



TRABAJO FIN DE MASTER: TIPO B

Evaluación de la Incertidumbre de los Parámetros y su efecto sobre la Incertidumbre Predictiva de un Modelo Hidrológico Distribuido, mediante un algoritmo *Markov Chain Monte Carlo* de tipo Adaptativo.

Intensificación: Sistemas de Recursos Hídricos

*Autor: Mario R. Hernández López
Director: Félix Francés García*



Evaluación de la Incertidumbre de los Parámetros y su efecto sobre la Incertidumbre Predictiva de un Modelo Hidrológico Distribuido, mediante un algoritmo *Markov Chain Monte Carlo* de tipo Adaptativo.

INDICE DE CONTENIDOS

1- *Introducción*

2- *La Distribución Posterior de los Parámetros*

3- *Algoritmos Markov Chain Monte Carlo*

4- *Caso Práctico*

5- *Conclusiones*





Evaluación de la Incertidumbre de los Parámetros y su efecto sobre la Incertidumbre Predictiva de un Modelo Hidrológico Distribuido, mediante un algoritmo *Markov Chain Monte Carlo* de tipo Adaptativo.

Evaluación de la Incertidumbre ...

Evaluación de la Incertidumbre de los Parámetros ...

... Incertidumbre Predictiva de un Modelo ...

... un algoritmo Markov Chain Monte Carlo de tipo Adaptativo.



Evaluación de la Incertidumbre...

❑ UNA DEFINICIÓN [Refsgaard et al. 2007]

UNA PROPIEDAD DE LA INFORMACIÓN EMPLEADA PARA TOMAR DECISIONES, O UNA MANIFESTACIÓN DE LA CONFIANZA QUE SE TIENE EN DICHA INFORMACIÓN

❑ HIDROLOGÍA

ES UNA CIENCIA INCIERTA POR FALTA DE CONOCIMIENTO DEL SISTEMA ANALIZADO, POR...

- ESCASEZ DE INFORMACIÓN
- DESCONOCIMIENTO DE LAS MECÁNICAS DE ALGUNOS PROCESOS

❑ FUENTES DE INCERTIDUMBRE EN MODELACIÓN HIDROLÓGICA [Mattot et al. 2009]

❑ EL MODELADOR

❑ LOS INPUTS

➤ FORCING DATA

➤ PARÁMETROS NO CALIBRADOS

➤ OBSERVACIONES DE LA RESPUESTA

❑ EL MODELO

➤ ESTRUCTURAL

➤ DE RESOLUCIÓN

➤ DE RESOLUCIÓN MUESTRAL

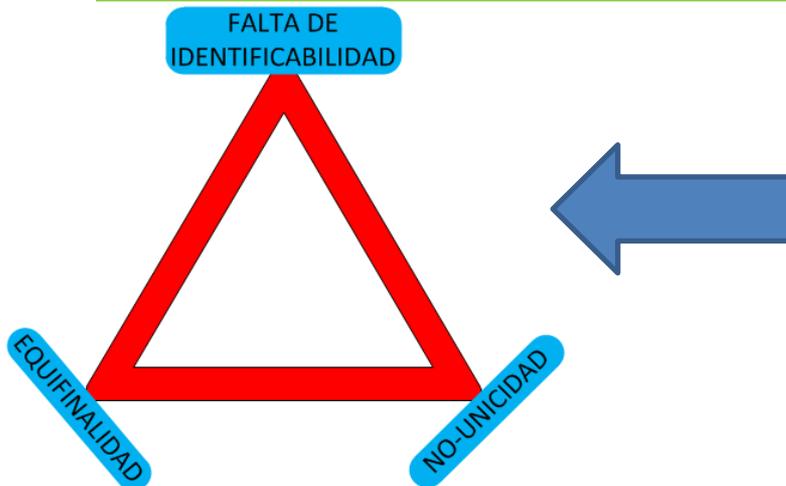


Temas Relacionados



ESA VARIEDAD DE TEMAS CLÁSICOS INTENTAN SOLUCIONAR PROBLEMAS QUE SE PODRÍAN RESUMIR EN UNO SOLO ...

Evaluación de la Incertidumbre de los Parámetros...



TIENEN SU ORIGEN EN...

LA PROYECCIÓN DE LAS DISTINTAS FUENTES DE INCERTIDUMBRE SOBRE LOS PARÁMETROS CALIBRADOS

PUESTO QUE EN SU VALOR HAY INCERTIDUMBRE...

❑ NINGÚN HIDRÓLOGO DEBERÍA CONFIAR EN UN ÚNICO CONJUNTO DETERMINISTA DE LOS VALORES DE LOS PARÁMETROS [KUKZERA & PARENT, 1998]



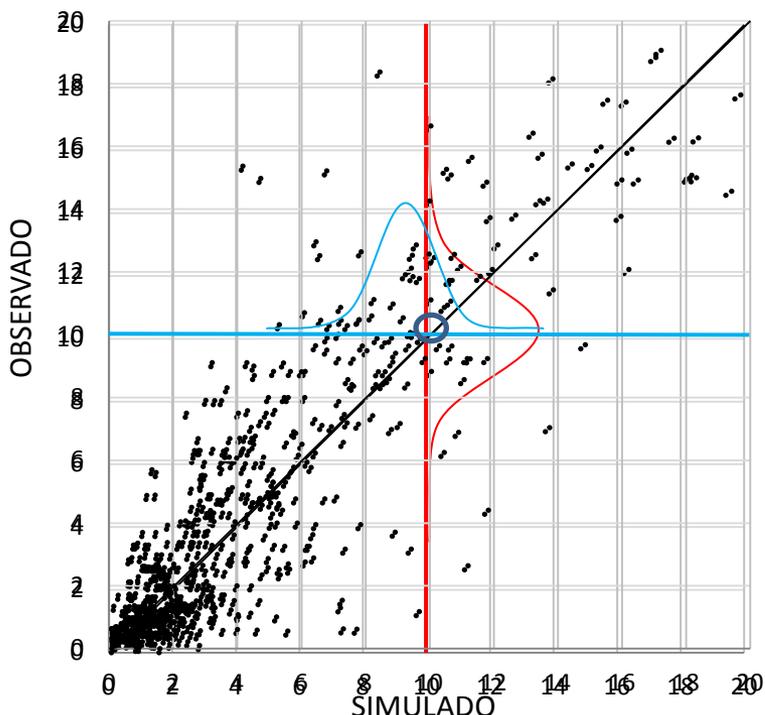
Evaluación de la Incertidumbre de los Parámetros y su efecto sobre

La Incertidumbre Predictiva de un Modelo...

INCERTIDUMBRE PREDICTIVA DEL MODELO

➤ INCERTIDUMBRE DE LA OBSERVACIÓN, DADO EL RESULTADO DEL MODELO

➤ “ADICIÓN” DE UN **MODELO DE ERROR** AL RESULTADO DEL MODELO HIDROLÓGICO



MODELO DE REGRESIÓN NO - LINEAL

OBSERVACIONES

PREDICCIÓN MODELO
HIDROLÓGICO

$$\tilde{Y} = E[Y | \theta, \tilde{X}] + \varepsilon_y$$

MODELO DE ERROR:

MODELA LA VARIABILIDAD DE LA RESPUESTA OBSERVADA, QUE NO ES CAPAZ DE EXPLICAR EL MODELO HIDROLÓGICO

INCERTIDUMBRE REPRODUCTIVA DEL MODELO

➤ INCERTIDUMBRE DEL SIMULADO, DADO EL OBSERVADO

➤ PROPAGACIÓN DE LA **INCERTIDUMBRE DE LOS PARÁMETROS** A TRAVÉS DEL MODELO



1- La Posterior: Evaluación de la Incertidumbre de los Parámetros

2- Métodos de Estimación de la Posterior

2.1- Métodos MCMC

2.2- Fundamentos de MCMC

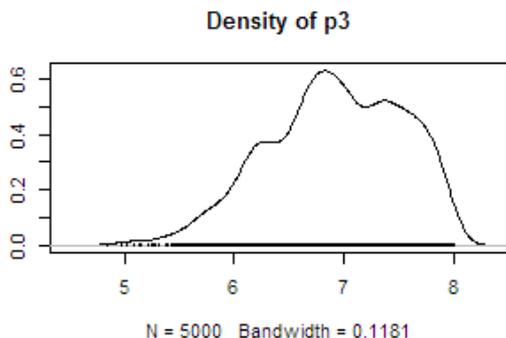


Evaluación de la Incertidumbre de los Parámetros...

LA POSTERIOR DE LOS PARAMETROS

DIST.PROBABILIDAD BASADA EN EL TMA. BAYES...

$$p(\theta, \eta | \tilde{Y}, \tilde{X}) = \frac{p(\tilde{Y} | \theta, \eta, \tilde{X}) p(\theta, \eta)}{p(\tilde{Y})} = T(\theta, \eta)$$



PROBABILIDAD CONJUNTA OBSERVACIONES-PARÁMETROS

$$p[(\theta, \eta), \tilde{Y}] = p(\tilde{Y} | \theta, \eta, \tilde{X}) p(\theta, \eta)$$

VEROSIMILITUD DE LOS PARÁMETROS

$$p(\tilde{Y} | \theta, \eta, \tilde{X}) = L(\theta, \eta | \tilde{Y}, \tilde{X}) = p(\boldsymbol{\varepsilon}_y)$$

PRIOR DE LOS PARÁMETROS

$$p(\theta, \eta)$$

PROBABILIDAD MARGINAL DE LAS OBSERVACIONES

$$p(\tilde{Y}) = \int p(\tilde{Y}, \theta, \eta) d\theta = \int p(\tilde{Y} | \theta, \eta, \tilde{X}) p(\theta, \eta) d\theta = K$$

FACTOR DE ESCALA

MAYOR PROBLEMA DEL CÁLCULO COMPUTACIONAL BAYESIANO

Métodos de Estimación de la Posterior

❑ ASIMILACIÓN DE LA POSTERIOR A UNA DISTRIBUCIÓN MULTINORMAL

EXCESIVA SIMPLIFICACIÓN EN HIDROLOGÍA POR LA ALTA NO-LINEALIDAD (DUAN ET AL 1992)

❑ BASADOS EN TEOREMA DE BAYES...

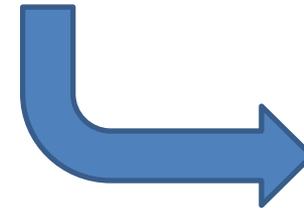
➤ CÁLCULO ANALÍTICO DE LA POSTERIOR

- EN PROBLEMAS SENCILLOS

➤ MÉTODOS MONTE CARLO DIRECTOS

- POCO EFICIENTES EN ESPACIOS PARAMÉTRICOS DE ALTA DIMENSIÓN

➤ **MÉTODOS MARKOV CHAIN MONTE CARLO (MCMC)**



Métodos *Markov Chain Monte Carlo*



AUTOR

Nicholas Metropolis 1915-1999
Físico Griego-Americano



OBJETO

- ❑ **AÑO 1953**
- ❑ **CÁLCULO DE INTEGRALES DE ALTA DIMENSIÓN** CUYA INTRATABILIDAD ANALÍTICA REQUIERE DE ...

