

UNIVERSIDAD DEL NORTE

TESIS DE MAESTRÍA

---

**Métodos de asimilación de datos  
secuenciales en el modelo Speedy con  
datos reales**

---

*Autor:*

Omar Angel MEJIA SUAREZ

*Asesor:*

Elias D. NIÑO-RUIZ, Ph.D.

*Una tesis presentada en cumplimiento de los requisitos  
por el grado de Maestría en Ingeniería de Sistemas y Computación*

*en el*

Departamento de Ingeniería de Sistemas y computación

9 de febrero de 2023



## Declaración de Autoría

Yo, Omar Angel MEJIA SUAREZ, Declaro que esta tesis titulada, «Métodos de asimilación de datos secuenciales en el modelo Speedy con datos reales» y el trabajo presentado en él son míos. Y confirmo que:

- Este trabajo fue realizado en su totalidad o en su mayor parte durante mi candidatura a un grado de investigación en esta Universidad.
- Cuando alguna parte de esta tesis se haya presentado previamente para obtener un título o cualquier otra calificación en esta Universidad o cualquier otra institución, esto se ha indicado claramente.
- Cuando he consultado el trabajo publicado de otros, esto siempre se atribuye claramente.
- Donde he citado el trabajo de otros, siempre se da la fuente. Excepto por tales citas, esta tesis es enteramente mi propio trabajo.
- He reconocido todas las principales fuentes de ayuda.
- Cuando la tesis se basa en un trabajo realizado por mí mismo junto con otros, he dejado claro exactamente qué fue hecho por otros y qué aporté yo mismo.

Firma:

---

Fecha:

---



*«Imagina un punto al que quieres ir y haz un plan para llegar allí. Pero sé honesto contigo mismo y establece tu punto de partida.»*

Jordan Belfort



UNIVERSIDAD DEL NORTE

## *Resumen*

División de Ingenierías  
Departamento de Ingeniería de Sistemas  
y  
Ciencias de la computación

Maestría en Ingeniería de Sistemas y Computación

**Métodos de asimilación de datos secuenciales en el modelo Speedy con datos reales**

por Omar Angel MEJIA SUAREZ

El proceso de asimilación de datos (DA) ha cobrado importancia en los últimos años debido a la mejora que ofrece en modelos numéricos complejos. Aun así, la mayoría de los fenómenos naturales tienen muchas correlaciones entre variables que son muy difíciles de capturar con tales modelos. Esta propuesta presenta una comparación entre modelos: el Ensemble Kalman Filter basado en una Descomposición de Cholesky Modificada, el Local Ensemble Kalman Filter y local ensemble transform Kalman filter. Tres formulaciones para el Ensemble Kalman Filter usando y sin usar el salto de paso para un Modelo de Circulación Atmosférica General, esto usando condiciones reales.





## *Agradecimientos*

En primer lugar, quiero agradecer a Dios por darme la sabiduría para lograr todo lo que he logrado. A mis amados padres, cuyo apoyo y confianza me inspiraron a lograrlo, a mis hermanas por su apoyo constante, y a mi sobrino quien es mi fuerza impulsora cada día. Estoy profundamente agradecido al asesor Ph.D. Elias Niño, por la confianza depositada en mí y en mis capacidades, por toda la enseñanza, paciencia y oportunidades brindadas.



# Índice general

<b>Declaración de Autoría</b>	<b>III</b>
<b>Resumen</b>	<b>VII</b>
<b>Agradecimientos</b>	<b>IX</b>
<b>1. Introducción</b>	<b>1</b>
<b>2. Marco Teórico</b>	<b>3</b>
2.1. Asimilación de datos . . . . .	3
2.1.1. Asimilación Variacional de Datos . . . . .	5
2.1.2. Asimilación Secuencial de Datos . . . . .	6
2.1.3. Ensemble Kalman Filter . . . . .	8
Ensemble Kalman Filter basado en la descomposicion de Cho-	
lesky Modificada . . . . .	9
Local Ensemble Kalman Filter . . . . .	10
Local Ensemble transform Kalman Filter . . . . .	13
2.2. Numerical Model SPEEDY . . . . .	14
<b>3. Objetivos</b>	<b>17</b>
3.1. Planteamiento del problema . . . . .	17
3.2. Contribuciones Esperadas . . . . .	17
3.3. Objetivos . . . . .	17
3.3.1. Objetivo General . . . . .	17
3.3.2. Objetivos Específicos . . . . .	18
3.4. Metodología . . . . .	18
<b>4. Implementación</b>	<b>19</b>
4.1. Diseñar e implementar herramienta de captura y redimensionado de	
datos reales para modelo SPEEDY . . . . .	19
4.1.1. Captura de datos reales . . . . .	19
4.1.2. Interpolación . . . . .	21
Redimensionado . . . . .	21
Hallar hora 01 . . . . .	21
4.2. Probar diferentes implementaciones de asimilación secuencial de da-	
tos con datos reales redimensionados a la resolución de una malla. . .	21
4.2.1. Ejecución de SPEEDY . . . . .	22
Preparación del modelo . . . . .	22
Ejecución del modelo con Asimilación de datos . . . . .	23
<b>5. Resultados</b>	<b>25</b>
5.1. Ensemble Kalman Filter basado en la descomposición de Cholesky . .	29
5.1.1. EnKF $p = 100\%$ $alpha = 1,02$ . . . . .	29
5.1.2. EnKF $p = 100\%$ $alpha = 1,04$ . . . . .	30

5.1.3.	EnKF $p = 100\%$ $\alpha = 1,06$	32
5.1.4.	EnKF $p = 100\%$ $\alpha = 1,08$	33
5.1.5.	EnKF $p = 25\%$ $\alpha = 1,02$	35
5.1.6.	EnKF $p = 25\%$ $\alpha = 1,04$	36
5.1.7.	EnKF $p = 25\%$ $\alpha = 1,06$	38
5.1.8.	EnKF $p = 25\%$ $\alpha = 1,08$	39
5.1.9.	EnKF $p = 11\%$ $\alpha = 1,02$	41
5.1.10.	EnKF $p = 11\%$ $\alpha = 1,04$	42
5.1.11.	EnKF $p = 11\%$ $\alpha = 1,06$	44
5.1.12.	EnKF $p = 11\%$ $\alpha = 1,08$	45
5.1.13.	EnKF $p = 6\%$ $\alpha = 1,02$	47
5.1.14.	EnKF $p = 6\%$ $\alpha = 1,04$	48
5.1.15.	EnKF $p = 6\%$ $\alpha = 1,06$	50
5.1.16.	EnKF $p = 6\%$ $\alpha = 1,08$	51
5.1.17.	EnKF $p = 4\%$ $\alpha = 1,02$	53
5.1.18.	EnKF $p = 4\%$ $\alpha = 1,04$	54
5.1.19.	EnKF $p = 4\%$ $\alpha = 1,06$	56
5.1.20.	EnKF $p = 4\%$ $\alpha = 1,08$	57
5.2.	Local Ensemble Transform Kalman Filter	59
5.2.1.	LETKF $p = 100\%$ $\alpha = 1,02$	59
5.2.2.	LETKF $p = 100\%$ $\alpha = 1,04$	60
5.2.3.	LETKF $p = 100\%$ $\alpha = 1,06$	62
5.2.4.	LETKF $p = 100\%$ $\alpha = 1,08$	63
5.2.5.	LETKF $p = 25\%$ $\alpha = 1,02$	65
5.2.6.	LETKF $p = 25\%$ $\alpha = 1,04$	66
5.2.7.	LETKF $p = 25\%$ $\alpha = 1,06$	68
5.2.8.	LETKF $p = 25\%$ $\alpha = 1,08$	69
5.2.9.	LETKF $p = 11\%$ $\alpha = 1,02$	71
5.2.10.	LETKF $p = 11\%$ $\alpha = 1,04$	72
5.2.11.	LETKF $p = 11\%$ $\alpha = 1,06$	74
5.2.12.	LETKF $p = 11\%$ $\alpha = 1,08$	75
5.2.13.	LETKF $p = 6\%$ $\alpha = 1,02$	77
5.2.14.	LETKF $p = 6\%$ $\alpha = 1,04$	78
5.2.15.	LETKF $p = 6\%$ $\alpha = 1,06$	80
5.2.16.	LETKF $p = 6\%$ $\alpha = 1,08$	81
5.2.17.	LETKF $p = 4\%$ $\alpha = 1,02$	83
5.2.18.	LETKF $p = 4\%$ $\alpha = 1,04$	84
5.2.19.	LETKF $p = 4\%$ $\alpha = 1,06$	86
5.2.20.	LETKF $p = 4\%$ $\alpha = 1,08$	87
5.3.	Local Ensemble Kalman Filter	89
5.3.1.	LEnKF $p = 100\%$ $\alpha = 1,02$	89
5.3.2.	LEnKF $p = 100\%$ $\alpha = 1,04$	90
5.3.3.	LEnKF $p = 100\%$ $\alpha = 1,06$	92
5.3.4.	LEnKF $p = 100\%$ $\alpha = 1,08$	93
5.3.5.	LEnKF $p = 25\%$ $\alpha = 1,02$	95
5.3.6.	LEnKF $p = 25\%$ $\alpha = 1,04$	96
5.3.7.	LEnKF $p = 25\%$ $\alpha = 1,06$	98
5.3.8.	LEnKF $p = 25\%$ $\alpha = 1,08$	99
5.3.9.	LEnKF $p = 11\%$ $\alpha = 1,02$	101
5.3.10.	LEnKF $p = 11\%$ $\alpha = 1,04$	102
5.3.11.	LEnKF $p = 11\%$ $\alpha = 1,06$	104

5.3.12. LEnKF $p = 11\%$ $\alpha = 1,08$ . . . . .	105
5.3.13. LEnKF $p = 6\%$ $\alpha = 1,02$ . . . . .	107
5.3.14. LEnKF $p = 6\%$ $\alpha = 1,04$ . . . . .	108
5.3.15. LEnKF $p = 6\%$ $\alpha = 1,06$ . . . . .	110
5.3.16. LEnKF $p = 6\%$ $\alpha = 1,08$ . . . . .	111
5.3.17. LEnKF $p = 4\%$ $\alpha = 1,02$ . . . . .	113
5.3.18. LEnKF $p = 4\%$ $\alpha = 1,04$ . . . . .	114
5.3.19. LEnKF $p = 4\%$ $\alpha = 1,06$ . . . . .	116
5.3.20. LEnKF $p = 4\%$ $\alpha = 1,08$ . . . . .	117
<b>6. Conclusions</b>	<b>119</b>
<b>Bibliografía</b>	<b>121</b>



# Índice de figuras

2.1. Premisas iniciales de Asimilación . . . . .	3
2.2. Inferencia Bayesiana . . . . .	4
2.3. Modelo de predictivo basado en las observaciones . . . . .	5
2.4. Asimilación secuencial de Datos . . . . .	7
2.5. Etiquetado . . . . .	11
2.6. Localización . . . . .	12
2.7. Dominios locales para diferentes radios de influencia $r$ . . . . .	13
2.8. Componentes del modelo para algunas resoluciones espectrales . . . . .	15
4.1. AMLCS-main . . . . .	20
4.2. Paquete data . . . . .	20
4.3. Paquete AMLCS . . . . .	22





# Índice de cuadros

2.1. Variables físicas del modelo AT-GCM Speedy . . . . .	14
5.1. Parámetros utilizados en los experimentos . . . . .	25
5.2. Nivel Superficie Temperatura . . . . .	26
5.3. Nivel Superficie componente V del viento . . . . .	27
5.4. Nivel Superficie componente U del viento . . . . .	28



# Lista de Abreviaciones

<b>EnKF</b>	<b>Ensemble Kalman Filter</b>
<b>LEnKF</b>	<b>Local Ensemble Kalman Filter</b>
<b>LETKF</b>	<b>Local Ensemble Transform Kalman Filter</b>
<b>SPEEDY</b>	<b>Simplified Parameterizations, primitive-Equation Dynamics</b>
<b>NOAA</b>	<b>National Oceanic and Atmospheric Administration</b>
<b>NCAR</b>	<b>National Center for Atmospheric Research</b>



# Lista de Simbolos

$\delta$	Error del modelo
$M$	Numero de observaciones en una ventana de tiempo
$\epsilon$	Error de las observaciones
$\mathcal{L}(\cdot, \cdot)$	Vecindad
$\mathbf{B}$	Matriz de covarianza de fondo
$\mathbf{R}$	Matriz de covarianza de las observaciones
$\mathbf{A}$	Matriz de covarianza del análisis
$\mathcal{H}$	Operador Observacional
$\mathbf{H}$	Operador observacional linealizado
$\langle \cdot \rangle$	Esperado



*A mis compañeros, profesores y todos aquellos que me  
apoyaron, gracias...*





## Capítulo 1

# Introducción

La Predicción Numérica del Clima es de gran relevancia en la actualidad, ya que nos permite modelizar la dinámica y evolución del clima global y local en una ventana de tiempo determinada, lo cual es fundamental en la prevención de desastres naturales, así como en la planificación agronómica, entre otras áreas de interés. Para realizar dichas previsiones, podemos utilizar Modelos de Circulación Global Atmosférica (AT-GCM) que son una discretización de un sistema de ecuaciones dinámicas que describe la física y dinámica atmosférica. (Hurrell, Deser y Phillips, 2019; Teixeira et al., 2014; Bauer, Thorpe y Brunet, 2015; Lorenc, 1986). Pero incluso los mejores modelos tienen incertidumbre asociada a sus estimaciones (ruido en la fuente de la observación, condiciones iniciales incorrectas, representación computacional, conocimiento incompleto sobre las leyes que rigen la dinámica del sistema, etc.), ya que es imposible capturar toda la variabilidad de los fenómenos de la vida real, es decir, cierto nivel de error está asociado al modelo y a la compleja dinámica de los fenómenos. (Liu et al., 2019; Verstraete, Aghezzaf y Desmet, 2020). Para mejorar la calidad de esas previsiones, podemos emplear técnicas de asimilación de datos que nos permitan gestionar esa incertidumbre y producir mejores estimaciones del estado futuro, dada la dinámica del sistema (Ott et al., 2004a; Saetrom y Omre, 2013; Kwiatkowski y Mandel, 2015; Wikle y Berliner, 2007) y las observaciones.

La estructura de esta propuesta es la siguiente, en la sección 2 se discuten temas relacionados con la Asimilación de Datos en métodos basados en conjuntos y su formulación, así como una breve descripción del modelo SPEEDY, en la sección 3 se detallan los resultados esperados de esta investigación y se establece el objetivo de la misma, En la sección 4 se presenta el método propuesto, el funcionamiento del paquete de descarga de datos que inyecta observaciones reales al modelo AT-GCM, finalmente en la sección 5 se presentan experimentos numéricos utilizando el Modelo de Circulación General Atmosférica SPEEDY; se comparan los resultados de la implementación propuesta frente a los obtenidos por las formulaciones de la literatura. Finalmente, en la sección 6 se presentan las conclusiones.



## Capítulo 2

# Marco Teórico

### 2.1. Asimilación de datos

En esta sección, presentamos una formulación del problema de asimilación de datos, en un marco bayesiano.

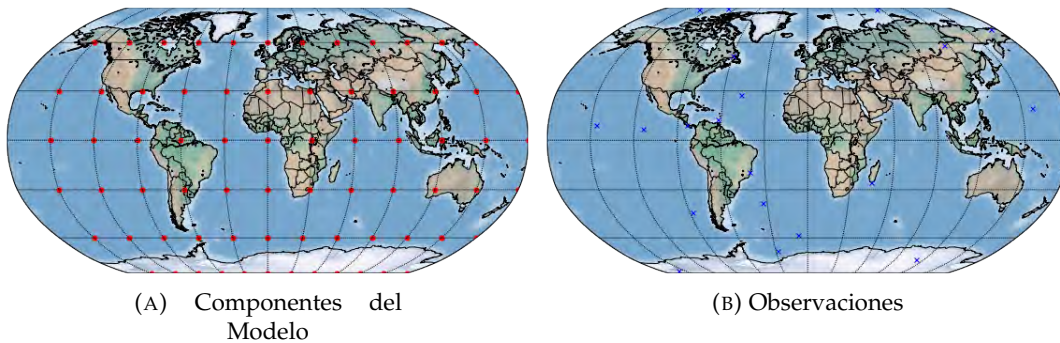


FIGURA 2.1: Premisas iniciales de Asimilación

Dado un modelo matemático o físico  $\mathcal{M}(\mathbf{x}_k)$  donde  $\mathbf{x}_{k+1} = \mathcal{M}(\mathbf{x}_k)_{t_k \rightarrow t_{k+1}}$ , donde  $\mathcal{M}$  es una discretización no lineal de los fenómenos reales (2.1a) que puede evolucionar dinámicamente en el tiempo, podemos prever el estado futuro del sistema ( $\mathbf{x}_{k+1}$ , también conocido como realización de modelos).

Estas previsiones se ven afectadas por su naturaleza caótica y estocástica, por lo que, incluso con unas condiciones iniciales perfectas, el sistema puede divergir de la realidad en unos pocos pasos de ejecución libre por numerosas razones, como ya se ha mencionado.

Además, tenemos algunas observaciones con ruido del estado del sistema, notadas como  $\mathbf{y}_k$ , normalmente desde satélites o sensores. Teniendo en cuenta que no todos los componentes observados coinciden con los componentes del modelo (como en 2.1b).

Con ese modelo y las observaciones, podemos obtener una mejor estimación del estado futuro mezclando esas dos fuentes en una nueva. Este proceso también se conoce como inferencia. Formalmente, tenemos:

$$\mathbf{x}_{k+1} = \mathcal{M}(\mathbf{x}_k)_{t_k \rightarrow t_{k+1}} + \delta \quad (2.1)$$

Donde  $\delta$  es el error del modelo computacional y se supone insesgado ( $\langle \delta \rangle = 0$ ). A esta realización del modelo la llamamos *background*, que será utilizado como *Prior* para estimar el estado futuro del sistema. Las observaciones pueden representarse mediante  $\mathbf{y} = \mathcal{H}(\mathbf{x}) + \epsilon_o$  donde  $\epsilon_o$  es un término estocástico (error de observación) y el operador observacional  $\mathcal{H}(\mathbf{x}) : \mathbb{R}^{n \times 1} \rightarrow \mathbb{R}^{m \times 1}$  que mapea desde el espacio de los

estados al espacio de las observaciones. Las observaciones se asimilan al pronóstico del modelo (*background*) por inferencia, utilizando el teorema de Bayes (Chen, 2003; Vetra-Carvalho et al., 2018):

$$\mathcal{P}(\mathbf{x}|\mathbf{y}) = \frac{\mathcal{P}(\mathbf{x}) \cdot \mathcal{P}(\mathbf{y}|\mathbf{x})}{\mathcal{P}(\mathbf{y})} \quad (2.2)$$

Donde:

- $\mathcal{P}(\mathbf{x})$  conocimiento previo sobre el estado del sistema (*Prior*).
- $\mathcal{P}(\mathbf{y}|\mathbf{x}) = \mathcal{L}(\mathbf{x}|\mathbf{y})$  cuantifica la distribución de los errores de las observaciones.
- $\mathcal{P}(\mathbf{y}) = \int \mathcal{P}(\mathbf{y}|\mathbf{x}) \cdot \mathcal{P}(\mathbf{x})d\mathbf{x}$  es una constante normalizadora.
- $\mathcal{P}(\mathbf{x}|\mathbf{y})$  da la estimación actualizada del estado verdadero, también llamada *Posterior*

Suponiendo que ambos se distribuyen normalmente. Esto nos da la mejor estimación (estado óptimo)  $\mathbf{x}^*$ , basado en las observaciones  $\mathbf{y}_k$  y la realización del modelo  $\mathbf{x}_k^b$

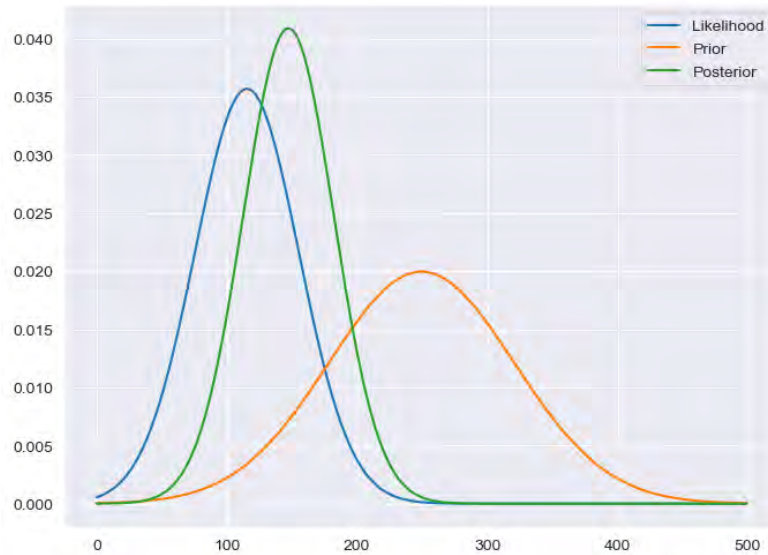


FIGURA 2.2: Inferencia Bayesiana

Esto se conoce como asimilación de datos: el proceso por el cual una previsión numérica imperfecta  $\mathbf{x}_k^b$  se ajusta en función de las observaciones ruidosas reales  $\mathbf{y}_k$  (Nino-Ruiz, 2018; Nino-Ruiz, Cheng y Beltran, 2018), donde  $\mathbf{x}_k^b \in \mathbb{R}^{n \times 1}$  y  $\mathbf{y}_k \in \mathbb{R}^{m \times 1}$  son el estado de fondo (*background*) y las observaciones en el paso  $k$ , para  $0 \leq k \leq M$ , donde,  $n$  es el tamaño del modelo o resolución del modelo,  $m$  denota el número de observaciones por paso de asimilación y  $M$  es el tamaño de la ventana de asimilación (el número de veces en que se dispone de observaciones).

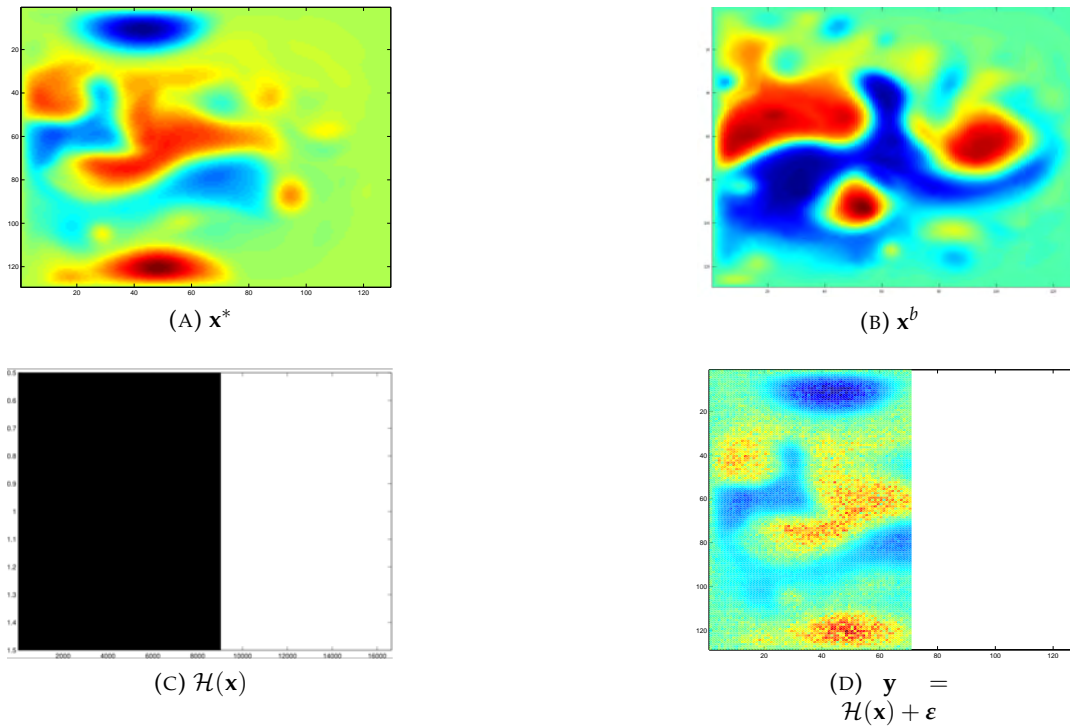


FIGURA 2.3: Modelo de predictivo basado en las observaciones

Mediante esta técnica, y el supuesto de gaussianidad, se obtiene una fórmula cercana para minimizar  $\mathcal{P}(\mathbf{x}|\mathbf{y})$  i.e, minimizar la diferencia entre las observaciones y las predicciones, ya se conoce. Llamamos a esta estimación *Analysis*. En este contexto, resolvemos el problema inverso, en el que utilizamos las observaciones para estimar la mejor condición inicial, que minimiza el error. Recordemos la ecuación 2.2 se utiliza para cuantificar la incertidumbre en nuestro modelo (Berger y Smith, 2019) mediante la actualización de los modelos. Al realizar estas actualizaciones, esperamos realizar pequeñas actualizaciones en las previsiones del modelo. Si las actualizaciones difieren demasiado de la previsión, podemos sospechar que el modelo no está funcionando como se esperaba. Podemos realizar esas actualizaciones utilizando dos enfoques:

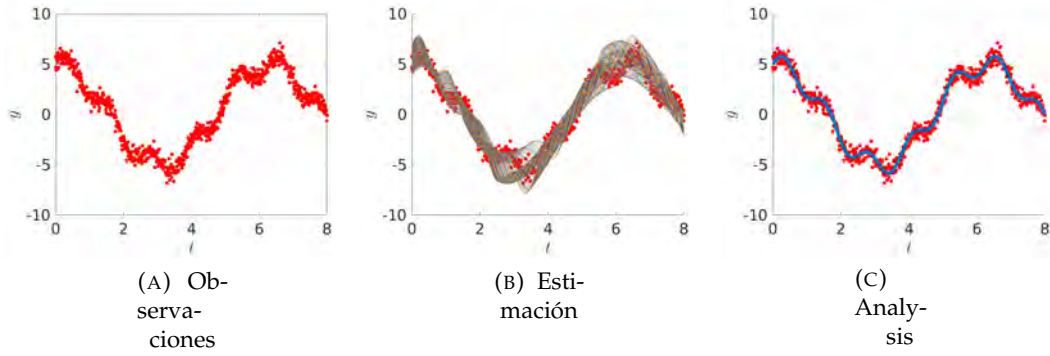
### 2.1.1. Asimilación Variacional de Datos

Dadas las múltiples observaciones, podemos asimilarlas en un solo paso de asimilación. El resultado posterior viene dado por

$$\mathcal{P}(\mathbf{x}_0 | \{\mathbf{y}_k\}_{k=0}^M) \propto \mathcal{P}(\mathbf{x}_0) \cdot \prod_{k=0}^M \mathcal{L}(\mathbf{x}_k | \mathbf{y}_k) \quad (2.3)$$

Y el proceso de optimización se realiza en una sola toma.

$$\mathbf{x}_0^a = \arg \max_{\mathbf{x}_0} \mathcal{P}(\mathbf{x}_0 | \{\mathbf{y}_k\}_{k=0}^M) \quad (2.4)$$



Teniendo  $\mathcal{J}(\mathbf{x})$  como (Lorenç, 2003b; Lorenç, 2003a):

$$\mathcal{J}(\mathbf{x}_0) = \left\| \mathbf{x}_0 - \mathbf{x}_0^b \right\|_{\mathbf{B}_0^{-1}}^2 + \sum_{k=0}^M \left\| \mathbf{y}_k - \mathcal{H}_k(\mathbf{x}_k) \right\|_{\mathbf{R}_k^{-1}}^2, \quad (2.5)$$

Para resolver este problema de optimización, se requiere el uso de Adjointos que son computacionalmente caros de calcular y complejos, más aún, de validar. Por lo cual se usa la asimilación de datos secuencial.

### 2.1.2. Asimilación Secuencial de Datos

Teniendo la ecuación 2.2, se puede decir que:

$$\mathcal{P}(\mathbf{x}|\mathbf{y}) \propto \mathcal{P}(\mathbf{x}) \cdot \mathcal{P}(\mathbf{y}|\mathbf{x}) \quad (2.6)$$

Esto significa que las actualizaciones del estado del modelo se realizan a medida que las observaciones están disponibles, y luego, se propaga la estimación. Recordemos la suposición gaussiana que hicimos, el **Analysis**  $\mathbf{x}_k^a$  se calcula resolviendo el siguiente problema de optimización

$$\mathbf{x}^a = \arg \max_{\mathbf{x}} \mathcal{P}(\mathbf{x}|\mathbf{y}). \quad (2.7)$$

Con:

$$\mathbf{x} \sim \mathcal{N}(\mathbf{x}^b, \mathbf{B}), \quad (2.8)$$

$$\mathbf{y} \sim \mathcal{N}(\mathcal{H}(\mathbf{x}), \mathbf{R}) \quad (2.9)$$

y errores:

$$\delta \sim \mathcal{N}(\mathbf{0}, \mathbf{B}), \quad (2.10)$$

$$\epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{R}) \quad (2.11)$$

Donde  $\mathbf{B} \in \mathbb{R}^{n \times n}$  es la matriz de covarianza del error de fondo, y  $\mathbf{R} \in \mathbb{R}^{m \times m}$  es la matriz de covarianza del error de las observaciones. Poniendo todo esto junto, nos queda:

$$\begin{aligned} \mathcal{P}(\mathbf{x}) &= \frac{1}{(2\pi)^{n/2} \cdot |\mathbf{B}|^{(1/2)}} \cdot \exp\left(-\frac{1}{2} \cdot \|\mathbf{x} - \mathbf{x}^b\|_{\mathbf{B}^{-1}}^2\right) \\ &\sim \exp\left((\mathbf{x} - \mathbf{x}^b)^T \cdot \mathbf{B}^{-1} \cdot (\mathbf{x} - \mathbf{x}^b)\right) \end{aligned} \quad (2.12)$$

$$\begin{aligned} \mathcal{P}(\mathbf{y}|\mathbf{x}) &= \frac{1}{(2\pi)^{m/2} \cdot |\mathbf{R}|^{(1/2)}} \cdot \exp\left(-\frac{1}{2} \cdot \|\mathbf{y} - \mathcal{H}(\mathbf{x})\|_{\mathbf{R}^{-1}}^2\right) \\ &\sim \exp\left((\mathbf{y} - \mathcal{H}(\mathbf{x}))^T \cdot \mathbf{R}^{-1} \cdot (\mathbf{y} - \mathcal{H}(\mathbf{x}))\right) \end{aligned} \quad (2.13)$$

Usando las ecuaciones 2.12 y 2.13 podríamos reescribir la ecuación 2.6 como  $\mathcal{P}(\mathbf{x}|\mathbf{y}) \sim \exp(-\mathcal{J}(\mathbf{x}))$  donde  $\mathcal{J}(\mathbf{x})$  es:

$$\mathcal{J}(\mathbf{x}) = \frac{1}{2} \cdot \|\mathbf{x} - \mathbf{x}^b\|_{\mathbf{B}^{-1}}^2 + \frac{1}{2} \cdot \|\mathbf{y} - \mathcal{H}(\mathbf{x})\|_{\mathbf{R}^{-1}}^2 \quad (2.14)$$

Convirtiendo el problema de optimización de la ecuación 2.7 en:

$$\mathbf{x}^a = \underset{\mathbf{x}}{\operatorname{argmax}} \mathcal{P}(\mathbf{x}|\mathbf{y}) = \underset{\mathbf{x}}{\operatorname{argmax}} \mathcal{J}(\mathbf{x}) \quad (2.15)$$

Una vez resuelto este problema, la moda posterior de la distribución de errores puede calcularse como:

$$\mathbf{x}^a = \mathbf{x}^b + \mathbf{A} \cdot \mathbf{H}^T \cdot \mathbf{R}^{-1} \cdot [\mathbf{y} - \mathbf{H} \cdot \mathbf{x}^b] \in \mathbb{R}^{n \times 1} \quad (2.16)$$

En el cual  $\mathcal{H}'(\mathbf{x}) \approx \mathbf{H}^T \in \mathbb{R}^{n \times m}$  es un operador lineal observacional (with  $\mathcal{H}(\mathbf{x}) \approx \mathcal{H}(\bar{\mathbf{x}}^b) + \mathbf{H} \cdot [\mathbf{x} - \bar{\mathbf{x}}^b]$ ). La matriz de covarianza del análisis  $\mathbf{A}$  queda como,

$$\mathbf{A} = [\mathbf{B}^{-1} + \mathbf{H}^T \cdot \mathbf{R}^{-1} \cdot \mathbf{H}]^{-1} \in \mathbb{R}^{n \times n} \quad (2.17)$$

Resolver el problema de optimización dado en la ecuación cuadrática descrita en 2.14 es equivalente a resolver el problema de Interpolación Óptima en una dimensión (Barker et al., 2004).

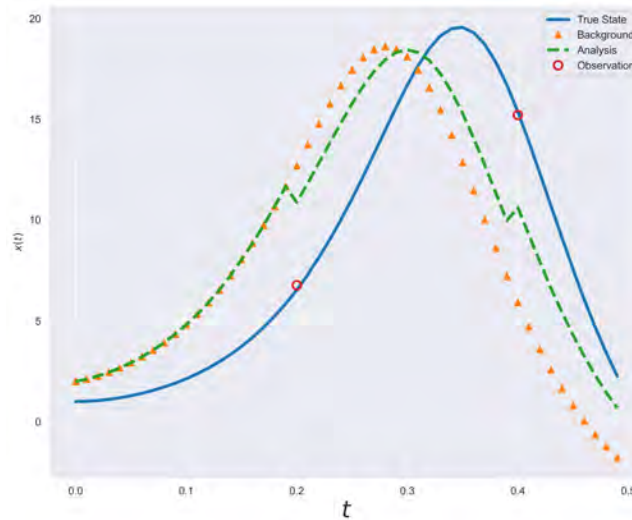


FIGURA 2.4: Asimilación secuencial de Datos

### 2.1.3. Ensemble Kalman Filter

Si se tiene un modelo  $\mathcal{M}$  con  $10^6$  componentes, para calcular  $\mathbf{B}$  se tendrían, en el peor de los casos, una complejidad de  $\mathcal{O}(n^3)$  y un espacio de memoria de 8Tb. Podemos ver que, a medida que el modelo crece en tamaño, es más inviable el cálculo explícito de esta matriz (Pourahmadi, 2011; Fan, Liao y Liu, 2016). Además, para resolver la ecuación 2.4 requerimos el uso de Adjointos y el cálculo del modelo Tangente, que son laboriosos y costosos computacionalmente (Gustafsson y Bojarova, 2014; Stengel et al., 2009). Así que, para tratar esto, consideremos  $N$  realizaciones del modelo de un estado inicial con una pequeña perturbación, que se llamará *Miembros del conjunto*, y estimamos la matriz de covarianza y su inversa o Matriz de precisión. Evitando así resolver el modelo Tangente o utilizar Adjointos (Kalnay, 2002). En el contexto del EnKF, se utiliza un conjunto de realizaciones del modelo, (Houtekamer y Mitchell, 1998; Stroud, Katzfuss y Wikle, 2018).

El filtro de Kalman, propuesto por primera vez en (Kalman, 1960; Kalman y Bucy, 1961) como resultado de la teoría de control, proporciona una descripción matemática del problema de asimilación, pero sufre de inconvenientes el Ensemble Kalman Filter (EnKF) trata de resolver (Evensen, 1994; Ito et al., 2016). Al igual que antes, el supuesto de gaussianidad debe cumplirse, para que el proceso de asimilación de datos sea significativo, y la predicción es la media de la posterior, y los momentos de la función de densidad de probabilidad gaussiana se estiman a partir de conjuntos de pequeño tamaño respecto a las dimensiones del espacio de estados (Nino Ruiz, Sandu y Anderson, 2014; Godinez y Moulton, 2012)

La popularidad del EnKF viene dada por su sencilla formulación y su relativa facilidad de aplicación (Lorenz, 2003b; Gillijns et al., 2006). En el EnKF, un conjunto de realizaciones del modelo  $N$ ,

$$\mathbf{X}^b = \left[ \mathbf{x}^{b[1]}, \mathbf{x}^{b[2]}, \dots, \mathbf{x}^{b[N]} \right] \in \mathbb{R}^{n \times N}, \quad (2.18)$$

Donde  $\mathbf{x}_k^{b[i]} \in \mathbb{R}^{n \times 1}$  se refiere al  $i$ -ésimo miembro del conjunto, para  $1 \leq i \leq N$ , en el momento  $k$ , para  $0 \leq k \leq M$ . Utilizamos este conjunto para estimar la distribución de errores a priori,

$$\mathbf{x} \sim \mathcal{N}(\mathbf{x}^b, \mathbf{B})$$

a través de los momentos empíricos del conjunto, en el que la media del conjunto es:

$$\mathbf{x}^b \approx \bar{\mathbf{x}}^b = \frac{1}{N} \cdot \sum_{e=1}^N \mathbf{x}^{b[e]} \in \mathbb{R}^{n \times 1}, \quad (2.19)$$

$$\mathbf{B} \approx \mathbf{P}^b = \frac{1}{N-1} \cdot \Delta \mathbf{X} \cdot \Delta \mathbf{X}^T \in \mathbb{R}^{n \times n}, \quad (2.20)$$

Donde  $\Delta \mathbf{X} \in \mathbb{R}^{n \times N}$  es la matriz de desviaciones de los miembros dada por,

$$\Delta \mathbf{X} = \mathbf{X}^b - \bar{\mathbf{x}}^b \cdot \mathbf{1}_N^T. \quad (2.21)$$

El paso de asimilación se realiza en realidad en el espacio abarcado por los miembros del conjunto, ya que esta es toda la información que realmente tenemos del modelo numérico. Cualquier elemento del espacio del conjunto puede escribirse como sigue:

$$\mathbf{x} = \mathbf{x}^b + \Delta \mathbf{X} \cdot \alpha, \quad (2.22)$$



El proceso de asimilación puede realizarse de la siguiente manera:

$$\mathbf{X}^a = \mathbf{X}^b + \mathbf{K} \cdot \mathbf{D} \in \mathbb{R}^{n \times N}, \quad (2.23)$$

$\mathbf{K}$  se conoce como la ganancia de Kalman,  $\mathbf{C}$  se conoce como la Matriz de la Innovación y  $\mathbf{D} \in \mathbb{R}^{m \times N}$  es la matriz de innovaciones sobre las observaciones sintéticas:

$$\begin{aligned} \mathbf{K} &= \mathbf{P}^b \cdot \mathbf{H}^T \cdot \mathbf{C}^{-1}, \\ \mathbf{C} &= \left[ \mathbf{R} + \mathbf{H} \cdot \mathbf{P}^b \cdot \mathbf{H}^T \right], \\ \mathbf{D} &= \mathbf{y} \cdot \mathbf{1}_N^T - \mathbf{H} \cdot \mathbf{X}^b + \mathbf{E}, \end{aligned} \quad (2.24)$$

en el que las columnas de  $\mathbf{E} \in \mathbb{R}^{m \times N}$  son muestras de una distribución normal de media cero con una matriz de covarianza de datos y errores  $\mathbf{R} \in \mathbb{R}^{m \times m}$  (Dovera y Della Rossa, 2011; Abaza et al., 2017).

Obteniendo como solución alternativa:

$$\begin{aligned} \mathbf{X}^a &= \mathbf{X}^b + \left[ \mathbf{B}^{-1} + \mathbf{H}^T \cdot \mathbf{R}^{-1} \cdot \mathbf{H} \right]^{-1} \cdot \mathbf{H}^T \cdot \mathbf{R}^{-1} \cdot \mathbf{D} \\ \mathbf{X}^a &= \left[ \mathbf{B}^{-1} + \mathbf{H}^T \cdot \mathbf{R}^{-1} \cdot \mathbf{H} \right]^{-1} \cdot \left[ \mathbf{B}^{-1} \cdot \mathbf{X}^b + \mathbf{H}^T \cdot \mathbf{R}^{-1} \cdot \mathbf{Y} \right] \end{aligned}$$

Una forma eficiente de calcular esas actualizaciones es la siguiente,

$$\mathbf{X}^a = \mathbf{X}^b + \mathbf{Z} \in \mathbb{R}^{n \times N}, \quad (2.25)$$

Donde  $\mathbf{Z} \in \mathbb{R}^{n \times N}$  se puede obtener mediante la solución del sistema lineal de ecuaciones,

$$\left[ \mathbf{B}^{-1} + \mathbf{H}^T \cdot \mathbf{R}^{-1} \cdot \mathbf{H} \right] \cdot \mathbf{Z} = \mathbf{H}^T \cdot \mathbf{R}^{-1} \cdot \Delta \mathbf{Y} \quad (2.26)$$

### Ensemble Kalman Filter basado en la descomposición de Cholesky Modificada

Otra implementación eficiente es un Filtro Kalman Ensemble basado en una descomposición Cholesky modificada (EnKF-MC). Esta implementación utiliza el concepto de predecesores espaciales para obtener estimadores dispersos de las matrices de precisión (Levina, Rothman, Zhu et al., 2008). Los predecesores del componente  $i$  del modelo, a partir de ahora  $\Pi(i, r)$ , para  $1 \leq i \leq n$  y un radio de influencia  $r \in \mathbb{Z}^+$ , vienen dadas por el conjunto de componentes cuyas etiquetas son menores que la del  $i$ -ésimo.

