente para realizar las simulaciones numéricas con el modelo WRF-Chem, sin asimilación de datos, pueden no reflejar el comportamiento real de T2 en esa región del área de estudio. La causa de esta situación puede estar relacionada con el hecho de que los datos obtenidos de NARR son campos de análisis promedio para la variable en cuestión y, además, con una resolución de 32 km que es mayor a la utilizada en esta investigación. No obstante, esto se puede corregir, incluyendo datos de observaciones en el proceso de asimilación.

En el caso de la ZMVM, ambas opciones de asimilación sugieren una corrección positiva en los valores de T2 mayor a 3°C, en el horario de las 12 UTC, mientras que, a las 00 UTC, los incrementos de análisis sugieren que se eleven también hasta en 3°C es la porción nordeste del área, pero indican la necesidad de disminuir los valores de T2 en el resto de la ZMVM, superando incluso los -9°C en zonas limítrofes con el Estado de México y Morelos. Esto es muy frecuente que ocurra en áreas urbanas, en las cuales hay presencia del efecto de isla de calor y que no puede ser simulado eficientemente por aquellos modelos numéricos que no incluyan urbanización dentro de sus parametrizaciones. Además, Barrales (2017) demuestra que el uso de suelo utilizado en el modelo es también un aspecto de gran importancia en la calidad de los resultados de las simulaciones y que tiene una influencia directa en los mismos.

Por otra parte, a las 00 UTC, el caso (PB+RD) propone una corrección mayor a los valores de T2 en comparación con PB, debido a que (PB+RD) incluye asimilación de datos de radiancia de satélites, con cobertura más amplia del área de estudio respecto a los datos PREPBUFR (en su mayoría observaciones en superficie separadas espacialmente entre sí), lo cual permite describir mejor el comportamiento de la variable. A las 12 UTC, los incrementos de análisis para ambas opciones de asimilación se asemejan más uno de otro, lo cual puede relacionarse con el mayor volumen de datos disponibles en ese horario y, por ende, las observaciones puntuales pueden representar mejor el estado meteorológico.

Las Figuras 11 y 12 muestran los campos de viento obtenidos en cuanto a su magnitud (VV) y dirección (DD), de forma similar a lo presentado para T2. En ambos horarios de inicio de la asimilación, el *background* indica la presencia de un flujo continuo sobre gran parte del dominio, prevaleciendo los vientos de dirección oeste y suroeste, aunque para el área sur de la ZMVM, a las 12 UTC, se observan también del noroeste. Se aprecia también la suavidad en los campos iniciales de VV y DD, en lo cual influye el hecho de que NARR es un modelo que ofrece campos promedio para las distintas variables que considera y refleja sus características en una escala mayor a la utilizada en esta investigación. Por tanto, el mejor rendimiento de NARR lo tiene a escala sinóptica, pero, a mesoescala, puede no describir el comportamiento del viento, con igual precisión, en el área de estudio.

Al realizarse la asimilación de datos, los campos de análisis muestran cambios en VV y DD en todo el dominio de estudio y para los casos presentados, con mayor variación a las 12 UTC. El comportamiento de estas variables es mucho más estructurado y presenta un mayor nivel de detalle, acorde con las características del terreno (flujos montaña-valle) y la resolución de la malla utilizada en las simulaciones. Poseer un mayor conocimiento de las características del viento mejora la descripción de la dispersión de contaminantes en el área, dada la influencia decisiva que tiene el mismo sobre este fenómeno (Bei et al. 2012).

El patrón de vientos obtenido para los campos de incremento a las 12 UTC es muy similar para ambas opciones de asimilación consideradas. Esto puede relacionarse con que, en ese horario sinóptico, hay mayor disponibilidad de datos de observaciones convencionales, las cuales tienen mayor relevancia en la opción PB, mientras que los datos de radiancia de satélite, asimilados en la opción (PB+RD), tienen menor influencia sobre los campos de las variables VV y DD, respecto a T2. Por su parte, a las 00 UTC, sí se aprecian mayores diferencias entre los campos de análisis obtenidos.

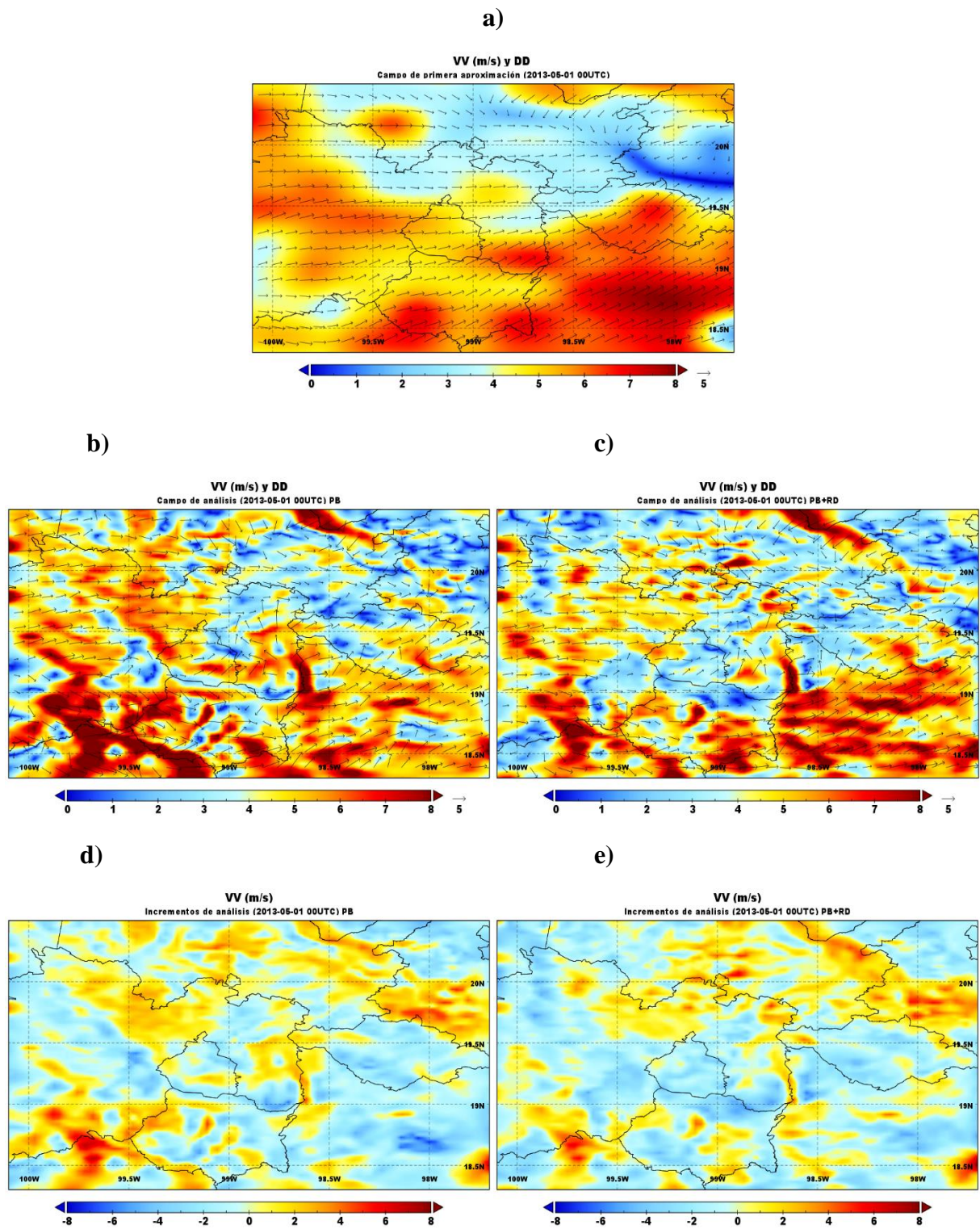
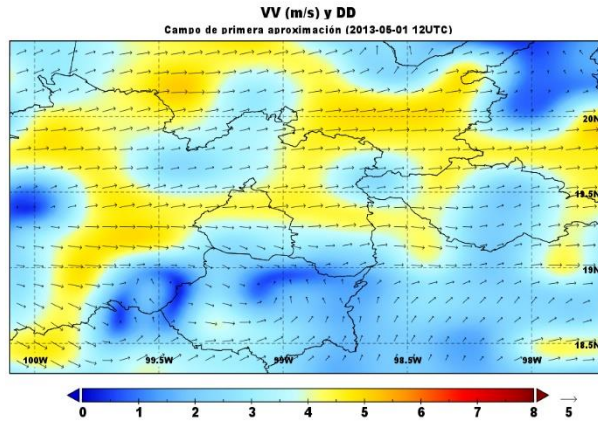
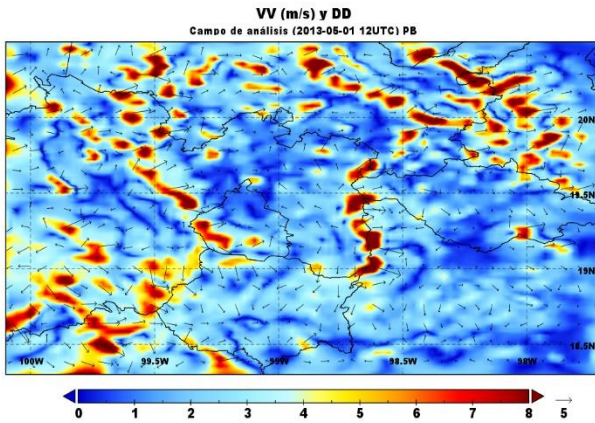


Figura 11. Campos de primera aproximación (a), de análisis (b, c) y de incrementos de análisis (d, e) para las variables VV y DD, para los casos PB y (PB+RD), el día 2013-05-01 a las 00 UTC

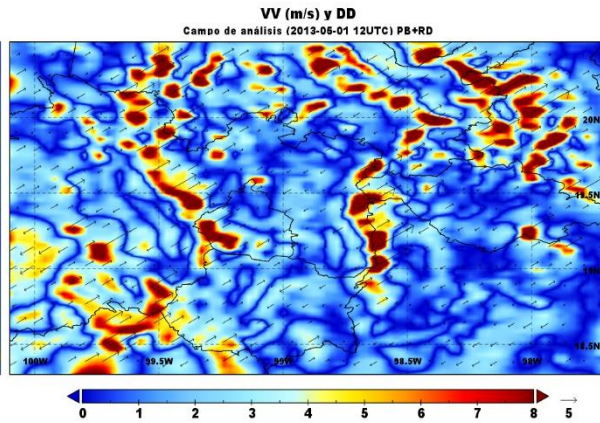
a)



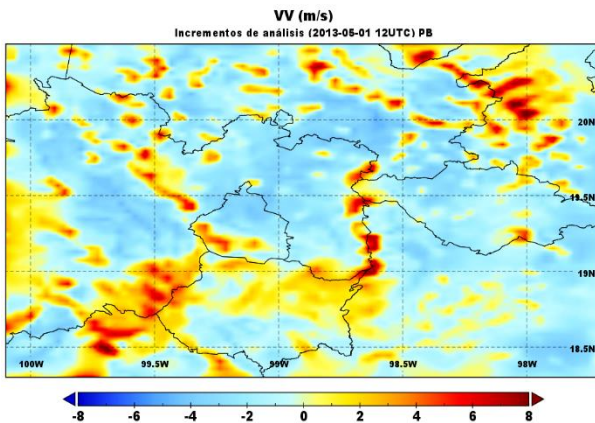
b)



c)



d)



e)

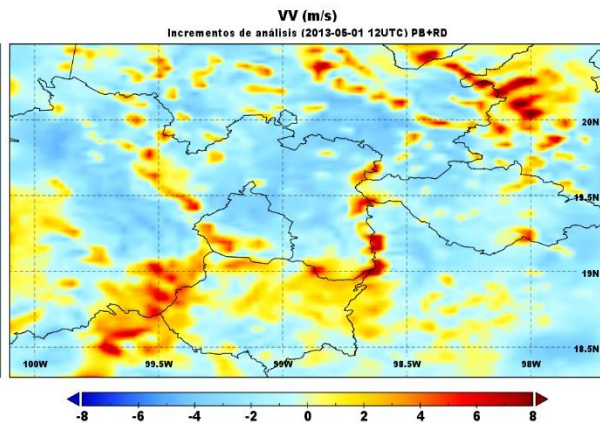


Figura 12. Campos de primera aproximación (a), de análisis (b, c) y de incrementos de análisis (d, e) para las variables VV y DD, para los casos PB y (PB+RD), el día 2013-05-01 a las 12 UTC

Las Figuras 13 y 14 representan, en cada caso, la distribución espacial de T2 y VV (a, b, c), y los incrementos de análisis para los casos PB (d, e, f) y (PB+RD) (g, h, i) para los ciclos de asimilación tercero, sexto y noveno (comenzando las simulaciones siempre a las 00 UTC). Estos gráficos muestran los cambios que ocurren en los campos de T2 y VV durante los ciclos indicados para cada caso de asimilación y también respecto a los demás en el transcurso del tiempo.

Para ambas variables, los valores absolutos de los incrementos son mucho menores en cada uno de los ciclos respecto al primero de ellos. Esto puede suceder porque el proceso de asimilación de datos comienza de forma *cold start*, es decir, los campos de análisis de NARR se utilizan para obtener las condiciones iniciales en el primer ciclo de asimilación ($t = 0$), pudiendo causar que se observen las mayores diferencias de incrementos en ese ciclo inicial.

Por otro lado, el resto de los ciclos del proceso se realizan de la forma *warm start*, es decir, el pronóstico inmediato anterior del modelo WRF-Chem se toma como condición inicial para el siguiente ciclo de asimilación y, por tanto, se espera que los incrementos de análisis sean menores (Hernández, 2013). Esto se debe a que dicho pronóstico se realiza considerando las condiciones iniciales ya ajustadas previamente, lo cual permite reflejar mejor el comportamiento de las variables modeladas, acorde a la resolución espacial del dominio, la topografía y los procesos físicos que tengan lugar, entre otros. Además, a medida que aumentan los pasos de tiempo, el modelo numérico se ajusta continuamente y los incrementos de análisis deberían disminuir al aumentar los ciclos de asimilación.

En el caso de T2, los mayores incrementos propuestos alcanzan hasta 4°C en el sexto ciclo de asimilación para (PB+RD), mientras que para VV, son menos pronunciados que los de T2 y, además, son siempre mayores en la opción (PB+RD) respecto a PB.

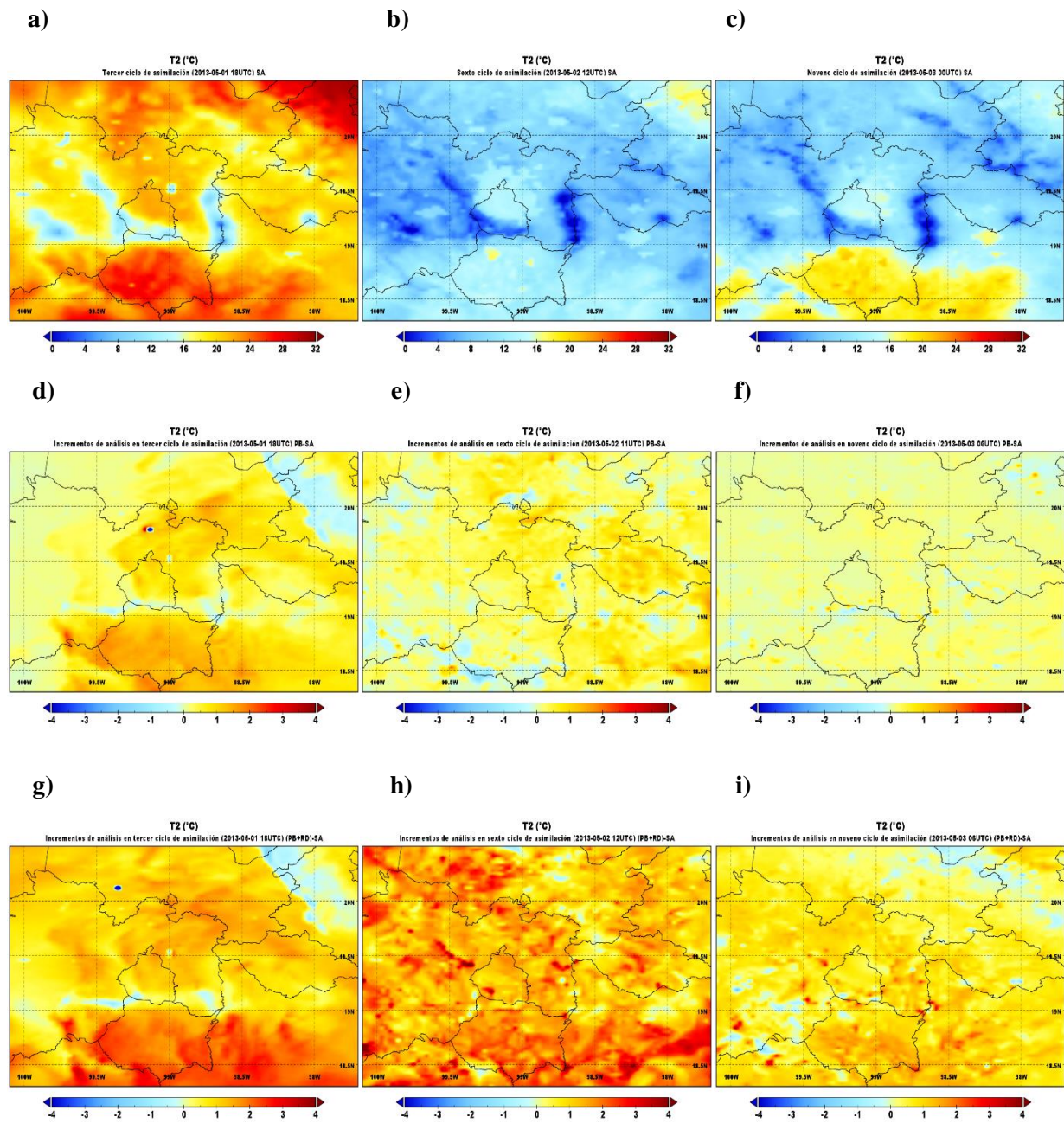


Figura 13. Evolución temporal de la distribución espacial (a, b, c) y los incrementos de análisis para la variable T2 y para los casos PB (d, e, f) y (PB+RD) (g, h, i) para el tercer, sexto y noveno ciclos de asimilación (comenzando desde las 00 UTC)

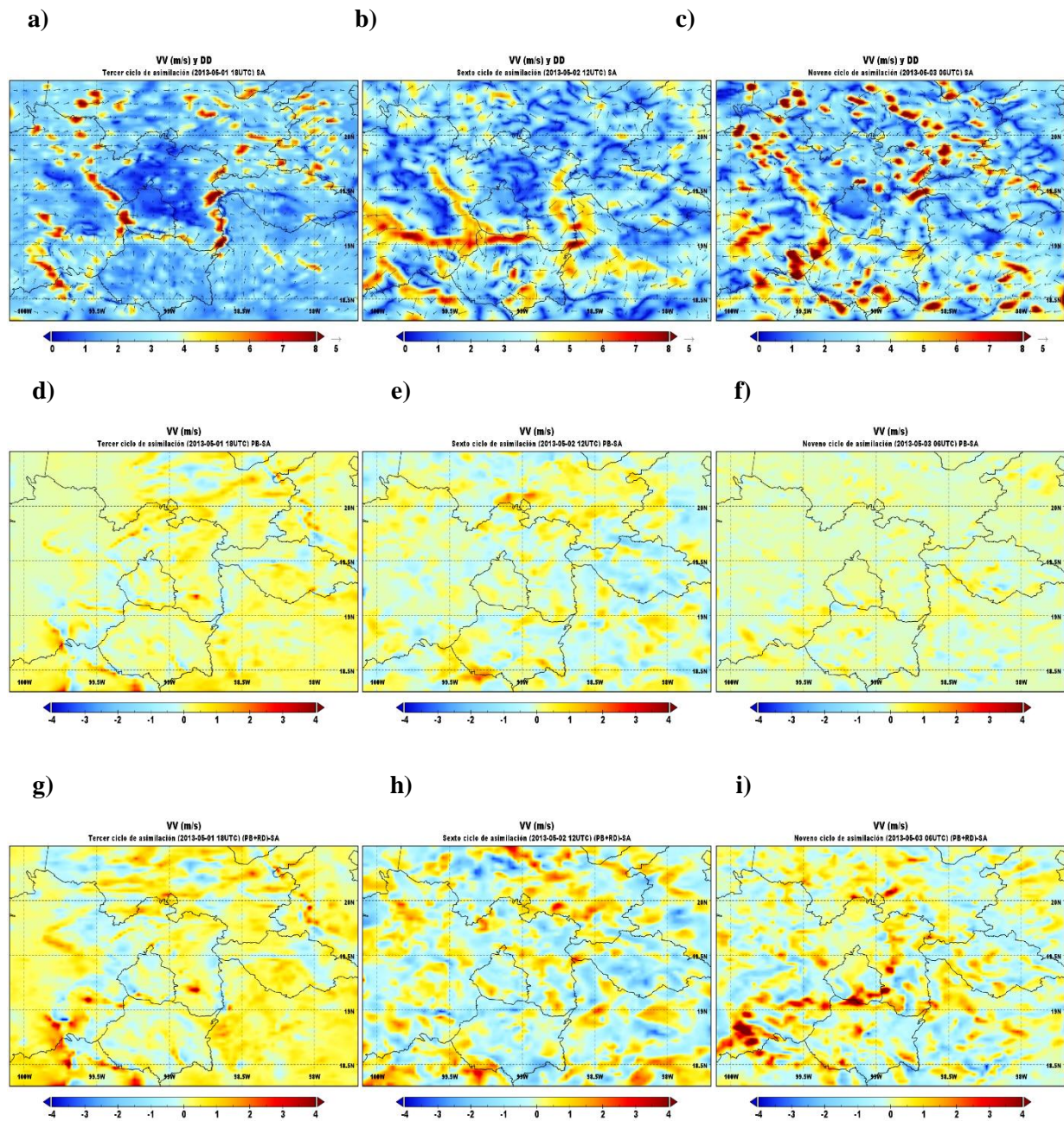


Figura 14. Evolución temporal de la distribución espacial (a, b, c) y los incrementos de análisis para la variable VV y para los casos PB (d, e, f) y (PB+RD) (g, h, i) para el tercer, sexto y noveno ciclos de asimilación (comenzando desde las 00 UTC)

Hasta aquí, es posible apreciar, que la asimilación de datos meteorológicos sí modifica los campos iniciales de las variables T2, VV y DD, principalmente la opción (PB+RD) que incluye radiancia de satélite. Los resultados obtenidos se asemejan a otras posibles fuentes de condiciones iniciales, en este caso, pronósticos de GFS y reanálisis de ERA5. Según se expresa en Kalnay (2003), esto es muy útil para la simulación numérica a mayor resolución que las que poseen modelos como NARR y puede contribuir a mejorar el desempeño del WRF-Chem sobre el área de estudio.

En la Figura 15 se muestran las distribuciones espaciales de O₃, principal contaminante atmosférico en la ZMVM, obtenidas con el modelo WRF-Chem sin realizar el proceso de asimilación (SA, a, b, c), así como las diferencias entre estos campos y los resultantes de la asimilación de datos meteorológicos con las opciones PB (d, e, f) y (PB+RD) (g, h, i) para los horarios del día correspondientes a los ciclos de asimilación tercero, sexto y noveno, comenzando las simulaciones numéricas a las 00 UTC.

Los campos de las diferencias, los cuales no son incrementos de análisis pues la variable química O₃ no fue asimilada propiamente, muestran que sí se producen cambios en las concentraciones de O₃ simuladas a lo largo de todo el dominio, predominando valores en torno a los 10 ppb. Estos cambios se evidencian al comparar entre sí las opciones de asimilación que se utilizaron, así como entre estas y el caso SA y también con el paso de los ciclos de asimilación.

Los resultados anteriores coinciden con Seaman (2000), el cual demuestra que existe una sensibilidad en el comportamiento de las variables químicas cuando se modifican las condiciones iniciales de tipo meteorológico, pues estas juegan un papel esencial como datos de entrada en los modelos de calidad del aire. En general, las emisiones se dispersan a una tasa dependiente de la intensidad y dirección del viento y al variar estos elementos entre los diferentes experimentos de asimilación, ocurre un cambio en la proporción de contaminantes primarios emitidos a la atmósfera y, por ende, la producción de O₃ en el dominio estudiado. Otros factores a considerar son el transporte horizontal y vertical, la humedad relativa, la radiación solar y la temperatura.

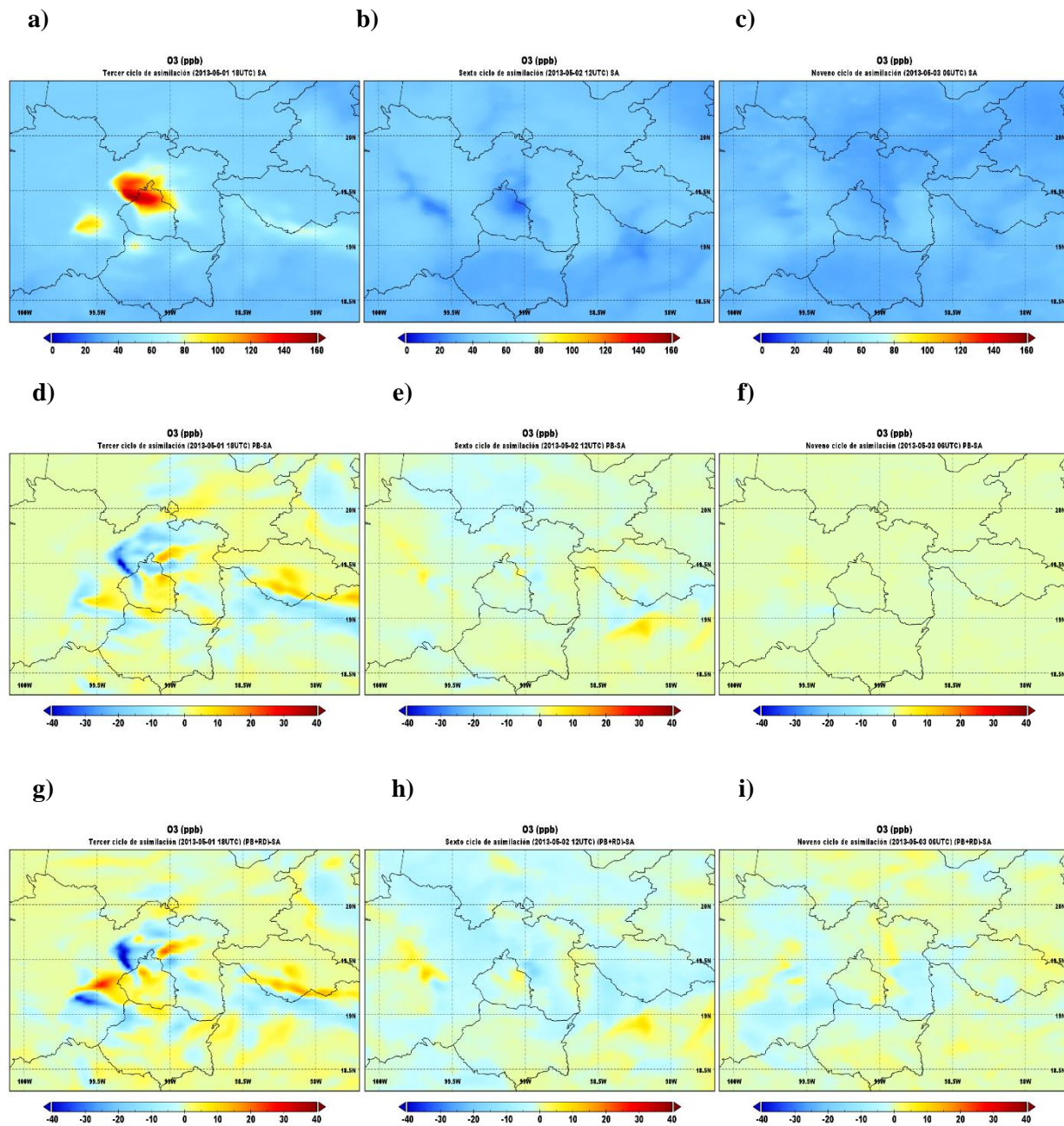


Figura 15. Evolución temporal de la distribución espacial (a, b, c) de O_3 y las diferencias entre los campos de O_3 respecto al campo SA para los casos PB (d, e, f) y (PB+RD) (g, h, i) para el tercer, sexto y noveno ciclos de asimilación (comenzando desde las 00 UTC)

Por su parte, de Foy et al. (2008) planteó la existencia de seis episodios generadores de O_3 sobre el Valle de México teniendo en cuenta las condiciones meteorológicas, los cuales fueron utilizados por Bei et al. (2010) para determinar la influencia de la asimilación de datos en la modelación de calidad del aire en cada uno de ellos.

Las mayores diferencias en las concentraciones de O₃ se observan en la ZMVM y las áreas circundantes, llegando a 40 ppb en el horario correspondiente al tercer ciclo de asimilación (18 UTC, o sea, 12 horas tiempo local), valor que se mantiene dentro de la escala característica de la variable. Estos resultados se asemejan a los obtenidos en Bei et al. (2008), quienes plantean que la asimilación de datos meteorológicos modificó hasta en 40 ppb los valores simulados con WRF-Chem en horarios de concentraciones máximas del contaminante y estuvieron entre 10 y 20 ppb en el resto del día.

3.2 Evaluación de desempeño del modelo WRF-Chem en la asimilación de datos meteorológicos con WRFDA.

Los resultados obtenidos en el análisis estadístico realizado para cada caso propuesto, se presentan a continuación.

Las Figuras 16 y 17 muestran, respectivamente, las series de tiempo para los valores observados y modelados de T2 y VV, en todas las estaciones (a-d) y particularmente MER (b-e) y SFE (c-f), para los horarios 00 UTC y 12 UTC, para los casos sin asimilación (SA), con asimilación de datos PREPBUFR (PB) y con asimilación de datos PB más radiancia de satélites (PB+RD). En el eje x están representados los pasos temporales desde el momento de inicio de la asimilación de datos.

En la Figura 16 se puede apreciar que el modelo es capaz de reproducir el ciclo diario de T2, alcanzando los valores máximos y mínimos, respectivamente, a las 21 UTC y a las 12 UTC, horarios en los cuales ocurren, climatológicamente, estos extremos. Sin embargo, el máximo se subestima en todos los casos, lo cual es muy frecuente en la simulación de la calidad del aire debido a los fenómenos de la isla de calor (Jauregui, 1997; Cui y de Foy, 2012) que afectan la ZMVM. La subestimación también puede verse influenciada por las parametrizaciones y el diseño propio del WRF, incluyendo el uso de suelos. La opción (PB+RD) muestra valores más próximos a los observados en las estaciones, principalmente a las 12 UTC, como se muestra en los gráficos de MER y SFE. Para todos los casos modelados, existe mayor correlación con las observaciones en los periodos en que se realiza la asimilación.

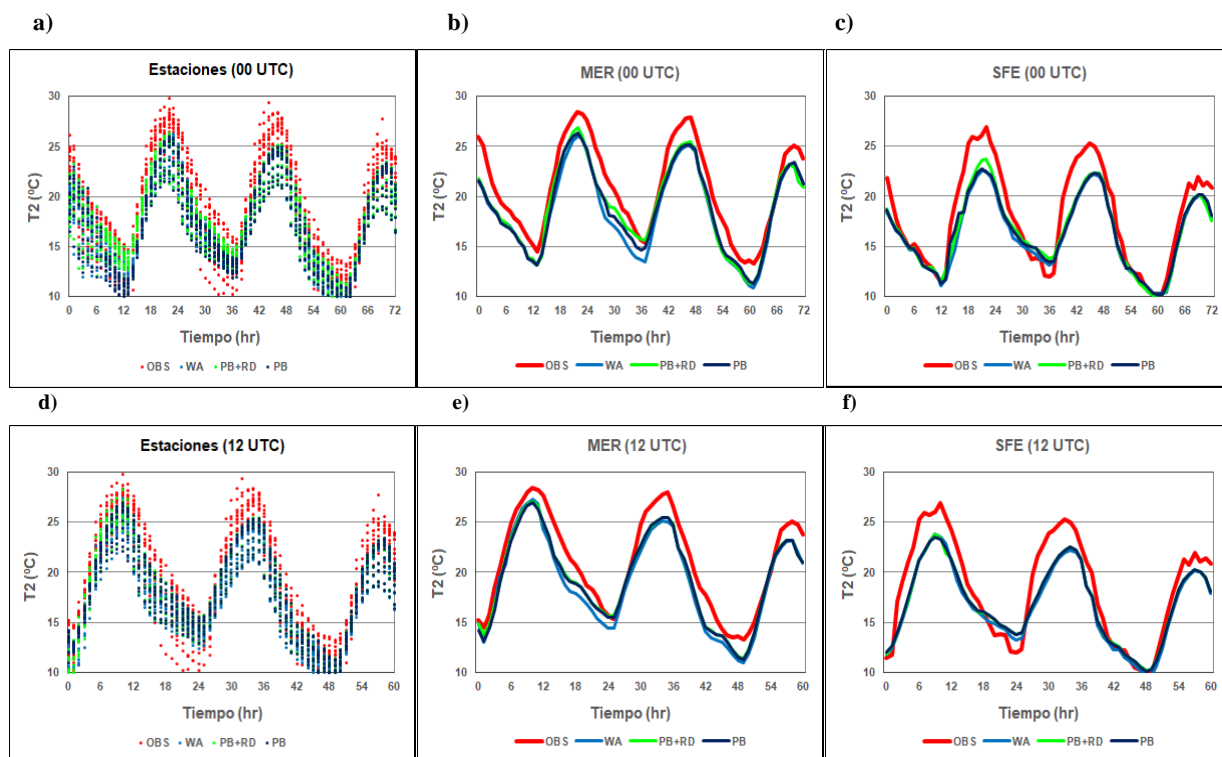


Figura 16. Series temporales de los valores simulados y observados de T2 sobre todas las estaciones (a, d), MER (b, e) y SFE (c, f), para todos los casos de asimilación (00 UTC y 12 UTC)

En la Figura 17 se observa que los valores modelados en cada caso de estudio se encuentran más alejados de las observaciones respecto a lo obtenido en la variable T2, es decir, existe una mayor dispersión entre los valores observados y los modelados.

