

# Diseño e Implementación de un Sistema de Asimilación de Datos Regulador para Modelos Cuasi-Geostróficos en Dos Dimensiones

Departamento de Ingeniería de Sistemas y Computación

Proyecto Final  
Informe Final

Sebastian Garrido Cepeda

Asesor:  
Elias David Niño Ruiz, PhD.

# Índice

<b>Abstract</b>	<b>2</b>
<b>Planteamiento y Justificación</b>	<b>3</b>
Contexto	3
Planteamiento del Problema	3
Justificación del Proyecto	3
Roles en el Proyecto	3
<b>Objetivos</b>	<b>5</b>
Objetivo General	5
Objetivos Específicos	5
<b>Alcances y Restricciones</b>	<b>6</b>
Alcance	6
Restricciones	6
<b>Metodología</b>	<b>7</b>
<b>Cronograma</b>	<b>8</b>
<b>Presupuesto</b>	<b>9</b>
<b>Requerimientos</b>	<b>10</b>
Requerimientos Funcionales	10
Requerimientos No Funcionales	10
<b>Antecedentes</b>	<b>11</b>
<b>Marco Conceptual</b>	<b>14</b>
<b>Modelo Propuesto</b>	<b>15</b>
<b>Esquema de la Solución</b>	<b>16</b>
<b>Diagrama de Componentes</b>	<b>17</b>
<b>Conceptualización de Otras Soluciones</b>	<b>18</b>
<b>Evaluación de la Solución</b>	<b>19</b>
<b>Análisis de Resultados</b>	<b>20</b>
<b>Conclusiones y Trabajos Futuros</b>	<b>21</b>
<b>Bibliografía</b>	<b>22</b>

# Abstract

# Planteamiento y Justificación

## Contexto

La asimilación de datos es un conjunto de técnicas que toman observaciones del mundo real y las inyectan a un modelo de predicción o simulación de un sistema físico. La finalidad de este procedimiento es combinar la información tanto del modelo como de las observaciones y, teniendo en cuentas sus diferentes grados de error, obtener una predicción óptima, es decir, que minimice la discrepancia con la realidad física.

## Planteamiento del Problema

En asimilación de datos, las observaciones son inyectadas a un sistema numérico imperfecto con el objetivo de ajustar su trayectoria con respecto a medidas reales del sistema dinámico de interés. Por ejemplo, podemos tener un modelo que prediga cómo cambia la temperatura de la ciudad. A medida que este modelo hace predicciones podemos medir la temperatura en la ciudad e inyectar esa información al modelo para corregir las predicciones que realiza.

Entre las múltiples estrategias de asimilación que existen en la literatura, los filtros de Kalman regularizados son una opción viable en la asimilación temporal de múltiples observaciones. Sin embargo, debido al costo computacional requerido para la realización (estadísticamente hablando) de una sola corrida del modelo, el número de muestras es mucho menor a la dimensión del modelo. Debido a esto, la calidad del estado resultante por medio del proceso de asimilación se ve afectada por el error de muestreo. La literatura propone mecanismos de localización para evitar tal efecto; sin embargo, algunos son computacionalmente costosos o simplemente imprácticos. Por tanto, es importante desarrollar esquemas de asimilación de datos que funcionen en un ambiente real con recursos computacionales limitados y un número pequeño (con respecto a las dimensiones del problema) de observaciones imperfectas.

## Justificación del Proyecto

El error de muestreo es típico en esquemas de asimilación de datos desde sus inicios. El costo computacional de una sola propagación del modelo es alto y por lo tanto, el número de muestras que pueden ser tomadas es restringida. Por tal motivo, es indispensable diseñar e implementar un sistema de asimilación de datos que permita asimilar observaciones minimizando el impacto del error de muestreo.

## Roles en el Proyecto

En el proyecto participan un Ph. D en Ciencias de la Computación que colaborará con la creación del esquema de asimilación de datos y un estudiante de Ingeniería de Sistemas que colaborará con la implementación de dicho esquema y con las pruebas sobre modelos

de predicción. El proyecto se realizará en colaboración con la Universidad del Norte que provee acceso a recursos computacionales y a base de datos para la revisión de la literatura científica.