Esta idea se explota en la formulación EnKF propuesta en (Nino-Ruiz, Sandu y Deng, 2017; Nino-Ruiz, Sandu y Deng, 2018) donde se emplea el siguiente estimador para aproximar la matriz de precisión (Bickel, Levina et al., 2008)

$$\hat{\mathbf{B}}^{-1} = \mathbf{L}^T \cdot \mathbf{D}^{-1} \cdot \mathbf{L} \in \mathbb{R}^{n \times n}, \quad (2.27)$$

Donde  $\mathbf{L} \in \mathbb{R}^{n \times n}$  es una matriz triangular inferior unitaria, y  $\mathbf{D} \in \mathbb{R}^{n \times n}$  es una matriz diagonal.

donde el factor Cholesky  $\mathbf{L}_k \in \mathbb{R}^{n \times n}$  es una matriz triangular inferior,

$$\{\mathbf{L}_k\}_{i,v} = \begin{cases} -\beta_{i,v,k} & , v \in P(i, r) \\ 1 & , i = v \\ 0 & , \text{otros casos} \end{cases}, \quad (2.28)$$

cuyos elementos triangulares inferiores estrictos  $\beta_{i,v,k}$  se obtienen ajustando los parámetros de diferentes modelos dados por,

$$\mathbf{x}_{[i]_k}^T = \sum_{v \in P(i, \delta)} \beta_{i,v,k} \cdot \mathbf{x}_{[v]_k}^T + \gamma_{ik} \in \mathbb{R}^{N \times 1}, \quad 1 \leq i \leq n, \quad (2.29)$$

Donde  $\mathbf{x}_{[i]_k}^T \in \mathbb{R}^{N \times 1}$  denota la  $i$ -ésima fila (componente del modelo) del conjunto, componentes del vector  $\gamma_{ik} \in \mathbb{R}^{N \times 1}$  son muestras de una distribución normal con media cero y varianza desconocida  $\sigma_k^2$ , y  $\mathbf{D}_k \in \mathbb{R}^{n \times n}$  es una matriz diagonal cuyos elementos diagonales son los siguientes

$$\begin{aligned} \{\mathbf{D}_k\}_{i,i} &= \widehat{\mathbf{var}} \left( \mathbf{x}_{[i]_k}^T - \sum_{v \in P(i, \delta)} \beta_{i,v,k} \cdot \mathbf{x}_{[v]_k}^T \right)^{-1} \\ &\approx \mathbf{var}(\gamma_{ik})^{-1} = \frac{1}{\sigma_k^2} > 0, \\ \text{con } \{\mathbf{D}_k\}_{1,1} &= \widehat{\mathbf{var}} \left( \mathbf{x}_{[1]_k}^T \right)^{-1}, \end{aligned}$$

Donde  $\mathbf{var}(\bullet)$  y  $\widehat{\mathbf{var}}(\bullet)$  denotan las varianzas real y empírica, respectivamente. En  $\mathbf{D}_k$  se almacenan los residuos del modelo de regresión que hemos utilizado. La estructura de  $\mathbf{L}$  puede ser dispersa, lo que implica un gran ahorro en términos de uso de memoria en las configuraciones actuales de asimilación de datos operativas en las que  $n$  puede ser muy grande. Además,  $\widehat{\mathbf{B}}^{-1}$  puede ser representado en términos de sus factores de Cholesky y por lo tanto, se pueden derivar maneras eficientes de calcular el conjunto Nino Ruiz, Sandu y Anderson, 2014.

En este contexto, el conjunto posterior se calcula como sigue:

$$\mathbf{X}^a = \mathbf{X}^b + \widehat{\mathbf{A}} \cdot \Delta \mathbf{Y} \in \mathbb{R}^{n \times N}, \quad (2.30)$$

donde la matriz de covarianza del error de análisis dice

$$\widehat{\mathbf{A}} = \left[ \widehat{\mathbf{B}}^{-1} + \mathbf{H}^T \cdot \mathbf{R}^{-1} \cdot \mathbf{H} \right]^{-1} \in \mathbb{R}^{n \times n}, \quad (2.31)$$

Usando identidades algebraicas podemos obtener:

$$\widehat{\mathbf{B}} = \mathbf{L}^{-1} \cdot \mathbf{D} \cdot \mathbf{L}^{-T} \in \mathbb{R}^{n \times n},$$

La estructura tanto de  $\widehat{\mathbf{B}}$  como de  $\widehat{\mathbf{B}}^{-1}$  está estrictamente relacionada con la estructura dispersa de  $\mathbf{L}$ , por lo que podemos aprovechar para obtener ambas sin su cómputo explícito, que es prohibitivo en contextos de mayor dimensión.

### Local Ensemble Kalman Filter

Para reducir los efectos del ruido de muestreo, se proponen técnicas de *localization*, tales métodos tratan las correlaciones espurias que podrían desarrollarse en el proceso de asimilación. (Houtekamer y Mitchell, 1998; Anderson, 2001; Buehner, 2011; Kondo y Miyoshi, 2016). Estas técnicas se basan en funciones de localización que definen la longitud de localización. En estos métodos, para cada componente del modelo  $1 \leq i \leq n$ , se define una vecindad  $P(i, \delta)$  basada en algún etiquetado de los componentes del modelo y un radio de influencia  $r \in \mathbb{R}^+$  (Nino-Ruiz, Sandu y Deng, 2018), por lo tanto,

$$j \in P(i, r) \Leftrightarrow d(x_i, x_j)^2 \leq r^2, \text{ y } j < i, \quad (2.32)$$

Donde  $d(\bullet, \bullet)$  denota una función de distancia. En la práctica, este etiquetado puede hacerse de muchas maneras diferentes, siendo la más común la de fila mayor y columna mayor. Esta definición se aplica para descartar componentes muy distanciados que no están relacionados, según los supuestos del modelo. De este modo, se impone una estructura matricial dispersa de la matriz de covarianza estimada. Esta técnica se basa en un conocimiento previo de la estructura de la covarianza. Por ejemplo, la distancia geográfica podría hacernos creer que los componentes distantes no están correlacionados.

1	5	9	13
2	6	10	14
3	7	11	15
4	8	12	16

FIGURA 2.5: Etiquetado

Una forma tradicional de conseguirlo es utilizando el producto de Schur denotado por  $\circ$ . Dada la matriz  $\hat{\Lambda}$ , que se define por

$$\{\hat{\Lambda}\}_{i,j} = \exp\left(-\frac{1}{2} \cdot \frac{d(i,j)^2}{r^2}\right), \text{ para } 1 \leq i, j \leq n, \quad (2.33)$$

Podemos localizar  $\mathbf{P}^b$  aplicando

$$\tilde{\mathbf{P}}^b = \hat{\Lambda} \circ \mathbf{P}^b \in \mathbb{R}^{n \times n}, \quad (2.34)$$

La función de distancia suele ser la distancia euclidiana. Este enfoque filtra las correlaciones espurias de largo alcance.

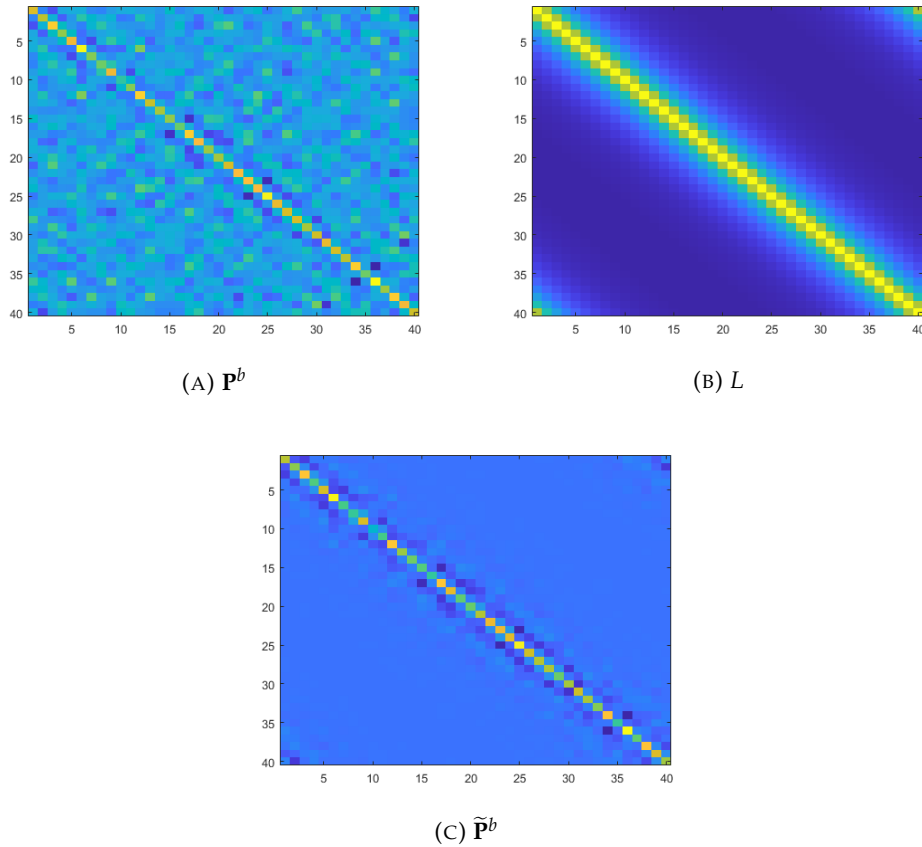


FIGURA 2.6: Localización

La selección correcta del radio de influencia es el núcleo en esta técnica, ya que un rango de influencia amplio significaría que mantenemos esas correlaciones espurias, pero si es demasiado corto, entonces las correlaciones dinámicas podrían perderse. En la práctica, los métodos de localización se utilizan comúnmente para aumentar artificialmente el rango de  $\mathbf{P}^b$  y para mitigar el impacto de las correlaciones espurias durante los pasos del análisis (Hamill, Whitaker y Snyder, 2001). Estos métodos de localización pueden ser: localización de la matriz de covarianza (**B**-localización) (Lei, Whitaker y Bishop, 2018), localización de dominios, y localización de observaciones (**R**-localización) (Anderson, 2001; Han, Zhang y Sun, 2018; Anderson, 2019).

También podemos pensar en una localización espacial en la que cada componente está rodeada por una caja local, y para cada caja local realizamos una asimilación local con toda la información contenida en ella, luego mapeada de nuevo al dominio global donde se realiza el análisis. Un ejemplo en 2D se puede ver en 2.7 donde el punto rojo es el componente y el cuadrado rojo son los componentes vecinos.

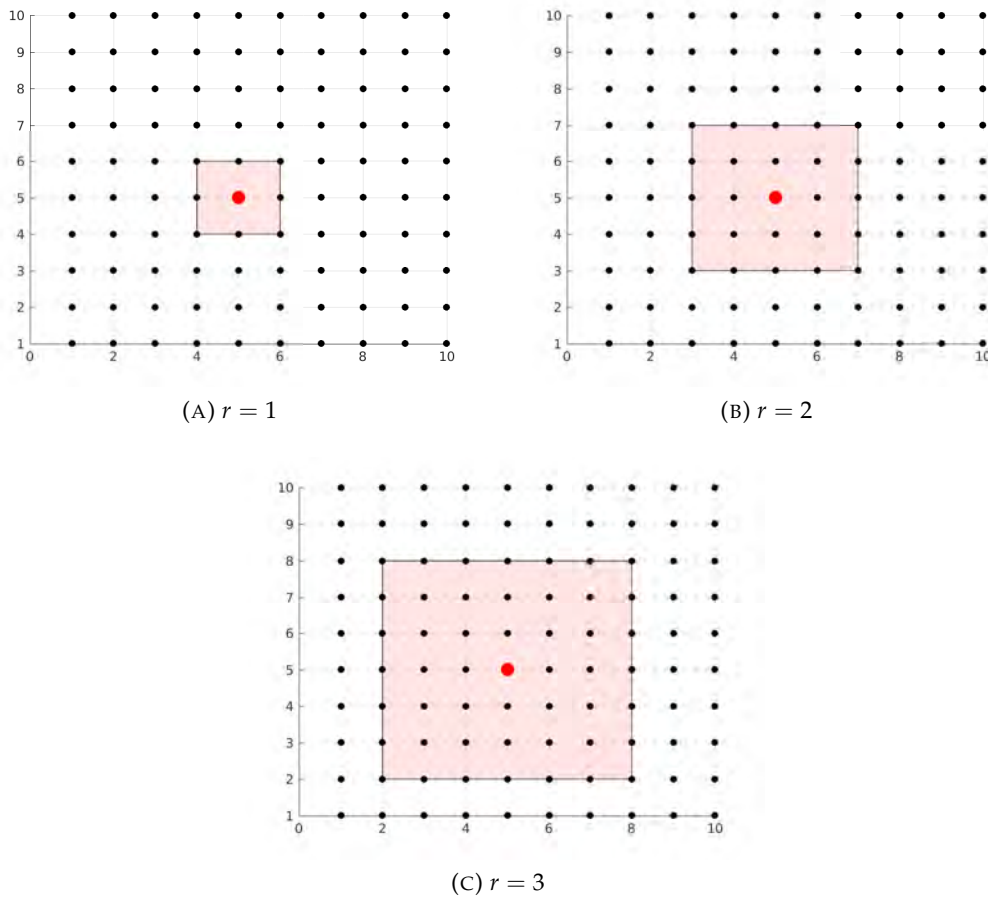


FIGURA 2.7: Dominios locales para diferentes radios de influencia  $r$ .

El local ensemble Kalman filter (LEnKF) (Ott et al., 2004b; Tong, 2018) explota la información local a través de una longitud de radio  $\delta$  para mitigar el impacto de las correlaciones espurias. Equivale a aplicar las ecuaciones EnKF (2.23) a cada componente del modelo dentro de su correspondiente caja local.

### Local Ensemble transform Kalman Filter

Otros métodos deterministas, como el filtro de Kalman de transformación de conjunto local (LETKF). El Filtro de Kalman de Transformación Local de Conjuntos, una formulación determinista del EnKF funciona utilizando esta técnica de localización, y ha sido ampliamente utilizado en los centros de asimilación de datos operacionales (Hunt, Kostelich y Szunyogh, 2007; Tippett et al., 2003). La media de la distribución de análisis se estima en el espacio del ensemble de la siguiente manera, evitando el uso de datos sintéticos para evitar inducir más error de muestreo durante los pasos de asimilación. En la formulación global del LETKF, el modo posterior se calcula como sigue:

$$\bar{\mathbf{x}}^a = \bar{\mathbf{x}}^b + \Delta \mathbf{X}^b \cdot \left[ (N-1) \cdot \mathbf{I} + \mathbf{Q}^T \cdot \mathbf{R}^{-1} \cdot \mathbf{Q} \right]^{-1} \cdot \mathbf{Q}^T \cdot \mathbf{R}^{-1} \cdot \left[ \mathbf{y} - \mathbf{H} \cdot \bar{\mathbf{x}}^b \right] \in \mathbb{R}^{n \times 1}, \quad (2.35a)$$

Donde  $\mathbf{Q} = \mathbf{H} \cdot \Delta \mathbf{X} \in \mathbb{R}^{m \times N}$ . El conjunto posterior se construye entonces sobre (2.35a) como sigue:

$$\mathbf{X}^a = \bar{\mathbf{x}}^a \cdot \mathbf{1}^T + \Delta \mathbf{X} \cdot \left[ (N-1) \cdot \mathbf{I} + \mathbf{Q}^T \cdot \mathbf{R}^{-1} \cdot \mathbf{Q} \right]^{-1/2}. \quad (2.35b)$$

De forma similar al LEnKF, el conjunto de ecuaciones (2.35) se aplica a cada componente del modelo dentro de su caja local.

El paso de asimilación se aplica a cada componente del modelo para un radio de influencia determinado  $r$ , con el que se obtendrá el estado de análisis global. Dado que el cálculo más costoso es la inversión de  $\widetilde{\mathbf{P}}^a$ , el esfuerzo computacional del LETKF se lee,

$$\mathcal{O}(\varphi \cdot n \cdot N^3),$$

donde  $\varphi$  denota el tamaño de las cajas locales.

## 2.2. Numerical Model SPEEDY

Los experimentos se realizan utilizando el Simplified Parameterizations, primitive-Equation DYNamics model (SPEEDY) (Molteni, 2003a; Bracco et al., 2004). El modelo SPEEDY es un modelo de circulación general atmosférica (AT-GCM) que imita el comportamiento de la atmósfera en ocho niveles de presión (Miyoshi, 2011): 30mb, 100mb, 200mb, 300mb, 500mb, 700mb, 850mb, y 925mb. Por defecto, este modelo emplea una resolución espectral T-30 (96 componentes zonales y 48 meridionales) en ocho capas numéricas Molteni, 2003b; Kucharski, Molteni y Bracco, 2006 pero, esta resolución puede ser modificada, por ejemplo, para utilizar T-21 (32 × 64 componentes de la red), T-47 (72 × 144), T-63 (96 × 192) y T-103 (160 × 320), cada uno con 8 capas numéricas. Este modelo implementa una integración de salto de dos pasos y los componentes de la malla espacial se muestran en la figura 2.8 para una sola capa y diferentes resoluciones numéricas. Para todas las resoluciones del modelo, las variables físicas se detallan en la Tabla 2.1 con sus correspondientes unidades y el número de capas numéricas.

Nombre	Notación	Unidad	Numero de Niveles
Temperatura	$T$	$K$	8
Componente Zonal del Viento	$u$	$m/s$	8
Componente Meridional del Viento	$v$	$m/s$	8
Humedad Especifica	$Q$	$g/kg$	8
Presión	$P$	$hPa$	1

CUADRO 2.1: Variables físicas del modelo AT-GCM Speedy .

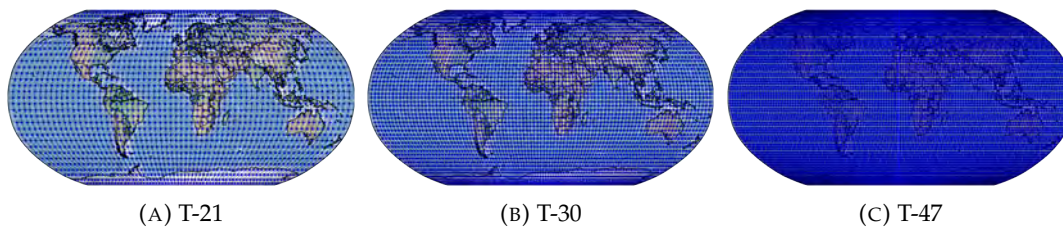


FIGURA 2.8: Componentes del modelo para algunas resoluciones espectrales

La resolución del modelo T-21 ( $32 \times 64$  componentes de malla) es la utilizada para la discretización espacial horizontal en estos experimentos. Cinco variables del modelo forman parte del proceso de asimilación: la temperatura ( $K$ ), las componentes del viento zonal y meridional ( $m/s$ ), la humedad específica ( $g/kg$ ) y la presión superficial ( $hPa$ ). El número total de componentes del modelo es  $n = 67,584$ . El número de miembros del conjunto es de  $N = 80$  para todos los escenarios. El espacio de estados del modelo es aproximadamente 844 veces mayor que el número de miembros del conjunto ( $n \gg N$ ), lo cual es muy común en los escenarios operativos de DA. A continuación se describen otros detalles de la configuración experimental, algunos de los cuales son similares a los detallados en (Miyoshi, Kondo e Imamura, 2014):

- Partiendo de un sistema en equilibrio, el modelo se integra durante un largo período de tiempo para obtener una condición inicial cuya dinámica es consistente con la del modelo SPEEDY.
- La condición inicial se perturba  $N$  veces y se propaga durante un largo período de tiempo a partir del cual se obtiene el conjunto inicial de fondo.
- Empleamos la trayectoria de la condición inicial como la de referencia. Esta trayectoria de referencia sirve para construir observaciones sintéticas.
- Dejamos las desviaciones estándar de los errores en las observaciones como
  - Temperatura  $1 K$ .
  - Componente del viento zonal  $1 m/s$ .
  - Componente del viento meridional  $1 m/s$ .
  - Humedad específica  $10^{-3} g/kg$ .
  - Presión  $100 hPa$ .
- Los experimentos se realizan bajo supuestos de modelo perfecto.
- El número de pasos de asimilación se lee  $M = 30$ .

Las diferentes formulaciones de EnKF se implementaron utilizando Python y bibliotecas especializadas como `scipy`, `numpy` y `NetCDF`, con los parámetros utilizados  $\delta$ ,  $p$  y  $\alpha$ . Donde  $\delta$  son los radios de influencia,  $p$  es el operador observacional, el que determina el peso de las observaciones y  $\alpha$  es la influencia. Consideramos redes observacionales dispersas y variamos el radio de influencia y la inflación para las formulaciones mencionadas anteriormente (EnKF-MC, LETKF, LEnKF), y para cada una de ellas, realizamos la asimilación teniendo en cuenta el paso Leapfrog (Leap) y asimilamos utilizando sólo el estado observado (NoLeap).





## Capítulo 3

# Objetivos

### 3.1. Planteamiento del problema

Teniendo en cuenta investigaciones anteriormente mencionadas, podemos destacar los siguientes problemas:

- Alta demanda computacional para realizar simulaciones meteorológicas.
- Alta sensibilidad a ruido de muestreo en metodos basados en conjuntos.
- Los datos reales que ajusten exactamente a la resolución de un modelo son difíciles de obtener.
- Debido a la alta complejidad computacional de los modelos replicar pruebas con diferentes modelos se vuelve muy complejo.
- El acceso rápido o en tiempo real a los datos meteorológicos en Colombia es difícil. Los datos comúnmente son obtenidos mediante peticiones a agencias como INVEMAR e IDEAM los cuales son cargados mensualmente.
- No existen suficientes modelos que permitan estimar componentes meteorológicas en Colombia lo que dificulta comprender mayormente nuestro ecosistemas.

### 3.2. Contribuciones Esperadas

Las contribuciones que se esperan obtener en esta propuesta son:

- Comparar las diferentes implementaciones del EnKF en cuestiones de tiempo y resultados.
- Implementar una forma sencilla de modificar el modelo numérico, así como las herramientas de captura de datos.
- Desarrollar una herramienta que permita obtener datos con poco retraso desde la NCAR para la resolución de malla especificada.

### 3.3. Objetivos

#### 3.3.1. Objetivo General

Comparar resultados de diferentes formulaciones eficientes de métodos basados en conjuntos para asimilación secuencial de datos reales ajustados a la resolución de la malla.

### 3.3.2. Objetivos Específicos

1. Diseñar e implementar herramienta de captura y redimensionado de datos reales para modelo SPEEDY.
2. Probar diferentes implementaciones de asimilación secuencial de datos con datos reales redimensionados a la resolución de una malla.
3. Validar las formulaciones a través de métricas de la literatura especializada.

## 3.4. Metodología

En esta sección, describimos brevemente los pasos para lograr los objetivos de la Sección 3.3.

1. Para diseñar e implementar herramienta de captura y redimensionado de datos reales para modelo SPEEDY.
  - Los datos son descargados mediante una petición FTP a la pagina de la NCAR descargando datos utilizados para entrenamientos de un Modelo Troposférico Global usado por la NCEP (National Centers for Environmental Prediction). Basado en (Li et al., 2012)
  - Se redimensiona los datos a la resolución de la malla especificada para el modelo usando interpolación lineal.
  - Los datos son organizados como observaciones y condiciones iniciales para la ventana de tiempo de los últimos 3 meses.
2. Probar diferentes implementaciones de asimilación secuencial de datos con datos reales redimensionados a la resolución de una malla.
3. Validar las formulaciones a través de métricas de la literatura especializada. Para validar los métodos de asimilación utilizados en este modelo.
  - Se compara el error entre las observaciones reales y los datos predichos por el modelo a partir de una condición inicial real. (Conway et al., 1994)

## Capítulo 4

# Implementación

### 4.1. Diseñar e implementar herramienta de captura y redimensionado de datos reales para modelo SPEEDY

Para llevar a cabo este proceso de obtener los datos reales, redimensionarlos e interpolarlos a las horas y resoluciones para la ejecución del modelo SPEEDY fue necesario realizar los siguientes pasos:

#### 4.1.1. Captura de datos reales

Inicialmente, se selecciono un repositorio web donde se pudiera realizar peticiones seguidas sobre archivos que contengan la mayor cantidad de información en el menor tiempo posible, para esto fue seleccionada el archivo de datos de investigación de la NCAR/UCAR con los datos brindados por la base de datos de la NCEP FNL Operational Model Global Tropospheric Analyses (National Centers for Environmental Prediction, National Weather Service, NOAA, U.S. Department of Commerce, 2000).

Estos datos constan con características que son utiles para esta implementación, como lo son:

- Datos reales alrededor del globo cada 6 horas, desde Julio de 1999.
- Resolución espectral bastante fina, y punto por cada grado de latitud y longitud ( $180 \times 360$ ).
- Los análisis están disponibles en la superficie, en 26 niveles de presión obligatorios desde 1000 milibares hasta 10 milibares.
- Los parámetros incluyen la presión en superficie, la presión a nivel del mar, la altura geopotencial, la temperatura, la temperatura de la superficie del mar, los valores del suelo, la cubierta de hielo, la humedad relativa y especifica, los vientos en u y v, el movimiento vertical, la vorticidad y el ozono.
- Los datos están en formato grib2 lo que los hace fácil de manejar y convertir en NETCDF4.
- Para la captura de estos datos es necesario, tener una cuenta creada en la pagina de NCAR/UCAR.

Para esto se utiliza el paquete AMLCS-main que consta con el modelo y las tres implementaciones del Ensemble Kalman Filter y el paquete data que es la implementación desarrollada para usar los datos reales en el modelo.

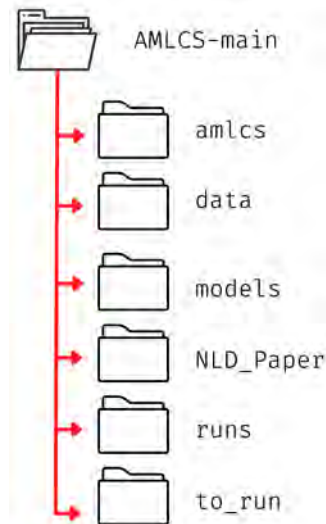


FIGURA 4.1: AMLCS-main

Para la descarga de estos archivos se utilizó el archivo que brinda la página de la NCAR para descargar los archivos que se seleccionen en Python, el cual fue modificado para descargar los archivos en las horas 00 y 06 de la fecha especificada (este archivo fue nombrado *download.py* dentro del paquete *data*).

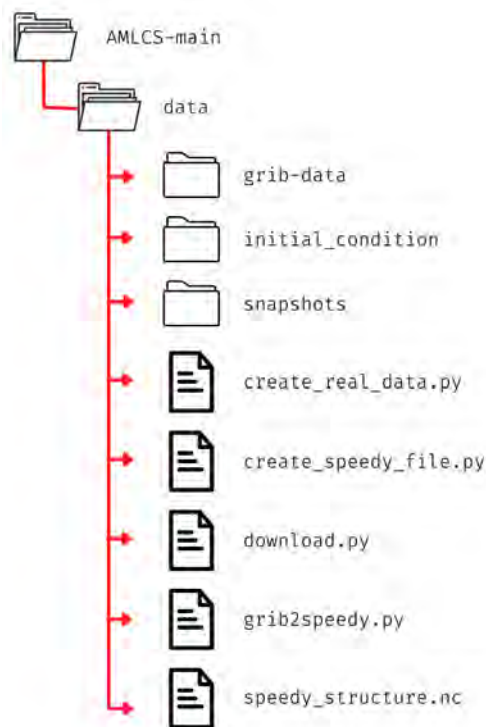


FIGURA 4.2: Paquete data

Todos los archivos descargados desde esta herramienta son almacenados en la carpeta *grib-data*. Posteriormente, se ejecuta el archivo *grib2speedy.py* especificando la fecha de los datos a seleccionar, esto teniendo en cuenta las variables utilizadas por SPEEDY para su funcionamiento como lo son Temperatura, Presión en la superficie, componente  $u$ , componente  $v$  del viento y humedad específica

### 4.1.2. Interpolación

Una vez se tienen seleccionada cada variable, se procede a hacer el redimensionado de la malla y hallar la hora 01 para esto se hace uso de la interpolación cubica y lineal respectivamente.

#### Redimensionado

Para el redimensionado en el archivo *grib2speedy.py* se tiene una función que con la ayuda de la librería XArray, redimensiona la malla numérica de  $180 \times 360$  a las medidas T-21 del modelo SPEEDY  $32 \times 64$  haciendo uso del método *dataset.interp* que interpola la malla mediante una aproximación cubica. Guardando estos datos en archivos de Numpy, por el nombre de la variable (*variable.npy*), recordando que esto para las horas 00 y 06.

#### Hallar hora 01

En este paso se usa el archivo *create\_speedy\_file.py*, el cual dada una fecha y el nombre del archivo (*initial\_condition.nc* o *reference\_solution\_i.nc* donde *i* es el identificador del miembro del conjunto) realiza los pasos indicados previamente, luego de esto realiza una interpolación lineal sencilla para hallar los valores de cada variable en la hora 01 teniendo como datos iniciales las hora 00 y 01. Después de realizado este paso, se procede a la creación del archivo con la estructura de los archivos de SPEEDY, guardándolos en su formato respectivo (NETCDF4) y en la carpeta correspondiente teniendo en cuenta el nombre del archivo.

Todo este proceso se puede automatizar ejecutando el archivo *create\_real\_data.py* creando los archivos de las soluciones de referencia para cada paso de asimilación y la condición inicial en sus respectivas carpetas para una ejecución para el día actual en las horas 00 y 01.

## 4.2. Probar diferentes implementaciones de asimilación secuencial de datos con datos reales redimensionados a la resolución de una malla.

Una vez creado las soluciones de referencia para cada paso de asimilación reales y la condición inicial real, se procede a ejecutar la preparación del modelo. En la cual se crea la carpeta *NLD\_Paper* y dentro de esta la carpeta *t21\_80\_0.05\_30* donde se almacenan los datos de configuración y ejecución del modelo previamente configurado, una vez realizado esto se cargan las soluciones de referencia para cada paso de asimilación y condición inicial, luego se crea el conjunto o ensemble inicial y los miembros de los conjuntos que se usaran en los pasos de asimilación. Para esto se hará uso del paquete AML-CS mencionado en (Consuegra Ortega et al., 2021).

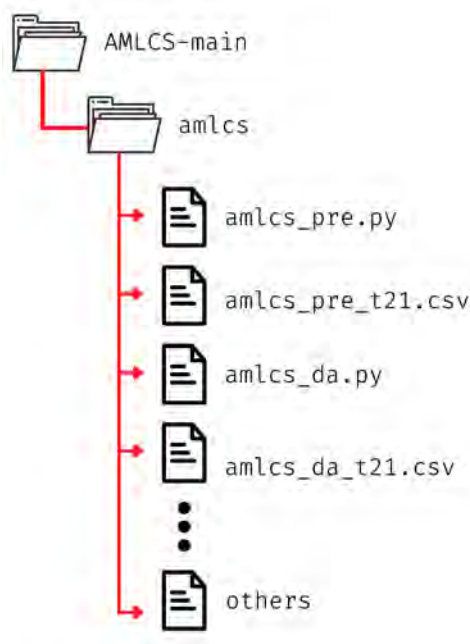


FIGURA 4.3: Paquete AMLCS

Para esto se realizarán pruebas con las siguientes configuraciones, para las implementaciones de EnKF basado en la descomposición de Cholesky modificada, LEnKF y LETKF. Guardando los resultados de cada ejecución para posteriormente ser analizados, tanto en tiempo de ejecución como en el error que hay entre las soluciones de referencia para cada paso de asimilación y lo predicho por el modelo.

#### 4.2.1. Ejecución de SPEEDY

##### Preparación del modelo

El paquete AMLCS consta de 2 archivos fundamentales para la ejecución de SPEEDY usando (o no) asimilación de datos, los cuales son *amlcs\_pre.py* y *amlcs\_da.py*.

- Configuración de preparación del modelo Para la configuración de preparación del modelo se usa el archivo *amlcs\_pre\_t21.csv* en el cual se especifica:
  - *Nens*, define el número de miembros del ensemble a utilizar en el proceso de asimilación.
  - *M*, define la ventana de tiempo en la que se ejecutará el modelo.
  - *res\_name*, elige la resolución (o especificación) del modelo que se puede utilizar.
  - *per*, establece el valor de la perturbación para las observaciones.
  - *obs\_steps*, establece cuantas veces se tomarán las observaciones. Puede ser horas, días o cualquier otra unidad del modelo.
  - *ini\_steps*, esto se refiere al número de días a utilizar en la construcción de las condiciones iniciales.
  - *ini\_imes*, se refiere al número de veces que se propagará la condición inicial.
  - *syn\_tests*, establece si se utilizarán observaciones sintéticas o no (un booleano).

- *folder\_prep*, la ruta en la que se colocará la inicialización. Puede ser absoluta o relativa.
  - *par* define si los conjuntos pueden propagarse en paralelo durante su construcción.
- Ejecutar script de preparación Una vez configurado el archivo anterior, se ejecuta el archivo *amlcs\_pre.py* el que se encarga de construir lo siguiente:
- Condición inicial del modelo (carpeta *initial\_condition*)
  - Ejecución libre del modelo sin Asimilación de datos (carpeta *free\_run*)
  - Solución de referencia para cada paso de asimilación (carpeta *snapshots*)
  - Miembros del conjunto (carpeta *ensemble\_0*)
  - Archivos del modelo con sus respectivas configuraciones.

### Ejecución del modelo con Asimilación de datos

- Configuración de asimilación del modelo Para configurar el archivo *amlcs\_da\_t21.csv* se debe tener en cuenta los siguientes parámetros:
- *r*, el radio de influencia utilizado en el proceso de localización
  - *s*, la escasez de observaciones.
  - *res\_name*, la resolución del modelo que ha sido compilada y ajustada previamente.
  - *method*, el nombre del método, según el disponible.
  - *exp\_settings*, la ruta en la que se ha realizado el pretratamiento. Puede ser relativa o absoluta.
  - *infla*, la inflación utilizada en el método (si se requiere)
  - *err\_obs*, el error en las observaciones de cada variable. Se utiliza para construir **R**.
  - La lista *obs\_plc* denotan qué variables se observan en:

$$[u_0, v_0, T_0, q_0, \rho_0, u_1, v_1, T_1, q_1, \rho_1],$$

por ejemplo, una lista de la forma:  $[1, 1, 1, 1, 1, 0, 0, 0, 0, 0]$  significa que se observan todas las variables del primer paso (en el sentido de salto)

- *list\_snapshots*, una lista elegida de momentos en el proceso de asimilación para ser almacenada, guardando el Análisis y el fondo para cada uno.
  - *option\_mask*, cuál de los ajustes de localización se utilizará, con respecto a lo mencionado en la sección anterior.
- Ejecutar script del modelo con asimilación de datos Este módulo produce los resultados de la asimilación para una configuración determinada, creando:
- Ejecución libre del modelo (sin DA).
  - Condición inicial utilizada.
  - Los errores para cada variable y para cada nivel, en su respectivo archivo CSV.
  - Solución de referencia utilizada.

- Las instantáneas para cada tiempo en *list\_snapshots*.

Una vez construida la configuración experimental, estamos listos para probar los métodos de asimilación de datos. Como hemos mencionado antes, en este paquete se utilizan tres métodos secuenciales: LEnKF, LETKF y EnKF-MC.



## Capítulo 5

# Resultados

Los resultados obtenidos en la ejecución del modelo SPEEDY fueron obtenidos con la siguiente configuración y sus distintas combinaciones, esto teniendo en cuenta que para cada predicción del modelo se uso datos reales, obtenidos de la pagina web de la NCAR.

Los parámetros utilizados son los indicados en la tabla 5.1

Parámetros	Valores
$\delta$	1
$p$	100 %,25 %,11 %,6 %,4 %
$\alpha$	1.02,1.04,1.06,1.08

CUADRO 5.1: Parámetros utilizados en los experimentos

Donde  $\delta$  son los radios de influencia,  $p$  es el operador observacional, el que determina el peso de las observaciones y  $\alpha$  es la influencia.

A continuación se muestran gráficas a nivel global de la diferencia entre lo predicho por el modelo, usando el método de asimilación mencionado (con la parametrización correspondiente) y los datos reales observados.