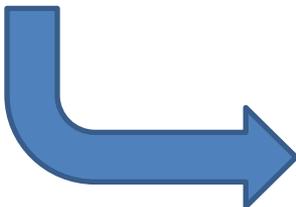
➤ **LA INTEGRACIÓN MONTE CARLO DE MUESTRAS ...**

➤ **... EXTRAÍDAS MEDIANTE UNA CADENA DE MARKOV**



**CÁLCULO DE LA POSTERIOR DE LOS PARAMETROS
EN ESPACIOS PARAMÉTRICOS DE ALTA DIMENSIÓN**

FUNDAMENTOS



Fundamentos de los Métodos MCMC



TEORÍA BAYESIANA FORMAL

INTEGRACIÓN MONTE CARLO

CADENAS DE MARKOV

NECESITA DE...

UN MODELO

(p.e. HIDROLÓGICO)

PARAMETROS

$$Y_h = h(\theta, \tilde{X})$$

SIMULADOS

OBSERVACIONES FORCING DATA

EL MODELO DE ERROR

$$\mathcal{E}_y(\eta) = f[e_i; \xi(0,1)]$$

PARAMETROS

$$\begin{cases} e_i = \tilde{Y}_i - Y_{h_i} \\ \xi(0,1) \end{cases}$$

Residuos

Ruido Blanco

INCERTIDUMBRE PREDICTIVA

LA FUNCIÓN DE VEROSIMILITUD

$$p(\tilde{Y} | \theta, \eta, \tilde{X}) = L(\theta, \eta | \tilde{Y}, \tilde{X}) = p(\mathcal{E}_y)$$

LA POSTERIOR DE LOS PARAMETROS

$$p(\theta, \eta | \tilde{Y}, \tilde{X}) = \frac{p(\tilde{Y} | \theta, \eta, \tilde{X}) p(\theta, \eta)}{p(\tilde{Y})} = T(\theta, \eta)$$

Fundamentos de los Métodos MCMC

TEORÍA BAYESIANA FORMAL

INTEGRACIÓN
MONTE CARLO

CADENAS DE MARKOV

- **CÁLCULO APROXIMADO** DE LA ESPERANZA DE UNA FUNCIÓN DE LOS PARÁMETROS :
➤ MEDIA, VARIANZA,...

SEA UNA FUNCIÓN DE LA POSTERIOR

$$f(\theta | \tilde{Y}, \tilde{X})$$

ESTIMACIÓN MONTE CARLO DE SU ESPERANZA

$$E[f(\theta | \tilde{Y}, \tilde{X})] = \int f(\theta | \tilde{Y}, \tilde{X}) T(\theta) d\theta \approx \frac{1}{n} \sum_{j=1}^n f(\theta_j)$$

MUESTREO DE SU DISTRIBUCIÓN



MCSE

PRECISIÓN DE
LA ESTIMACIÓN

(Tamaño de Muestra)

(VARIANZA DEL ESTIMADOR)

Fundamentos de los Métodos MCMC

CADENAS DE MARKOV

DEFINICIÓN

- ❑ **PROCESO ESTOCÁSTICO** DISCRETO EN EL TIEMPO QUE SE MUEVE EN TORNO A UN CONJUNTO DE POSIBLES VALORES LLAMADO **ESPACIO DE ESTADOS (CONTINUO)**
- ❑ **LA PRINCIPAL PROPIEDAD ES LA FALTA DE MEMORIA: EL ESTADO FUTURO DEL PROCESO SOLO DEPENDE DEL ESTADO PRESENTE**

SU PAPEL ES SER EL "MUESTREADOR" DE LA DISTRIBUCIÓN POSTERIOR

PROPIEDADES NECESARIAS

- ❑ **ERGODICIDAD DE LA CADENA**
- ❑ **EXISTENCIA DE LA DISTRIBUCIÓN DE EQUILIBRIO DE LA CADENA**
- ❑ **REVERSIBILIDAD DE LA CADENA**

❑ **ECUACIÓN DE BALANCE DETALLADO**

PARA UNA CADENA DE MARKOV REVERSIBLE Y ERGÓDICA...

DISTRIB. CANDIDATA

➤ DISTRIBUCIÓN DE PROBABILIDAD DE TRANSICIÓN **q** QUE SATISFAGA ESTA ECUACIÓN...

➤ ...LA CADENA TENDRÁ COMO **DISTRIBUCIÓN DE EQUILIBRIO**

DISTRIB. POSTERIOR

$$T(\theta | y) \cdot q(\theta, \theta') = T(\theta' | y) \cdot q(\theta', \theta)$$



1- Eficiencia de los Algoritmos: Factores

2- Algoritmo de Metropolis

3- Algoritmos Adaptativos

... mediante un algoritmo Markov Chain Monte Carlo de tipo Adaptativo.

Factores que influyen en la Eficiencia

❑ LA PRECISIÓN MONTE CARLO REQUERIDA EN LA INFERENCIA

❑ LA DISTRIBUCIÓN CANDIDATA O GENERADORA O “ DE SALTO” QUE SE EMPLEA

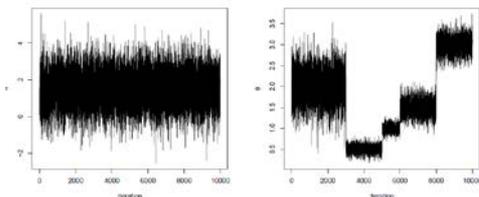
$$q(\theta_{t+1} | \theta_t) = q(\theta_t, \theta_{t+1})$$

➤ SUS PROPIEDADES DE MEZCLADO

➤ MUESTREO CON PROPORCIONES CORRECTAS

VELOCIDAD DE EXPLORACIÓN
DEL ESPACIO DE PARÁMETROS

CAPACIDAD DE EVITAR LOS ATRACTORES LOCALES



❑ LAS CARACTERÍSTICAS DEL ESPACIO DE PARÁMETROS

LOS MODELOS HIDROLÓGICOS COMPLEJOS,
PRESENTAN UNA SUPEFICIE DE RESPUESTA DE
LOS PARAMETROS CON MUCHOS SUB-ÓPTIMOS
O ATRACTORES LOCALES (DUAN ET AL. 1992)

**NÚMERO DE SIMULACIONES NECESARIAS PARA OBTENER
UNA MUESTRA ACEPTABLE DE LA POSTERIOR**



Algoritmo de Metropolis

El algoritmo de Metropolis está considerado entre los 10 más importantes del S.XX
(*Computing in Science and Engineering*)

❑ ES EL ALGORITMO BÁSICO DE LOS MÉTODOS MCMC

❑ EMPLEA UNA DIST. CANDIDATA MULTINORMAL Y SIMÉTRICA (RANDOM WALK)

❑ SU LOGRO CONCEPTUAL: LA PROBABILIDAD DE ACEPTACIÓN DE MOVIMIENTO (α)

$$\alpha(\theta_{t+1} | \theta_t) = \text{Min} \left[\frac{T(\theta_{t+1})}{T(\theta_t)}, 1 \right]$$

➤ ESTA PROBABILIDAD ES LA QUE EVOLUCIONA LA CADENA ("MOTOR" DEL MÉTODO)

X => **Y1**

MOVIMIENTO ACEPTADO CON PROB=1

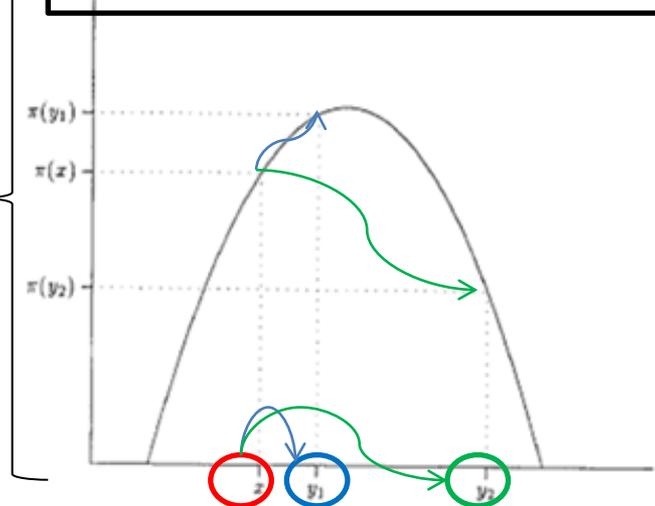
X => **Y2**

MOVIMIENTO ACEPTADO CON PROB<1

➤ EVITA EL CÁLCULO DEL FACTOR DE ESCALA DE LA POSTERIOR

$$p(\theta | \tilde{Y}, \tilde{X}) = \frac{p(\tilde{Y} | \theta, \tilde{X})p(\theta)}{p(\tilde{Y})} = \frac{\pi(\theta)}{p(\tilde{Y})} = T(\theta)$$

EL "MOTOR" DE LA CADENA DE MARKOV NO ES OPTIMIZAR UNA F.OBJETIVO





Algoritmos Adaptativos

❑ OBJETO: MEJORAR LA EFICIENCIA DEL ALGORITMO METROPOLIS

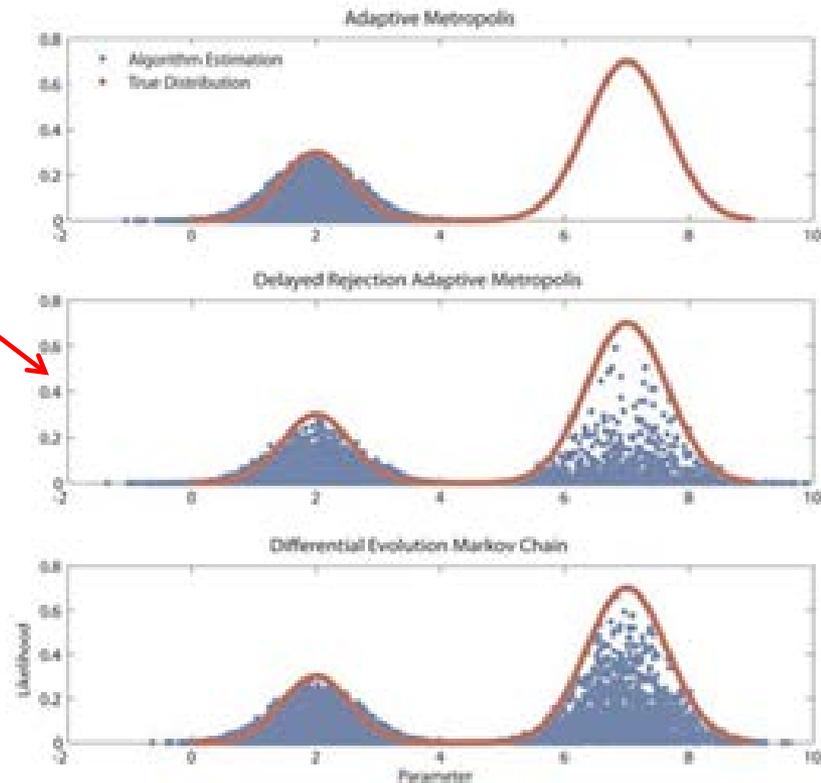
❑ ¿CÓMO?: ADAPTACIÓN DINÁMICA DE LA DISTRIBUCIÓN CANDIDATA “MIRANDO” LA EVOLUCIÓN DE LA CADENA

