

Taller Entrega 3: Modelo estocástico para estudiar el comportamiento del régimen de caudales de un río.

Omar Mercado

1. Modelo conceptual

Para los análisis de este trabajo el comportamiento de la variable caudal de un río se representa con un sencillo modelo conceptual, donde no se tiene en cuenta ni el almacenamiento ni cuerpos de agua reservorios:

- Variables de entrada del modelo:
 - Precipitación: Datos tomados de la reconstrucción del campo de precipitación a partir de estaciones meteorológicas dentro del área de aporte del río.
 - Infiltración: Un porcentaje de la cantidad de la precipitación que se infiltra en el suelo.
 - Evapotranspiración: Cantidad de agua que se evapora desde el suelo y se transpira de las plantas formula de Hargreaves.
 - Escorrentía Superficial: Un porcentaje de la cantidad de precipitación que no se infiltra y fluye sobre la superficie hacia el río.
 - Variables de Salida del modelo:
 - Caudal medio del Río: Cantidad de agua que fluye por el lecho del río en una unidad de tiempo.
 - Derivaciones: Cantidad de agua extraída del río para uso humano.
 - Caudal ambiental: Cantidad mínima de agua que debe conservar el río.
 - Parámetros del modelo:
 - Pendiente del terreno: Mapa de coeficientes de pendientes
 - Aforos del río: Mediciones de la forma y características del lecho del río que afectan la velocidad y el volumen del caudal.
 - Cobertura Vegetal: Coeficientes que determinan la interceptación de la precipitación, la infiltración y la evapotranspiración.
 - Tipo del Suelo: Coeficientes que determinan la escorrentía y la infiltración
 - Clima: Variaciones estacionales y cambios en la temperatura.
-
-

2. Modelo numérico con enfoque bayesiano

Se construye el modelo numérico para la obtención de la variable de caudales medios de un río usando un enfoque bayesiano que permite inferir variables de entrada del modelo a partir de los datos de precipitación. Luego, se realiza un proceso de optimización de los parámetros del modelo teniendo como parámetro de ajuste las mediciones de aforo realizadas en los ríos, mientras que los parámetros como la pendiente, cobertura vegetal y tipo de suelo que son coeficientes conocidos y establecidos a priori, se colocan a variar aleatoriamente dentro del modelo a partir de distribuciones dadas. A continuación, se detallan los pasos que se ejecutan dentro del modelo:

- Datos Observados

Se introducen los valores observados de precipitación y caudales medidos junto con sus desviaciones estándar. Asumiendo inicialmente un modelo lluvia-escorrentía se infieren las cantidades porcentuales de infiltración y escorrentía y se calibra un balance hídrico a partir de los datos de evapotranspiración calculada por Hargreaves.

- Preparación de los Datos para el Modelo
 - Se organizan los datos observados en una lista para pasarlos al modelo durante la inferencia.
- Modelo de inferencia Bayesiana

Se define el modelo bayesiano en Stan, que incluye:

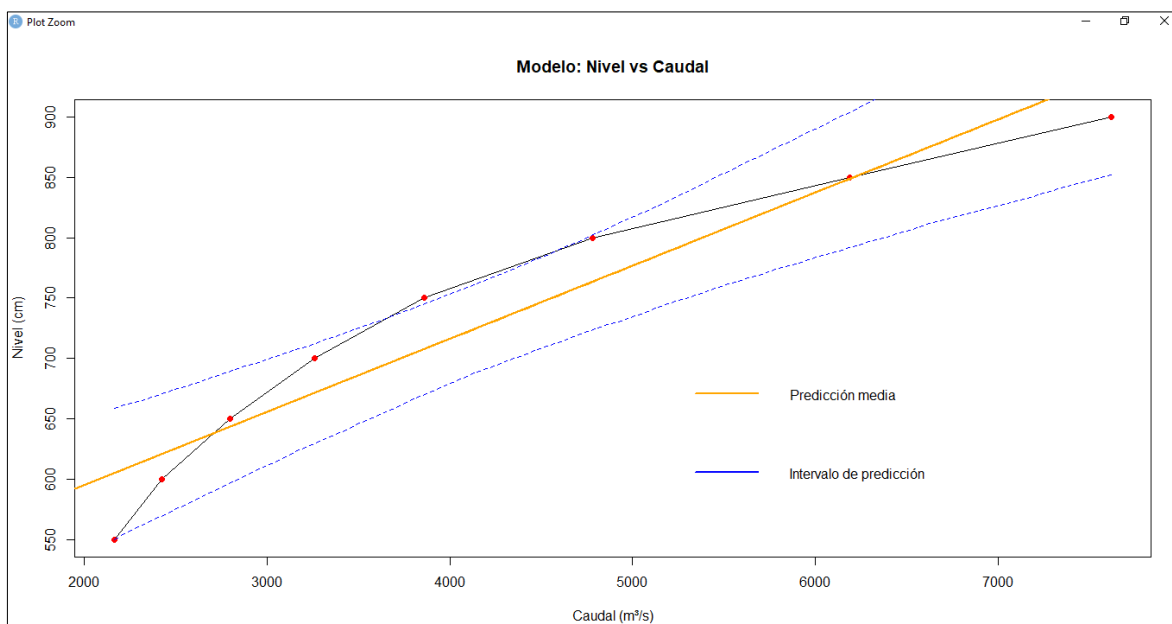
- Distribuciones a priori para los parámetros.
- Función de verosimilitud que relaciona los datos observados con los parámetros del modelo.
- Compilación del Modelo
 - Cantidades Inferidas: Predicción de datos de infiltración y escorrentía basada en los parámetros estimados.
- Ajuste del Modelo a los Datos
 - Se ajusta el modelo a los datos utilizando el muestreo de cadenas de Markov (sampling), especificando múltiples cadenas para mejorar la convergencia.
- Resumen de los Resultados
 - Se imprime un resumen de los resultados obtenidos del ajuste del modelo, que incluye estimaciones de los parámetros y diagnósticos de convergencia.
- Modelo de Optimización bayesiana para los resultados
 - Extracción de las distribuciones posteriores de los parámetros estimados (Pendiente, cobertura vegetal, tipo de suelo), lo que permite analizar y visualizar estos parámetros en función de los datos de entrada.
- Visualización de la Distribución Posterior del Caudal Medio
 - Cálculo del mínimo, máximo, media y desviación estándar de la distribución posterior del caudal medio.

3. Análisis de incertidumbre de variables

El proceso de muestreo bayesiano comienza con la simulación e inferencia de las variables de infiltración y escorrentía utilizando un modelo estadístico complejo conocido como Stan y luego realiza el cálculo de los caudales medio teniendo en cuenta distribuciones aleatorias para los parámetros del modelo pendiente, cobertura y tipo de suelo. El resumen de los resultados del muestreo proporciona una visión detallada de varios parámetros del modelo. El objetivo del modelo estimar los valores medios de caudal y la log-probabilidad del modelo.

Estas estimaciones no solo se presentan en términos numéricos, sino que también se acompañan de intervalos de confianza y medidas de convergencia, como el tamaño efectivo de muestra y el factor de reducción de escala potencial, lo que asegura la robustez y confiabilidad de los resultados obtenidos.

Para complementar el análisis numérico, se generan gráficos que visualizan las distribuciones posteriores de los parámetros clave, proporcionando una representación visual de la incertidumbre y la variabilidad inherente en las estimaciones.