P	a	Metodo Asimilacion	Predicho	Observado
100	1.02	EnKF	196.587466	206.417383
		LETKF	201.950946	210.333133
		LEnKF	215.948973	222.419507
	1.04	EnKF	180.434960	194.134779
		LETKF	201.950946	210.333133
		LEnKF	215.948973	222.419507
	1.06	EnKF	168.061928	185.841462
		LETKF	201.950946	210.333133
		LEnKF	215.948973	222.419507
	1.08	EnKF	158.586618	180.510092
		LETKF	201.950946	210.333133
		LEnKF	215.948973	222.419507
11	1.02	EnKF	1278.749364	1245.282725
		LETKF	487.302564	487.803084
		LEnKF	1259.160822	1234.650558
	1.04	EnKF	1690.321243	1631.910018
		LETKF	487.302564	487.803084
		LEnKF	1259.160822	1234.650558
	1.06	EnKF	2293.126996	2192.649841
		LETKF	487.302564	487.803084
		LEnKF	1259.160822	1234.650558
	1.08	EnKF	3205.281702	3028.983424
		LETKF	487.302564	487.803084
		LEnKF	1259.160822	1234.650558
25	1.02	EnKF	1664.815292	1611.908754
		LETKF	457.041578	458.170030
		LEnKF	1670.990508	1638.634431
	1.04	EnKF	2286.061498	2191.092349
		LETKF	457.041578	458.170030
		LEnKF	1670.990508	1638.634431
	1.06	EnKF	3206.386843	3048.221588
		LETKF	457.041578	458.170030
		LEnKF	1670.990508	1638.634431
	1.08	EnKF	4579.865026	4319.782901
		LETKF	457.041578	458.170030
		LEnKF	1670.990508	1638.634431
4	1.02	EnKF	787.586615	778.004703
		LETKF	496.437040	496.581468
		LEnKF	870.671082	858.208413
	1.04	EnKF	960.262177	933.537273
		LETKF	496.437040	496.581468
		LEnKF	870.671082	858.208413
	1.06	EnKF	1234.865856	1166.289203
		LETKF	496.437040	496.581468
		LEnKF	870.671082	858.208413
	1.08	EnKF	835.008822	812.117905
		LETKF	1683.444635	1531.357016
		LEnKF	870.671082	858.208413
6	1.02	EnKF	835.008822	812.117905
		LETKF	496.393236	496.667490
		LEnKF	986.419519	968.129072
	1.04	EnKF	1070.149382	1014.878959
		LETKF	496.393236	496.667490
		LEnKF	986.419519	968.129072
	1.06	EnKF	1516.239956	1399.868390
		LETKF	0.000000	0.000000
		LEnKF	986.419519	968.129072
	1.08	EnKF	2295.923665	2079.633459
		LETKF	496.393236	496.667490
		LEnKF	986.419519	968.129072

CUADRO 5.2: Nivel Superficie Temperatura

P	a	Metodo Asimilacion	Predicho	Observado
100	1.02	EnKF	353.364343	365.208414
		LETKF	357.916263	366.402133
		LEnKF	363.298993	369.682731
	1.04	EnKF	342.108507	361.637186
		LETKF	357.916263	366.402133
		LEnKF	363.298993	369.682731
	1.06	EnKF	330.584341	359.372467
		LETKF	357.916263	366.402133
		LEnKF	363.298993	369.682731
	1.08	EnKF	319.443958	358.350652
		LETKF	357.916263	366.402133
		LEnKF	363.298993	369.682731
11	1.02	EnKF	455.507861	456.622574
		LETKF	455.413098	456.144702
		LEnKF	456.108728	456.659985
	1.04	EnKF	455.530782	457.589960
		LETKF	455.413098	456.144702
		LEnKF	456.108728	456.659985
	1.06	EnKF	456.757865	460.028793
		LETKF	455.413098	456.144702
		LEnKF	456.108728	456.659985
	1.08	EnKF	459.201892	465.222106
		LETKF	455.413098	456.144702
		LEnKF	456.108728	456.659985
25	1.02	EnKF	442.141168	444.578646
		LETKF	440.191941	441.968887
		LEnKF	443.368064	444.784337
	1.04	EnKF	441.398839	445.049571
		LETKF	440.191941	441.968887
		LEnKF	443.368064	444.784337
	1.06	EnKF	444.211518	449.276663
		LETKF	440.191941	441.968887
		LEnKF	443.368064	444.784337
	1.08	EnKF	454.918903	460.304126
		LETKF	440.191941	441.968887
		LEnKF	443.368064	444.784337
4	1.02	EnKF	463.573643	463.862669
		LETKF	464.145204	464.381631
		LEnKF	464.046069	464.179554
	1.04	EnKF	463.416251	463.964003
		LETKF	464.145204	464.381631
		LEnKF	464.046069	464.179554
	1.06	EnKF	463.153898	463.992680
		LETKF	464.145204	464.381631
		LEnKF	464.046069	464.179554
	1.08	EnKF	463.098428	464.348526
		LETKF	464.145204	464.381631
		LEnKF	464.046069	464.179554
6	1.02	EnKF	458.099328	459.180739
		LETKF	457.948477	458.534792
		LEnKF	459.110388	459.633387
	1.04	EnKF	458.011120	459.846782
		LETKF	457.948477	458.534792
		LEnKF	459.110388	459.633387
	1.06	EnKF	458.845277	461.283189
		LETKF	0.000000	0.000000
		LEnKF	459.110388	459.633387
	1.08	EnKF	461.984995	464.346462
		LETKF	457.948477	458.534792
		LEnKF	459.110388	459.633387

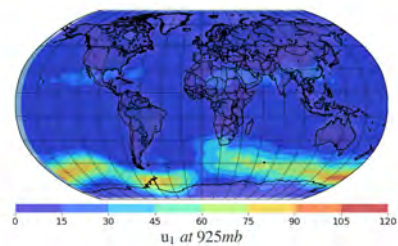
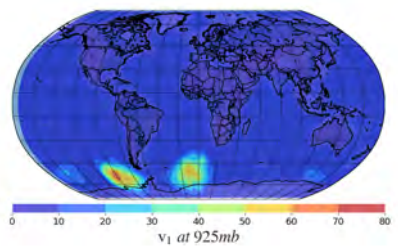
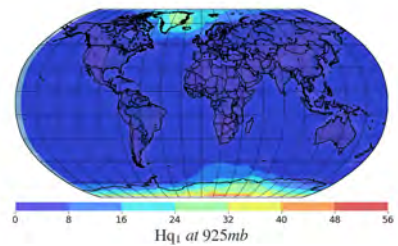
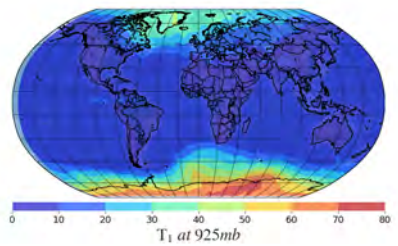
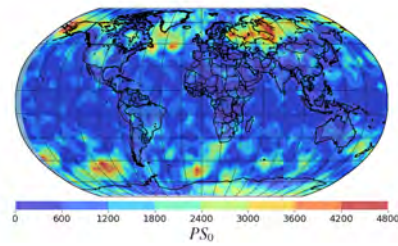
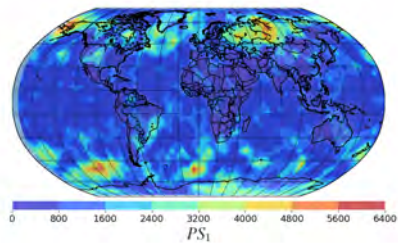
CUADRO 5.3: Nivel Superficie componente V del viento

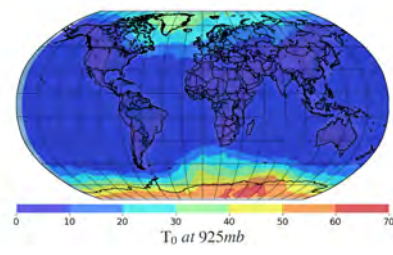
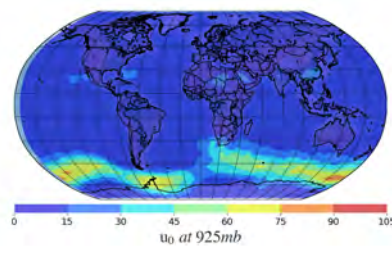
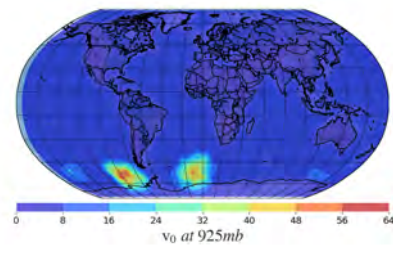
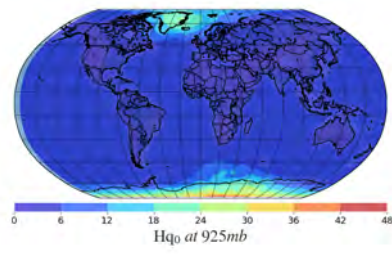
P	a	Metodo Asimilacion	Predicho	Observado
100	1.02	EnKF	302.862056	313.811636
		LETKF	309.329772	317.436827
		LEnKF	315.256790	321.393230
	1.04	EnKF	289.609751	306.982901
		LETKF	309.329772	317.436827
		LEnKF	315.256790	321.393230
	1.06	EnKF	276.837728	301.671983
		LETKF	309.329772	317.436827
		LEnKF	315.256790	321.393230
	1.08	EnKF	264.906123	297.692274
		LETKF	309.329772	317.436827
		LEnKF	315.256790	321.393230
11	1.02	EnKF	432.619426	433.107160
		LETKF	427.935129	428.639947
		LEnKF	430.500566	430.834159
	1.04	EnKF	435.843582	436.819741
		LETKF	427.935129	428.639947
		LEnKF	430.500566	430.834159
	1.06	EnKF	438.465748	439.927364
		LETKF	427.935129	428.639947
		LEnKF	430.500566	430.834159
	1.08	EnKF	446.990477	448.573132
		LETKF	427.935129	428.639947
		LEnKF	430.500566	430.834159
25	1.02	EnKF	418.726778	420.512689
		LETKF	418.501267	419.674031
		LEnKF	419.345013	420.276907
	1.04	EnKF	418.972719	421.868055
		LETKF	418.501267	419.674031
		LEnKF	419.345013	420.276907
	1.06	EnKF	422.881003	426.494175
		LETKF	418.501267	419.674031
		LEnKF	419.345013	420.276907
	1.08	EnKF	437.260484	439.967733
		LETKF	418.501267	419.674031
		LEnKF	419.345013	420.276907
4	1.02	EnKF	435.767361	436.113827
		LETKF	435.683831	436.022443
		LEnKF	435.353419	435.536566
	1.04	EnKF	436.176105	436.565191
		LETKF	435.683831	436.022443
		LEnKF	435.353419	435.536566
	1.06	EnKF	437.182807	437.708739
		LETKF	435.683831	436.022443
		LEnKF	435.353419	435.536566
	1.08	EnKF	438.976094	439.332201
		LETKF	435.683831	436.022443
		LEnKF	435.353419	435.536566
6	1.02	EnKF	437.457655	437.763446
		LETKF	435.271672	435.525011
		LEnKF	436.463234	436.648003
	1.04	EnKF	438.057032	438.513012
		LETKF	435.271672	435.525011
		LEnKF	436.463234	436.648003
	1.06	EnKF	438.157412	438.846244
		LETKF	0.000000	0.000000
		LEnKF	436.463234	436.648003
	1.08	EnKF	439.814167	440.087956
		LETKF	435.271672	435.525011
		LEnKF	436.463234	436.648003

CUADRO 5.4: Nivel Superficie componente U del viento

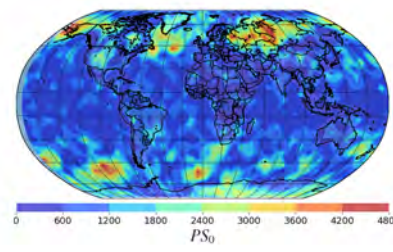
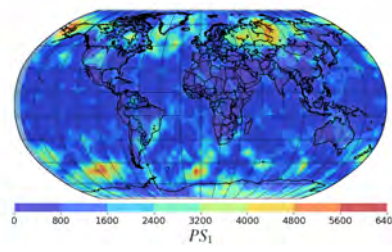
## 5.1. Ensemble Kalman Filter basado en la descomposición de Cholesky

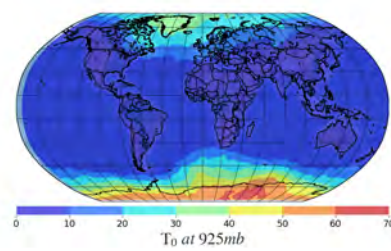
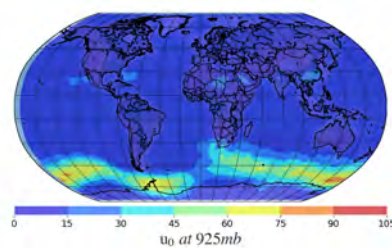
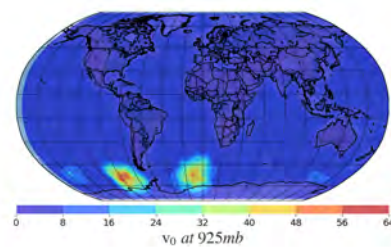
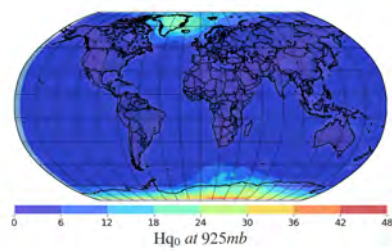
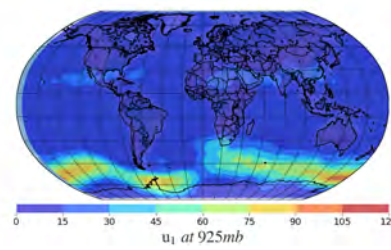
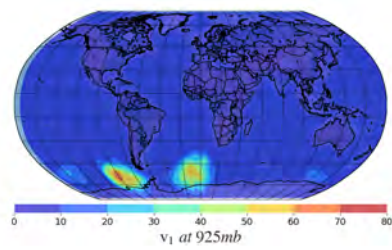
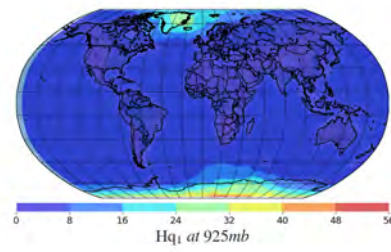
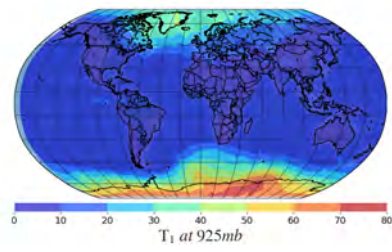
### 5.1.1. EnKF $p = 100\%$ $\alpha = 1,02$



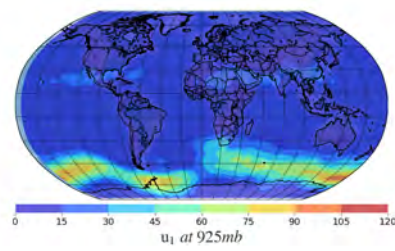
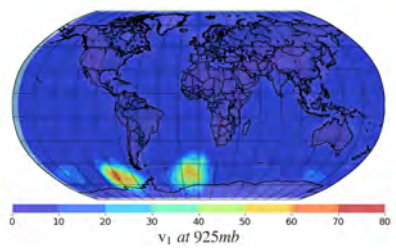
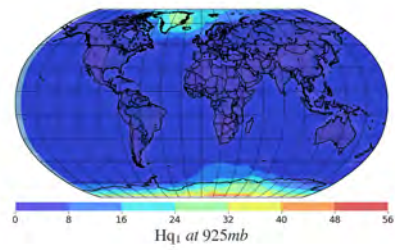
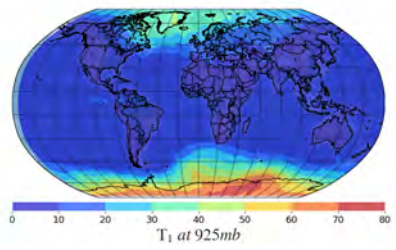
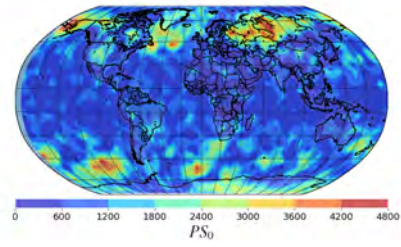
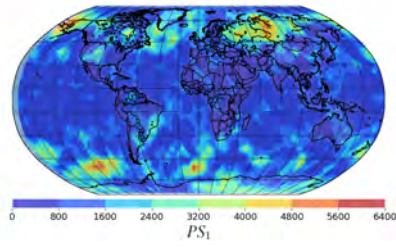


### 5.1.2. EnKF $p = 100\%$ $\alpha = 1,04$

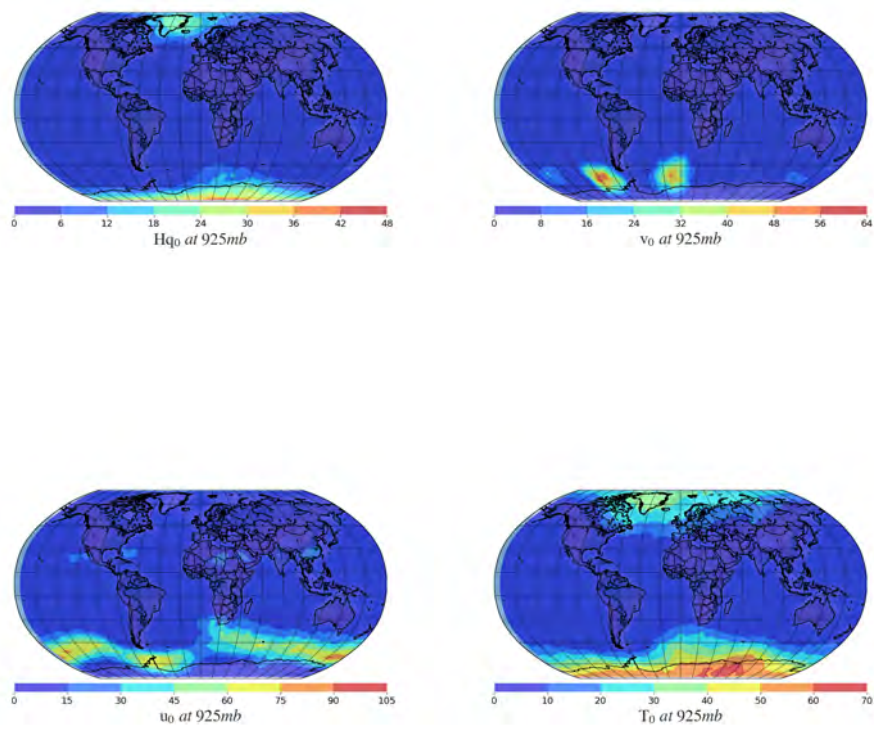




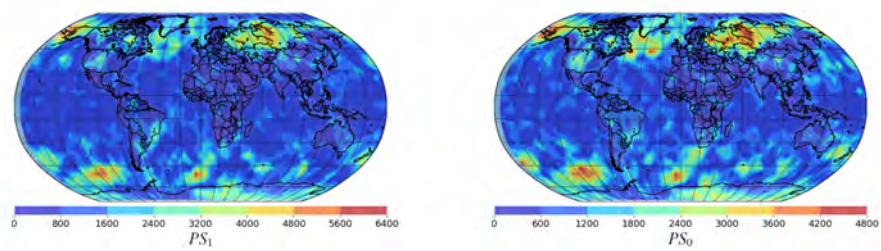


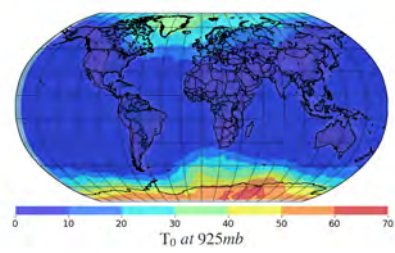
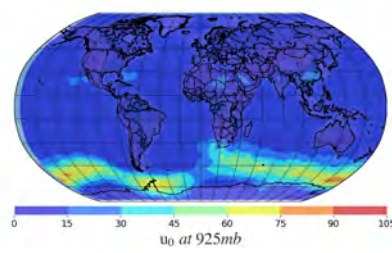
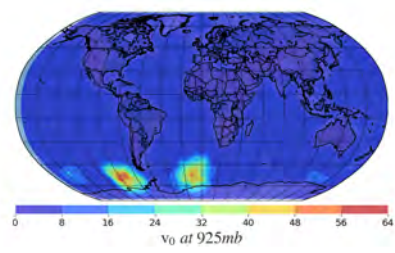
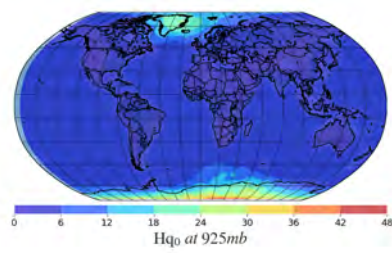
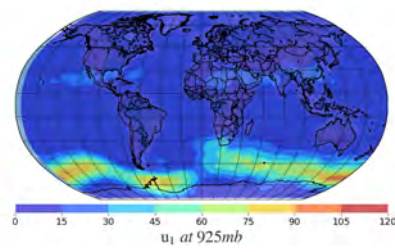
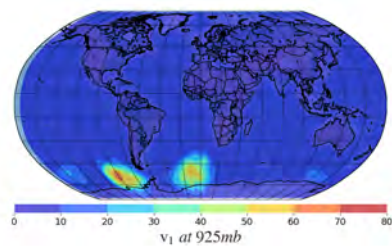
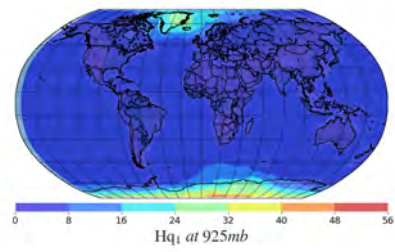
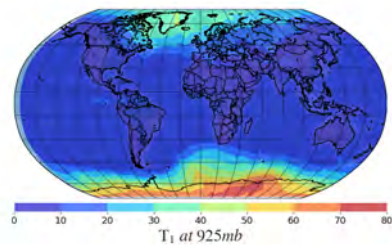
**5.1.3. EnKF  $p = 100\%$   $\alpha = 1,06$** 

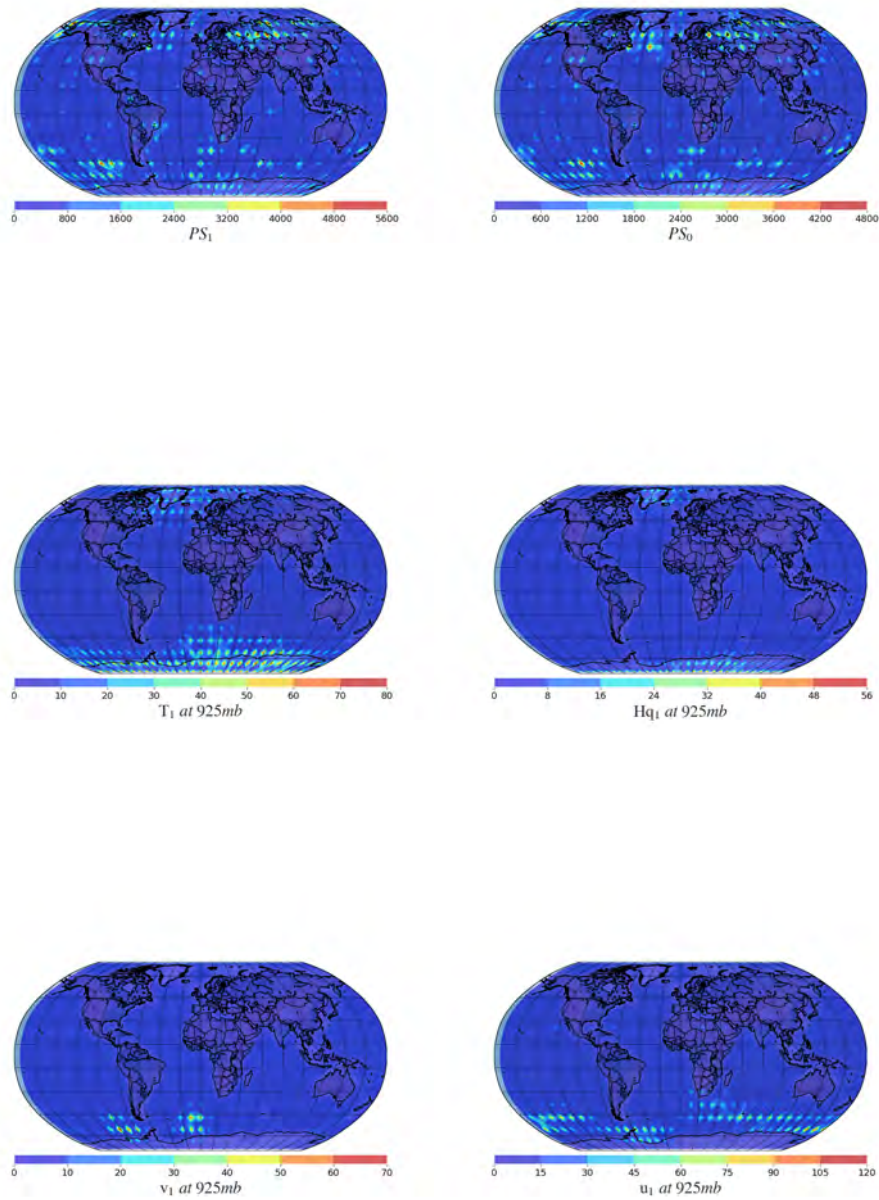


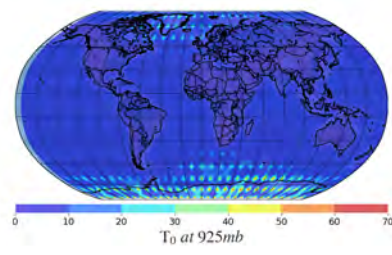
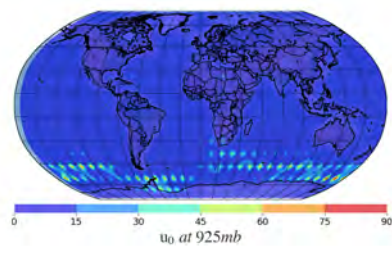
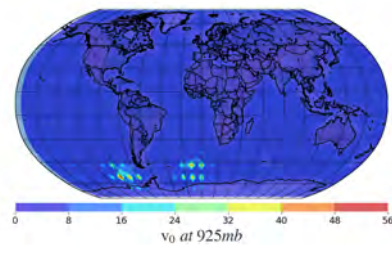
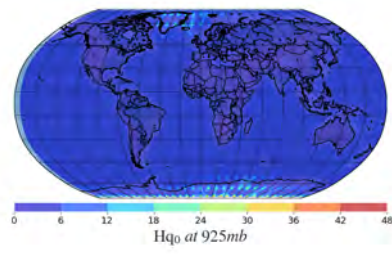


#### 5.1.4. EnKF $p = 100\%$ $\alpha = 1,08$

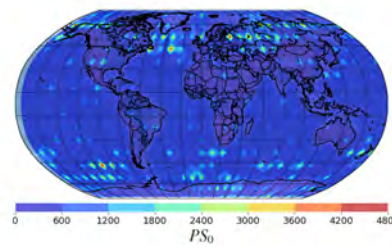
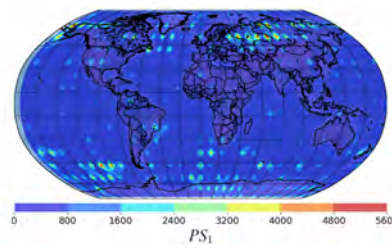




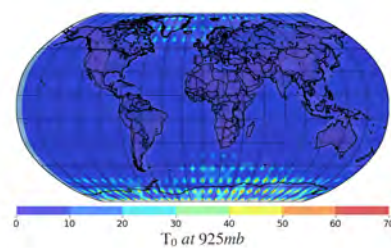
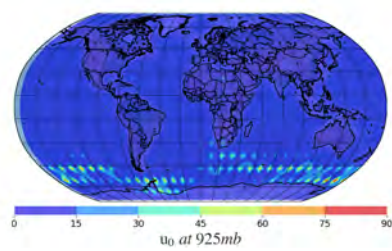
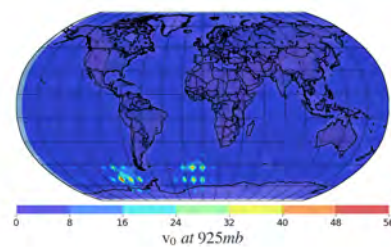
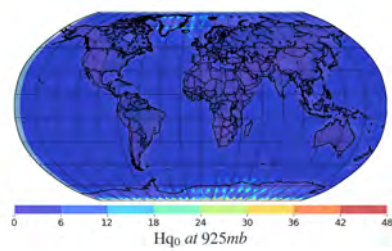
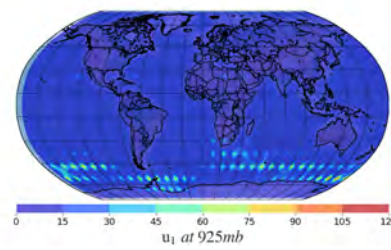
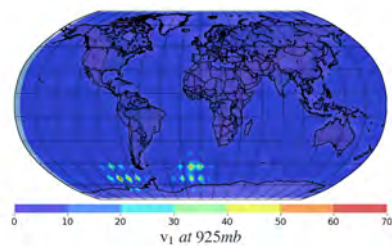
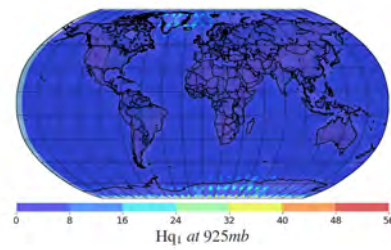
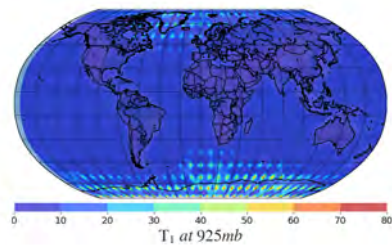
5.1.5. EnKF  $p = 25\%$   $\alpha = 1,02$ 

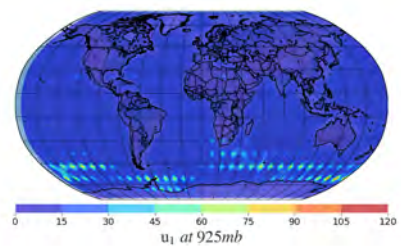
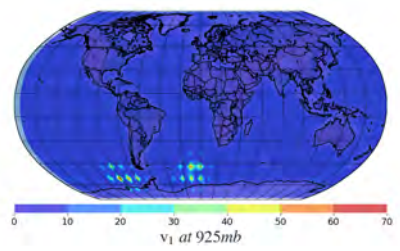
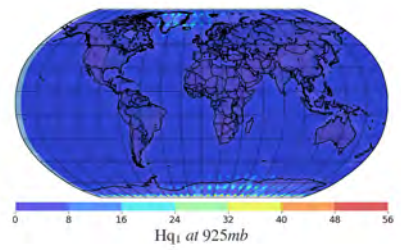
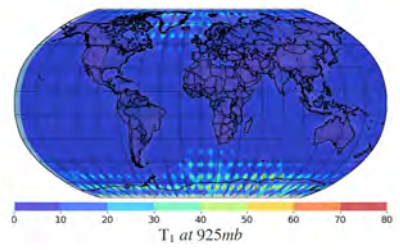
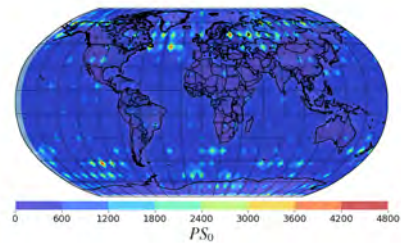
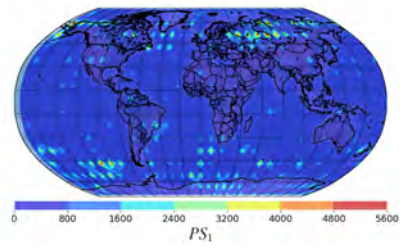


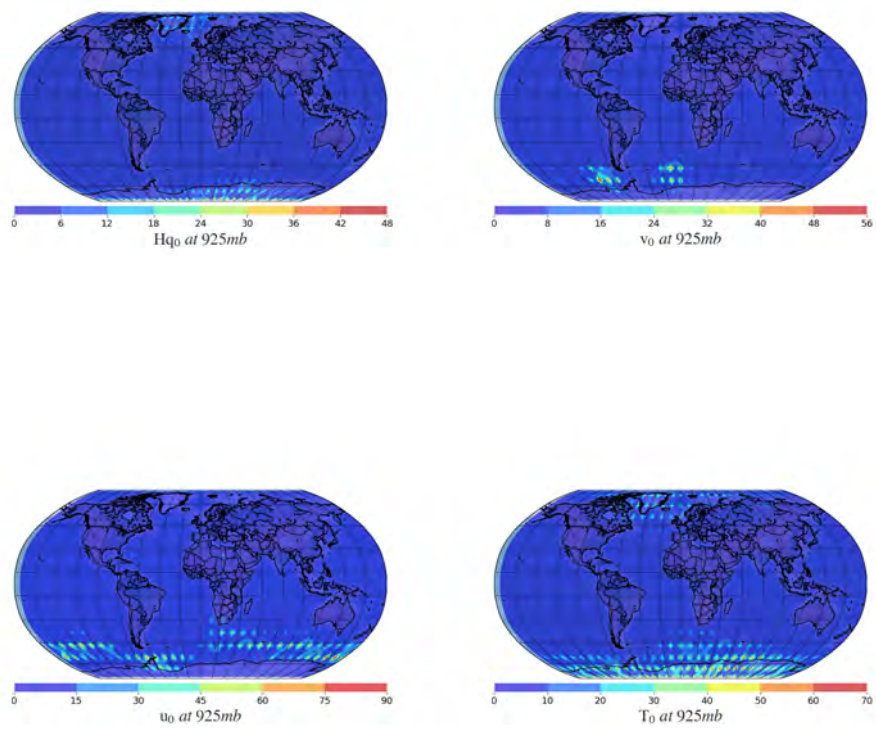
### 5.1.6. EnKF $p = 25\%$ $\alpha = 1,04$



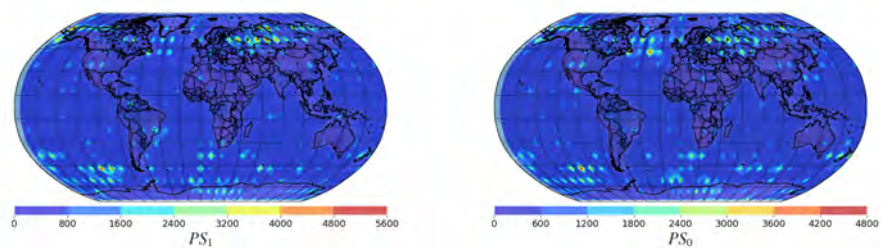




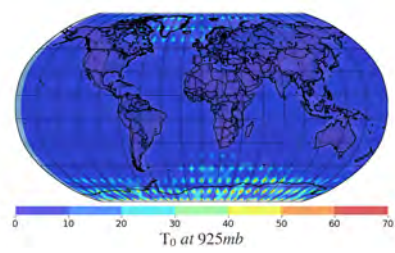
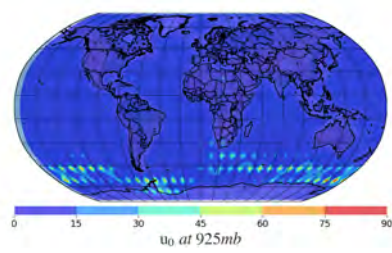
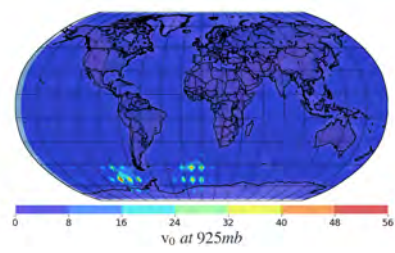
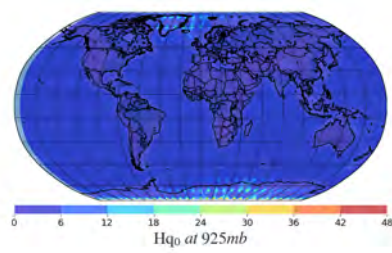
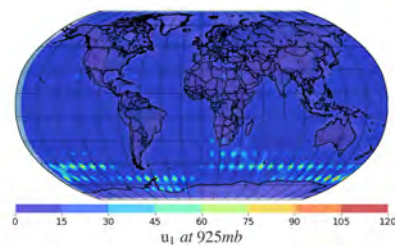
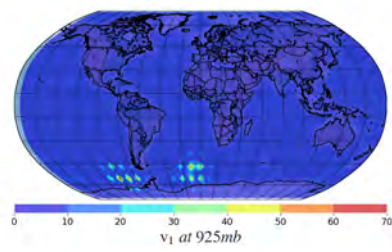
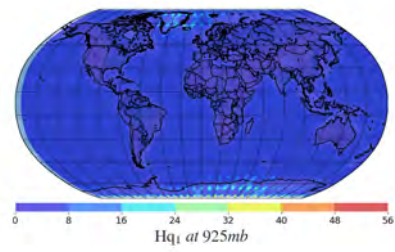
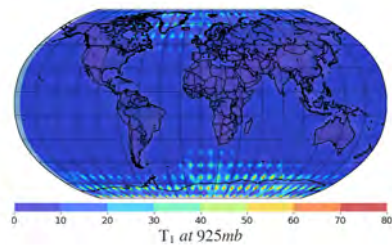
5.1.7. EnKF  $p = 25\%$   $\alpha = 1,06$ 



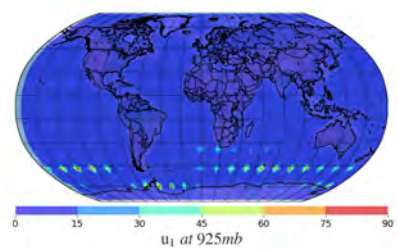
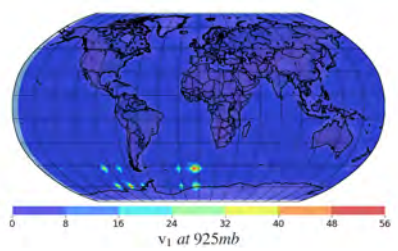
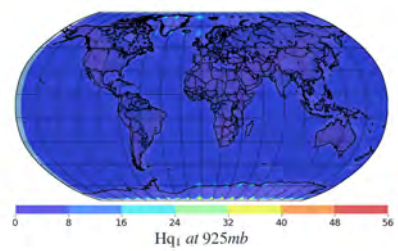
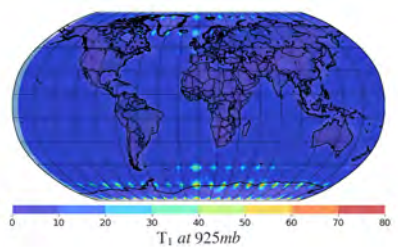
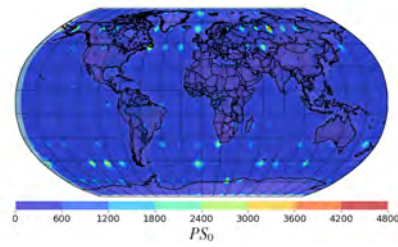
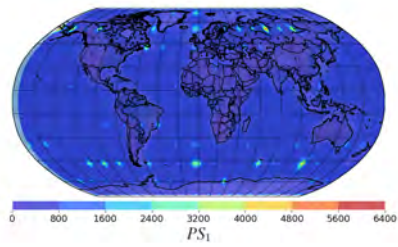
### 5.1.8. EnKF $p = 25\%$ $\alpha = 1,08$

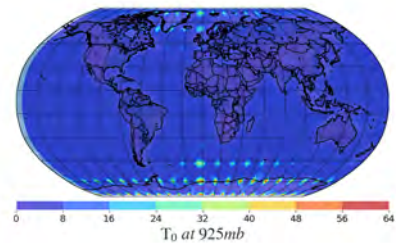
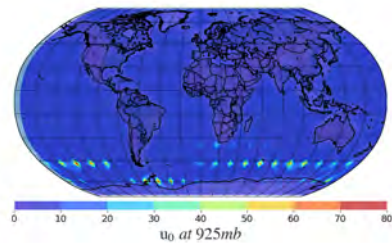
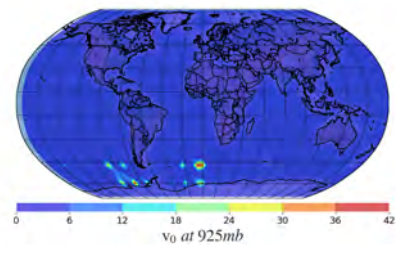
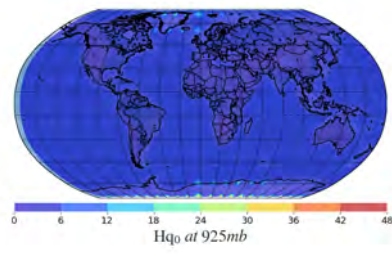




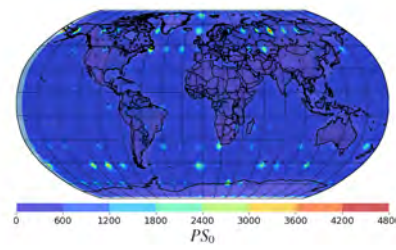
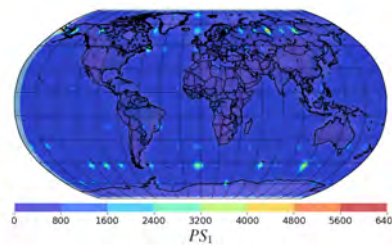


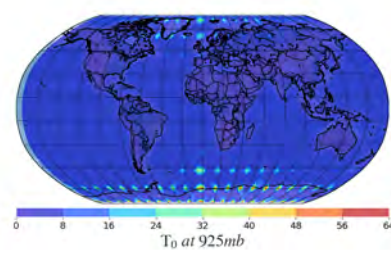
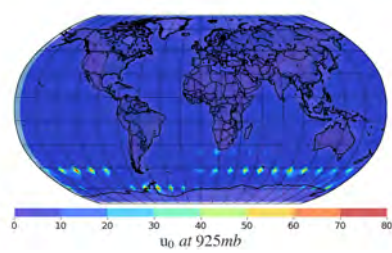
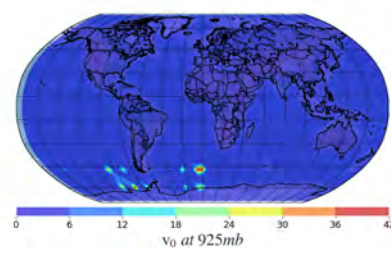
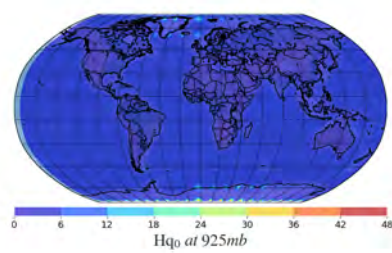
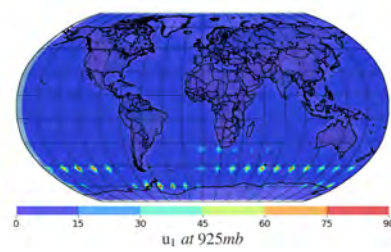
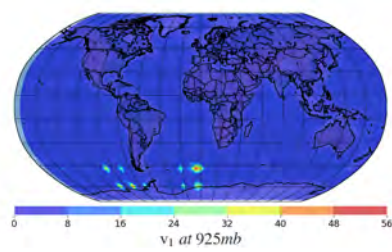
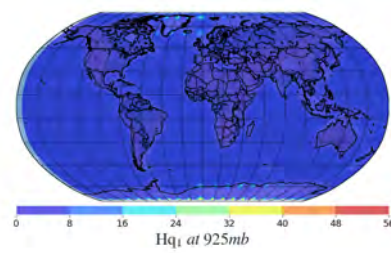
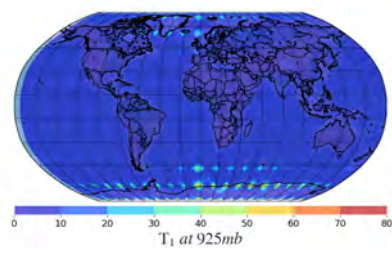