# Objetivos

## Objetivo General

Diseñar e implementar un sistema de asimilación de datos basado en filtros de Kalman y muestreo para la asimilación de observaciones en modelos cuasi-geostróficos en dos dimensiones.

## Objetivos Específicos

- Diseñar un sistema de asimilación de datos basado en muestreo y filtros de Kalman para la síntesis de observaciones en dominios temporales y espaciales bidimensionales.
- Implementar el sistema de asimilación de datos propuesto en MATLAB.
- Parametrizar el modelo cuasi-geostrófico de acuerdo al sistema propuesto.
- Comparar los resultados obtenidos con los filtros de Kalman basado en localización espacial y temporal: local ensemble transform Kalman filter & iterative ensemble Kalman smoother.

# Alcances y Restricciones

## Alcance

- Debe realizarse un repaso de la literatura científica para definir los criterios y seleccionar las técnicas a utilizar.
- Debe escogerse una fuente de datos para generar los valores de las observaciones de prueba y valores iniciales del sistema físico.
- Debe implementarse un esquema de asimilación de datos para un modelo cuasi-geostrófico.
- Deben evaluarse los resultados de la efectividad del esquema de asimilación de datos.

## Restricciones

- El código se desarrollará en Matlab. Entonces, se hace necesaria una licencia de este software.
- Los datos se obtendrán de las siguientes fuentes: [Por definir]. La obtención de los datos entonces estará sujeta a la disponibilidad de estas plataformas.
- Por la naturaleza caótica de los sistemas que estamos prediciendo, las predicciones sólo son válidas (con un margen de error pequeño) en una ventana de tiempo de [Por definir].
- A pesar de que el método escogido tiene aplicaciones en varias áreas, solo probaremos su efectividad en el área específica de asimilación de datos para modelos cuasi-geostróficos.

# Metodología

En el artículo de Chavarriga [1] se realiza una propuesta para proyectos con fases de investigación y desarrollo. Con base en esto, y tomando lineamientos de la metodología XP [2], podemos definir cinco etapas principales como la metodología con la que vamos a trabajar:

1. **Seleccionar las herramientas para la recolección de datos.** Esto incluye investigar las APIs o bases de datos abiertas que nos ofrezcan acceso a datos reales que podamos usar en los modelos cuasi-geostróficos que harán parte de nuestros esquemas de asimilación de datos.
2. **Limpieza de datos.** En esta etapa tomamos los datos adquiridos y removemos la información innecesaria o redundante al mismo tiempo que hacemos normalización de datos que se encuentren en formatos que no nos sean de utilidad.
3. **Selección del modelo de asimilación de datos.** Existen diversos modelos de asimilación de datos donde cada uno presenta un conjunto de ventajas y desventajas diferentes. De acuerdo a nuestras restricciones y al alcance del proyecto determinaremos cuál de estos modelos es más probable que presente resultados positivos.
4. **Implementación.** Se realizará una implementación en el framework de desarrollo numérico y científico Matlab del modelo de asimilación de datos escogido. Para esto tomaremos un modelo cuasi-geostrófico existente que devuelve datos en formato NetCDF como base.
5. **Resultados y conclusiones.** Analizaremos el rendimiento obtenido de la ejecución del esquema de asimilación de datos utilizando un modelo cuasi-geostrófico y datos reales del comportamiento de una cuadrilla de un sistema físico.



