

Procesos gaussianos: metodología y aplicaciones
Taller de Tesis I - Entrega 4
Maestría en Explotación de Datos y Descubrimiento del
Conocimiento – UBA 2023 cuat. 1

G. Sebastián Pedersen

sebasped@gmail.com

Última actualización: 24 de julio de 2023

Resumen

En este trabajo se estudian a los procesos gaussianos y su aplicabilidad para predecir series de tiempo, en particular series de tiempo de producción de hidrocarburos. Los procesos gaussianos son un método no paramétrico probabilístico de regresión. En este trabajo por un lado se indaga metodológicamente a los procesos gaussianos, explicitándose sus virtudes y limitaciones teóricas (modelado por gaussianas, incertidumbre natural, etc.). Y por otro lado se aplican los procesos gaussianos para predecir la producción mensual de hidrocarburos en la Argentina, encontrándose las dificultades típicas del caso (diseño del kernel, elección y optimización de los parámetros, etc.). Los resultados de este trabajo, si bien con las dificultades antes mencionadas, por lo menos hacen que la técnica merezca más tiempo de estudio y experimentación, pues a primer análisis y experimentación predice dentro de límites razonables (o, si no es el caso hay indicios suficientes para suponer que podría hacerlo).

Índice

1. Introducción	2
2. Marco teórico: procesos gaussianos	2
2.1. Trabajos previos	2
2.2. Qué es un proceso gaussiano (\mathcal{GP})	2
2.2.1. Idea y relación con la regresión clásica	2
2.2.2. Definición formal de proceso gaussiano (\mathcal{GP})	2
2.3. Muestreo de un \mathcal{GP}	3
2.3.1. Muestreo <i>prior</i>	3
2.3.2. Predicción <i>posterior</i> sin ruido	4
2.3.3. Predicción <i>posterior</i> con ruido	5
3. Metodología	6
3.1. Datos: descripción, preprocesamiento y análisis exploratorio	6
3.2. Técnicas de análisis y modelado	7
4. Resultados y discusión	8
4.1. Resultados: presentación, análisis y discusión	8
4.2. Limitaciones y mejoras	10
5. Conclusiones	10
Referencias	11

1. Introducción

En este trabajo se estudian a los procesos gaussianos tanto desde un punto de vista teórico, así como desde un punto de vista práctico mediante una aplicación en concreto. Los procesos gaussianos son un método no paramétrico probabilístico de regresión. Los objetivos son dos. Por un lado hay un objetivo metodológico sobre los procesos gaussianos que involucra estudiarlos, entenderlos y ahondar en ellos. Por otro lado se investiga las bondades de los procesos gaussianos para realizar predicciones sobre series temporales, tomando como ejemplo práctico datos de producción mensual de petróleo y gas desde 2006 publicada por la Secretaría de Energía de la Nación. Los datos son abiertos y accesibles desde:

<https://datos.gob.ar/dataset/energia-produccion-petroleo-gas-por-pozo-capitulo-iv>

Como referencia general al problema de forecasting (predicción sobre series temporales) se puede consultar [5]. Como referencias generales a procesos gaussianos se pueden consultar [1], [2], [3] y [4]. Como referencia particular de procesos gaussianos aplicados a la predicción de producción de petróleo y gas se puede consultar [6].

2. Marco teórico: procesos gaussianos

2.1. Trabajos previos

- Como trabajos previos específicos: hacer referencia a [6] y por las dudas buscar otros.
- Como referencias relevantes generales: buscar las partes correspondientes de [1], [3], [4], [2].

2.2. Qué es un proceso gaussiano (\mathcal{GP})

2.2.1. Idea y relación con la regresión clásica

El problema de regresión clásica busca encontrar una $f(x) = y$ que mejor ajuste a un conjunto de datos $\mathcal{D} = \{(x_i, y_i)\}_{i=1}^N \subset \mathbb{R}^2$. La f se propone parametrizada por una cantidad fija de parámetros (sería la elección del modelo) y el “mejor ajuste” viene dado por encontrar los parámetros que minimizan una función de costo o función objetivo (típicamente el error cuadrático) entre la f y los datos \mathcal{D} . Por ejemplo en el problema de regresión lineal clásico la cantidad de parámetros está dada por la cantidad de covariables de los datos a utilizar (incluyendo un posible feature engineering), y la f se propone como una función lineal en los parámetros (podría ser no lineal en los datos).

Los modelos que utilizan una cantidad fija de parámetros se llaman *paramétricos*. En contraste a éstos existe los modelos *no paramétricos* cuya cantidad de parámetros no es fija (por ejemplo KNN o SVM). Por supuesto siempre hablando de parámetros del modelo a aprender. También existen los hiperparámetros del modelo que sí pueden ser fijos, independientemente de si el modelo es paramétrico o no. Por ejemplo en KNN un hiperparámetro del modelo sería el radio elegido para considerar a un punto vecino de otro.