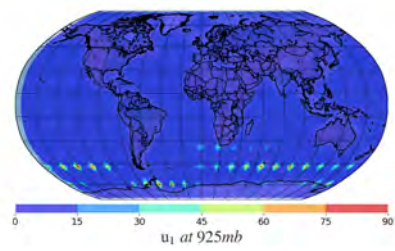
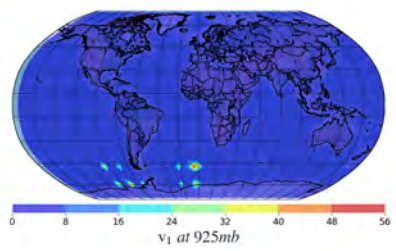
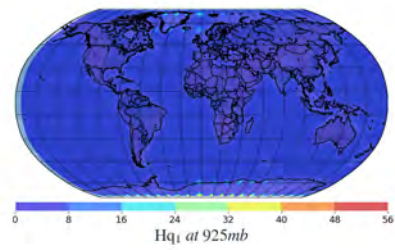
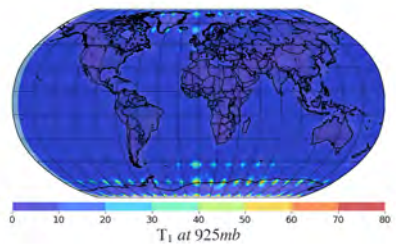
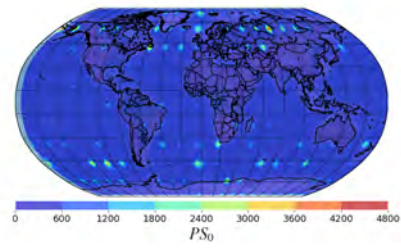
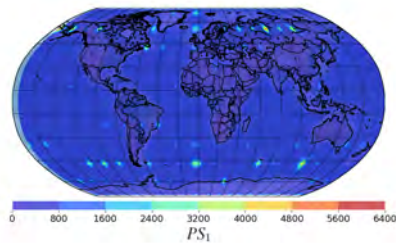
5.1.9. EnKF  $p = 11\%$   $\alpha = 1,02$ 



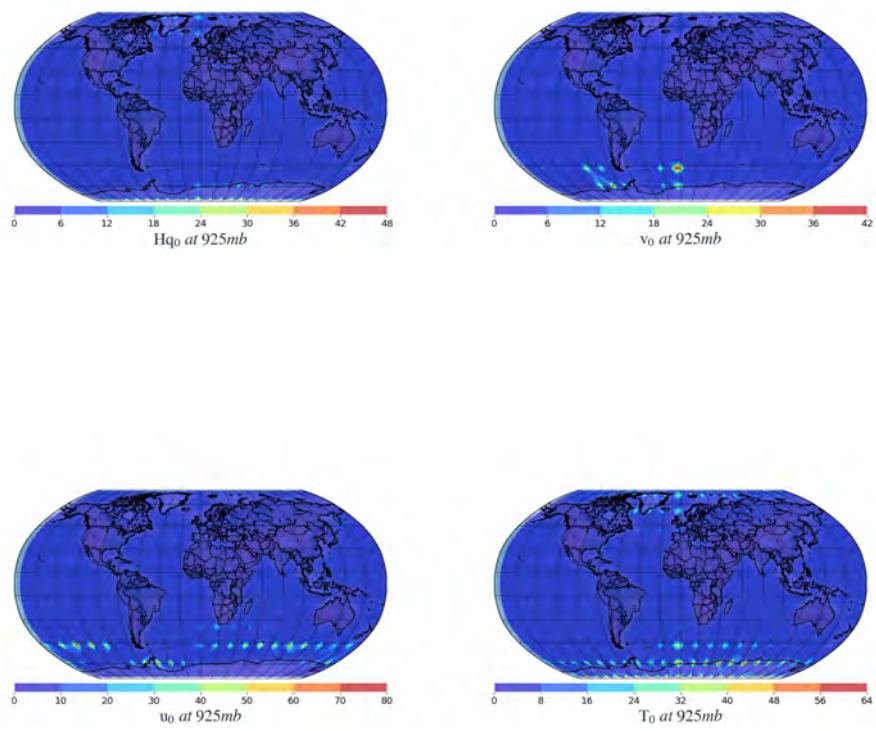
### 5.1.10. EnKF $p = 11\%$ $\alpha = 1,04$



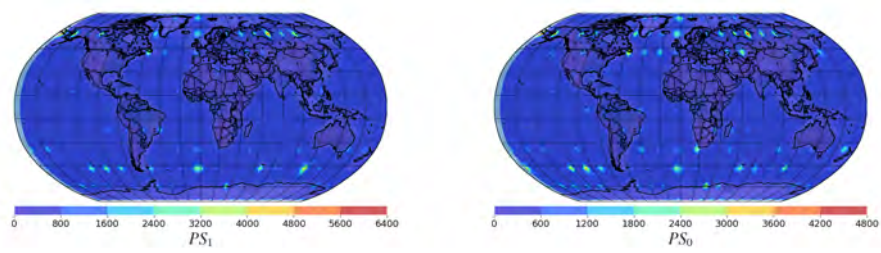


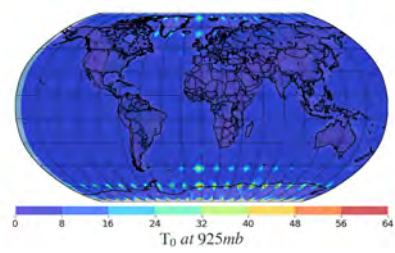
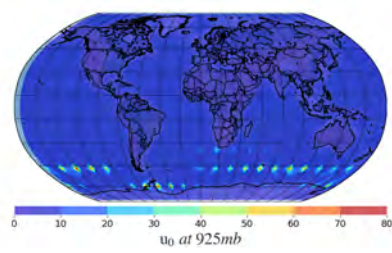
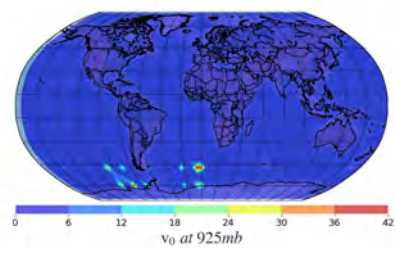
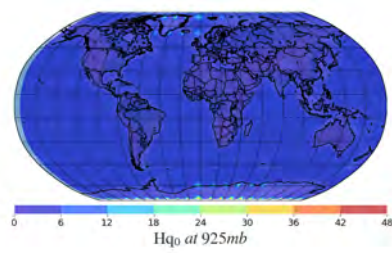
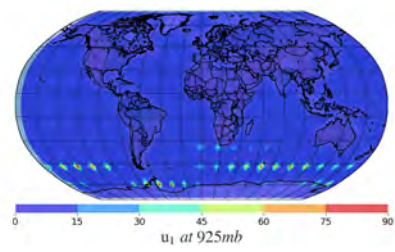
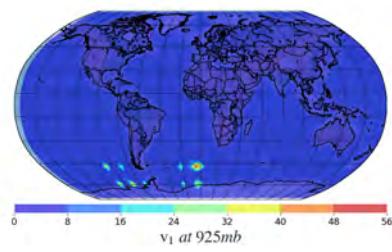
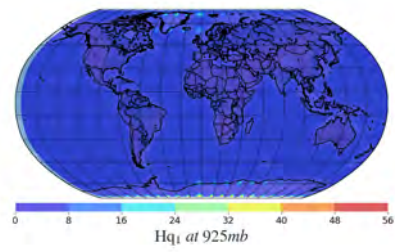
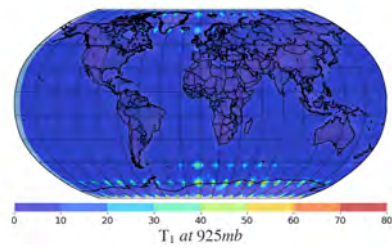
5.1.11. EnKF  $p = 11\%$   $\alpha = 1,06$ 

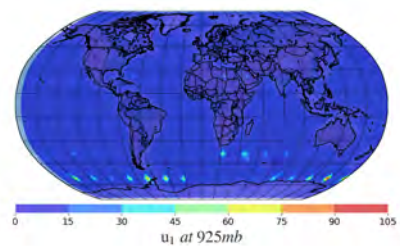
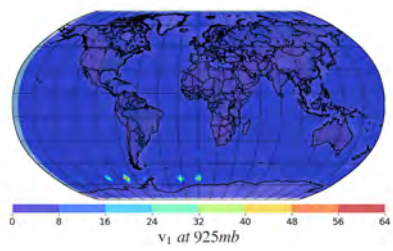
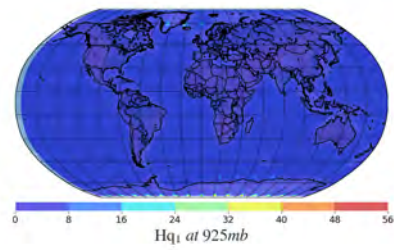
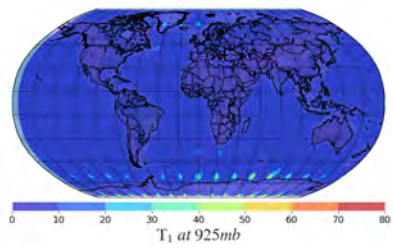
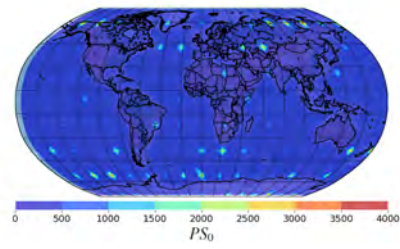
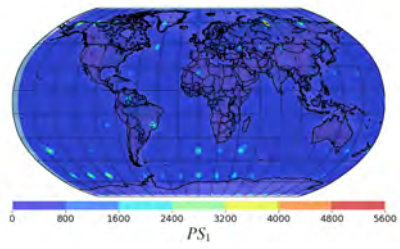


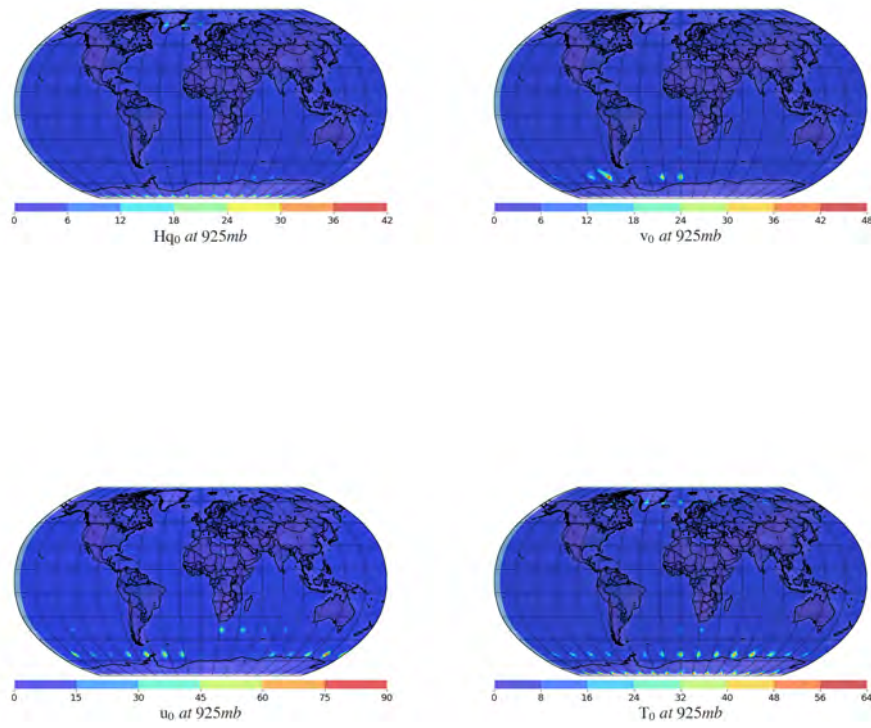


### 5.1.12. EnKF $p = 11\%$ $\alpha = 1,08$

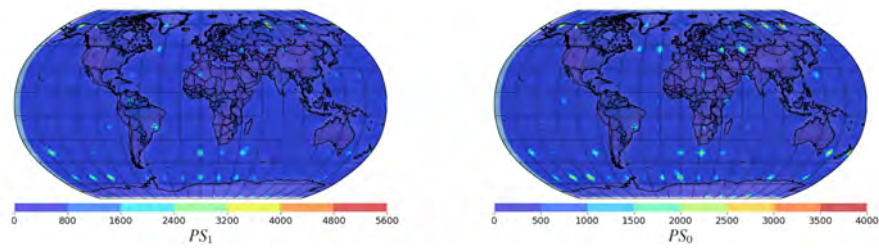




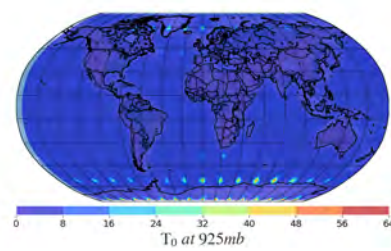
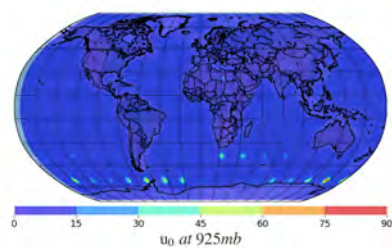
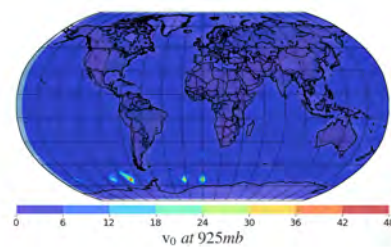
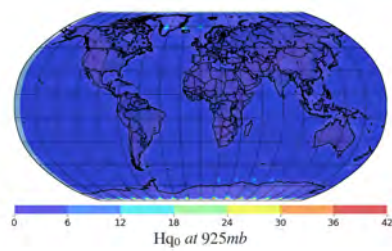
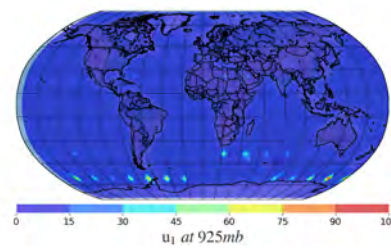
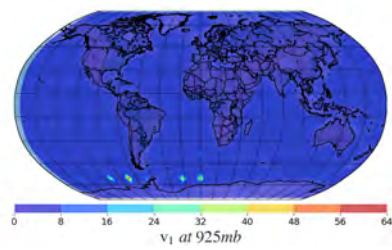
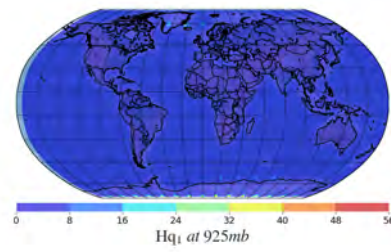
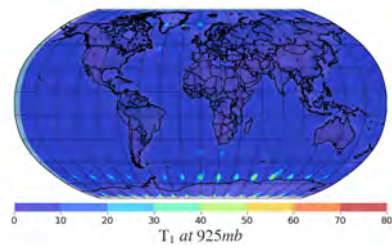
5.1.13. EnKF  $p = 6\%$   $\alpha = 1,02$ 

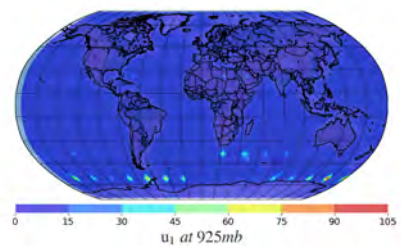
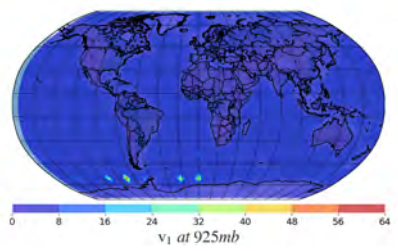
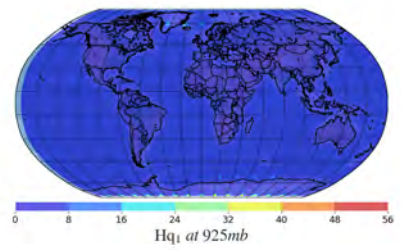
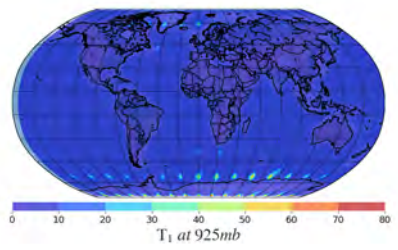
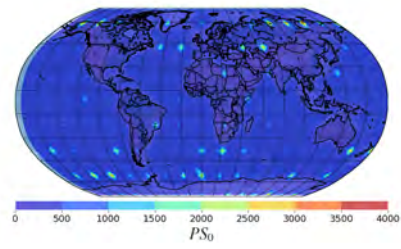
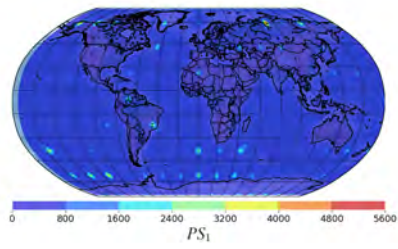


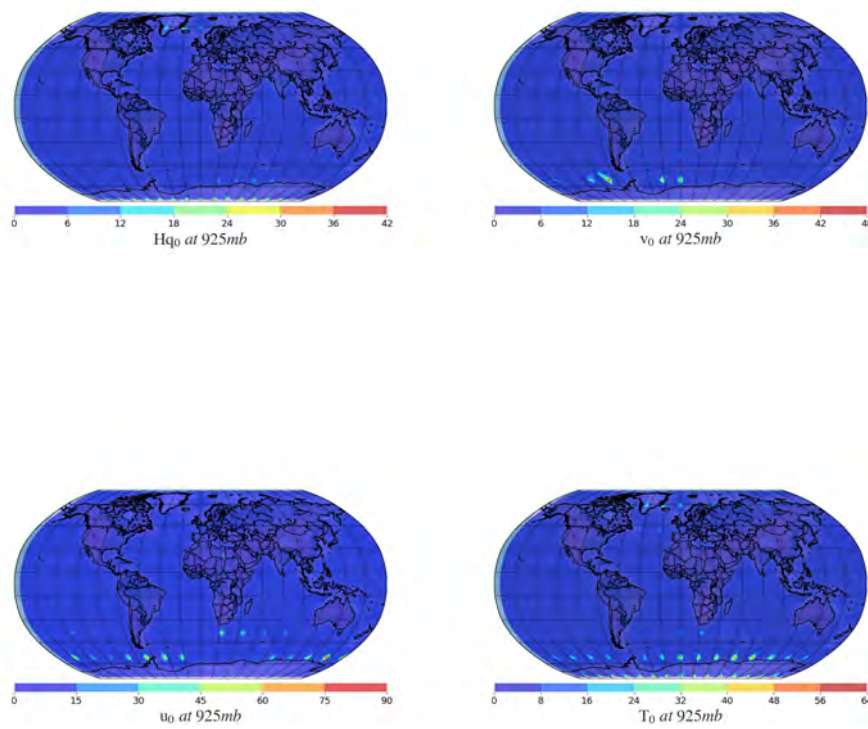
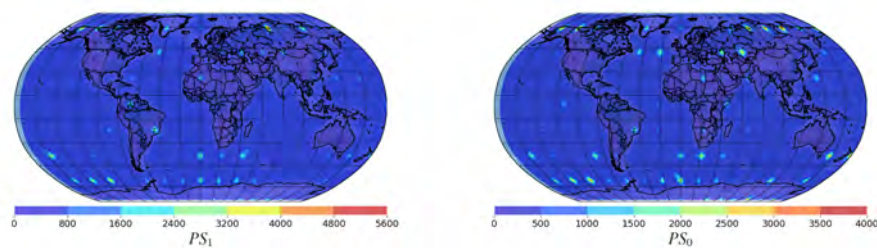
5.1.14. EnKF  $p = 6\%$   $\alpha = 1,04$

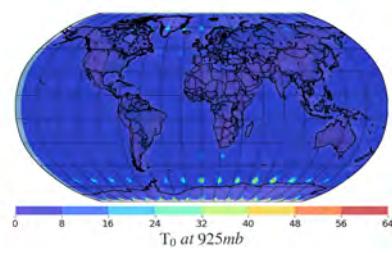
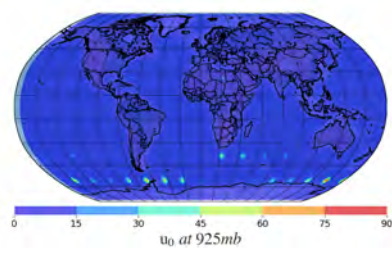
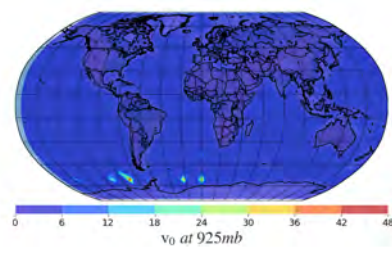
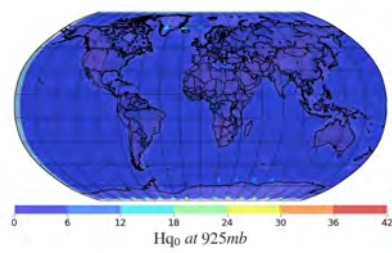
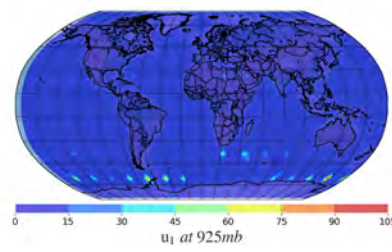
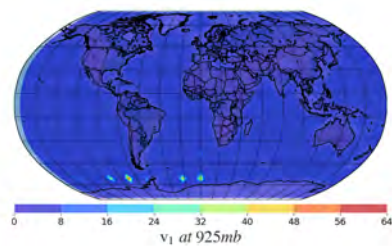
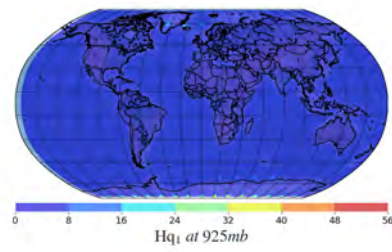
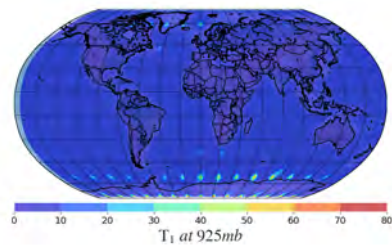




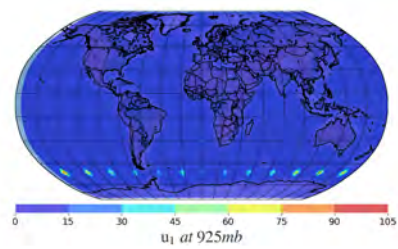
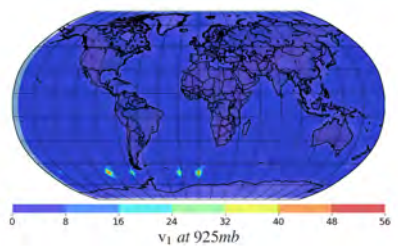
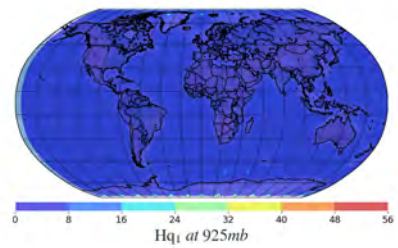
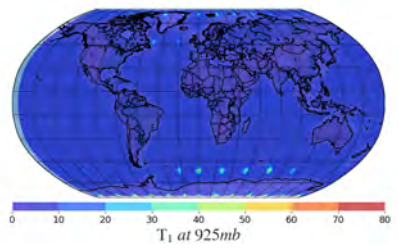
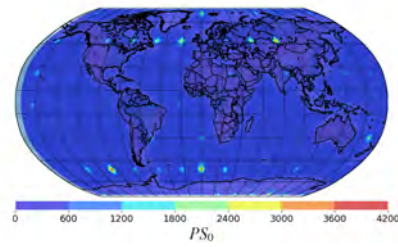
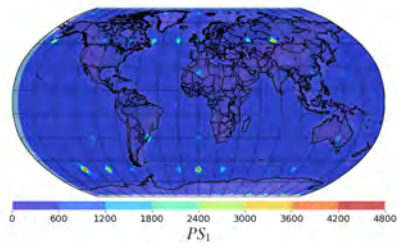


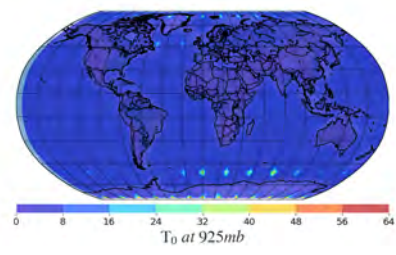
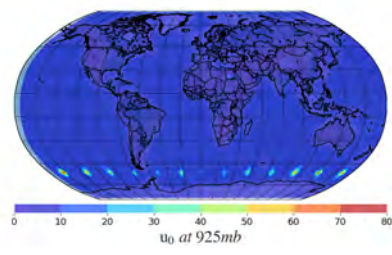
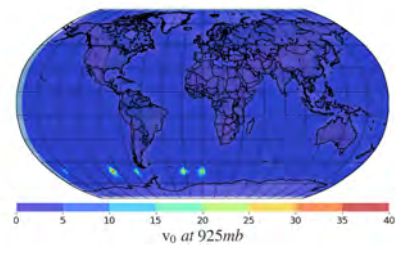
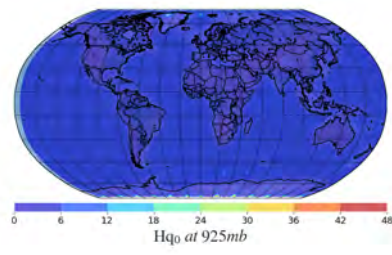
5.1.15. EnKF  $p = 6\%$   $\alpha = 1,06$ 

**5.1.16. EnKF  $p = 6\%$   $\alpha = 1,08$** 

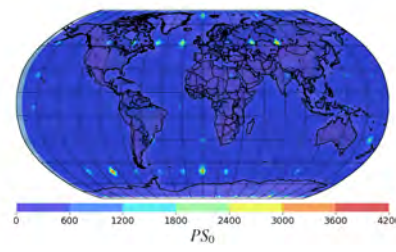
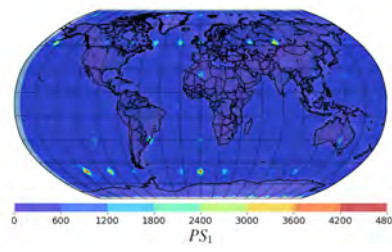


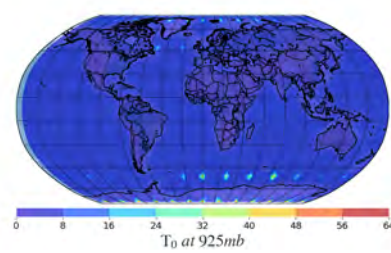
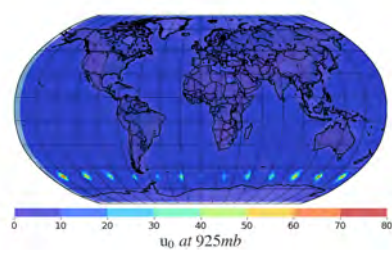
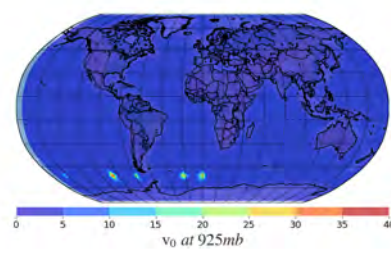
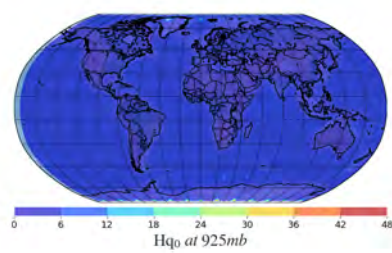
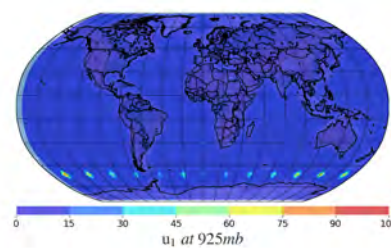
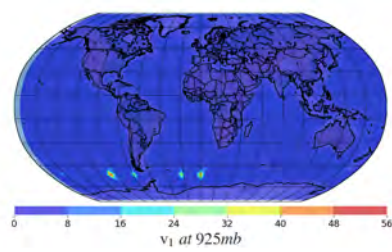
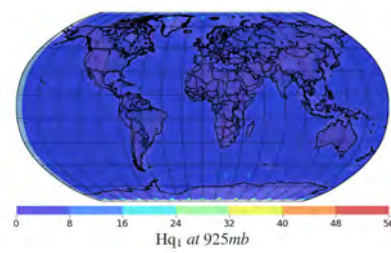
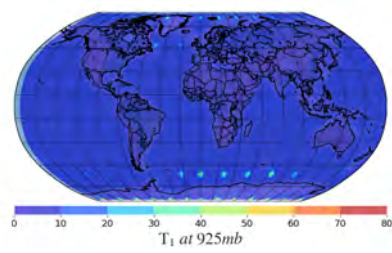


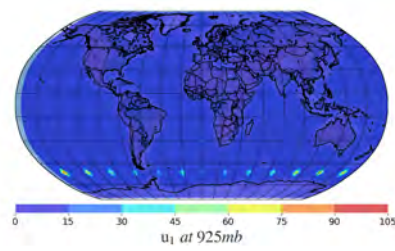
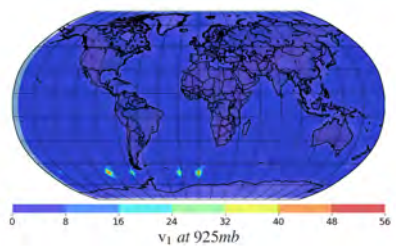
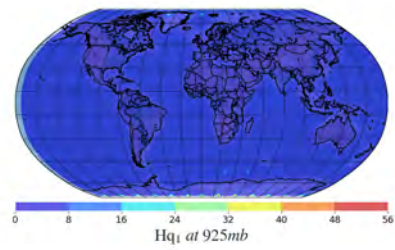
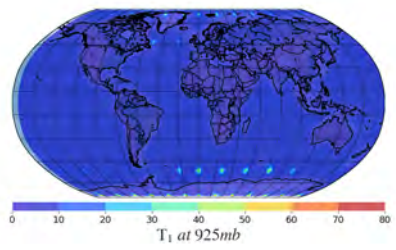
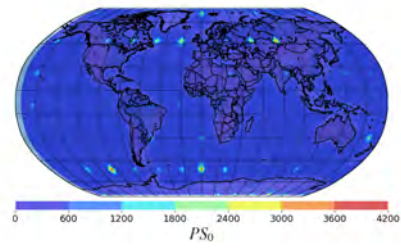
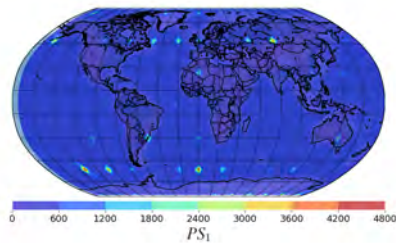
5.1.17. EnKF  $p = 4\%$   $\alpha = 1,02$ 



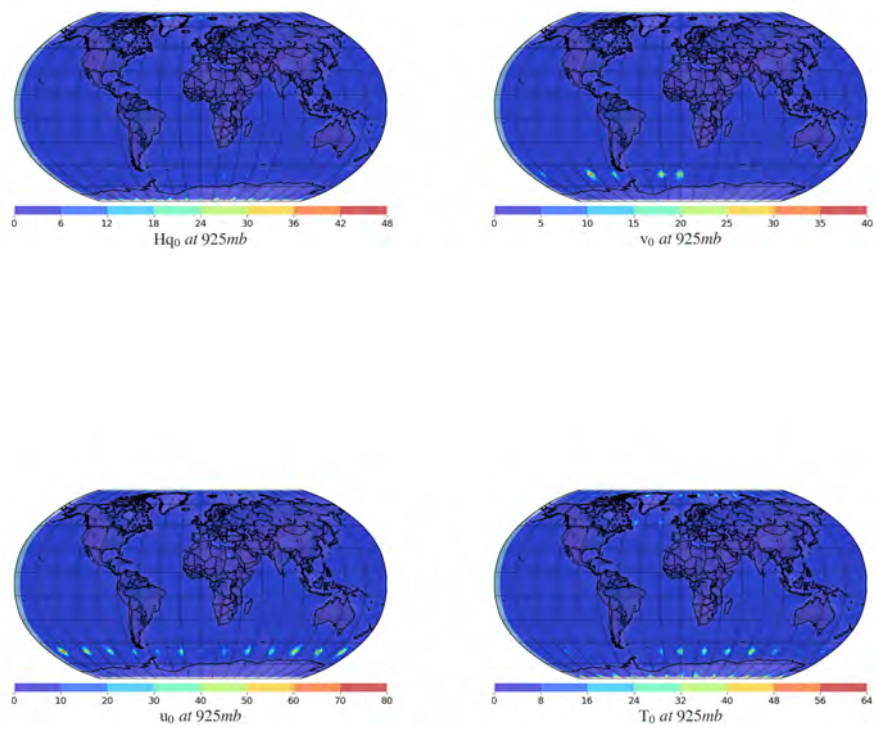
5.1.18. EnKF  $p = 4\%$   $\alpha = 1,04$



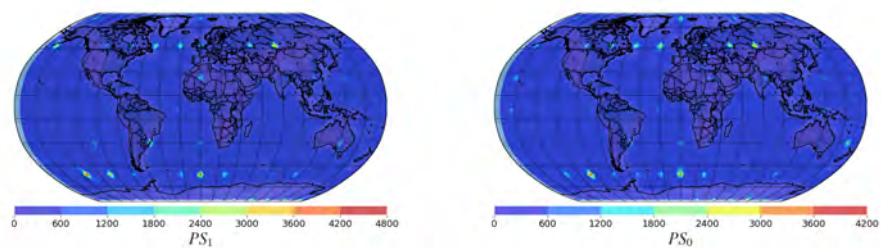


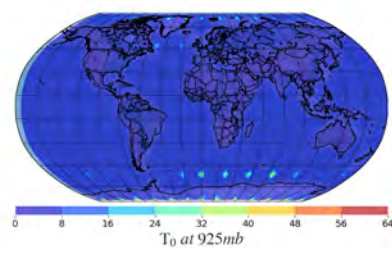
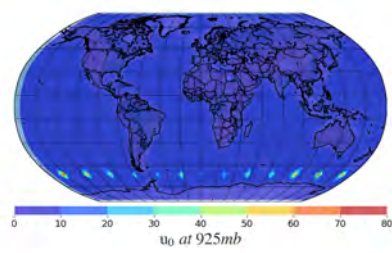
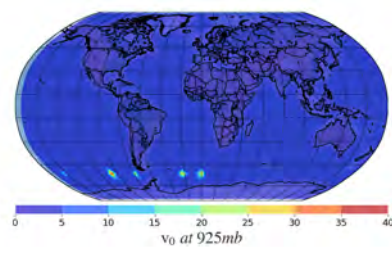
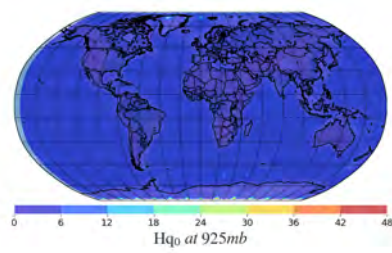
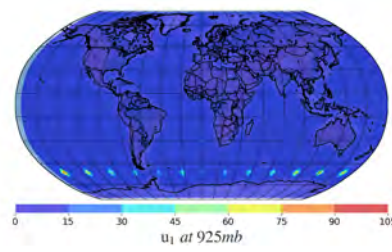
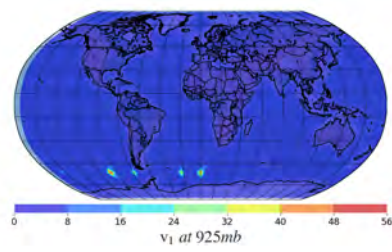
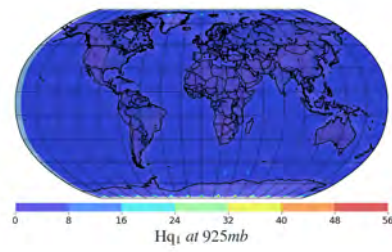
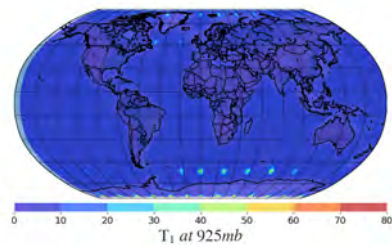
5.1.19. EnKF  $p = 4\%$   $\alpha = 1,06$ 





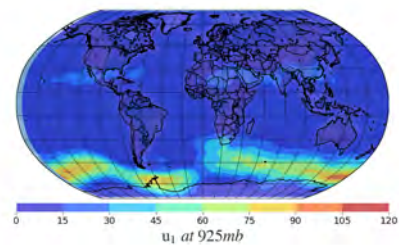
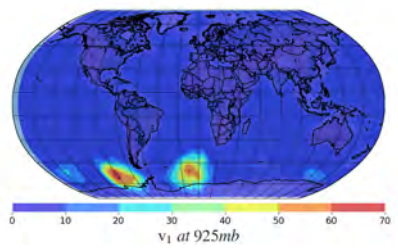
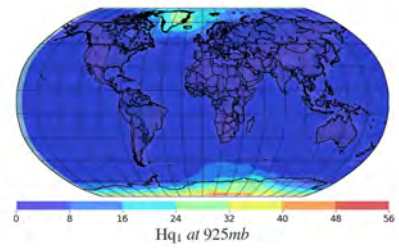
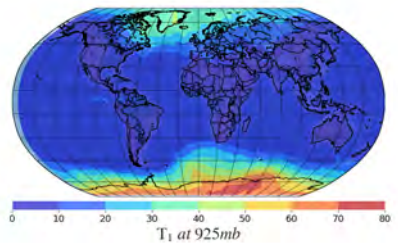
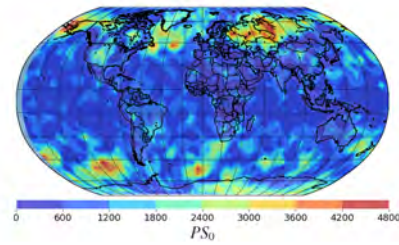
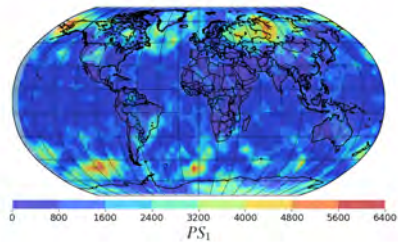
### 5.1.20. EnKF $p = 4\%$ $\alpha = 1,08$

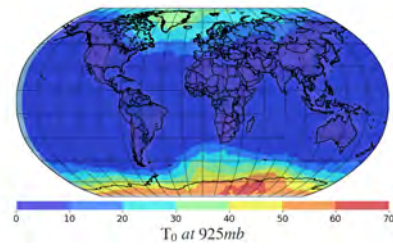
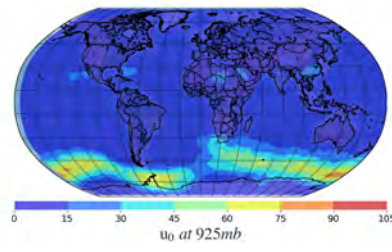
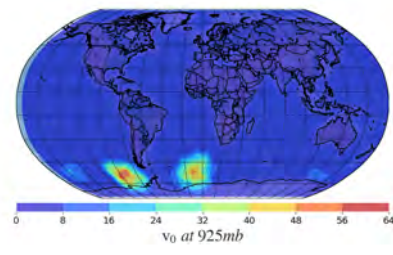
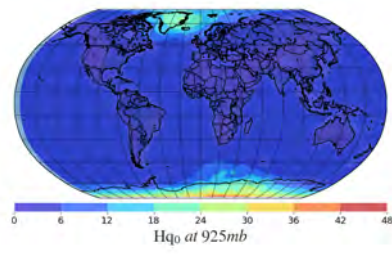