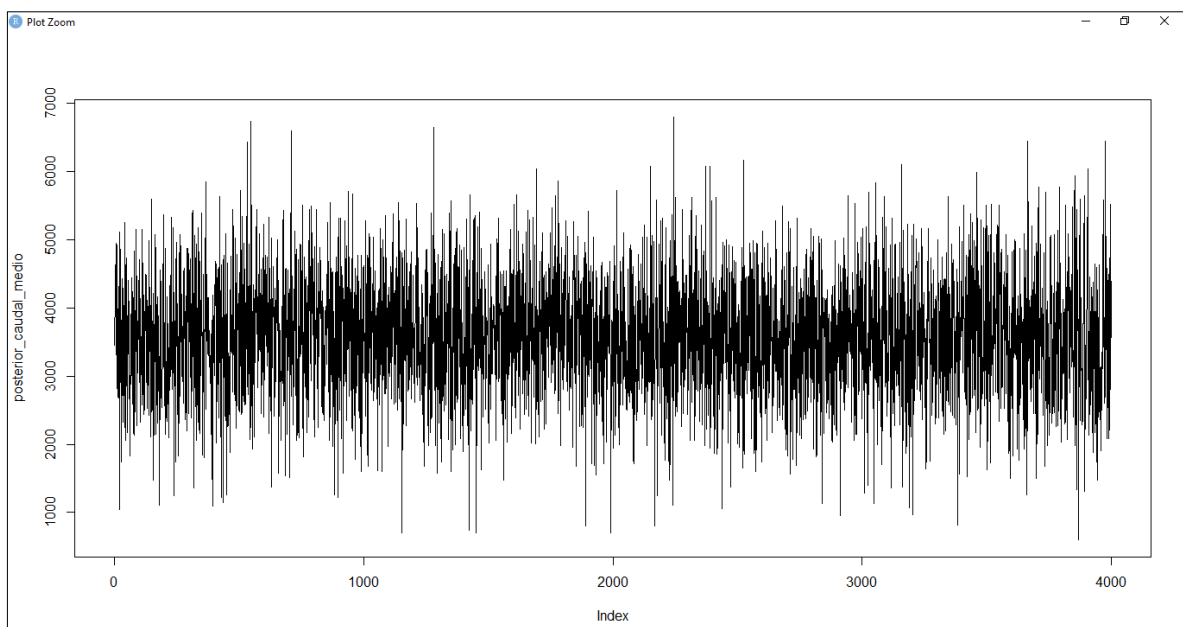
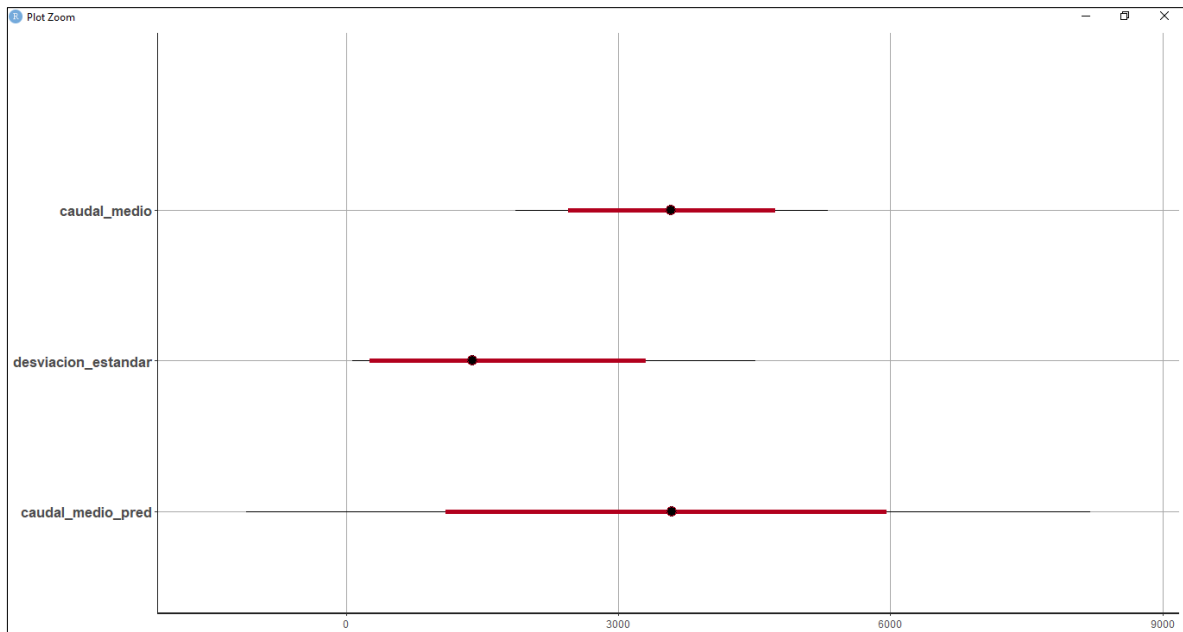


```
Inference for Stan model: anon_model.
4 chains, each with iter=2000; warmup=1000; thin=1;
post-warmup draws per chain=1000, total post-warmup draws=4000.
```