Los procesos gaussianos (\mathcal{GP}) son un método no paramétrico probabilístico de regresión. Este modelo busca encontrar ya no una única f que mejor ajusta a los datos, sino una distribución sobre funciones $P(f)$ que mejor ajusten a los datos, donde $f : \mathcal{X} \rightarrow \mathbb{R}$ es una función del espacio de entrada \mathcal{X} (espacio en donde vivirán los datos \mathcal{D}). Como su nombre ya lo indica en los \mathcal{GP} la distribución de las f será gaussiana. Una de los principales beneficios de los \mathcal{GP} es que se obtiene una cuantificación sobre la incertidumbre del ajuste, dada naturalmente por la distribución $P(f)$.

2.2.2. Definición formal de proceso gaussiano (\mathcal{GP})

Dado que el espacio de entrada \mathcal{X} de la $f : \mathcal{X} \rightarrow \mathbb{R}$ bien puede ser infinito, se puede pensar a la f como un vector infinito dimensional. Como en la práctica solamente se pueden trabajar con finitos valores, podemos pensar que dados $\{x_i\}_{i=1}^n$ la f es un vector evaluando en cada x_i , es decir $f(\mathbf{x}) = (f(x_1), \dots, f(x_n))$, y en donde la $f(\mathbf{x})$ tendrá distribución gaussiana. Más precisamente:

Definición (proceso gaussiano): un proceso gaussiano (\mathcal{GP}) es una colección de variables aleatorias tal que cualquier subconjunto finito tiene distribución (conjunta) gaussiana.

Para entender esta definición, apliquémosla al problema de regresión formulado anteriormente en 2.2.1. Será entonces $P(f)$ un \mathcal{GP} si para cualquier subconjunto $\{x_i\}_{i=1}^n \subset \mathcal{X} = \mathbb{R}$ la distribución $P(f(\mathbf{x}))$ de $f(\mathbf{x}) = (f(x_1), \dots, f(x_n))$ es una gaussiana multivariada. Las variables aleatorias representan, en este caso, los valores $f(x_i)$ para cada x_i .

Como cualquier gaussiana multivariada queda completamente caracterizada por su media y su covarianza, para un \mathcal{GP} definimos sus funciones media y covarianza como:

$$m(x) = \mathbb{E}\{f(x)\} \quad (1)$$

$$K(x_1, x_2) = \mathbb{E}\{[f(x_1) - m(x_1)][f(x_2) - m(x_2)]\} \quad (2)$$

De esta forma podemos escribir al \mathcal{GP} y a su marginal como:

$$f \sim \mathcal{GP}(m, K) \quad (3)$$

$$f(\mathbf{x}) \sim \mathcal{N}(m(\mathbf{x}), K(\mathbf{x}, \mathbf{x})) \quad (4)$$

Como esta marginalización a una gaussiana debe cumplirse para cualquier subconjunto finito de variables aleatorias, y el conjunto \mathcal{X} original bien puede ser infinito, se pide que el \mathcal{GP} cumpla con esta propiedad de marginalización (llamada propiedad de consistencia). Notemos que esta propiedad se cumple para la función de covarianza definida en la ecuación 2.

Elecciones de funciones de media y covarianza: es normal asumir que $m = 0$ pues siempre se pueden referenciar los datos al promedio. Para la función de covarianza K , también llamada kernel, hay que pedir que sea semi definida positiva (para cada subconjunto finito). La elección del kernel en general dependerá del problema en particular que se desee estudiar. Una elección popular es el kernel *exponencial cuadrático* (square exponential), también conocido como radial basis function (RBF) o kernel gaussiano:

$$K_{SE}(x_1, x_2) = \sigma^2 \exp\left(-\frac{(x_1 - x_2)^2}{2l^2}\right) \quad (5)$$

donde σ y l son parámetros interpretables.

Observación: en principio cualquier elección de kernel que cumpla la definición es teóricamente aceptable. En la práctica existen unos cuantos modelos de kernels populares y su elección está relacionada con la naturaleza del problema y los datos (por ejemplo ver el ch. 2 de [7]). Además los valores de los parámetros de los kernels son relevantes a la hora de predecir y en general se estiman a partir de los datos de entrenamiento mediante máxima verosimilitud o MCMC (por ejemplo ver sec. 8.3 de [2]).