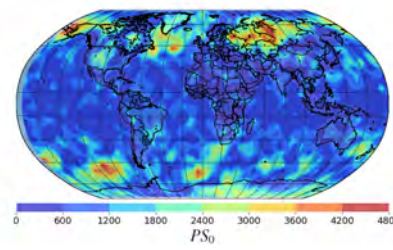
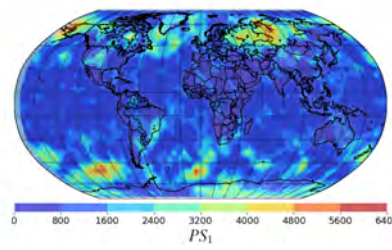
## 5.2. Local Ensemble Transform Kalman Filter

### 5.2.1. LETKF $p = 100\%$ $\alpha = 1,02$

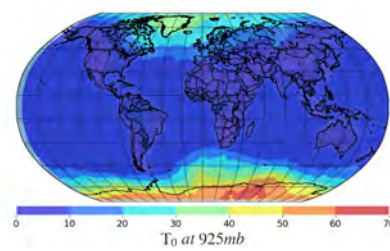
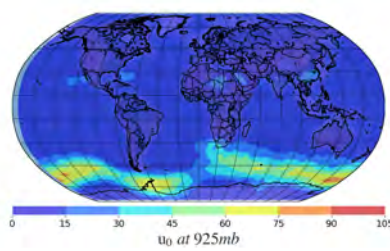
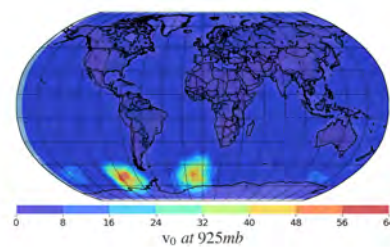
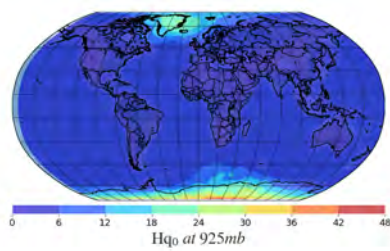
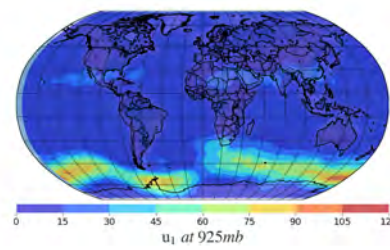
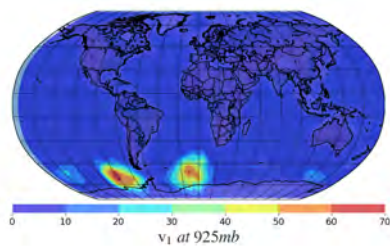
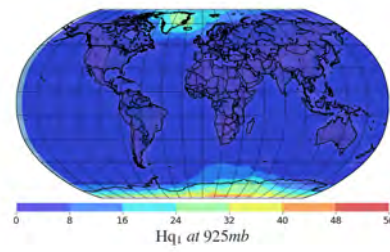
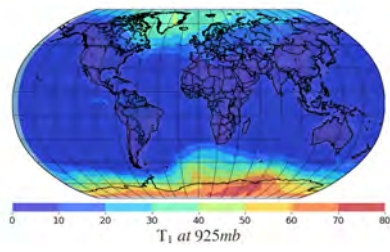


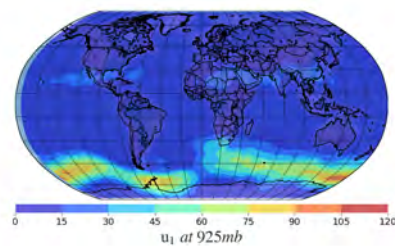
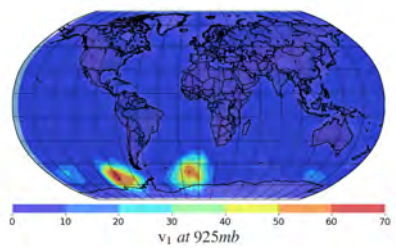
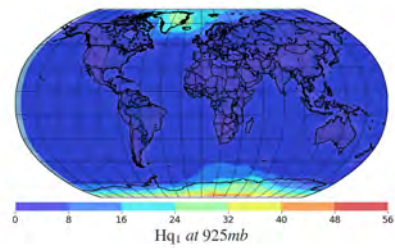
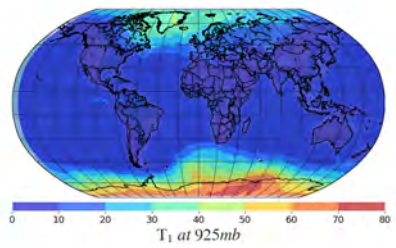
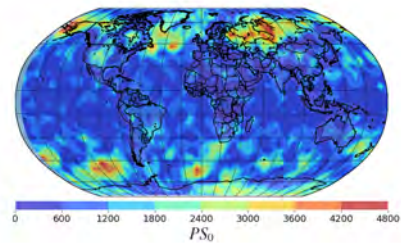
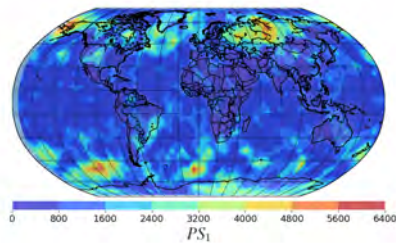


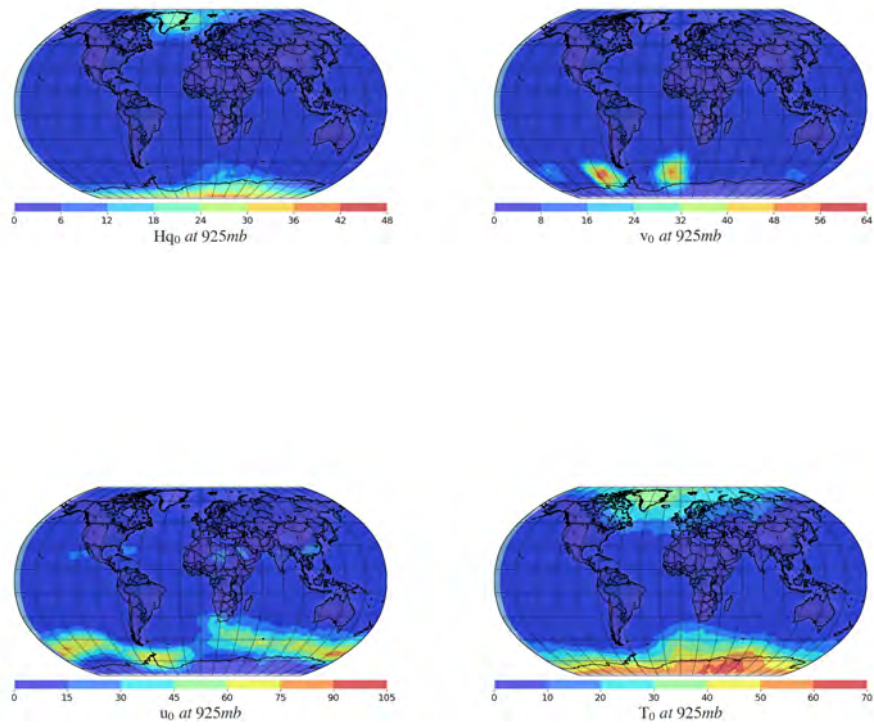
### 5.2.2. LETKF $p = 100\%$ $\alpha = 1,04$



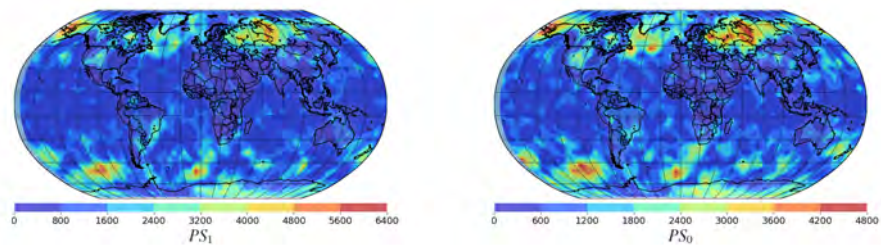




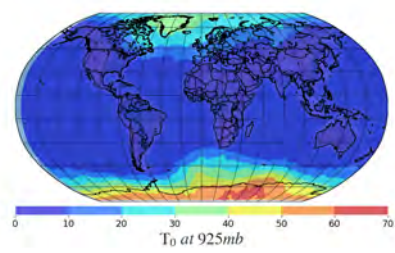
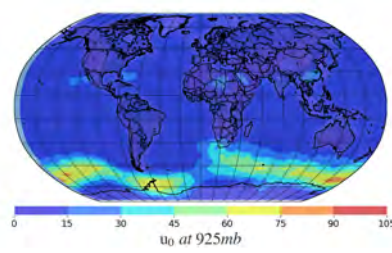
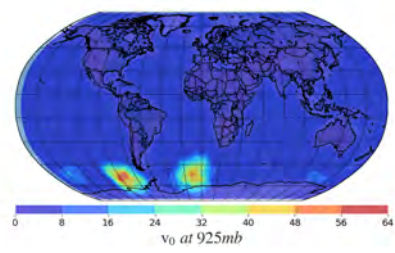
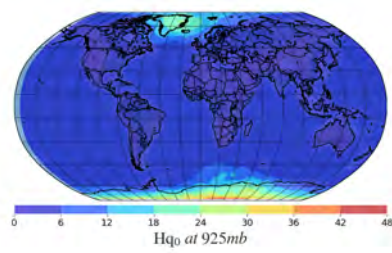
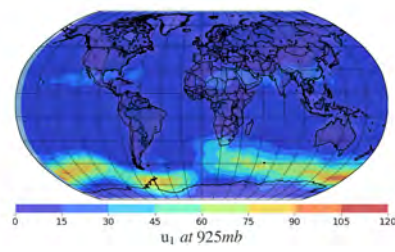
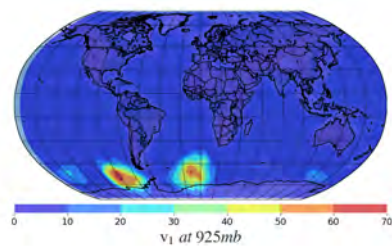
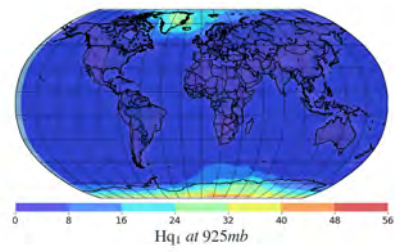
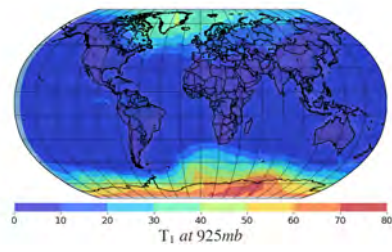
5.2.3. LETKF  $p = 100\%$   $\alpha = 1,06$ 



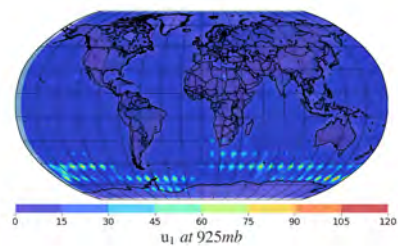
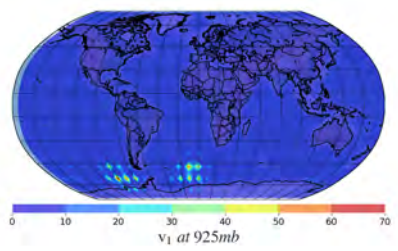
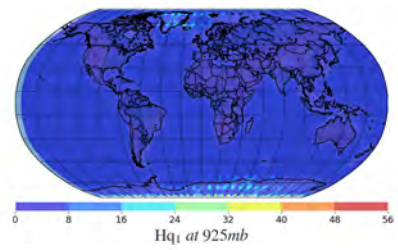
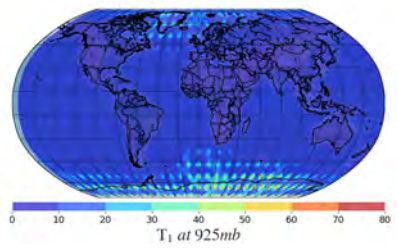
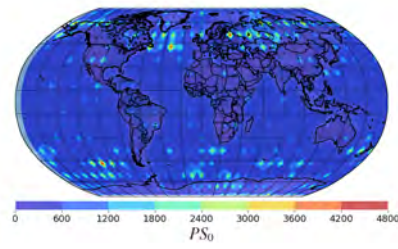
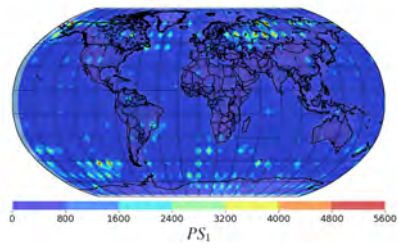
#### 5.2.4. LETKF $p = 100\%$ $\alpha = 1,08$

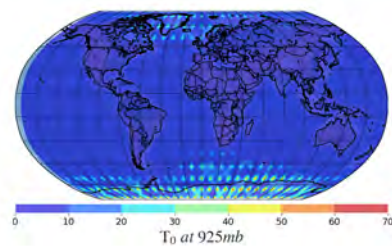
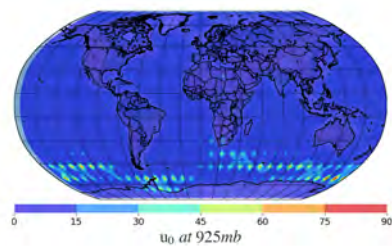
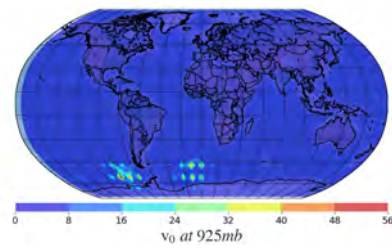
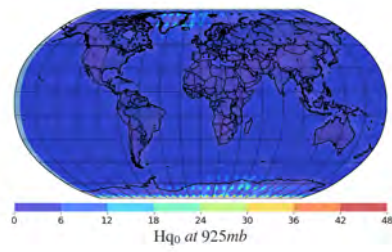




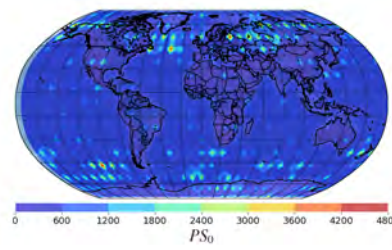
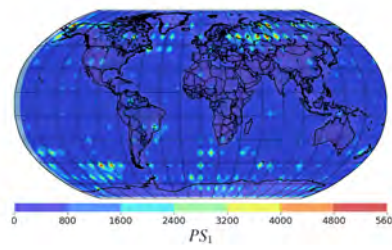


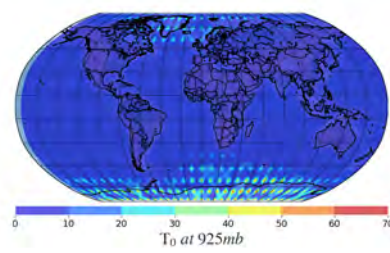
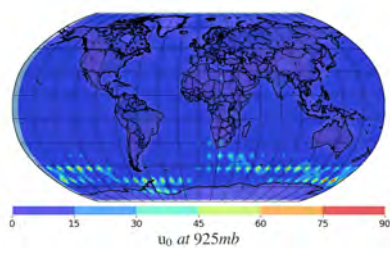
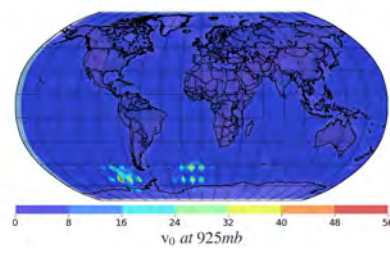
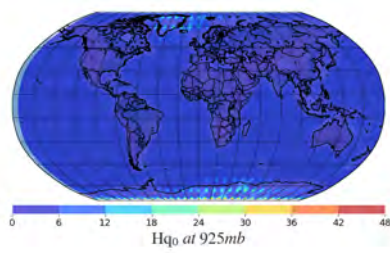
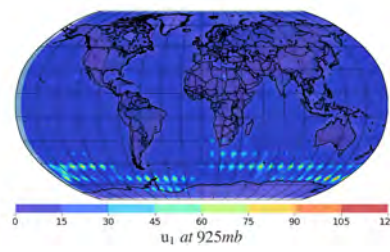
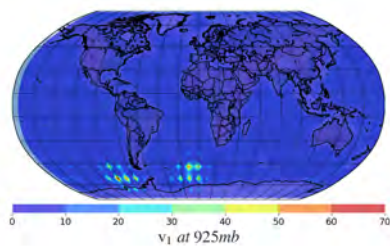
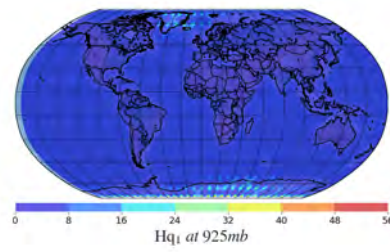
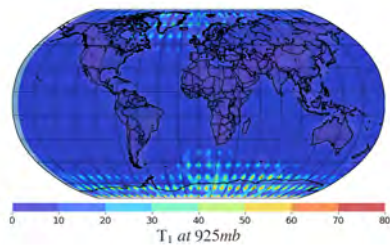


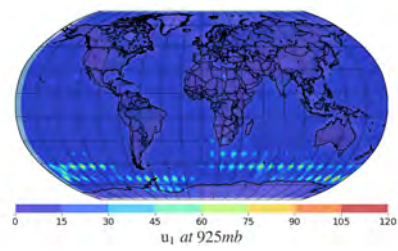
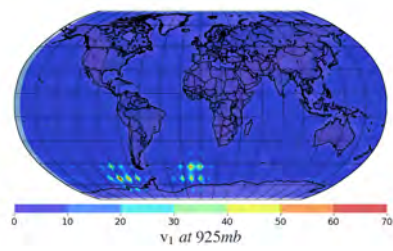
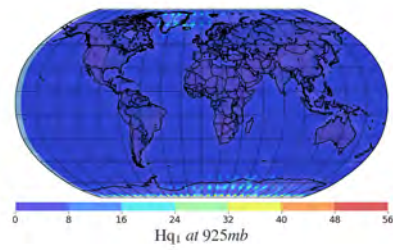
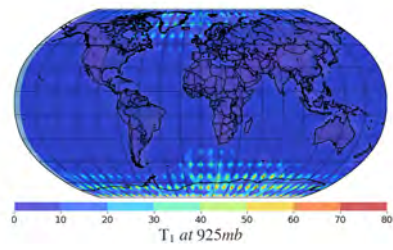
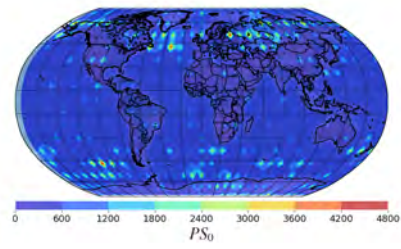
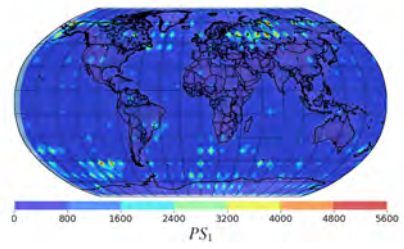
5.2.5. LETKF  $p = 25\%$   $\alpha = 1,02$ 



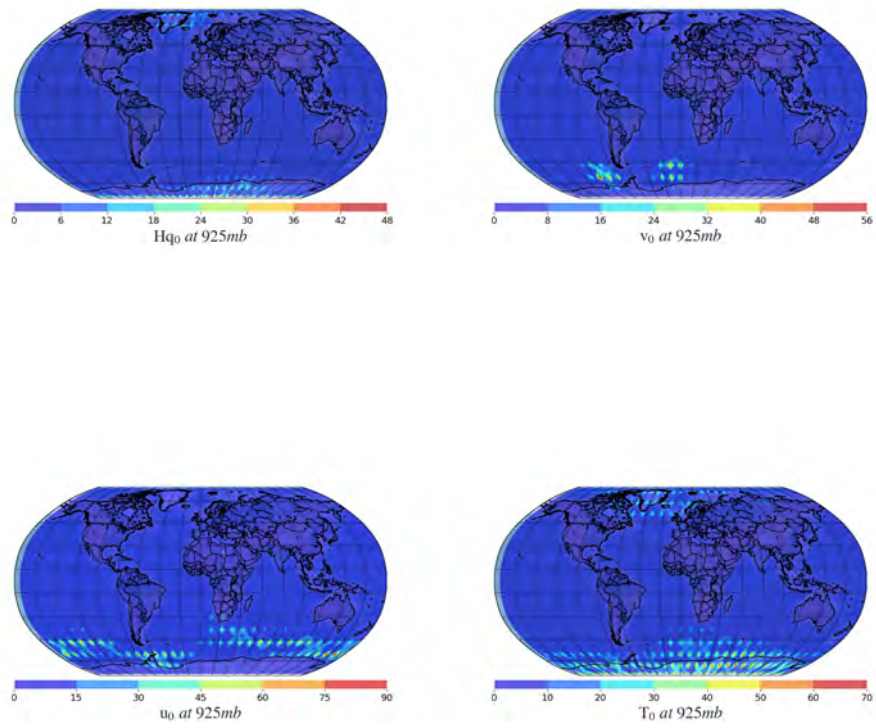
### 5.2.6. LETKF $p = 25\%$ $\alpha = 1,04$



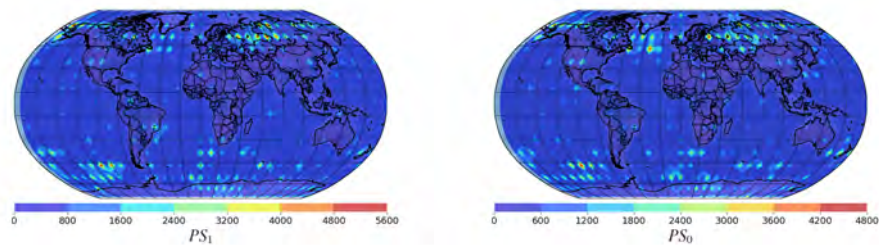


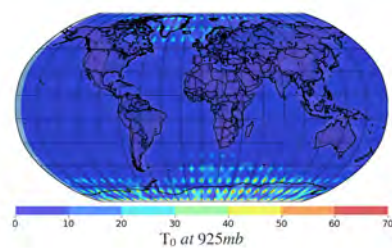
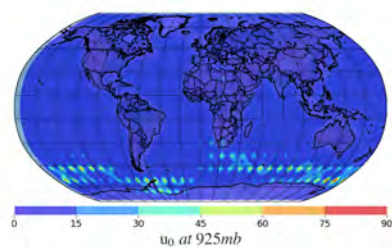
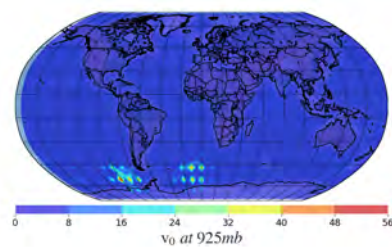
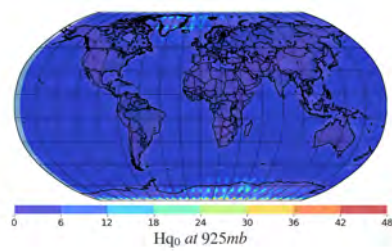
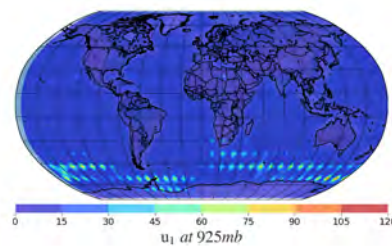
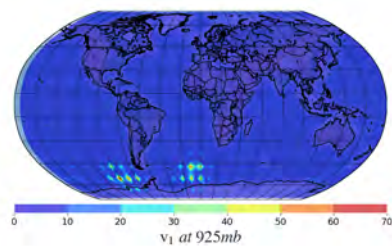
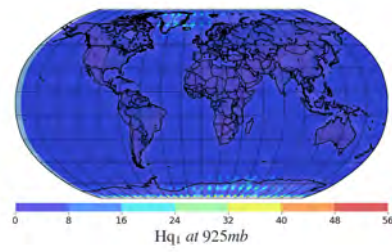
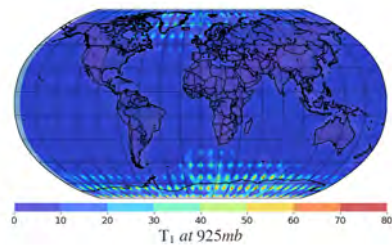
5.2.7. LETKF  $p = 25\%$   $\alpha = 1,06$ 

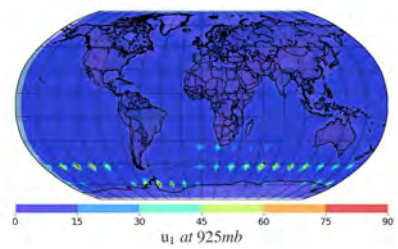
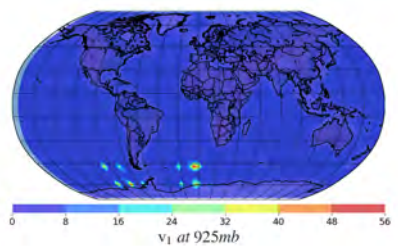
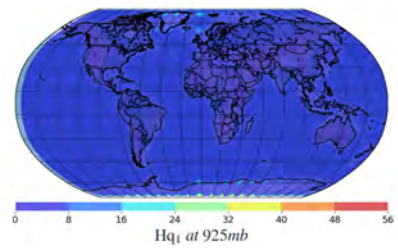
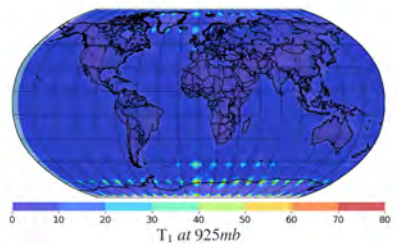
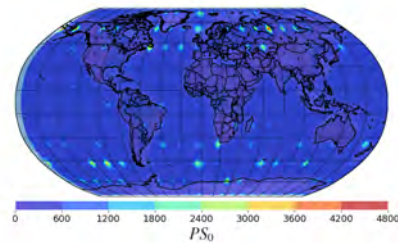
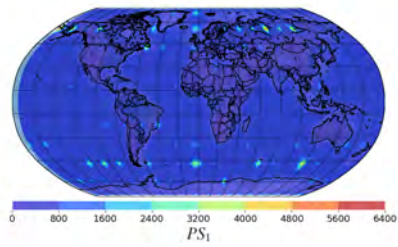


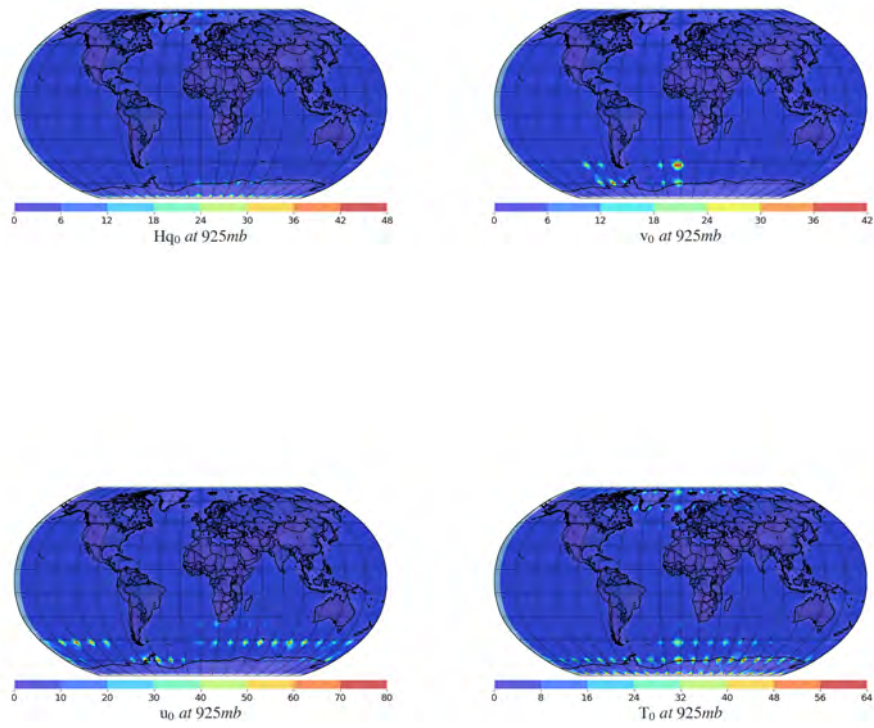


### 5.2.8. LETKF $p = 25\%$ $\alpha = 1,08$

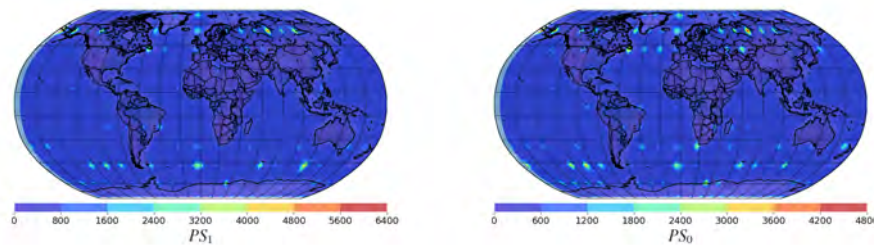




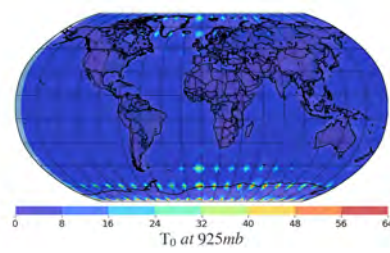
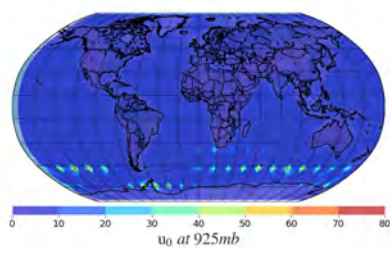
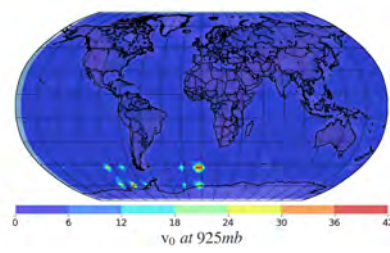
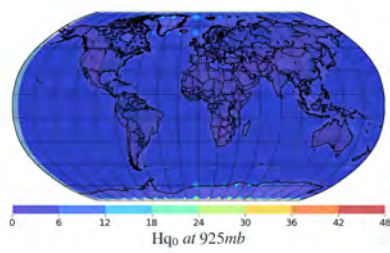
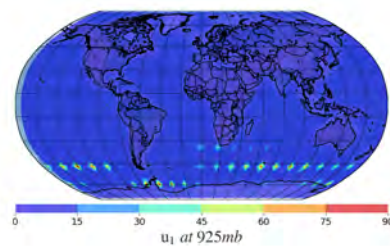
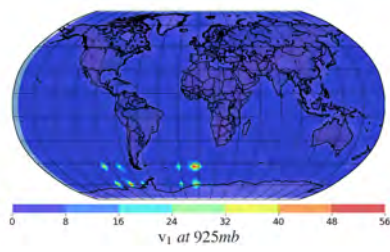
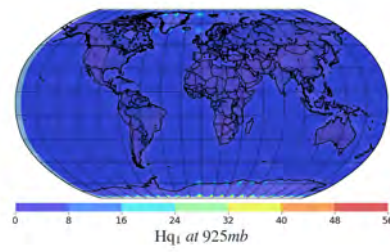
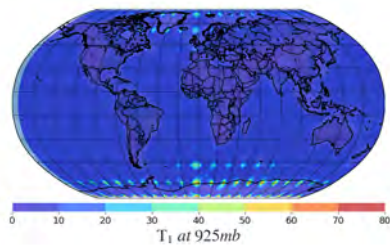
5.2.9. LETKF  $p = 11\%$   $\alpha = 1,02$ 

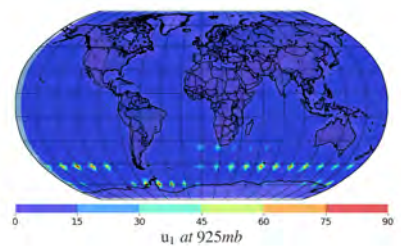
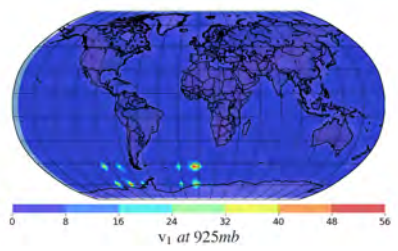
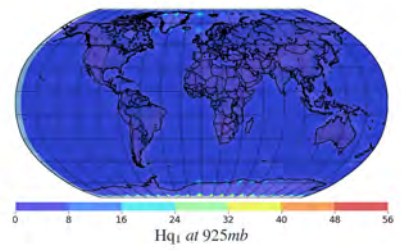
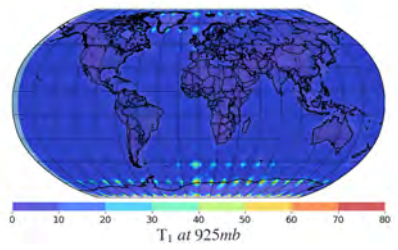
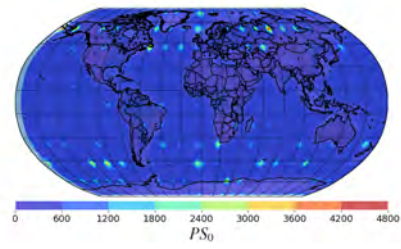
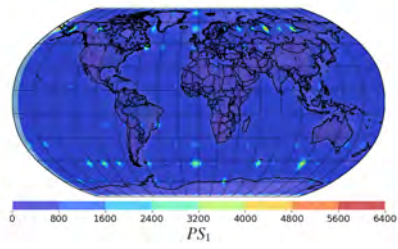


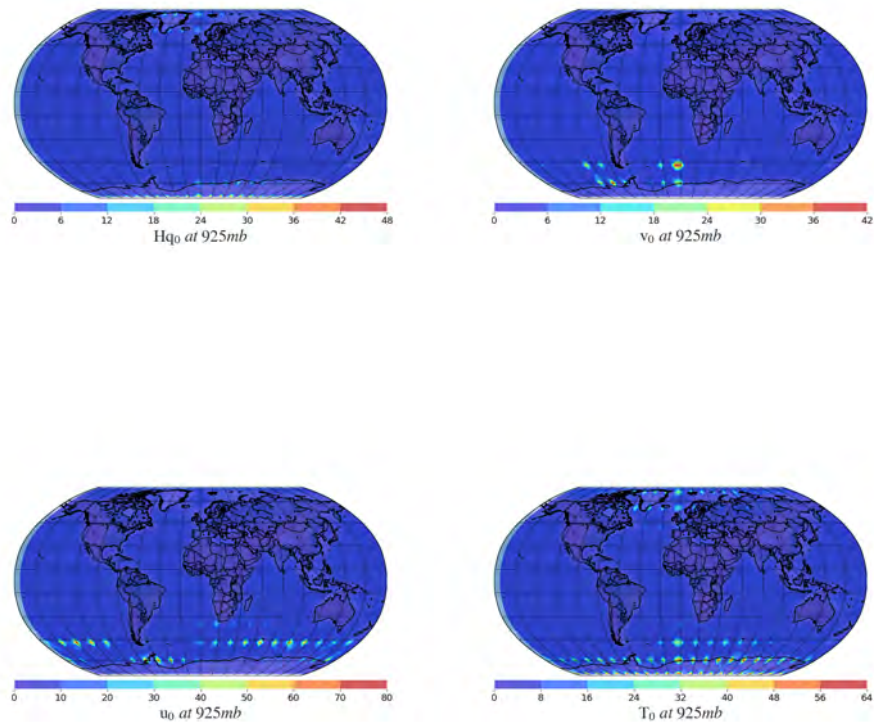
5.2.10. LETKF  $p = 11\%$   $\alpha = 1,04$



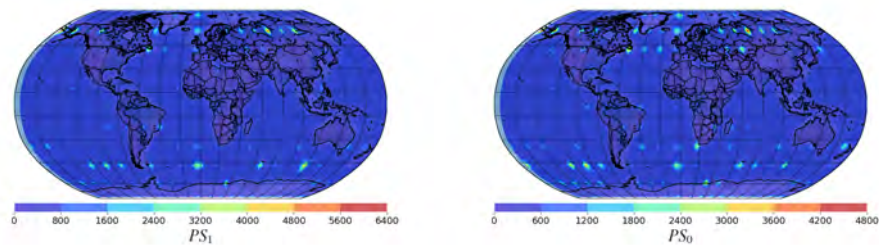


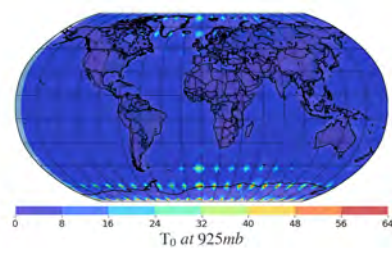
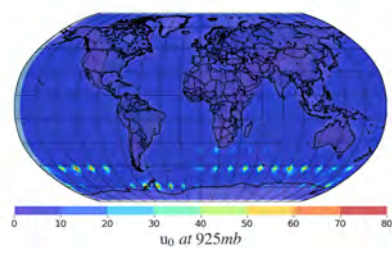
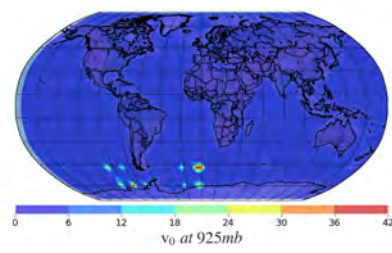
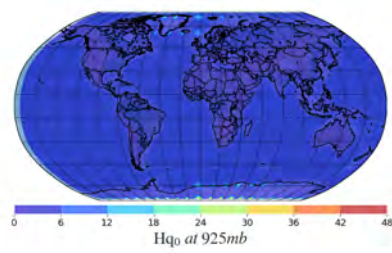
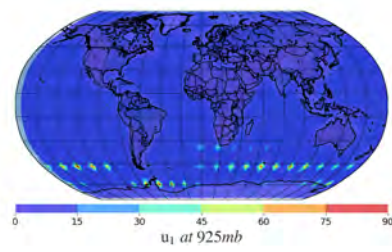
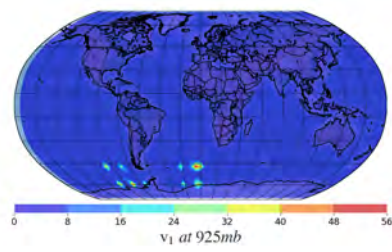
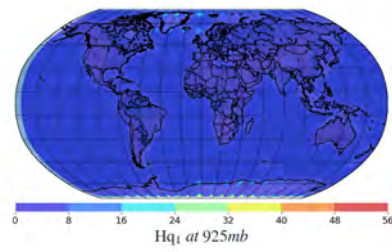
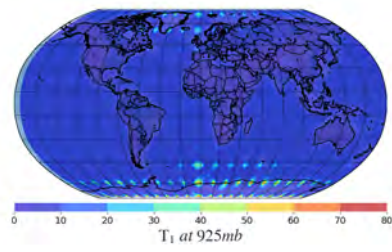


5.2.11. LETKF  $p = 11\%$   $\alpha = 1,06$ 

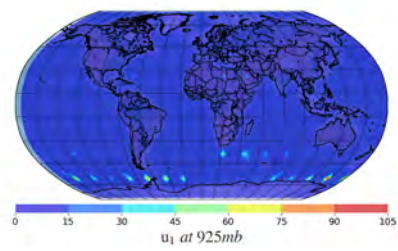
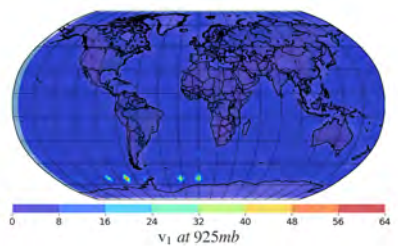
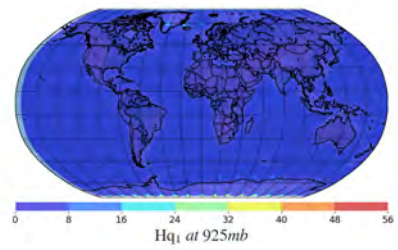
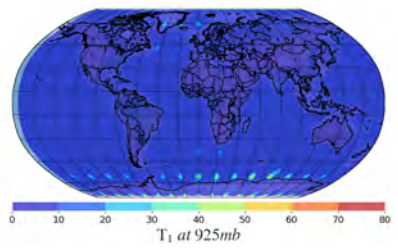
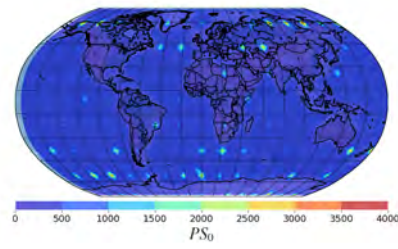
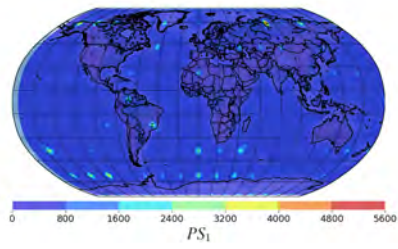


