

# ASIMILACIÓN DE DATOS Y ESTIMACIÓN DE PARÁMETROS EN MODELOS DE QUÍMICA ACOPLADOS Y NO ACOPLADOS

Félix Carrasco Galleguillos<sup>1,2</sup>, Juan José Ruiz<sup>1,2</sup>, Axel Osses<sup>4,5</sup>, Celeste Saulo<sup>1,2,3</sup>

[felix.carrasco@cima.fcen.uba.ar](mailto:felix.carrasco@cima.fcen.uba.ar)

<sup>1</sup> Centro de Investigaciones del Mar y la Atmósfera, CONICET

<sup>2</sup> Departamento de Ciencias de la Atmósfera y los Océanos. FCEyN UBA

<sup>3</sup> Servicio Meteorológico Nacional

<sup>4</sup> Centro de Modelación Matemática, Santiago, Chile

<sup>5</sup> Departamento de Ingeniería Civil Matemática, Universidad de Chile

## RESUMEN

El presente trabajo tiene como objetivo comparar la asimilación de constituyentes inertes en la atmósfera, cuando la misma se realiza en forma acoplada con la asimilación de las variables atmosféricas o en forma desacoplada. Dicho impacto se evalúa tanto para la estimación de las concentraciones como así también para la estimación de las fuentes. Para ello se utiliza la metodología de experimentos gemelos con un modelo de dispersión, construido en base a las variables del modelo de Lorenz 96. Los resultados son evaluados en diversos escenarios en los que se varía la cantidad de observaciones disponibles. Por otra parte se evalúan también los parámetros de inflación y localización óptimos para la asimilación acoplada y desacoplada.

Los resultados muestran que ambas estrategias permiten realizar una estimación de las variables que representan el flujo atmosférico como así también de las concentraciones. No obstante, la implementación acoplada muestra una clara ventaja sobre la desacoplada. Dicha ventaja se encuentra tanto en las concentraciones como en las variables de flujo. La implementación desacoplada mejora significativamente si se incorpora de alguna manera la incertidumbre en los flujos dentro del sistema de asimilación para las concentraciones. No obstante, aún en este caso los resultados son inferiores a los que se obtienen con el método acoplado. La ventaja del método acoplado es mayor cuando la cantidad de observaciones de concentración es limitada. Cuando se realiza la asimilación para la estimación conjunta de concentraciones y las fuentes, también se encuentran mejores resultados para el caso acoplado.

## ABSTRACT

This work compare data assimilation (DA) of inert atmospheric constituents in the atmosphere using two different approaches: one in which the dispersion model is coupled with the atmospheric model and one in which both models are uncoupled. The impact of these two approaches upon the estimation of the concentrations as well as the source terms in the model is addressed. A twin experiments methodology has been implemented with simple model consisting of coupling the Lorenz 96 system with a continuity equation for the concentration of one inert constituent. The comparison is performed using observation networks with different observation number and spatial distribution. The optimal values for the inflation factor and localization length used in the DA scheme are estimated for the coupled and uncoupled cases. It was found that optimal localization scales for the concentration variables may be different from the ones used for the

meteorological variables.

Overall, the results show that both strategies produce a reasonable estimation of the atmospheric variables and of the concentrations. The coupled configuration shows a clearly advantage over the uncoupled for both concentration and atmospheric variables. An implementation of the uncoupled approach in which the uncertainty of the atmospheric variables is included improves the results with respect to the standard uncoupled implementation. Nonetheless, the results for the coupled configuration are better than the uncoupled particularly, when the amount of concentration observation is reduced. We also found that the coupled approach produces a better estimation of the sources of constituents.