2.3. Muestreo de un \mathcal{GP}

2.3.1. Muestreo *prior*

Mirando la definición de la sección 2.2.2 podemos entonces pensar que un \mathcal{GP} define un *prior* sobre funciones, y por lo tanto ya podemos realizar muestreos para cada subconjunto finito $\{x_i\}_{i=1}^n \subset \mathcal{X} = \mathbb{R}$, previa elección de la forma funcional de la covarianza K , pues tendremos que:

$$f(\mathbf{x}) \sim \mathcal{N}(m(\mathbf{x}), K(\mathbf{x}, \mathbf{x}))$$

donde $\mathbf{x} = (x_1, \dots, x_n)$, $f(\mathbf{x}) = (f(x_1), \dots, f(x_n))$ y $K(x_i, x_j)$ supongamos, para fijar ideas, viene dado por K_{SE} de la ecuación 5. Asumiendo media igual a cero, tomando $\sigma = l = 1$ para los parámetros del K_{SE} , y muestreando tres veces para 100 valores de x equiespaciados entre 0 y 10, obtenemos 3 muestras del correspondiente *prior* del \mathcal{GP} como se puede ver en la figura 1. El intervalo de confianza corresponde a moverse 1,96 devíos estándares para cada valor $f(x_i)$.

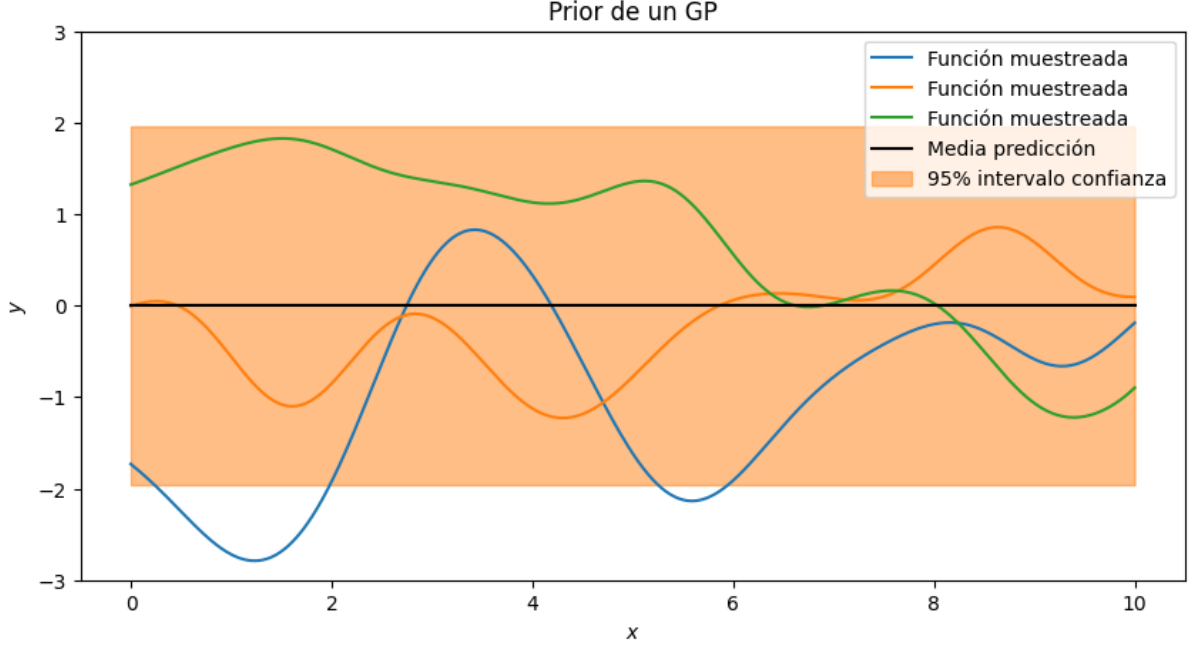


Figura 1. Prior de un \mathcal{GP} . A modo ilustrativo se grafican 3 muestras del prior del \mathcal{GP} .

2.3.2. Predicción *posterior* sin ruido

Supongamos que tenemos observaciones $(X, f(X))$ donde $X = (x_1, \dots, x_n)$ y queremos realizar una predicción sobre un conjunto de valores X^* de tamaño n^* (X y X^* pueden o no compartir valores). De la definición de \mathcal{GP} se desprende inmediatamente que la distribución conjunta es:

$$\begin{bmatrix} f(X) \\ f(X^*) \end{bmatrix} \sim \mathcal{N} \left(\begin{bmatrix} m(X) \\ m(X^*) \end{bmatrix}, \begin{bmatrix} K(X, X) & K(X, X^*) \\ K(X^*, X) & K(X^*, X^*) \end{bmatrix} \right) \quad (6)$$

Como para realizar la predicción necesitamos la distribución de $f(X^*)$ (condicional a los datos conocidos, por supuesto), nos valeremos del siguiente resultado:

Proposición: dado un prior \mathcal{GP} , la posterior también es un \mathcal{GP} y se puede condicionar sobre los datos conocidos para obtener:

$$f(X^*)|f(X), X \sim \mathcal{N}(m_{X^*|X}, \Sigma_{X^*|X}) \quad (7)$$

Donde la media y covarianza son:

$$m_{X^*|X} = m(X^*) + K(X^*, X)K^{-1}(X, X)(f(X) - m(X)) \quad (8)$$

$$\Sigma_{X^*|X} = K(X^*, X^*) - K(X^*, X)K^{-1}(X, X)K(X, X^*) \quad (9)$$

Observemos que tanto la media como la covarianza de la distribución de $f(X^*)$ dependen tanto de las observaciones conocidas (que juegan el papel de datos de entrenamiento) como de los valores de X^* en donde deseamos realizar las predicciones $f(X^*)$.

Veamos un ejemplo de esto en acción. Supongamos nuevamente media cero y kernel SE con $\sigma = l = 1$. Tomemos a X como 6 datos entre 0 y 10, y las $f(X)$ dadas por $f(x) = x \sin(x)$. Tomemos ahora para predecir a X^* como 100 datos equiespaciados entre 0 y 10. Realizamos 3 muestreos de la $f(X^*)|f(X), X$ que se pueden ver en la figura 2.

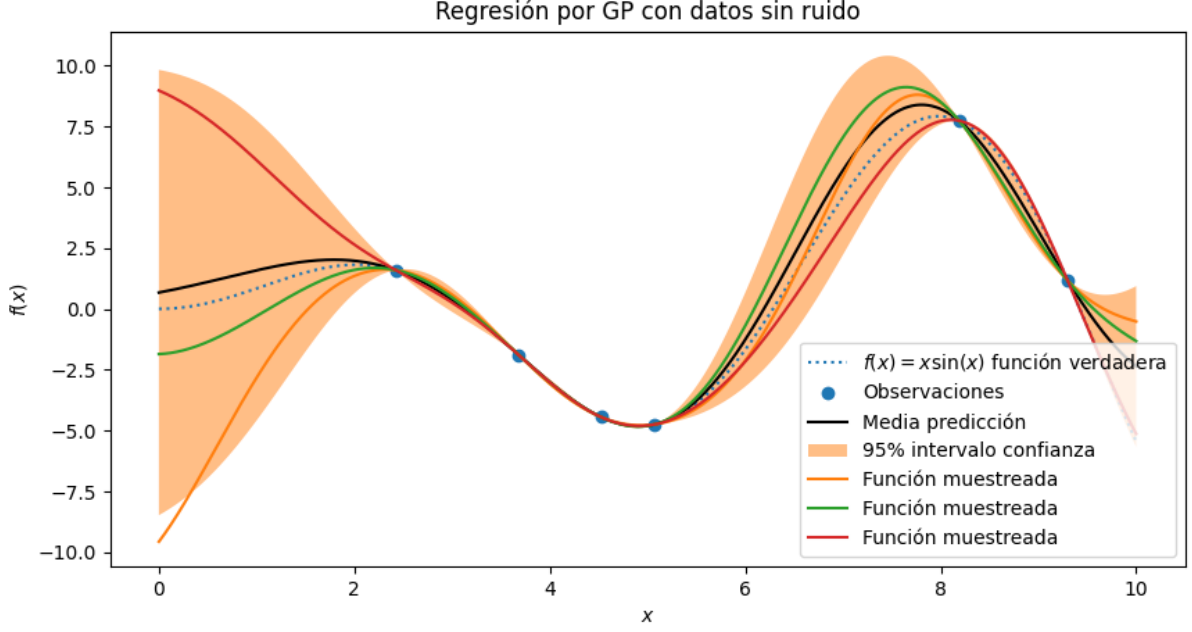


Figura 2. Regresión por \mathcal{GP} con datos sin ruido. A modo ilustrativo se grafican 3 muestras del posterior del \mathcal{GP} .

Observemos que la predicción puede realizarse tanto interpolando como extrapolando a los datos conocidos. Además el intervalo de confianza es más angosto cuando los datos conocidos están menos distantes.

2.3.3. Predicción *posterior* con ruido

En aplicaciones reales es normal tener mediciones con ruido. Supongamos que tenemos observaciones $(X, f(X))$ donde $X = (x_1, \dots, x_n)$, pero ahora nuestras observaciones tiene ruido $y_i = f(x_i) + \eta$, donde suponemos que $\eta \sim \mathcal{N}(0, \sigma_n^2)$ y que el ruido es i.i.d. Por lo tanto ahora nuestro conjunto de datos es de la forma (X, Y) donde $Y = f(X) + \eta$.