	mean	se_mean	sd	2.5%	25%	50%	75%	97.5%
caudal_medio	3584.52	23.79	887.93	1864.95	2982.93	3582.73	4184.56	5314.92
desviacion_estandar	1613.48	26.30	1195.99	63.33	658.38	1385.83	2340.81	4515.97
caudal_medio_pred	3566.36	38.65	2216.43	-1098.36	2493.16	3584.18	4655.71	8206.85
lp__	14.08	0.04	1.17	10.74	13.62	14.41	14.92	15.23

	n_eff	Rhat
caudal_medio	1393	1.00
desviacion_estandar	2069	1.00
caudal_medio_pred	3288	1.00
lp__	798	1.01

En particular, la distribución posterior del caudal medio se presenta en un gráfico de densidad, destacando las regiones más probables y ofreciendo una comprensión intuitiva de los posibles valores que puede tomar el caudal medio.



Una vez compilado el modelo bayesiano en Stan, se especifican la distribución a priori para el caudal medio y la desviación estándar desde los datos medidos en la estación hidrológica, y establece la función de verosimilitud que relaciona los datos observados con estos parámetros.

La optimización bayesiana comienza a buscar el mejor valor para la media a priori que maximice la verosimilitud de los datos observados.

```
elapsed = 0.48 Round = 1    pri_mean = 3449.4249 Value = -34.1079
elapsed = 0.30 Round = 2    pri_mean = 3159.8587 Value = -264.9912
elapsed = 0.28 Round = 3    pri_mean = 3169.4482 Value = -243.7549
elapsed = 0.30 Round = 4    pri_mean = 3380.3073 Value = -63.23675
elapsed = 0.28 Round = 5    pri_mean = 3724.4287 Value = -21.65153
elapsed = 0.29 Round = 6    pri_mean = 3663.0200 Value = -8.014278
```

elapsed = 0.28	Round = 7	pri_mean = 3560.2849	Value = -6.276767
elapsed = 0.32	Round = 8	pri_mean = 4000.0000	Value = -201.0053
elapsed = 0.28	Round = 9	pri_mean = 3602.0402	Value = -3.936012
elapsed = 0.29	Round = 10	pri_mean = 3595.8560	Value = -4.045923
elapsed = 0.39	Round = 11	pri_mean = 3884.1692	Value = -100.3441
elapsed = 0.33	Round = 12	pri_mean = 3604.6959	Value = -3.91101
elapsed = 0.38	Round = 13	pri_mean = 3609.9187	Value = -3.943899
elapsed = 0.29	Round = 14	pri_mean = 3610.2000	Value = -3.926447
elapsed = 0.34	Round = 15	pri_mean = 3600.0836	Value = -3.964081
elapsed = 0.33	Round = 16	pri_mean = 3601.1108	Value = -3.932598
elapsed = 0.34	Round = 17	pri_mean = 3604.7717	Value = -3.914688
elapsed = 0.30	Round = 18	pri_mean = 3603.7193	Value = -3.926866
elapsed = 0.30	Round = 19	pri_mean = 3595.5811	Value = -4.032697
elapsed = 0.29	Round = 20	pri_mean = 3599.4302	Value = -3.95937
elapsed = 0.28	Round = 21	pri_mean = 3604.6652	Value = -3.901646
elapsed = 0.28	Round = 22	pri_mean = 3610.1718	Value = -3.938356
elapsed = 0.28	Round = 23	pri_mean = 3607.5566	Value = -3.888511
elapsed = 0.28	Round = 24	pri_mean = 3000.0000	Value = -445.8944
elapsed = 0.27	Round = 25	pri_mean = 3608.0253	Value = -3.927331