**Palabras clave:** LETKF, Estimación de parámetros, química atmosférica.

## 1) INTRODUCCIÓN

Durante los últimos 20 años la predicción del tiempo químico ha tenido muchos avances y aplicaciones, llegando incluso a implementaciones operativas en distintos países, tal como se muestra en el trabajo de Kukkonen et al. (2012) que coloca en perspectiva los desarrollos en esta materia en la comunidad europea señalando el esfuerzo conjunto para tener pronósticos de compuestos químicos y aerosoles provenientes de distintos modelos que se corren en distintos centros a lo largo de Europa. Este trabajo aborda la experimentación de técnicas de asimilación de datos en el contexto de la química atmosférica, para la búsqueda de una mejora en los pronósticos de las concentraciones. Si bien en la temática de la predicción numérica del tiempo se viene utilizando la asimilación hace varios años, recientemente las mismas técnicas comenzaron a utilizarse en la química atmosférica para obtener mejoras tanto en la estimación de la concentración de diferentes compuestos y en su pronóstico, así como también en la estimación de las fuentes para la generación de los inventarios de emisión, en la cual podemos incluir los trabajos de Bocquet (2011), Kang et al. (2011), Saide et al. (2011) y Khade et al. (2012).

Uno de los parámetros importantes para correr un modelo de dispersión corresponde a los inventarios de emisión de cada compuesto, es decir, el parámetro que modula cuanto se emite, donde se emite y que se emite. Debido a las distintas metodologías que se adoptan en distintos países, resulta complejo tener un inventario unificado y comparable, lo que aumenta la incertidumbre de la modelación. Por ello una de las aplicaciones de la asimilación de datos en el área de química atmosférica, es la estimación de las fuentes de emisión utilizando observaciones de los compuestos en la atmósfera. Existen varios trabajos en los cuales se estiman diferentes tipos de fuentes y a distintas escalas espaciales y temporales. Uno de ellos es el trabajo de Bocquet (2011) en el cual se estima una fuente puntual asociada al accidente de Chernobyl usando una metodología variacional (4Dvar) aplicada a un modelo desacoplado. El trabajo de Kang et al. (2011) utiliza una versión simplificada de un modelo atmosférico acoplado con una ecuación de dispersión para simular la distribución del Dióxido de Carbono (CO<sub>2</sub>) y utilizando la técnica LETKF (Local Ensemble Transform Kalman Filter) para obtener conjuntamente análisis de concentración y de la estimación de fuentes de CO<sub>2</sub> a escala global. Finalmente podemos mencionar el trabajo de Saide et al. (2011) en el cual se estima el inventario de emisión para la ciudad de Santiago de Chile pero considerando transformaciones no gaussianas en el ciclo de asimilación, también de forma desacoplada. Dentro de estos trabajos, podemos reconocer que existen dos metodologías abordar la estimación de las concentraciones. La primera es considerando el trabajo de Bocquet (2011), en el cual solo considera un ciclo de asimilación sólo para las variables de concentración, totalmente desacoplado de lo que ocurre con las variables

meteorológicas. La segunda, la podemos identificar con el trabajo de Kang et al. (2011) en el cual el ciclo de asimilación se realiza conjuntamente para variables meteorológicas y de concentración. Ahora bien, en este trabajo y tal como se realiza en varios trabajos de asimilación, se utilizará un modelo simplificado para comprobar la habilidad de una técnica de asimilación en el contexto de dispersión de contaminantes (e.g Kang et al., 2011; Yang et al., 2009).

Bajo este contexto se propone estudiar una técnica de asimilación (LETKF) propuesta en el trabajo de Hunt et al. (2007), en un modelo de dispersión simple. En este caso, se ha escogido el LETKF al ser una técnica actual, eficiente y aplicable independiente del modelo que se utilice y que se ha utilizado ampliamente en distintos modelos. El modelo simple que se utiliza en este trabajo corresponde a acoplar las variables del modelo de Lorenz 96 (Lorenz, 1996), usadas como una representación del viento, con una ecuación de continuidad para un compuesto ficticio, en el cual se incluyen los procesos de transporte, decaimiento y emisión. Este modelo fue propuesto originalmente en el trabajo de Bocquet y Sakov (2013) investigando otra técnica de asimilación y comparándola con el LETKF. Sin embargo, al ser una representación simplificada de un modelo de transporte de contaminantes, resulta de interés profundizar los resultados de este modelo al utilizar el LETKF para una futura implementación en un modelo de mayor complejidad como lo es el modelo Weather Research Forecast-Chem (Grell et al., 2005).