❑ EJEMPLO DE EFICIENCIA EXPLORATORIA
[Smith and Marshall, 2008]

❑ ALGORITMO DRAM
[Haario, 2005]

ADAPTACIÓN LOCAL
(**D**elayed **R**ejection)

ADAPTACIÓN GLOBAL
(**A**daptive **M**etropolis)





1- Herramientas y Área Estudio

2- Incertidumbre de los Parámetros

2.1- Incertidumbre Reproductiva

2.2- Sensibilidad Global

3- Incertidumbre Predictiva





Herramientas y Área de Estudio

HERRAMIENTA DE ANÁLISIS HIDROLÓGICO

- ❑ EL MODELO DISTRIBUIDO CONCEPTUAL TETIS (*GIMHA, UPV*) 

- ❑ LA PRODUCCIÓN DE ESCORRENTÍA SE HACE MEDIANTE ESQUEMAS CONCEPTUALES SIMPLES ADAPTADOS A LA **ESCALA DE CELDA** Y AL **INTERVALO DE TIEMPO** QUE SE CONSIDERA EN LA MODELACIÓN

- ❑ PRESENTA UNA ESTRUCTURA DE **PARÁMETRO EFECTIVO** SEPARADA EN DOS PARTES:
 - VALOR ESTIMADO EN CADA CELDA: **MAPA DEL PARÁMETRO**

 - FUNCIÓN DE CORRECCIÓN: **FACTOR DE CORRECCIÓN GLOBAL DE CADA MAPA**

- ❑ **ESTIMACIÓN DEL PARÁMETRO** MEDIANTE CALIBRACIÓN DEL **FACTOR CORRECTOR**



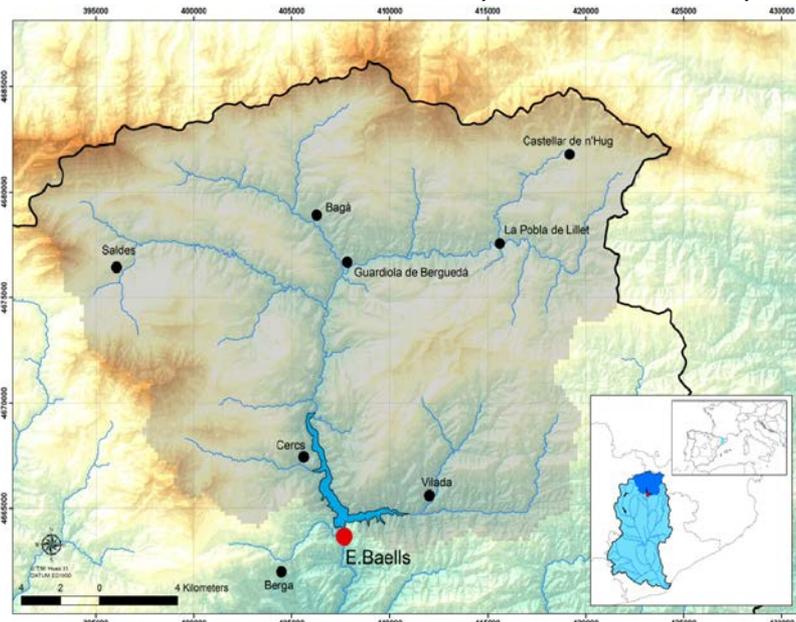
Herramientas y Área de Estudio

HERRAMIENTA DE ANÁLISIS MCMC

FME, Flexible Modeling Environment
[Soetaert and Petzoldt 2012; Wu and Liu 2012]

ÁREA DE ESTUDIO

CUENCA DE **LA BAELLS**, DE 505 KM², SUBCUENCA DEL LLOBREGAT (BARCELONA, ESPAÑA)



- CLIMA ALTA MONTAÑA - MEDITERRÁNEO
- VEGETACIÓN PREDOMINANTE: BOSQUES DE PINO
- CELDAS DE **200x200 m**
- MODELACIÓN A **ESCALA DIARIA**

EL PERIODO DE CALIBRACIÓN CONSTA DE **731 DÍAS**.

➤ ENTRE EL 1 DE OCTUBRE DE 2002 Y EL 30 DE SEPTIEMBRE DE 2004

PERIODO DE CALENTAMIENTO DE 1 AÑO (2001-2002)

Incertidumbre de los Parámetros

METODOLOGÍA

EN 4 PASOS ...

1- ANÁLISIS EXPLORATORIO DE PARÁMETROS: A. DE SENSIBILIDAD LOCAL

2- SIMULACIÓN MCMC

3- TESTEO DE LA SIMULACIÓN MCMC

TESTEO DE ESTACIONARIEDAD
DE LA CADENA

TRATAMIENTO DE LA NO
INDEPENDENCIA MUESTRAL

4- INFERENCIA SOBRE LA MUESTRA DE LA POSTERIOR DE LOS PARÁMETROS



Incertidumbre de los Parámetros

ANÁLISIS EXPLORATORIO DE PARÁMETROS: SENSIBILIDAD LOCAL

OBJETIVO: DESCARTAR PARÁMETROS QUE NO SON ÚTILES EN EL MODELO



CRITERIO DE EVALUACIÓN: LA FUNCIÓN **SSR** (suma residuos al cuadrado)

ANÁLISIS DE LA VARIACIÓN “LOCAL” DE LA **SSR**

GRÁFICAS PARÁMETRO - **SSR**

¿LOCAL?:

➤ SE PARTE DE UN PUNTO CONCRETO DEL ESPACIO DE PARÁMETROS

➤ MODIFICACIÓN DE CADA PARÁMETRO, DEJANDO FIJOS EL RESTO

▪ SOLO NOS MOVEMOS EN UNA DIRECCIÓN DENTRO DEL HIPERESPACIO DE PARÁMETROS

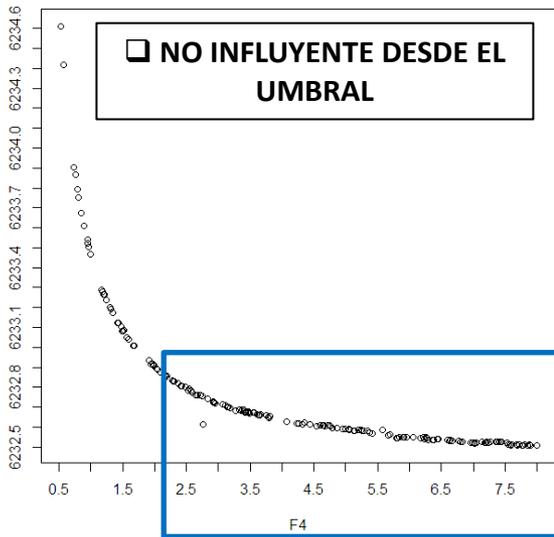
➤ **NO** SE TIENEN EN CUENTA POSIBLES **INTERACCIONES** ENTRE PARÁMETROS

Incertidumbre de los Parámetros

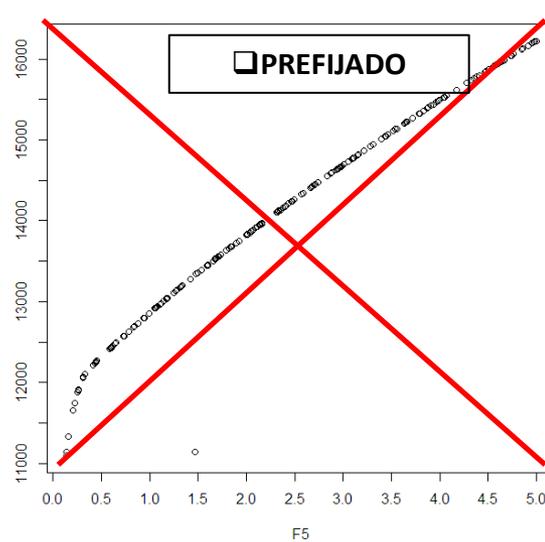
ANÁLISIS EXPLORATORIO DE PARÁMETROS: SENSIBILIDAD LOCAL

GRÁFICAS PARÁMETRO - SSR

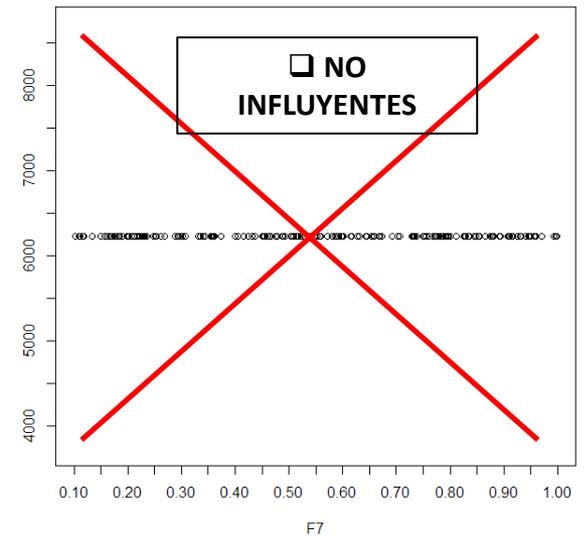
SENSIBILIDAD PARAMETRO A PARAMETRO, POR MONTECARLO.



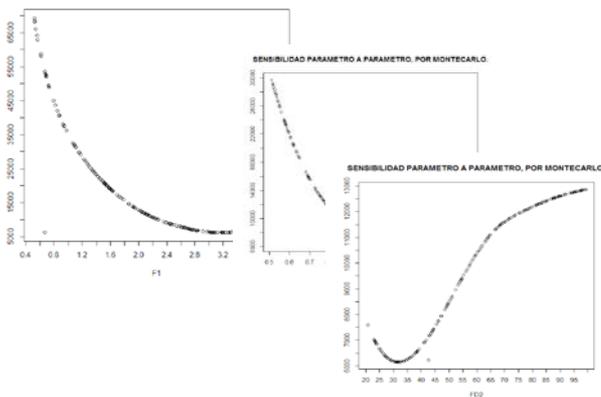
SENSIBILIDAD PARAMETRO A PARAMETRO, POR MONTECARLO



SENSIBILIDAD PARAMETRO A PARAMETRO, POR MONTECARLO.