Es fácil ver que este ruido equivale, en nuestro modelo, a agregar un término diagonal a la covarianza:

$$\text{cov}(Y) = K(X, X) + \sigma_n^2 \mathbb{I}$$

Nuevamente queremos predecir sobre un conjunto X^* , y en este caso la conjunta queda como:

$$\begin{bmatrix} Y \\ f(X^*) \end{bmatrix} \sim \mathcal{N} \left(\begin{bmatrix} m(X) \\ m(X^*) \end{bmatrix}, \begin{bmatrix} K(X, X) + \sigma_n^2 \mathbb{I} & K(X, X^*) \\ K(X^*, X) & K(X^*, X^*) \end{bmatrix} \right) \quad (10)$$

Al igual que en el caso sin ruido tenemos un resultado sobre la distribución condicional que nos permite realizar predicciones:

Proposición:

$$f(X^*)|Y, X \sim \mathcal{N}(m_{X^*|X}, \Sigma_{X^*|X}) \quad (11)$$

Donde la media y covarianza son:

$$m_{X^*|X} = m(X^*) + K(X^*, X) [K(X, X) + \sigma_n^2 \mathbb{I}]^{-1} (Y - m(X)) \quad (12)$$

$$\Sigma_{X^*|X} = K(X^*, X^*) - K(X^*, X) [K(X, X) + \sigma_n^2 \mathbb{I}]^{-1} K(X, X^*) \quad (13)$$

En la figura 3 se puede ver este caso en acción, tomando los mismos parámetros y datos que para el ejemplo sin ruido, y tomando un ruido gaussiano de varianza 1.

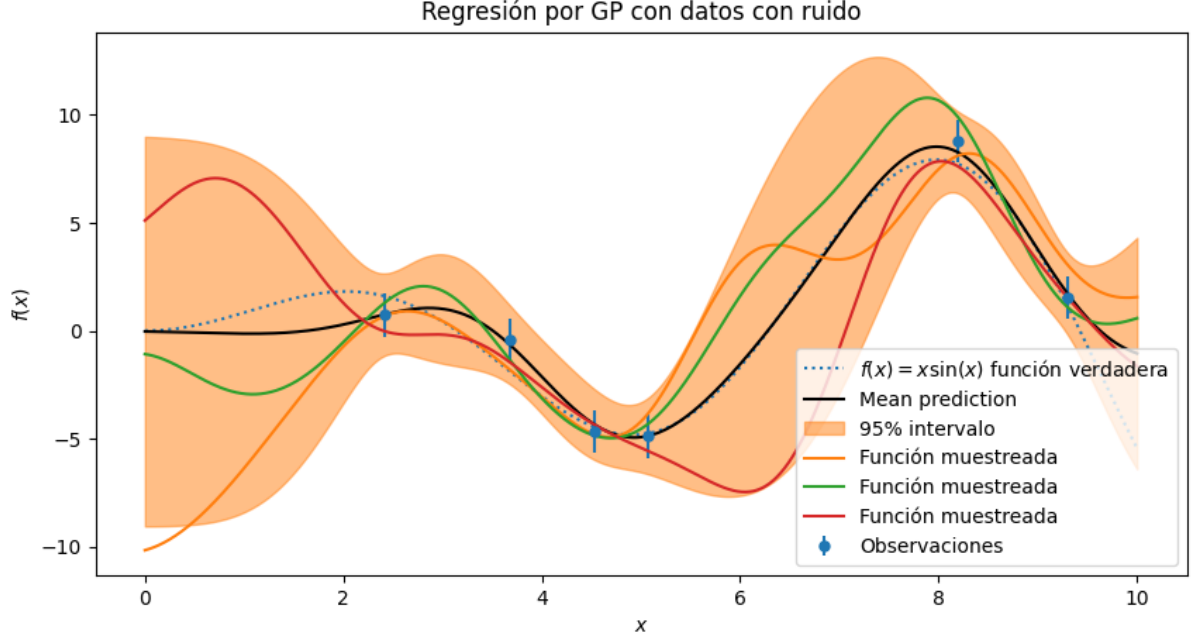


Figura 3. Regresión por \mathcal{GP} con datos *con* ruido. A modo ilustrativo se grafican 3 muestras del posterior del \mathcal{GP} .

Si comparamos la figura 3 con la figura 2 observamos que ahora que los datos tienen ruido, la media de la gaussiana ya no necesariamente interpola a las observaciones (léase no interpola la esperanza de las observaciones).