## 2) MODELOS DE DISPERSIÓN

El modelo de dispersión simple es presentado en el trabajo de Bocquet y Sakov (2013) y consiste en acoplar las ecuaciones del sistema Lorenz 96 (Lorenz, 1996), con una ecuación de continuidad de masa para un compuesto inerte. Las ecuaciones para el sistema de Lorenz corresponden a:

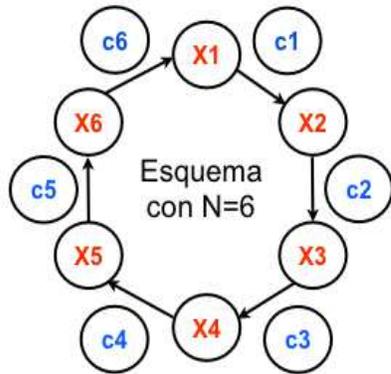
$$\begin{aligned} \frac{dx_i}{dt} &= x_{i-1}(x_{i+1} - x_{i-2}) - x_i + F \\ i &= 1, 2, \dots, N \end{aligned} \quad (1)$$

donde  $x_i$  representa el valor de una variable meteorológica en el  $i$ -ésimo punto de retícula y  $F$  representa un forzante que en este trabajo se supone constante en el espacio y en el tiempo y con un valor igual a 8. En el caso particular de este trabajo, consideraremos esta ecuación, como una representación del viento, a la cual acoplamos la siguiente ecuación de transporte para representar las variaciones en la concentración de un compuesto inerte:

$$\begin{aligned} \frac{dc_{i+\frac{1}{2}}}{dt} &= \Phi_i - \Phi_{i+1} - \lambda c_{i+\frac{1}{2}} + E_{i+\frac{1}{2}} \\ \Phi_i &= x_i c_{i-\frac{1}{2}}, \text{ si } x_i \geq 0 \\ \Phi_i &= x_i c_{i+\frac{1}{2}}, \text{ si } x_i < 0 \end{aligned} \quad (2)$$

donde  $c_i$ , representa la concentración del compuesto,  $\Phi_i$  representan el flujo del compuesto en el  $i$ -ésimo punto de retícula,  $\lambda$  representa la tasa de decaimiento (que puede ser producida por decaimiento químico o depósito) y  $E_{i+1/2}$  representa la emisión del compuesto entre los puntos  $i$  e  $i+1$ . La Figura 1 muestra un esquema del modelo cuando el valor de  $N$  es seis. En este caso es posible ver como se ordena cada una de las variables, notando que entre cada variable de viento, se intercala una de concentración.

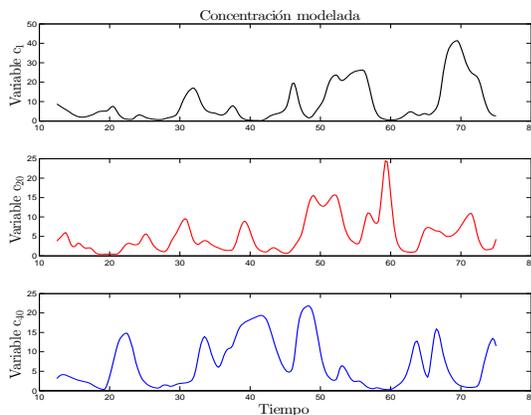
Para resolver este sistema de ecuaciones se utiliza un método Runge-Kutta de orden cuatro con paso de tiempo  $dT = 0.01$ . En este trabajo se consideró  $N = 40$  al igual que en Lorenz 1996. La Tabla I resume los valores de los parámetros utilizados en los experimentos realizados en este trabajo. La Figura 2, muestran la evolución temporal del sistema utilizando los valores presentados en la tabla I. Los valores de concentración obtenidos por este modelo muestran una importante variabilidad espacial y temporal que se asemeja a la variabilidad temporal y espacial de la concentración de algunos compuestos traza en la atmósfera.