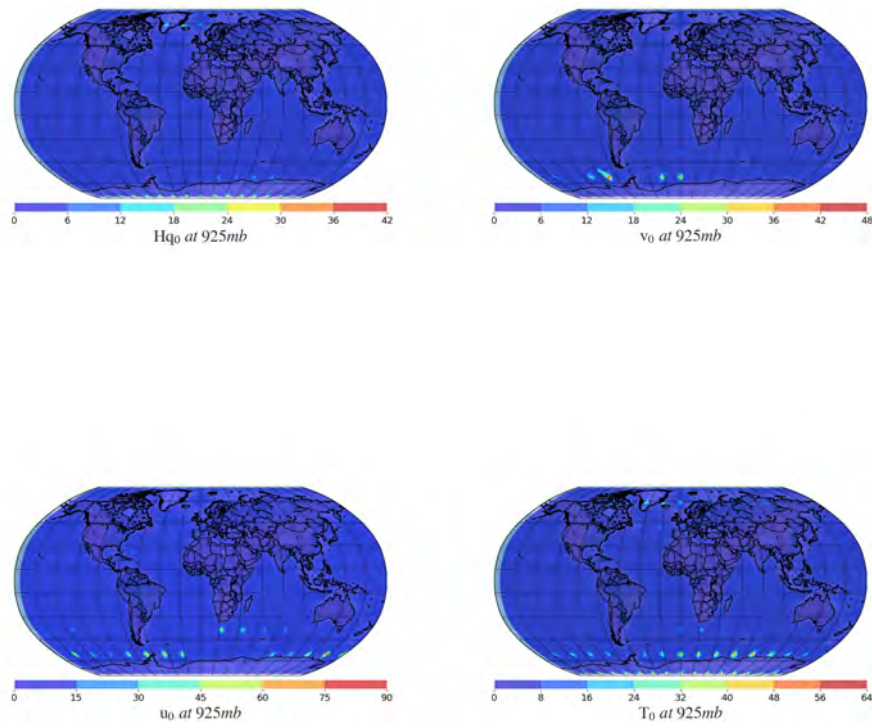
### 5.2.12. LETKF $p = 11\%$ $\alpha = 1,08$



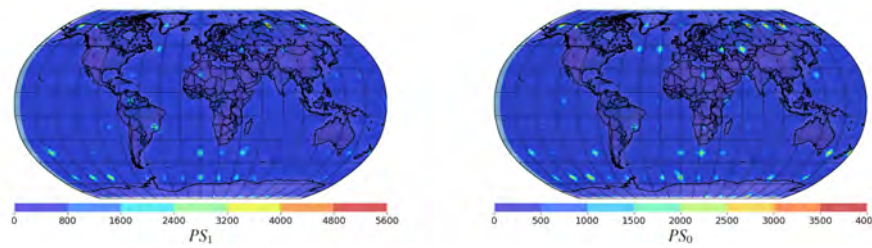


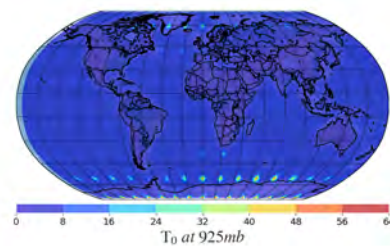
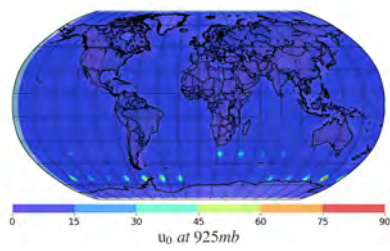
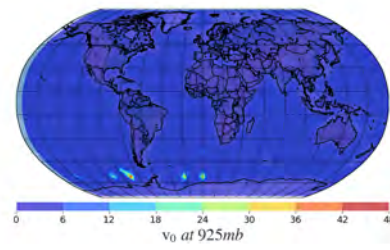
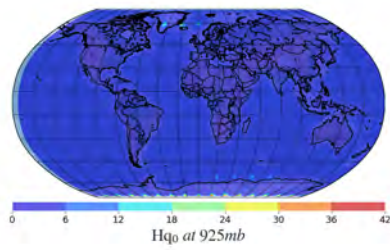
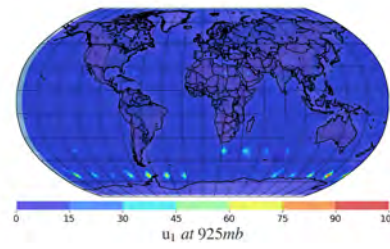
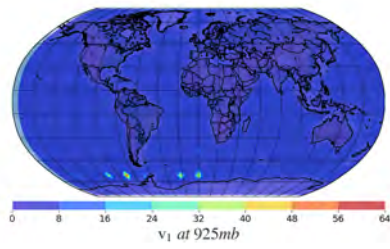
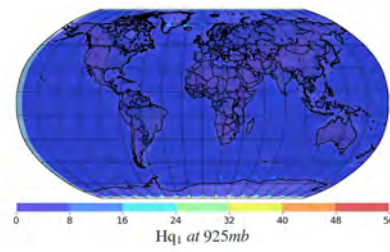
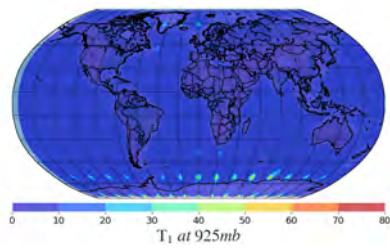


5.2.13. LETKF  $p = 6\%$   $\alpha = 1,02$ 

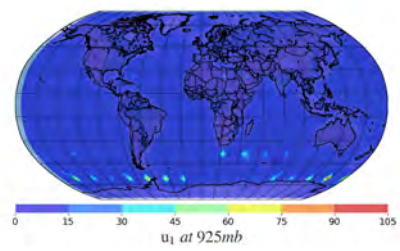
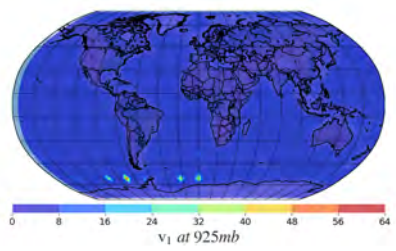
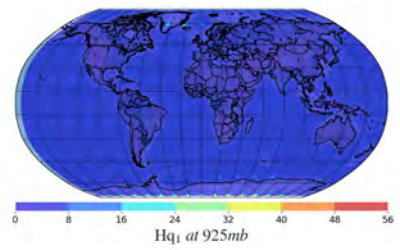
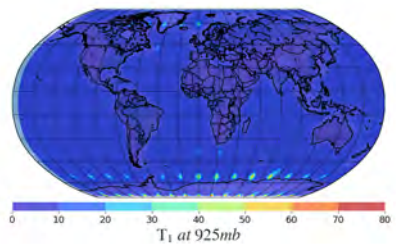
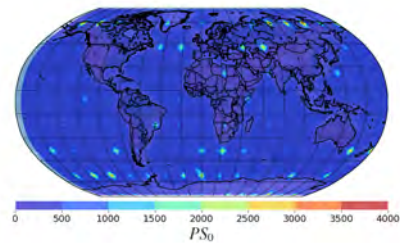
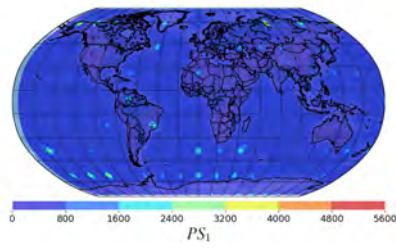


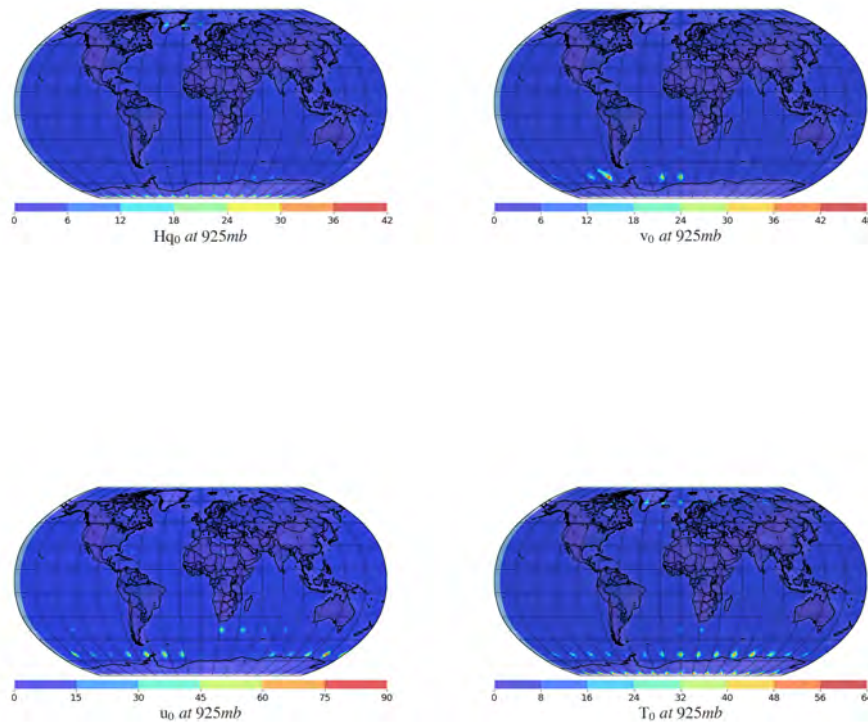
#### 5.2.14. LETKF $p = 6\%$ $\alpha = 1,04$



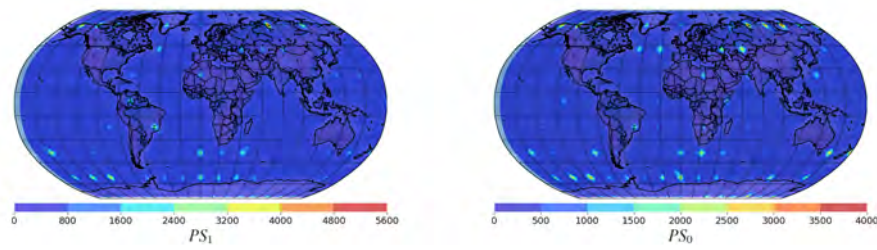


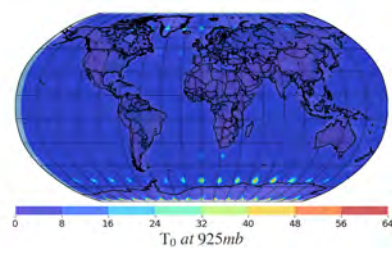
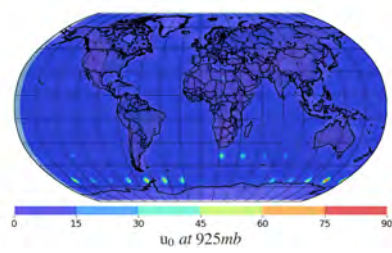
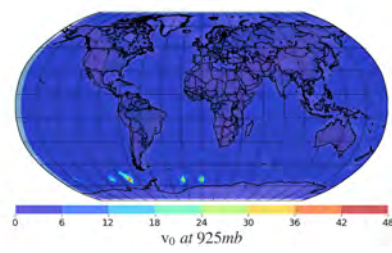
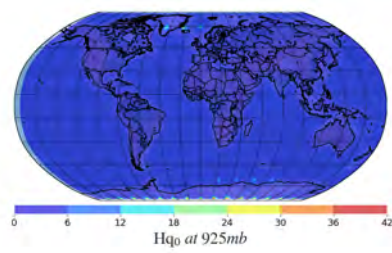
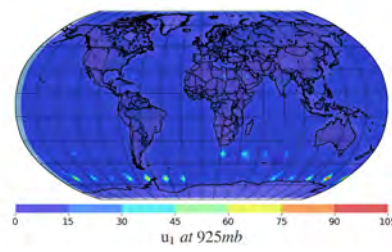
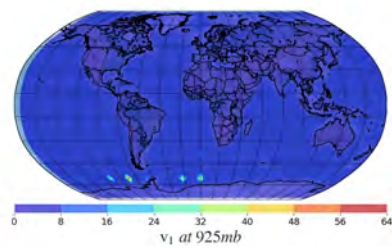
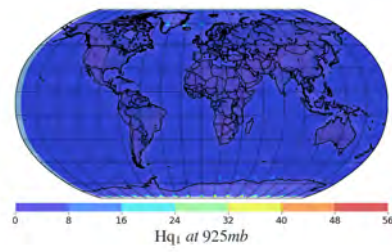
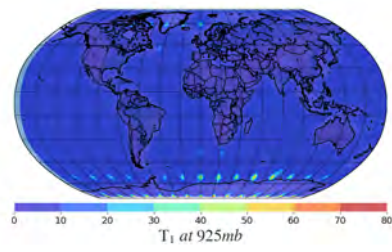


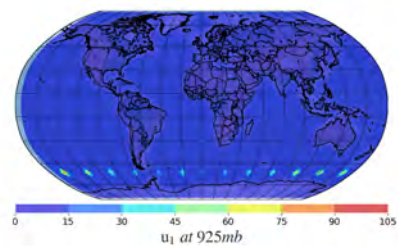
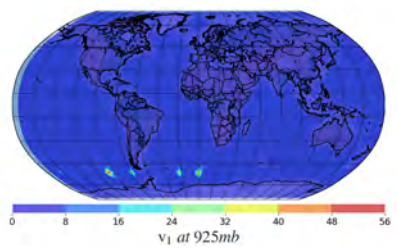
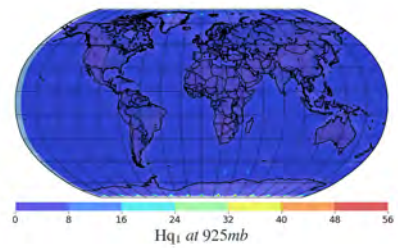
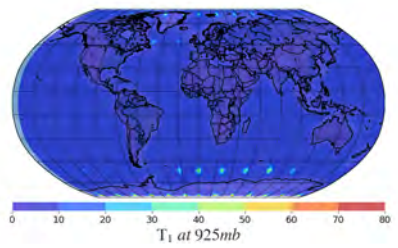
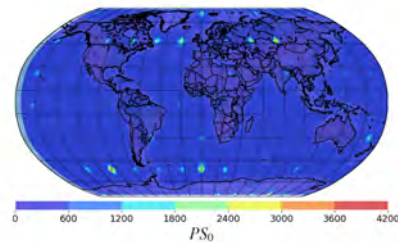
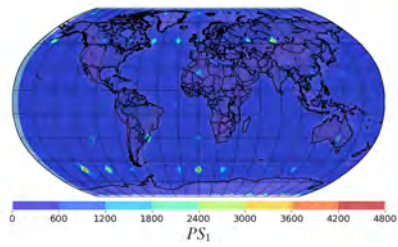
5.2.15. LETKF  $p = 6\%$   $\alpha = 1,06$ 



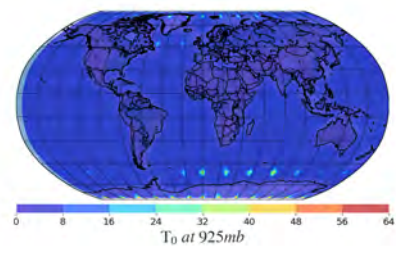
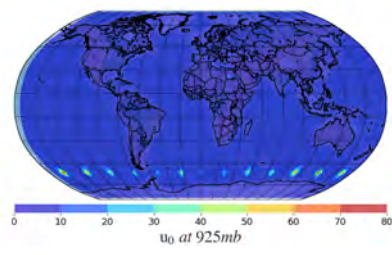
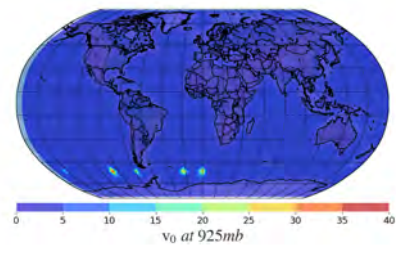
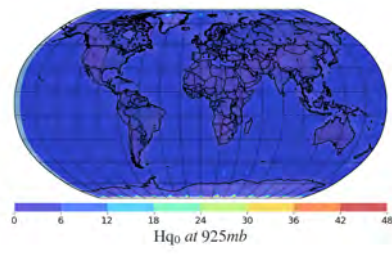
### 5.2.16. LETKF $p = 6\%$ $\alpha = 1,08$



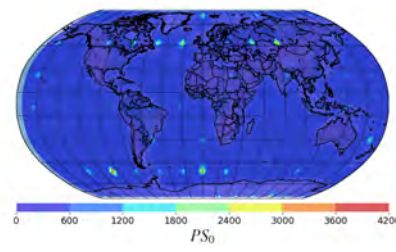
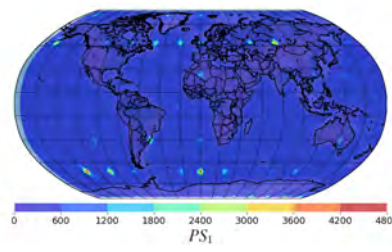


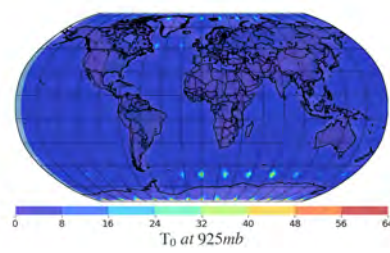
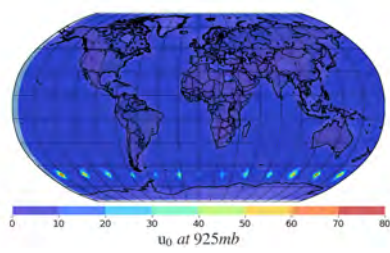
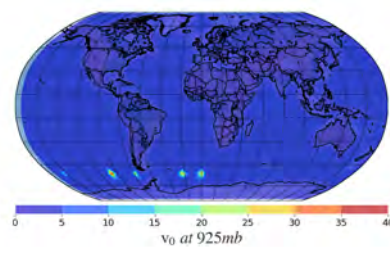
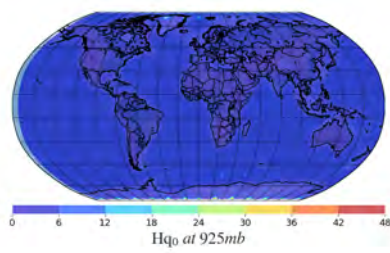
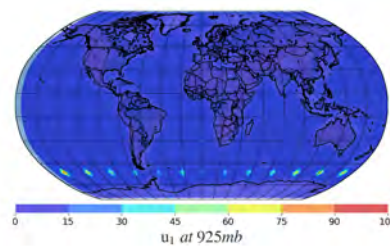
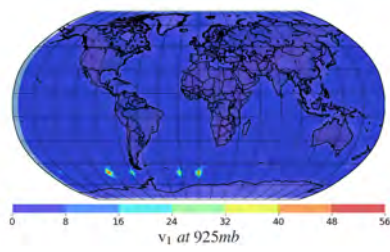
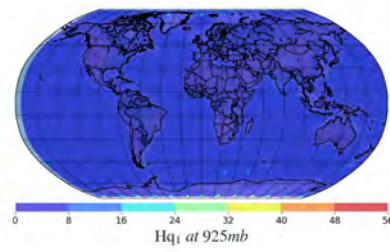
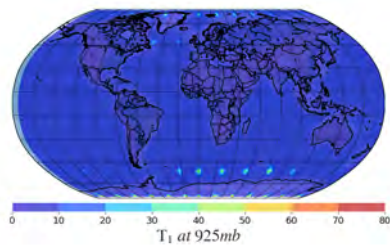
5.2.17. LETKF  $p = 4\%$   $\alpha = 1,02$ 

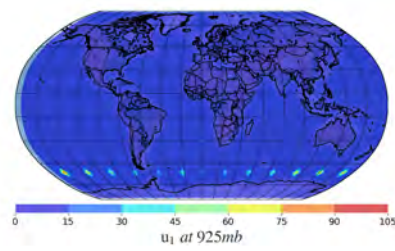
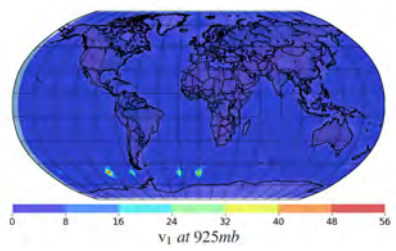
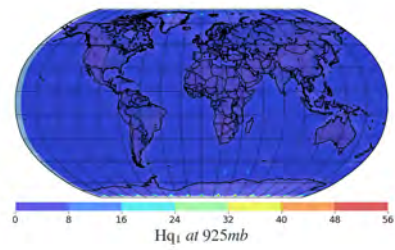
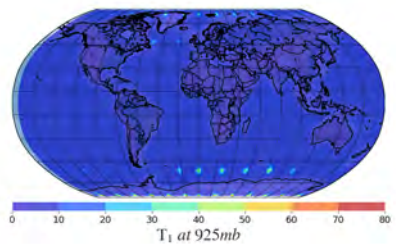
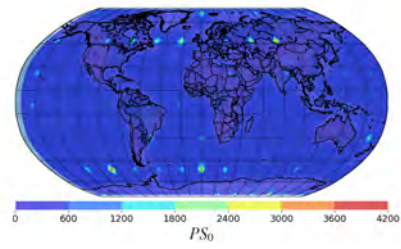
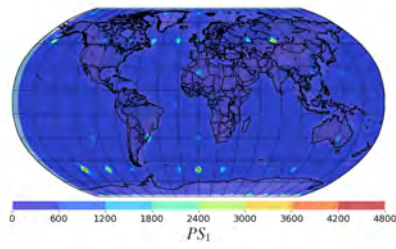




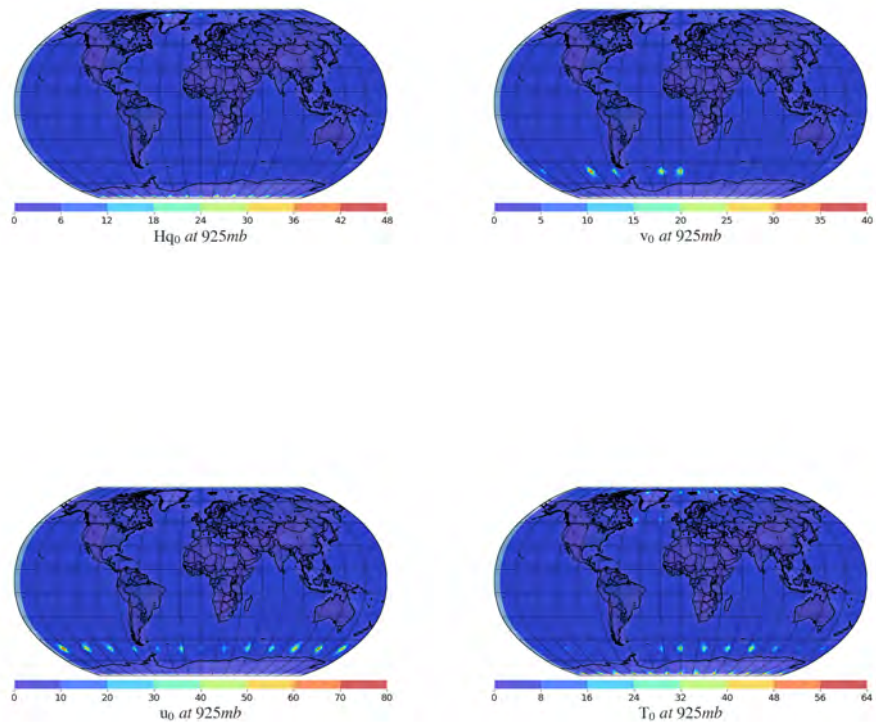
### 5.2.18. LETKF $p = 4\%$ $\alpha = 1,04$



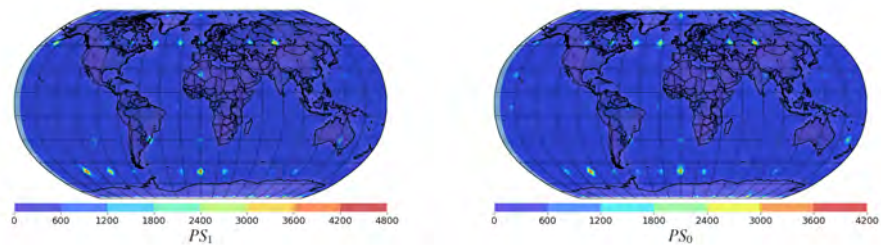


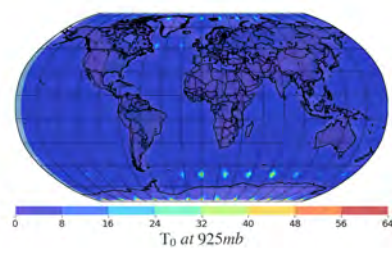
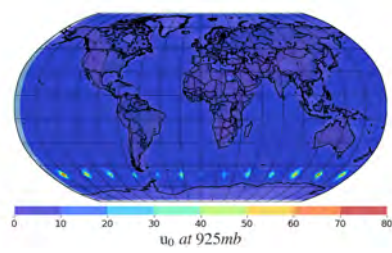
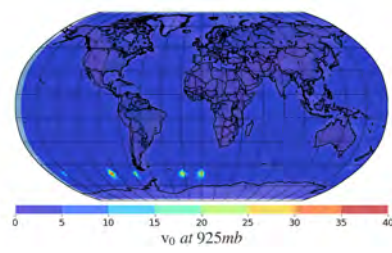
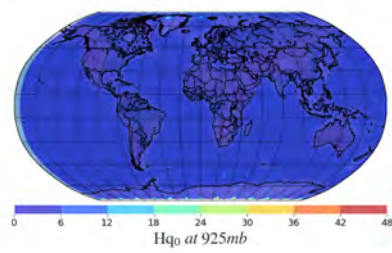
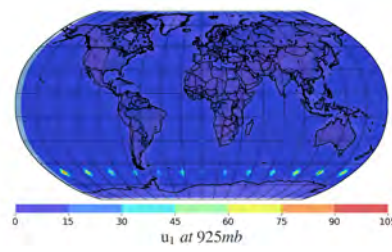
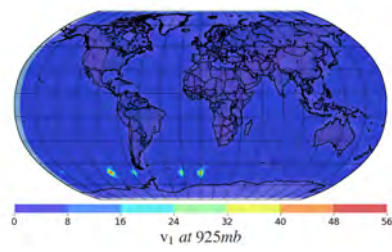
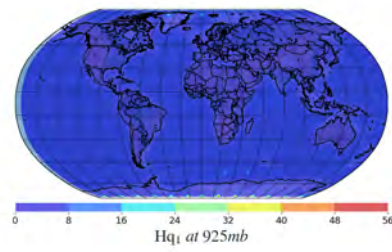
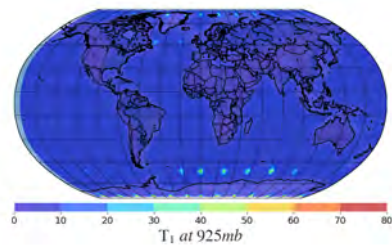
5.2.19. LETKF  $p = 4\%$   $\alpha = 1,06$ 





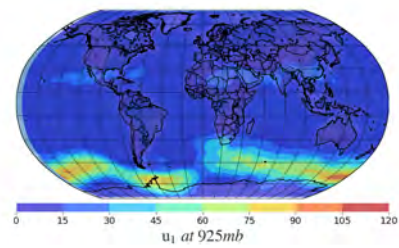
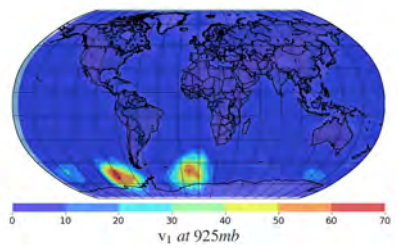
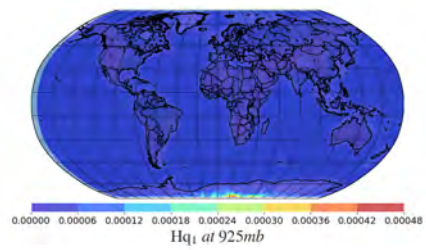
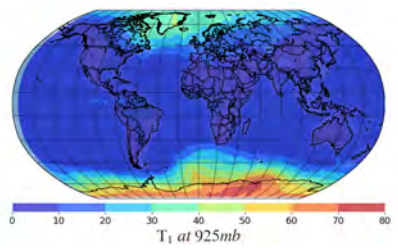
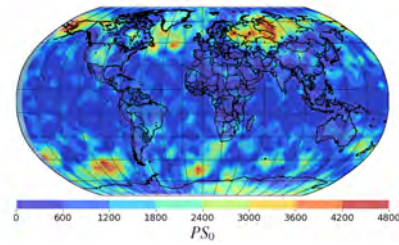
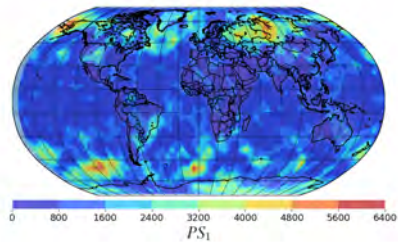
### 5.2.20. LETKF $p = 4\%$ $\alpha = 1,08$

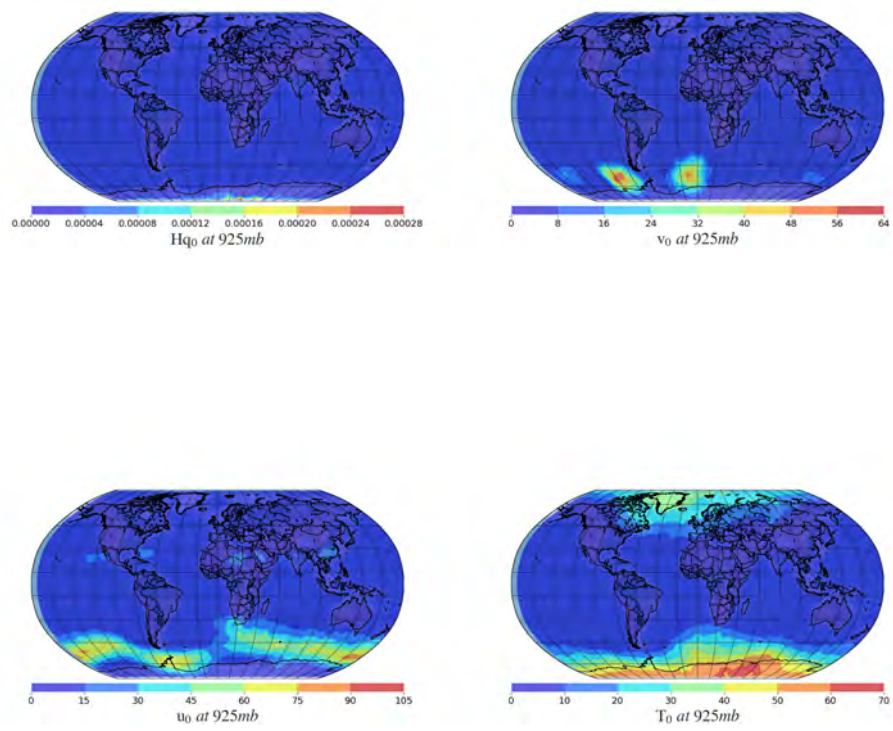




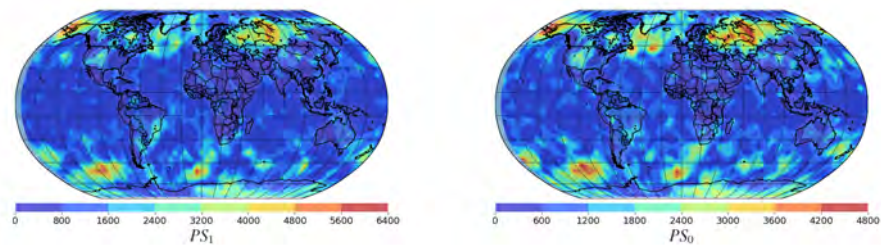
### 5.3. Local Ensemble Kalman Filter

#### 5.3.1. LEnKF $p = 100\%$ $\alpha = 1,02$

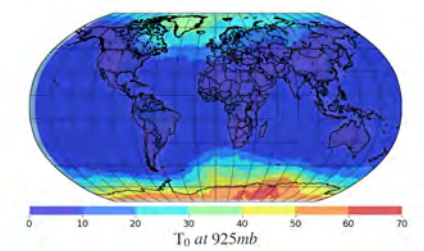
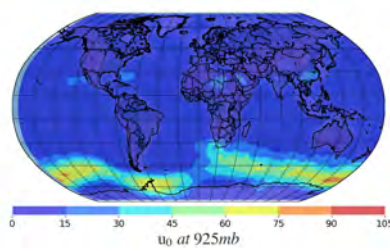
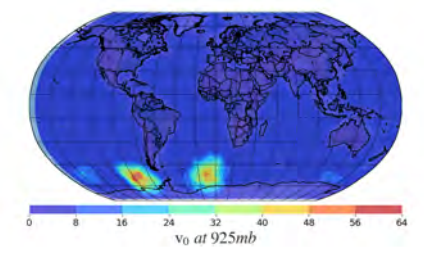
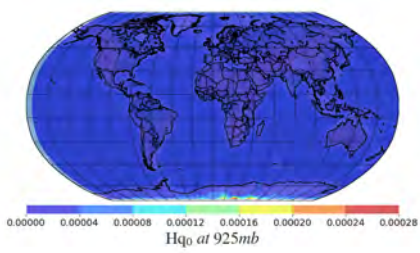
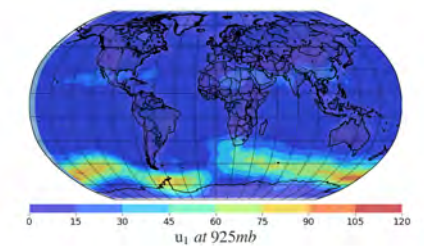
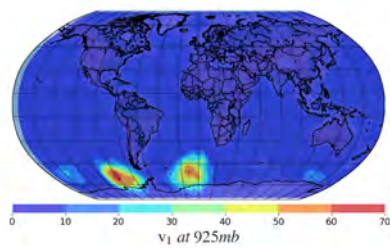
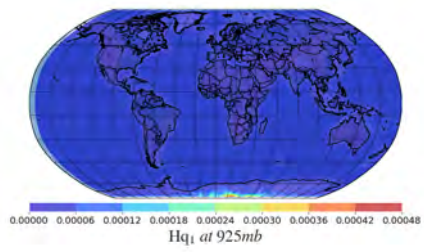
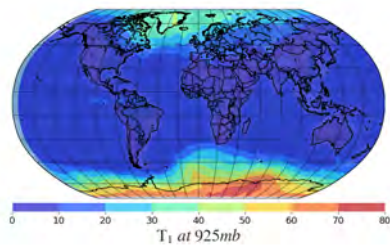


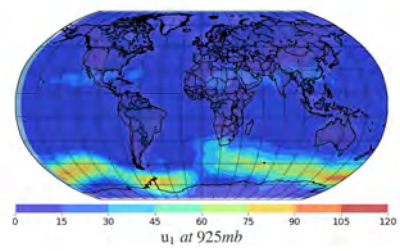
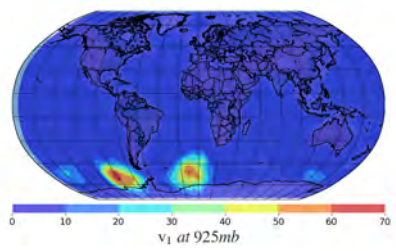
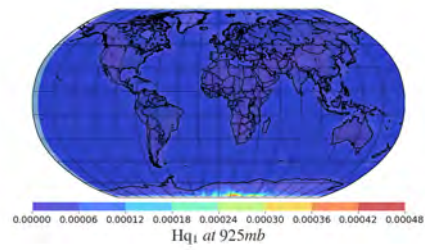
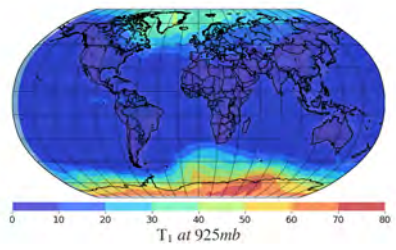
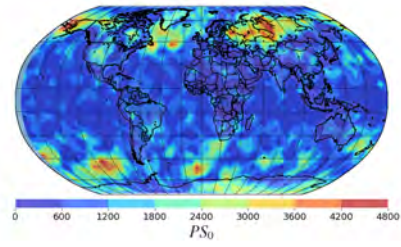
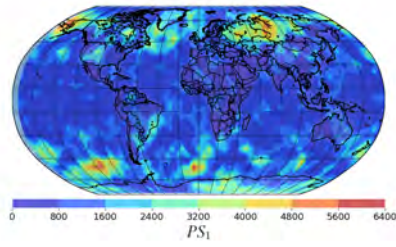


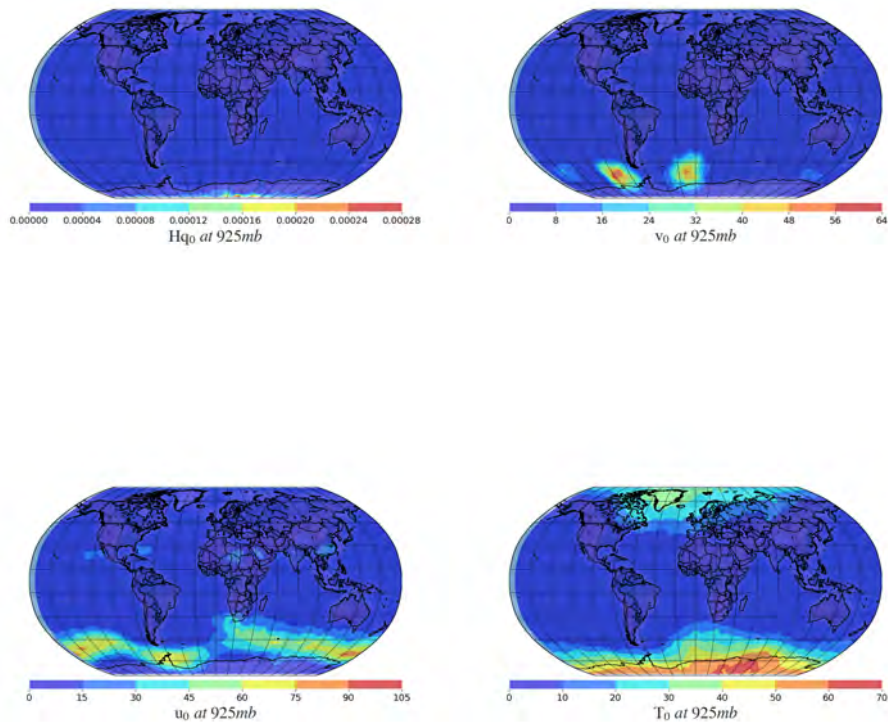
### 5.3.2. LEnKF $p = 100\%$ $\alpha = 1,04$