3. Metodología

3.1. Datos: descripción, preprocesamiento y análisis exploratorio

Como mencionamos en la sección 1 los datos son de producción de petróleo y gas mensual desde 2006 publicada por la Secretaría de Energía de la Nación, abiertos y accesibles desde:

<https://datos.gob.ar/dataset/energia-produccion-petroleo-gas-por-pozo-capitulo-iv>

Estos datos tienen información mensual discriminada por pozo, a saber: producción de petróleo, gas y agua; inyección de agua, gas, CO_2 u otro; yacimiento, cuenca, concesión y provincia a la que pertenece; además de datos sobre la empresa operadora del pozo. En total son 971.882 datos.

Como primera medida se procedió a explorar la calidad de los datos, detectándose casos en donde una o más covariables tenían datos faltantes y esto derivaba en una mala carga del dato pues el faltante siempre ocurría en las covariables finales y, observando los demás datos de buena calidad, se podía deducir que había ocurrido un “corrimiento” hacia las covariables primeras. Es decir, por poner un ejemplo, si el dato constaba de x_1, \dots, x_9 y x_9 tenía dato faltante, mirando datos de buena calidad, se podía deducir que el faltante en realidad estaba entre las x_1, \dots, x_8 y no en la x_9 . Este tipo de datos se descartaron, siendo su cantidad no significativa sobre el total, por la imposibilidad en muchos casos de identificar adecuadamente a cuál o cuáles covariables pertenecía el faltante.

Como uno de los objetivos de este trabajo es realizar predicciones sobre series temporales de la producción de petróleo y gas mensual, pero a primer acercamiento sin discriminar por pozo (ni por cuenca, provincia, etc.) entonces se procedió a sumar para cada mes la producción de todos los pozos, obteniéndose la producción mensual total de petróleo y gas. Cabe aclarar que un pozo puede, y en general lo hace, producir al mismo tiempo tanto petróleo como gas.

Como los procesos gaussianos que se desean usar en este trabajo son alimentados por series de datos ordenados, por ejemplo por el tiempo en el caso de producción de petróleo y gas, como ya se ejemplificó

en la sección 2.3, posterior a la obtención de la producción total por mes se procedió a inspeccionar los datos para corroborar que los mismos eran adecuados para usarse con la técnica antes mencionada. Las figuras 4 y 5 muestran la producción mensual desde Enero de 2006 hasta Marzo de 2023 de petróleo (en m^3) y de gas (en miles de m^3), comprobando que efectivamente estos datos sirven para alimentar a los procesos gaussianos descriptos en la sección 2.2 y ejemplificados en la sección 2.3.

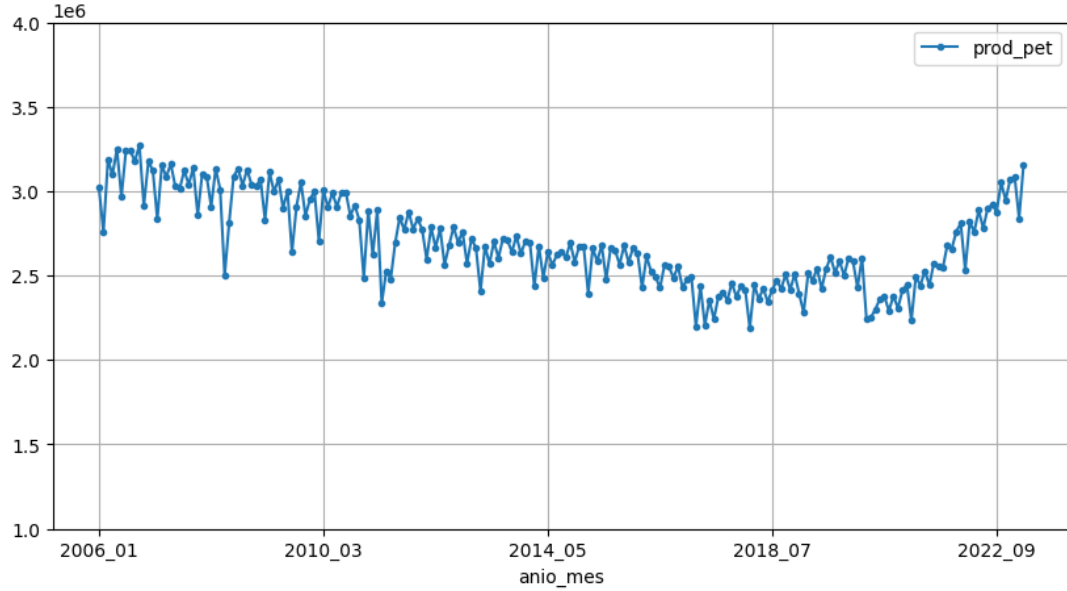


Figura 4. Producción total mensual de petróleo en m^3 en Argentina, desde Enero-2006 hasta Marzo-2023 según la Secretaría de Energía de la Nación.