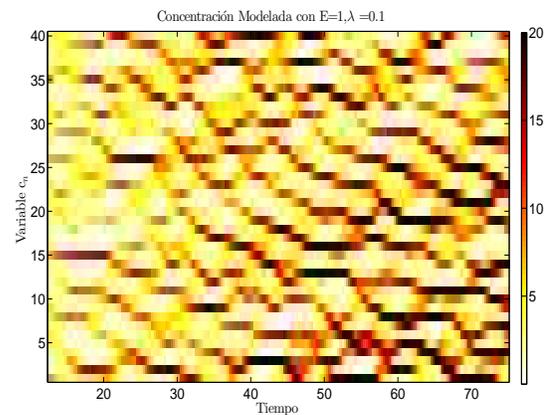
**Figura 1: Esquema simplificado del modelo en el caso N=6.**

Parámetros	Valor
Número de Variables ( $N$ )	40
Paso de tiempo ( $dT$ )	0.01
Forzante de Lorenz ( $F$ )	8
Decaimiento ( $\lambda$ )	0.1
Emisión ( $E_{i+1/2}$ )	1

**Tabla 1: Valores de parámetros para resolver el modelo.**



(a)



(b)

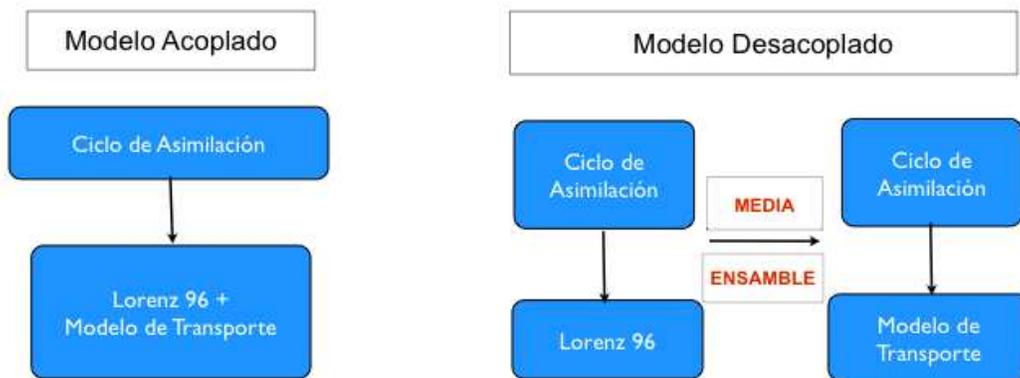
**Figura 2: Series de tiempo para (a) concentración en 3 ubicaciones diferentes y (b) diagrama espacio tiempo de la concentración incluyendo todos los puntos del modelo.**

### 3) METODOLOGÍA

Antes de estudiar el desempeño del LETKF en el modelo de transporte simple, procedemos a dar una breve descripción del ciclo de asimilación. El LETKF (Hunt et al., 2007) es una técnica derivada del filtro de Kalman (Kalman, 1960), pero en donde el modelo es no lineal, se considera un efecto localizado de la influencia de las observaciones en las variables y la evolución de la covarianza se estima utilizando un ensamble de realizaciones del modelo.