El horario 00 UTC ofrece mejores resultados tanto en el gráfico que hace referencia a todas las estaciones como en los correspondientes a las estaciones individuales MER y SFE, especialmente en los primeros pasos de ejecución. El caso (PB+RD) es la fuente de asimilación que muestra los mejores resultados para VV. Los valores máximos se subestiman en los últimos pasos temporales del horario 12 UTC, es decir, entre las 7 y 19 horas tiempo local, del día 2 de mayo de 2013 (2013-05-02), período en el cual ocurrió la contingencia ambiental en la ZMVM. Teniendo en cuenta que la advección es un factor que contribuye a la distribución de los contaminantes atmosféricos en un área, si el modelo no reflejó de manera adecuada los valores de VV respecto a las observaciones, al asimilarse estos datos conlleva a que otras variables como la concentración de O₃ se modifiquen.

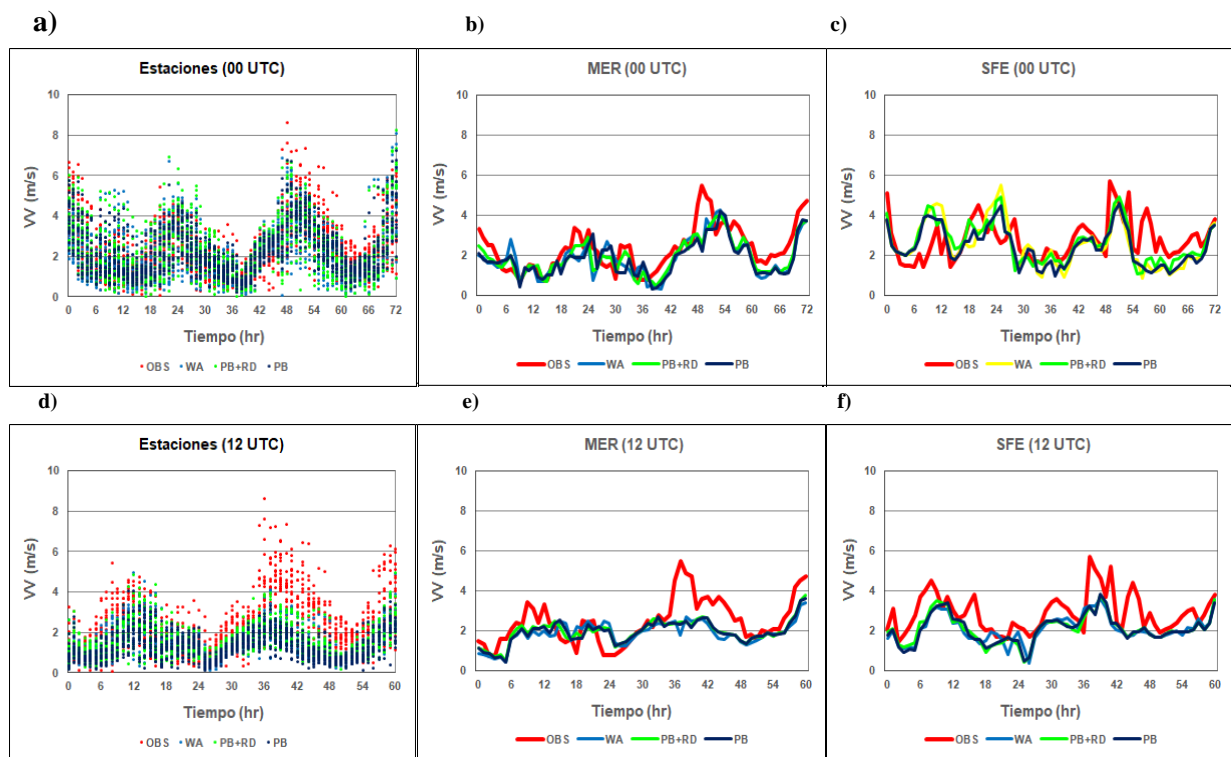


Figura 17. Series temporales de los valores simulados y observados de VV sobre todas las estaciones (a, d), MER (b, e) y SFE (c, f), para todos los casos de asimilación (00 UTC y 12 UTC)

Las métricas IOA y RSME para T2 y VV se muestran en la Figura 18 para todos los casos de asimilación considerados y para los horarios de las 00 UTC y 12 UTC. Las estaciones que se presentan fueron escogidas según los criterios establecidos en la Sección 2.6. Los resultados obtenidos se aproximan a los mencionados en las figuras 16 y 17. Para la variable T2, se observan valores de IOA superiores a 0.9 en todas las estaciones, para cada caso y tiempo, lo que significa que existe una buena relación entre los valores modelados y las observaciones. La opción (PB+RD) tiene el IOA más alto en seis de las siete estaciones para las 00 UTC y en todas las estaciones para las 12 UTC. Los valores de RMSE para T2 siempre están entre 1 y 3 °C (los más bajos se encontraron a las 12 UTC) y TPN es la estación con el peor desempeño de todas. En todos los casos, excepto en la estación HGM, (PB+RD) tiene los valores más bajos de RSME, siendo la mejor fuente de asimilación.

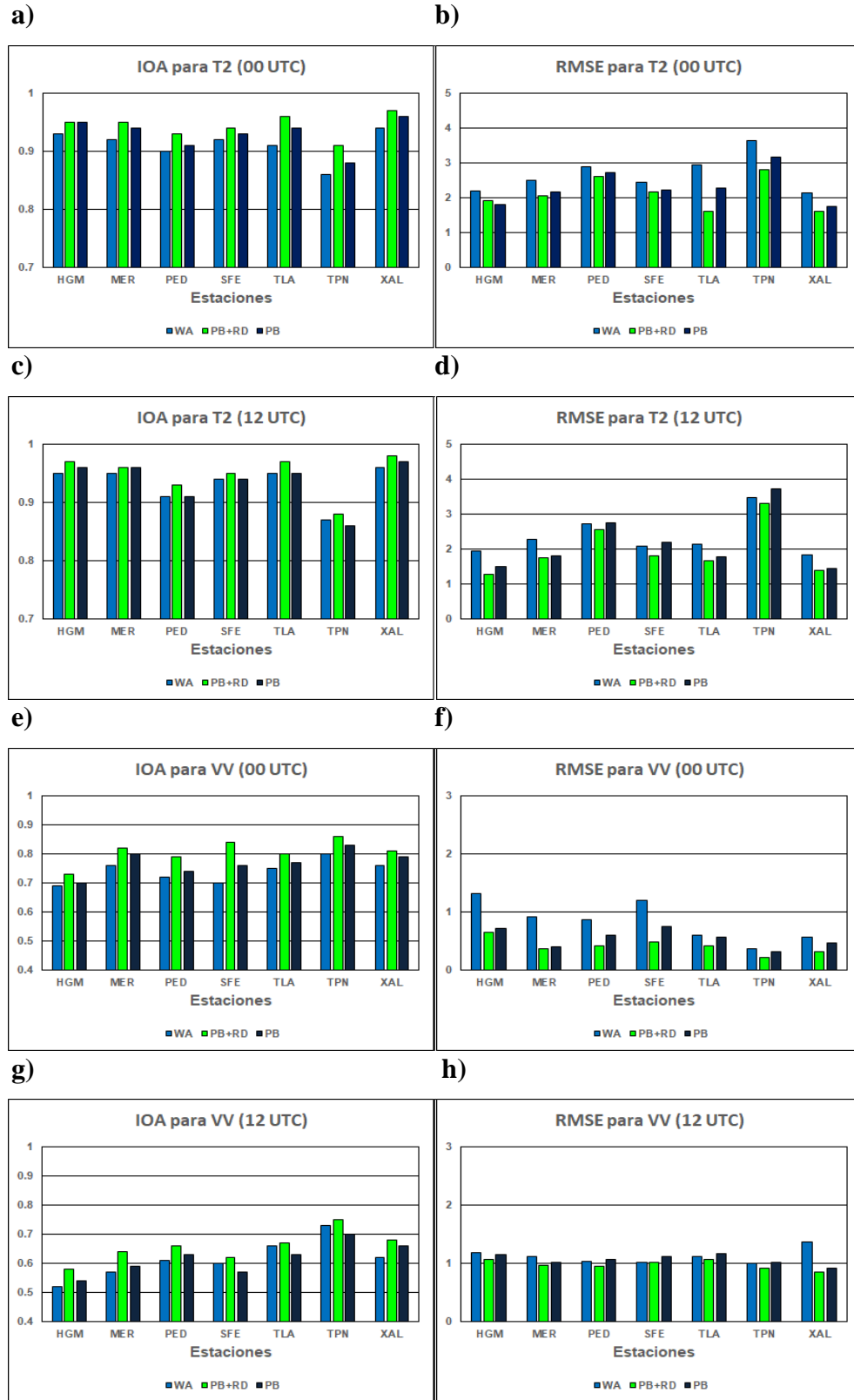


Figura 18. Valores de IOA y RMSE para las variables T2 y VV para cada una de las estaciones seleccionadas y para todos los casos de asimilación (00 UTC y 12 UTC)

En el caso de VV, las estadísticas de IOA muestran valores más cercanos a 1 en el horario de las 00 UTC, siempre superiores a 0.7 en las opciones PB y (PB+RD), mientras que SA presenta las cifras más bajas. La estación con mejor desempeño es TPN, precisamente la de peores resultados para el caso de T2, mientras que HGM presentó las mayores diferencias respecto a las observaciones, en el horario de las 12 UTC. Los valores de RMSE para VV siempre son inferiores a 1,5 m/s, superiores cualitativamente a las 00 UTC y con el caso de asimilación (PB+RD).

La Figura 19 muestra la evolución temporal de las concentraciones de O₃, simuladas y observadas, para todas las estaciones seleccionadas y, particularmente, para MER y SFE, a modo de ejemplo, para los horarios de las 00 UTC y las 12 UTC, incluidos todos los casos de asimilación. Esto se complementa con la Figura 20, en la cual se muestran las métricas IOA y RMSE para todas las estaciones, horarios y casos.

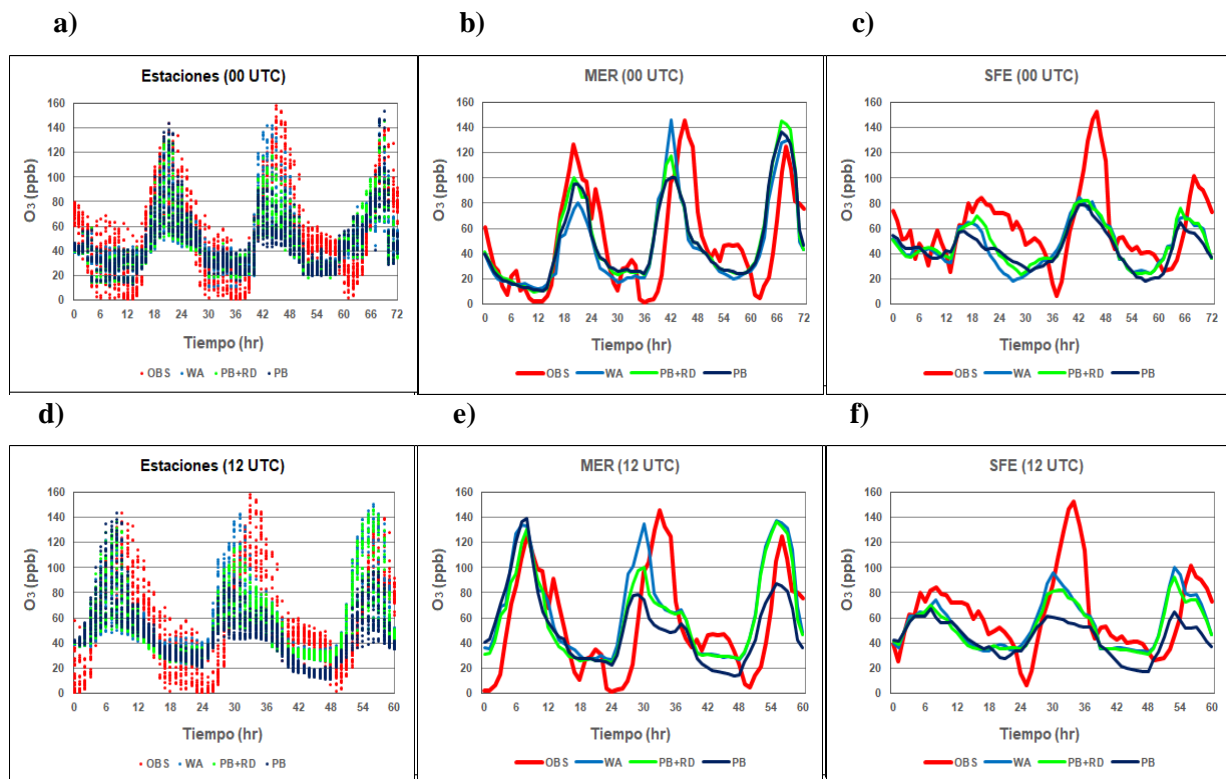


Figura 19. Series temporales de las concentraciones de O₃, simuladas y observadas, para todas las estaciones (a, d), MER (b, e) y SFE (c, f) para todos los casos de asimilación (00 UTC y 12 UTC)

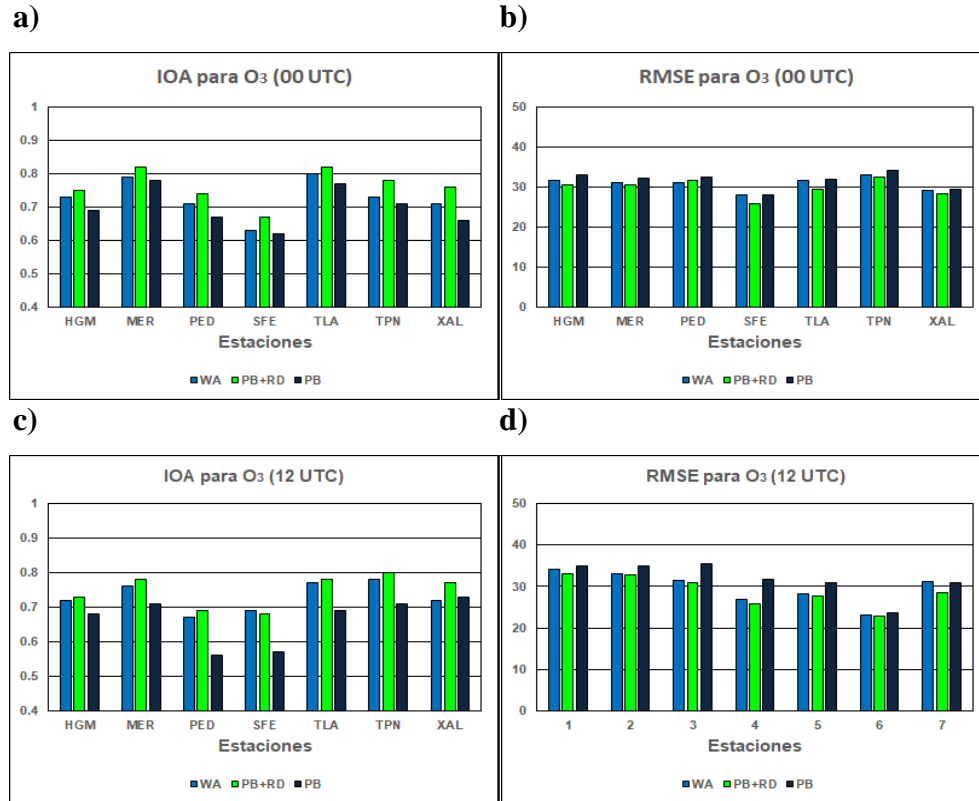


Figura 20. Valores de IOA y RMSE para las concentraciones de O₃ para cada una de las estaciones seleccionadas y para todos los casos de asimilación (00 UTC y 12 UTC)

En general, el modelo representa el ciclo diario de las concentraciones de O₃, aunque la correspondencia entre los valores modelados y observados no es tan alta como en el caso de las variables meteorológicas T2 y VV. Para ambos horarios, el modelo es capaz de simular acertadamente los máximos de O₃, pero tiende a sobreestimar los mínimos, los cuales presentan cifras, en su mayoría, por debajo de las concentraciones de fondo (40 ppb) de la variable modelada. Esto puede estar relacionado con el mecanismo químico empleado en el WRF-Chem y su interacción con el inventario de emisiones, provocando que el modelo no tenga buen desempeño en concentraciones de O₃ dentro del rango de valores entre 0 y 40 ppb (García et al., 2018).

Por otra parte, el rendimiento general para cada fuente de asimilación es diferente del que se muestra para T2 y VV. Por ejemplo, aunque la opción (PB+RD) mantiene el mejor desempeño, se puede ver que, en algunos intervalos de series de tiempo, los valores obtenidos en las simulaciones sin asimilación de datos (SA) están más

cercanos a las observaciones que (PB+RD). El caso PB siempre presenta los peores resultados. En cuanto a los valores mostrados para las estaciones MER y SFE, estos son muy contrastantes. La serie temporal de las concentraciones de O₃ en la estación MER muestra perfiles similares entre los valores simulados y observados, mientras que en SFE se observan diferencias significativas. Los resultados de IOA y RMSE muestran una relación aceptable entre el modelo y las mediciones.

En el caso de IOA para las 00 UTC, seis de las siete estaciones tienen valores superiores a 0.7, la cual se considera una cifra aceptable, cualitativamente, para esta variable, según lo reportado en la modelación de O₃ con WRF-Chem realizada por Zabkar et al. (2015), Kryza et al. (2020) y Sun et. al (2021). En estas investigaciones se obtuvieron índices de IOA entre 0.4 y 0.8, para zonas fuera de México, mientras que en Bei et al. (2008) y Bei et al. (2010) se reportaron valores entre 0.5 y 0.7 para esta métrica, en dominios de modelación ubicados dentro del país.

Los mejores resultados se observan en todas las estaciones que comienzan la asimilación de datos a las 12 UTC, con valores de IOA más altos. Para el RMSE, todos los valores se encuentran alrededor de los 30 ppb. El caso PB tiene el rendimiento más bajo, mientras que (PB+RD) continúa siendo la mejor opción de asimilación.

La Tabla VII muestra los valores promedio de los estadígrafos encontrados para todas las estaciones, tanto para las 00 UTC como para las 12 UTC y para todas las opciones de asimilación, pudiendo observar las tendencias generales de cada variable estudiada. Para el caso de T2, los resultados de las métricas para ambos horarios son similares y con buenos resultados para esta variable. El mejor rendimiento se obtiene mediante la combinación de datos de (PB+RD) a las 12 UTC. Los valores de BIAS indican que el modelo subestima las observaciones en todos los casos y el RMSE siempre es menor que 2.5°C mientras que ρ e IOA son siempre mayores que 0.9.

Tabla VII. Valores de IOA, BIAS, RMSE y ρ para T2, VV and concentraciones de O₃ para los casos SA, PB y (PB+RD) (00 UTC y 12 UTC)

		00 UTC			12 UTC		
		SA	PB	(PB+RD)	SA	PB	(PB+RD)
T2	IOA	0.91	0.93	0.94	0.93	0.94	0.95
	BIAS	-2.33	-1.94	-1.62	-1.76	-1.30	-1.28
	RMSE	2.49	2.41	2.32	2.46	2.24	2.22
	P	0.93	0.94	0.96	0.95	0.95	0.96
VV	IOA	0.75	0.77	0.80	0.64	0.63	0.66
	BIAS	0.02	-0.03	-0.05	0.07	0.13	0.05
	RMSE	0.83	0.54	0.40	1.19	1.21	1.18
	P	0.72	0.76	0.79	0.61	0.59	0.63
O₃	IOA	0.71	0.67	0.72	0.72	0.68	0.74
	BIAS	-9.61	-11.56	-8.57	-8.74	-10.65	-7.95
	RMSE	27.71	29.98	26.39	26.94	29.44	25.53
	P	0.65	0.62	0.67	0.64	0.64	0.68

Para el caso de VV, se observa un mejor desempeño para la opción (PB+RD) a las 00 UTC. Los valores IOA y ρ , comenzando la asimilación a las 00 UTC, son mayores que 0.7 en todos los casos, mientras que, comenzando a las 12 UTC, son menores que 0.66 y 0.63, respectivamente. Estos resultados, para VV, son satisfactorios, aunque se pudo esperar que fueran mejores en el horario de las 12 UTC. Sin embargo, las no linealidades inherentes a esta variable meteorológica provocan que la misma sea muy compleja de modelar, por lo que algunos resultados presentan irregularidades, las cuales se reflejan en las métricas estadísticas calculadas. En cuanto al BIAS, este indica que el modelo tiende a sobrestimar ligeramente las observaciones.

Para el caso de las concentraciones de O₃, la opción (PB+RD) a las 12 UTC es la de mejor desempeño nuevamente. Los valores de ρ se encuentran entre 0.6 y 0.7, mientras que IOA siempre es mayor que 0.67 y RMSE se ubica entre 25 y 30 ppb, lo cual refleja una buena correspondencia, en general, entre los valores simulados y observados. El BIAS indica que el modelo WRF-Chem subestima las concentraciones de O₃ medidas en todas las estaciones seleccionadas.

3.3 Evaluación del desempeño del modelo WRF-Chem en la asimilación de datos con WRFDA para las técnicas 3DVAR, 4DVAR y En3DVAR

Luego de realizarse los experimentos numéricos para dos períodos de estudio con diferentes condiciones de contaminación en el área de interés, es decir, con presencia de contingencia ambiental (1-4 de mayo de 2013, es decir, 2013-05-01 al 2013-05-04) y sin ella (4-7 de mayo de 2013, o sea, 2013-05-04 al 2013-05-07), empleando las técnicas 3DVAR, 4DVAR y En3DVAR, se obtienen resultados que permiten evaluar el desempeño numérico de estas técnicas en la asimilación de datos y la respuesta del modelo WRF-Chem ante los cambios que las mismas producen en las condiciones iniciales, en cuanto a la correspondencia que existe entre sus salidas numéricas y las observaciones de las estaciones disponibles.

3.3.1 Caso de estudio con contingencia ambiental

A continuación, se muestran los resultados para el periodo del 1 al 4 de mayo de 2013 (2013-05-01 al 2013-05-04), donde se activó la contingencia ambiental en la ZMVM.

En la Figura 21 se muestran las series de tiempo de valores observados y modelados para T2, considerando las técnicas de asimilación de datos utilizadas. Se observa que todas representan la distribución temporal de la variable y su ciclo diario, aunque tienden a subestimar los valores extremos máximos observados y sobreestimar los mínimos, sobre todo en las primeras horas de la simulación. Desde el punto de vista gráfico, no se aprecian diferencias significativas entre las técnicas empleadas, sin embargo, sí se observa que, como resultado de la modelación con 4DVAR, se obtienen valores de T2 más próximos a las mediciones, principalmente en los picos máximos.

Se observa que las estaciones HGM, MER y TLA presentan los mejores desempeños del modelo WRF-Chem, mientras que TPN, los menos favorables. La técnica 4DVAR brinda al modelo la posibilidad de obtener resultados más próximos a las observaciones pues, como se demuestra en Kalnay (2003), esta opción de asimilación es la más consistente y genera condiciones iniciales más cercanas al estado real.

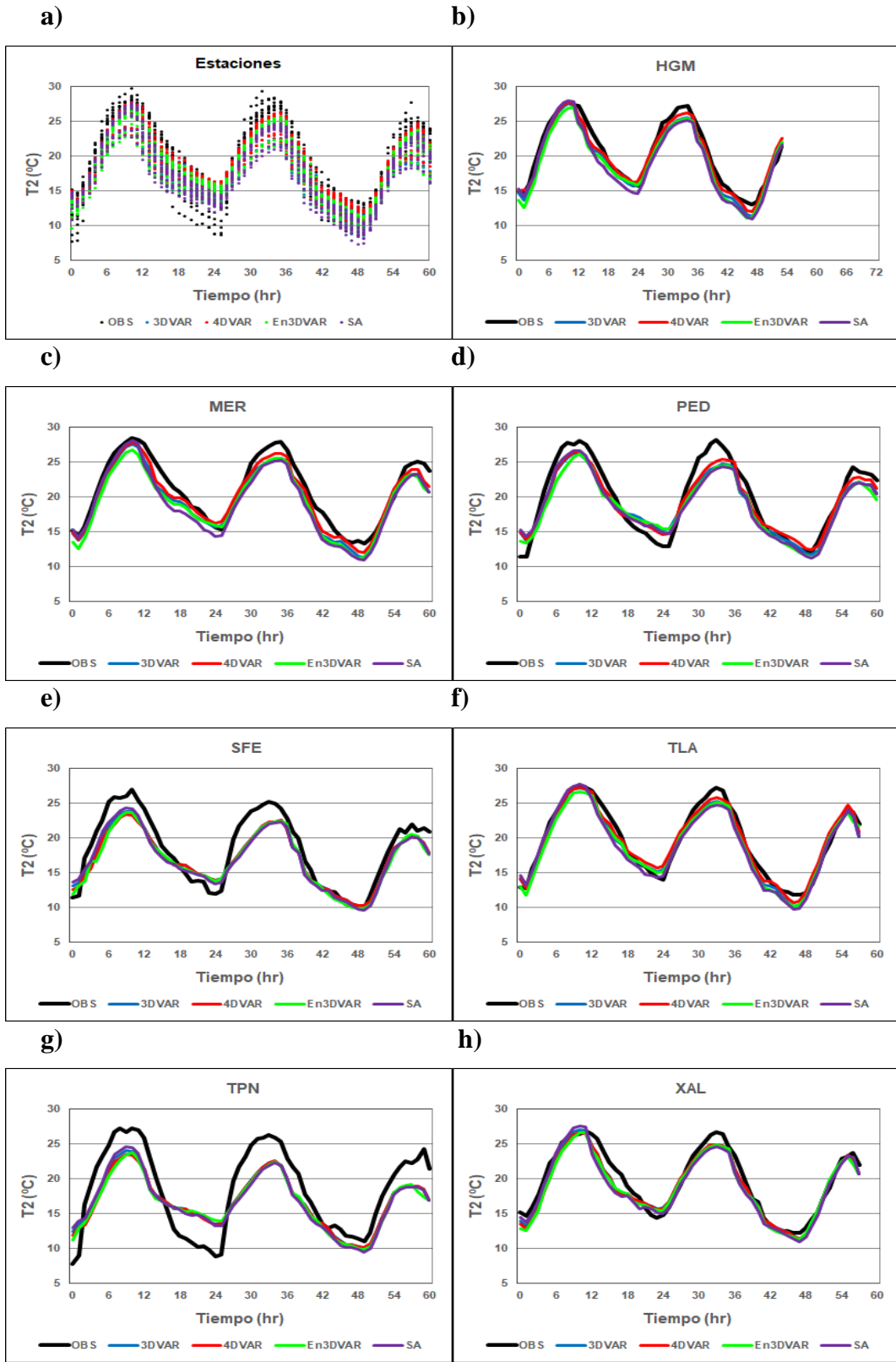


Figura 21. Series temporales de los valores observados y modelados de T2 para el conjunto de todas las estaciones y las seleccionadas, a partir de las diferentes técnicas de asimilación de datos

En la Tabla VIII, aparecen los resultados de las métricas estadísticas calculadas para todos los experimentos realizados en el conjunto de todas las estaciones disponible, en forma de promedio, así como de manera individual para aquellas seleccionadas, teniendo en cuenta la técnica de asimilación de datos que se aplica en cada caso.

Tabla VIII. Valores de IOA, BIAS, RMSE y ρ para la variable T2 y para las técnicas de asimilación empleadas: SA, 3DVAR, 4DVAR y En3DVAR

		SA	3DVAR	4DVAR	En3DVAR
HGM	IOA	0.94	0.95	0.97	0.96
	BIAS	-1.73	-1.29	-1.09	-1.14
	RMSE	2.39	2.28	2.01	2.17
	ρ	0.94	0.94	0.97	0.95
MER	IOA	0.94	0.95	0.96	0.96
	BIAS	-1.74	-1.21	-1.10	-1.15
	RMSE	2.47	2.21	2.10	2.11
	ρ	0.93	0.95	0.96	0.95
PED	IOA	0.93	0.94	0.95	0.95
	BIAS	-1.67	-1.34	-1.21	-1.17
	RMSE	2.42	2.37	2.18	2.16
	ρ	0.95	0.94	0.95	0.96
SFE	IOA	0.92	0.94	0.94	0.93
	BIAS	-1.86	-1.34	-1.20	-1.82
	RMSE	2.57	2.31	2.26	2.49
	ρ	0.93	0.94	0.95	0.93
TLA	IOA	0.94	0.94	0.96	0.95
	BIAS	-1.79	-1.37	-1.11	-1.53
	RMSE	2.44	2.24	2.06	2.32
	ρ	0.92	0.94	0.96	0.93
TPN	IOA	0.91	0.93	0.92	0.93
	BIAS	-1.96	-1.83	-1.87	-1.84
	RMSE	2.76	2.48	2.54	2.59
	ρ	0.92	0.93	0.93	0.92
XAL	IOA	0.92	0.94	0.95	0.93
	BIAS	-1.81	-1.52	-1.36	-1.78
	RMSE	2.51	2.37	2.33	2.45
	ρ	0.94	0.93	0.94	0.93
PROMEDIO	IOA	0.93	0.95	0.96	0.95
	BIAS	-1.76	-1.28	-1.13	-1.24
	RMSE	2.49	2.32	2.18	2.24
	ρ	0.94	0.96	0.96	0.95

Se observa que 4DVAR presenta valores de estadísticos cualitativamente superiores al resto de las técnicas, lo cual indica que existe un mejor desempeño del modelo WRF-Chem en la simulación numérica de la variable T2. Los valores de IOA y ρ iguales a 0.96, para el caso de 4DVAR, comprueban que existe muy buena correspondencia y relación lineal entre las salidas del modelo y las observaciones de las estaciones de monitoreo, lo cual se complementa con un RMSE = 2.18, el más bajo de todos los obtenidos. Para todas las técnicas y estaciones, los valores de BIAS son negativos, lo cual indica que el modelo subestima las mediciones, aunque en un rango aceptable para la escala de la variable. El caso SA presenta siempre los resultados con mayor diferencia respecto a las observaciones, indicando que para T2 la asimilación de datos mejora el rendimiento del modelo independientemente de la técnica empleada

Se obtuvieron también resultados para la concentración de O₃, principal contaminante responsable de la activación de las contingencias ambientales en la ZMVM. En la Figura 22 se muestran las series de tiempo de los valores observados y modelados para esta variable, empleando las técnicas de asimilación de datos disponibles. Se observa que, para todos los casos, el modelo representa de forma satisfactoria la distribución temporal de O₃ en el área de estudio, así como su ciclo diario. No se aprecia, gráficamente, una marcada tendencia a sobrestimar o subestimar los valores extremos de la variable, principalmente en los primeros días de modelación, donde la influencia de la asimilación de datos es superior respecto al resto del periodo de tiempo considerado. Existe una superioridad cualitativa de 4DVAR respecto al resto de las opciones empleadas, lo cual ratifica a esta técnica de asimilación como la de mejores resultados para el caso de estudio que se analiza.

En cuanto a las estaciones seleccionadas se aprecia que, en todos los casos, la técnica 4DVAR describe un comportamiento de la variable más próximo a los valores observados, en comparación con el resto de las variantes aplicadas. Además, el modelo WRF-Chem tuvo un mejor desempeño en las estaciones TPN, TLA y SFE, mientras que el resto tiene menor relevancia cualitativa.

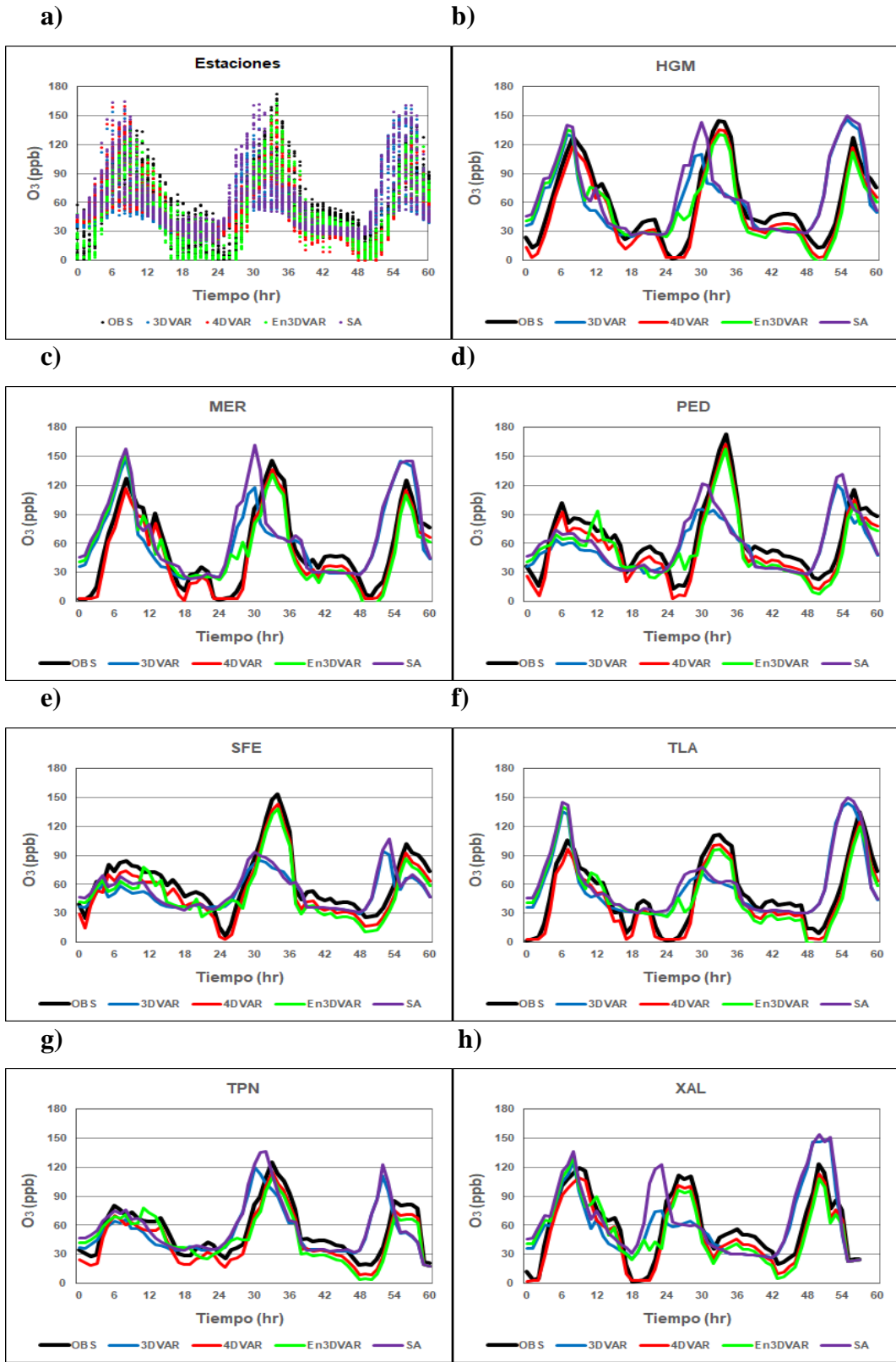


Figura 22. Series temporales de los valores observados y modelados de O_3 para el conjunto de todas las estaciones y las seleccionadas, a partir de las diferentes técnicas de asimilación de datos

En la Tabla IX se presentan los valores obtenidos para cada una de las métricas estadísticas calculadas en los distintos experimentos que se realizaron para la variable O₃, según la técnica de asimilación de datos aplicada en cada uno de ellos, tanto los valores promedios como para cada estación analizada.