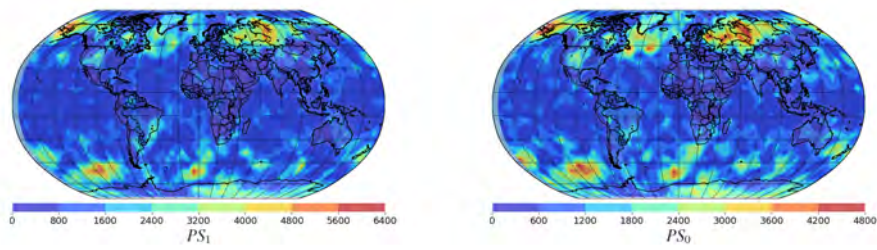




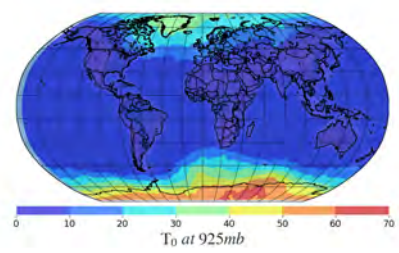
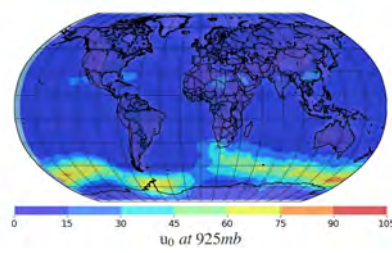
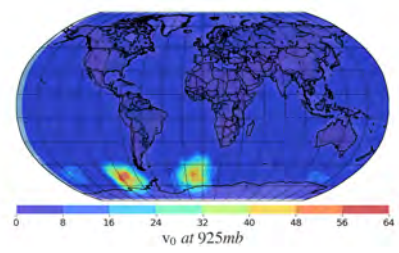
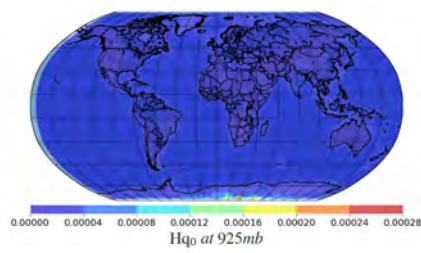
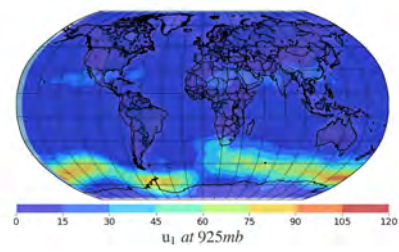
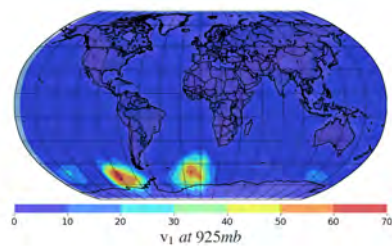
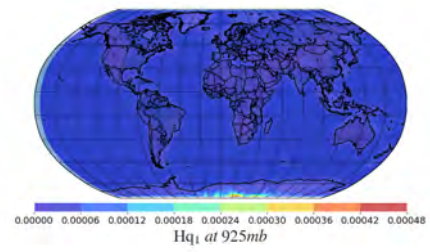
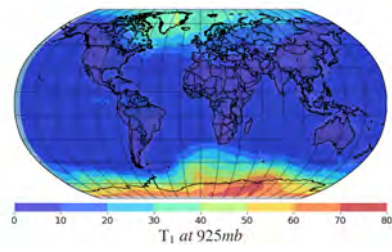
5.3.3. LEnKF  $p = 100\%$   $\alpha = 1,06$ 

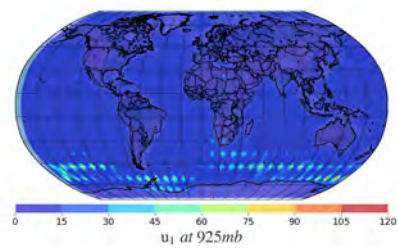
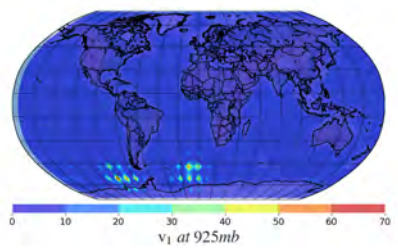
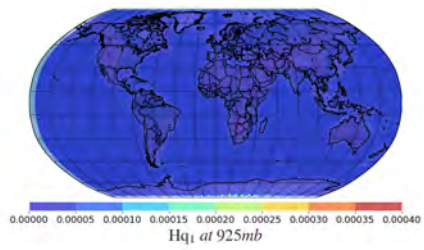
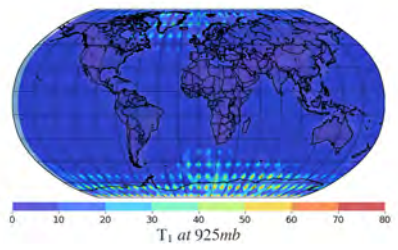
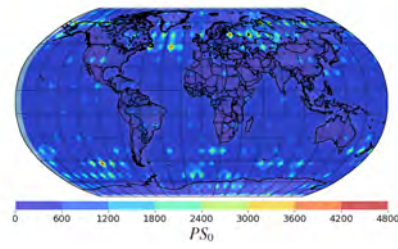
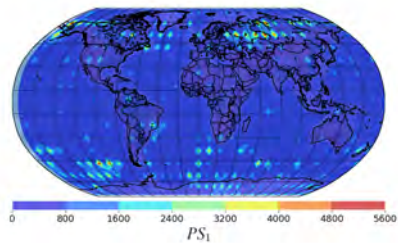


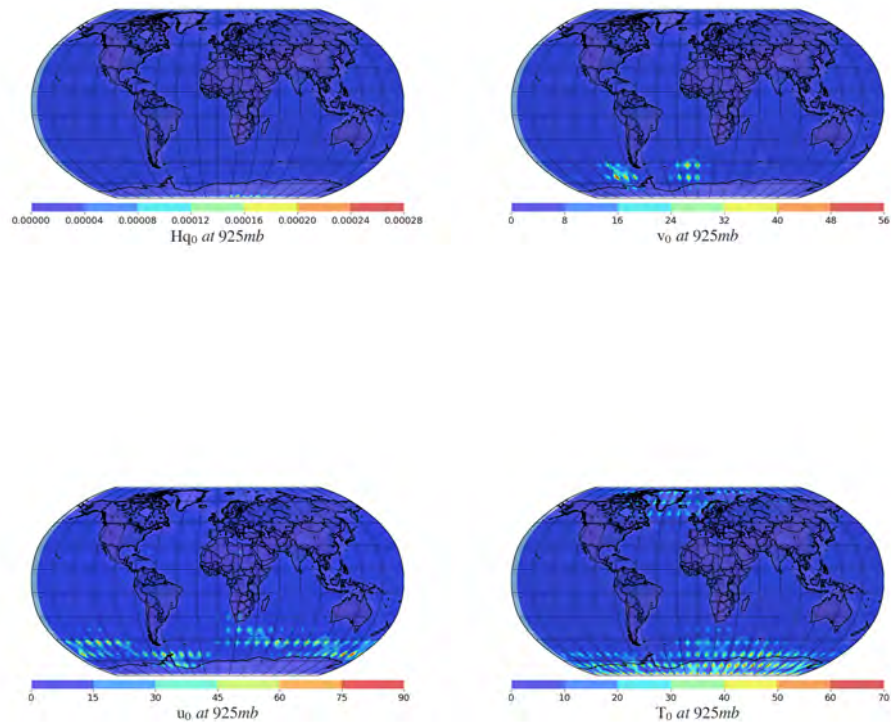
#### 5.3.4. LEnKF $p = 100\%$ $\alpha = 1,08$



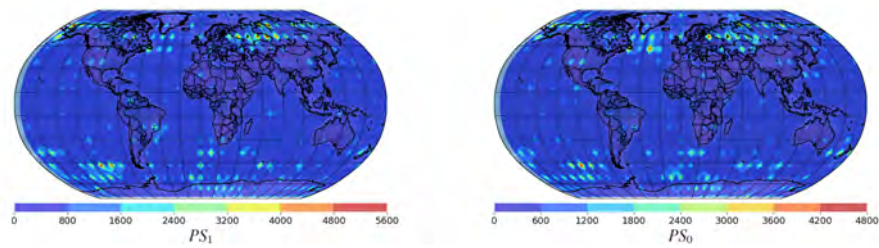




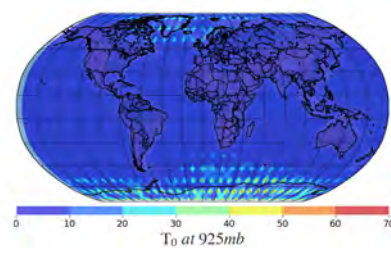
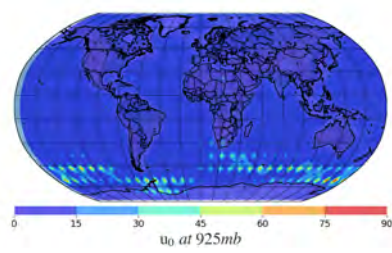
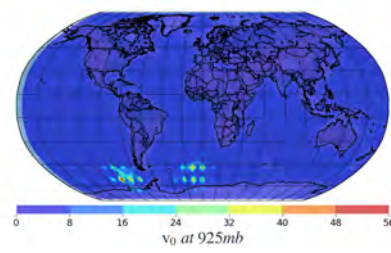
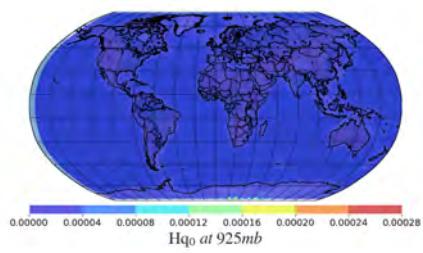
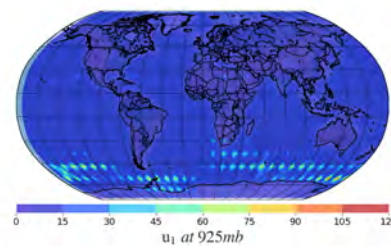
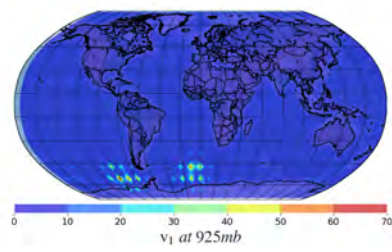
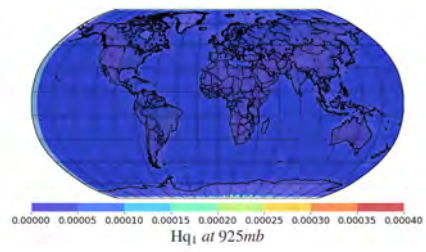
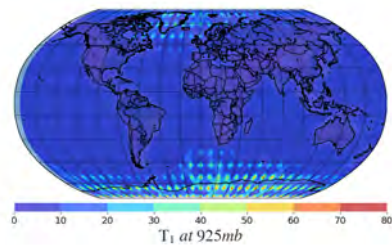
5.3.5. LEnKF  $p = 25\%$   $\alpha = 1,02$ 

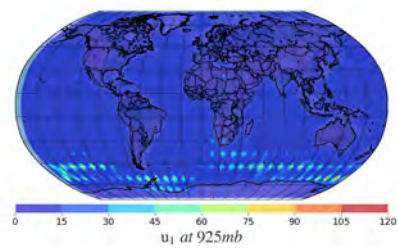
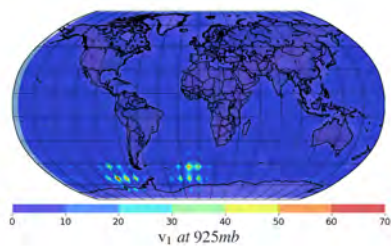
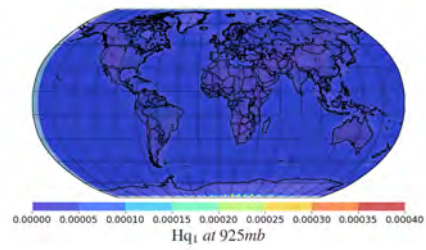
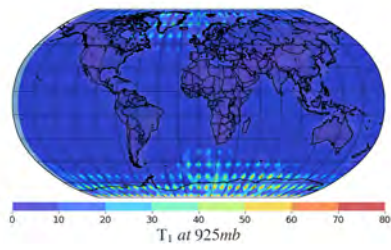
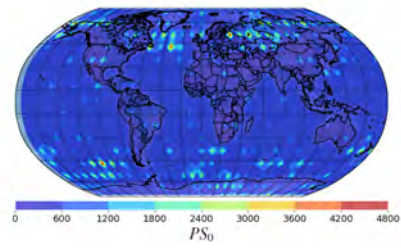
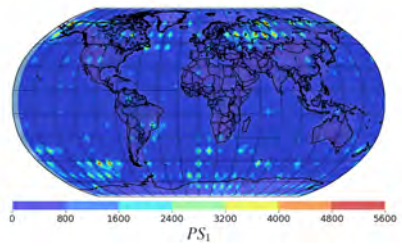


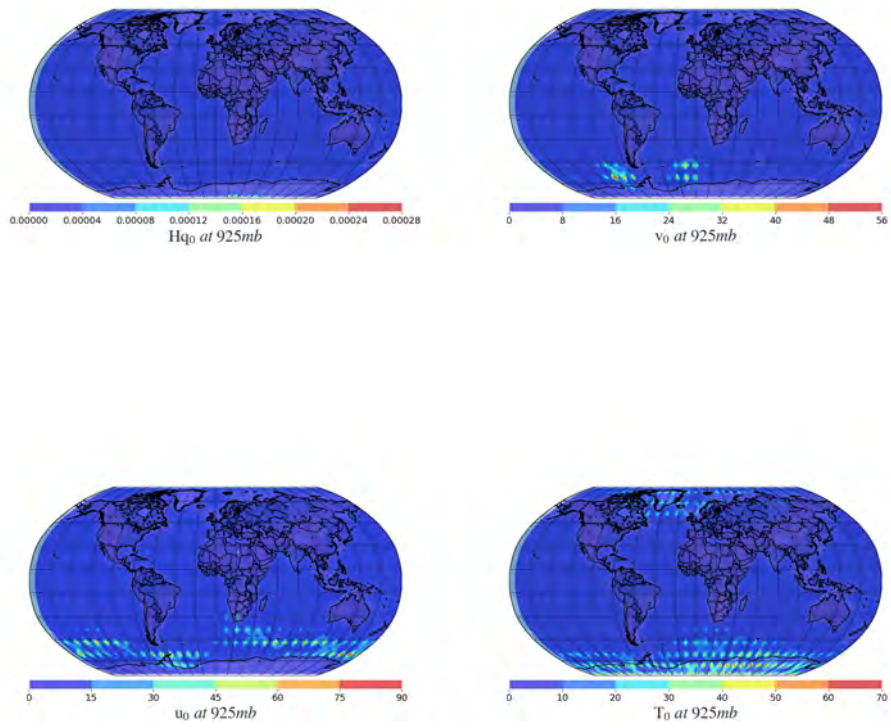
### 5.3.6. LEnKF $p = 25\%$ $\alpha = 1,04$



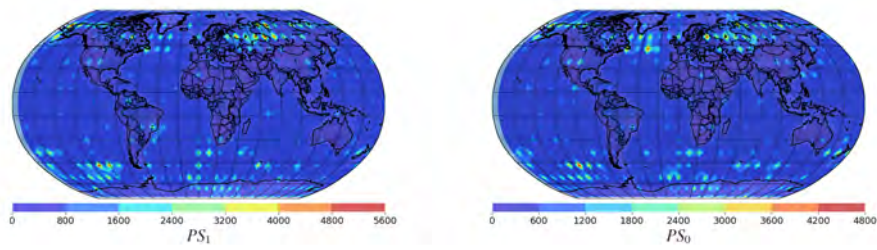




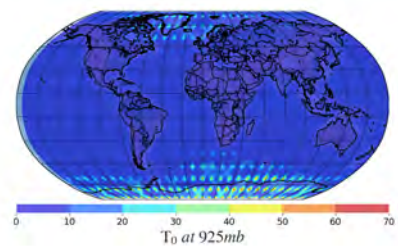
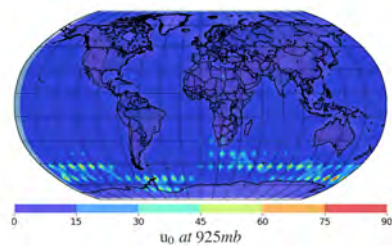
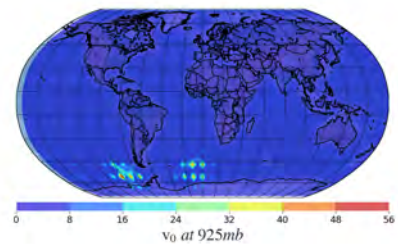
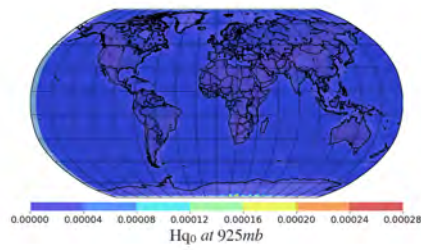
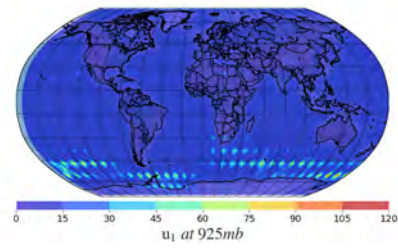
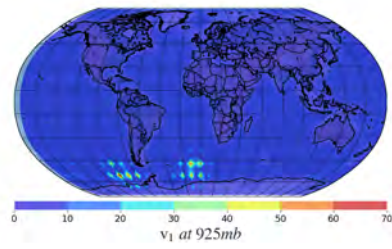
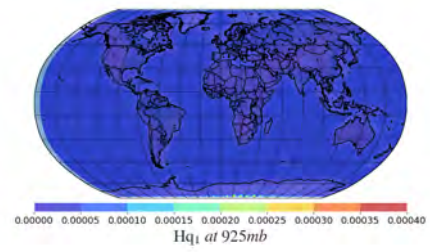
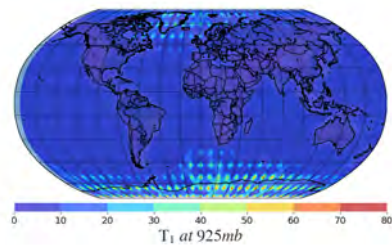
5.3.7. LEnKF  $p = 25\%$   $\alpha = 1,06$ 

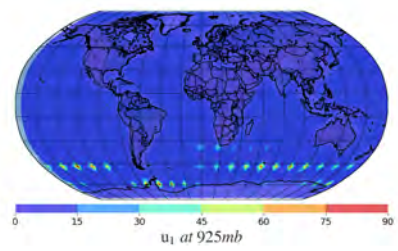
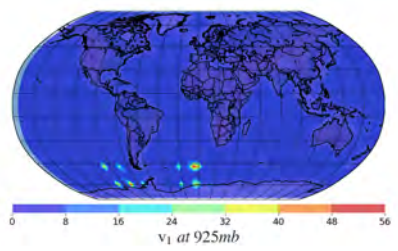
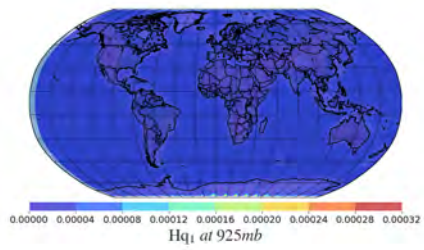
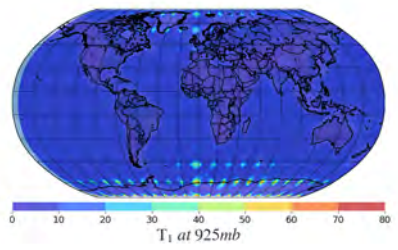
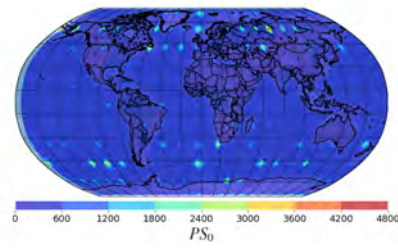
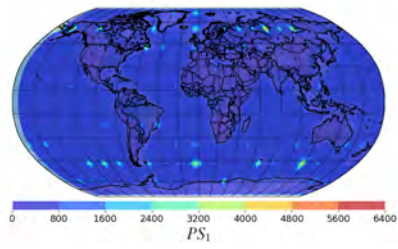


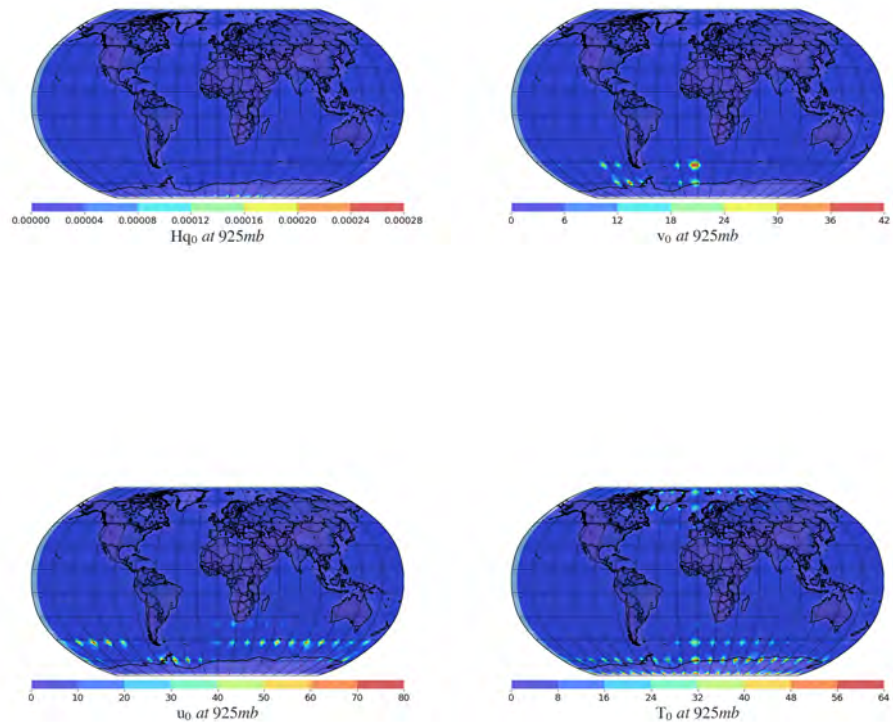
### 5.3.8. LEnKF $p = 25\%$ $\alpha = 1,08$



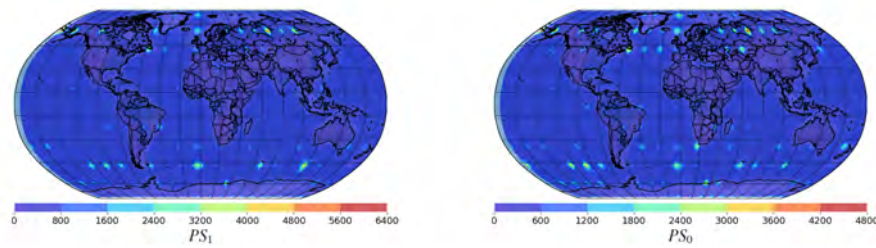


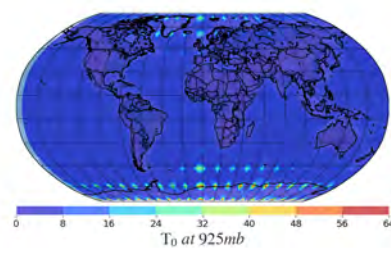
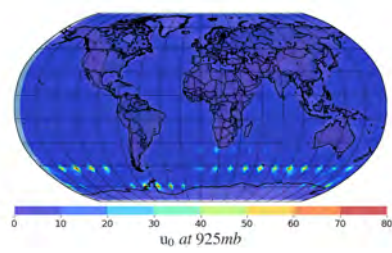
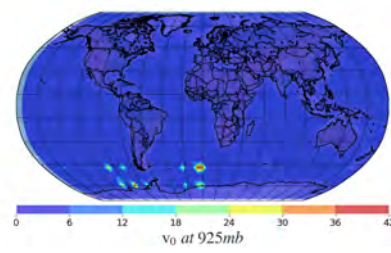
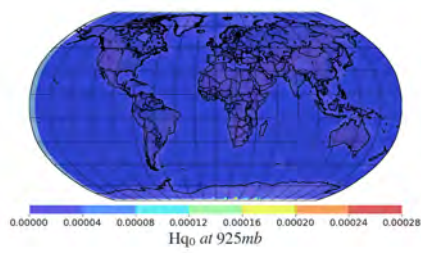
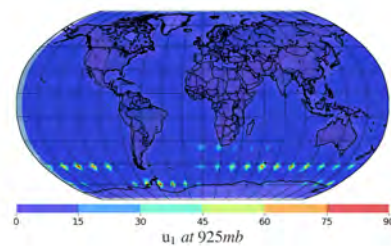
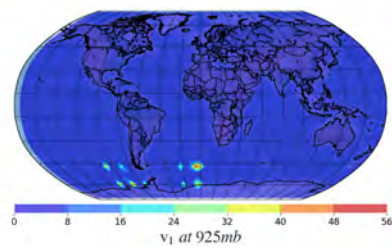
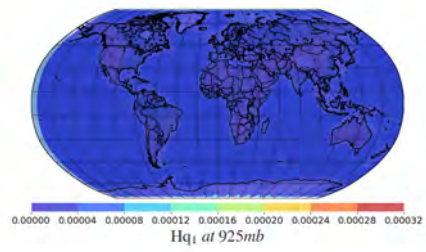
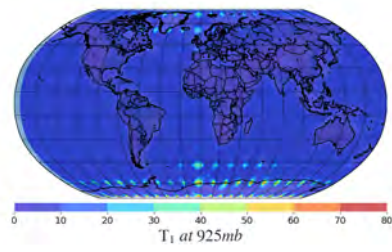


5.3.9. LEnKF  $p = 11\%$   $\alpha = 1,02$ 

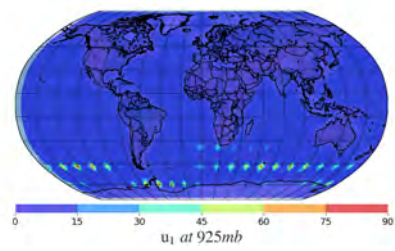
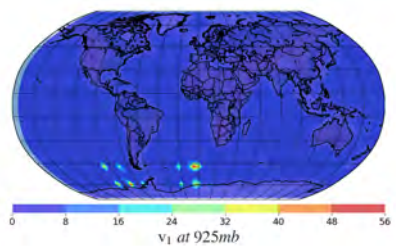
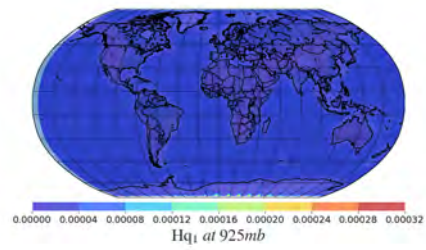
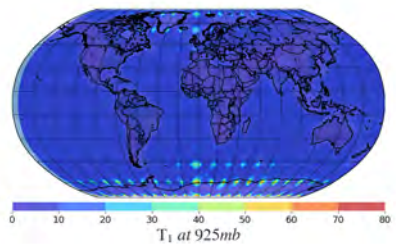
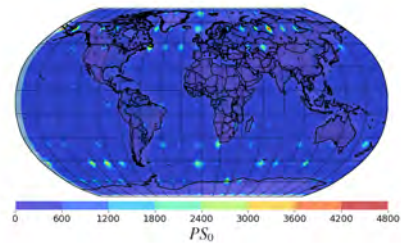
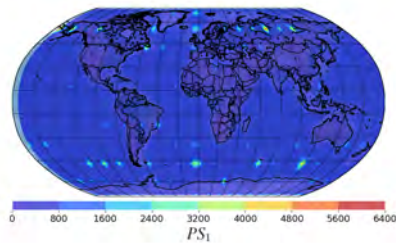


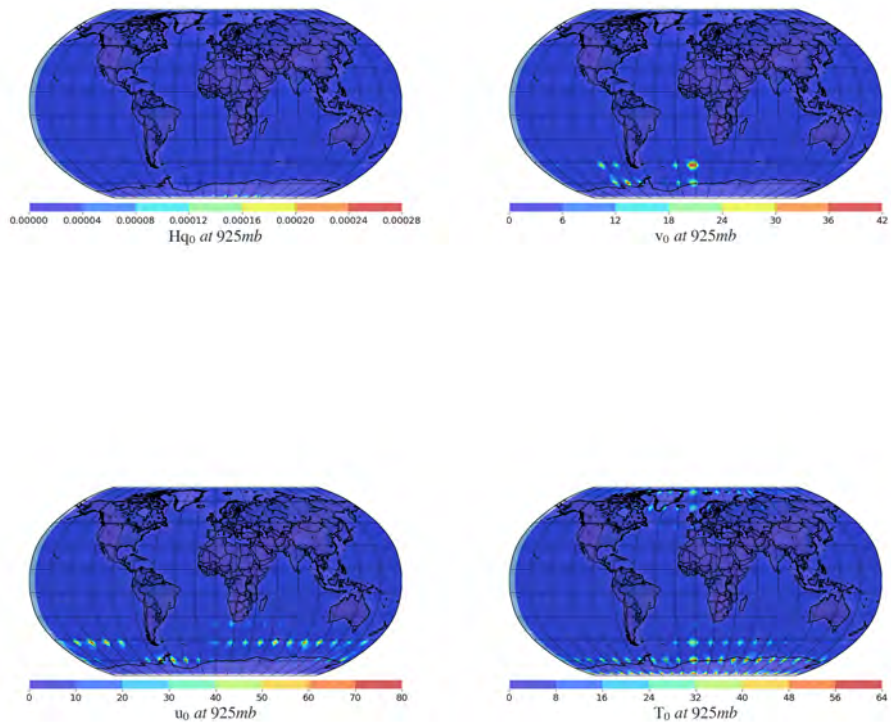
5.3.10. LEnKF  $p = 11\%$   $\alpha = 1,04$



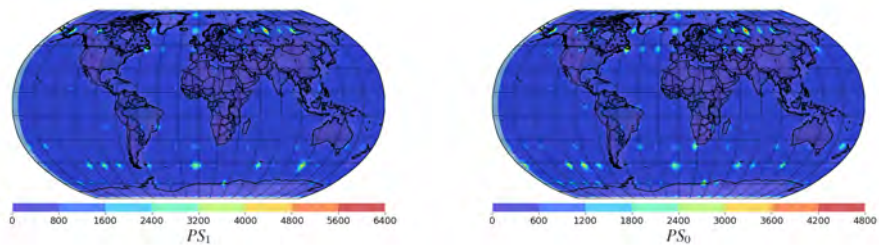




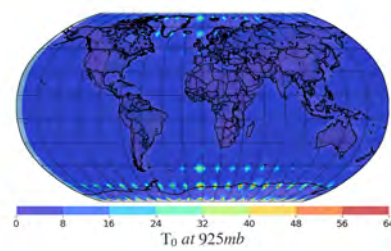
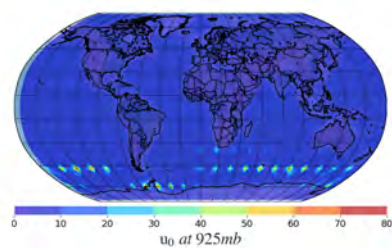
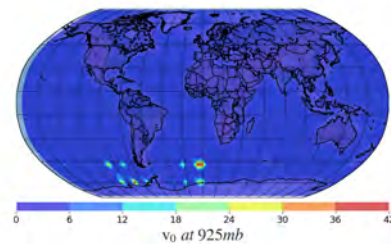
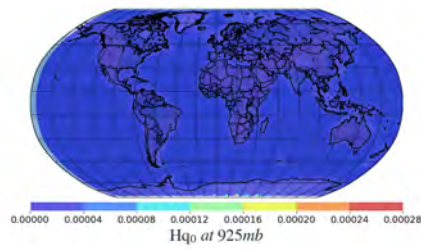
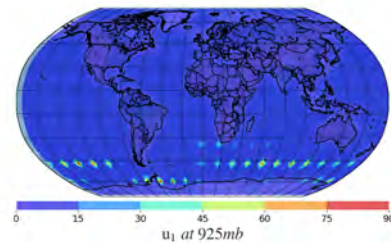
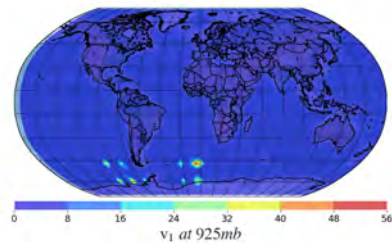
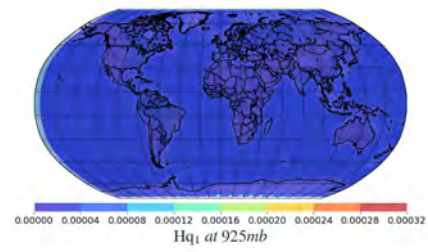
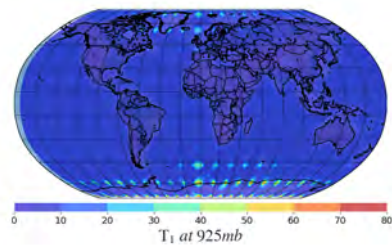
5.3.11. LEnKF  $p = 11\%$   $\alpha = 1,06$ 

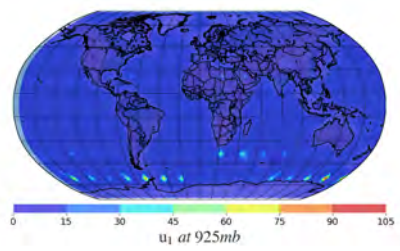
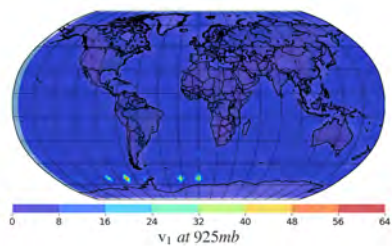
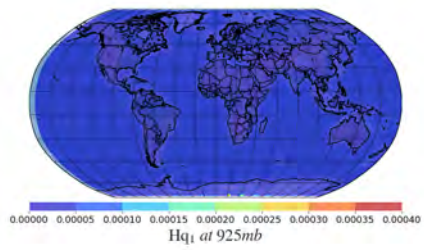
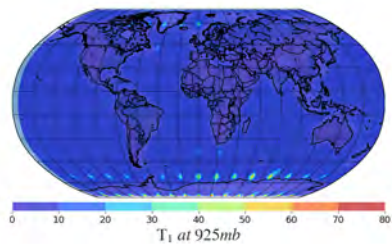
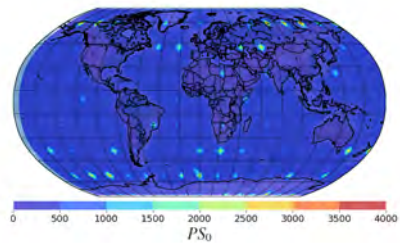
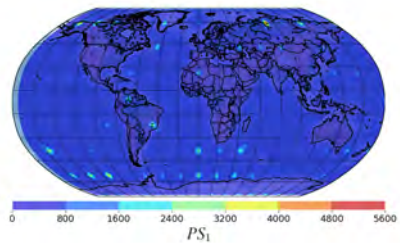


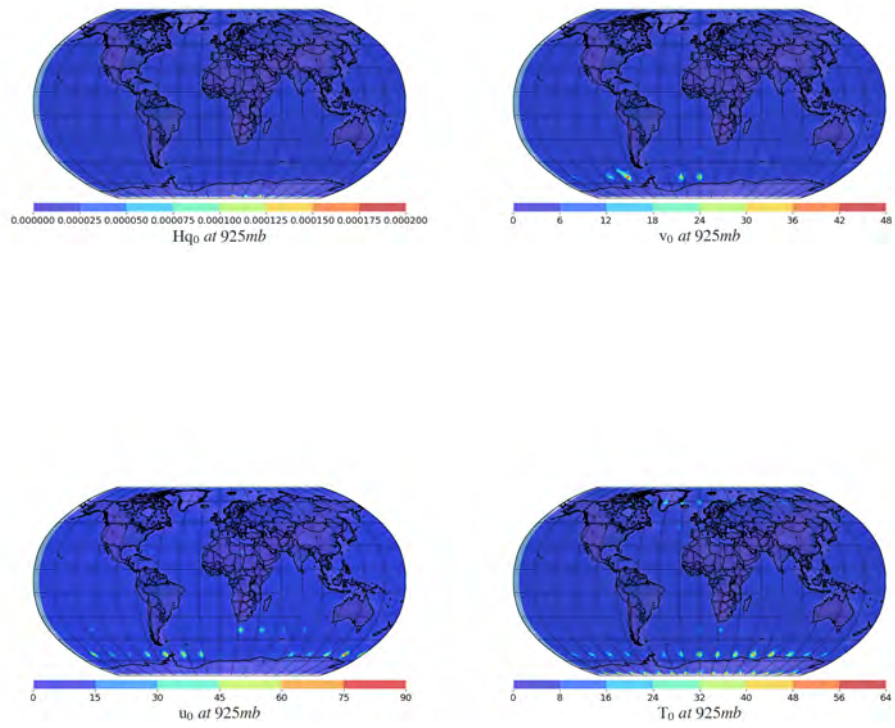
### 5.3.12. LEnKF $p = 11\%$ $\alpha = 1,08$



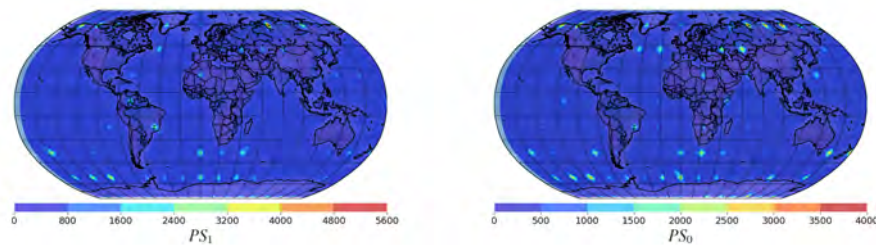


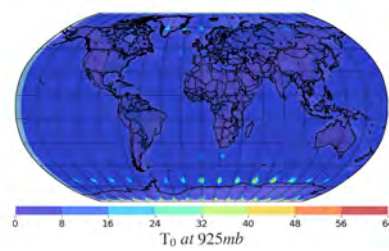
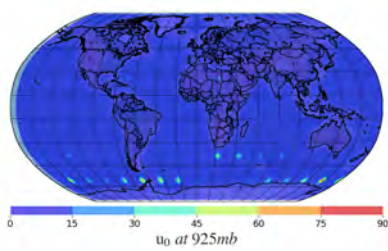
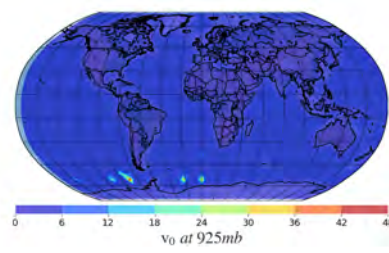
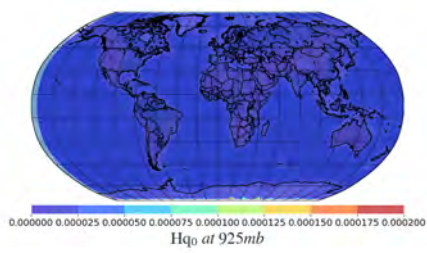
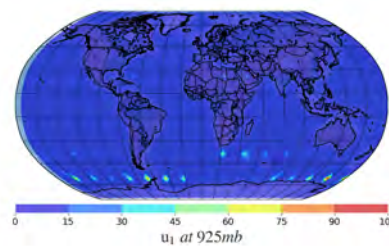
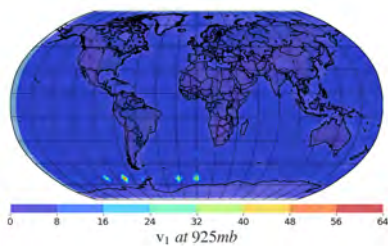
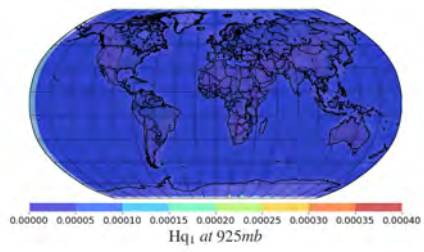
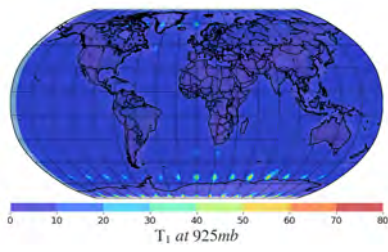