# Cronograma

WBS	Nombre	Inicio	Fin	Duracion	Predecessores	Febrero				Marzo					Abril				Mayo				
						1	8	15	22	1	8	15	22	29	5	12	19	26	3	10	17	24	31
1	Planteamiento Problema, Objetivos	1/02/2018	4/02/2018	4d																			
2	Selección Herramienta	5/02/2018	7/02/2018	3d	1																		
3	Definición de Metodología	8/02/2018	14/02/2018	7d	2																		
4.1	Busqueda y recolección de datos	15/02/2018	28/02/2018	14d	3																		
4.2	Limpieza de los datos	1/03/2018	7/03/2018	7d	4.1																		
4.3	Definición de alcance, restricciones, supuestos y requerimientos	8/03/2018	14/03/2018	7d	4.2																		
4.4	Estado del arte	15/03/2018	21/03/2018	7d	4.3																		
4.5.1	Planteamiento y selección de técnicas para utilizar en el esquema de asimilación de datos	15/03/2018	21/03/2018	7d	4.4																		
4.5.2	Formalización del esquema de asimilación de datos	22/03/2018	28/03/2018	7d	4.4																		
4.6.1	Implementación de los modelos físicos de prueba	22/03/2018	28/03/2018	7d	4.4																		
4.6.2	Implementación de los sistemas de captura de las muestras	29/03/2018	4/04/2018	7d	4.4																		
4.6.3	Implementación del esquema de asimilación	5/04/2018	18/04/2018	14d	4.5.2																		
5.1	Diseño e Implementación de pruebas de Validación	19/04/2018	2/05/2018	14d	4.6																		
5.2	Validación de efectividad con las pruebas	3/05/2018	16/05/2018	14d	5.1																		
5.3	Documentación del modelo	17/05/2018	23/05/2018	7d	5.2																		
6	Sustentación	24/05/2018	30/05/2018	7d	5.3																		

# Presupuesto

Rubros	Valor		Financiación		Total
	Valor Und.	#	Universidad	Estudiante	
<b>Personal</b>					<b>\$19,000,000</b>
Ingeniero de Sistema	\$7,000,000	1		\$7,000,000	
Doctor en Ciencias de la Computación	\$12,000,000	1	\$12,000,000		
<b>Software</b>					<b>\$7,900,000</b>
API	\$0	1			
Licencia Matlab	\$7,900,000	1	\$7,900,000		
Licencia NetCDF	\$0	1			
Licencia Ubuntu Escritorio	\$0	1			
<b>Bases de Datos</b>					
Acceso a Artículo Científico	\$94,500	20	\$1,890,000		<b>\$1,890,000</b>
<b>Equipos</b>					<b>\$5,000,000</b>
Computador de Escritorio con Procesador Intel i7	\$5,000,000	1	\$5,000,000		
<b>Materiales</b>					<b>\$250,000</b>
Papelería	\$150,000	1		\$500,000	
Utensilios	\$100,000	1		\$150,000	
<b>Otros</b>					<b>\$450,000</b>
Imprevistos	\$450,000	1		\$450,000	
<b>Total</b>					<b>\$34,490,000</b>

# Requerimientos

## Requerimientos Funcionales

- El software debe realizar asimilación de observaciones físicas a modelos cuasi-geostróficos.
- La implementación debe reducir la incertidumbre en las predicciones de los modelos físicos entre el tiempo actual y la siguiente observación.
- El software debe realizar la asimilación de las observaciones en tiempo real, es decir, debe responder a los cambios en el conjunto de observaciones a medida que vayan pasando.

## Requerimientos No Funcionales

- El esquema de asimilación de datos debe operar correctamente dado cualquier modelo, estado inicial y conjunto de observaciones.
- El esquema de asimilación de datos debe funcionar de la manera más eficiente posible, permitiendo su utilización en ambientes de producción donde pueden presentarse condiciones subóptimas de cómputo.
- El software debe estar propiamente documentado con el fin de que sea fácil de utilizar para un público técnico con experiencia en el modelamiento de procesos físicos.

# Antecedentes

Para entender el movimiento de fluidos donde entra en juego el efecto Coriolis, como en las corrientes marítimas y vientos se crearon los modelos geostróficos [3]. Estos modelos asumen que el efecto Coriolis es balanceado por un diferencial de presión en el fluido, produciendo movimientos en el agua a medida que esta se traslada de áreas con alta presión a áreas con baja presión. Es tanto el éxito de estos modelos al realizar predicciones en espacios relativamente cortos de tiempo que Wust en 1924 [4] demostró que las predicciones de este modelo no se alejaban de realidad por mucho al aplicarlo a corrientes oceánicas reales y aun este mismo ejemplo es utilizado en libros académicos sobre oceanografía.