Tabla IX. Valores de IOA, BIAS, RMSE y ρ para la variable O₃ y para las técnicas de asimilación empleadas: SA, 3DVAR, 4DVAR y En3DVAR

		SA	3DVAR	4DVAR	En3DVAR
HGM	IOA	0.70	0.72	0.75	0.72
	BIAS	-8.87	-8.61	-7.32	-8.49
	RMSE	27.61	26.24	24.73	25.86
	ρ	0.66	0.68	0.70	0.69
MER	IOA	0.69	0.71	0.75	0.73
	BIAS	-9.13	-8.51	-7.22	-7.84
	RMSE	28.76	26.27	24.34	25.81
	ρ	0.64	0.68	0.71	0.68
PED	IOA	0.68	0.70	0.75	0.71
	BIAS	-9.54	-8.99	-7.36	-8.92
	RMSE	29.83	27.71	24.38	27.24
	ρ	0.62	0.66	0.71	0.66
SFE	IOA	0.77	0.79	0.82	0.79
	BIAS	-6.42	-6.20	-5.18	-6.12
	RMSE	24.93	23.56	20.46	23.09
	ρ	0.73	0.74	0.80	0.75
TLA	IOA	0.78	0.79	0.82	0.80
	BIAS	-6.38	-6.01	-5.14	-5.49
	RMSE	24.41	22.97	20.15	22.21
	ρ	0.73	0.76	0.81	0.77
TPN	IOA	0.79	0.81	0.83	0.81
	BIAS	-5.52	-5.44	-5.02	-5.35
	RMSE	23.07	21.96	20.04	21.70
	ρ	0.76	0.77	0.80	0.78
XAL	IOA	0.72	0.75	0.77	0.74
	BIAS	-8.62	-7.83	-6.26	-7.71
	RMSE	26.64	25.44	23.83	25.02
	ρ	0.67	0.68	0.75	0.69
PROMEDIO	IOA	0.72	0.74	0.78	0.75
	BIAS	-8.74	-7.95	-6.12	-7.48
	RMSE	26.94	25.53	23.69	24.71
	ρ	0.66	0.68	0.75	0.70

La técnica 4DVAR presenta nuevamente valores superiores desde el punto de vista cualitativo para todas las métricas respecto al resto de las opciones, con una mejor correspondencia y relación lineal de las salidas del modelo WRF-Chem, respecto a las estaciones de monitoreo, afirmación sustentada en las cifras de IOA = 0.78 y R = 0.75. Los valores negativos del BIAS indican que el modelo tiende a subestimar ligeramente las mediciones en un rango promedio desde -6.12 ppb (4DVAR) hasta -8.74 (SA). Los valores de RMSE se encuentran dentro de un rango aceptable para la variable, destacando 4DVAR con 23.69. El caso 3DVAR presenta los resultados menos favorables dentro de las opciones de técnicas empleadas, aunque superior al caso SA.

3.3.2 Caso de estudio sin contingencia ambiental

Los resultados obtenidos para el periodo del 4 al 7 de mayo de 2013 (2013-05-04 al 2013-05-07), en el cual no hubo ocurrencia de ninguna contingencia ambiental en la ZMVM, se muestran a continuación.

En la Figura 23 aparecen las series temporales de la variable T2 en cuanto a sus valores observados y modelados a partir de las técnicas de asimilación de datos que se emplearon en esta investigación. En general, todas las técnicas representan la distribución temporal de T2 acorde a su ciclo diario y a las observaciones de las estaciones disponibles en el periodo de interés, aunque el uso de 4DVAR conllevó a un desempeño del modelo ligeramente superior, sobre todo en los valores extremos. Se aprecia, nuevamente, una subestimación por parte del WRF-Chem de los valores máximos observados y una sobreestimación de los picos mínimos de la variable, principalmente en las primeras horas de la simulación.

Se presentan además las series temporales de T2 para las distintas estaciones de la RAMA analizadas, en cuanto a los valores observados en las mismas y modelados con WRF-Chem, empleando las técnicas de asimilación ya mencionadas. Se observa un comportamiento similar a lo descrito anteriormente, siendo la opción 4DVAR la que resulta en un mejor desempeño para el modelo. Las estaciones TLA, XAL y MER tienen mejor correspondencia entre las simulaciones y las observaciones, mientras que, para TPN, el modelo presenta un peor desempeño.

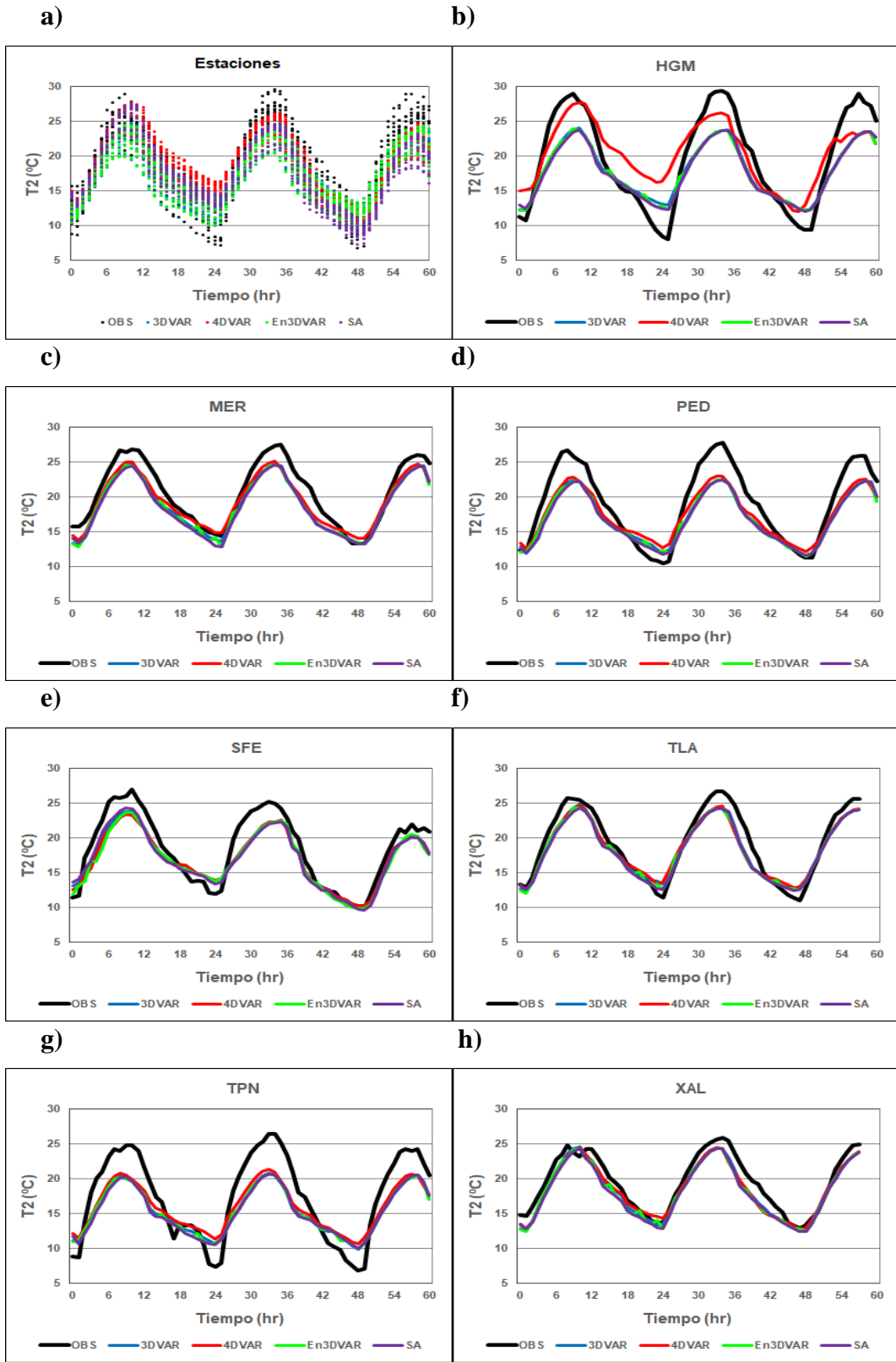


Figura 23. Series temporales de los valores observados y modelados de T2 para el conjunto de todas las estaciones y las seleccionadas, a partir de las diferentes técnicas de asimilación de datos

La Tabla X muestra los valores de los estadígrafos calculados para cada experimento realizado, según la técnica de asimilación de datos que se aplica, presentando datos promedio y para cada estación de monitoreo analizada.

Tabla X. Valores de IOA, BIAS, RMSE y ρ para la variable T2 y para las técnicas de asimilación empleadas: SA, 3DVAR, 4DVAR y En3DVAR

		SA	3DVAR	4DVAR	En3DVAR
HGM	IOA	0.93	0.94	0.94	0.94
	BIAS	-1.67	-1.58	-1.43	-1.70
	RMSE	2.43	2.51	2.39	2.39
	ρ	0.94	0.93	0.94	0.93
MER	IOA	0.93	0.94	0.95	0.94
	BIAS	-1.78	-1.30	-1.19	-1.27
	RMSE	2.51	2.29	2.16	2.23
	ρ	0.93	0.94	0.95	0.95
PED	IOA	0.92	0.93	0.94	0.94
	BIAS	-1.84	-1.49	-1.29	-1.59
	RMSE	2.51	2.37	2.26	2.42
	ρ	0.93	0.92	0.95	0.94
SFE	IOA	0.93	0.94	0.95	0.94
	BIAS	-1.63	-1.42	-1.28	-1.68
	RMSE	2.47	2.44	2.31	2.41
	ρ	0.93	0.93	0.94	0.94
TLA	IOA	0.94	0.95	0.95	0.94
	BIAS	-1.62	-1.27	-1.18	-1.26
	RMSE	2.49	2.25	2.11	2.29
	ρ	0.94	0.95	0.96	0.95
TPN	IOA	0.93	0.93	0.93	0.94
	BIAS	-1.85	-1.75	-1.85	-1.72
	RMSE	2.53	2.54	2.50	2.41
	ρ	0.93	0.93	0.92	0.94
XAL	IOA	0.94	0.94	0.96	0.96
	BIAS	-1.78	-1.33	-1.14	-1.18
	RMSE	2.46	2.35	2.11	2.20
	ρ	0.93	0.94	0.96	0.95
PROMEDIO	IOA	0.94	0.94	0.96	0.95
	BIAS	-1.61	-1.58	-1.04	-1.16
	RMSE	2.41	2.37	2.16	2.22
	ρ	0.94	0.95	0.96	0.95

De modo general, la técnica 4DVAR muestra los mejores resultados con valores de IOA y R de 0.96, similares al caso con contingencia ambiental, ratificando que existe buena correspondencia y relación lineal entre las simulaciones y las observaciones, aunque el RMSE = 2.16 es ligeramente más favorable que el caso anterior y, además, respecto al resto de las técnicas. Los valores de BIAS negativos, para todas las opciones, indican que el modelo tiende a subestimar las mediciones. El caso SA presenta los resultados más discretos, mientras que 3DVAR lo hace para las opciones de técnicas disponibles.

De igual forma, se obtuvieron resultados para la variable química concentración de O₃ para este caso de estudio sin contingencia ambiental. En la Figura 24 se muestran las series temporales de los valores observados de O₃ y simulados con cada una de las técnicas de asimilación de datos disponibles. Para todos los casos, el modelo representa de forma satisfactoria la distribución temporal de O₃ en el área de estudio, así como su ciclo diario, apreciándose además una ligera superioridad cualitativa de la técnica 4DVAR respecto al resto de las opciones disponibles.

Se muestran también las series temporales de O₃ para estaciones seleccionadas. Se observa que la técnica 4DVAR describe un comportamiento de la variable más próximo a los valores de las observaciones. El modelo WRF-Chem tuvo un mejor desempeño en las estaciones XAL y SFE, según se aprecia en las gráficas, mientras que TPN y TLA, tienen menor correspondencia entre las observaciones y las simulaciones.

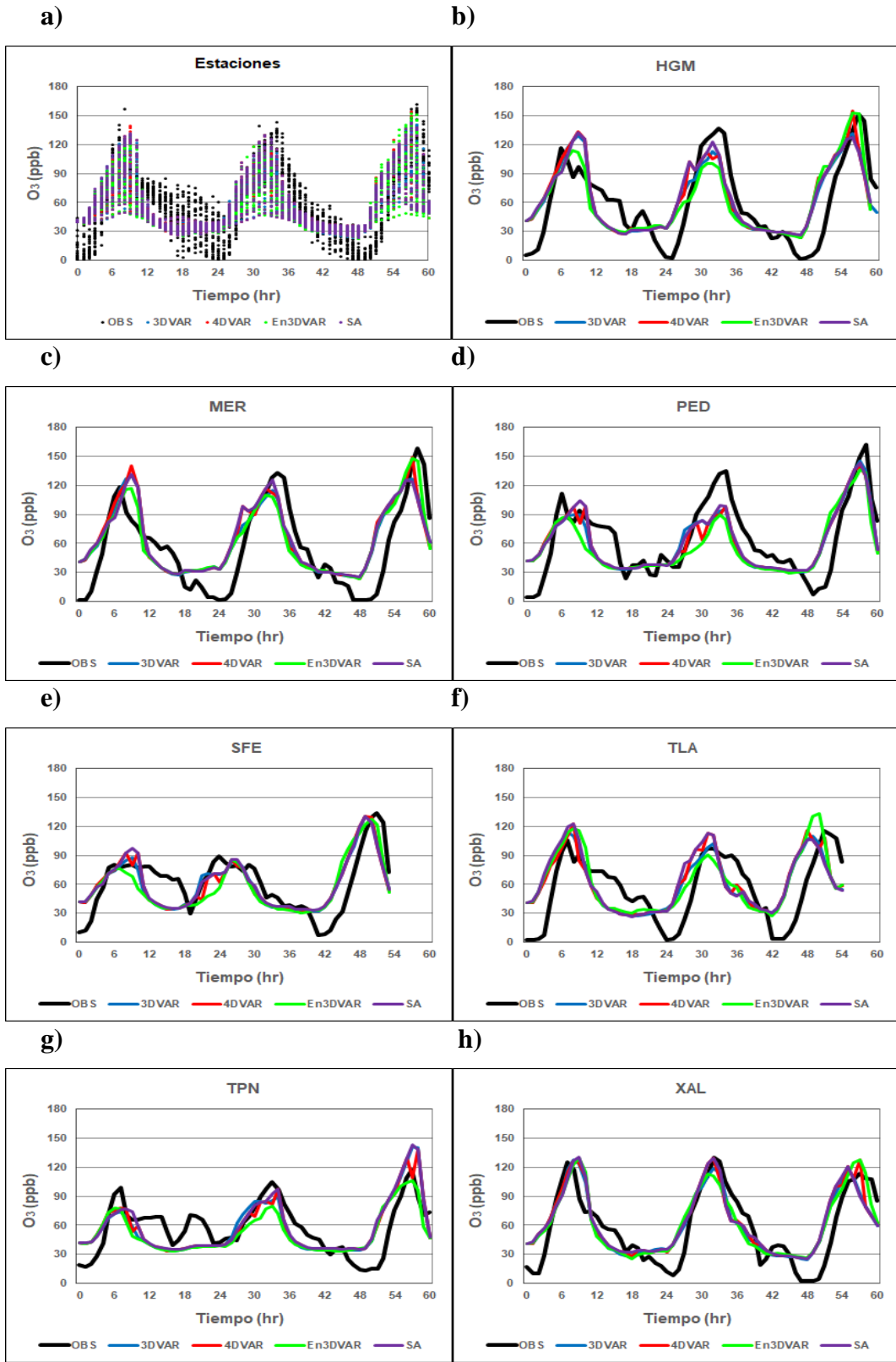


Figura 24. Series temporales de los valores observados y modelados de O₃ para el conjunto de todas las estaciones y las seleccionadas, a partir de las diferentes técnicas de asimilación de datos

Como complemento de lo anterior, en la Tabla XI se presentan los valores obtenidos para cada una de las métricas estadísticas calculadas en los casos de estudio considerados, según la técnica de asimilación de datos aplicada en cada uno de ellos.

Tabla XI. Valores de IOA, BIAS, RMSE y ρ para la variable O_3 y para las técnicas de asimilación empleadas: SA, 3DVAR, 4DVAR y En3DVAR

		SA	3DVAR	4DVAR	En3DVAR
HGM	IOA	0.76	0.75	0.80	0.77
	BIAS	-7.02	-8.15	-5.76	-6.68
	RMSE	23.74	25.09	22.18	23.36
	P	0.70	0.69	0.79	0.73
MER	IOA	0.75	0.74	0.79	0.76
	BIAS	-7.42	-7.73	-5.86	-6.49
	RMSE	24.36	25.64	22.28	23.52
	P	0.69	0.67	0.79	0.74
PED	IOA	0.74	0.75	0.79	0.75
	BIAS	-7.31	-7.18	-6.09	-6.76
	RMSE	24.79	24.65	23.12	23.91
	P	0.70	0.70	0.77	0.73
SFE	IOA	0.80	0.80	0.84	0.82
	BIAS	-6.32	-6.14	-4.64	-5.37
	RMSE	23.56	23.21	19.75	21.53
	P	0.76	0.77	0.83	0.79
TLA	IOA	0.71	0.73	0.77	0.75
	BIAS	-8.73	-7.69	-6.42	-7.37
	RMSE	29.16	26.97	23.94	24.61
	P	0.66	0.69	0.73	0.70
TPN	IOA	0.70	0.72	0.76	0.73
	BIAS	-9.14	-8.09	-6.72	-7.60
	RMSE	29.63	26.56	23.78	25.44
	P	0.65	0.68	0.73	0.68
XAL	IOA	0.81	0.80	0.85	0.84
	BIAS	-5.58	-5.94	-4.52	-4.85
	RMSE	22.02	22.60	19.07	20.59
	P	0.79	0.78	0.83	0.81
PROMEDIO	IOA	0.75	0.73	0.80	0.77
	BIAS	-7.24	-8.26	-5.69	-6.54
	RMSE	24.52	26.21	21.95	23.38
	P	0.70	0.66	0.79	0.73

La técnica 4DVAR presenta valores cualitativamente superiores, con IOA = 0.80 y R = 0.79, los cuales comprueban que existe buena correspondencia y relación lineal entre las simulaciones numéricas y las observaciones. Esto se complementa, de manera favorable, con un RMSE = 21.95, el más bajo de los obtenidos. Además, los valores negativos de BIAS indican que el modelo subestima las mediciones, aunque en un rango aceptable para la escala de la variable. El caso 3DVAR presenta los resultados menos favorables, encontrándose por debajo de los promedios obtenidos en cada uno de los estadísticos calculados, incluso al compararlo con el caso SA.

3.4 Rendimiento numérico de los métodos de asimilación de datos utilizados en el módulo WRFDA

Con el objetivo de tener una medida del comportamiento numérico de las técnicas de asimilación de datos empleadas, en esta sección se analizan algunos parámetros computacionales sugeridos para ello en el Manual de Usuario del módulo WRFDA (https://www2.mmm.ucar.edu/wrf/users/docs/user_guide_v4/v4.0/users_guide_chap6.html) En este caso, se consideran las funciones de costo y gradiente (Kalnay, 2003).

Por otra parte, se muestra el comportamiento de los incrementos de análisis desde un enfoque diferente a la Sección 3.1. En lugar de la distribución espacial de VV, DD y T2, dentro del dominio de modelación, aquí se presentan los datos de algunas variables de estado análogas y que son modificadas en el proceso de asimilación con el módulo WRFDA: u (componente zonal del viento), v (componente meridional del viento) y t (temperatura). Los valores de estas variables las describen en cada nivel vertical de la atmósfera considerado por WRF-Chem. Para el cálculo se tiene en cuenta el período sin contingencia ambiental del 4 al 7 de mayo de 2013 (2013-05-04 a 2013-05-07).

En la Figura 25 aparecen los valores de las funciones de costo y gradiente, calculados como el promedio de cada una de estas variables en el total de ciclos de asimilación que se realizaron durante el período de tiempo considerado. Los resultados obtenidos se muestran para cada una de las iteraciones y en cada método empleado en el proceso de asimilación de datos.

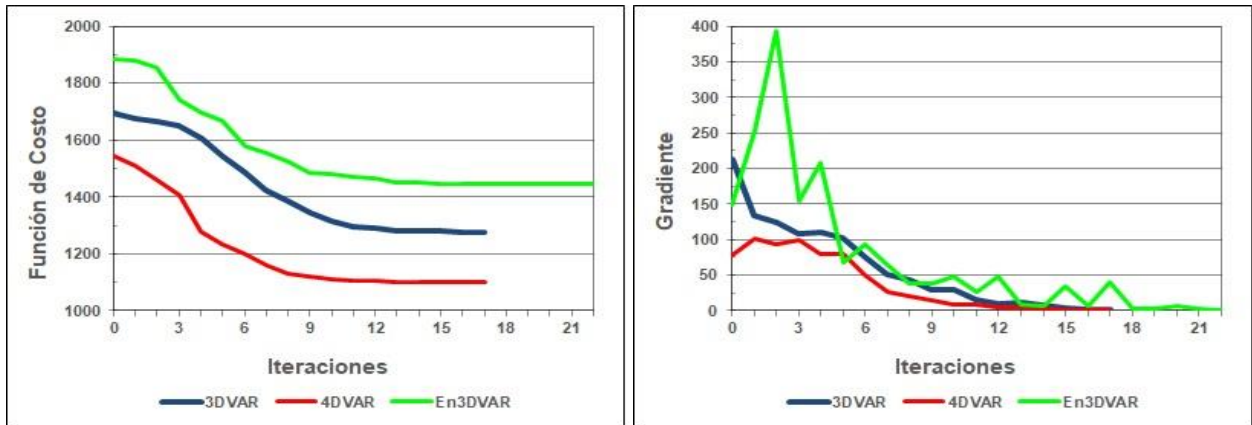


Figura 25. Funciones de costo (a) y gradiente (b) para cada método de asimilación de datos

Se puede observar que 4DVAR es el método que converge mejor a la solución y, por ende, presenta un menor valor de la función de costo, además de que el gradiente tiende a cifras más cercanas a cero. El peor desempeño lo tiene En3DVAR, en la cual se necesita un mayor número de iteraciones, 22 en total, para converger a la solución.

En la Figura 26 aparecen los valores de los incrementos de análisis calculados como el promedio de la variable en el total de ciclos de asimilación y en los 29 niveles verticales del modelo WRF-Chem y para u, v y t.

En el caso de t, los incrementos de análisis son mayores, modularmente, en los niveles más bajos de la atmósfera, lo cual puede relacionarse con que los campos de las variables meteorológicas en la capa límite se comportan de un modo más irregular y, por ende, el *background* puede no representar el estado atmosférico real existente. Estos incrementos sugieren una disminución en los valores de las condiciones iniciales de la variable, llegando a ser de -1.2°C como cifra promedio extrema con la técnica 3DVAR, mientras que las correcciones positivas que se proponen son ligeras y apenas alcanzan los 0.4°C . Lo anterior indica que los campos de primera aproximación tienden a sobreestimar los valores de t y pueden ser corregidos con la asimilación de datos.

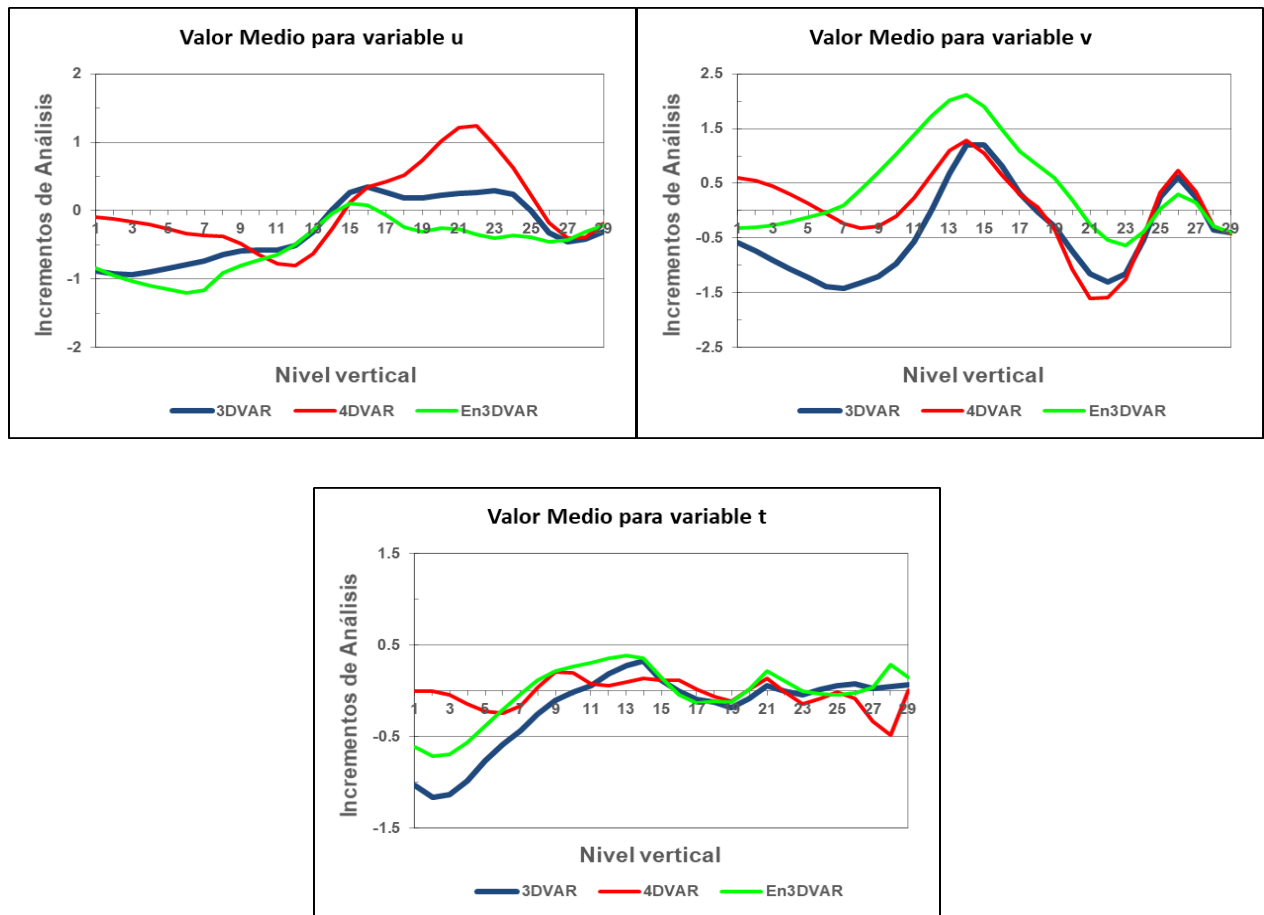


Figura 26. Valores medios de los incrementos de análisis de las variables u (a), v (b) y t (c) para los diferentes niveles verticales calculados por el modelo WRF-Chem

En sentido general, el método 4DVAR sugiere los menores cambios desde el punto de vista cuantitativo para la variable t, pues este tiene en cuenta, de forma consistente, los balances de calor, energía y humedad durante el proceso de minimización. Por tanto, se puede esperar que la curva que describe los incrementos de análisis propuestos por esta técnica tenga menores gradientes a lo largo de los niveles verticales de la atmósfera. Aquí influye, además, la manera en que se calculan las matrices de errores de covarianza del *background* y de las observaciones, lo cual se explicó anteriormente.

Por su parte, las componentes del viento, u y v, presentan comportamientos diferentes entre sí. En el caso de u, los incrementos de análisis son negativos en los niveles bajos de la atmósfera, sugiriendo una disminución de los valores de la variable, incluso, en más de 1 m/s.

A partir del nivel 14, los incrementos pasan a ser positivos en 3DVAR y 4DVAR y, para la técnica En3DVAR, se mantienen inferiores a 0m/s. En cuanto a la variable de estado v , los incrementos se mueven, indistintamente, en rangos positivos y negativos a lo largo de los diferentes niveles de la atmósfera, siendo el método En3DVAR el que propone los mayores cambios, superando los 2 m/s en el nivel 14.

De forma complementaria, en la Figura 27 se muestran los valores medios de RMSE en todos los niveles verticales calculados y para las variables u , v y t . Se observa que, para las componentes zonal y meridional del viento, el método En3DVAR muestra los valores más altos de esta métrica estadística en los primeros 14 niveles, alcanzando incluso los 2.5 m/s, mientras que 4DVAR presenta las cifras más bajas, confirmando que es la técnica con mejores resultados en la presente investigación.

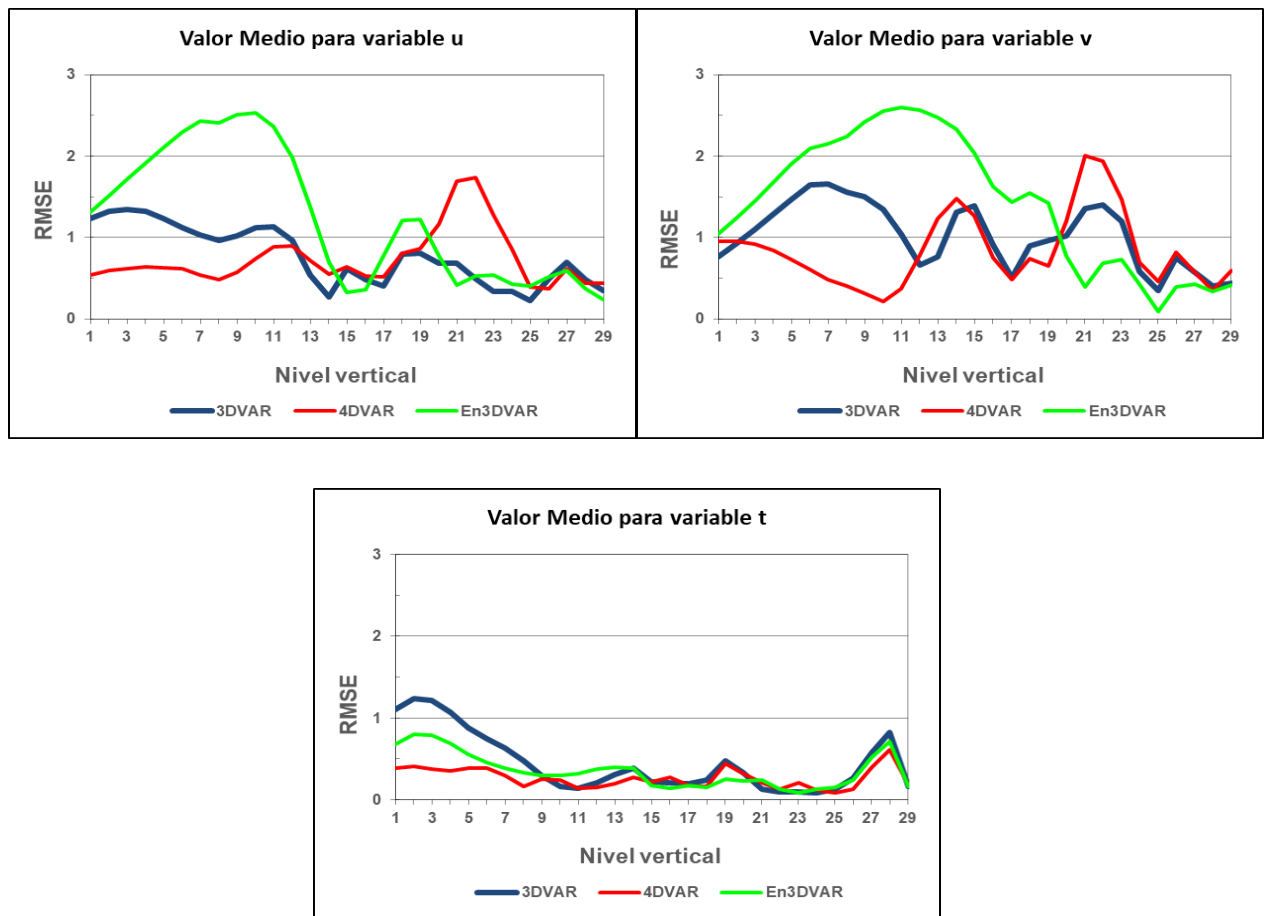


Figura 27. Valores medios de RMSE de las variables u (a), v (b) y t (c) para los diferentes niveles verticales calculados por el modelo WRF-Chem

En el caso de t , los tres métodos se comportan de forma muy similar, con un RMSE inferior a 1 °C en casi todos los niveles, aunque en aquellos más cercanos a la superficie, la técnica 4DVAR mantiene un desempeño superior respecto a las demás.

En la Tabla XII se presentan algunas cifras que describen el costo computacional asociado a cada una de las técnicas utilizadas, para un solo ciclo de asimilación, esto es, el tiempo de ejecución empleado y el espacio en memoria necesario para almacenar sus salidas. Se observa que, aunque 4DVAR es el método que permite obtener un mejor desempeño del modelo WRF-Chem, como se ha visto, es también el que requiere un mayor costo computacional, dadas las características propias de la técnica, siendo entonces En3DVAR la que muestra mejor balance costo-beneficio.

Tabla XII. Costo computacional asociado a las técnicas de asimilación de datos utilizadas para un solo ciclo

Métodos	3DVAR	4DVAR	En3DVAR (10 miembros del ensamble)
Tiempo de ejecución	240 seg	148,000 seg	2,500 seg
Espacio en memoria	0.8 GB	1.7 GB	3.7 GB

3.5 Resultados de las simulaciones numéricas con el sistema WRF-Chem/DART

Primeramente, con el propósito de validar la correcta implementación del sistema WRF-Chem/DART, se realizaron simulaciones numéricas para el caso de Prueba, descrito en la Sección 2.3 (ver Apéndice I). Una vez validado, se llevó a cabo el mismo procedimiento para el caso de Prueba-México, teniendo como objetivo principal el de comprobar que los cambios realizados seguían siendo consistentes con el sistema WRF-Chem/DART. A continuación, se describen los principales resultados obtenidos, los cuales están basados en la comparación de las simulaciones con asimilación de datos químicos y meteorológicos respecto a las salidas numéricas sin asimilación.