Para testear el desempeño de esta técnica de asimilación en el modelo de transporte simple, se utiliza la técnica de experimentos gemelos, la cual consiste en realizar una simulación con el modelo que será considerada como la verdadera evolución del sistema. A partir de dicha simulación se genera un conjunto de observaciones de viento y de concentración, tomando los valores de viento y concentración simulados por el modelo en diferentes puntos. A fin de representar el efecto del error observacional, a dichos valores se los perturba con un error aleatorio con distribución normal de promedio cero y varianza unitaria. En el caso de que las observaciones de concentración adquiriera un valor negativo debido a esta perturbación se le asigna el valor 0 a la observación, sin embargo esta situación es raramente observada en estos experimentos. Luego se procede a realizar los experimentos de asimilación de 200 ciclos cada uno en donde se asimilan las observaciones generadas. La longitud del ciclo de asimilación se fija en 0.05 unidades de tiempo que para el modelo de Lorenz, representa una tasa de crecimiento del error similar al que se observa en plazos de 6 horas en modelos de atmósfera.

Los experimentos de asimilación se llevaron a cabo con una versión acoplada del modelo de dispersión y con una versión desacoplada. La figura 3 muestra las diferencias entre ambos esquemas. En el esquema acoplado, el modelo de dispersión y el modelo meteorológico se encuentran acoplados y la asimilación se realiza en forma simultánea para las variables de viento y para las variables de concentración. El caso desacoplado, el modelo de concentración y viento están desacoplados. En primera instancia se realiza el pronóstico y asimilación para las variables de viento. Luego los valores de viento obtenidos son utilizados para forzar el modelo de concentración y realizar el pronóstico y asimilación para la estimación de la concentración. En el caso desacoplado, se implementaron dos variantes: una en la cual todos los miembros del ensamble de concentración utilizan la media del ensamble de vientos como forzante (desacoplado-media), y otra en la que se utiliza cada uno de los miembros del ensamble de viento para forzar cada uno de los miembros del ensamble de concentración (desacoplado-ensamble). Esta ultima incluye una representación de la incertidumbre que se tiene en el viento al asimilar las variables de concentración. El ensamble de modelos de viento tiene la misma cantidad de miembros que el ensamble de modelos de concentración.



**Figura 3: Esquema de configuraciones acoplado y desacoplado**

En primer lugar se realizaron experimentos de asimilación para determinar los valores óptimos de la inflación multiplicativa ( $ro$ ) (Anderson y Anderson 1999) y la escala de localización utilizados en el

LETKF. El factor de inflación compensa parcialmente la subestimación de la covarianza que se produce en los filtros de Kalman por ensambles, mientras que la escala de localización determina el radio de influencia de las observaciones y se utiliza para reducir el efecto de valores espurios que puedan aparecer en la matriz de covarianza de los errores del pronóstico debido a que la misma se estima a partir de un número reducido de miembros en el ensamble.

Dentro de los objetivos de este trabajo está el probar la habilidad del LETKF para la estimación de parámetros dentro del modelo Lorenz-Transporte, para lo cual se escogió la emisión ( $E_{i+1/2}$ ). La estimación de la emisión se realizó incorporando dicho parámetro como una variable más a ser estimada en el proceso de asimilación. En los experimentos que se presentan en este trabajo el valor de la emisión se considero espacialmente uniforme, es decir que solo se estimó su valor global y no su valor en cada punto de retícula. La estimación se llevó a cabo utilizando las diferentes variantes (acoplada, desacoplada-media y desacoplada-ensamble). En estos experimentos los parámetros descritos en la sección anterior como el tamaño del ensamble, el factor de inflación y la localización, se aplican también en esta serie de experimentos.