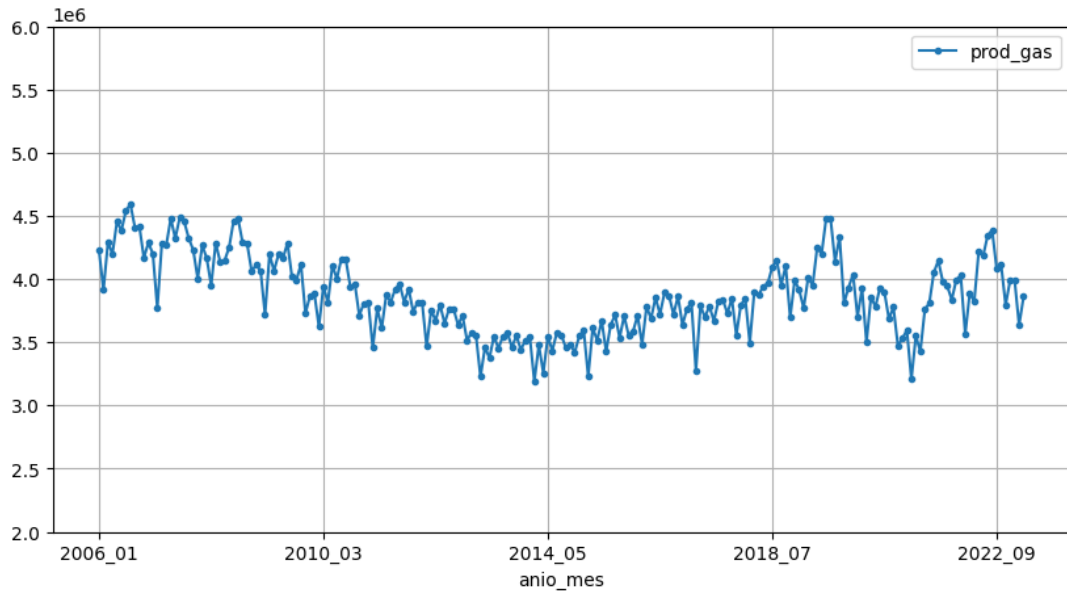


Figura 5. Producción total mensual de gas en miles de m^3 en Argentina, desde Enero-2006 hasta Marzo-2023 según la Secretaría de Energía de la Nación.

3.2. Técnicas de análisis y modelado

Se utilizaron los datos descriptos en la sección 3.1 como serie temporal insumo del modelado y técnica de predicción por \mathcal{GP} descripta y ejemplificada en la sección 2. El split train/test consistió en tomar datos

de entrenamiento desde Enero-2006 y hasta n meses antes del último dato de Marzo-2023 (típicamente $n = 6$ o 12) y predecir tanto interpolando como extrapolando hasta Marzo-2023. El paso de interpolación típico varió entre un décimo y un medio de mes.

Al utilizar \mathcal{GP} como técnica de análisis y modelado, naturalmente se obtiene una incertumbre en la predicción dada por la distribución de la posterior (ver sec. 2.3). Más precisamente si $X^* = (x_1, \dots, x_n)$ son los meses en donde se predice y $f(X^*) = (f(x_1), \dots, f(x_n))$ es la gaussiana multivariada correspondiente, entonces la diagonal de la covarianza contiene la información sobre la incertidumbre de cada $f(x_i)$ cuantificada por la correspondiente varianza, y brindando naturalmente la misma un intervalo de confianza sobre la predicción.

El diseño del kernel es un factor clave a la hora de modelar correctamente por \mathcal{GP} , y más allá de la definición teórica, existen variadas propuestas populares. En este sentido es una limitación y a la vez una libertad, en donde la elección del kernel dependerá en gran medida de la naturaleza del problema y de los datos. Como tanto suma como productos de kernels es un kernel en sí mismos, existen muchas variantes posibles a analizar, que intentarán capturar términos de tendencia a largo plazo, términos de estacionalidad, términos de tendencia a corto plazo, ruido, etc. Para fijar ideas se puede consultar <https://www.cs.toronto.edu/~duvenaud/cookbook/>. De naturaleza más práctica y computacional se puede consultar https://scikit-learn.org/stable/auto_examples/gaussian_process/plot_gpr_co2.html#design-the-proper-kernel. Una referencia más teórica y profunda es [7], en especial el capítulo 2.

Una vez elegido el diseño del kernel, el mismo posee parámetros que influirán de manera relevante en la calidad de la predicción. Los mismos se estiman y optimizan utilizando los datos de entrenamiento elegidos, típicamente mediante máxima verosimilitud o mediante MCMC (por ejemplo se puede consultar la sección 8.3 de [2]). Por supuesto siempre existe la posibilidad de elegirlos por experimentación y/o por conocimiento de dominio.