En las Figuras 28, 29 y 30, se muestran las distribuciones espaciales de las variables CO, O₃ y T₂, respectivamente, representadas a través de la media del ensamble en el primer paso de pronóstico, en cada caso, para las corridas (a) con asimilación de datos químicos procedentes de MOPITT y meteorológicos de PREPBUFR GDAS (*Global Data Assimilation System*) y (b) sin asimilación de datos, así como el campo de diferencias (incrementos de observación), entre ambos casos considerados (c).

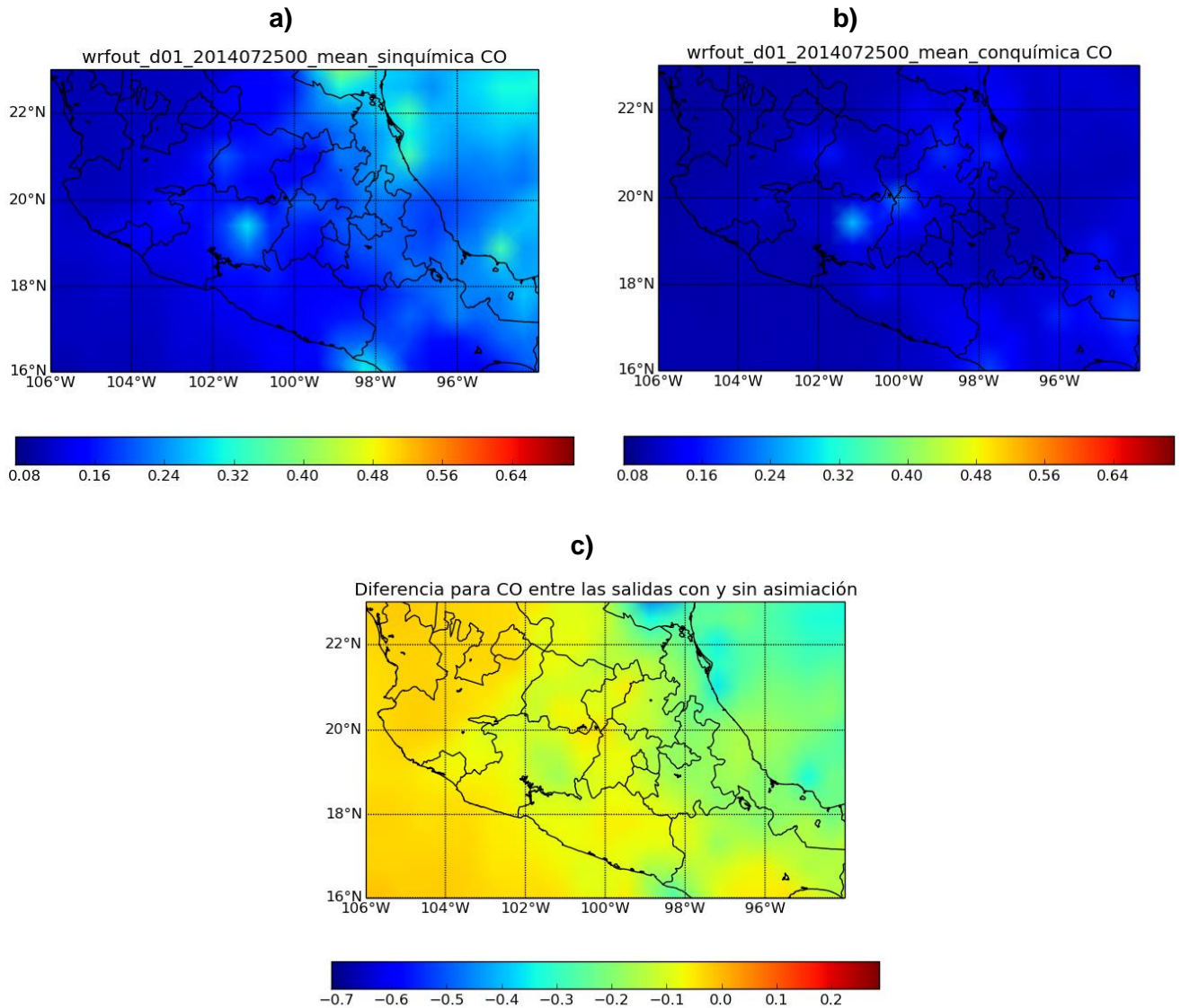


Figura 28. Distribución espacial de CO (media del ensamble en el primer paso de pronóstico) para el dominio Centro México con asimilación de datos químicos de MOPITT (a), sin asimilación (b) e incrementos de observación (c)

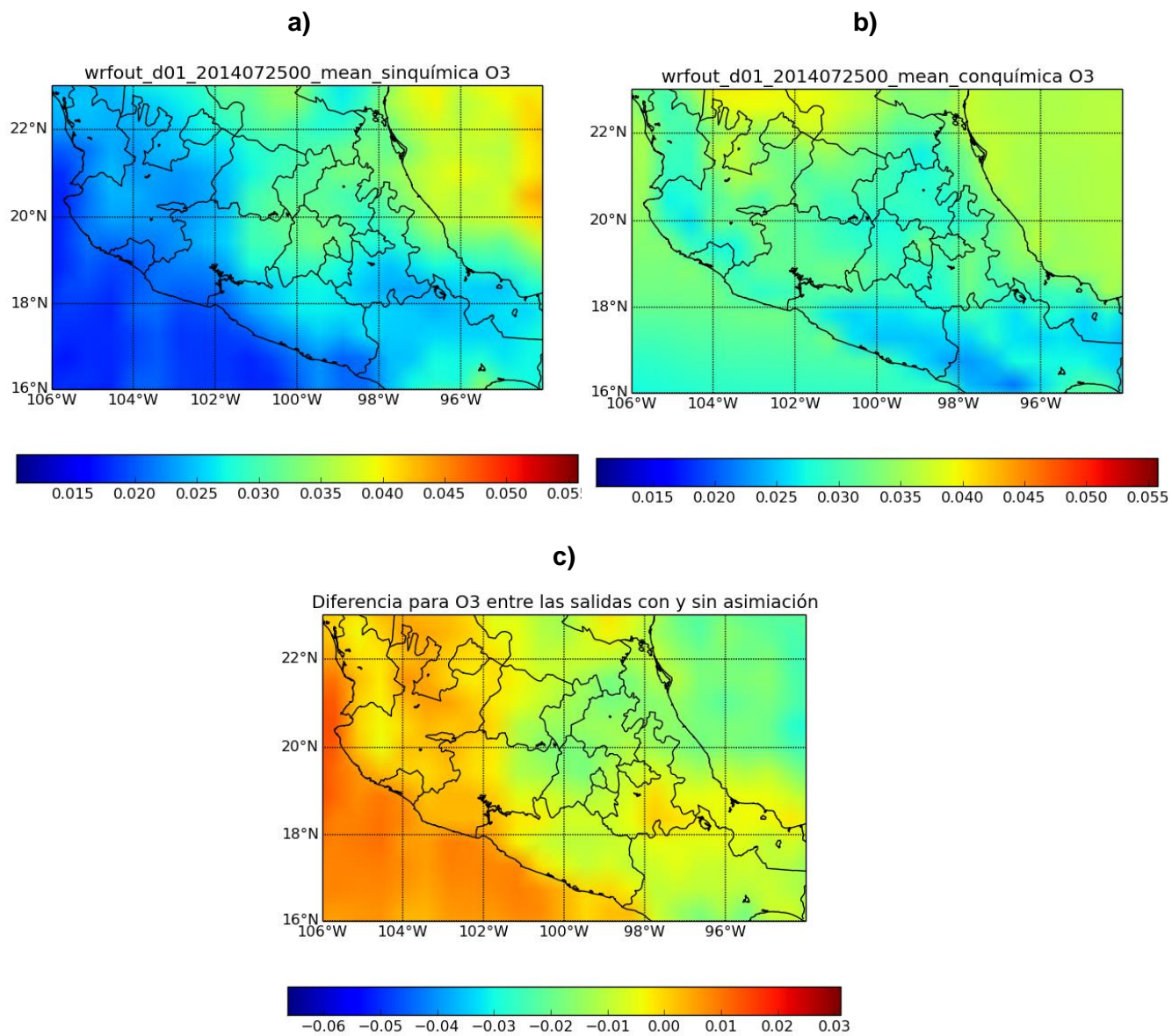


Figura 29. Distribución espacial de O₃ (media del ensamble en el primer paso de pronóstico) para el dominio Centro de México con asimilación de datos químicos de MOPITT (a), sin asimilación (b) e incrementos de observación (c)

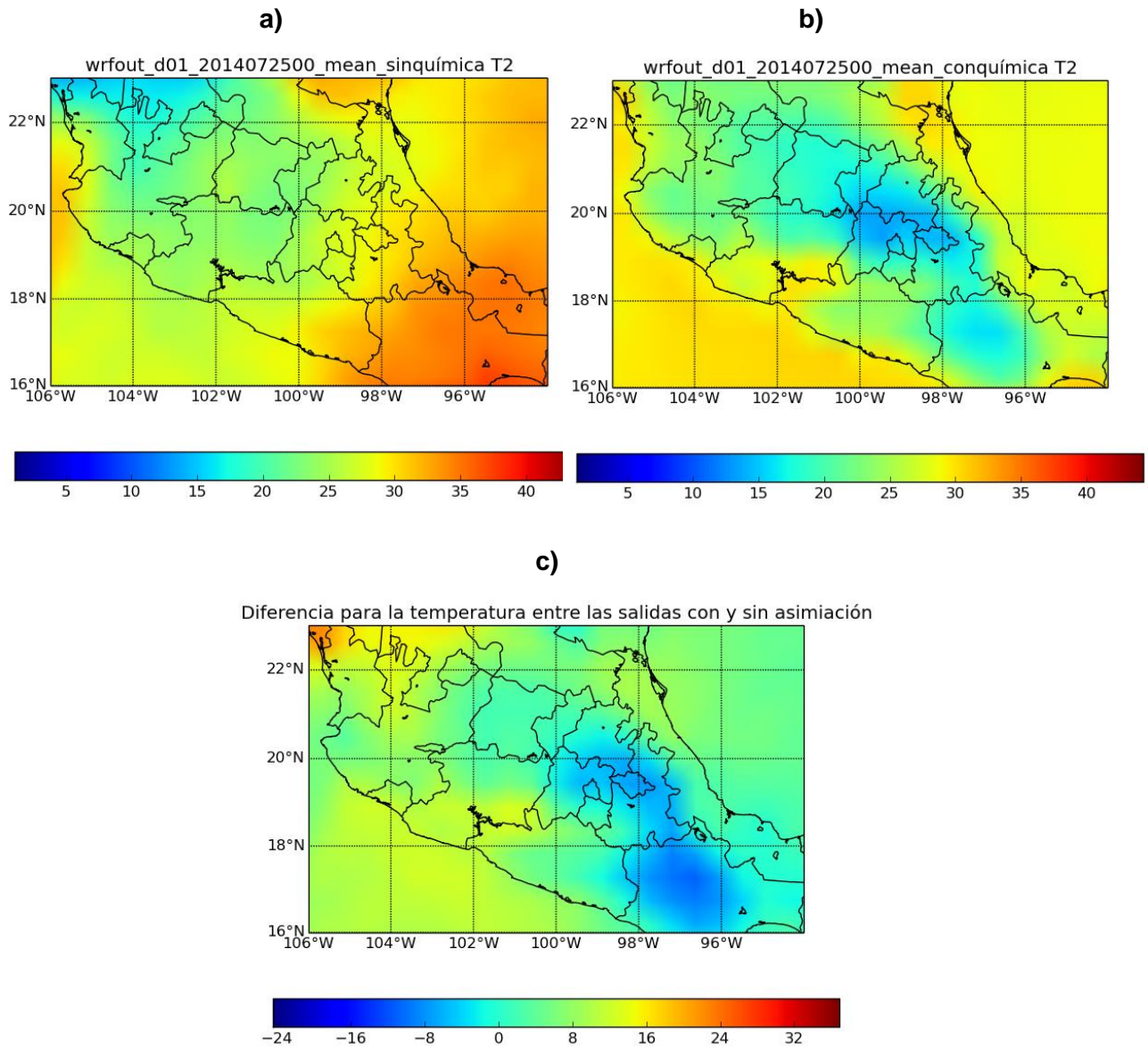


Figura 30. Distribución espacial de T2 (media del ensemble en el primer paso de pronóstico) para el dominio Centro de México con asimilación de datos químicos de MOPITT (a), sin asimilación (b) e incrementos de observación (c)

En general, para todos los casos, se observa que sí existen diferencias entre todas las simulaciones realizadas, con y sin asimilación de datos, principalmente en la variable T2. Encontrar estas diferencias era, precisamente, el objetivo esencial de esta etapa de la investigación e indica que la asimilación de datos sí influye en los resultados de la modelación numérica.

Lo anterior puede comprobarse nuevamente en la Figura 31, en la cual se presentan los campos correspondientes a las diferencias entre las condiciones iniciales de las variables CO (a) y O₃ (b), a modo de ejemplo, obtenidas de restar los campos pertenecientes al primer ciclo de asimilación, en cada caso, respecto al ciclo inicial.

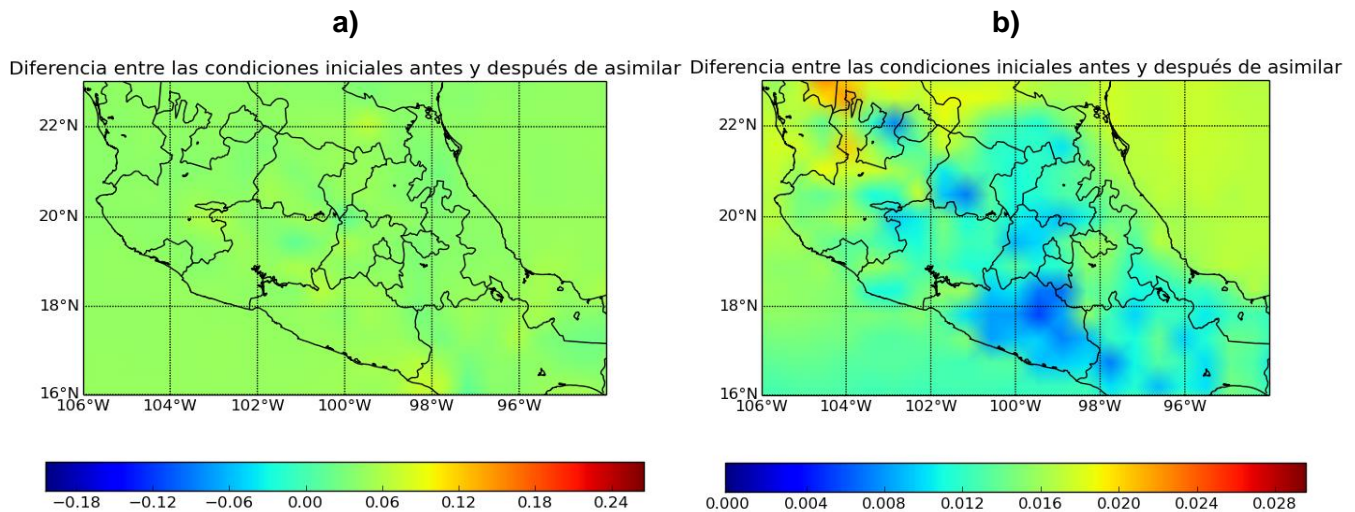


Figura 31: Campos de diferencias entre las condiciones iniciales antes de después de la asimilación de datos para las variables CO (a) y O₃ (b)

Se observa que, para todos los casos mostrados, sí existen diferencias entre los campos que se indican, lo cual está en correspondencia con lo obtenido por Mizzi (2020) y descrito en Mitzi et al. (2016)

Por otra parte, con el objetivo de comprobar la influencia que tiene el proceso de asimilación de datos químicos procedentes de MOPITT (específicamente la variable CO) en el desempeño del modelo WRF-Chem, se realizó una evaluación del mismo para el caso de Prueba-México, a partir del cálculo de las métricas estadísticas anteriormente empleadas. Como se observa en la Figura 32, la distribución espacial de los perfiles verticales de este sensor para el momento en que se inicia el primer ciclo de asimilación (2014-07-25 a las 00 UTC) muestra que no se cubre el área de interés de la presente investigación, es decir, el Centro de México.

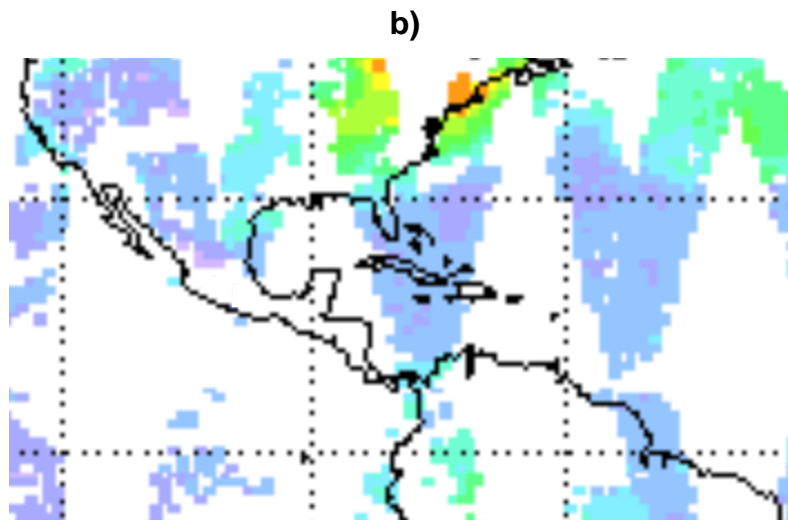
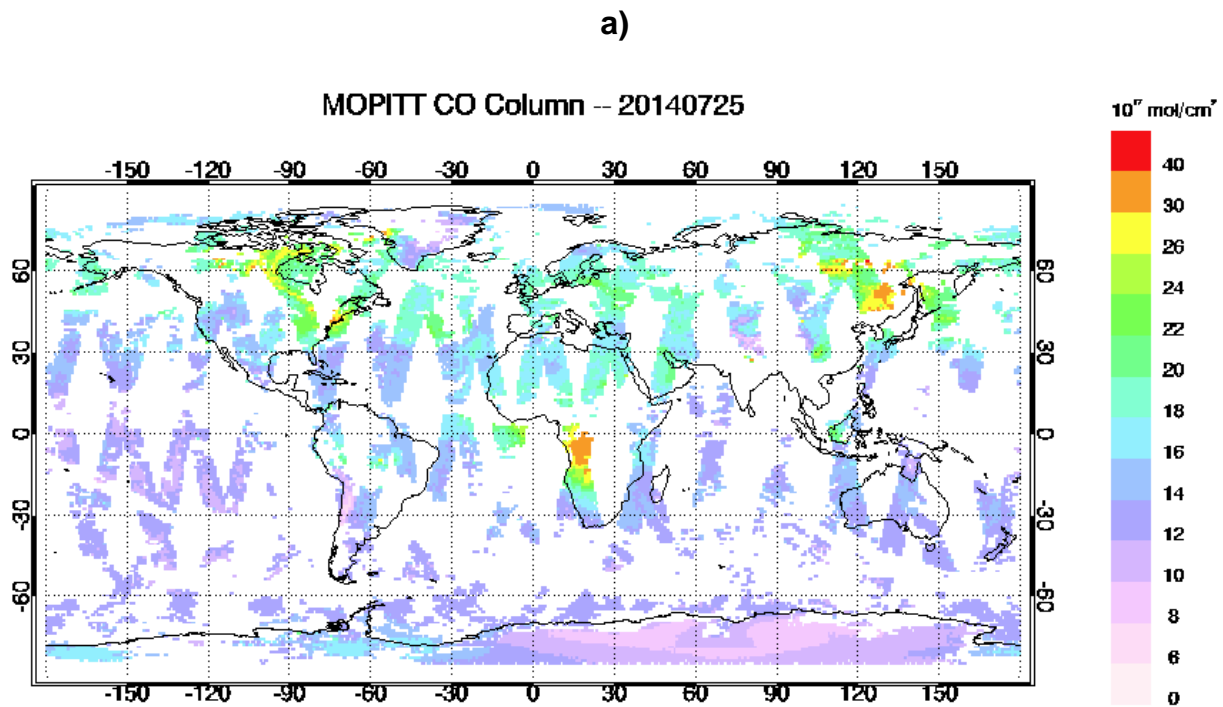


Figura 32. Distribución espacial de los perfiles verticales de MOPITT CO para el momento en que se inicia el primer ciclo de asimilación (20140725 a las 00 UTC), mapa completo (a) y acercamiento al dominio de modelación (b)

Al no existir observaciones dentro del dominio o cercanas a este, los dos parámetros que determinan aquellas consideradas durante el proceso de minimización, es decir, el radio de influencia (120 km) y la localización (asociada al método EAKF en el módulo DART) no tienen influencia sobre los cambios producidos en el campo de análisis que se obtiene luego de la asimilación de datos en el dominio de modelación que abarca el Centro de México. Debido a que el dominio de modelación del caso Prueba-México cubre todo el país, se decide elegir una región incluida en el barrido del satélite con observaciones que se puedan asimilar. Para tener un criterio consistente, se realiza una evaluación comparativa entre las salidas del modelo WRF-Chem con respecto a las observaciones de las estaciones disponibles entre el 24 y el 25 de julio de 2014 (2014-07-24 y 2014-07-25), a modo de ejemplo, en la Red de Monitoreo del Área Metropolitana de Monterrey.

En la Tabla XIII se muestra el promedio de los resultados de las métricas estadísticas calculadas para la variable química CO en comparación con un caso sin asimilación de datos, para las estaciones de monitoreo de Monterrey, mencionadas en la Sección 2.6 y representadas en la Figura 8.

Tabla XIII. Valores de IOA, BIAS, RMSE y ρ para la variable CO en simulaciones numéricas con asimilación de datos MOPITT CO y sin asimilación para estaciones de monitoreo en Monterrey (24 al 25 de julio de 2014)

	SA	MOPITT CO
IOA	0.56	0.68
BIAS	-67.95	-55.17
RMSE	173.37	67.58
P	0.47	0.60

Los resultados son superiores para el caso de asimilación con MOPITT con IOA = 0.68 y ρ = 0.60, indicando una buena correspondencia entre los valores modelados y las observaciones de O₃. El valor del RMSE es relativamente alto, pero acorde con la escala de la variable CO y logra disminuirse de 173.37 a 67.58, mientras que el BIAS = -55.17 indica que el modelo tiende a subestimar las mediciones, pero de un modo menos notable que en el caso SA.

En el trabajo realizado por Chamu (2020) se encontró una deriva en los datos de CO de las estaciones de monitoreo en Monterrey la cual no se tuvo en cuenta en los resultados de la presente investigación. Además, los resultados en el IOA siempre fueron inferiores a 0.5, mientras que los valores RSME estuvieron en un orden superior a los mostrados en la Tabla XI. De esta forma, se evidencia la influencia positiva que tiene la asimilación de datos químicos y meteorológicos, conjuntamente, en un área de estudio determinada. El ajuste en el inventario de emisiones de Monterrey es otro punto importante a considerar en el rendimiento del modelo y que no está optimizado.

3.6 Resultados de la modelación con el sistema WRF-Chem/DART con asimilación de datos procedentes de observaciones de la RAMA

A continuación, se presentan los resultados de las simulaciones numéricas con el sistema WRF-Chem/DART para el dominio de modelación principal, correspondiente al caso de estudio Centro de México (descrito en la Sección 2.3), el cual fue utilizado también en WFRDA. La particularidad de esta sección radica, además del cambio de dominio, en la inclusión de las observaciones puntuales procedentes de las estaciones de monitoreo atmosférico de la RAMA, como parte de los datos que son asimilados.

En las Figuras 33, 34, 35 y 36, aparecen los incrementos de análisis obtenidos para las variables T2, CO, O₃ y SO₂, respectivamente, en los ciclos de asimilación: segundo (2014-07-25, 06 UTC - 01 horas tiempo local), cuarto (2014-07-25, 18 UTC - 13 horas tiempo local), sexto (2014-07-26, 06 UTC - 01 horas tiempo local) y octavo (2014-07-26, 12 UTC - 13 horas tiempo local).

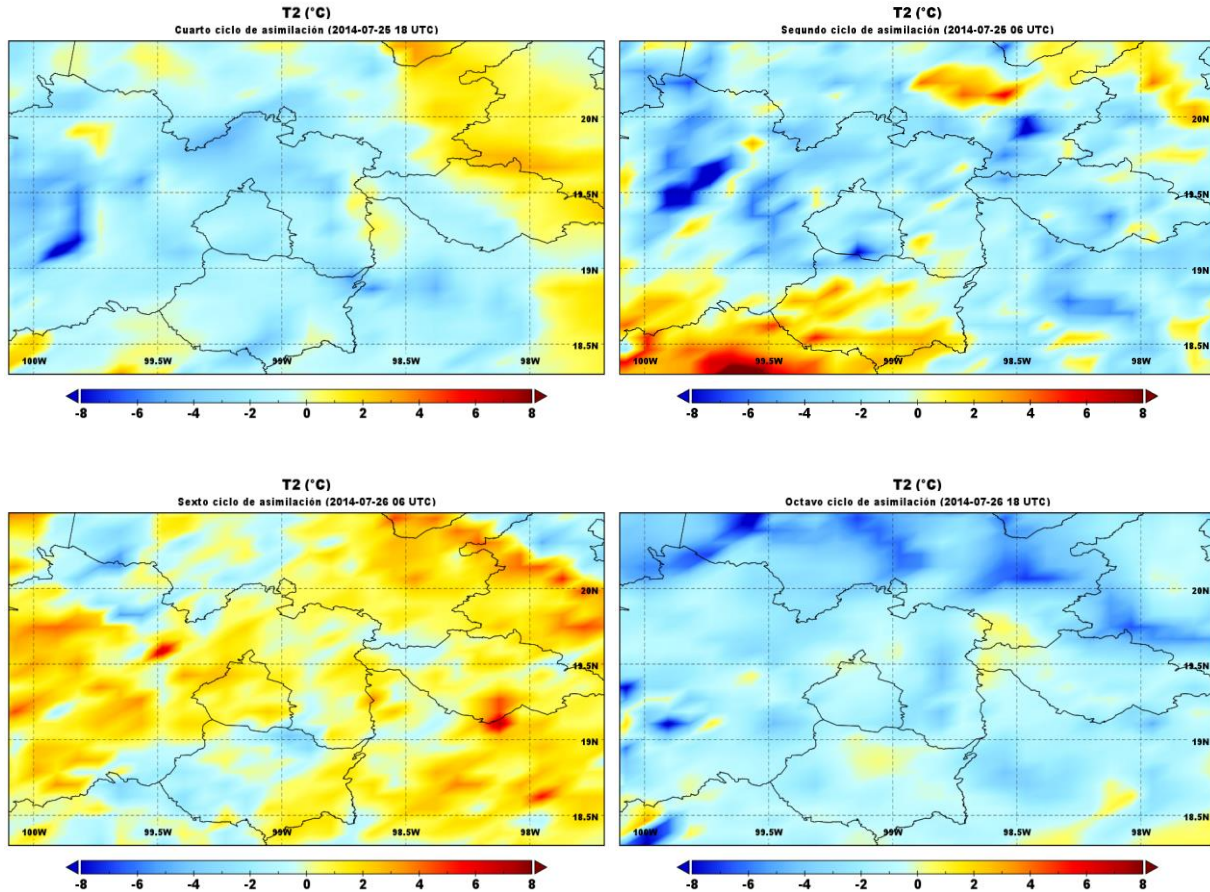


Figura 33. Incrementos de análisis para la variable T2 en el segundo, cuarto, sexto y octavo ciclos de asimilación con el sistema WRF-Chem/DART

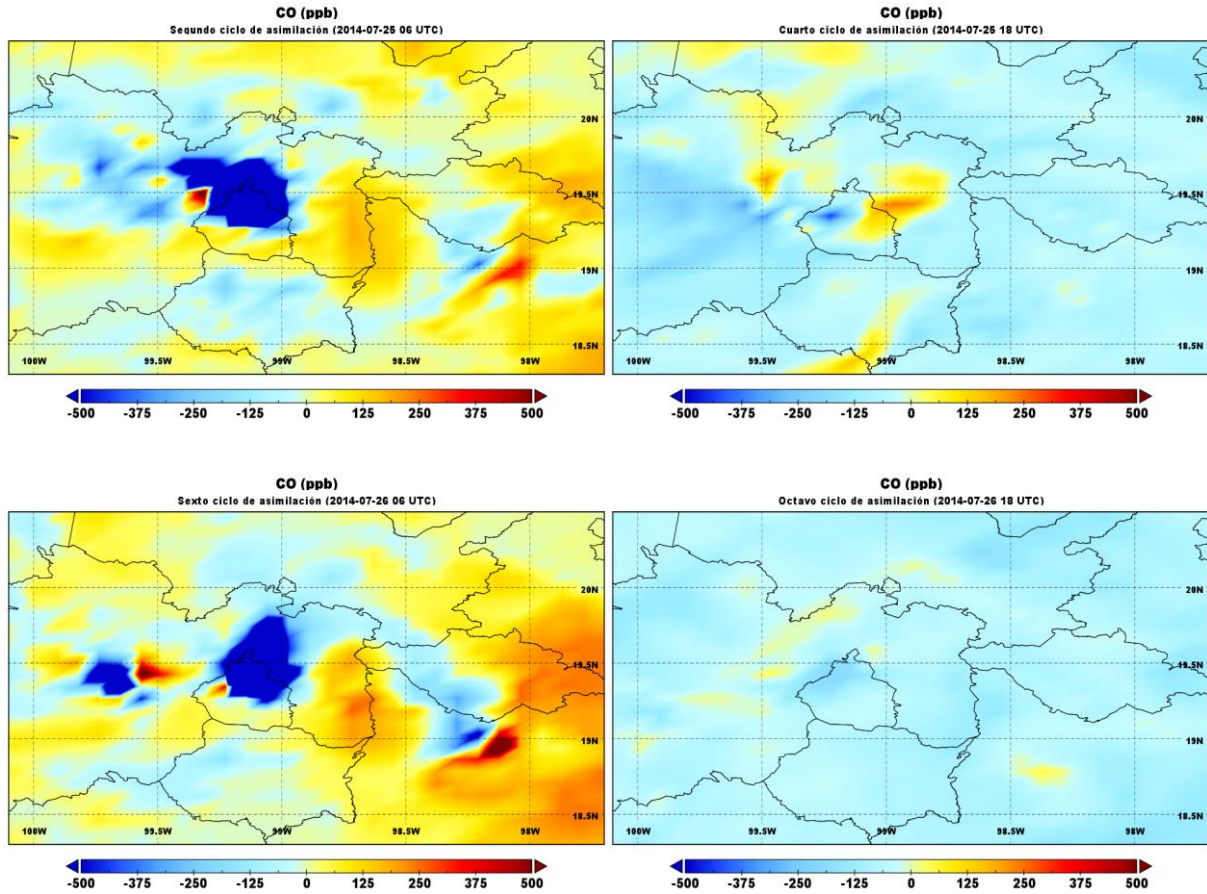


Figura 34. Incrementos de análisis para la variable CO en el segundo, cuarto, sexto y octavo ciclos de asimilación con el sistema WRF-Chem/DART

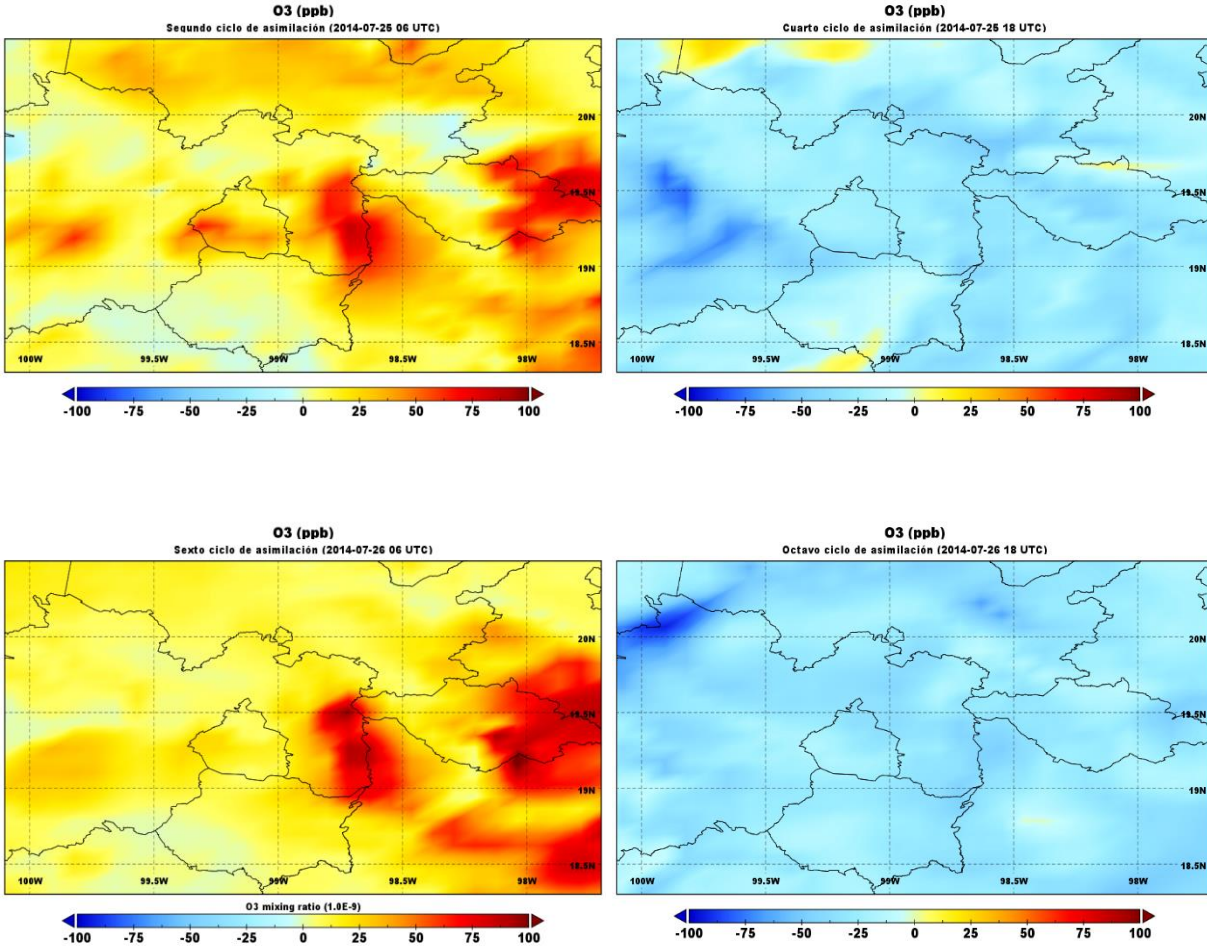


Figura 35. Incrementos de análisis para la variable O₃ en el segundo, cuarto, sexto y octavo ciclos de asimilación con el sistema WRF-Chem/DART