SENSIBILIDAD PARAMETRO A PARAMETRO, POR MONTECARLO.



**REDUCCION DE 14 A 11 PARÁMETROS
SOBRE LOS QUE SE ESTIMARÁ SU
INCERTIDUMBRE**



Incertidumbre de los Parámetros

SIMULACIÓN MCMC

- ❑ **PARA INICIAR** LA SIMULACIÓN DE LA CADENA DE MARKOV ES NECESARIO:
 - TENER UNA **MATRIZ INICIAL DE COVARIANZAS** DE LA DISTRIBUCIÓN **CANDIDATA**
 - TENER UN **VECTOR INICIAL DE PARÁMETROS** (SOPORTE DE LA POSTERIOR)
 - DISPONER UNA **PRIOR DE LOS PARÁMETROS**: EN NUESTRO CASO, **NO-INFORMATIVA**

- ❑ **SE COMIENZA CON UNA MATRIZ DE COVARIANZAS DIAGONAL**
 - SE ASUME INICIALMENTE UNA HIPÓTESIS DE **INDEPENDENCIA ENTRE PARÁMETROS**
 - **VARIANZAS** SE DETERMINAN POR **CONOCIMIENTO EXPERTO**

- ❑ **MÉTODO MCMC ADAPTATIVO**
 - **MATRIZ INICIAL** SE VA MODIFICANDO DURANTE PROCESO DE **CALENTAMIENTO** (“BURNING”)

- ❑ **25000 SIMULACIONES DE CALENTAMIENTO**
 - NINGUNA DE ESTAS SIMULACIONES ES MUESTREO DE LA CADENA DE MARKOV

- ❑ **MUESTREO DE LA CADENA DE MARKOV**
 - TRAS LA ADAPTACIÓN, **5.000 SIMULACIONES MÁS**

Incertidumbre de los Parámetros

INFERENCIA SOBRE LA MUESTRA MCMC OBTENIDA

R> summary(MC)

```
Iterations = 1:5000
Thinning interval = 1
Number of chains = 1
Sample size per chain = 5000
```

1. Empirical mean and standard deviation for each variable, plus standard error of the mean:

	Mean	SD	Naive SE	Time-series SE
p1	4.197014	1.134e-01	1.604e-03	8.503e-03
p2	1.839625	5.035e-02	7.120e-04	4.044e-03
p3	6.923788	6.122e-01	8.657e-03	4.369e-02
p4	5.033751	1.069e+00	1.512e-02	7.435e-02
p5	241.179784	2.072e+01	2.931e-01	1.852e+00
p6	0.048093	1.791e-03	2.533e-05	1.357e-04
p7	3.223441	6.211e-01	8.783e-03	5.889e-02
p8	7.081434	4.761e-01	6.733e-03	4.176e-02
p9	-0.984668	4.469e-02	6.320e-04	3.093e-03
p10	0.004353	6.596e-05	9.329e-07	4.876e-06
p11	43.845504	1.717e+00	2.429e-02	1.485e-01

2. Quantiles for each variable:

	2.5%	25%	50%	75%	97.5%
p1	3.983237	4.117656	4.192587	4.267327	4.429281
p2	1.735466	1.808873	1.842349	1.874543	1.934949
p3	5.693817	6.523597	6.954430	7.415422	7.894978
p4	3.158247	4.202924	4.956089	5.792617	7.150535
p5	204.975272	225.821932	239.998664	255.217815	283.540004
p6	0.044449	0.046866	0.048160	0.049365	0.051416
p7	1.987604	2.808479	3.236030	3.665951	4.406630
p8	6.116830	6.750596	7.106272	7.441466	7.881216
p9	-1.072917	-1.013790	-0.983127	-0.954271	-0.895950
p10	0.004231	0.004308	0.004353	0.004399	0.004482
p11	40.389690	42.842877	43.840115	45.002904	46.942114

3. Varianza del Ruido

```
mean 7.4843404
sd 0.3803322
min 6.1878854
max 9.0260970
q025 7.2271093
q050 7.4835378
q075 7.7402213
```

DIFERENCIAS EN LA INCERTIDUMBRE DE LA MEDIA, DEBIDAS AL EFECTO DE AUTOCORRELACIÓN

4. Ratio de Aceptación MCMC

37.91667 %



Incertidumbre de los Parámetros

TESTEO DE LA SIMULACIÓN MCMC

OBJETO: COMPROBAR UN MUESTREO ACEPTABLE DE LA POSTERIOR

¿ALCANZÓ EL EQUILIBRIO LA CADENA DE MARKOV?

➤ **SOLO SE PODRÁ AFIRMAR QUE NO HAY EVIDENCIA DE LO CONTRARIO**

➤ INSPECCIÓN VISUAL DE LAS ITERACIONES

➤ TEST DE GEWEKE (1992) BASADO EN LAS TEORÍA DE SERIES TEMPORALES

Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to calculating posterior moments. In *Bayesian Statistics 4*, (ed. J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith). Clarendon Press, Oxford, UK

¿GRADO DE DEPENDENCIA DE LOS ELEMENTOS DE LA CADENA ?

➤ **INDEPENDENCIA MUESTRAL: CONDICIÓN INDISPENSABLE PARA PODER REALIZAR INFERENCIA**

➤ INSPECCIÓN VISUAL DE LA FUNCIÓN DE AUTOCORRELACIÓN

➤ TEST DE RAFTERY-LEWIS (1992)

Raftery, A.E. and Lewis, S.M. (1992). How many iterations in the Gibbs sampler? In *Bayesian Statistics 4* (J.M. Bernardo et al., editors), Oxford University Press, pp. 763-773

Incertidumbre de los Parámetros

TESTEO DE LA SIMULACIÓN MCMC

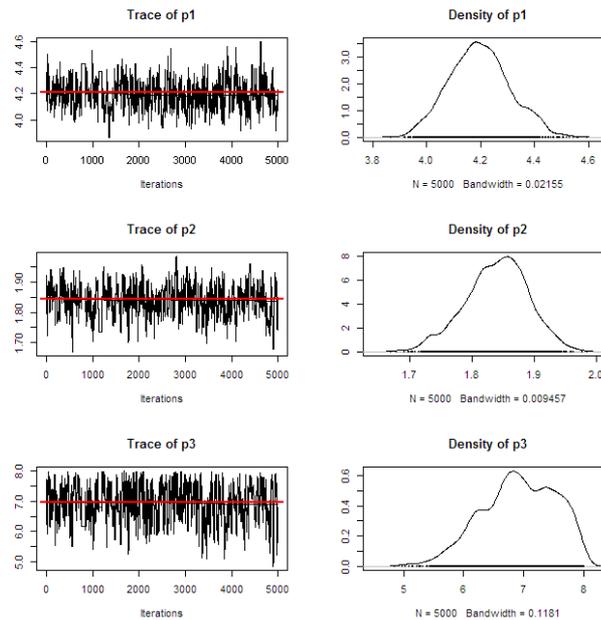
❑ ¿ALCANZÓ EL EQUILIBRIO LA CADENA DE MARKOV?

❑ INSPECCIÓN VISUAL DE LAS ITERACIONES

❑ SE PERSIGUE:

➤ COMPROBAR EL BUEN MEZCLADO DE LA CADENA

➤ COMPROBAR TENDENCIA MEDIA DE LOS VALORES ESTABILIZADA



Test OK !

Incertidumbre de los Parámetros

TESTEO DE LA SIMULACIÓN MCMC

¿ALCANZÓ EL EQUILIBRIO LA CADENA DE MARKOV?

TEST GEWEKE (1992)

FUNDAMENTO

SI LA CADENA ALCANZÓ LA DISTRIBUCIÓN DE EQUILIBRIO, LA MEDIA DEL PRINCIPIO DEBE SER APROXIMADAMENTE IGUAL A LA MEDIA DEL FINAL

```
R> geweke.diag(MC, frac1=0.1, frac2=0.5)
```

#Z-SCORES DEBEN SER EN VALOR ABSOLUTO APROX. MENORES A 2

```
Fraction in 1st window = 0.1
```

```
Fraction in 2nd window = 0.5
```

p1	p2	p3	p4	p5	p6	p7	p8	p9	p10	p11
1.3499	1.1748	1.1901	1.1774	1.5815	2.0341	0.6196	1.7123	0.9128	0.7883	-1.9873

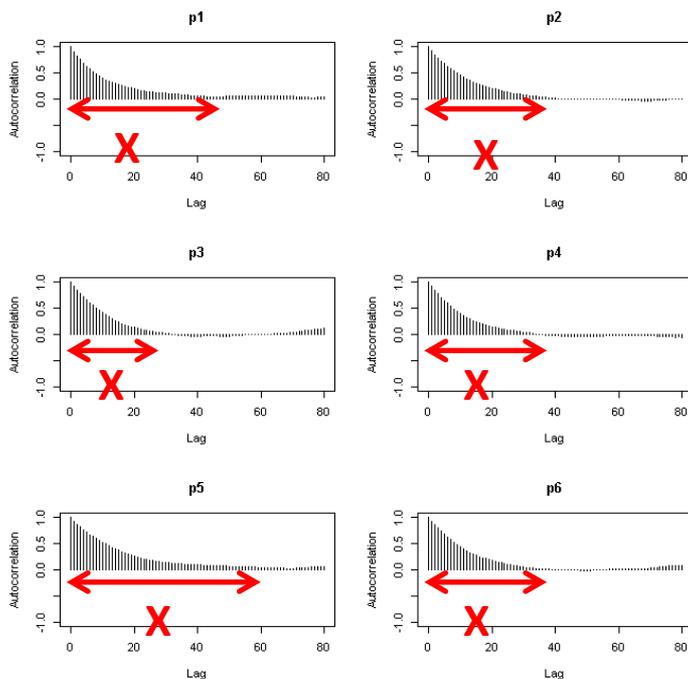
Test OK !

Incertidumbre de los Parámetros

TESTEO DE LA SIMULACIÓN MCMC

❑ ¿GRADO DE DEPENDENCIA DE LOS ELEMENTOS DE LA CADENA?

❑ INSPECCIÓN VISUAL DE LA FUNCIÓN DE AUTOCORRELACIÓN DE LAS CADENAS



➤ SE OBSERVA AUTOCORRELACION DEL TIPO AR(1):
PROCESO AUTOREGRESIVO DE ORDEN 1
(MARKOVIANO)

➤ LA **MUESTRA EFECTIVA** SERÁ TANTO MENOR
 CUANTO MAYOR SEA EL GRADO DE
 AUTOCORRELACIÓN

➤ SE NECESITA “ADELGAZAR” LA MUESTRA TOTAL, TOMANDO 1 VALOR DE LA CADENA POR CADA X : ***THINNING***

➤ ¿CUÁNTO VALE X ?