4. Análisis de métricas de ajuste en cada uno de los procesos de modelación: calentamiento, calibración, validación y pronóstico

En este caso, se usan cuatro cadenas de Markov para realizar el muestreo, cada una de ellas llevando a cabo un total de 2000 iteraciones. Las primeras 1000 iteraciones se dedican al calentamiento, lo que permite al modelo ajustarse y estabilizarse, mientras que las siguientes 1000 iteraciones se utilizan para obtener las muestras finales que conformarán la base del análisis. Cada cadena sigue un proceso de evaluación gradual del gradiente y ajusta los pasos de “leapfrog” necesarios para recorrer el espacio de los parámetros del modelo. Este proceso se repite hasta que todas las iteraciones necesarias se completan, con tiempos de ejecución que varían ligeramente entre cada cadena debido a las diferencias en las evaluaciones de gradientes.

```
bayes_result <- sampling(stan_model, data = stan_data, chains = 4)

SAMPLING FOR MODEL 'anon_model' NOW (CHAIN 1).
Chain 1:
Chain 1: Gradient evaluation took 1.7e-05 seconds
Chain 1: 1000 transitions using 10 leapfrog steps per transition would take 0.17 seconds.
Chain 1: Adjust your expectations accordingly!
Chain 1:
Chain 1:
Chain 1: Iteration: 1 / 2000 [ 0%] (warmup)
Chain 1: Iteration: 200 / 2000 [ 10%] (warmup)
Chain 1: Iteration: 400 / 2000 [ 20%] (warmup)
Chain 1: Iteration: 600 / 2000 [ 30%] (warmup)
Chain 1: Iteration: 800 / 2000 [ 40%] (warmup)
Chain 1: Iteration: 1000 / 2000 [ 50%] (warmup)
Chain 1: Iteration: 1001 / 2000 [ 50%] (Sampling)
Chain 1: Iteration: 1200 / 2000 [ 60%] (Sampling)
Chain 1: Iteration: 1400 / 2000 [ 70%] (Sampling)
Chain 1: Iteration: 1600 / 2000 [ 80%] (Sampling)
Chain 1: Iteration: 1800 / 2000 [ 90%] (Sampling)
Chain 1: Iteration: 2000 / 2000 [100%] (Sampling)
Chain 1:
Chain 1: Elapsed Time: 0.032 seconds (warm-up)
Chain 1: 0.022 seconds (Sampling)
Chain 1: 0.054 seconds (Total)
```

```
Chain 1:
SAMPLING FOR MODEL 'anon_model' NOW (CHAIN 2).
Chain 2:
Chain 2: Gradient evaluation took 1.4e-05 seconds
Chain 2: 1000 transitions using 10 leapfrog steps per transition would take 0.14 seconds.
Chain 2: Adjust your expectations accordingly!
Chain 2:
Chain 2:
Chain 2: Iteration: 1 / 2000 [ 0%] (warmup)
Chain 2: Iteration: 200 / 2000 [ 10%] (warmup)
Chain 2: Iteration: 400 / 2000 [ 20%] (warmup)
Chain 2: Iteration: 600 / 2000 [ 30%] (warmup)
Chain 2: Iteration: 800 / 2000 [ 40%] (warmup)
Chain 2: Iteration: 1000 / 2000 [ 50%] (warmup)
Chain 2: Iteration: 1001 / 2000 [ 50%] (Sampling)
Chain 2: Iteration: 1200 / 2000 [ 60%] (Sampling)
Chain 2: Iteration: 1400 / 2000 [ 70%] (Sampling)
Chain 2: Iteration: 1600 / 2000 [ 80%] (Sampling)
Chain 2: Iteration: 1800 / 2000 [ 90%] (Sampling)
Chain 2: Iteration: 2000 / 2000 [100%] (Sampling)
Chain 2:
Chain 2: Elapsed Time: 0.02 seconds (warm-up)