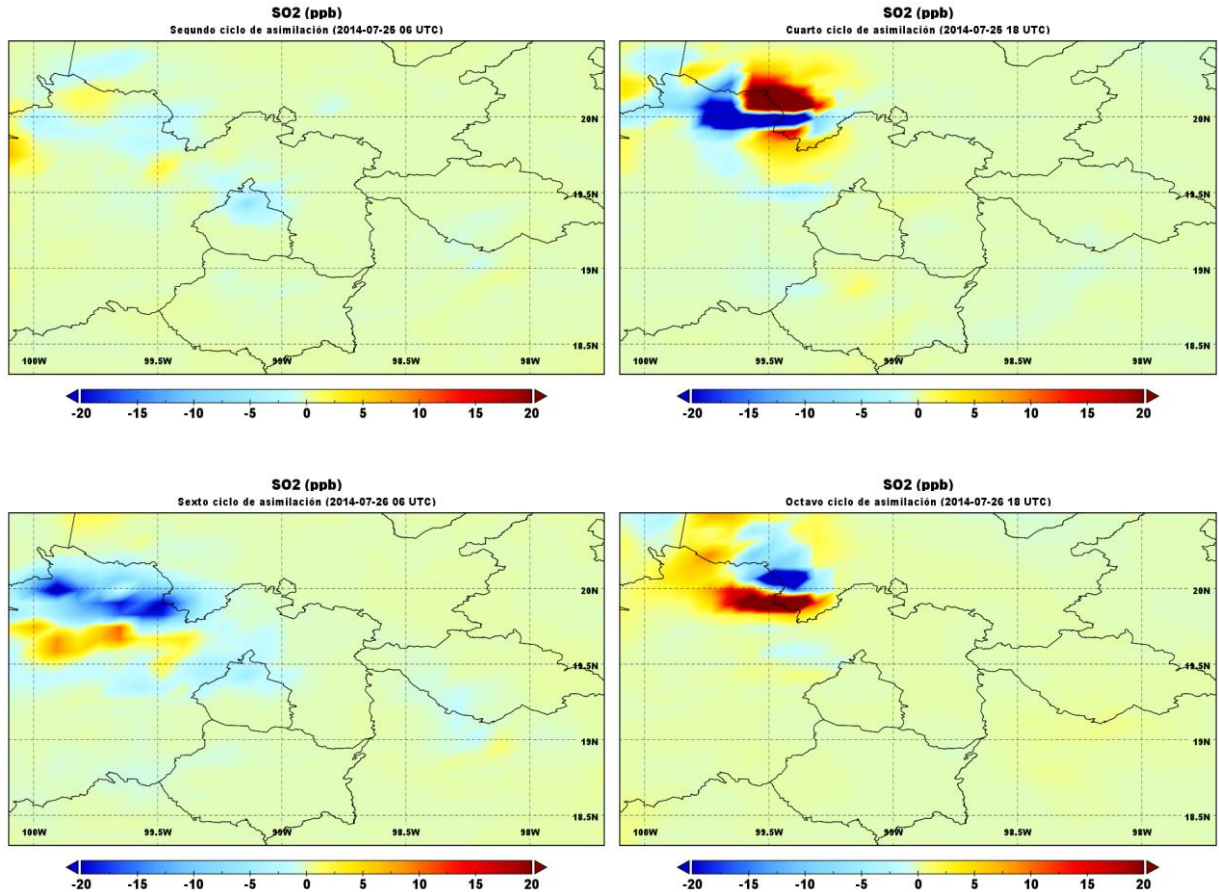


Figura 36. Incrementos de análisis para la variable SO₂ en el segundo, cuarto, sexto y octavo ciclos de asimilación con el sistema WRF-Chem/DART

Según se puede observar, para el caso de T2, los incrementos de análisis proponen disminuir los valores en gran parte del dominio de modelación, en un orden de magnitud que se encuentra en concordancia con la escala característica de la variable. En algunas áreas, se alcanzan valores extremos superiores a -12 °C de correcciones, principalmente en las zonas montañosas. Esto demuestra que los datos obtenidos de GFS no siempre describen los valores de T2 en este tipo de relieve, comportamiento similar al descrito en el caso de NARR para la inicialización de WRFDA. La resolución espacial de las salidas de GFS (0.25 grados) es otro factor que influye en que los campos de análisis presenten las diferencias antes mencionadas, ya que al ser mucho menor que la del dominio de modelación no representa de forma adecuada el nivel de detalle de los procesos a escala regional que se modelan.

Del mismo modo, en áreas de la ZMVM, se observan correcciones positivas de la variable en un rango de valores que supera los 3 °C, similar a lo obtenido para el caso de WRFDA, lo cual puede estar relacionado con la cobertura de suelo.

En el caso de las variables químicas CO y O₃, ambas presentan un comportamiento similar, proponiéndose un aumento en las concentraciones para el segundo y sexto ciclo de asimilación, en gran parte del dominio, incluyendo la ZMVM, en un rango que va desde 0 hasta 500 ppb para CO y entre 0 y 100 ppb para O₃. En el caso del O₃, los máximos valores se observan en el horario de las 01 horas tiempo local.

En cuanto a la ubicación hacia el este del dominio de los máximos incrementos de O₃, sugiere que estos fueron advectados hacia esa zona durante el todo el día y que la presencia de valles impidió que se dispersaran. Sin embargo, en los ciclos cuarto y octavo, se sugiere disminuir las concentraciones de ambas variables, hasta en 250 ppb para CO y 100 ppb para O₃. Por su parte, en el caso del SO₂, los incrementos de análisis presentan indistintamente valores positivos y negativos en todo el dominio, durante los diferentes ciclos de asimilación, en un rango que oscila entre -40 y 40 ppb, observándose diferencias importantes en la región influenciada por las emisiones provenientes de la zona industrial de Tula, en el estado de Hidalgo.

De manera general, los cambios más variados son propuestos en los ciclos segundo y sexto, cuyos horarios coinciden con el momento del día en que la radiación solar empieza a alcanzar su máximo. La distribución espacial de los incrementos de análisis en los ciclos cuarto y octavo es más homogénea, mientras que las correcciones propuestas al campo de primera aproximación son menores, aspectos que pueden estar relacionados con el horario de estos ciclos (madrugada), en los cuales, las concentraciones de los contaminantes presentados, como promedio, son menores y más cercanos a las concentraciones de fondo. Estos valores son bien conocidos climatológicamente y pueden ser descritos adecuadamente por los campos iniciales utilizados para realizar la asimilación de datos.

3.7 Evaluación de desempeño del sistema WRF-Chem/DART en la asimilación conjunta de datos meteorológicos y químicos

Los resultados obtenidos en el análisis estadístico realizado luego de las simulaciones numéricas con el sistema WRF-Chem/DART, con asimilación de datos meteorológicos y químicos procedentes de la RAMA, se presentan a continuación.

La Figura 37 muestra las series de tiempo para los valores observados y modelados de las variables T₂, SO₂, O₃ y CO, a modo de ejemplo, para todas las estaciones disponibles. En el eje x están representados los pasos temporales desde el momento de inicio de la asimilación de datos.

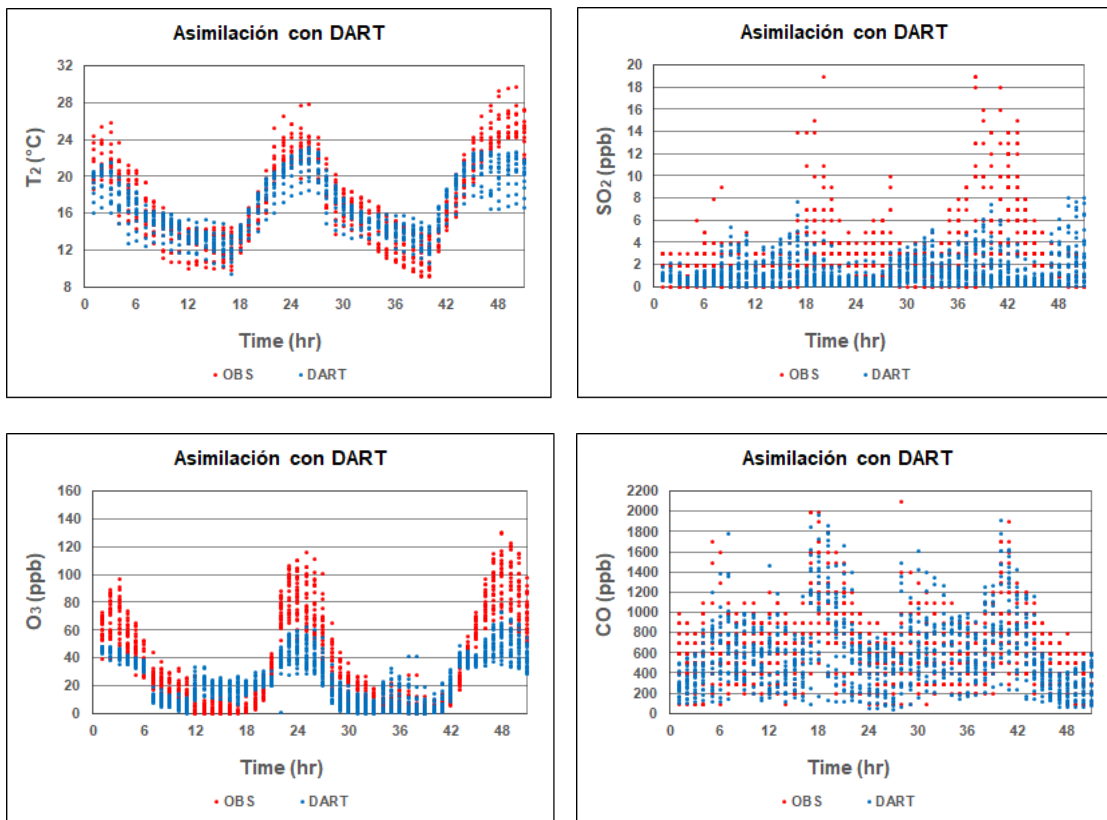


Figura 37. Series temporales de los valores simulados y observados de T₂ (a), SO₂ (b), O₃ (c) y CO (d) para todas las estaciones de la RAMA disponibles

Como se puede apreciar, el modelo reproduce los ciclos diarios de todas las variables presentadas, principalmente los casos de T2 y CO y en menor medida para SO₂, alcanzando los valores extremos, máximo y mínimo, acorde con las observaciones, lo cual ocurre en menor medida para SO₂ y O₃. De forma general, el sistema WRF-Chem/DART tiende a subestimar las mediciones de las estaciones.

En la Figura 38 se muestran los indicadores IOA y RSME, respectivamente, para las variables T2, SO₂, O₃ y CO en cuatro de las siete estaciones de la RAMA seleccionadas, cuyos valores, de forma general, confirman lo descrito anteriormente y siempre son superiores, cualitativamente, cuando se asimila datos procedentes de MOPITT CO con respecto al caso SA, en todas las estaciones y variables analizadas.

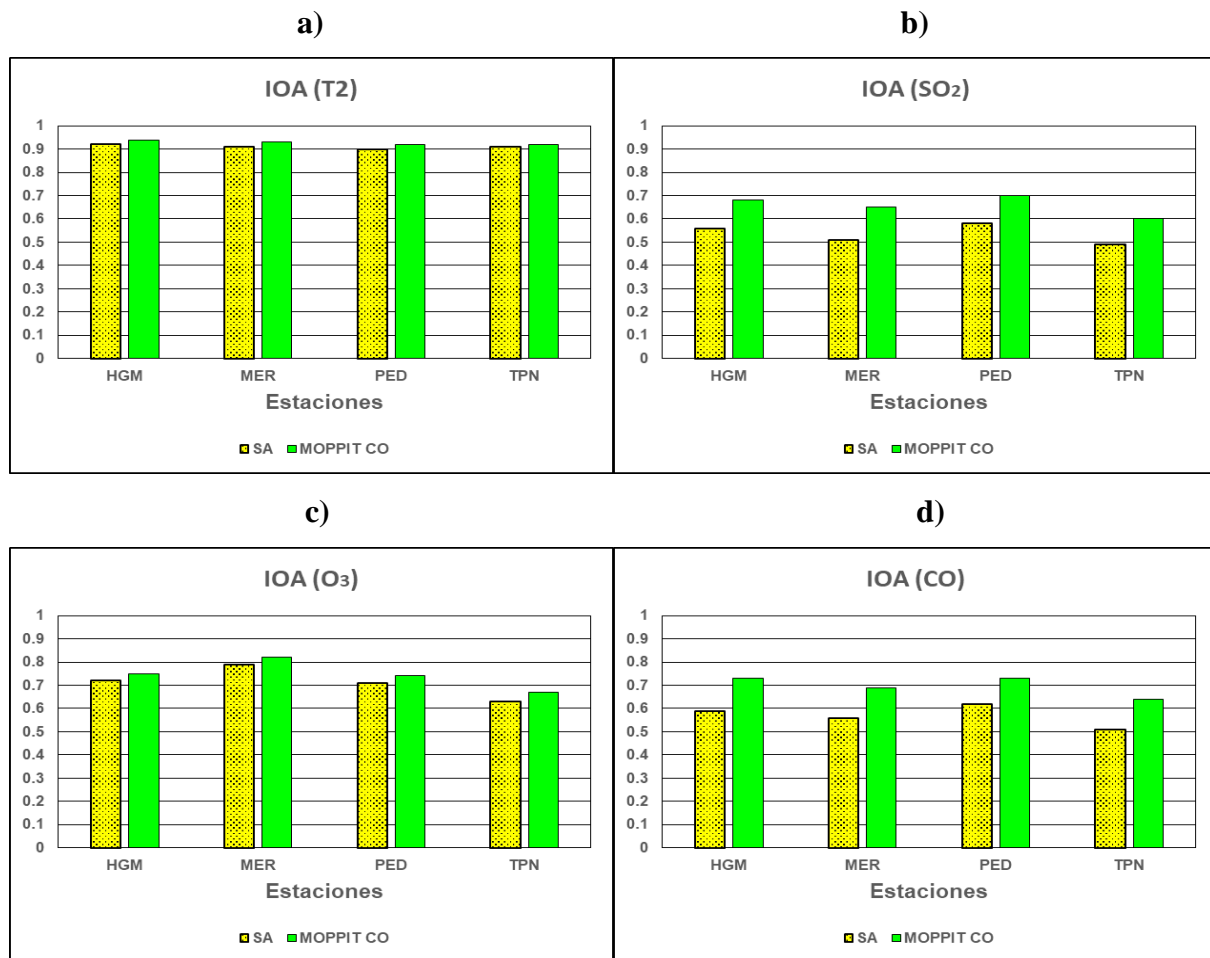


Figura 38. Valores de IOA para las variables T2 (a), SO₂ (b), O₃ (c) y CO (d) para cada una de las estaciones de la RAMA seleccionadas

Como se observa, los valores de IOA para la variable T2 son superiores a 0.9 en todas las estaciones, lo que significa que existe una buena correspondencia entre los valores modelados y las observaciones. En el caso de O₃, excepto la estación TPN, el resto superan el valor de 0.7 por IOA, mientras que, para las variables SO₂ y CO, una vez realizada la asimilación de datos, los valores de IOA siempre están por encima de 0.65, salvo la estación TPN, demostrando que existe una mejoría, de forma general, en el desempeño del modelo WRF-Chem.

A manera de complemento, la Tabla XIV muestra los valores promedio de los estadísticos calculados, en este caso: IOA, RMSE, BIAS y ρ , considerando el conjunto de todas las estaciones disponibles y para cada una de las variables T2, SO₂, O₃ y CO. Se puede observar, en sentido general, que existe buena correspondencia entre los resultados de la modelación y las observaciones, indicado por los valores de IOA y ρ obtenidos para cada variable y complementado con los bajos valores de RMSE que se muestran en cada caso, acordes con las respectivas escalas de las variables. Los valores negativos de BIAS indican que, nuevamente, el modelo tiene a subestimar las mediciones de las estaciones disponibles.

Tabla XIV. Valores medios de IOA, BIAS, RMSE y ρ calculados considerando todas las estaciones de la RAMA disponibles y para las variables T2, SO₂, O₃ y CO

	T2	SO₂	O₃	CO
IOA	0.92	0.66	0.74	0.69
BIAS	-1.97	-1.64	-7.19	-52.35
RMSE	2.83	4.86	25.68	61.78
ρ	0.90	0.63	0.72	0.65

CONCLUSIONES

Teniendo como objetivo principal la incorporación de la asimilación de datos de variables meteorológicas y químicas a la modelación de la calidad del aire con WRF-Chem en el Centro de México, se desarrolla la presente investigación. Para ello, se utilizan dos módulos de asimilación: WRFDA y DART, el primero para definir un criterio sobre el comportamiento de la asimilación sólo de variables meteorológicas y su posible influencia sobre las químicas, mientras que el segundo, para realizar la asimilación combinada de ambos tipos de variables, es decir, de manera conjunta meteorológicas y químicas, y tener una primera aproximación a la influencia de este proceso en la modelación de la calidad del aire en el área de interés, tema de investigación nunca antes abordado en la literatura científica.

Se adapta el módulo de asimilación WRFDA al modelo WRF-Chem, asimilándose datos BUFR procedentes de radiancia del satélite y PREPBUFR de NCEP a las 00 UTC y a las 12 UTC, utilizando la técnica 3DVAR. Se definen seis casos de sensibilidad para contingencias ambientales reportadas en la ZMVM en 2013, entre el 1 y el 4 de mayo (2013-05-01 al 2013-05-04), para determinar la mejor combinación que incluya la fuente de datos y la hora de inicio de las simulaciones numéricas.

De los incrementos de análisis calculados, se obtiene una medida de la variación que experimentan los campos iniciales T2 y VV, a modo de ejemplo, una vez realizado el proceso de minimización y también permiten mostrar si los nuevos campos obtenidos mediante asimilación de datos representan mejor el comportamiento de estas variables meteorológicas en el dominio de estudio y describir los procesos físicos que ocurren en él, principalmente relacionados con la resolución de la malla. En este sentido, se obtiene, que los valores de incremento disminuyen a medida que aumentan los ciclos de asimilación, lo que está en concordancia con el uso de la opción *warm start* para el paso inicial y *cold start* para los siguientes pasos de tiempo, así como el ajuste propio del modelo a través de los diferentes pasos de tiempo de simulación. Estos valores de incremento presentan un continuo aumento cuando hay más datos disponibles.

Los resultados estadísticos muestran que la asimilación de datos meteorológicos influye positivamente en los valores obtenidos en la modelación con WRF-Chem para variables relacionadas, tales como T2, mostrando que existe buena concordancia entre los datos modelados y observados, con valores de IOA superiores a 0.90 para todos los casos de estudio. El modelo, originalmente, tiende a subestimar los valores de T2, dados por los valores negativos de BIAS observados en el caso SA, que se corrige de alguna manera mediante la asimilación de datos, alcanzando BIAS = -1.28 para el caso (PB+RD) a las 12 UTC, es decir, de manera general, la mejor opción de combinación para la variable T2. Se observa una mejoría en los campos de VV, principalmente para el caso (PB+RD) a las 00 UTC, aunque no en igual magnitud que para T2. Los valores de RMSE y BIAS están cerca de 0 para la opción mencionada, demostrando la influencia positiva de la asimilación de datos, con valores de IOA y ρ superiores a 0.75 y 0.70, respectivamente. Para los casos iniciados a las 12 UTC, se obtienen resultados con menor correspondencia respecto a las observaciones.

Sobre la influencia de la asimilación de datos meteorológicos en las variables químicas, se observan cambios en el comportamiento del O₃ luego de ejecutar la simulación, en correspondencia con los resultados obtenidos por Bei et al. (2010), Bei, Li y Molina (2012), Liu et al. (2017) y Mizzi (2017). Las métricas analizadas indican una mejoría para el caso (PB+RD); mientras que, para PB, en general, los valores muestran mayores diferencias en el rendimiento del modelo WRF-Chem. Además, los mejores resultados se logran iniciando las ejecuciones a las 12 UTC, lo que puede estar relacionado con el mayor volumen de datos de observaciones en superficie disponibles en ese momento, mientras que la mejor opción de datos es (PB+RD), que muestra los beneficios que brinda la asimilación de datos de radiancia satelital, sobre todo para T2.

Por tanto, la asimilación de datos químicos (en lugar de meteorológicos) utilizando 3DVAR, 4DVAR o EnKF, constituye un primer paso para mejorar el desempeño del modelo WRF-Chem en la simulación numérica de la concentración de contaminantes en la ZMVM, teniendo en cuenta la influencia de la base de datos de emisiones utilizada en la modelación para obtener resultados más cercanos a las observaciones, que forma parte de las condiciones químicas iniciales y está en constante mejora.

De forma general, el período sin contingencia ambiental presenta mejor desempeño en la modelación, principalmente para O_3 y no existen diferencias significativas entre los métodos de asimilación para el caso de T2. A pesar del alto costo computacional asociado a la técnica 4DVAR, su empleo permite obtener resultados superiores y le sigue En3DVAR, por lo que, de desear un mejor balance entre el costo computacional y la calidad de los resultados, la técnica a utilizar sería En3DVAR.

En cuanto a la asimilación de datos meteorológicos PREPBUFR y químicos (CO de MOPITT y observaciones de la RAMA) con el módulo DART, se observa una buena correspondencia lineal entre los valores observados y modelados, además de comprobarse que existen diferencias entre los campos obtenidos con y sin asimilación. Para el caso de la asimilación combinada de datos de MOPITT con meteorología, se comprueba la necesidad de considerar las resoluciones espacial y temporal del sensor utilizado, ya que cuando el dominio de modelación es muy pequeño, pueden obviarse observaciones dentro del área de interés. No obstante, los resultados de las métricas obtenidas para las estaciones de la Red del Área Metropolitana de Monterrey indicaron que la asimilación de datos influye positivamente en el desempeño del modelo.

Por otra parte, en la asimilación combinada de observaciones puntuales de la RAMA con datos meteorológicos, se obtuvieron mejores resultados en todas las variables analizadas, con valores de IOA mayores a 0.90 para T2 y 0.60 en las químicas.

La asimilación conjunta química-meteorología constituye una herramienta de gran utilidad en la modelación de la calidad del aire, siendo el módulo DART el primero en ofrecer esta funcionalidad de forma eficiente. La propia complejidad del proceso y sus configuraciones, requieren de un alto costo computacional asociado, por lo que, aunque sus aplicaciones en el diagnóstico e investigación son incalculables, su uso para el pronóstico operativo aún no se recomienda.

No obstante, como aporte científico principal de la investigación, esta poderosa herramienta ya se encuentra implementada y verificada para el área del Centro de México, pudiéndose utilizar, incluso, en todo el país.

RECOMENDACIONES

- Teniendo en cuenta los resultados obtenidos en la presente investigación para el sistema de modelación WRF-Chem/WRFDA se recomienda continuar trabajando para poder implementación el mismo en el pronóstico operativo de la calidad del aire, así como la evaluación de su desempeño en esta función.
- Evaluar el comportamiento de otras variables meteorológicas y químicas diferentes a las presentadas en los resultados de esta tesis, con el objetivo de tener una idea mucho más amplia de la influencia que tiene la asimilación de datos en el área de estudio.
- Incluir dentro de la comparación entre las técnicas de asimilación, el método ETKF el cual fue incluido para WRFDA en una versión posterior a la utilizada en esta investigación.
- Para el caso del DART, es necesario incluir otras fuentes de asimilación espacial ya disponibles para este módulo, tales como: CO de IASI y MODIS AOD.
- Actualizar el inventario de emisiones antropogénicas utilizado para disminuir las diferencias en el O₃ debido a cambios en la dilución por vientos más cercanos a la realidad, obtenidos con la asimilación de datos.

REFERENCIAS BIBLIOGRÁFICAS

- Anderson J.L. 2001. An Ensemble Adjustment Kalman Filter for Data Assimilation. *Mon. Wea. Rev.* 129(12):2884-2903
DOI: [https://doi.org/10.1175/1520-0493\(2001\)129<2884:AEAKFF>2.0.CO;2](https://doi.org/10.1175/1520-0493(2001)129<2884:AEAKFF>2.0.CO;2)
- Almanza V.H., Molina L.T. y Sosa G. 2012. Soot and SO₂ contribution to the supersites in the MILAGRO campaign from elevated flares in the Tula Refinery. *Atmos. Chem. Phys.* 12(21):10583- 10599. DOI: <https://doi.org/10.5194/acp-12-10583-2012>
- Anderson J.L. 2003. A Local Least Squares Framework for Ensemble Filtering. *Mon. Wea. Rev.* 131(4):634-642
DOI: [https://doi.org/10.1175/1520-0493\(2003\)131<0634:ALLSFF>2.0.CO;2](https://doi.org/10.1175/1520-0493(2003)131<0634:ALLSFF>2.0.CO;2)
- Anderson J.L. y Collins N.S. 2007. Scalable implementations of ensemble filtering algorithms for data assimilation. *J. Atmos. Ocean. Technol.* 24:1452-1463
DOI: <https://doi.org/10.1175/JTECH2049.1A>
- Anderson J.L., Hoar T., Raeder K., Liu H., Collins N., Torn R. y Arellano A. 2009. The Data assimilation Research Testbed: A community facility. *Bull. Amer. Meteor. Soc.* 90:1283-1296. DOI: <https://doi.org/10.1175/2009BAMS2618.1>
- Archer S., Lowe D., Utembe S., Allan J., Zaveri R., Fast J., Hodnebrog O., Gon H., McFiggans G. 2014. Gaseous chemistry and aerosol mechanism developments for version 3.5.1 of the online regional model, WRF-Chem. *Geosci. Model Devel.* 7(6):2557-2579. DOI: <https://doi.org/10.5194/gmd-7-2557-2014>
- Baklanov A., Schlünzen K., Suppan P., Baldasano J., Brunner D., Aksoyoglu S., Carmichael G., Douros J., Flemming J., Forkel R., Galmarini S., Gauss M., Grell G., Hirtl M., Joffre S., Jorba O., Kaas E., Kaasik M., Kallos G., Kong X., Korsholm U., Kurganskiy A., Kushta J., Lohmann U., Mahura A., Manders-Groot A., Maurizi A., Moussiopoulos N., Rao S.T., Savage N., Seigneur C., Sokhi R.S., Solazzo E., Solomos S., Sorensen B., Tsegas G., Vignati E., Vogel B. y Zhang Y. 2014. Online

coupled regional meteorology chemistry models in Europe: current status and prospects. *Atmos. Chem. Phys.* 14:317–398. DOI: <https://doi.org/10.5194/acp-14-317-2014>

Barbu A.L., Segers A.J., Schaap M., Heemink A.W. y Builtjes P.J.H. 2009. A multi-component data assimilation experiment directed to sulphur dioxide and sulphate over Europe. *Atmos. Environ.* 43:1622–1631

DOI: <https://doi.org/10.1016/j.atmosenv.2008.12.005>

Barker D.M., Huang W., Guo Y.R., Bourgeois A.J. y Xiao Q.N. 2004. A three-dimensional variational data assimilation system for MM5: Implementation and initial results. *Mon. Wea. Rev.* 132(4):897–914

DOI: [https://doi.org/10.1175/1520-0493\(2004\)132<0897:ATVDAS>2.0.CO;2](https://doi.org/10.1175/1520-0493(2004)132<0897:ATVDAS>2.0.CO;2)

Barnes S.L. 1973. Mesoscale objective map analysis using weighted time-series observations. NOAA Tech. Memo. ERL NSSL-62, Norman, OK, 60 pp

Bei N., de Foy B., Lei W., Zavala M. y Molina L.T. 2008. Using 3DVAR data assimilation system to improve ozone simulations in the Mexico City basin. *Atmos. Chem. Phys.* 8:7353–7366. DOI: <https://doi.org/10.5194/acp-8-7353-2008>

Bei N., Lei W., Zavala M. y Molina L.T. 2010. Ozone predictabilities due to meteorological uncertainties in the Mexico City basin using ensemble forecast. *Atmos. Chem. Phys.* 10:6295-6309. DOI: <https://doi.org/10.5194/acp-10-6295-2010>

Bei N., Li G. y Molina L.T. 2012. Uncertainties in SOA simulations due to meteorological uncertainties in Mexico City during MILAGRO-2006 field campaign. *Atmos. Chem. Phys.* 12:11295-11308. DOI: <https://doi.org/10.5194/acp-12-11295-2012>

Bei N., Li G., Meng Z., Weng Y., Zavala M. y Molina L.T. 2014. Impacts of using an ensemble kalman filter on air quality simulations along the California-Mexico border region during Cal-Mex 2010 field campaign. *Sci. Total Environ.* 499:141-153. DOI: <https://doi.org/10.1016/j.scitotenv.2014.07.121>

- Bergthorsson P. y Doos B. 1955. Numerical weather map analysis. *Tellus* 7:329-340.
DOI: <https://doi.org/10.3402/TELLUSA.V7I3.8902>
- Bocquet M. 2012. Parameter field estimation for atmospheric dispersion: Applications to the Chernobyl accident using 4D-Var. *Q. J. Roy Meteorol. Soc.* 138:664–681. DOI: <https://doi.org/10.1002/qj.961>
- Bocquet M., Elbern H., Eskes H., Hirtl M., Žabkar R, Carmichael G.R., Flemming J. e Inness A. 2015: Data assimilation in atmospheric chemistry models: current status and future prospects for coupled chemistry meteorology models. *Atmos. Chem. Phys.* 15:5325–5358. DOI: <https://doi.org/10.5194/acp-15-5325-2015>
- Borrego C., Coutinho M., Costa A.M., Ginja J., Ribeiro C., Monteiro A., Ribeiro I., Valente J., Amorim J.H., Martins H., Lopes D. y Miranda A.I. 2015. Challenges for a new air quality directive: the role of monitoring and modelling techniques. *Urban Climate*. DOI: <https://doi.org/10.1016/j.uclim.2014.06.007>
- Bouttier F. y Rabier F. 1997. The operational implementation of 4D-Var. *European Centre for Medium-Range Weather Forecasts (ECMWF) Newsletter No.* 78:2-5
- Bouttier F. y Courtier P. 1999. Data assimilation concepts and methods. *Meteorological Training Course Lecture Series, European Centre for Medium-Range Weather Forecasts (ECMWF)*. Disponible en:
http://www.ecmwf.int/newsevents/training/rcourse_notes/DATA_ASSIMILATION/ASSIM_CONCEPTS/Assim_concepts.html)
- Bratseth A.M. 1986. Statistical interpolation by means of successive corrections. *Tellus* 38^a(5):439-447. DOI: <https://doi.org/10.1111/j.1600-0870.1986.tb00476.x>
- Burgers G., van Leeuwen P.J. y Evensen G. 1998. Analysis Scheme in the Ensemble Kalman Filter. *Mon. Wea. Rev.* 126(6):1719-1724
DOI: [https://doi.org/10.1175/1520-0493\(1998\)126<1719:ASITEK>2.0.CO;2](https://doi.org/10.1175/1520-0493(1998)126<1719:ASITEK>2.0.CO;2)

- CAME. 2018. La megalópolis de la Zona Metropolitana del Valle de México. Comisión Ambiental de la Megalópolis, Ciudad de México, México. Disponible en: <https://www.gob.mx/comisionambiental/articulos/la-megalopolis-de-la-zmvm?idiom=es>
- Candiani G., Carnevale C., Finzi G., Pisoni E. y Volta M. 2013. A comparison of reanalysis techniques: Applying optimal interpolation and Ensemble Kalman Filtering to improve air quality monitoring at mesoscale. *Sci. Total Environ.* 458-460C:7-14. DOI: 10.1016/j.scitotenv.2013.03.089
- Chai T.F., Carmichael G.R., Sandu A., Tang Y.H. y Daescu D.N. 2006. Chemical data assimilation of transport and chemical evolution over the Pacific (TRACE-P) aircraft measurements *J. Geophys. Res.* 111:D02301
DOI: <https://doi.org/10.1029/2006JD007763>
- Chamu D. (2020) Estudio de la dispersión y destino de los contaminantes en Monterrey mediante modelación de la calidad del aire. Tesis de Licenciatura en Ingeniería Química. Universidad Nacional Autónoma de México, México.
- Charney J. 1949. On a physical basis for numerical prediction of large-scale motions in the atmosphere. *J. Meteor.* 6(6):372-385. DOI: [https://doi.org/10.1175/1520-0469\(1949\)006<0372:OAPBFN>2.0.CO;2](https://doi.org/10.1175/1520-0469(1949)006<0372:OAPBFN>2.0.CO;2)
- Charney J.G., Fjortoft R. y von Neuman J. 1950. Numerical integration of the barotropic vorticity equation. *Tellus* 2(4):237-254. DOI: <https://doi.org/10.1111/j.2153-3490.1950.tb00336.x>
- Charney J.G. 1951. Dynamic Forecasting by Numerical Process. En: Malone T.F. (eds) *Compendium of Meteorology*. Amer. Meteor. Soc., Boston, MA. DOI: https://doi.org/10.1007/978-1-940033-70-9_40
- Chen F. y Dudhia J. 2001. Coupling an advanced land surface–hydrology model with the Penn State–NCAR MM5 modeling system. Part I: Model implementation and sensitivity, *Mon. Wea. Rev.* 129:569–585
DOI: [https://doi.org/10.1175/1520-0493\(2001\)129<0569:CAALSH>2.0.CO;2](https://doi.org/10.1175/1520-0493(2001)129<0569:CAALSH>2.0.CO;2)