Incertidumbre de los Parámetros

TESTEO DE LA SIMULACIÓN MCMC

¿GRADO DE DEPENDENCIA DE LOS ELEMENTOS DE LA CADENA?

```
R> boa.randl(MC, 0.025, 0.05, 0.95, 0.001)
```

	Thin	Burn-in	Total	Lower Bound	Dependence Factor
p1	7	42	574	38	15.10526
p2	15	75	1425	38	37.50000
p3	9	45	684	38	18.00000
p4	7	35	434	38	11.42105
p5	8	40	536	38	14.10526
p6	11	66	902	38	23.73684
p7	11	44	605	38	15.92105
p8	9	54	648	38	17.05263
p9	7	42	490	38	12.89474
p10	8	32	480	38	12.63158
p11	15	180	1875	38	49.34211

```
R> boa.randl(MC, 0.975, 0.05, 0.95, 0.001)
```

	Thin	Burn-in	Total	Lower Bound	Dependence Factor
p1	6	30	252	38	6.631579
p2	11	66	726	38	19.105263
p3	6	24	282	38	7.421053
p4	4	20	212	38	5.578947
p5	10	50	470	38	12.368421
p6	9	54	648	38	17.052632
p7	9	45	540	38	14.210526
p8	5	25	295	38	7.763158
p9	6	30	330	38	8.684211
p10	6	30	312	38	8.210526
p11	5	20	225	38	5.921053

TEST DE RAFTERY-LEWIS (1992)

INFORMACIÓN NECESARIA PARA OBTENER UNA MUESTRA APROX. ALEATORIA

- CON LA QUE REALIZAR LA INFERENCIA DE UN DETERMINADO **CUANTIL** DE LA POSTERIOR
- FIJANDO UN ERROR STD. MONTE CARLO

INTERVALO DE ADELGAZAMIENTO

EL MAYOR INTERVALO DE ADELGAZAMIENTO O **THINNING** REQUERIDO ES 15, Y ES EL QUE HA DE TOMARSE PARA ADELGAZAR LA CADENA

Incertidumbre de los Parámetros

TESTEO DE LA SIMULACIÓN MCMC

¿GRADO DE DEPENDENCIA DE LOS ELEMENTOS DE LA CADENA?

```
R> boa.randl(AA, 0.025, 0.05, 0.95, 0.001)
```

	Thin	Burn-in	Total	Lower Bound	Dependence Factor
p1	1	3	53	38	1.394737
p2	1	2	42	38	1.105263
p3	1	5	68	38	1.789474
p4	1	3	53	38	1.394737
p5	1	3	53	38	1.394737
p6	1	6	87	38	2.289474
p7	1	2	42	38	1.105263
p8	1	5	68	38	1.789474
p9	1	3	53	38	1.394737
p10	1	2	42	38	1.105263
p11	1	9	114	38	3.000000

➤ **TRAS EL PROCESO DE ADELGAZAMIENTO DE LA CADENA, LA MUESTRA RESULTANTE PRESENTA 322 VECTORES DE PARÁMETROS EN LUGAR DE LOS 5000 INICIALES**

➤ **REPETICIÓN DEL TEST RAFTERY:**

LOS 322 VECTORES DE PARÁMETROS YA SE PUEDEN CONSIDERAR UNA MUESTRA APROX. ALEATORIA DE LA DISTRIBUCIÓN POSTERIOR.

YA SE PUEDEN REALIZAR CÁLCULOS DE INFERENCIA CON ELLOS

```
R> boa.randl(AA, 0.975, 0.05, 0.95, 0.001)
```

	Thin	Burn-in	Total	Lower Bound	Dependence Factor
p1	1	9	114	38	3.000000
p2	1	5	68	38	1.789474
p3	1	2	42	38	1.105263
p4	1	5	40	38	1.052632
p5	1	5	68	38	1.789474
p6	1	3	53	38	1.394737
p7	1	5	68	38	1.789474
p8	1	3	53	38	1.394737
p9	1	2	42	38	1.105263
p10	1	5	68	38	1.789474
p11	1	2	42	38	1.105263



Incertidumbre de los Parámetros

INFERENCIA SOBRE LA MUESTRA DE LA POSTERIOR DE LOS PARÁMETROS

```
R> summary(AA)
```

```
Iterations = 1:322
```

```
Thinning interval = 1
```

```
Number of chains = 1
```

```
Sample size per chain = 322
```

1. Empirical mean and standard deviation for each variable, plus standard error of the mean:

2.	Mean	SD	Naive SE	Time-series SE
p1	4.199244	1.131e-01	6.304e-03	9.651e-03
p2	1.837322	5.059e-02	2.819e-03	4.278e-03
p3	6.913651	6.283e-01	3.501e-02	4.833e-02
p4	5.035049	1.083e+00	6.030e-02	8.069e-02
p5	240.828980	2.048e+01	1.141e+00	1.862e+00
p6	0.048034	1.833e-03	1.021e-04	1.512e-04
p7	3.218921	6.247e-01	3.481e-02	6.077e-02
p8	7.078877	4.799e-01	2.674e-02	4.841e-02
p9	-0.985387	4.531e-02	2.525e-03	3.427e-03
p10	0.004352	6.637e-05	3.699e-06	5.015e-06
p11	43.860377	1.704e+00	9.497e-02	1.580e-01

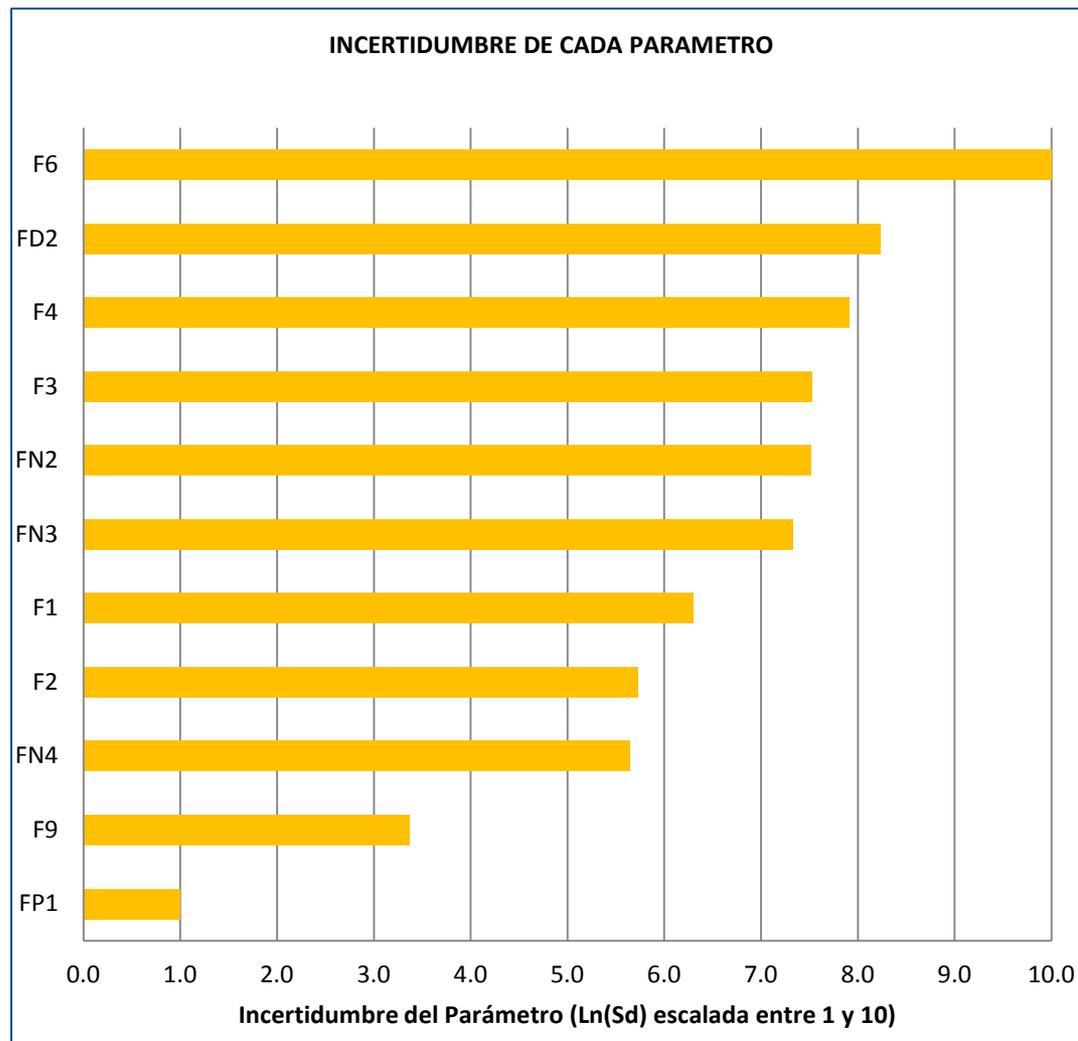
**AUTOCORRELACIÓN DESPRECIABLE:
LAS INCERTIDUMBRES DE LA MEDIA, YA
SON DEL MISMO ORDEN DE MAGNITUD**

2. Quantiles for each variable:

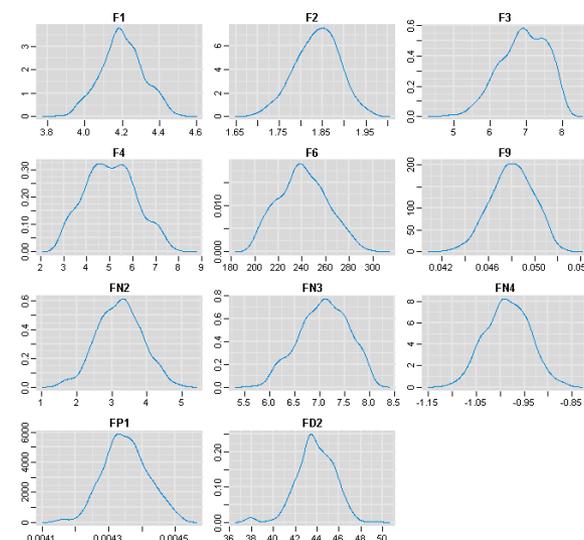
	2.5%	25%	50%	75%	97.5%
p1	3.970288	4.12626	4.191025	4.269767	4.429059
p2	1.735210	1.80120	1.840135	1.872292	1.930961
p3	5.632615	6.47848	6.960535	7.428221	7.889694
p4	3.099452	4.21885	4.970229	5.797013	7.150535
p5	204.016963	225.87801	239.906816	254.773741	282.082903
p6	0.044306	0.04686	0.048100	0.049353	0.051171
p7	1.978643	2.78436	3.231754	3.656588	4.371547
p8	6.116999	6.74788	7.109351	7.439576	7.892651
p9	-1.069221	-1.01515	-0.986919	-0.954480	-0.897490
p10	0.004231	0.00431	0.004348	0.004393	0.004481
p11	40.690234	42.90559	43.798672	45.053126	46.784611