#### 4) RESULTADOS

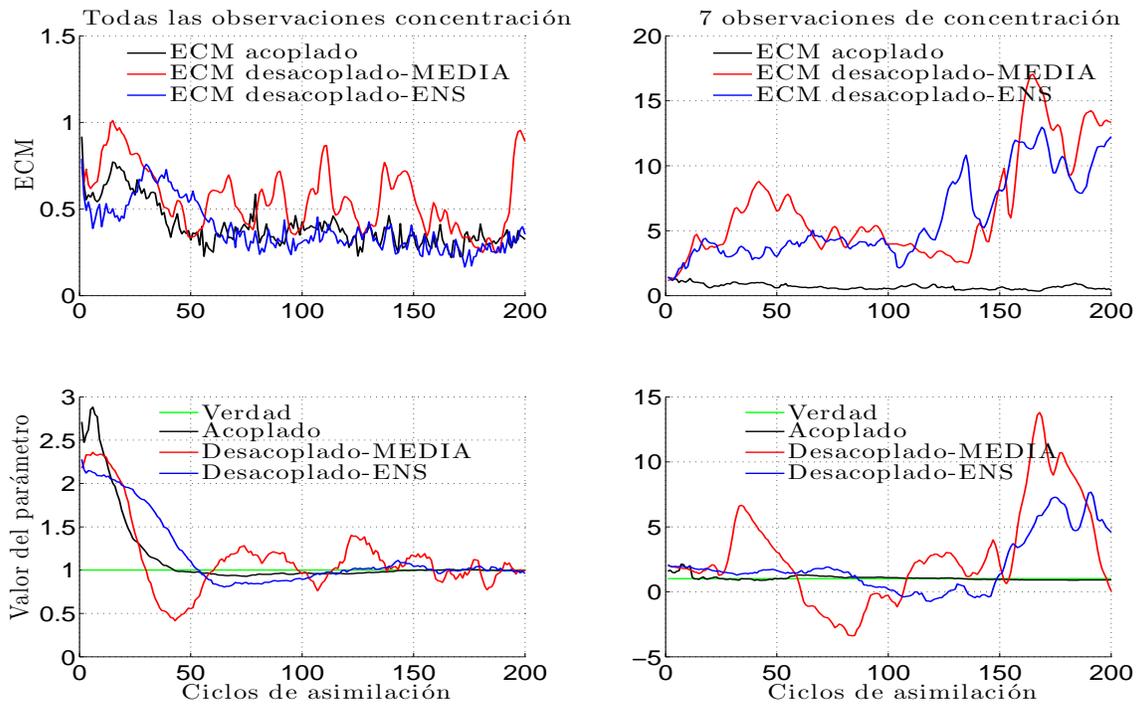
A continuación se muestran algunos de los resultados obtenidos. Los primeros experimentos que se plantearon fueron para determinar el valor óptimo de los parámetros de inflación y localización es decir, aquellos que producen un valor mínimo en términos del error cuadrático medio (ECM). Para ello se realizan experimentos de asimilación en los cuales se varía el parámetro de inflación ( $r_0$ ) y el radio de localización ( $l$ ) para finalmente calcular el ECM promediado temporal y espacialmente para cada experimento. En todos ellos, la cantidad de observaciones es igual a la cantidad de variables del sistema. Las figuras 4 y 5 muestran los valores de ECM en función del parámetro de inflación y el parámetro de localización para un tamaño de ensamble igual a veinte para el modelo acoplado y desacoplado-media.

Por un lado podemos destacar que en el caso desacoplado-media, al no tener una representación adecuada de la variabilidad del modelo por sólo contar con una realización de los vientos, es necesario utilizar un valor alto para el factor de inflación ( $r_0 > 1.6$ ). También es destacable que en ambas zonas, la influencia de la longitud de localización es mayor en el modelo acoplado que en el modelo desacoplado-media, que se puede explicar a través de las diferentes relaciones en las variables que se utilizan en la formulación del modelo de Lorenz y del modelo de transporte.