- Chen D., Liu Z., Schwartz C. S., Lin H. C., Cetola J. D., Gu Y., Xue L. 2014. The impact of aerosol optical depth assimilation on aerosol forecasts and radiative effects during a wild fire event over the United States. *Geosci. Model. Dev.* 7:2709–2715. DOI: <https://doi.org/10.5194/gmd-7-2709-2014>
- Cressman G.P. 1959. An operational objective analysis system. *Mon. Wea. Rev.* 87(10):367-374
DOI: [https://doi.org/10.1175/1520-0493\(1959\)087<0367:AOOAS>2.0.CO;2](https://doi.org/10.1175/1520-0493(1959)087<0367:AOOAS>2.0.CO;2)
- Cui Y.Y. y de Foy B. 2012. Seasonal variations of the urban heat island at the surface and the near-surface and reductions due to urban vegetation in Mexico City. *J. Appl. Meteor. Climatol.* 51:855-868. DOI: <https://doi.org/10.1175/JAMC-D-11-0104.1>
- Curier R. L., Timmermans R., Calabretta-Jongen S., Eskes H., Segers A., Swart D., Schaap, M. 2012. Improving ozone forecasts over Europe by synergistic use of the LOTOS-EUROS chemical transport model and in-situ measurements. *Atmos. Environ.* 60:217–226. DOI: <https://doi.org/10.1016/j.atmosenv.2012.06.017>
- Daley R. 1991. *Atmospheric Data Analysis*, Cambridge Atmospheric and Space Science Series, Cambridge University Press, Cambridge, UK
- de Foy B., Fast J.D., Paech S.J., Phillips D., Walters J.T., Coulter R.L., Martin T.J., Pekour M.S., Shaw W.J., Kastendeuch P.P., Marley N.A., Retama A. y Molina L.T., 2008. Basin-scale wind transport during the MILAGRO field campaign and comparison to climatology using cluster analysis. *Atmos. Chem. Phys.* 8:1209-1224. DOI: <https://doi.org/10.5194/acp-8-1209-2008>
- de Foy B., Krotkov N.A., Bei N., Herndon S.C., Huey L.G., Martínez A.P., Ruiz L.G., Word E.C., Zavala M. y Molina L.T. 2009. Hit from both sides: tracking industrial and volcanic plumes in Mexico City with surface measurements and OMI SO₂ retrievals during the MILAGRO field campaign. *Atmos. Chem. Phys.* 9:9599–9617. DOI: <https://doi.org/10.5194/acp-9-9599-2009>

- Doran J.C. y Zhong S. 2000. Thermally Driven Gap Winds into the Mexico City Basin. *J. Appl. Meteorol.* 39(8): 1330-1340
DOI: [https://doi.org/10.1175/1520-0450\(2000\)039<1330:TDGWIT>2.0.CO;2](https://doi.org/10.1175/1520-0450(2000)039<1330:TDGWIT>2.0.CO;2)
- Dudhia J. 1989. Numerical study of convection observed during the winter monsoon experiment using a mesoscale two-dimensional model. *J. Atmos. Sci.* 46:3077-3107. DOI: [https://doi.org/10.1175/1520-0469\(1989\)046<3077:NSOCOD>2.0.CO;2](https://doi.org/10.1175/1520-0469(1989)046<3077:NSOCOD>2.0.CO;2)
- Elbern H. y Schmidt H. 1999. A four-dimensional variational chemistry data assimilation scheme for Eulerian chemistry transport modeling. *J. Geophys. Res.* 104(D15):18583-18598. DOI: <https://doi.org/10.1029/1999JD900280>
- Elbern H., Schmidt H. 2001. Ozone episode analysis by four-dimensional variational chemistry data assimilation. *J. Geophys. Res.* 106:3569–3590
DOI: <https://doi.org/10.1029/2000JD900448>
- Elbern H., Strunk A., Schmidt H. y Talagrand O. 2007. Emission rate and chemical state estimation by 4-dimensional variational inversion. *Atmos. Chem. Phys.* 7:3749-3769. DOI: <https://doi.org/10.5194/acp-7-3749-2007>
- Emmons L., Walters S., Hess P., Lamarque J., Pfister G., Fillmore D., Granier C., Guenther A., Kinnison D., Laepple T., Orlando J., Tie X., Tyndall G., Wiedinmyer C., Baughcum S. y Kloster S. 2010. Description and evaluation of the Model for Ozone and Related chemical Tracers, version 4 (MOZART-4). *Geosci. Model Dev.* 3:43-67. DOI: <https://doi.org/10.5194/gmd-3-43-2010>
- Engelen R.J. y Bauer P. 2014 The use of variable CO₂ in the data assimilation of AIRS and IASI radiances. *Q. J. Roy. Meteorol. Soc.* 140:958–965
DOI: <https://doi.org/10.1002/qj.919>
- Eltahan M. y Alahmadi S. 2019. Numerical Dust Storm Simulation Using Modified Geographical Domain and Data Assimilation: 3DVAR and 4DVAR (WRF-Chem/WRFDA). *IEEE Access* 7:128980-128989
DOI: <https://doi.org/10.1109/ACCESS.2019.2930812>

- Evensen G. 1994. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte-Carlo methods to forecast error statistics. *J. Geophys. Res. Oceans* 99(C5). DOI: <https://doi.org/10.1029/94JC00572>
- Evensen G. 2009. *Data Assimilation: The Ensemble Kalman Filter*. Springer-Verlag Berlin Heidelberg XXIII, 307 pp. DOI: <https://doi.org/10.1007/978-3-642-03711-5>
- Fast J.D., Gustafson W.I.Jr., Easter R.C., Zaveri R.A., Barnard J.C., Chapman E.G., Grell G.A. y Peckham S.E. 2006. Evolution of ozone, particulates, and aerosol direct radiative forcing in the vicinity of Houston using a fully coupled meteorology-chemistry-aerosol model. *J. Geophys. Res.* 111:D21305
DOI: <https://doi.org/10.1029/2005JD006721>
- Forkel R., Werhahn J., Hansen A.B., McKeen S., Peckham S., Grell G. y Suppan P. 2012. Effect of aerosol-radiation feedback on regional air quality – A case study with WRF/Chem. *Atmos. Environ.* 53:202-211
DOI: <https://doi.org/10.1016/j.atmosenv.2011.10.009>
- Fox DG. 1981. Judging air quality model performance: A summary of the AMS workshop on dispersion model performance. *Bull. Amer. Meteor. Soc.* 62:599-609.
DOI: [https://doi.org/10.1175/1520-0477\(1981\)062<0599:JAQMP>2.0.CO;2](https://doi.org/10.1175/1520-0477(1981)062<0599:JAQMP>2.0.CO;2)
- Gandin L.S. 1963. *Objective analysis of meteorological fields*. Gidrometeorologicheskoe Izdatelstvo, Leningrad. English translation by Israeli Program for Scientific Translations, Jerusalem
- García J.A., Ruiz L.G., García J.S. y Reséndiz N.A. 2009. Comportamiento de los contaminantes en cuencas atmosféricas: metodología y estudio de caso. Reporte Técnico. Instituto Nacional de Ecología, Ciudad de México, México
DOI: <https://doi.org/10.13140/RG.2.1.1725.7203>
- García J.S., García J.A., Jazcilevich A. y Ruíz L.G. 2014. The influence of the Tula, Hidalgo complex on the air quality of the Mexico City Metropolitan Area. *Rev. Atmos.* 27(2):215-225. DOI: [https://doi.org/10.1016/S0187-6236\(14\)71111-7](https://doi.org/10.1016/S0187-6236(14)71111-7)

- García J.A y Mora M.A. 2017. Implementation of the Unified Post Processor (UPP) and the Model Evaluation Tools (MET) for WRF-Chem evaluation performance. *Rev. Atmos.* 30(3). DOI: <https://doi.org/10.20937/atm.2017.30.03.06>
- García J.A., Mar B.E. y Ruiz L.G. 2018. Modelo de distribución espacial, temporal y de especiación del inventario de emisiones de México (año base 2008) para su uso en modelización de calidad del aire (DiETE). *Rev. Int. Contam. Ambient.* DOI: <https://doi.org/10.20937/RICA.2018.34.04.07>
- Grell G.A., Peckham S.E., Schmitz R., McKeen S.A., Frost G., Skamarock W.C. y Eder B. 2005. Fully coupled “online” chemistry within the WRF model. *Atmos. Environ.* 39:6957-6975. DOI: <https://doi.org/10.1016/j.atmosenv.2005.04.027>
- Guenther A., Zimmerman P. y Wildermuth M. 1994. Natural volatile organic compound emission rate estimates for U.S. Woodland landscapes. *Atmos. Environ.* 28:1197-1210. DOI: [https://doi.org/10.1016/1352-2310\(94\)90297-6](https://doi.org/10.1016/1352-2310(94)90297-6)
- Guenther A., Karl T., Harley P., Wiedinmyer C., Palmer P.I. y Geron C. 2006. Estimates of global terrestrial isoprene emissions using MEGAN (Model of Emissions of Gases and Aerosols from Nature). *Atmos. Chem. Phys.* 6:3181–3210
DOI: <https://doi.org/10.5194/acp-6-3181-2006>
- Guerrette J.J. y Henze D.K. 2015. Development and application of the WRFPLUS-Chem online chemistry adjoint and WRFDA-Chem assimilation system. *Geosci. Model Dev.* 8, 1857-1876. DOI: <https://doi.org/10.5194/gmd-8-1857-2015>
- Henze D. 2016. Implementation and application to inverse modeling of black carbon Emissions
- Hernandez R. 2013. Asimilación de observaciones meteorológicas con el modelo físico-matemático ARPS. Tesis de Licenciatura en Meteorología. Instituto Superior de Tecnologías y Ciencias Aplicadas. La Habana, Cuba
- Herrera G. 2016. Notas de clases de la asignatura: “Asimilación de datos usando Filtro de Kalman Ensamblado”

- Hong S.Y., Dudhia J. y Chen S.H. 2004. A revised approach to ice microphysical processes for the bulk parameterization of clouds and precipitation. *Mon. Wea. Rev.* 132:103-120
DOI: [https://doi.org/10.1175/15200493\(2004\)132<0103:ARATIM>2.0.CO;2](https://doi.org/10.1175/15200493(2004)132<0103:ARATIM>2.0.CO;2)
- Hong S. Y. y Dudhia J. 2003. Testing of a new non-local boundary layer vertical diffusion scheme in numerical weather prediction applications, En: 20th Conference on Weather Analysis and Forecasting/16th Conference on Numerical Weather Prediction, Seattle, SA
- Houtekamer P.L. y Mitchell H.L. 2001. A Sequential Ensemble Kalman Filter for Atmospheric Data Assimilation. *Mon. Wea. Rev.* 129(1):123-137. DOI: [https://doi.org/10.1175/1520-0493\(2001\)129<0123:ASEKFF>2.0.CO;2](https://doi.org/10.1175/1520-0493(2001)129<0123:ASEKFF>2.0.CO;2)
- INECC. 2016. Diagnóstico sobre la calidad del aire en Cuencas atmosféricas de México. Informe Final. Instituto Nacional de Ecología y Cambio Climático. Secretaría de Medio Ambiente y Recursos Naturales. Ciudad de México, México, 72 pp
- INECC. 2018. Informe Nacional de Calidad del Aire 2017, México. Instituto Nacional de Ecología y Cambio Climático, Secretaría de Medio Ambiente y Recursos Naturales, Ciudad de México, México, 317 pp
- Jauregui E. 1997. Heat island development in Mexico City. *Atmos. Environ.* 31(22):3821-3831. DOI: [https://doi.org/10.1016/S1352-2310\(97\)00136-2](https://doi.org/10.1016/S1352-2310(97)00136-2)
- Jiang Z., Liu Z., Wang T., Schartz C.S., Lin H.C. y Jiang, F. 2013. Probing into the impact of 3DVAR assimilation of surface PM10 observations over China using process analysis. *J. Geophys. Res. Atmos.* 118:6738-6749
DOI: <https://doi.org/10.1002/jgrd.50495>.
- Kain J.S. y Fritsch J.M. 1993. Convective parameterization for mesoscale models: The Kain–Fritsch scheme. The Representation of Cumulus Convection in Numerical Models. *Meteor. Monogr. No. 24*, Amer. Meteor. Soc., 165-170

- Kalnay E. 2003. Atmospheric modeling: Data assimilation and predictability, Cambridge University Press, Cambridge, UK. ISBN: 9780521796293
DOI: <https://doi.org/10.1017/CBO9780511802270>
- Koohkan M.R., Bocquet M., Roustan Y., Kim Y. y Seigneur C. 2013. Estimation of volatile organic compound emissions for Europe using data assimilation. *Atmos. Chem. Phys.* 13:5887–5905. DOI: <https://doi.org/10.5194/acp-13-5887-2013>
- Kryza M., Werner M., Dudek J. y Dore A.J. 2020. The Effect of Emission Inventory on Modelling of Seasonal Exposure Metrics of Particulate Matter and Ozone with the WRF-Chem Model for Poland. *Sustainability*, 12(5414)
DOI: <https://doi.org/10.3390/su12135414>
- Lahoz W., Khattatov B. y Ménard, R. (eds.). 2010. Data assimilation - Making sense of observations. Springer Science and Business Media, 718 pp
DOI: <https://doi.org/10.1007/978-3-540-74703-1>
- Lauvaux T., Schuh A.E., Bocquet M., Wu L., Richardson S., Miles N. y Davies K.J. 2012. Network design for mesoscale inversions of CO₂ sources and sinks, *Tellus B*, 64(1): 17980. DOI: <https://doi.org/10.3402/tellusb.v64i0.17980>
- Liu Z., Liu Q., Lin H.C., Schwartz C.S., Lee Y.H. y Wang T. 2011. Three-dimensional variational assimilation of MODIS aerosol optical depth: Implementation and application to a dust storm over East Asia. *J. Geophys. Res.* 116(D23). DOI: <https://doi.org/10.1029/2011JD016159>
- Liu X., Mizzi A.P., Anderson J.L., Fung I.Y. y Cohen R.C. 2017. Assimilation of satellite NO₂ observations at high spatial resolution using OSSEs. *Atmos. Chem. Phys.* 17: 7067-7081. DOI: <https://doi.org/10.5194/acp-17-7067-2017>
- Liu X., Mizzi A. y Anderson J. 2016. Assimilation of satellite NO₂ observations at high spatial resolution, *Atmos. Chem. Phys. Discuss.* DOI: <https://doi.org/10.5194/acp-2016-770>

- LTM Center for Energy and the Environment. 2017. Evaluación de los impactos en la concentración de ozono por la aplicación de estrategias integradas de control de emisiones en la Megalópolis. Technical Report. INECC/LP-009/2017, México. En: https://www.gob.mx/cms/uploads/attachment/file/328623/CAMe_Ozono_Modelacion_Reporte_Final.pdf
- Ma C., Wang T., Mizzi A., Anderson J., Zhuang B., Xie M. y Wu R. (2019a). Multiconstituent Data Assimilation With WRF-Chem/DART: Potential for Adjusting Anthropogenic Emissions and Improving Air Quality Forecasts Over Eastern China. *J. Geophys. Res. Atmos.* 124(13):7393-7412
DOI: <https://doi.org/10.1029/2019JD030421>
- Ma C., Wang T, Jiang Z., Wu H., Zhao M., Zhuang B., Li S., Xie M., Li M., Liu J. y Wu R. (2019b). Importance of Bias Correction in Data Assimilation of Multiple Observations Over Eastern China Using WRF-Chem/DART. *J. Geophys. Res. Atmos.* 125(1) e2019JD031465. DOI: <https://doi.org/10.1029/2019JD031465>
- Mesinger F. 1997. Dynamics of limited-area models: Formulation and numerical methods. *Meteorol. Atmos. Phys.* 63, 3-14. DOI: <https://doi.org/10.1007/BF01025360>
- Mesinger F., DiMego G., Kalnay E., Mitchell K., Shafran P.C., Ebisuzaki W., Jovic D., Woollen J., Rogers E., Berbery E.H., Ek M.B., Fan Y., Grumbine R., Higgins W., Li H., Lin Y., Manikin G., Parrish D., Shi W. 2006. North American Regional Reanalysis. *Bull. Amer. Meteorol. Soc.* 87:343-360
DOI: <https://doi.org/10.1175/BAMS-87-3-343>
- Mijling B. y van der A R.J. 2012. Using daily satellite observations to estimate emissions of short-lived air pollutants on a mesoscopic scale. *J. Geophys. Res.* 117:D17302
DOI: <https://doi.org/10.1029/2012JD017817>

- Mizzi A.P., Arellano A.F., Edwards D.P., Anderson J.L., Pfister G.G. 2016. Assimilating compact phase space retrievals of atmospheric composition with WRF-Chem/DART: a regional chemical transport/ensemble Kalman filter data assimilation system. *Geosci. Model Dev.* 9:965-978
DOI: <https://doi.org/10.5194/gmd-9-965-2016>
- Mizzi A.P. 2017. Introduction to Assimilation of Atmospheric Composition Observations. Disponible en: https://ruc.noaa.gov/wrf/wrf-chem/wrf_tutorial_2018/DA.pdf
- Mizzi A.P. 2018. Introduction to Assimilation of Atmospheric Composition Observations with WRF-Chem. WRF-Chem Tutorial. Disponible en: https://ruc.noaa.gov/wrf/wrf-chem/wrf_tutorial_2018/DA.pdf
- Mizzi A.P., Edwards D. y Anderson J. 2018. Assimilating compact phase space retrievals (CPSRs): Comparison with independent observations (MOZAIC in situ and IASI retrievals) and extension to assimilation of truncated retrieval profiles. *Geosci. Model Dev.* 11:3727-3745. DOI: <https://doi.org/10.5194/gmd-11-3727-2018>
- Mizzi A.P. 2020. Ensemble Chemical Data Assimilation/Forecast Cycling with WRF-Chem/DART and Localization Tutorial. UNAM Air Quality Forecasting / Data Assimilation Workshop, Ciudad de México, Mexico. Disponible en: http://www.ltmce2.org.mx/images/taller_2020/UNAM_AQ_Workshop_Localization_Tutorial_Jan_2020_low.pdf
- Mlawer E.J., Taubman S.J., Brown P.D., Iacono M.J., Clough S.A. 1997. Radiative transfer for inhomogeneous atmosphere: RRTM, a validated correlated-k model for the long-wave. *J. Geophys. Res.* 102:16663–16682
DOI: <https://doi.org/10.1029/97JD00237>
- Molina L.T., Madronich S., Gaffney J.S., Apel E., de Foy B., Fast J., Ferrare R., Herndon S, Jimenez J.L., Lamb B., Osornio-Vargas A.R., Russell P., Schauer J.J., Stevens P.S., Zavala M. 2010. An overview of the MILAGRO 2006 campaign: Mexico City emissions and their transport and transformation. *Atmos. Chem. Phys. Discuss.* 10:8697-8760. DOI: <https://doi.org/10.5194/acp-10-8697-2010>

- OJEU. 2008. Directive 2008/50/EC of the European Parliament and of the Council of 21 May 2008. Official Journal of the European Union, L 152/1, 11 de Junio de 2008
- OMS. 2014. 7 millones de muertes cada año debidas a la contaminación atmosférica. Organización Mundial de la Salud. Comunicado de Prensa. Ginebra, Suiza. 25 de marzo de 2014. Disponible en: <https://www.who.int/es/news-room/detail/25-03-2014-7-million-premature-deaths-annually-linked-to-air-pollution>
- Oropeza P., Cruz P., Díaz O., Zavala J y Castulo R. 2020. Implementación del Clúster Ometeotl en el CCA-UNAM. Coloquio de Supercómputo 2020. Centro de Ciencias de la Atmósfera, Universidad Nacional Autónoma de México, México. Disponible en: <https://coloquio.super.unam.mx/2020/materiales/carteles/CS2020AV03.pdf>
- Pagowski M. y Grell G.A. 2012. Experiments with the assimilation of fine aerosols using an Ensemble Kalman Filter. J. Geophys. Res. 117(D21)
DOI: <https://doi.org/10.1029/2012JD018333>
- Pagowski M., Grell G.A., McKeen S.A., Peckham S.E. y Devenyi D. 2010. Three-dimensional variational data assimilation of ozone and fine particulate matter observations: some results using the Weather Research and Forecasting-Chemistry model and Grid-point Statistical Interpolation. Q. J. Roy. Meteorol. Soc. 136:2013-2024. DOI: <https://doi.org/10.1002/qj.700>
- Parrish D.F. y Derber J.C. 1992. The National Meteorological Center's Spectral Statistical-Interpolation Analysis System. Mon. Wea. Rev. 120(8):1747-1763
DOI: [https://doi.org/10.1175/1520-0493\(1992\)120<1747:TNMCSS>2.0.CO;2](https://doi.org/10.1175/1520-0493(1992)120<1747:TNMCSS>2.0.CO;2)
- Peckham S.E., Grell G.A., McKeen S.A., Ahmadov R., Wong K.Y., Barth M., Pfister G., Wiedinmyer C., Fast J.D., Gustafson W.I. , Ghan S.J., Zaveri R., Easter R.C., Chapman E., Hewson M., Schmitz R., Salzmann M., Beck V. y Freitas S.R. 2017. WRF/Chem Version 3.8.1 User's Guide, NOAA, PNNL, NCAR, CIRES, CIRA, University of Queensland, University of Chile, Max Planck Institute for Chemistry, Centro de Previsao de Tempo e Estudos Climaticos
DOI: <https://doi.org/10.7289/V5/TM-OAR-GSD-48>

- Peng Z., Liu Z., Chen D. y Ban J. 2017. Improving PM 2.5 forecast over China by the joint adjustment of initial conditions and source emissions with an ensemble Kalman filter. *Atmos. Chem. Phys.* 17:4837-4855
DOI: <https://doi.org/10.5194/acp-17-4837-2017>
- Peng Z., Lei L., Liu Z., Sun J., Ding A., Ban J., Chen D., Kou X. y Chu K. 2018. The impact of multi-species surface chemical observation assimilation on air quality forecasts in China, *Atmos. Chem. Phys.* 18:17387-17404
DOI: <https://doi.org/10.5194/acp-18-17387-2018>
- Petrie R. E. 2008. Localization in the ensemble Kalman Filter. Tesis de Maestría en Atmosfera, Océano y Clima. Universidad de Reading, Inglaterra
- Pielke R. A. 1984. *Mesoscale Meteorological Modeling*. Academic Press, London
- Pu Z. y Kalnay E. 2019. Numerical Weather Prediction Basics: Models, Numerical Methods, and Data Assimilation. In: Duan Q., Pappenberger F., Wood A., Cloke H., Schaake J. (eds) *Handbook of Hydrometeorological Ensemble Forecasting*. Springer, Berlin, Heidelberg. DOI: https://doi.org/10.1007/978-3-642-39925-1_11
- Rayner R.J. 2004. Optimizing CO₂ observing networks in the presence of model error: results from TransCom 3. *Atmos. Chem. Phys.* 4:413-421
DOI: <https://doi.org/10.5194/acp-4-413-2004>
- Richardson L.F. 1922. *Weather prediction by numerical process*. Cambridge University Press, Cambridge. Reprinted by Dover (1965, New York) with a new introduction by Sydney Chapman
- Rivera C., Sosa G., Wöhrnschimmel H., de Foy B., Johansson M. y Galle B. 2009. Tula industrial complex (Mexico) emissions of SO₂ and NO₂ during the MCMA 2006 field campaign using a mobile mini-DOAS system. *Atmos. Chem. Phys.* 9:6351-6361. DOI: <https://doi.org/10.5194/acp-9-6351-2009>

- Rodriguez-Zas J.A y García-Reynoso J.A. 2016. Modelación de contaminantes atmosféricos con WRF-Chem en la Zona Metropolitana de Guadalajara. Tesis de Maestría en Ciencias de la Tierra. Universidad Nacional Autónoma de México, Ciudad de México, México
- Roustan Y. y Bocquet M. 2006. Inverse modelling for mercury over Europe. *Atmos. Chem. Phys.* 6(10):3085-3098. DOI: <https://doi.org/10.5194/acp-6-3085-2006>
- Ruiz L.G., Torres R., Woernschimmel H., Steinbrecher R., Junkerman W., Nigenda E., García J.A., Melamed M., Jazcilevich A., Mar B.E., Grutter M. y Molina L.T. 2010. In depth analysis of the role of the mountain gap south of the Valley of Mexico on the air quality in Mexico City. European Geosciences Union 2010-14043-1, General Assembly, Viena, Austria
- Russell A.G. y Dennis R. 2000. NARSTO critical review of photochemical models and modeling. *Atmos. Environ.* 34(12):2283-2324. DOI: [https://doi.org/10.1016/S1352-2310\(99\)00468-9](https://doi.org/10.1016/S1352-2310(99)00468-9)
- Salcedo D., Castro T., Ruiz L.G., García J.A., Torres R., Torres A., Mar B.E., Salcido A., Celada A.T., Carreón S., Martínez A.P., Fentanes O.A., Deustúa E., Ramos R., Retama A., Saavedra M.I. y Suárez M. 2012. Study of the regional air quality south of Mexico City (Morelos state). *Sci. Total Environ.* 414:417-432
DOI: <https://doi.org/10.1016/j.scitotenv.2011.09.041>
- Sasaki Y. 1970. Some basic formalisms in numerical variational analysis. *Mon. Wea. Rev.* 98:875-883
DOI: [https://doi.org/10.1175/1520-0493\(1970\)098<0875:SBFINV>2.3.CO;2](https://doi.org/10.1175/1520-0493(1970)098<0875:SBFINV>2.3.CO;2)
- Schwartz C.S., Liu Z., Lin H.C. y McKeen S.A. 2012. Simultaneous three-dimensional variational assimilation of surface fine particulate matter and MODIS aerosol optical depth. *J. Geophys. Res. Atmos.* 117:D13202
DOI: <https://doi.org/10.1029/2011JD017383>

- Schwartz C.S., Liu Z., Lin H.C., Cetola J.D. 2014. Assimilating aerosol observations with a “hybrid” variational-ensemble data assimilation system. *J. Geophys. Res. Atmos.* 119:4043–4069. DOI: <https://doi.org/10.1002/2013JD020937>
- Schutgens N.A.J., Miyoshi T., Takemura T. y Nakajima T. 2010. Applying an ensemble Kalman filter to the assimilation of AERONET observations in a global aerosol transport model. *Atmos. Chem. Phys.* 10:2561-2576
DOI: <https://doi.org/10.5194/acp-10-2561-2010>
- Seinfeld J.H. y Pandis S.N. 2006. *Atmospheric Chemistry and Physics – from Air Pollution to Climate Change*, Chapter 23: Atmospheric Chemical Transport Models, Wiley-Interscience, New York, NY
- Seaman N.L. 2000. Meteorological modeling for air quality assessments. *Atmos. Environ.* 34:2231–2259. DOI: [https://doi.org/10.1016/S1352-2310\(99\)00466-5](https://doi.org/10.1016/S1352-2310(99)00466-5)
- Seaman N.L. 2003. Future directions of meteorology related to air-quality research. *Environ. Int.* 29:245–252. DOI: [https://doi.org/10.1016/S0160-4120\(02\)00183-6](https://doi.org/10.1016/S0160-4120(02)00183-6)
- Semane N., Peuch V.H., Pradier S., Desroziers G., El Amraoui L., Brousseau P., Massart S., Chapnik B. y Peuch A. 2009. On the extraction of wind information from the assimilation of ozone profiles in Météo-France 4-D-Var operational NWP suite. *Atmos. Chem. Phys.* 9:4855-4867. DOI: <https://doi.org/10.5194/acp-9-4855-2009>
- SEMARNAT. 2016. Informe de la Situación del Medio Ambiente en México. Compendio de Estadísticas Ambientales. Indicadores Clave, de Desempeño Ambiental y de Crecimiento Verde. Edición 2015. Secretaría de Medio Ambiente y Recursos Naturales, México, 498 pp
- Skamarock W.C., Klemp J.B., Dudhia J., Gill D.O., Barker D.M., Duda M.G., Huang X.Y., Wang W. y Powers J.G. 2008. A Description of the Advanced Research WRF Version 3. Technical Note, National Center for Atmospheric Research (NCAR), Boulder, Colorado

- SMN. 2013. Reporte del Clima en México. Mayo 2013. Año 3, Número 5. Servicio Meteorológico Nacional. Comisión Nacional del Agua. Ciudad de México, México. Disponible en:
<https://smn.conagua.gob.mx/tools/DATA/Climatolog%C3%ADa/Diagn%C3%B3stico%20Atmosf%C3%A9rico/Reporte%20del%20Clima%20en%20M%C3%A9xico/R-C-Mayo13.pdf>
- Sosa G., Vega E., González E., Mora V. y López D. 2013. Air Pollutant Characterization in Tula Industrial Corridor, Central Mexico, during the MILAGRO Study. *BioMed Res. Int.* 2013:ID521728. DOI: <https://doi.org/10.1155/2013/521728>
- Stockwell, W.R., Middleton P., Chang J.S. y Xiaoyan T. 1990. The second generation regional acid deposition model chemical mechanism for regional air quality modeling. *J. Geophys. Res.* 95(D10):16343-16367
DOI: <https://doi.org/10.1029/JD095iD10p16343>
- Sun J., Shen Z., Wang R., Li G., Zhang Y., Zhang B., He K., Tang Z., Xu H., Qu L., Sai Hang Ho S., Liu S. y Cao L. 2021. A comprehensive study on ozone pollution in a megacity in North China Plain during summertime: Observations, source attributions and ozone sensitivity. *Environ. Int.* 146:106279
DOI: <https://doi.org/10.1016/j.envint.2020.106279>
- Tombette M., Mallet V. y Sportisse B. 2009. PM₁₀ data assimilation over Europe with the optimal interpolation method. *Atmos. Chem. Phys.* 9:57-70
DOI: [10.5194/acp-9-57-2009](https://doi.org/10.5194/acp-9-57-2009)
- Turtós L., Capote G., Fonseca Y., Álvarez L., Sánchez M., Bezanilla A., Borrajero I., Meneses R. y Pire S. 2013. Assessment of the Weather Research and Forecasting model implementation in Cuba addressed to diagnostic air quality modeling. *Atmos. Pollut. Res.* 4(1):64-74
DOI: <https://doi.org/10.5094/APR.2013.007>

- Vira J. y Sofiev M. 2012. On variational data assimilation for estimating the model initial conditions and emission fluxes for short-term forecasting of Sox concentrations. *Atmos. Environ.* 46:318-328. DOI: <https://doi.org/10.1016/j.atmosenv.2011.09.066>
- Wang X.L. 2008a. Accounting for autocorrelation in detecting mean shifts in climate data series using the penalized maximal t or F test. *J. Appl. Meteorol. Climatol.* 47(9):2423–2444. DOI: <https://doi.org/10.1175/2008JAMC1741.1>
- Wang X.L. 2008b. Penalized maximal F test for detecting undocumented mean shift without trend change. *J. Atmos. Ocean. Technol.* 25:368–384
DOI: <https://doi.org/110.1175/2007JTECHA982.1>
- Wang X., Mallet V., Berroir J.P. y Herlin I. 2011. Assimilation of OMI NO₂ retrievals into a regional chemistry transport model for improving air quality forecasts over Europe. *Atmos. Environ.* 45:485-492
DOI: <https://doi.org/10.1016/j.atmosenv.2010.09.028>
- Wang X., Barker D.M., Snyder C. y Hamill T.M. 2008. A hybrid ETKF–3DVAR data assimilation scheme for the WRF model. Part I: Observing system simulation experiment. *Mon. Wea. Rev.* 136:5116-5131.
DOI: <https://doi.org/10.1175/2008MWR2444.1>
- Wang X., Barker D.M., Snyder C. y Hamill T.M. 2008. A hybrid ETKF–3DVAR data assimilation scheme for the WRF model. Part II: Real Observation experiments. *Mon. Wea. Rev.* 136:5132-5147. DOI: <https://doi.org/10.1175/2008MWR2445.1>
- Wiedinmyer C., Akagi S.K., Yokelson R.J.; Emmons L.K., Al-Saadi J.A., Orlando J.J., Soja A.J. 2010. The Fire INventory from NCAR (FINN): a high resolution global model to estimate the emissions from open burning. *Geosci. Model Dev.* 4:625-641. DOI: <https://doi.org/10.5194/gmd-4-625-2011>
- Willmott C.J. 1981. On the validation of models. *Phys. Geog.* 2(2):184-194
DOI: <https://doi.org/10.1080/02723646.1981.10642213>

- Willmott C.J. 1982. Some comments on the evaluation of model performance. Bull. Amer. Meteor. Soc. 63(11):1309-1313
DOI: [https://doi.org/10.1175/1520-0477\(1982\)063<1309:SCOTEO>2.0.CO;2](https://doi.org/10.1175/1520-0477(1982)063<1309:SCOTEO>2.0.CO;2)
- Wu L., Mallet V., Bocquet M. y Sportisse B. 2008. A comparison study of data assimilation algorithms for ozone forecasts. J. Geophys. Res. 113:D20310
DOI: <https://doi.org/10.1029/2008JD009991>
- Wu L., Bocquet M. y Chevallier M. 2010. Optimal Reduction of the Ozone Monitoring Network over France. Atmos. Environ. 44:3071-3083
DOI: <https://doi.org/10.1016/j.atmosenv.2010.04.012>
- Wu L. y Bocquet M. 2010. Optimal Redistribution of the Background Ozone Monitoring Stations over France. Atmos. Environ. 45:772-783
DOI: <https://doi.org/10.1016/j.atmosenv.2010.08.038>
- Xu M., Fu K. y Lv X. 2017. Application of Adjoint Data Assimilation Method to Atmospheric Aerosol Transport Problems. Adv. Math. Phys. 2017:5865403
DOI: <https://doi.org/10.1155/2017/5865403>
- Yumimoto K. y Takemura T. 2013. The SPRINTARS version 3.80/4D-Var data assimilation system: development and inversion experiments based on the observing system simulation experiment framework. Geosci. Model Dev. 6:2005-2022. DOI: <https://doi.org/10.5194/gmd-6-2005-2013>
- Yumimoto K., Uno I., Sugimoto N., Shimizu A., Hara Y. y Takemura T. 2012. Size-resolved adjoint inversion of Asian dust. Geophys. Res. Lett. 39:L24808
DOI: <https://doi.org/10.1029/2012GL053890>
- Zabkar R., Honzak L., Skok G., Forkel R., Rakovec J., Ceglar A. y Zagar N. 2015. Evaluation of the high resolution WRF-Chem (v3.4.1) air quality forecast and its comparison with statistical ozone predictions. Geosci. Model Dev. 8:2119-2137. DOI: <https://doi.org/10.5194/gmd-8-2119-2015>

Zhang D. y Anthes R.A. 1982. A High-Resolution Model of the Planetary Boundary Layer – Sensitivity Tests and Comparisons with SESAME-70 Data. *J. Appl. Meteorol.* 21: 1594-1609

DOI: [https://doi.org/10.1175/1520-0450\(1982\)021<1594:AHRMOT>2.0.CO;2](https://doi.org/10.1175/1520-0450(1982)021<1594:AHRMOT>2.0.CO;2)

Zhang Y. 2008. Online-coupled meteorology and chemistry models: history, current status, and outlook. *Atmos. Chem. Phys.* 8:2895-2932

DOI: <https://doi.org/10.5194/acp-8-2895-2008>

Zhang Y., Wen X.Y. y Jang C.J. 2010. Simulating chemistry aerosol cloud radiation climate feedbacks over the continental U.S. using the online-coupled Weather Research Forecasting Model with chemistry (WRF/Chem)

Zitácuaro A. y Caetano E. 2004. Simulación numérica de la circulación atmosférica local en el Valle de México. Congreso de la Asociación Española de Climatología (AEC): El Clima, entre el Mar y la Montaña. Santander, España. 2-5 Noviembre.