Chain 2: 0.021 seconds (Sampling)
Chain 2: 0.041 seconds (Total)
Chain 2:

SAMPLING FOR MODEL 'anon_model' NOW (CHAIN 3).
Chain 3:
Chain 3: Gradient evaluation took 1e-05 seconds
Chain 3: 1000 transitions using 10 leapfrog steps per transition would take 0.1 seconds.
Chain 3: Adjust your expectations accordingly!
Chain 3:
Chain 3:
Chain 3: Iteration: 1 / 2000 [ 0%] (warmup)
Chain 3: Iteration: 200 / 2000 [ 10%] (warmup)
Chain 3: Iteration: 400 / 2000 [ 20%] (warmup)
Chain 3: Iteration: 600 / 2000 [ 30%] (warmup)
Chain 3: Iteration: 800 / 2000 [ 40%] (warmup)
Chain 3: Iteration: 1000 / 2000 [ 50%] (warmup)
Chain 3: Iteration: 1001 / 2000 [ 50%] (Sampling)
Chain 3: Iteration: 1200 / 2000 [ 60%] (Sampling)
Chain 3: Iteration: 1400 / 2000 [ 70%] (Sampling)
Chain 3: Iteration: 1600 / 2000 [ 80%] (Sampling)
Chain 3: Iteration: 1800 / 2000 [ 90%] (Sampling)
Chain 3: Iteration: 2000 / 2000 [100%] (Sampling)
Chain 3:
Chain 3: Elapsed Time: 0.024 seconds (warm-up)
Chain 3: 0.023 seconds (Sampling)
Chain 3: 0.047 seconds (Total)
Chain 3:

SAMPLING FOR MODEL 'anon_model' NOW (CHAIN 4).
Chain 4:
Chain 4: Gradient evaluation took 8e-06 seconds
Chain 4: 1000 transitions using 10 leapfrog steps per transition would take 0.08 seconds.
Chain 4: Adjust your expectations accordingly!
Chain 4:
Chain 4:
Chain 4: Iteration: 1 / 2000 [ 0%] (warmup)
Chain 4: Iteration: 200 / 2000 [ 10%] (warmup)
Chain 4: Iteration: 400 / 2000 [ 20%] (warmup)
```

```

Chain 4: Iteration: 600 / 2000 [ 30%] (warmup)
Chain 4: Iteration: 800 / 2000 [ 40%] (warmup)
Chain 4: Iteration: 1000 / 2000 [ 50%] (warmup)
Chain 4: Iteration: 1001 / 2000 [ 50%] (Sampling)
Chain 4: Iteration: 1200 / 2000 [ 60%] (Sampling)
Chain 4: Iteration: 1400 / 2000 [ 70%] (Sampling)
Chain 4: Iteration: 1600 / 2000 [ 80%] (Sampling)
Chain 4: Iteration: 1800 / 2000 [ 90%] (Sampling)
Chain 4: Iteration: 2000 / 2000 [100%] (Sampling)
Chain 4:
Chain 4: Elapsed Time: 0.023 seconds (warm-up)
Chain 4:                  0.024 seconds (Sampling)
Chain 4:                  0.047 seconds (Total)
Chain 4:

```

5. Análisis de sensibilidad paramétrico

En este análisis, se busca determinar cómo diferentes valores de las variables de entrada y de distribución a priori de los parámetros afectan la distribución posterior del caudal medio en un modelo bayesiano. Al compilar el modelo utilizando `stan_model` se definen los datos de observación de caudal y precipitación junto con sus desviaciones estándar. A continuación, se ajusta el modelo a los datos utilizando diferentes valores para las variables de entrada. Se itera sobre estos valores, ajustando el modelo para cada uno y almacenando los resultados.

Este análisis permite entender cómo las elecciones de las medias a priori pueden influir en las estimaciones posteriores en un modelo bayesiano.