Una vez obtenidos los valores óptimos de estos parámetros, se comparan las series de tiempo de ECM tanto para variables de Lorenz, como de concentración, obtenidos con estos valores. Los resultados de esta comparación se muestran en la Figura 6 en la cual se puede apreciar que para variables de Lorenz, las diferentes configuraciones (acoplado y desacoplado) no tienen un impacto significativo. Cuando se analizan las series temporales de ECM para la concentración, se encuentra que el mayor error se comete cuando se utiliza la configuración desacoplado-media y que tanto la configuración acoplada como la desacoplada-ensamble tienen valores de ECM comparables y no se aprecia una diferencia significativa.



de esta figura, se puede mirar primero las tres curvas entre el 50% y 100%, donde las configuraciones de tipo acoplado y desacoplado-ensamble tienen un desempeño similar y se encuentran bajo el umbral de ECM dado por la varianza de las observaciones que definimos con valor unitario y está graficado con color verde. La curva correspondiente al desempeño del modelo desacoplado-media siempre esta sobre las curvas de las otras dos configuraciones incluyendo su desviación estándar. Esto se debe a que no se incluye una representación de la variabilidad del modelo de Lorenz 96 en el ciclo de asimilación, dado que sólo se utiliza el promedio de los ensambles. Cuando se consideran los porcentajes que están entre el 10% y el 50%, la única curva que se mantiene bajo el umbral de ECM de las observaciones, es la curva del modelo acoplado, las otras dos configuraciones mantienen valores bastante parecidos tanto en promedio, como en desviación. De esta forma es posible ver que el desempeño del modelo acoplado en el caso de una red de observaciones menos densa es mucho mejor que los ambos casos desacoplados. Un hecho destacable sobre los modelos desacoplados es que la desviación estándar aumenta a medida que el número de observaciones disminuye, mostrando el impacto que tiene en el resultado la posición de las observaciones.



**Figura 8: Resultados de estimación de fuente constante con alta (columna de la izquierda) y baja densidad de observaciones (columna de la derecha).**

Los resultados obtenidos en los experimentos en donde se realiza la estimación de la fuente, se muestran en la Figura 8. Los dos paneles izquierdos presentan los resultados cuando se asimilan observaciones de concentración en todos los puntos de retícula. El panel superior muestra la evolución temporal del ECM para la concentración cuando se estima conjuntamente con la emisión. En este caso es posible ver que hay un deterioro del resultado cuando se utiliza alguna de las configuraciones desacopladas respecto del desempeño de la configuración acoplada. El panel inferior muestra las estimaciones del parámetro en las distintas configuraciones, mostrando también una mejor estimación, cuando la configuración es de tipo acoplada. Los paneles de la derecha, muestran los resultados cuando sólo se cuenta con siete obser-

vaciones de concentración equiespaciadas en el dominio. En este caso es posible observar que el mejor desempeño es obtenido por la configuración de tipo acoplada, siendo las diferencias entre los errores del caso acoplado y desacoplado, mucho mayores que cuando la disponibilidad de observaciones era mayor.

#### **4) CONCLUSIONES Y TRABAJO FUTURO**

En este trabajo se ha podido estudiar a través de un modelo simplificado, el impacto de una técnica de asimilación (LETKF) para estimar variables de viento, concentración y fuente utilizando una metodología de tipo acoplado y desacoplado. También es importante destacar que la técnica fue probada en distintos escenarios en los cuales se realizaron variaciones de los parámetros de asimilación como la escala de localización y el factor de inflación así como también en la densidad de las observaciones. En los experimentos y distintas configuraciones se puede ver que la técnica de asimilación tiene un buen rendimiento, sin embargo la configuración de tipo acoplada obtiene mejores resultados en términos del ECM que las otras dos configuraciones. Sin embargo la configuración de tipo desacoplada-ensamble que incluye una representación de la incertidumbre en el modelo de Lorenz, logra resultados similares a la configuración acoplada cuando se considera una cantidad mayor a 50% de observaciones de concentración. Aún así los resultados muestran que bajo este porcentaje de observación, el mejor desempeño se obtiene utilizando la configuración de tipo acoplada.