APÉNDICES

APÉNDICE A: Métodos iniciales de asimilación de datos

Método de aproximaciones sucesivas

Los métodos de aproximaciones sucesivas fueron introducidos por Bergthorsson y Döös (1955), al que le sucedieron otras versiones desarrolladas por Cressman (1959), Barnes (1973) y Bratseth (1986). Plantean que el campo de toda variable puede ser modificado de manera iterativa y que, en cada paso, se actualiza el valor de las variables en los puntos de la malla, basándose en el *background* y observaciones cercanas al punto. Cuando se ha recorrido todo el dominio, se va al siguiente paso.

El campo en el nodo de la malla es actualizado de acuerdo a la siguiente ecuación:

$$f_i^{m+1} = f_i^m + \frac{\sum_{k=1}^{k_i^\infty} \omega_{ik}^m (O_k - f_k^m)}{\sum_{k=1}^{k_i^\infty} \omega_{ik}^m + \varepsilon^2}$$

donde f_i^m es el valor de la variable en el punto i -ésimo en la m -ésima iteración, O_k es la k -ésima observación cercana al nodo de la malla, ω_{ik}^m es la función de peso que depende de la distancia de la observación al nodo de la malla y ε^2 es la razón de error estimada entre la observación y el *background* (si el *background* fuera perfecto entonces $\varepsilon^2 = 0$).

Estos métodos tienen dos efectos indeseables: el análisis siempre converge al dato, lo cual solo debe ocurrir cuando no existen errores en las observaciones; además, el *background* y las regiones de alta densidad de datos, tienen mucho peso en comparación con las zonas de baja densidad. En el caso del esquema de Cressman, se define la función de peso en términos del radio de influencia R_m de la forma:

$$\omega_{ik}^m = \begin{cases} \frac{R_m^2 - r_{ik}^2}{R_m^2 + r_{ik}^2} \text{ si } r_{ik}^2 \leq R_m^2 \\ 0 \text{ si } r_{ik}^2 > R_m^2 \end{cases}$$

donde r_{ik} es la distancia entre la observación y el nodo de la malla.

Las observaciones fuera del radio de influencia no son utilizadas en la actualización del campo de la variable, mientras que aquellas que sí están dentro, son tomadas en función de cuán cerca estén del nodo de la malla, considerando que las más cercanas tienen un mayor peso en el análisis. Se supone que las observaciones son perfectas ($\varepsilon^2 = 0$) y, por tanto, determinarán el campo de la variable. Además, el radio de influencia usualmente decrece con cada paso iterativo, siendo la corrección a mayor escala en los primeros pasos y a menor escala en los últimos.

En el caso del esquema de Barnes, el más utilizado cuando no se cuenta con el campo de primera aproximación, lo cual se produce, usualmente, en el análisis de fenómenos de pequeña escala, es similar al de Cressman, excepto en la función de peso utilizada:

$$\omega_{ik}^m = -\exp\left(\frac{-r_{ik}^2}{2R_m^2}\right)$$

Por su parte, Bratseth, plantea que, si las funciones de peso se escogen apropiadamente en lugar de las relaciones empíricas, el método de corrección sucesiva converge a la interpolación estadística. La variable s , es analizada en el nodo de la malla x usando la observación s_j^o . Para la n -ésima iteración se tiene:

$$s_x(n) = s_x(n-1) + \sum_{j=1}^{nobs} \alpha_{xj} [s_j^o - s_j(n-1)]$$

donde la misma ecuación ha sido aplicada en el punto de la observación i , para llegar al análisis en este punto por la iteración previa, s_i :

$$s_i(n) = s_i(n-1) + \sum_{j=1}^{nobs} \alpha_{ij} [s_j^o - s_j(n-1)]$$

donde $\alpha_{xj} = \frac{\rho_{xj}}{m_j}$ y $\alpha_{ij} = \frac{\rho_{ij + \sigma_n^2} \delta_{ij}}{m_j}$ son las funciones de peso para cada observación, σ_n es la razón entre la varianza de error de la observación y del background, δ_{ij} es la función delta de Kronecker. La correlación espacial es modelada como una Gaussiana:

$$\rho_{ij} = \exp\left(\frac{-|r_{ij}|^2}{R_h^2}\right) \exp\left(\frac{-|\Delta z_{ij}|^2}{R_z^2}\right)$$

Las funciones de peso son normalizadas por la densidad de observaciones alrededor de la localización de cada nodo de la malla:

$$m_j = \sigma_n^2 + \sum_{j=1}^{nobs} \rho_{ij}$$

En el paso inicial ($n = 1$), el análisis en el lugar de la observación $s_j^o(0)$, está dado por una interpolación trilineal de los datos de entrada.

Para hacer un análisis isentrópico, se toma la correlación vertical como una función del diferencial de la temperatura potencial en lugar de la altura, z :

$$\rho_{ij} = \exp\left(\frac{-|r_{ij}|^2}{R_h^2}\right) \exp\left(\frac{-|\Delta\theta_{ij}|^2}{R_\theta^2}\right)$$

Para acelerar la convergencia, el factor de correlación de la distancia horizontal (R) es generalmente reducido desde el paso 1 al último.

$$R(n) = k_n k_v R_o$$

Por ejemplo, $k_n = (2; 0; 1; 0; 0; 8; 0; 4)$ y k_v representa la diferencia en la correlación del modelo entre las variables, permitiendo distancias de correlación menores para variables como la humedad.

Esta técnica soluciona los defectos clásicos de los métodos de corrección sucesiva usando la función de correlación para los errores de los pronósticos para determinar su peso en el análisis, este se reduce en lugares de alta densidad de datos observacionales. Adicionalmente, el método de Bratseth, requiere incluso menor costo computacional que los de interpolación óptima.

Métodos de nudging

También conocidos como de inicialización dinámica, los métodos de *nudging* adicionan un término a la ecuación principal del modelo. La simulación es inicializada con el campo de primera aproximación y el término de *nudging* hace que la integración de las ecuaciones converja al valor de la observación en el punto. Este método balancea las condiciones iniciales suponiendo que los campos serán ajustados geostrofica e hidrostáticamente quedando la ecuación:

$$\frac{\partial u}{\partial t} = -\vec{V} \cdot \nabla u + f_v - \frac{\partial \Phi}{\partial x} + \frac{u_{obs} - u}{\tau_u}$$

donde u es el intervalo de relación. Este método es más adecuado para la asimilación de observaciones de pequeña escala como las observaciones de radar.

Método de interpolación óptima

El método de interpolación óptima es el método de mínimos cuadrados más utilizado en la asimilación de datos y difiere de los anteriores esquemas en que a las observaciones se les da un peso de acuerdo a estimados estadísticos de los errores en lugar de relaciones empíricas. Cada fuente de datos tendrá un peso diferente basándose en los errores conocidos del instrumento. Intenta minimizar el error total de todas las observaciones dándole el peso ideal. Cuando coinciden varias variables y cientos de nodos de malla y observaciones el proceso se vuelve muy complejo, aunque es la forma más científica, flexible y simple de incluir las observaciones de diferentes fuentes. Mientras que los otros esquemas de asimilación tratan a todas las fuentes por igual, este considera sus errores intrínsecos (Gandin, 1963).

Como se mencionó, este método se basa en la técnica de interpolación de mínimos cuadrados sin tener en cuenta restricciones por lo que se puede considerar como un problema de aproximación lineal de funciones. El análisis o mejor estimación $x = x_i$, del vector que representa el estado real del sistema (las observaciones) $x^t = x_i^t$, es representado a partir de la combinación lineal entre el background x^f y la diferencia entre los errores ponderados de los datos observados y del background, siendo:

$$x_i = W_i^f + \sum_{\alpha} W_{i\alpha} (d_{\alpha}^0 - d_{\alpha}^f)$$

El campo de análisis es actualizado cada 6 horas para coincidir con el horario de muchas de las observaciones meteorológicas de tipo sinóptico, por lo que se usan todas las observaciones en una ventana de tiempo de 6 horas centrada en el tiempo de análisis.

La matriz ponderada $W_{i\alpha}$ se determina por la siguiente condición: que el error medio cuadrado $\langle (x_i - x_i^t)^2 \rangle$ para cada componente del campo analizado sea minimizado. El campo se representa por:

$$W_{i\alpha} = D_{\alpha\beta}^{-1} \langle x_i^f d_{\beta}^f \rangle$$

donde:

$$D_{\alpha\beta} = \langle (d_{\alpha}^f - d_{\alpha}^0) (d_{\beta}^f - d_{\beta}^0) \rangle = \langle d_{\alpha}^f d_{\beta}^f \rangle + \langle d_{\alpha}^0 d_{\beta}^0 \rangle.$$

El cálculo de W es independiente, respecto al índice i , del estado de las componentes del vector. Se asume, en la segunda parte de la ecuación, que los errores en el campo de primera aproximación y las observaciones no están correlacionados.

La ecuación que representa a x_i , es aplicable para cualquier tipo de datos. Además, el cálculo de las matrices de los errores de covarianza, generalmente, requiere de una estadística minuciosa, la cual no siempre se encuentra disponible. En la mayoría de las aplicaciones, estas expresiones son reemplazadas por formas funcionales, utilizando escalas estimadas de correlación, las cuales pueden manejarse de forma analítica.

APÉNDICE B: Métodos variacionales de asimilación de datos

Los métodos variacionales, muy empleados en la modelación numérica de la atmósfera, permiten eliminar el paso de inicialización e incluyen dos opciones: 3DVAR (Sasaki, 1970) y 4DVAR (Bouttier y Rabier, 1997).

El método 3DVAR logra la asimilación a través de la reducción iterativa de la función de costo. Las diferencias entre el análisis y las observaciones del campo inicial son penalizadas de acuerdo a sus errores estimados. De esta forma, partiendo de una variable de control, se busca el mínimo de una función $J(x)$, el cual mide el tamaño de v y la diferencia que existe entre cada observación y^0 (distribuidas a través de una ventana de tiempo desde $T - 3$ hasta $T + 3$) y su correspondiente predicción y .

$$J(v) = \frac{1}{2} vv^T + \frac{1}{2} (y^0 - y)R^{-1}(y^0 - y)^T$$

$$y^0 = (y_{t1}^T, \dots, y_{ti}^T, \dots, y_{tm}^T)^T$$

$$y = (y_{t1}, \dots, y_{ti}, \dots, y_{tm})^T$$

En la técnica 3DVAR más básica no se considera el tiempo real de cada observación; el pronóstico usa tanto la suposición como el incremento en $T + 0$, los cuales son interpolados horizontalmente a la posición de la observación y luego verticalmente a las variables observadas. La suficiencia en el operador H está denotada por sus partes, específicamente cuando representa la interpolación horizontal de todos los campos, \tilde{H}_h es la interpolación horizontal de los incrementos y H_{ob} es el operador de las observaciones de la columna vertical del modelo. Sobre estos y sobre v, y^0 , y x^g está definido su tiempo de validez, es decir, $T - 3, T + 0, T + 3$ horas. S^{-I} es la inversa generalizada del operador de simplificación incremental:

$$y_{ti} = H_{ob} [H_h(x_0^g) + S^{-I} \{ \tilde{H}_h U v_0 \}]$$

Una variante de este tipo de métodos lo constituye 3DVAR FGAT (*First Guess at Appropriate Time*), la cual se considera una opción intermedia entre 3DVAR y 4DVAR, pues los valores del pronóstico en la posición de la observación son obtenidos por interpolación 4D de los valores del pronóstico a corto plazo, dado por:

$$y_{ti} = H_{0b}[H_{hti}(x_{-3}^g, \dots, x_0^g, \dots, x_3^g) + S^{-1}\{\tilde{H}_h U v_0\}]$$

En el caso específico de 4DVAR, existen dos versiones principales: 4DVAR sinóptico y 4DVAR en su forma básica. En el caso de 4DVAR sinóptico, se mueve la variable de control al inicio de la ventana temporal, calculando los incrementos en $T + 0$ usando los incrementos de pronóstico de $M(-3,0)$, por lo que son usados como en 3DVAR FGAT:

$$y_{ti} = H_{0b}[H_{hti}(x_{-3}^g, \dots, x_0^g, \dots, x_3^g) + S^{-1}\{\tilde{H}_h M_{(-3,0)} U v_{-3}\}]$$

Por su parte, en 4DVAR básico, el incremento de pronóstico se usa en el tiempo actual de cada observación:

$$y_{ti} = H_{0b}[H_{hti}(x_{-3}^g, \dots, x_0^g, \dots, x_3^g) + S^{-1}\{\tilde{H}_h M_{(-3,ti)} U v_{-3}\}]$$

El uso de la variable de control v es un método eficiente para remover los errores de covarianza B del background, lo cual se hace a través de la transformación de B en una matriz identidad, de modo que minimizar $J(v)$ es lo mismo q minimizar:

$$J(w') = \frac{1}{2} w' B^{-1} w^T + \frac{1}{2} (y^0 - y) R^{-1} (y^0 - y)^T$$

$$y = (y_{t1}, \dots, y_{ti}, \dots, y_{tm})^T = H(x^g, w')$$

con

$$w' = Uv, B = UU^T$$

En los métodos 4DVAR la variable de control y los incrementos son válidos en $T + 0$, lo cual se denota como:

$$w'_0 = Uv_0$$

Para el método 4DVAR básico, en el cual se cuenta con un pronóstico al inicio de la ventana, es decir, en $M(-3, ti)$ como parte de H en $J(w')$, se tiene:

$$w'_{-3} = Uv_{-3}$$

En el caso de 4DVAR sinóptico, se puede asumir la misma expresión anteriores, pero con el término $M(-3,0)$ en H . Este se relaciona más con 3DVAR que con FGAT, teniendo en cuenta que el pronóstico forma parte de la variable de transformación y el mismo es considerado en la covarianza:

$$w'_0 = M_{(-3,0)}Uv_{-3}$$

APÉNDICE C: Métodos basados en el Filtro de Kalman

El Filtro de Kalman (KF), de modo formal, parte de un caso lineal y luego pasa a su versión vectorial, proceso que se repite, sucesivamente, mediante transformaciones matemáticas descritas en Burgers et al. (1998) y Evensen (2009), hasta llegar a versiones de KF más actuales y avanzadas. Estas últimas, integran los módulos de asimilación de datos utilizados en la modelación numérica y se describen a continuación.

Filtro de Kalman Discreto (DKF)

Si se considera un sistema discreto cuyo estado en el tiempo t_n se denota por x_n , y su evolución se describe a través del modelo lineal:

$$x_{n+1} = A_n x_n + b_n + \Gamma_n w_{n+1}$$

donde x_n y b_n son vectores columna de $N \times 1$, A_n y Γ_n matrices de $N \times N$. El vector b_n y las matrices A_n y Γ_n son deterministas y conocidos. Cada vector w_n es el error del modelo en el tiempo dado.

Sea $\{z_n; n = 1; 2 \dots\}$ una sucesión de mediciones de los estados del sistema, cada vector z_n es el conjunto de L mediciones al tiempo t_n . Estas mediciones están relacionadas con el estado a través de las ecuaciones lineales de medición:

$$z_n = H_n x_n + v_n$$

Los errores del modelo $\{w_{n+1}; n = 0; 1 \dots\}$ y de medición $\{v_n; n = 1; 2 \dots\}$ son sucesiones blancas gaussianas, con media cero y matriz de covarianza dadas por Q_{n+1} y R_n , respectivamente. La sucesión de errores de medición $\{v_n\}$ y el estado x_n son independientes, mientras que la distribución de x_0 es conocida y gaussiana.

Desde el punto de vista vectorial, se puede demostrar que el estimador lineal insesgado de mínima varianza, denotado por x_{kn} , para la variable de estado en el tiempo t_n , dadas las mediciones $z_1 \dots, z_n$, es el valor esperado del estado x_k dados los datos, o sea:

$$x_n^k = E\{x_n/z_1; z_2, \dots, z_k\}$$

En esta notación el subíndice indica el tiempo en que se estima el estado y el superíndice el número de mediciones que se utilizan para obtener la estimación. La matriz de covarianza del error de esta estimación es:

$$P_n^k = E\{(x_n - x_n^k) (x_n - x_n^k)^T / z_1; z_2, \dots, z_k\}$$

donde T denota el transpuesto.

Así mismo, dada la estimación del estado en el tiempo t_n utilizando los datos disponibles:

$$x_n^n = E\{x_n/z_1; z_2; \dots; z_n\},$$

se puede hacer un pronóstico del estado en t_{n+1} usando además la ecuación del sistema:

$$x_{n+1}^n = A_n x_n^n + b_n$$

$$P_{n+1}^n = A_n P_n^n A_n^T + \Gamma_n Q_{n+1} \Gamma_n^T$$

Cuando se disponga de la próxima medición z_{n+1} esto se puede actualizar con:

$$x_{n+1}^{n+1} = x_{n+1}^n + K_{n+1}(z_{n+1} - H_{n+1}x_{n+1}^n)$$

$$P_{n+1}^{n+1} = P_{n+1}^n - K_{n+1}H_{n+1}P_{n+1}^n$$

$$K_{n+1} = P_{n+1}^n H_{n+1}^T \{H_{n+1} P_{n+1}^n H_{n+1}^T + R_{n+1}\}^{-1}$$

donde K es la ganancia de Kalman.

Estas fórmulas se utilizan secuencialmente, comenzando a partir de una estimación a priori dada, x_{00} , para el estado en el tiempo t_0 y su matriz de covarianza P_{00} . Primero se corrige la predicción, lo cual se realiza a través del cálculo de la ganancia de Kalman, K_{n+1} . Este factor de ponderación o ganancia se selecciona de tal manera que minimiza la varianza del error de la nueva estimación del estado. Luego, se incorpora la medición del proceso, z_{n+1} , para generar una nueva estimación del estado mediante la

ecuación y finalmente, se obtiene una nueva estimación de la covarianza del error utilizando la ecuación. Después de aplicar el pronóstico y las ecuaciones de corrección, se repite el proceso partiendo de la nueva estimación del estado y la nueva covarianza del error.

Filtro de Kalman Extendido (EKF)

Para modelos no lineales, aun cuando se tengan errores gaussianos con características requeridas por el KF, el estado no será gaussiano. Por tanto, para obtener el primer y segundo momentos del estado, será necesario propagar la función de densidad de probabilidad completa, lo que en general no es práctico porque requiere de un alto costo computacional. Una posibilidad es aplicar el Filtro de Kalman Extendido (EKF), en el cual se usa una ecuación de aproximación linealizada.

El proceso no lineal se describe mediante la ecuación estocástica no lineal n dimensional:

$$x_{n+1} = f(x_n) + w_{n+1}$$

y la relación entre las mediciones y los datos está dada por $z_n = h(x_n) + v_n$ donde $f(x_n)$ y $h(x_n)$ son funciones no lineales.

El modelo del proceso y la ecuación de mediciones son linealizadas respecto de la estimación actual x_n^n . Se utiliza la expansión en series de Taylor de primer orden para aproximar estas funciones.

Las fórmulas del EKF son:

$$x_{n+1}^n = f(x_n^n)$$

$$P_{n+1}^n = Vf(x_n^n)P_n^nVf(x_n^n)^T + Q_{n+1}$$

Si se dispone de la próxima medición, z_{n+1} , la última predicción se puede actualizar con:

$$x_{n+1}^{n+1} = x_{n+1}^n + K_{n+1}(z_{n+1} - \nabla h(x_n^n)x_{n+1}^n);$$

$$P_{n+1}^{n+1} = P_{n+1}^n - K_{n+1} + \nabla h(x_n^n) P_{n+1}^n$$

donde

$$k_{n+1} = P_{n+1}^n V h(x_n^n)^T \{ r h(x_n^n) P_{n+1}^n V h(x_n^n)^T + R_{n+1} - 1$$

Filtro de Kalman Ensamblado (EnFK)

Otro método para la asimilación secuencial de datos es el Filtro de Kalman Ensamblado (EnKF; Evensen, 1994), como alternativa de tipo Monte Carlo al EKF. Las matrices de covarianza del error para la predicción y la estimación actualizada en KF son:

$$P_{n+1}^n = E\{(x_{n+1} - x_{n+1}^n)(x_{n+1} - x_{n+1}^n)^T\}$$

$$P_{n+1}^{n+1} = E\{(x_{n+1} - x_{n+1}^{n+1})(x_{n+1} - x_{n+1}^{n+1})^T\}$$

donde x_{n+1} representa el estado real en t_{n+1} .

Es posible expresar estas covarianzas en términos del ensamble como:

$$P_n = \overline{(x_n + 1 - x_{nn} + 1)(x_n + 1 - x_{nn} + 1)^T}$$

$$P_n = \overline{((x_n + 1 - x_n + 1)(x_n + 1 - x_n + 1))^T}$$

En vez de almacenar una matriz de covarianza completa, podemos representar los estadísticos de error utilizando un conjunto apropiado de realizaciones del modelo. La información contenida en una función de densidad de probabilidad completa se puede representar exactamente con un ensamble infinito de estados del modelo.

El EnFK aplica el método Cadena de Markov Monte Carlo (MCMC) para resolver la ecuación diferencial parcial para la función de densidad de probabilidad, la cual se representa con un gran ensamble de estados del modelo, cuya integración hacia adelante en el tiempo, de acuerdo con la dinámica del modelo, hace que la predicción del ensamble sea similar a resolver la ecuación de Fokker Planck por un método MCMC.

El EnFK utiliza el modelo para generar realizaciones del estado, las cuales se corrigen usando la medición z_{n+1} y las ecuaciones de predicción de KF. Luego, se sustituyen las ecuaciones de predicción con cálculos de ensamble (Herrera, 2016)

APÉNDICE D: Datos PREPBUFR y BUFR

Datos PREPBUFR

Las observaciones meteorológicas mundiales de aire superior y superficie, en formato PREPBUFR, constituyen un sistema global de datos, recolectados y procesados operacionalmente por NCEP y que representan el paso final en la preparación de la mayoría de las observaciones convencionales para ser asimiladas por sistemas de asimilación y reanálisis. Incluyen informes en superficie terrestre, radiosondeos y aviones del Sistema Mundial de Telecomunicaciones, otros derivados de radares estadounidenses, datos de viento por satélite, además de presión, altura geopotencial, temperatura, temperatura del punto de rocío y dirección y velocidad del viento, procedentes de la Organización Meteorológica Mundial (OMM). Los intervalos de tiempo varían de forma horaria.

Datos de radiancia procedentes de satélites meteorológicos en formato BUFR

Para la representación de datos meteorológicos de forma operacional, fue aprobado por la OMM el formato BUFR desde 1988, como un código flexible y reducido que disminuye el espacio de almacenamiento y posibilita un mejor intercambio de datos. Una de sus formas de utilización radica en la creación y manejo del fichero PREPBUFR, el cual incluye todo tipo de observaciones meteorológicas para la inicialización de los modelos numéricos, entre ellas: observaciones de sondeos, barcos y boyas, observaciones sinópticas, satelitales, entre otras. Ha sido utilizado este formato para almacenar e intercambiar todo tipo de datos de observaciones en centros operacionales incluyendo el NCEP y bajo este código, se encuentran disponibles las observaciones satelitales organizadas por sensores que se encuentran ensamblados en los propios satélites

APÉNDICE E: Implementación del sistema WRF-Chem/DART

Para ejecutar el sistema WRF-Chem/DART, primeramente, es necesario instalar los modelos numéricos WRFV3 y WRF-Chem, los módulos de asimilación WRFDA y DART, así como todos los scripts de *shell* asociados a DART. De igual forma, para realizar la evaluación de los resultados obtenidos, se pueden utilizar las mismas herramientas que se emplean en esta investigación para evaluar el desempeño del WRF-Chem, las cuales se describen en la Sección 2.6, o en su defecto, un conjunto de scripts de MATLAB que fueron desarrollados específicamente para el diagnóstico del módulo por sus creadores.

Se comienza con el paso de pronóstico, en el cual DART ejecuta la subrutina necesaria para realizar un pronóstico de tipo ensamble al siguiente tiempo de análisis. A continuación, con las condiciones iniciales obtenidas, se ejecuta el modelo WRF, cuyas salidas en archivos *wrfout* se convierten al vector de estado de DART.

Luego, se realiza la actualización de la condición de frontera, para lo cual se construye un conjunto de archivos LBC (*Low Boundary Condition*), uno por cada miembro del ensamble, teniendo en cuenta una fuente externa especificada, como el modelo *Global Forecast System* (GFS), como media de estas LBC. Así mismo, las condiciones iniciales de cada miembro son perturbadas utilizando ruido correlacionado espacialmente, el cual es producido por el módulo de asimilación WRFDA a través de la opción "RANDOMCV".

Posteriormente, se realiza la primera predicción por conjunto, utilizando el modelo no lineal completo, para el siguiente tiempo de análisis y las condiciones iniciales obtenidas se introducen en el paso de análisis, proceso que da conclusión al ciclo inicial en WRF-Chem/DART.

El paso de análisis comienza, precisamente, a partir de este conjunto de predicciones obtenidas en el paso de pronóstico, adicionándole las observaciones que se asimilarán, para las cuales debe existir un conjunto de operadores implementados para su tratamiento (lectura, procesos y operaciones con ellas por parte de DART).

Una vez leídas todas estas observaciones en formato DART, para cada miembro del ensamble y por cada conjunto de pronósticos para cada variable observada, se calcula un conjunto de análisis a través del algoritmo EAKF. Luego, se pasa al conjunto de interfaces entre WRF-Chem y DART para lo cual se convierte el vector de estado DART al formato de las variables de WRF-Chem y se calcula la distancia desde la ubicación dada por el modelo a cualquier elemento del vector de estado. Finalmente, se ejecuta el modelo WRF-Chem para realizar el pronóstico deseado.

APÉNDICE F: Sensores satelitales utilizados

Los datos de radiancia de satélites que ofrece NCEP, proceden de diferentes sensores o instrumentos instalados en los mismos. A continuación, se describen aquellos que se utilizaron en la presente investigación, tanto para variables meteorológicas (AMSU-A, MHS y AIRS) como para las químicas (MOPITT).

Advanced Microwave Sounding Unit-A (AMSU-A)

AMSU-A es una sonda de 15 canales de microondas diseñado principalmente para obtener perfiles de temperatura en la atmósfera superior, especialmente en la estratosfera y proporciona información sobre el agua atmosférica en todas sus formas, con excepción de las pequeñas partículas de hielo que son transparentes a las frecuencias trabajadas.

Microwave Humidity Sounder (MHS)

MHS es uno de los sensores ubicados en satélites meteorológicos operativos (MetOp) de la *European Space Agency (ESA)* y la *European Organisation for the Exploitation of Meteorological Satellites (EUMETSAT)*. Constituye un radiómetro de microondas de 5 canales, diseñado para medir la aparente radiación de microondas en bandas de frecuencia específicas. Dado que la humedad en la atmósfera (hielo, nubosidad, lluvia y nieve) atenúa la radiación de microondas emitida desde la superficie de la Tierra, es posible, a partir de las observaciones realizadas por MHS, obtener detalles de la humedad atmosférica con los diferentes canales relacionados a diferentes altitudes. También se puede determinar la temperatura en la superficie de la Tierra.

Sounder Atmospheric Infrared (AIRS)

AIRS es un sensor satelital lanzado en órbita terrestre el 4 de mayo de 2002, a bordo del satélite Aqua de la NASA y realiza mediciones de la atmósfera y la superficie de la Tierra para mejorar la predicción del tiempo y observar cambios en el clima del planeta. Inaugurando una nueva generación de monitores atmosféricos operacionales, AIRS, AMSU y el Humidity Sounder Brazil (HSB) están programados para volar juntos en la segunda plataforma del Sistema de Observación de la Tierra (EOS Aqua) de la NASA.

Measurement of Pollution in the Troposphere (MOPITT)

MOPITT es un instrumento lanzado en órbita terrestre por la NASA a bordo del satélite Terra en 1999. Está diseñado para monitorear los cambios en los patrones de contaminación y su efecto en la baja atmósfera de la Tierra. El instrumento fue financiado por la *Canadian Space Agency (CSA)* y mide el CO troposférico en la escala global, permitiendo a los científicos analizar la distribución, el transporte, las fuentes y los sumideros de este contaminante atmosférico.

APÉNDICE G: Pasos para la ejecución de las técnicas de asimilación de datos 3DVAR, 4DVAR y En3DVAR en el módulo WRFDA

A continuación, se describen los elementos de ejecución a tener en cuenta para cada una de las técnicas de asimilación de datos utilizadas con WRFDA en las simulaciones numéricas con el modelo WRF-Chem, para los diferentes casos de estudio.

WRFDA/3DVAR

Una vez instalado el módulo de asimilación WRFDA, así como todas las librerías y paquetes necesarios para su ejecución y asimilación de las diferentes fuentes de datos, se procede a la configuración de los diferentes elementos del sistema para la ejecución de la técnica 3DVAR. Para ello se deben realizar los siguientes pasos:

- 1- Establecer una variable de entorno para el directorio de trabajo y para aquel donde se encuentran los datos
- 2- Editar el `namelist.input.3dvar`
- 3- Crear los links a los datos que serán asimilados (radiancia de satélite y PREPBUFR de NCEP)
- 4- Ejecutar `WRFDA 3DVAR`

WRFDA/4DVAR

Para la implementación de WRFDA/4DVAR, es necesario instalar el modelo WRFPLUS, el cual constituye una versión simplificada del modelo WRF y contiene los modelos adjunto y de la tangente lineal. Además, WRDPLUS incluye algunos módulos de parametrizaciones físicas en forma simplificada, tales como: procesos de arrastre en superficie, condensación y precipitación a gran escala y parametrización de cúmulos.

Adicionalmente, se debe considerar que, para el método de asimilación, la ventana de tiempo es tomada de forma diferente al caso de 3DVAR, es decir, se incluye un mayor volumen de datos iniciales a asimilar para un tiempo específico. En el caso de los datos de radiancia se toman 6 horarios partiendo de la fecha del análisis, mientras que para los PREBUFR se utilizan 2 archivos, debido a que el formato de estos datos sólo incluye 3 horas hacia adelante y hacia atrás respecto al horario de análisis (Figura G1)

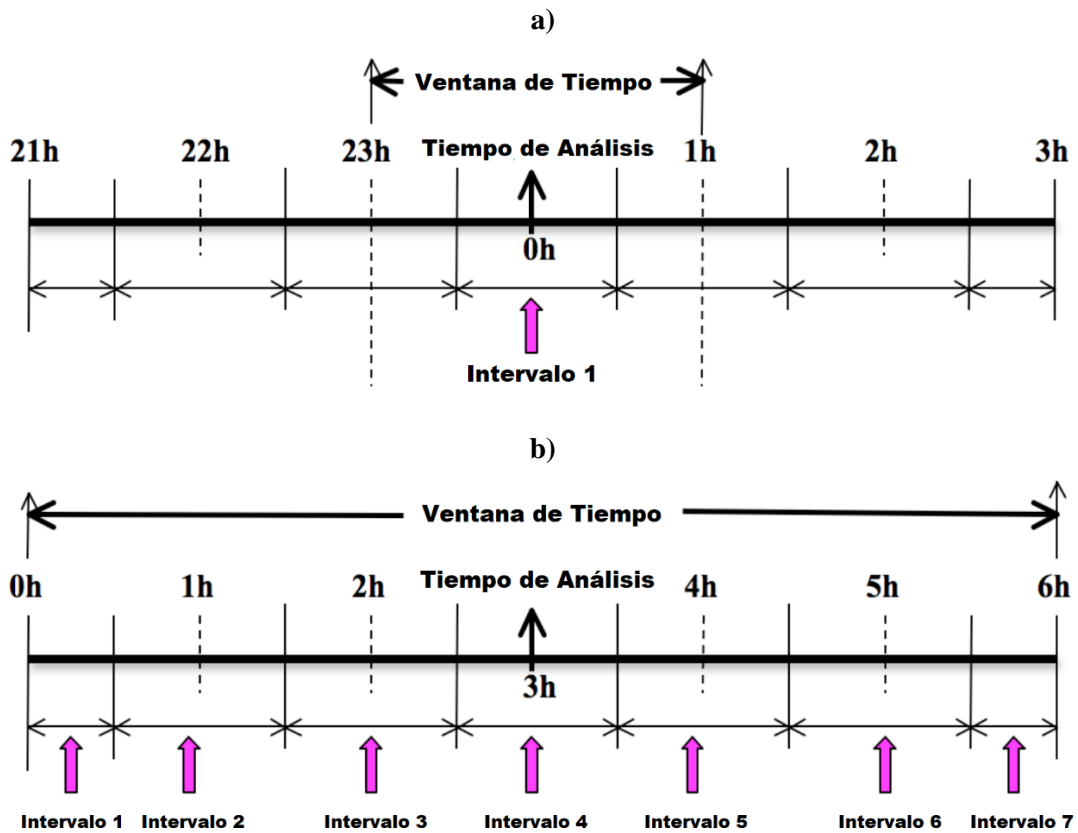


Figura G1. Ventanas de tiempo para 3DVAR (a) y 4DVAR (b)

Los pasos a desarrollar son los siguientes:

- 1- Establecer una variable de entorno para el directorio de trabajo y aquel donde se encuentran los datos.
- 2- Crear el *link* al ejecutable de WRFDA 4DVAR
- 3- Crear los *links* a las observaciones, campo de primera aproximación, BE y LANDUSE.TBL, así como a los demás archivos necesarios
- 4- Editar el `namelist.input.4dvar`
- 5- Ejecutar WRFDA 4DVAR

WRFDA/En3DVAR

Además de los métodos de asimilación de datos variacionales mencionados, el módulo WRFDA incluye otros que se basan en técnicas ensambles, específicamente ETKF, así como métodos híbridos que combinan las técnicas ensambles con las variacionales.