Incertidumbre de los Parámetros

DISTRIBUCIÓN DE PROBABILIDAD POSTERIOR DE LOS PARÁMETROS



DISTRIBUCIONES POSTERIORES DE LOS PARAMETROS



		Mean	SD
F6	p5	240.82898	2.05E+01
FD2	p11	43.860377	1.70E+00
F4	p4	5.035049	1.08E+00
F3	p3	6.913651	6.28E-01
FN2	p7	3.218921	6.25E-01
FN3	p8	7.078877	4.80E-01
F1	p1	4.199244	1.13E-01
F2	p2	1.837322	5.06E-02
FN4	p9	-0.985387	4.53E-02
F9	p6	0.048034	1.83E-03
FP1	p10	0.004352	6.64E-05



Incertidumbre Reproductiva del Modelo

METODOLOGÍA

❑ SE EXTRAEN **100 MUESTRAS** (VECTORES DE PARÁMETROS) DEL TOTAL (322) DE LA CADENA DE MARKOV ESTACIONARIA Y “ADELGAZADA”

❑ CON ELLAS SE EJECUTA 100 VECES EL MODELO, OBTENIENDO ASÍ UNA SERIE DE **100 HIDROGRAMAS RESULTANTES**

❑ CON ELLOS SE ESTIMA:

- HIDROGRAMA MEDIO
- CUANTILES DEL 5% Y 95%.
- HIDROGRAMA DE QMÁXIMOS
- HIDROGRAMA DE QMÍNIMOS

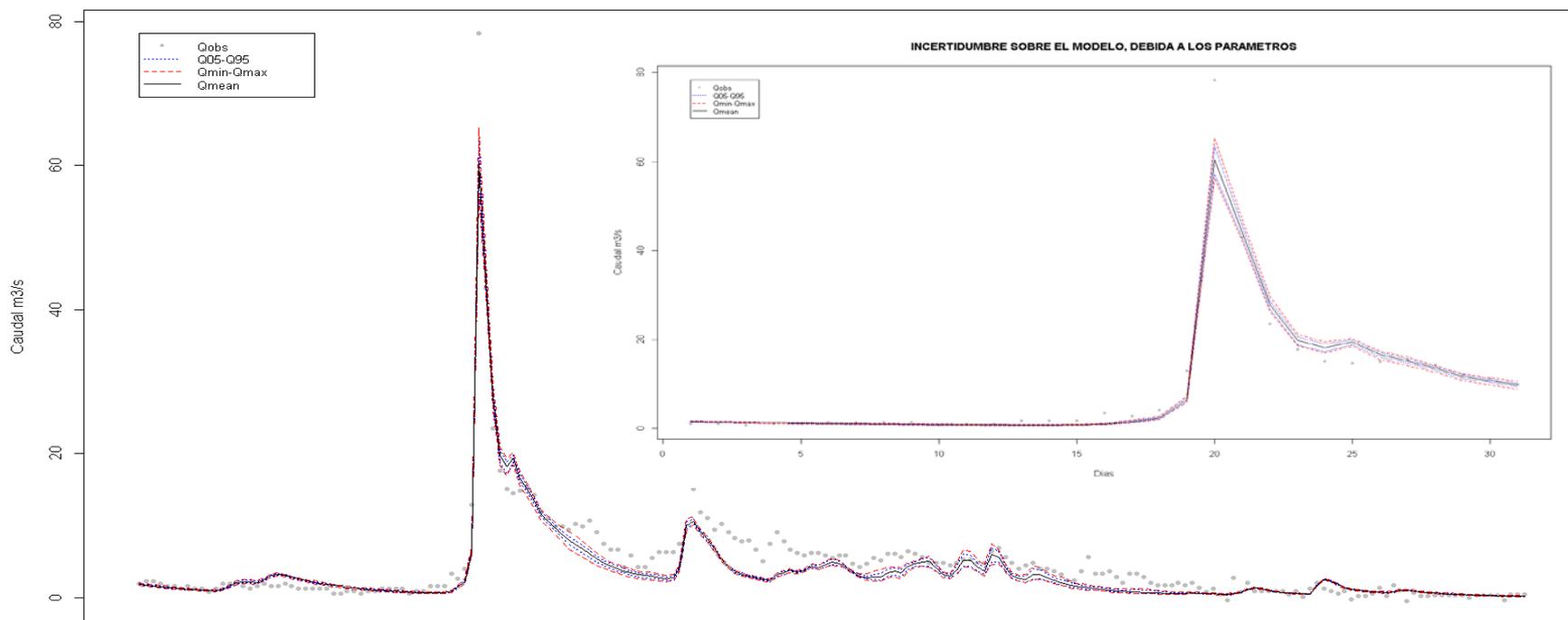


Incertidumbre Reproductiva del Modelo

ANÁLISIS DEL RESULTADO

- ❑ LAS BANDAS DE INCERTIDUMBRE NO ENGLOBALAN A LA MAYORÍA DE LAS OBSERVACIONES.
- ❑ EL EFECTO DE LA INCERTIDUMBRE DE LOS PARÁMETROS SOBRE LA RESPUESTA DEL MODELO SE MUESTRA POCO IMPORTANTE.
- ❑ **EXISTE OTRA INCERTIDUMBRE, CUYO EFECTO PARECE MÁS IMPORTANTE, Y QUE NO ES EXPLICADA POR LA INCERTIDUMBRE DE LOS PARÁMETROS.**

INCERTIDUMBRE SOBRE EL MODELO, DEBIDA A LOS PARAMETROS



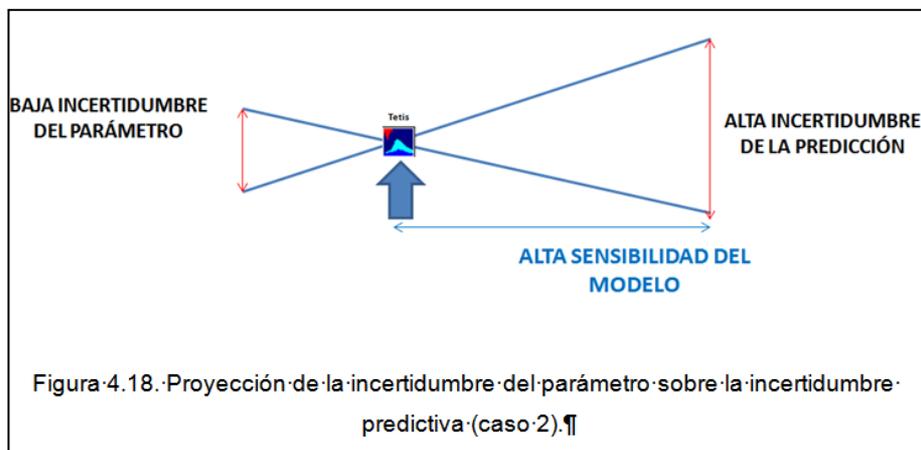
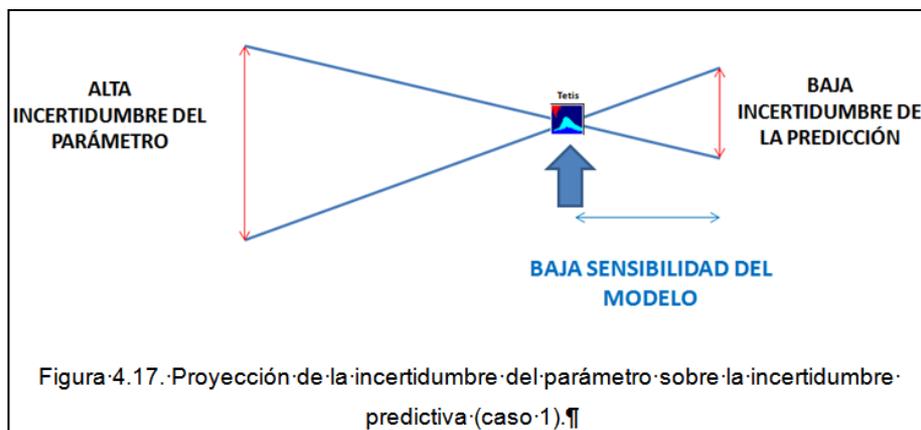
DETALLE ENTRE LOS DIAS 100 Y 300 DE SIMULACIÓN Y ENTRE LOS DIAS 130 Y 160.

Sensibilidad Global

...DEL MODELO A UN PARÁMETRO, SE PUEDE ENTENDER COMO LA CONTRIBUCIÓN DEL PARÁMETRO A LA INCERTIDUMBRE DE LA PREDICCIÓN

LA INCERTIDUMBRE DEL PARÁMETRO SOLO ES IMPORTANTE, SI AFECTA DE MANERA SIGNIFICATIVA AL RESULTADO DEL MODELO

3 COMPORTAMIENTOS



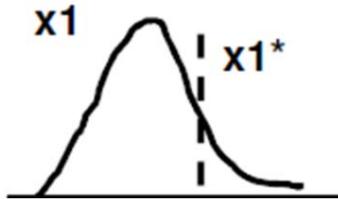


Sensibilidad Global

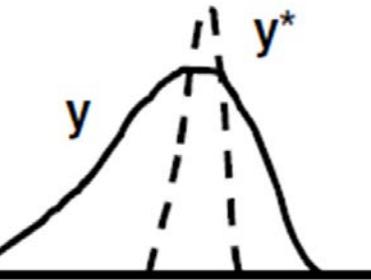
METODOLOGÍA

PARÁMETROS x_1, x_2

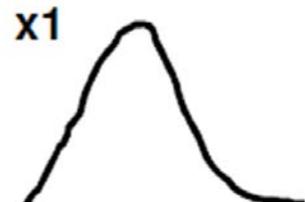
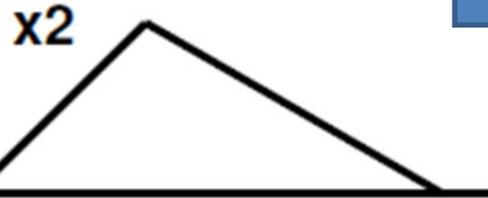
PREDICCIÓN (media)