Cuando se analizan los resultados al estimar conjuntamente la concentración y la fuente, también se observa una clara diferencia en los resultados en términos de la cantidad de observaciones, donde la configuración acoplada obtiene mejores resultados sobre todo cuando la disponibilidad de observaciones es pequeña.

Algunos puntos que no se han estudiado en este trabajo son el efecto de la disminución de densidad de observaciones para las variables de Lorenz, tanto para estimar concentraciones como para estimar fuentes. En este trabajo se considero el mismo tamaño del ensamble para todos los experimentos, por lo que resta saber cuál es el impacto que este tendría en los resultados, utilizando las distintas configuraciones.

En el caso del estudio de estimación de fuentes es importante notar que por definición, la fuente es una variable positiva, por lo que la suposición de gaussianidad que necesita el filtro no se satisface lo que motiva a estudiar que tipo de técnicas se pueden aplicar en este caso. Finalmente un punto importante que no se ha tenido en cuenta al estudiar la estimación de fuentes con las distintas configuraciones, es la variación temporal y espacial que dicho parámetro podría tener y que para este estudio sólo se consideró constante a lo largo del tiempo.

#### **5) AGRADECIMIENTOS**

Este trabajo ha sido financiado con una beca doctoral del CONICET, categoría: DOC. C/PAISES LATINOAMERICANOS. También se agradecen los valiosos comentarios del revisor anónimo del CONGREMET.

## REFERENCIAS

- Anderson, J. L., Anderson, S. L., 1999:** A Montecarlo implementation of the nonlinear filtering problem to produce ensemble assimilation and forecasts. *Monthly Weather Review*, 127, 2741-2758.
- Bocquet, M., 2011:** Parameter-field estimation for atmospheric dispersion: application to the Chernobyl accident using 4D-Var. *Q. J. R. Meteorol. Soc*, 138, 664-681.
- Bocquet, M., Sakhov, P., 2013:** Joint state parameter estimation with iterative ensemble Kalman smoother. *Nonlinear Processes Geophys.*, 20, 803-818.
- Grell, G. A., Peckham, S. E., Schmitz, R., McKeen, S. A., Frost, G., Skamarock, W. C., Eder, B., 2005:** Fully coupled “online” chemistry within the WRF model, *Atmospheric Environment*, 39, 6957-6975.
- Hunt, B. R., Kostelich, E. J., Szunyogh, I., 2007:** Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter. *Physica D*, 230, 112–126.
- Kalman, R. E., 1960:** A new approach to linear filtering and prediction problems. *Journal Basic Eng*, 82, 35-45.
- Kang, J.-S., Kalnay, E., Liu, J., Fung, I., Miyoshi, T., Ide, K., 2011:** “Variable localization” in an ensemble Kalman filter: Application to the carbon cycle data assimilation, *J. Geophys. Res.*, 116.
- Khade, V. M., Hansen, J. A., Reid, J. S., Wetphal, D. L., 2012:** Ensemble filter based estimation of spatially distributed parameters in a mesoscale dust model: experiments with simulated and real data. *Atmos. Chem. Phys.*, 13, 3481-3500.
- Kukkonen, J. Olsson, T., Schultz, D. M., Baklanov, A., et al., 2012:** A review of operational, regional-scale, chemical weather forecasting model in Europe. *Atmos. Chem. Phys.*, 12, 1-87.
- Lorenz, E. 1996:** Predictability, a problem partly solved. *Seminar on predictability*, vol 1, ECMWF
- Saide, P., Bocquet, M., Osses, A., Gallardo, L., 2011:** Constraining surface emissions of air pollutants using inverse modeling: Method intercomparison and a new two-step multiscale approach. *Tellus, Ser. B*, 63, 360–370.
- Yang, S-Ch., Corazza, M., Carrassi, A., Kalnay, E., and Miyoshi, T., 2009:** Comparison of Local Ensemble Transform Kalman Filter, 3DVAR, and 4DVAR in a Quasigeostrophic Model. *Mon. Wea. Rev.*, 137, 693–709.