Para utilizar estos métodos, debe tenerse un conjunto de pronósticos que contribuya a la generación de las condiciones iniciales de tipo ensamble. WRFDA tiene un método incorporado para ello, denominado RANDOMCV, cuya función principal es agregar ruido de forma aleatoria al análisis en el espacio de la variable de control. Viene dado como una capacidad del ejecutable principal de módulo en sí, por lo que se ejecuta configurando un grupo de variables en el `namelist.input` correspondiente y luego, ejecutando `da_wrfvar.exe` como suele hacerse con cualquier ejecución de la asimilación de datos del sistema.

Debido a que las perturbaciones se producen en el espacio de la variable de control, el patrón general que presentan las perturbaciones dependerá del error correspondiente al *background*. Por lo tanto, es posible utilizar cualquiera de las opciones para el error del *background* (CV3, CV5, CV6 o CV7) sin que interfiera en el proceso que se realiza con RANDOMCV.

En cuanto a las técnicas híbridas En3DVAR y En4DVAR, su peso radica en la capacidad variacional. No obstante, la diferencia entre los esquemas híbridos y los variacionales WRFDA/3DVAR y WRFDA/4DVAR radica principalmente en que, para estos últimos, el modelo usado para estimar los errores de covarianza del *background* es estático, mientras que el híbrido usa una combinación de errores de covarianza estáticos y otros obtenidos mediante ensambles para incorporar una estimación de los errores de covarianza del *background* dependiente del flujo (Wang et al., 2008a y Wang et al. 2008b)

Se utilizan cuatro ejecutables en el sistema híbrido que se encuentran en el directorio WRFDA / var / build:

- 1- WRFDA / var / build / gen_be_ensmean.exe
- 2- WRFDA / var / build / gen_be_ep2.exe
- 3- WRFDA / var / build / da_wrfvar.exe
- 4- WRFDA / var / build / gen_be_vertloc.exe

Adicionalmente, `gen_be_ensmean.exe` se utiliza para calcular la media del ensamble, `gen_be_ep2.exe` para calcular las perturbaciones del ensamble, mientras que `gen_be_vertloc.exe` se usa para la localización vertical. Al igual que con los métodos variacionales, `da_wrfvar.exe` es el programa principal de WRFDA también para este caso, aunque debe tenerse en cuenta que `da_wrfvar.exe` se ejecutará en modo híbrido.

Para el caso de En3DVAR, el procedimiento es similar al utilizado para ejecutar 3DVAR, con la excepción de algunos archivos de entrada adicionales y configuraciones del `namelist`. Los archivos de entrada básicos para WRFDA son: `LANDUSE.TBL`, `ob.ascii` u `ob.bufnr` (según el formato de observación que utilice) y `be.dat` (errores estadísticos del background), mientras que los adicionales necesarios son: un archivo llamado `fg` que representa la media del ensamble, utilizado como campo de primera aproximación, así como otro que contenga las perturbaciones de este ensamble, para representar los errores del background.

Los pasos a seguir en el proceso de asimilación son:

- 1- Establecer una variable de entorno para el directorio de trabajo y el directorio donde se encuentran los datos
- 2- Calcular la media del ensamble
- 3- Calcular perturbaciones del ensamble
- 4- Crear el archivo de entrada para que funcione la localización vertical
- 5- Ejecutar WRFDA en modo híbrido

Después de completar el proceso de asimilación de datos y con independencia de la técnica utilizada, se crea un archivo nombrado `wrfvar_output` (el archivo de análisis de WRFDA, es decir, la nueva condición inicial para WRF) el cual aparece en el directorio de trabajo junto con varios archivos de diagnóstico que sirven para evaluar el rendimiento.

Una vez obtenido el campo de análisis, antes de ejecutar un pronóstico con WRF, es esencial que sean actualizadas las condiciones de frontera laterales del modelo, contenidas en el archivo `wrfbdy_d01`, creado por `real.exe`, para que coincidan con su nuevo análisis.

El dominio principal (`wrfbdy_d01`), debe actualizarse para que sea coherente con la nueva condición inicial (análisis) de WRFDA. Esto es imprescindible para evitar discontinuidades en el pronóstico para dicho dominio, así como resultados adversos en el límite. Para dominios anidados (dominio-2, dominio-3, etc.), las condiciones de los límites laterales las proporcionan sus respectivos dominios principales, por lo que no se necesita una actualización de las fronteras laterales para ellos.

El procedimiento de actualización lo realiza la utilidad de WRFDA denominada `da_update_bc.exe`, la cual puede encontrarse, luego de la compilación del módulo, en el directorio `$ WRFDA_DIR / var / build`. Para la ejecución de `da_update_bc.exe`, se requiere de tres archivos de entrada: el análisis WRFDA (`wrfvar_output`), el archivo `wrfbdy` de `real.exe` y un `namelist.input` nombrado `parame.in`. El resultado de este proceso es la actualización de los archivos `wrfvar_output` y `wrfbdy_d01` en su directorio de trabajo de WRFDA, los cuales constituyen las nuevas condiciones iniciales y de contorno ya con asimilación de datos que serán empleadas en las simulaciones numéricas con WRF-Chem.

APÉNDICE H: Principales directorios del sistema WRF-Chem/DART

Son tres los directorios principales del sistema WRF-Chem/DART para realizar el proceso de asimilación/modelación.

El primero de ellos, denominado `FRAPPE_DIR` o `WORK_DIR` (Figura H1) es donde se encuentran los programas compilados de WRF-Chem/DART y los scripts necesarios para su ejecución, así como también los scripts para las corridas de WRF, WRFDA, WPS y WRF-Chem. Además, aquí se realiza el post procesamiento y evaluación de los resultados obtenidos en las simulaciones numéricas.

```
[evelyn@ometeotl RUN_REAL_TIME]$ ll -rt
total 26328
-rwxrwxr-x+ 1 evelyn users      9808 Feb 27 14:27 real_time_wrf_chem.ksh
-rwxrwxr-x+ 1 evelyn users    116110 Feb 27 14:27 real_time_parameters_nested_v3.6.1.ksh
-rwxrwxr-x+ 1 evelyn users    113999 Feb 27 14:27 real_time_parameters_nested.ksh
-rwxrwxr-x+ 1 evelyn users     98620 Feb 27 14:27 real_time_parameters.ksh
-rwxrwxr-x+ 1 evelyn users    119348 Feb 27 14:27 real_time_parameters_FRAPPE_v3.6.1.ksh
-rwxrwxr-x+ 1 evelyn users    178295 Feb 27 14:27 real_time_PANDA_RETR_IDIR_SV.ksh
-rwxrwxr-x+ 1 evelyn users    196614 Feb 27 14:27 real_time_KORUS_RETR_GFS.ksh
-rwxrwxr-x+ 1 evelyn users    170970 Feb 27 14:27 real_time_FRAPPE_RETR_TEST.ksh
-rwxrwxr-x+ 1 evelyn users    171262 Feb 27 14:27 real_time_FRAPPE_RETR_TEST_FCST.ksh
-rwxrwxr-x+ 1 evelyn users    170766 Feb 27 14:27 real_time_FRAPPE_RETR_FINE_SCALE.ksh
-rwxrwxr-x+ 1 evelyn users    198359 Feb 27 14:27 real_time_FRAPPE_CPSR.ksh
-rwxrwxr-x+ 1 evelyn users    171262 Feb 27 14:27 real_time_FRAPPE_CPSR_FCST.ksh
-rwxrwxr-x+ 1 evelyn users    171028 Feb 27 14:27 real_time_FRAPPE_CNTL.ksh
-rwxrwxr-x+ 1 evelyn users    170811 Feb 27 14:27 real_time_FRAPPE_CNTL_FINE_SCALE.ksh
-rwxrwxr-x+ 1 evelyn users    197760 Feb 27 14:27 real_time_2008CASE_CPSR.ksh
-rwxrwxr-x+ 1 evelyn users    214620 Mar  6 12:32 real_time_PANDA_RETR_MA.ksh
-rwxrwxr-x+ 1 evelyn users    208889 Mar  6 12:32 real_time_PANDA_RETR.ksh
-rwxrwxr-x+ 1 evelyn users    199249 Mar  6 12:32 real_time_FRAPPE_RETR_MA.ksh
-rwxrwxr-x+ 1 evelyn users    202302 Mar  6 12:32 real_time_FRAPPE_RETR.ksh
drwxrwxr-x+ 3 evelyn users      4096 Mar  6 12:32 FINAL_TEST_SCRIPTS
-rwxrwxr-x+ 1 evelyn users    212353 Mar 30 09:38 real_time_PANDA_RETR_RELEASE_TEST_MD.ksh
-rw-rw-r--+ 1 evelyn users         0 Apr  4 13:50 dart_log.out
-rwxrwxr-x+ 1 evelyn users    214818 Apr 26 10:10 real_time_PANDA_RETR_RELEASE_TEST.ksh-
-rwxrwxr-x+ 1 evelyn users       810 May 17 13:13 do_dart.sh
-rw-rw-r--+ 1 evelyn users         0 May 17 13:13 export
-rw-rw-r--+ 1 evelyn users    19038240 May 17 13:59 dart.o76376
-rwxrwxr-x+ 1 evelyn users    219016 May 17 15:43 real_time_PANDA_RETR_RELEASE_TEST_MizzID.ksh
-rw-rw-r--+ 1 evelyn users    3297527 May 24 16:42 nohup.out
-rw-rw-r--+ 1 evelyn users        26 May 29 15:45 dart_log
-rwxrwxr-x+ 1 evelyn users    221285 May 30 15:42 real_time_PANDA_RETR_RELEASE_TEST.ksh
-rw-rw-r--+ 1 evelyn users    28011 May 30 15:42 input.nml
-rwxrwxr-x+ 1 evelyn users    574776 May 30 15:42 advance_time
[evelyn@ometeotl RUN_REAL_TIME]$
```

Figura H1. Directorio `FRAPPE_DIR` o `WORK_DIR`

El segundo directorio, denominado `real_PANDA_RELEASE_TEST` (Figura H2) es el relacionado con las salidas de cada uno de los componentes de WRF-Chem/DART.

```

[evelyn@ometeotl ~]$ cat /etc/issue: Broken pipe
evelyn@evelyn-pc:~$
evelyn@evelyn-pc:~$
evelyn@evelyn-pc:~$
evelyn@evelyn-pc:~$
evelyn@evelyn-pc:~$
evelyn@evelyn-pc:~$ ssh -l evelyn 132.248.8.195 -p 9022
evelyn@132.248.8.195's password:
Connection closed by 132.248.8.195
evelyn@evelyn-pc:~$ ssh -l evelyn 132.248.8.195 -p 9022 -Y
evelyn@132.248.8.195's password:
Last login: Mon Jun 11 23:32:18 2018 from 189.130.19.176
Bienvenido a Ometeotl

Para consultar la documentación ir a:
http://grupo-10a.atmosfera.unam.mx/hpc/ometeotl/guia_ometeotl.pdf

usuario: cca-hpc
password: cca.cwkx0*

Para sesiones interactivas:
  2 Nodos:
    srun -N 2 --ntasks-per-node 24 --pty /bin/bash
  5 Nodos:
    srun -N 5 --ntasks-per-node 24 --pty /bin/bash

[evelyn@ometeotl ~]$ cd /LUSTRE/ID/FQA/evelyn/manhattan/models/wrf_chem/run_scripts/RUN_REAL_TIME
[evelyn@ometeotl RUN_REAL_TIME]$ cd /LUSTRE/ID/FQA/evelyn/real_PANDA_RETR_VARLOC_RELEASE_TEST/2014072500/
[evelyn@ometeotl 2014072500]$ cd ..
[evelyn@ometeotl real_PANDA_RETR_VARLOC_RELEASE_TEST]$ ls
2014072418 2014072500  geogrid  pert_file_emiss  pert_file_icbc  real_PANDA_RETR_VARLOC_RELEASE_TEST_MtzzID
[evelyn@ometeotl real_PANDA_RETR_VARLOC_RELEASE_TEST]$
[evelyn@ometeotl real_PANDA_RETR_VARLOC_RELEASE_TEST]$

```

```

drwxrwxr-x+ 2 evelyn users 4096 May 23 14:15 seasons_wes
drwxrwxr-x+ 2 evelyn users 4096 May 23 14:15 exo_coldens
drwxrwxr-x+ 2 evelyn users 4096 May 23 15:12 wrfchem_bto
drwxrwxr-x+ 2 evelyn users 12288 May 23 15:13 wrfchem_fire
drwxrwxr-x+ 2 evelyn users 12288 May 23 16:14 wrfchem_chem_icbc
drwxrwxr-x+ 2 evelyn users 4096 May 23 16:19 wrfchem_cheml
drwxrwxr-x+ 2 evelyn users 20480 May 23 16:20 wrfchem_chem_emiss
drwxrwxr-x+ 2 evelyn users 4096 May 24 12:46 mopitt_co_obs
drwxrwxr-x+ 2 evelyn users 4096 May 24 14:02 prepbuf_met_obs
drwxrwxr-x+ 2 evelyn users 4096 May 24 14:03 combine_obs
drwxrwxr-x+ 2 evelyn users 4096 May 24 14:03 preprocess_obs
drwxr-xr-x+ 12 evelyn users 4096 May 24 16:13 wrfchem_initial
drwxr-xr-x+ 2 evelyn users 12288 May 29 13:30 real
[evelyn@ometeotl 2014072418]$ cd ..
[evelyn@ometeotl real_PANDA_RETR_VARLOC_RELEASE_TEST]$ cd real_PANDA
[evelyn@ometeotl 2014072418_MtzzID]$ ll -rt
total 116
drwxr-xr-x+ 2 evelyn users 4096 Apr 3 00:00 ungrib
drwxr-xr-x+ 2 evelyn users 4096 Apr 3 00:00 metgrid
drwxr-xr-x+ 2 evelyn users 12288 Apr 3 00:01 real
drwxr-xr-x+ 2 evelyn users 4096 Apr 3 01:09 wrfchem_met_bc
drwxrwxr-x+ 2 evelyn users 4096 Apr 9 14:40 wrfchem_cheml
drwxrwxr-x+ 2 evelyn users 4096 Apr 16 16:17 exo_coldens
drwxrwxr-x+ 2 evelyn users 12288 Apr 16 16:27 wrfchem_fire
drwxrwxr-x+ 2 evelyn users 4096 Apr 17 14:27 seasons_wes
drwxrwxr-x+ 2 evelyn users 12288 Apr 18 16:14 wrfchem_chem_icbc
drwxr-xr-x+ 22 evelyn users 12288 Apr 18 16:43 wrfchem_met_ic
drwxrwxr-x+ 2 evelyn users 4096 Apr 18 17:12 wrfchem_bto
drwxrwxr-x+ 2 evelyn users 20480 Apr 19 14:46 wrfchem_chem_emiss
drwxrwxr-x+ 2 evelyn users 4096 Apr 19 15:35 mopitt_co_obs
drwxrwxr-x+ 2 evelyn users 4096 Apr 19 16:26 prepbuf_met_obs
drwxrwxr-x+ 2 evelyn users 4096 Apr 19 16:26 combine_obs
drwxrwxr-x+ 2 evelyn users 4096 Apr 19 16:31 preprocess_obs
drwxr-xr-x+ 12 evelyn users 4096 Apr 24 14:13 wrfchem_initial
[evelyn@ometeotl 2014072418_MtzzID]$
drwxr-xr-x+ 22 evelyn users 12288 Apr 18 16:43 wrfchem_met_ic
drwxrwxr-x+ 2 evelyn users 4096 Apr 18 17:12 wrfchem_bto
drwxrwxr-x+ 2 evelyn users 20480 Apr 19 14:46 wrfchem_chem_emiss
drwxrwxr-x+ 2 evelyn users 4096 Apr 19 15:35 mopitt_co_obs
drwxrwxr-x+ 2 evelyn users 4096 Apr 19 16:26 prepbuf_met_obs
drwxr-xr-x+ 12 evelyn users 4096 Apr 24 14:13 wrfchem_initial
[evelyn@ometeotl 2014072418_MtzzID]$ cd ..
[evelyn@ometeotl real_PANDA_RETR_VARLOC_RELEASE_TEST]$ cd 2014072500/
[evelyn@ometeotl 2014072500]$ ll -rt
total 140
drwxr-xr-x+ 2 evelyn users 4096 May 29 15:09 ungrib
drwxr-xr-x+ 2 evelyn users 4096 May 29 15:14 metgrid
drwxr-xr-x+ 22 evelyn users 12288 May 29 16:32 wrfchem_met_ic
drwxr-xr-x+ 2 evelyn users 4096 May 29 16:34 wrfchem_met_bc
drwxrwxr-x+ 2 evelyn users 4096 May 29 16:34 exo_coldens
drwxr-xr-x+ 2 evelyn users 12288 May 30 11:45 real
drwxrwxr-x+ 2 evelyn users 4096 May 30 14:09 seasons_wes
drwxrwxr-x+ 2 evelyn users 4096 May 30 15:06 wrfchem_bto
drwxrwxr-x+ 2 evelyn users 4096 May 30 15:15 wrfchem_fire
drwxrwxr-x+ 2 evelyn users 4096 May 30 15:22 wrfchem_cheml
drwxrwxr-x+ 2 evelyn users 12288 May 30 15:25 wrfchem_chem_icbc
drwxrwxr-x+ 2 evelyn users 20480 May 30 15:26 wrfchem_chem_emiss
drwxrwxr-x+ 2 evelyn users 4096 May 30 15:32 mopitt_co_obs
drwxrwxr-x+ 2 evelyn users 4096 May 30 15:45 prepbuf_met_obs
drwxrwxr-x+ 2 evelyn users 4096 May 30 15:46 combine_obs
drwxrwxr-x+ 2 evelyn users 4096 May 30 15:47 preprocess_obs
drwxr-xr-x+ 22 evelyn users 4096 May 30 17:13 dart_filter
drwxr-xr-x+ 2 evelyn users 4096 May 30 17:14 update_bc
drwxr-xr-x+ 12 evelyn users 4096 May 30 18:00 wrfchem_cycle_cr
drwxr-xr-x+ 2 evelyn users 12288 May 30 18:07 ensemble_mean_input
drwxr-xr-x+ 2 evelyn users 12288 Jun 4 16:27 ensemble_mean_output
[evelyn@ometeotl 2014072500]$

```

Figura H2. Directorio real_PANDA_RELEASE_TEST

Por último, en el directorio `PANDA_REAL_TIME_DATA`, se almacenan todas las fuentes de observación a asimilar y las condiciones iniciales (Figura H3).

```
drwxr-xr-x  4 gmp_a agr_g 4096 Mar 12 00:45 RUN_REAL_TIME
[gmp_a@mn328 run_scripts]$ cd
[gmp_a@mn328 ~]$ cd $TMPU
[gmp_a@mn328 gmp_a]$ cd PANDA_REAL_TIME_DATA/
[gmp_a@mn328 PANDA_REAL_TIME_DATA]$ ll -rt
total 68
drwxr-xr-x  2 gmp_a agr_g  4096 Apr  5  2017 airnow_csv_data
drwxr-xr-x  2 gmp_a agr_g 16384 Apr  5  2017 anthro_emissions
drwxr-xr-x  4 gmp_a agr_g  4096 Apr  5  2017 bio_emissions
drwxr-xr-x  6 gmp_a agr_g  4096 Apr  5  2017 fire_emissions
drwxr-xr-x 14 gmp_a agr_g  4096 Apr  5  2017 gfs_forecasts
drwxr-xr-x  2 gmp_a agr_g  4096 Apr  5  2017 hist_io_files
drwxr-xr-x  3 gmp_a agr_g  4096 Apr  5  2017 iasi_co_hdf_data
drwxr-xr-x  2 gmp_a agr_g  4096 Apr  5  2017 iasi_o3_hdf_data
drwxr-xr-x  5 gmp_a agr_g  4096 Apr  5  2017 modis_aod_hdf_data
drwxr-xr-x  2 gmp_a agr_g  4096 Apr  5  2017 mozart_forecasts
drwxr-xr-x  2 gmp_a agr_g  4096 Apr  5  2017 mopitt_co_hdf_data
drwxr-xr-x 18 gmp_a agr_g  4096 Apr  5  2017 met_obs_prep_data
drwxr-xr-x  2 gmp_a agr_g  4096 Apr  5  2017 static_files
drwxr-xr-x  2 gmp_a agr_g  4096 Nov 15  2017 wes_coldens
[gmp_a@mn328 PANDA_REAL_TIME_DATA]$
```

Figura H3. Directorio `PANDA_REAL_TIME_DATA`

APÉNDICE I: Resultados de las simulaciones numéricas con el sistema WRF-Chem/DART para el caso de prueba

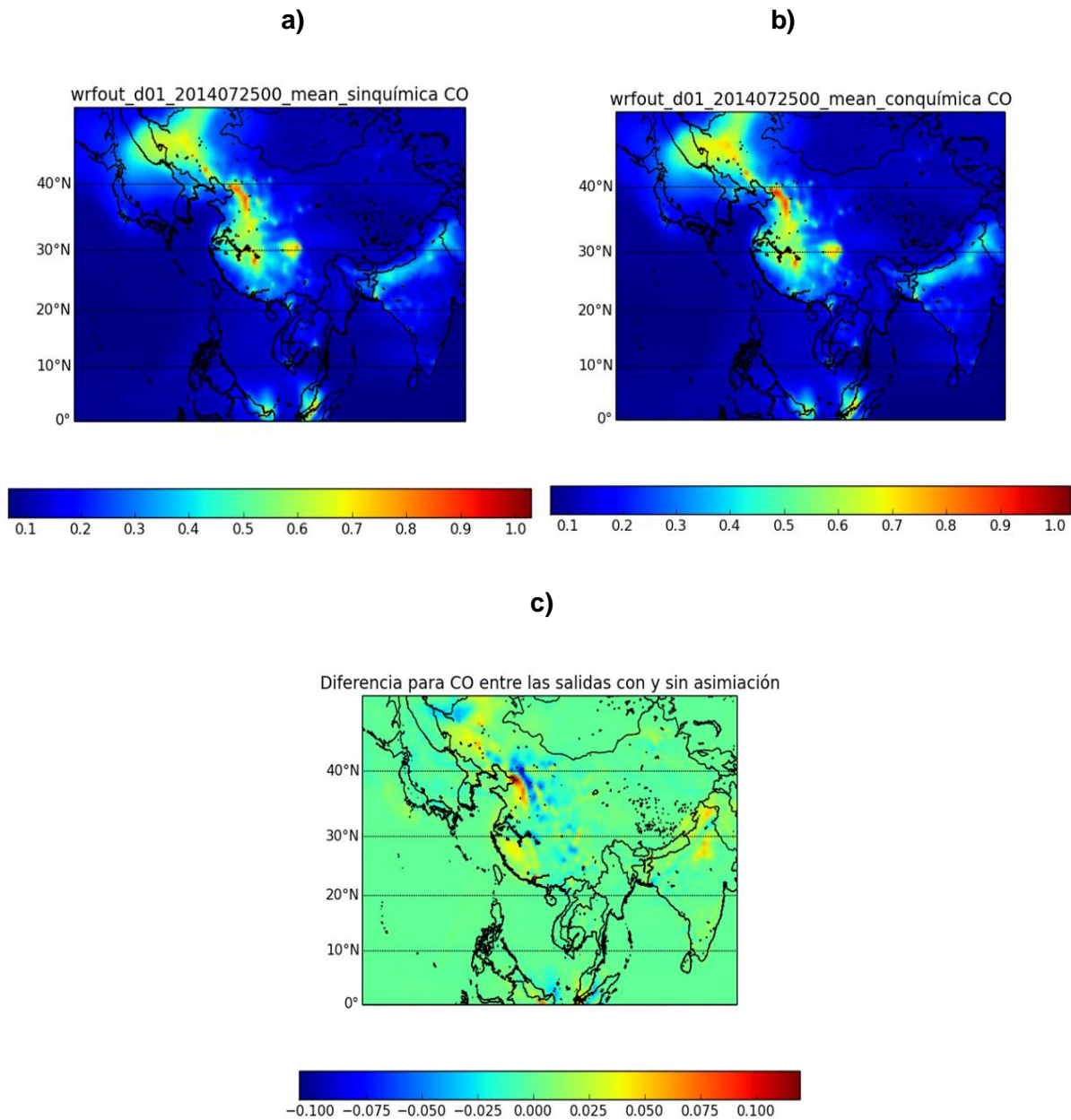


Figura I1. Distribución espacial de CO (media del ensamble en el primer paso de pronóstico) para el dominio de Prueba con asimilación de datos químicos de MOPITT (a), sin asimilación (b) e incrementos de observación (c)

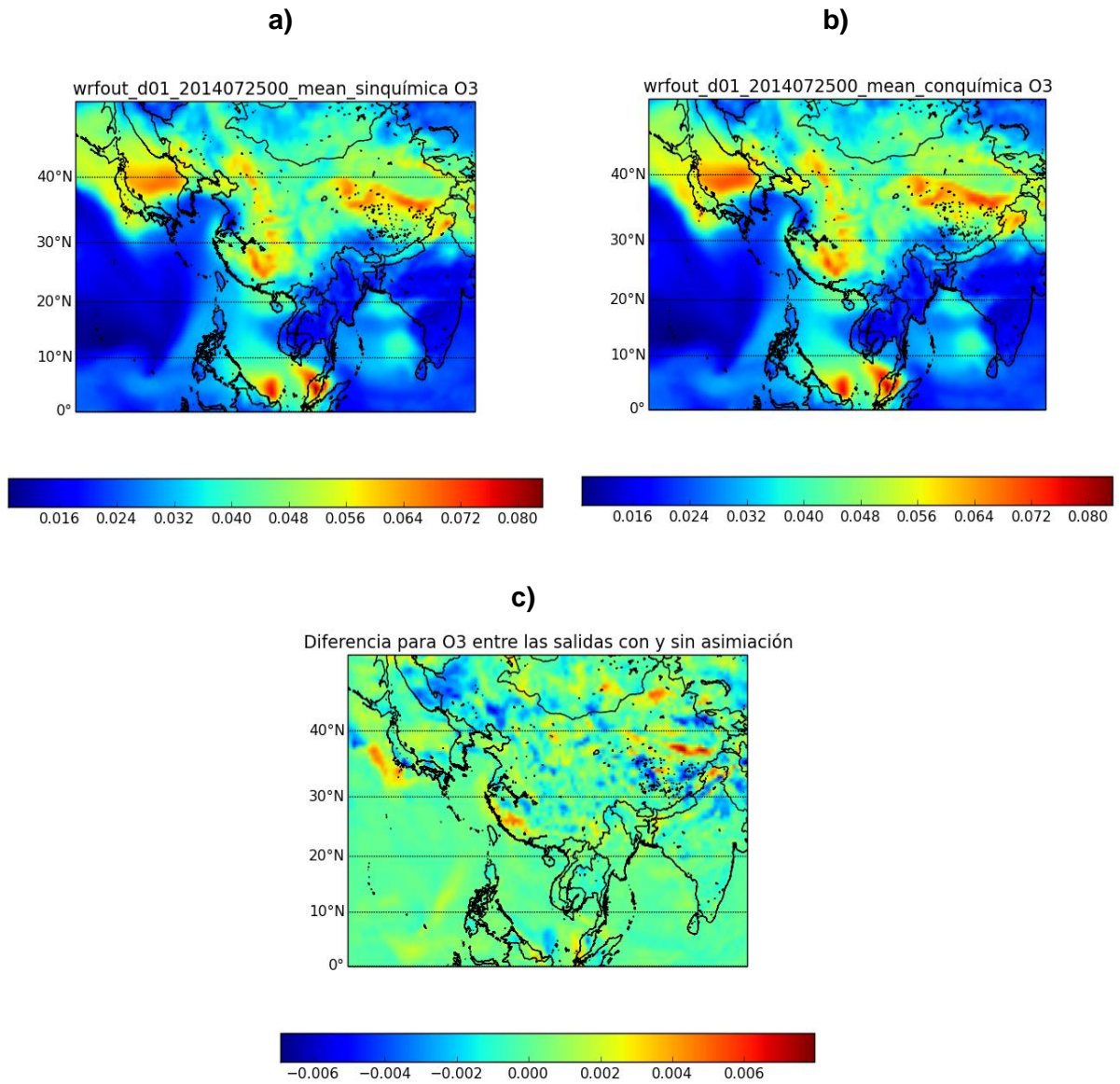


Figura I2. Distribución espacial de O₃ (media del ensamble en el primer paso de pronóstico) para el dominio de Prueba con asimilación de datos químicos de MOPITT (a), sin asimilación (b) e incrementos de observación (c)

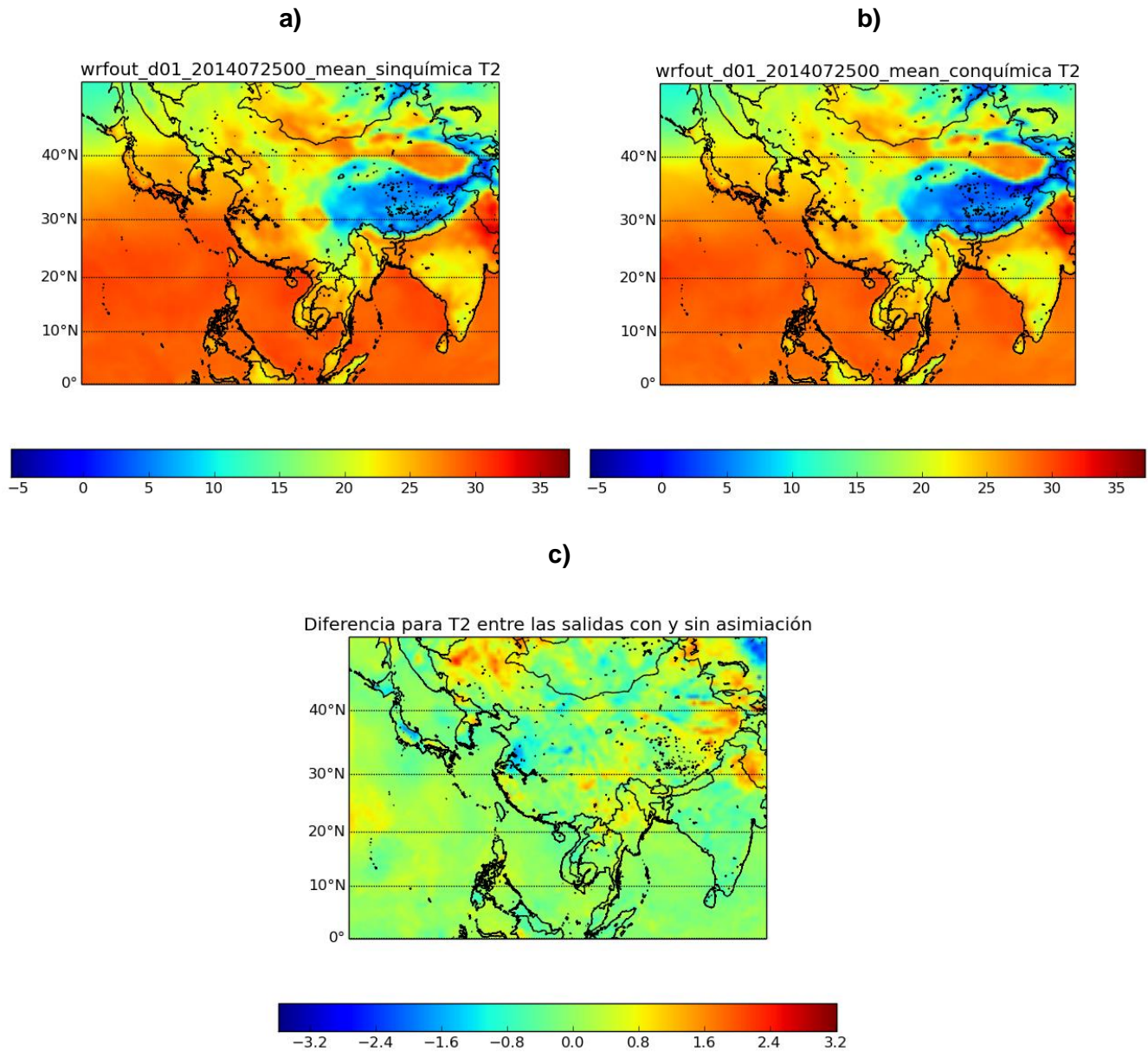


Figura I3. Distribución espacial de T2 (media del ensamble en el primer paso de pronóstico) para el dominio de Prueba con asimilación de datos químicos de MOPITT (a), sin asimilación (b) e incrementos de observación (c)

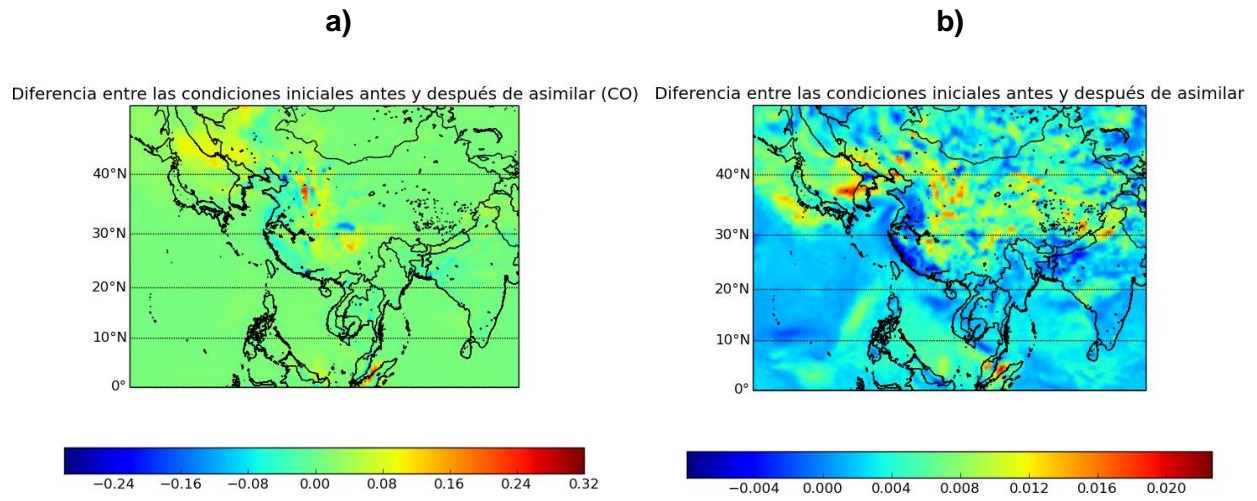


Figura I4. Campos de diferencias entre las condiciones iniciales antes de después de la asimilación de datos para las variables CO (a) y O₃ (b)