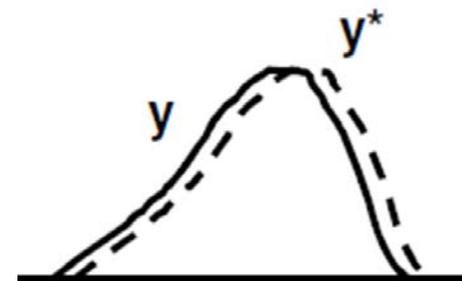
MODELO



MODELO MUY
SENSIBLE A x_1



MODELO



MODELO POCO
SENSIBLE A x_2

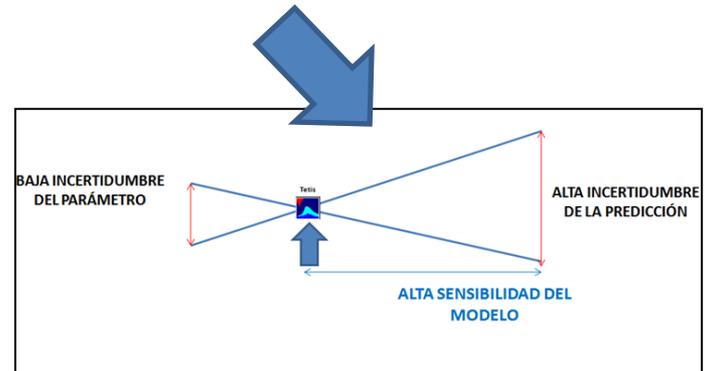
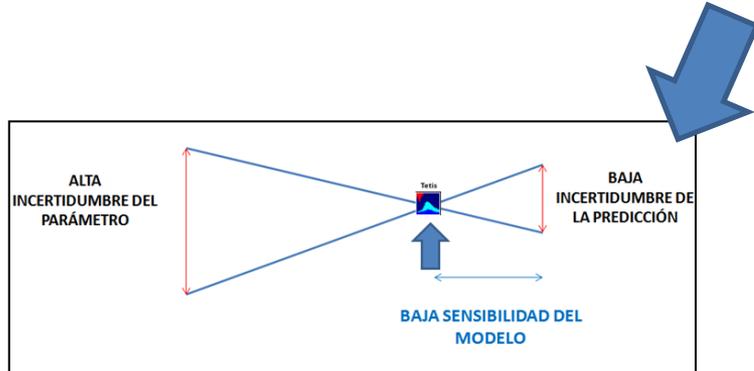
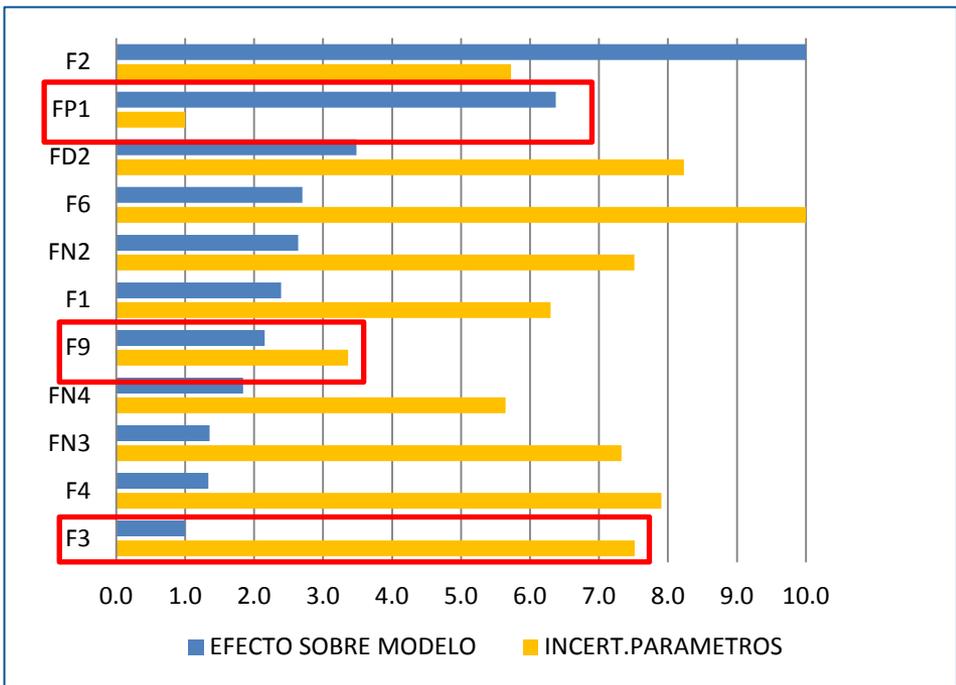
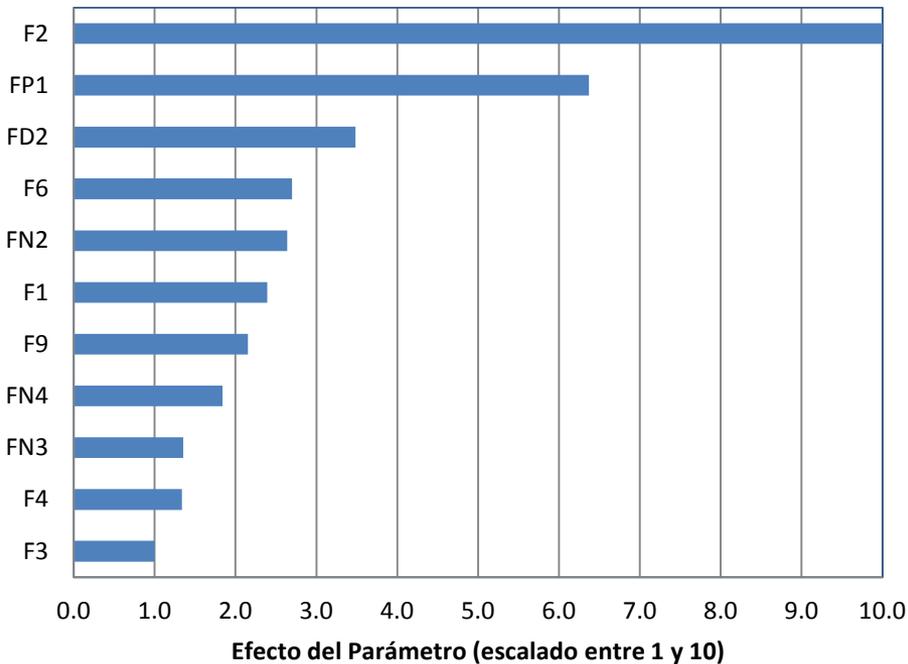
x_2

x_2^*



Sensibilidad Global

ANÁLISIS DEL RESULTADO



Incertidumbre Predictiva del Modelo

METODOLOGÍA (I)

MODELO DE ERROR

HIPÓTESIS :
RUIDO ADITIVO GAUSSIANO INDEPENDIENTE
Y HOMOCEDÁSTICO

$$\xi \sim \mathbf{N}(\mathbf{0}; \mathbf{I}_{N \times N} \sigma_i^2)$$

$$e_i \in \xi$$

$$e_i = y_i^{obs} - y_i^{mod}$$

$$\sigma_i^2 = \sigma^2$$

VARIANZA CONSTANTE
 ÚNICO PARÁMETRO DEL
 MODELO DE ERROR

POSTERIOR DE TODOS LOS PARAMETROS

$$p(\theta, \sigma | \tilde{Y}, \tilde{X}) = \frac{p(\tilde{Y} | \theta, \sigma, \tilde{X}) p(\theta, \sigma)}{p(\tilde{Y})}$$

POSTERIOR DEL PARÁMETRO DEL MODELO DE ERROR

VARIANZA RESIDUAL
PARÁMETRO NUISANCE: MUESTREO DE SU DIST. MARGINAL

$$p(\sigma^{-2}) \sim \Gamma\left(\frac{n_0}{2}, \frac{n_0}{2} S_0^2\right)$$

$$p(\sigma^{-2} | y, \theta) = \Gamma\left(\frac{n_0 + n}{2}, \frac{n_0 S_0^2 + SS(\theta)}{2}\right)$$

POSTERIOR PARÁMETROS MODELO HIDROLÓGICO

MUESTREO MCMC

$$p(\theta | \tilde{Y}, \tilde{X}) = \frac{p(\tilde{Y} | \theta, \tilde{X}) p(\theta)}{p(\tilde{Y})}$$



Incertidumbre Predictiva del Modelo

METODOLOGÍA (y II)

□ CADA VECTOR DE PARÁMETROS DEL MODELO HIDROLÓGICO (MUESTRA DE LA POSTERIOR) LLEVA ASOCIADA UN VALOR DE VARIANZA RESIDUAL DEL MODELO DE ERROR

□ DEFINIDO EL MODELO DE ERROR, SE PUEDEN ESTIMAR LAS BANDAS DE INCERTIDUMBRE PREDICTIVA DEL MODELO

□ PROCEDIMIENTO

$$y_{\text{mod } j} + \xi_{kj}$$

$$j = 1 \text{ to } 100 \\ k = 1 \text{ to } 500$$

$$\xi_{kj} \sim N(0; \sigma_j^2)$$

□ **50.000 HIDROGRAMAS**

➤ Hidrograma Medio

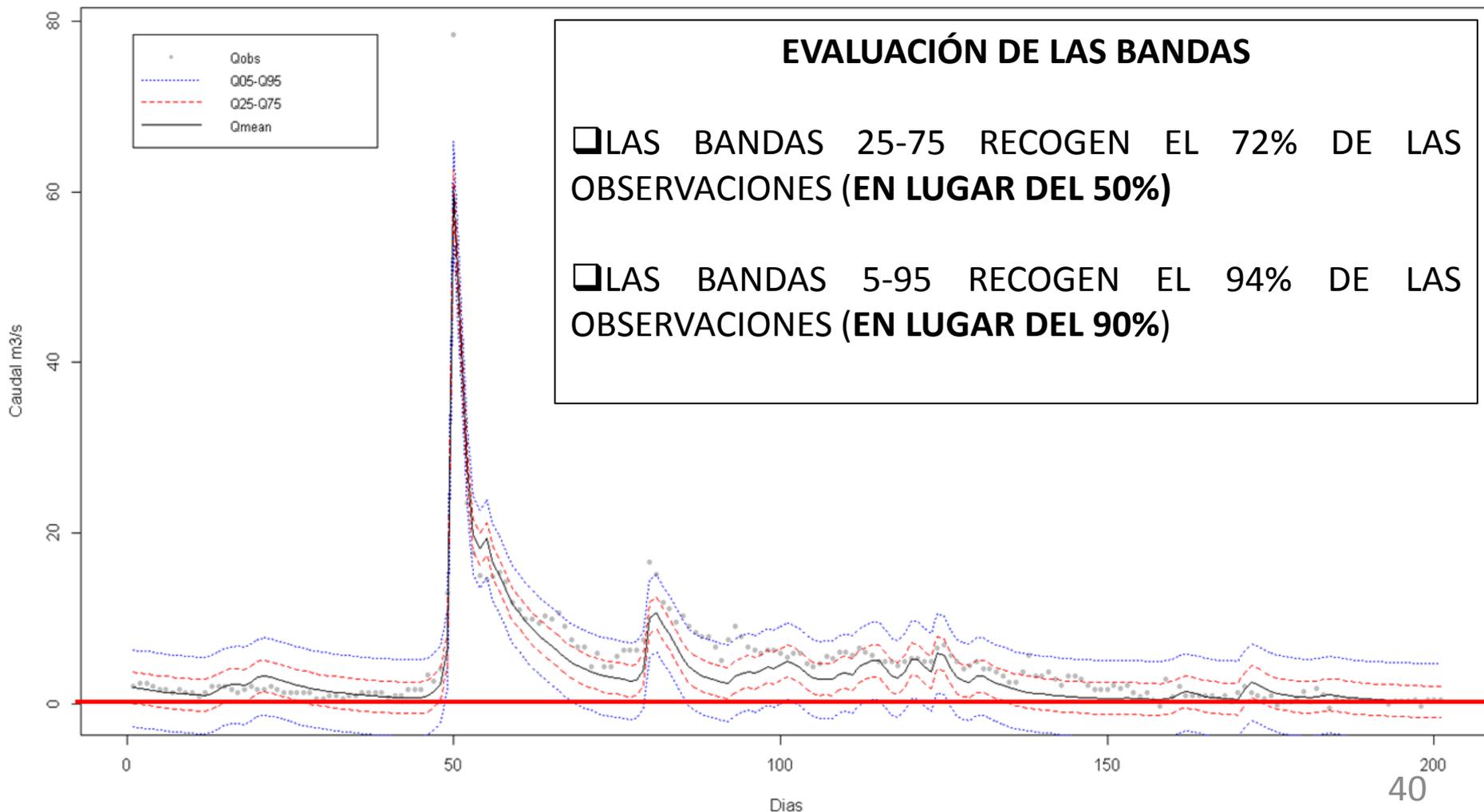
➤ Cuantiles del 5%, 25%, 75% y 95%.