5.3.13. LEnKF  $p = 6\%$   $\alpha = 1,02$ 

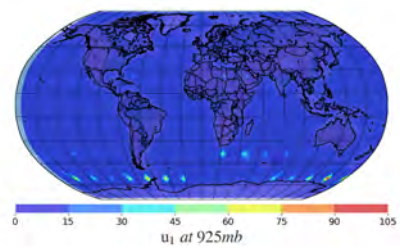
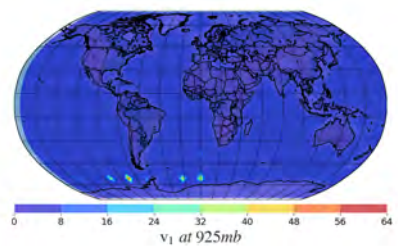
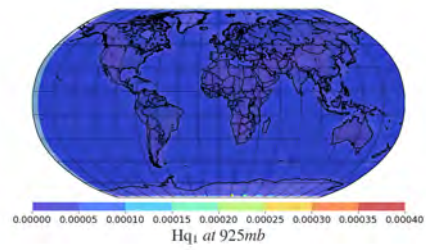
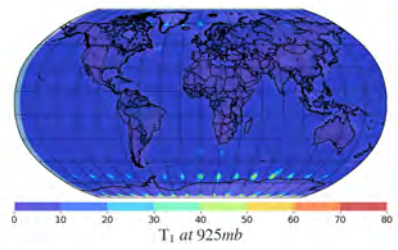
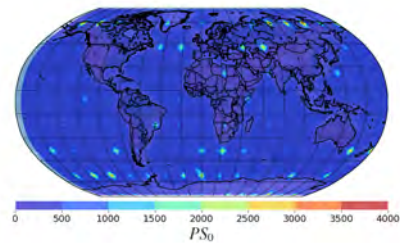
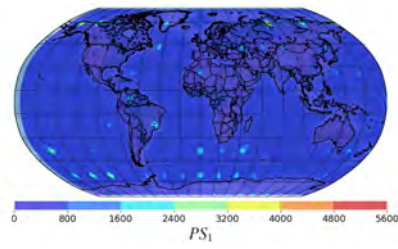


### 5.3.14. LEnKF $p = 6\%$ $\alpha = 1,04$

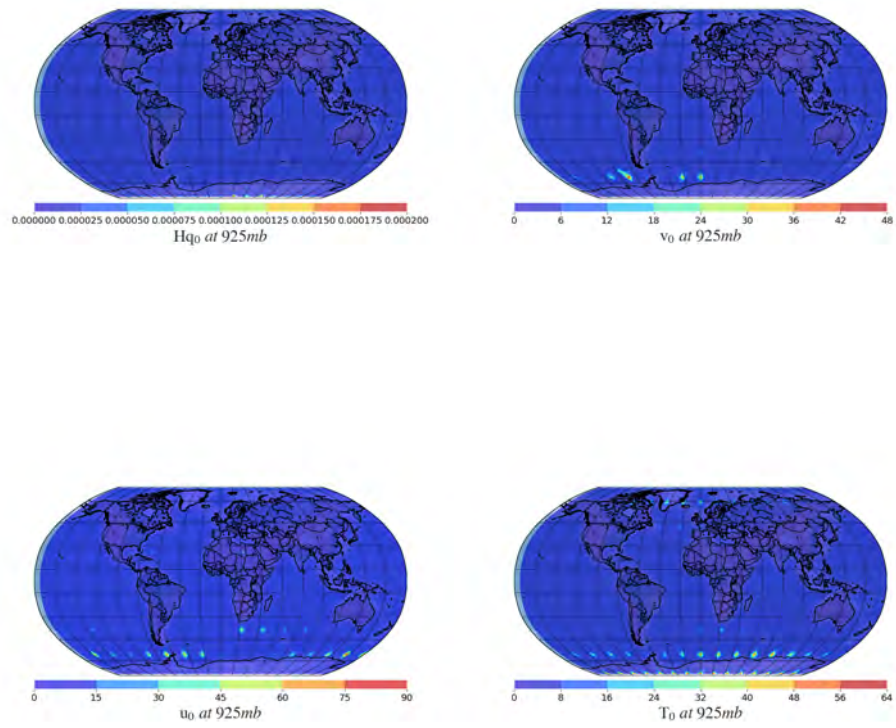




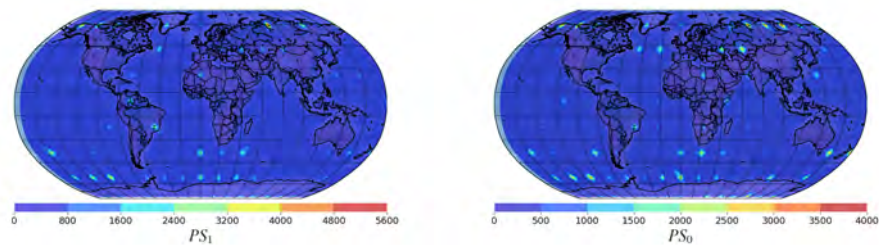


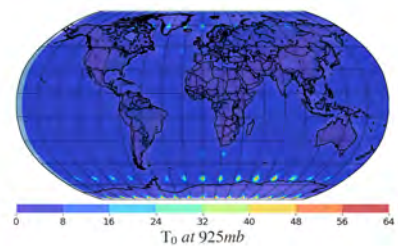
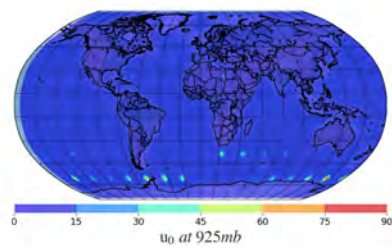
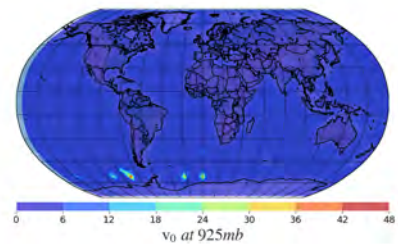
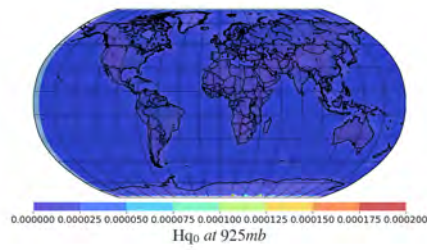
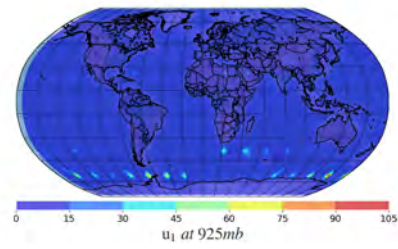
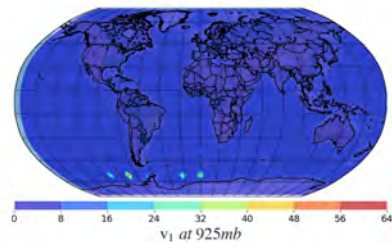
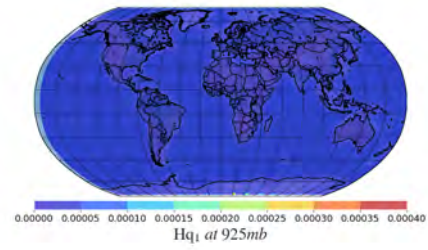
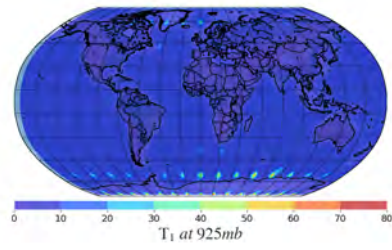
5.3.15. LEnKF  $p = 6\%$   $\alpha = 1,06$ 

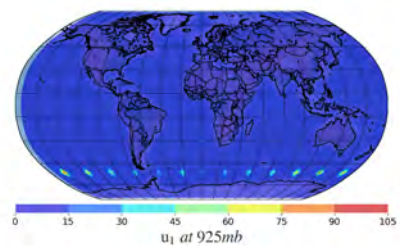
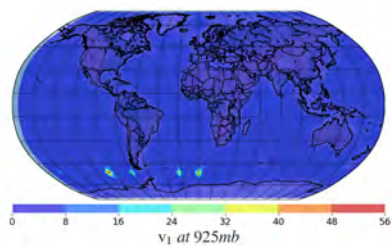
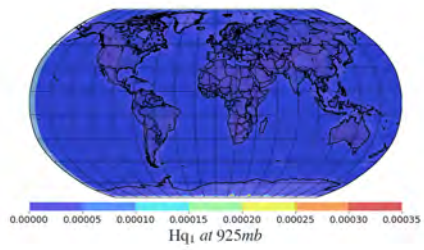
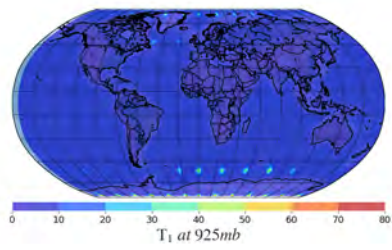
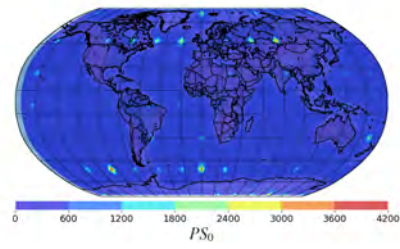
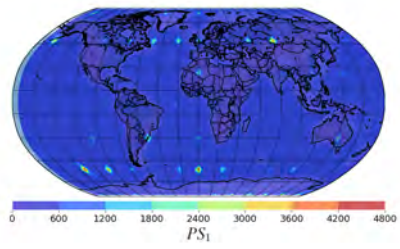


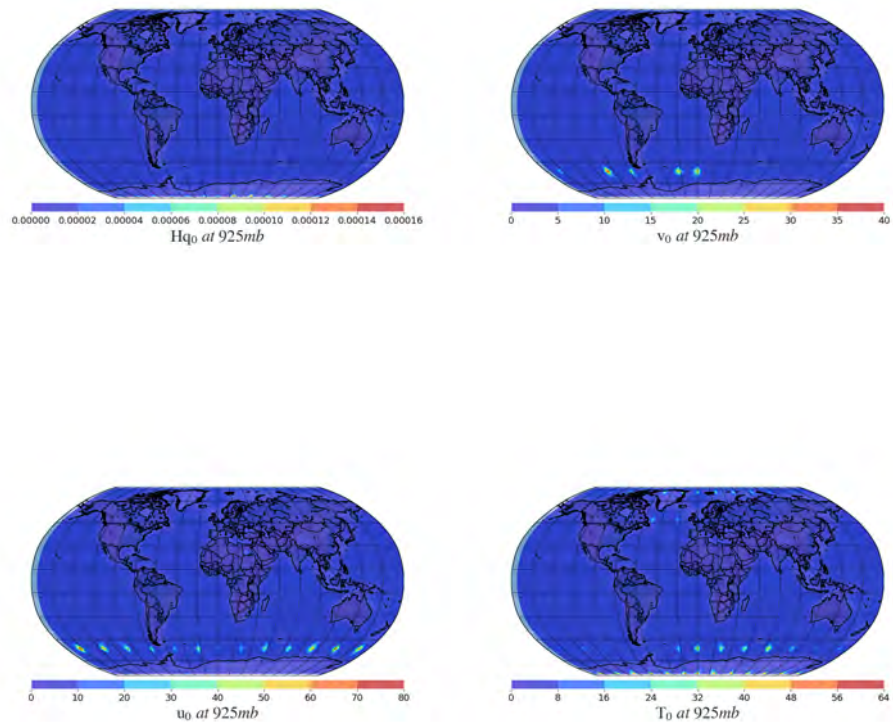


### 5.3.16. LEnKF $p = 6\%$ $\alpha = 1,08$

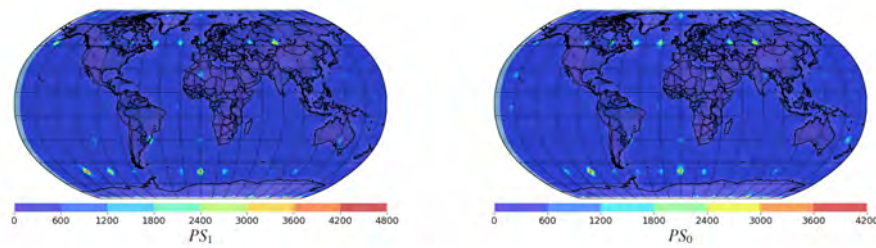




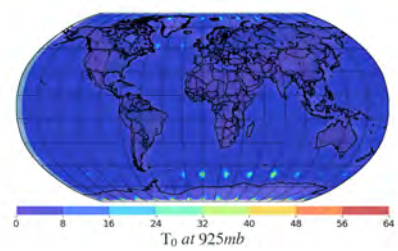
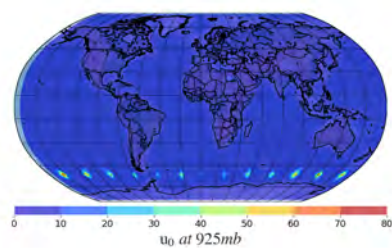
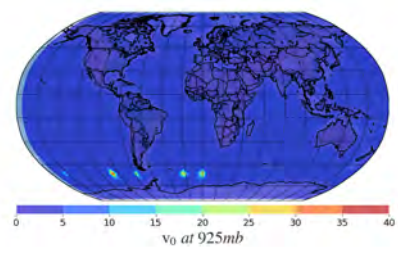
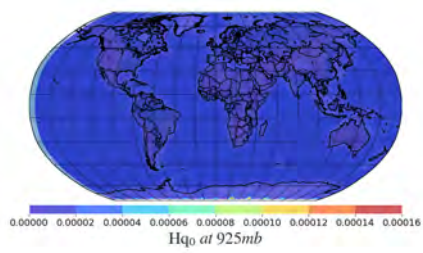
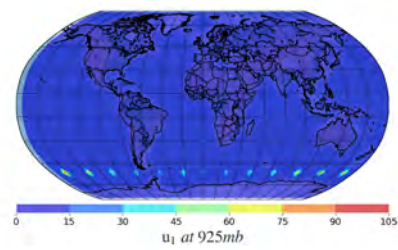
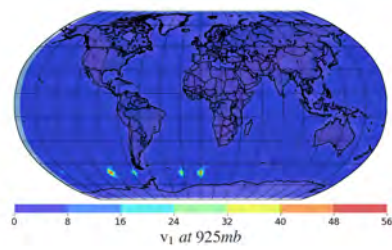
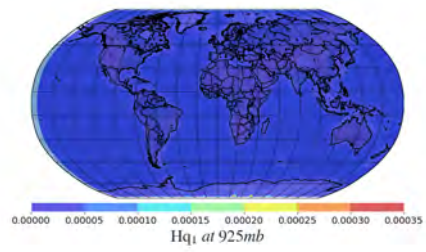
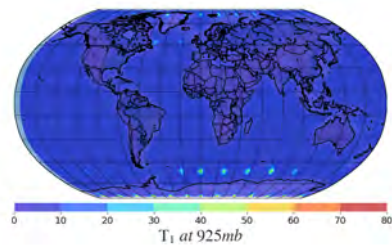
5.3.17. LEnKF  $p = 4\%$   $\alpha = 1,02$ 



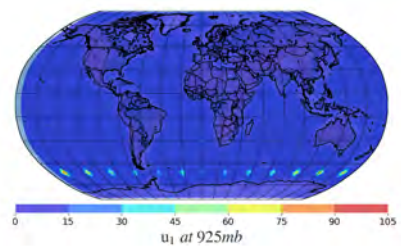
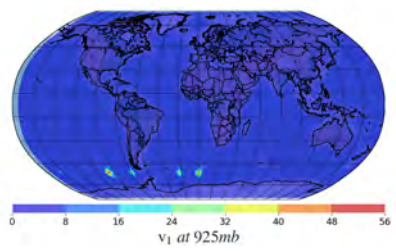
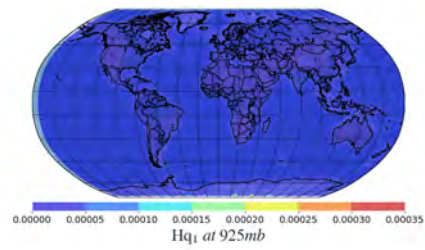
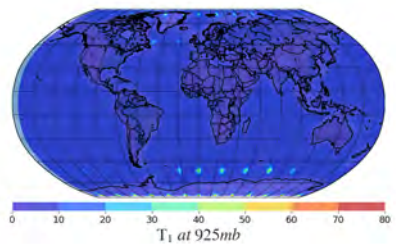
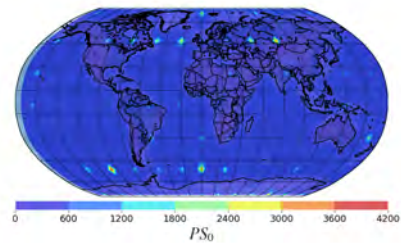
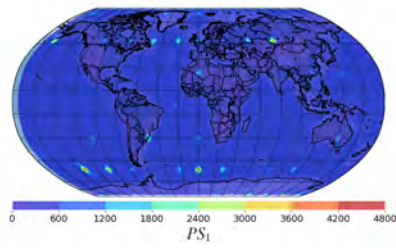
### 5.3.18. LEnKF $p = 4\%$ $\alpha = 1,04$

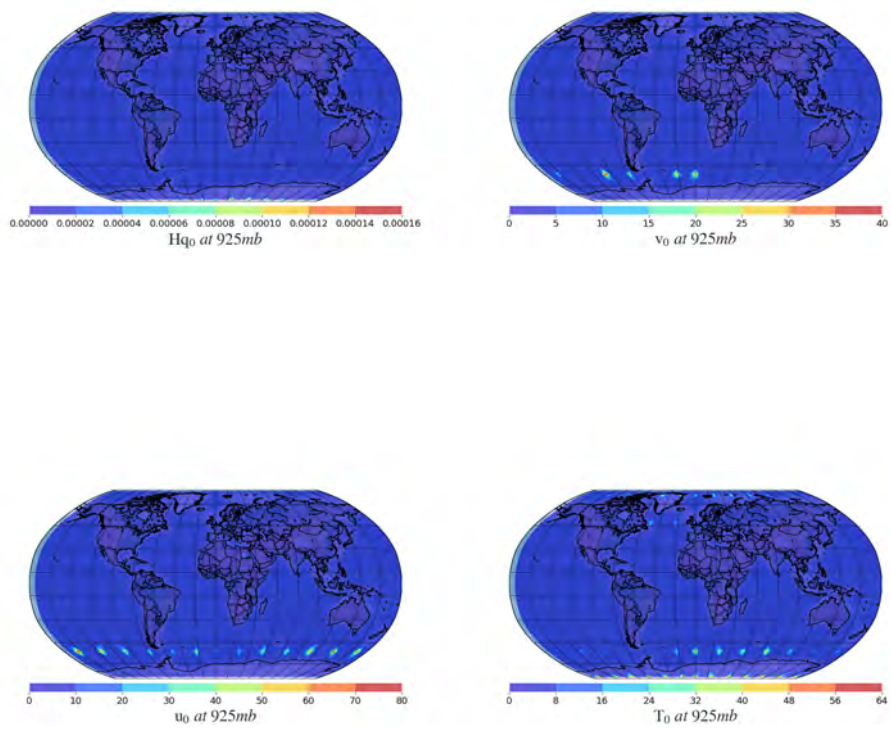




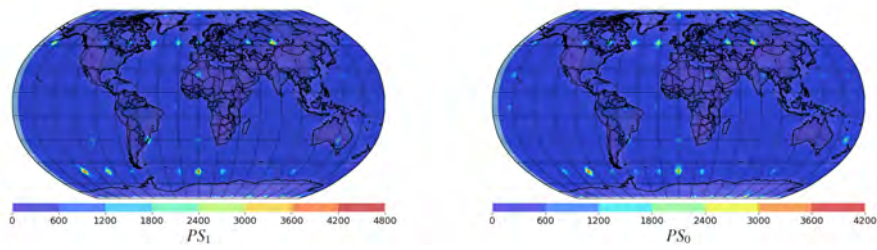


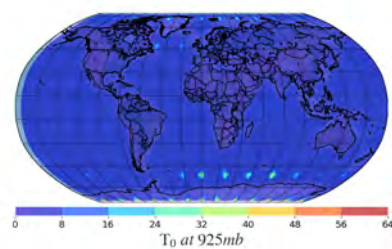
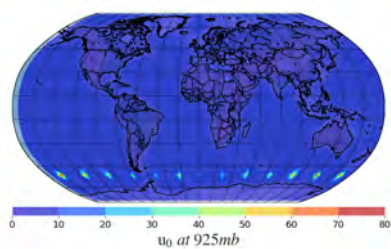
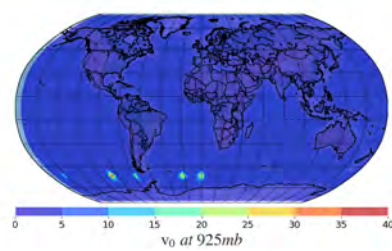
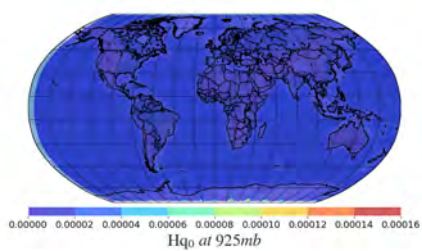
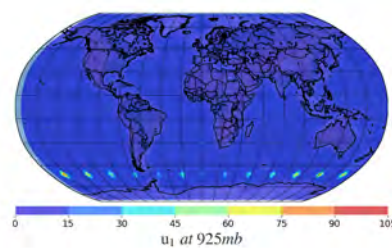
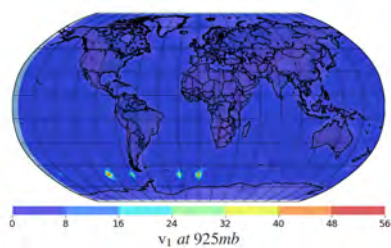
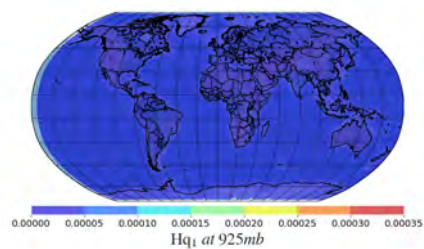
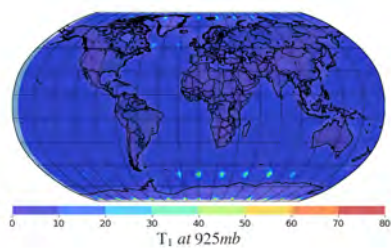


5.3.19. LEnKF  $p = 4\%$   $\alpha = 1,06$ 



### 5.3.20. LEnKF $p = 4\%$ $\alpha = 1,08$





## Capítulo 6

# Conclusions

Después de realizadas distintas pruebas se llegó a la conclusión que para un radio de uno y un peso del 100 % a las observaciones, nos da resultados con gran cantidad de ruido, algo que no pasa para un peso menor como lo es el 25 %. También se obtuvo que algoritmos como el Local Ensemble Transform Kalman Filter tuvo un tiempo de ejecución menor a diferencia de los otros algoritmos de asimilación como el Ensemble Kalman Filter basado en la descomposición de Cholesky Modificada y el Local Ensemble Kalman Filter, por lo cual se podría sugerir que para una mayor profundidad en este estudio o estudios próximos, se aconseja ampliar el radio de localización para las pruebas, así como también realizar estas en entornos computacionales con mayor capacidad, esto debido a que fue una limitante al momento de llevar a cabo los experimentos.

Por ultimo se recomienda en caso de usar esta implementación para una herramienta de alerta temprana configurar esta con los parámetros más óptimos obtenidos. Teniendo en cuenta que únicamente se mencionan la inflación de los datos, el radio de localización y el peso de la red observacional, podrían cambiar datos como el tiempo entre cada paso, o el tiempo entre la condición inicial y las observaciones. Para este caso, se obtuvo que algoritmo de asimilación Local Ensemble Transform Kalman Filter con  $\alpha = 1,04$  y  $p = 25\%$  obtuvo buenos resultados, además del poco tiempo que este lleva en realizar esta predicción en comparación de los otros algoritmos de asimilación previamente mencionados. En el trabajo utilizamos 30 miembros del conjunto para lograr un error bajo en las predicciones debido a las limitaciones de hardware que teníamos, pero se recomienda aumentar el paso en la medida de lo posible para lograr mejores resultados.





# Bibliografía

- Abaza, Mabrouk et al. (2017). «On the incidence of meteorological and hydrological processors: effect of resolution, sharpness and reliability of hydrological ensemble forecasts». En: *Journal of Hydrology*.
- Anderson, Jeffrey L (2001). «An ensemble adjustment Kalman filter for data assimilation». En: *Monthly weather review* 129.12, págs. 2884-2903.
- (2019). «A Nonlinear Rank Regression Method for Ensemble Kalman Filter Data Assimilation». En: *Monthly Weather Review* 147.8, págs. 2847-2860.
- Barker, D. M. et al. (2004). «A Three-Dimensional Variational Data Assimilation System for MM5: Implementation and Initial Results». En: *Monthly Weather Review* 132.4, págs. 897 -914. DOI: [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). URL: [https://journals.ametsoc.org/view/journals/mwre/132/4/1520-0493\\_2004\\_132\\_0897\\_atvdas\\_2.0.co\\_2.xml](https://journals.ametsoc.org/view/journals/mwre/132/4/1520-0493_2004_132_0897_atvdas_2.0.co_2.xml).
- Bauer, Peter, Alan Thorpe y Gilbert Brunet (sep. de 2015). «The quiet revolution of numerical weather prediction». En: *Nature* 525, págs. 47-55. DOI: [10.1038/nature14956](https://doi.org/10.1038/nature14956).
- Berger, James O. y Leonard A. Smith (2019). «On the Statistical Formalism of Uncertainty Quantification». En: *Annual Review of Statistics and Its Application* 6.1, págs. 433-460. DOI: [10.1146/annurev-statistics-030718-105232](https://doi.org/10.1146/annurev-statistics-030718-105232). eprint: <https://doi.org/10.1146/annurev-statistics-030718-105232>. URL: <https://doi.org/10.1146/annurev-statistics-030718-105232>.
- Bickel, Peter J, Elizaveta Levina et al. (2008). «Regularized estimation of large covariance matrices». En: *The Annals of Statistics* 36.1, págs. 199-227.
- Bracco, Annalisa et al. (2004). «Internal variability, external forcing and climate trends in multi-decadal AGCM ensembles». En: *Climate Dynamics* 23.6, págs. 659-678.
- Buehner, Mark (2011). «Evaluation of a Spatial/Spectral Covariance Localization Approach for Atmospheric Data Assimilation». En: *Monthly Weather Review* 140.2, págs. 617-636. DOI: [10.1175/MWR-D-10-05052.1](https://doi.org/10.1175/MWR-D-10-05052.1).
- Chen, Zhe (2003). «Bayesian Filtering: From Kalman Filters to Particle Filters, and Beyond». En: *Statistics* 182. DOI: [10.1080/02331880309257](https://doi.org/10.1080/02331880309257).
- Consuegra Ortega, Randy Steven et al. (2021). «Sequential data assimilation methods for atmospheric general circulation models». En.
- Conway, Thomas J et al. (1994). «Evidence for interannual variability of the carbon cycle from the National Oceanic and Atmospheric Administration/Climate Monitoring and Diagnostics Laboratory global air sampling network». En: *Journal of Geophysical Research: Atmospheres* 99.D11, págs. 22831-22855.
- Dovera, Laura y Ernesto Della Rossa (2011). «Multimodal ensemble Kalman filtering using Gaussian mixture models». En: *Computational Geosciences* 15.2, págs. 307-323.
- Evensen, Geir (1994). «Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics». En: *Journal of Geophysical Research: Oceans* 99.C5, págs. 10143-10162.
- Fan, Jianqing, Yuan Liao y Han Liu (2016). «An overview of the estimation of large covariance and precision matrices». En: *The Econometrics Journal* 19.1, págs. C1-C32. DOI: <https://doi.org/10.1111/ectj.12061>. eprint: <https://onlinelibrary>.

- wiley.com/doi/pdf/10.1111/ectj.12061. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1111/ectj.12061>.
- Gillijns, S et al. (2006). «What is the Ensemble Kalman Filter and How Well Does It Work?» En: *American Control Conference, 2006*, 6 pp.-. DOI: [10.1109/ACC.2006.1657419](https://doi.org/10.1109/ACC.2006.1657419).
- Godinez, Humberto C y J David Moulton (2012). «An efficient matrix-free algorithm for the ensemble Kalman filter». En: *Computational Geosciences* 16.3, págs. 565-575.
- Gustafsson, Nils y J Bojarova (2014). «Four-dimensional ensemble variational (4D-En-Var) data assimilation for the high resolution limited area model (HIRLAM)». En: *Nonlinear Processes in Geophysics* 21.4, págs. 745-762.
- Hamill, Thomas M, Jeffrey S Whitaker y Chris Snyder (2001). «Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter». En: *Monthly Weather Review* 129.11, págs. 2776-2790.
- Han, Yunfeng, Jucheng Zhang y Dajun Sun (2018). «Error control and adjustment method for underwater wireless sensor network localization». En: *Applied Acoustics* 130, págs. 293-299.
- Houtekamer, Peter L y Herschel L Mitchell (1998). «Data assimilation using an ensemble Kalman filter technique». En: *Monthly Weather Review* 126.3, págs. 796-811.
- Hunt, Brian R., Eric J. Kostelich e Istvan Szunyogh (2007). «Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter». En: *Physica D: Nonlinear Phenomena* 230.1. Data Assimilation, págs. 112-126. ISSN: 0167-2789. DOI: <https://doi.org/10.1016/j.physd.2006.11.008>. URL: <https://www.sciencedirect.com/science/article/pii/S0167278906004647>.
- Hurrell, J.W., C. Deser y A.S. Phillips (2019). «North Atlantic Oscillation (NAO)». En: *Encyclopedia of Ocean Sciences (Third Edition)*. Ed. por J. Kirk Cochran, Henry J. Bokuniewicz y Patricia L. Yager. Third Edition. Oxford: Academic Press, págs. 447-454. ISBN: 978-0-12-813082-7. DOI: <https://doi.org/10.1016/B978-0-12-409548-9.11621-5>. URL: <https://www.sciencedirect.com/science/article/pii/B9780124095489116215>.
- Ito, Shin-ichi et al. (2016). «Data assimilation for massive autonomous systems based on a second-order adjoint method». En: *Physical Review E* 94.4, págs. 43307.
- Kalman, R. E. (mar. de 1960). «A new approach to linear filtering and prediction problems». En: *Journal of Fluids Engineering, Transactions of the ASME* 82.1, págs. 35-45. ISSN: 1528901X. DOI: [10.1115/1.3662552](https://doi.org/10.1115/1.3662552). URL: <https://doi.org/10.1115/1.3662552>.
- Kalman, R. E. y R. S. Bucy (mar. de 1961). «New results in linear filtering and prediction theory». En: *Journal of Fluids Engineering, Transactions of the ASME* 83.1, págs. 95-108. ISSN: 1528901X. DOI: [10.1115/1.3658902](https://doi.org/10.1115/1.3658902). URL: <https://doi.org/10.1115/1.3658902>.
- Kalnay, Eugenia (2002). *Atmospheric Modeling, Data Assimilation and Predictability*. Cambridge University Press. DOI: [10.1017/CB09780511802270](https://doi.org/10.1017/CB09780511802270).
- Kondo, Keiichi y Takemasa Miyoshi (2016). «Impact of Removing Covariance Localization in an Ensemble Kalman Filter: Experiments with 10 240 Members Using an Intermediate AGCM». En: *Monthly Weather Review* 144.12, págs. 4849-4865. DOI: [10.1175/MWR-D-15-0388.1](https://doi.org/10.1175/MWR-D-15-0388.1). URL: <https://doi.org/10.1175/MWR-D-15-0388.1>.
- Kucharski, Fred, Franco Molteni y Annalisa Bracco (2006). «Decadal interactions between the western tropical Pacific and the North Atlantic Oscillation». En: *Climate dynamics* 26.1, págs. 79-91.
- Kwiatkowski, Evan y Jan Mandel (abr. de 2015). «Convergence of the Square Root Ensemble Kalman Filter in the Large Ensemble Limit». En: *SIAM/ASA Journal on*

- Uncertainty Quantification* 3.1, págs. 1-17. DOI: [10.1137/140965363](https://doi.org/10.1137/140965363). URL: <http://arxiv.org/abs/1404.4093><http://dx.doi.org/10.1137/140965363>.
- Lei, Lili, Jeffrey S Whitaker y Craig Bishop (2018). «Improving assimilation of radiance observations by implementing model space localization in an ensemble Kalman filter». En: *Journal of Advances in Modeling Earth Systems* 10.12, págs. 3221-3232.
- Levina, Elizaveta, Adam Rothman, Ji Zhu et al. (2008). «Sparse estimation of large covariance matrices via a nested lasso penalty». En: *The Annals of Applied Statistics* 2.1, págs. 245-263.
- Li, Hua et al. (2012). «Evaluation of MODIS and NCEP atmospheric products for land surface temperature retrieval from HJ-1B IRS thermal infrared data with ground measurements». En: *2012 IEEE International Geoscience and Remote Sensing Symposium*. IEEE, págs. 5057-5060.
- Liu, Zhenkun et al. (nov. de 2019). «A combined forecasting model for time series: Application to short-term wind speed forecasting». En: *Applied Energy*, pág. 114137. ISSN: 0306-2619. DOI: [10.1016/J.APENERGY.2019.114137](https://doi.org/10.1016/j.apenergy.2019.114137). URL: <https://www.sciencedirect.com/science/article/abs/pii/S0306261919318240>.
- Lorenc, A. C. (1986). «Analysis methods for numerical weather prediction». En: *Quarterly Journal of the Royal Meteorological Society* 112.474, págs. 1177-1194. DOI: <https://doi.org/10.1002/qj.49711247414>. eprint: <https://rmets.onlinelibrary.wiley.com/doi/pdf/10.1002/qj.49711247414>. URL: <https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.49711247414>.
- Lorenc, Andrew C (2003a). «Modelling of error covariances by 4D-Var data assimilation». En: *Quarterly Journal of the Royal Meteorological Society: A journal of the atmospheric sciences, applied meteorology and physical oceanography* 129.595, págs. 3167-3182.
- (2003b). «The potential of the ensemble Kalman filter for NWP—a comparison with 4D-Var». En: *Quarterly Journal of the Royal Meteorological Society: A journal of the atmospheric sciences, applied meteorology and physical oceanography* 129.595, págs. 3183-3203.
- Miyoshi, Takemasa (2011). «The Gaussian approach to adaptive covariance inflation and its implementation with the local ensemble transform Kalman filter». En: *Monthly Weather Review* 139.5, págs. 1519-1535.
- Miyoshi, Takemasa, Keiichi Kondo y Toshiyuki Imamura (2014). «The 10,240-member ensemble Kalman filtering with an intermediate AGCM». En: *Geophysical Research Letters* 41.14, págs. 5264-5271.
- Molteni, Franco (ene. de 2003a). «Atmospheric simulations using a GCM with simplified physical parametrizations. I: Model climatology and variability in multi-decadal experiments». En: *Climate Dynamics* 20, págs. 175-191. DOI: [10.1007/s00382-002-0268-2](https://doi.org/10.1007/s00382-002-0268-2).
- (2003b). «Atmospheric simulations using a GCM with simplified physical parametrizations. I: Model climatology and variability in multi-decadal experiments». En: *Climate dynamics* 20.2, págs. 175-191.
- National Centers for Environmental Prediction, National Weather Service, NOAA, U.S. Department of Commerce (2000). *NCEP FNL Operational Model Global Tropospheric Analyses, continuing from July 1999*. Boulder CO. URL: <https://doi.org/10.5065/D6M043C6>.
- Nino-Ruiz, Elias D (2018). «Implicit surrogate models for trust region based methods». En: *Journal of Computational Science*.
- Nino-Ruiz, Elias D, Haiyan Cheng y Rolando Beltran (2018). «A Robust Non-Gaussian Data Assimilation Method for Highly Non-Linear Models». En: *Atmosphere* 9.4, pág. 126.

- Nino-Ruiz, Elias D, Adrian Sandu y Xinwei Deng (2017). «A parallel implementation of the ensemble Kalman filter based on modified Cholesky decomposition». En: *Journal of Computational Science*.
- (2018). «An ensemble Kalman filter implementation based on modified Cholesky decomposition for inverse covariance matrix estimation». En: *SIAM Journal on Scientific Computing* 40.2, A867-A886.
- Nino Ruiz, EliasD., Adrian Sandu y Jeffrey Anderson (2014). «An Efficient Implementation of the Ensemble Kalman Filter Based on an Iterative Sherman–Morrison Formula». En: *Statistics and Computing*, págs. 1-17. ISSN: 0960-3174. DOI: [10.1007/s11222-014-9454-4](https://doi.org/10.1007/s11222-014-9454-4). URL: <http://dx.doi.org/10.1007/s11222-014-9454-4>.
- Ott, Edward et al. (2004a). «A local ensemble Kalman filter for atmospheric data assimilation». En: *Tellus A: Dynamic Meteorology and Oceanography* 56.5, págs. 415-428. DOI: [10.3402/tellusa.v56i5.14462](https://doi.org/10.3402/tellusa.v56i5.14462). eprint: <https://doi.org/10.3402/tellusa.v56i5.14462>. URL: <https://doi.org/10.3402/tellusa.v56i5.14462>.
- (2004b). «A local ensemble Kalman filter for atmospheric data assimilation». En: *Tellus A: Dynamic Meteorology and Oceanography* 56.5, págs. 415-428. DOI: [10.3402/tellusa.v56i5.14462](https://doi.org/10.3402/tellusa.v56i5.14462). eprint: <https://doi.org/10.3402/tellusa.v56i5.14462>. URL: <https://doi.org/10.3402/tellusa.v56i5.14462>.
- Pourahmadi, Mohsen (2011). «Covariance Estimation: The GLM and Regularization Perspectives». En: *Statistical Science* 26.3. ISSN: 0883-4237. DOI: [10.1214/11-sts358](https://doi.org/10.1214/11-sts358). URL: <http://dx.doi.org/10.1214/11-STS358>.
- Saetrom, Jon y Henning Omre (dic. de 2013). «Uncertainty Quantification in the Ensemble Kalman Filter». En: *Scandinavian Journal of Statistics* 40. DOI: [10.1111/sjos.12039](https://doi.org/10.1111/sjos.12039).
- Stengel, M et al. (2009). «Assimilation of SEVIRI infrared radiances with HIRLAM 4D-Var». En: *Quarterly Journal of the Royal Meteorological Society: A journal of the atmospheric sciences, applied meteorology and physical oceanography* 135.645, págs. 2100-2109.
- Stroud, Jonathan R, Matthias Katzfuss y Christopher K Wikle (2018). «A Bayesian adaptive ensemble Kalman filter for sequential state and parameter estimation». En: *Monthly Weather Review* 146.1, págs. 373-386.
- Teixeira, Joao et al. (2014). «Atmospheric General Circulation Models». En: *Encyclopedia of Remote Sensing*. New York, NY: Springer New York, págs. 35-37. ISBN: 978-0-387-36699-9. DOI: [10.1007/978-0-387-36699-9\\_8](https://doi.org/10.1007/978-0-387-36699-9_8). URL: [https://doi.org/10.1007/978-0-387-36699-9\\_8](https://doi.org/10.1007/978-0-387-36699-9_8).
- Tippett, Michael K et al. (jul. de 2003). «Ensemble Square Root Filters». En: *Monthly Weather Review* 131.7, págs. 1485-1490. ISSN: 0027-0644. DOI: [10.1175/1520-0493\(2003\)131<1485:ESRF>2.0.CO;2](https://doi.org/10.1175/1520-0493(2003)131<1485:ESRF>2.0.CO;2).
- Tong, Xin T (2018). «Performance analysis of local ensemble Kalman filter». En: *Journal of Nonlinear Science* 28.4, págs. 1397-1442.
- Verstraete, Gyliau, El-Houssaine Aghezzaf y Bram Desmet (ene. de 2020). «A leading macroeconomic indicators' based framework to automatically generate tactical sales forecasts». En: *Computers & Industrial Engineering* 139, págs. 106169. ISSN: 0360-8352. DOI: [10.1016/J.CIE.2019.106169](https://www.sciencedirect.com/science/article/pii/S0360835219306382). URL: <https://www.sciencedirect.com/science/article/pii/S0360835219306382>.
- Vetra-Carvalho, Sanita et al. (2018). «State-of-the-art stochastic data assimilation methods for high-dimensional non-Gaussian problems». En: *Tellus A: Dynamic Meteorology and Oceanography* 70.1, págs. 1-43. DOI: [10.1080/16000870.2018.1445364](https://doi.org/10.1080/16000870.2018.1445364). URL: <https://doi.org/10.1080/16000870.2018.1445364>.
- Wikle, Christopher K y L Mark Berliner (2007). «A Bayesian tutorial for data assimilation». En: *Physica D: Nonlinear Phenomena* 230.1-2, págs. 1-16.