Dado que este modelo es una simplificación de la realidad, es posible realizarle extensiones que lo hagan más fiable en la práctica. Es por esto que Charney (1949) [5] introduce el modelo cuasi-geostrófico el cual toma en cuenta la inercia como un factor en los cálculos para realizar las predicciones. Junto con este cambio Charney introduce un método para el cálculo numérico de estos fenómenos, dando inicio al uso de los métodos numéricos en la oceanografía y las predicciones de corrientes de viento.

A pesar de la utilidad de estos modelos ellos ignoran varios fenómenos físicos al momento de realizar los cálculos. Estas discrepancias con la realidad producen incertidumbre en las predicciones y a medida que estas se proyectan en el tiempo, la incertidumbre crece, produciendo un error considerable. Para aliviar esta limitación realizamos mediciones periódicas del estado actual del sistema físico. Lastimosamente estas mediciones cuentan con varias restricciones. Primero, son costosas de realizar ya que requieren equipos especializados que recogen una gran cantidad de datos. Segundo, estos innumerables datos luego deben ser transportados a los centros donde van a ser realizados para las operaciones computacionales. Tercero, el mantenimiento y reparación de estos dispositivos medidores hace que sea prohibitivamente costoso mantener una cantidad grande de ellos. Tercero, muchas veces la observación no posee toda la información que maneja el modelo de predicción, forzando la estimación de datos faltantes.

Por esto es ideal poseer técnicas que a partir de una cantidad limitada de observaciones, limitadas tanto en la frecuencia con la cual se pueden realizar como en la extensión del espacio observado, te permitan realizar predicciones con un grado de incertidumbre pequeño. Históricamente el método usado por meteorólogos para ajustar las predicciones a medida que los errores se propagaban en el tiempo era con el uso del “análisis subjetivo”, donde, usando su conocimiento del comportamiento de los fenómenos meteorológicos ajustaban los modelos “al ojo”. El primer intento para realizar un “análisis objetivo” consto de interpolación simple. Solo fue en 1963 cuando L. Gandin [6] introdujo por primera vez la “interpolación estadística”. Esta técnica es una forma reducida del Filtro de Kalman, método que luego sería el más popular en el campo de las NWP (Numerical Weather Predictions).

Los Filtros de Kalman fueron introducidos y desarrollados parcialmente por Swerling en 1958 y Kalman en 1960, pero su formalización se dio en 1961 con el trabajo de Kalman y Bucy [7]. Esta técnica fue recibida con escepticismo en un principio, pero luego de su aplicación exitosa en los sistemas de trayectoria de las misiones Apollo de la NASA se popularizó e implementó en diversos campos.

El Filtro de Kalman es, como su nombre lo indica, un filtro estadístico. Este filtro funciona encontrando la corrección al estado del sistema que maximice la probabilidad a posteriori de esta corrección dada una observación. Podemos demostrar que bajo la hipótesis de linealidad del operador que evoluciona el estado del sistema, linealidad del operador que transforma la predicción del sistema al espacio de observaciones y bajo hipótesis de errores gaussianos las ecuaciones que rigen esta corrección son:

$$K = P^b H^T (R + H P^b H^T)$$

$$x^a = x^b + K y$$

$$P^a = (I - K H) P^b (I - K H)^T + K R K^T$$

Con base en la optimización de la siguiente función de costo:

$$J(x) = \|x - x^b\|_{B^{-1}}^2 + \|y - Hx\|_{R^{-1}}^2$$

Una de las limitaciones iniciales de esta técnica era la gran cantidad de ecuaciones que debían resolverse en el proceso. Esto hacía que la técnica fuese impráctica para sistemas con un gran número de variables. Es por esto que se desarrollaron nuevas técnicas con costos computacionales menores pero con soluciones aproximadas o subóptimas, entre ellas, EnKF (Ensemble Kalman Filter) y 4DVar [8].