Incertidumbre Predictiva del Modelo

ANÁLISIS DEL RESULTADO

DETALLE ENTRE LOS DIAS 100 Y 300 DE SIMULACIÓN

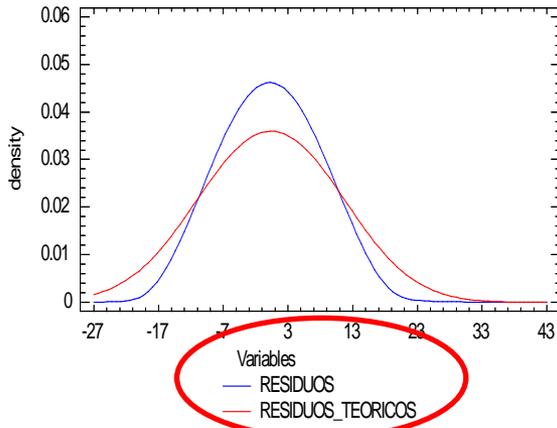


Incertidumbre Predictiva del Modelo

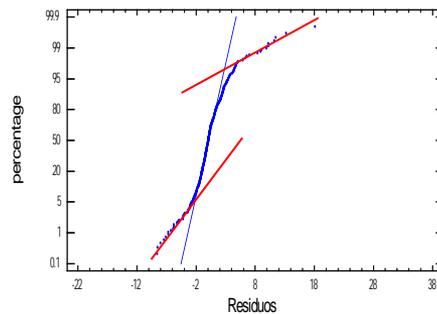
VERIFICACIÓN DEL MODELO DE ERROR

ERROR GAUSSIANO

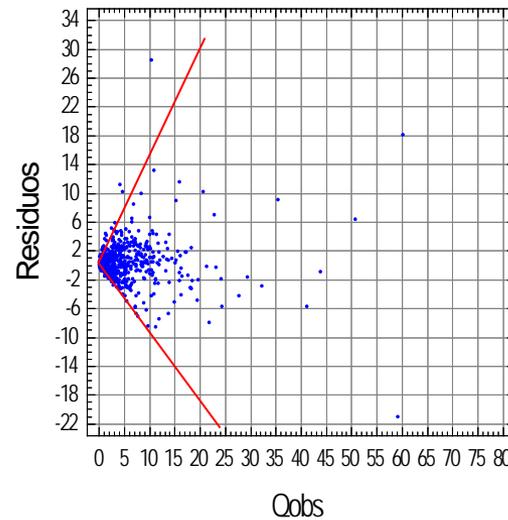
Density Traces



Normal Probability Plot

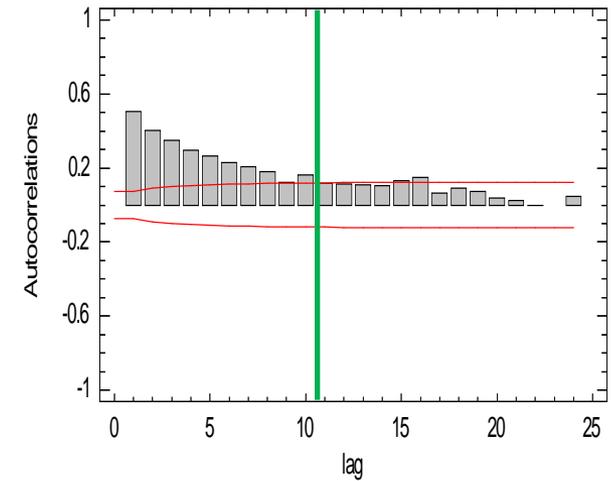


RESIDUOS HOMOCEDÁSTICOS



RESIDUOS INDEPENDIENTES

Estimated Autocorrelations for RESIDUOS



HIPÓTESIS :
RUIDO GAUSSIANO INDEPENDIENTE Y HOMOCEDÁSTICO



1- Sobre la Metodología

2- Sobre los Resultados

3- Aportes del Trabajo

4- Líneas Futuras





Sobre la Metodología

DEBILIDADES

- ALGORITMOS **MCMC** CADA VEZ MÁS **EFICIENTES** AUNQUE COMPUTACIONALMENTE **MUY INTENSIVOS**

- LA COMBINACIÓN DE SOFTWARE EMPLEADO **NO** HA PERMITIDO LA **SIMULACIÓN SIMULTÁNEA DE VARIAS CADENAS DE MARKOV**
 - **METODOLOGÍA MÁS ROBUSTA Y EFICIENTE** PARA “ASEGURAR” QUE SE HA LLEGADO AL EQUILIBRIO DE LA CADENA



Sobre la Metodología

FORTALEZAS

- ❑ METODOLOGÍA CON LA QUE ES POSIBLE ALCANZAR **MÚLTIPLES OBJETIVOS**
 - **CALIBRACIÓN DEL MODELO MÁS EFECTIVA** QUE OTROS MÉTODOS BASADOS EN LA OPTIMIZACIÓN DE FUNCIONES OBJETIVO
 - OBTENCIÓN DE LA **POSTERIOR (INCERTIDUMBRE)** DE LOS **PARÁMETROS**, EN LUGAR DE UNOS VALORES “ÓPTIMOS”
 - **EVALUAR LA INCERTIDUMBRE REPRODUCTIVA** DEL MODELO
- ❑ **PERMITE EL EMPLEO DEL CONOCIMIENTO PREVIO** SOBRE EL VALOR DEL PARÁMETRO
- ❑ METODOLOGÍA **ABIERTA A LA EVALUACIÓN POR SEPARADO** DE LAS DISTINTAS FUENTES DE INCERTIDUMBRE



Sobre los Resultados

EL MODELO DE ERROR

❑ ES **INCORRECTO** ASUMIR QUE EL **ERROR DE LOS MODELOS HIDROLÓGICOS** SE PUEDE MODELAR MEDIANTE UN MODELO DE ERROR CUYO **ÚNICO COMPONENTE SEA EL RUIDO BLANCO**, TAL Y COMO SE HA VISTO EN ESTE TRABAJO

➤ **LA INCERTIDUMBRE PREDICTIVA** ESTÁ CONDICIONADA AL MODELO DE ERROR EMPLEADO

❑ LA **PROPAGACIÓN DE LA INCERTIDUMBRE DE LOS PARÁMETROS** A TRAVÉS DEL MODELO HIDROLÓGICO HA EXPLICADO MUY POCO SOBRE LA INCERTIDUMBRE DEL MODELO

➤ EL MODELO DE ERROR ESCOGIDO TAMBIÉN INFLUYE EN EL CÁLCULO DE LA POSTERIOR DE LOS PARÁMETROS Y POR TANTO EN **LA INCERTIDUMBRE REPRODUCTIVA**

SENSIBILIDAD GLOBAL

❑ LA **SENSIBILIDAD GLOBAL DEL MODELO A CADA PARÁMETRO** DETERMINA LA FORMA EN QUE UN PARÁMETRO PROPAGA SU INCERTIDUMBRE.



Aportes del Trabajo

❑ **ANÁLISIS DE INCERTIDUMBRE DE LOS PARÁMETROS DE UN MODELO DISTRIBUÍDO MEDIANTE UN MÉTODO BAYESIANO FORMAL**

❑ **SE HA COMPROBADO UN MEJOR COMPORTAMIENTO DEL ALGORITMO BAYESIANO ADAPTATIVO DRAM (HAARIO, 2005) RESPECTO AL ALGORITMO DE OPTIMIZACIÓN SCE-UA (DUAN ET AL., 1993), EN LA CALIBRACIÓN DEL MODELO ANALIZADO**

➤ *Vrug et al., 2009* **LLEGAN A CONCLUSIÓN SIMILAR**

❑ **DESARROLLO DEL CÓDIGO DE PROGRAMACIÓN (R & MATLAB) QUE PERMITE LA UTILIZACIÓN DEL PAQUETE FME DE R, EN COMBINACIÓN CON EL MODELO TETIS**



Líneas Futuras de Trabajo

- ❑ **CONSIDERACIÓN DE MODELOS DE ERROR** QUE EXPLIQUEN MEJOR LOS RESIDUOS DE UN MODELO HIDROLÓGICO

- ❑ EXTENSIÓN DEL **ANÁLISIS DE INCERTIDUMBRE A OTROS PARÁMETROS** DEL MODELO **TETIS**, PERTENECIENTES A MÓDULOS EN DESARROLLO (SEDIMENTACIÓN, VEGETACIÓN, NIEVE,...)

- ❑ PROFUNDIZAR SOBRE LA METODOLOGÍA EMPLEADA Y EN GENERAL SOBRE LA **INFERENCIA BAYESIANA APLICADA A LA MODELACIÓN HIDROLÓGICA DISTRIBUIDA**

- ❑ **PARALELIZACIÓN DEL CÓDIGO DE TETIS** PARA ABORDAR EN TIEMPOS MÁS RAZONABLES, ANÁLISIS COMO EL EFECTUADO



TRABAJO FIN DE MASTER: TIPO B

Evaluación de la Incertidumbre de los Parámetros y su efecto sobre la Incertidumbre Predictiva de un Modelo Hidrológico Distribuido, mediante un algoritmo *Markov Chain Monte Carlo* de tipo Adaptativo.

Gracias por su atención.