Para el caso de EnKF, en vez de evolucionar la matriz de covarianzas de los errores, esta se estima a partir de una muestra de observaciones. Esto es debido al alto costo computacional que implican las ecuaciones que actualizan dicha matriz. Las ecuaciones para actualizar la matriz de covarianzas asociada con esta técnica son las siguientes:

$$P^b = \frac{1}{N-1} \sum_{i=1}^N (x_i^b - \bar{x}^b)(x_i^b - \bar{x}^b)^T$$

En el caso de 4DVar, esta técnica busca optimizar con base a una función de costo diferente a la del Filtro de Kalman estándar. En esta función de costo tomamos en cuenta no solo el error en la última observación, sino en todas las observaciones previas. Esta

técnica nos permite encontrar trayectorias para la evolución del sistema que a la larga tienden a divergir menos de la realidad. La función de costo asociada es:

$$J(x) = \|x - x^b\|_{B^{-1}}^2 + \sum_t \|y - Hx_t\|_{R^{-1}}^2$$

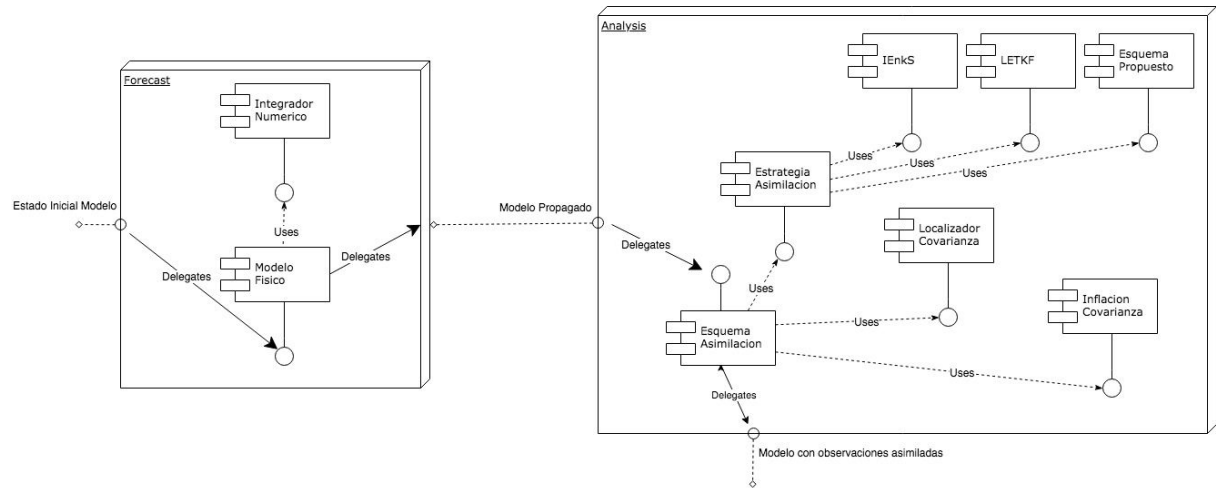
## Marco Conceptual

## Modelo Propuesto



## Esquema de la Solución

# Diagrama de Componentes



## Conceptualización de Otras Soluciones

## Evaluación de la Solución

## Análisis de Resultados

## Conclusiones y Trabajos Futuros

# Bibliografía

- [1] J. A. Chavarriaga, L. Hugo, and F. Arboleda, "Modelo de Investigación en Ingeniería del Software: Una propuesta de investigación tecnológica."
- [2] Newkirk, J. and Martin, R. (2002). La programación extrema en la práctica. [Madrid]: Addison Wesley.
- [3] R. Stewart. "Introduction to physical oceanography". [College Station, Tex.]: [Texas A & M University], 2008, pp. 153-154.
- [4] Wüst, Georg. "Florida-und Antillenstrom: eine hydrodynamische Untersuchung". ES Mittler, 1924.
- [5] J. G. Charney. "On a Physical Basis for Numerical Prediction of Large-Scale Motions in the Atmosphere", Journal of Meteorology, Vol. 6, No. 6, December 1949.
- [6] L. Gandin. "Objective Analysis of Meteorological Fields", 1963.
- [7] R. Kalman and R. Bucy, "New Results in Linear Filtering and Prediction Theory", Journal of Basic Engineering, vol. 83, no. 1, p. 95, 1961.
- [8] R. Bannister, "A review of operational methods of variational and ensemble-variational data assimilation", Quarterly Journal of the Royal Meteorological Society, vol. 143, no. 703, pp. 607-633, 2017, January 2017.