4. Resultados y discusión

4.1. Resultados: presentación, análisis y discusión

A partir de los datos de producción mensual de petróleo preprocesados como se describió en 3.1, se estandarizaron y se procedió a realizar predicciones usando \mathcal{GP} s. Al tener ahora los datos media cero, solamente hace falta modelar la covarianza (o kernel).

Como elección de diseño básica se utilizó un kernel RBF. Para este caso de elección de diseño de kernel, es natural asumir que los datos posean algo de ruido, a pesar de que en realidad no lo tengan (al menos en teoría), pues el RBF posee dos parámetros: una de varianza, que intenta capturar la tendencia a largo plazo de la serie temporal (o al menos su ruido intrínseco), y otro de escala que intenta capturar la correlación temporal entre datos (es decir el período temporal de influencia).

La magnitud del ruido supuesto para los datos influirá en la optimización de los parámetros del kernel RBF al entrenar el \mathcal{GP} , pues intuitivamente a menor ruido menor será el período de correlación temporal y por lo tanto también menor la tendencia a largo plazo. Dicho de otra forma a menor ruido el kernel se optimizará mucho más localmente.

Recordemos que los datos disponibles van desde Enero de 2006 hasta Marzo de 2023. Supongamos que queremos predecir los últimos seis meses. Asumiendo un ruido en los datos con varianza de 0.05 (recordar que los datos están estandarizados) vemos que el kernel se optimizó de manera muy local, como muestra la figura 6.

Si ahora asumimos que el ruido tiene una varianza de 0.1 ya vemos el efecto menos local de la predicción, como se puede observar en la figura 7. Este efecto se ve intensificado asumiendo ruido con varianza 0.5 como muestra la figura 8, en donde ya vemos que la predicción comienza a capturar mejor la tendencia de los datos.

Esencialmente al aumentar el ruido de los datos, la optimización del parámetro l del kernel RBF es mayor, interpretándose esto como una correlación a mayor período temporal. Esto se puede ver de la definición del kernel RBF que a continuación recordamos:

$$K_{SE}(x_1, x_2) = \sigma^2 \exp\left(-\frac{(x_1 - x_2)^2}{2l^2}\right) \quad (14)$$

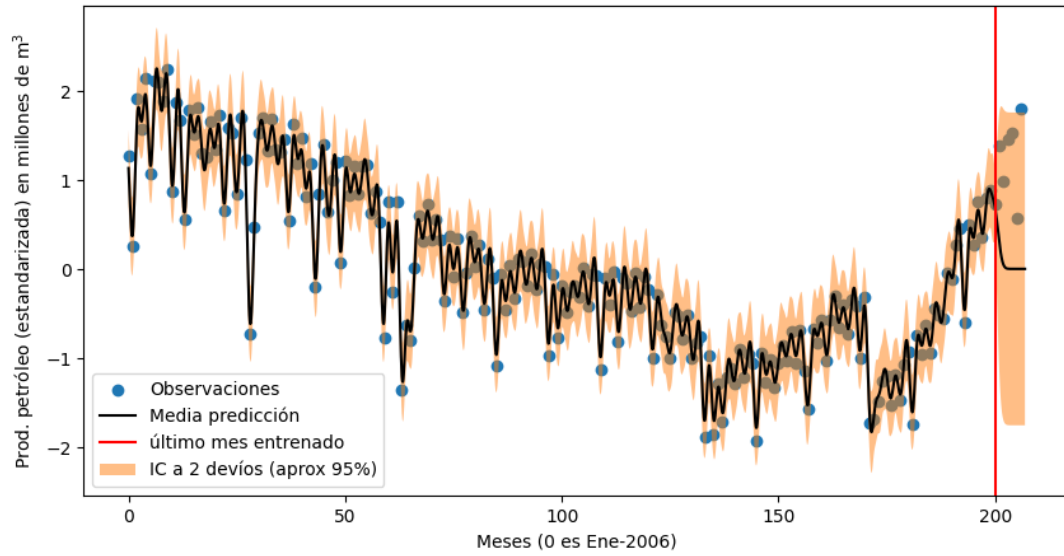


Figura 6. Predicción de los últimos 6 meses de la producción total mensual de petróleo.

Podemos observar que a mayor valor de l mayor deberá ser la diferencia entre x_1 y x_2 para que la misma sea una correlación con influencia relevante. Esto también se puede relacionar con la influencia de la covarianza en la media posterior dada por la ecuación 12.

De estos resultados podemos verificar cómo el kernel RBF al ganar precisión local pierde predicción sobre la tendencia a más largo plazo (figura 6), y viceversa, es decir al ganar sobre tendencia a más largo plazo pierde la tendencia más local (si se quiere la periodicidad local en este caso, como se ve en la figura 8).

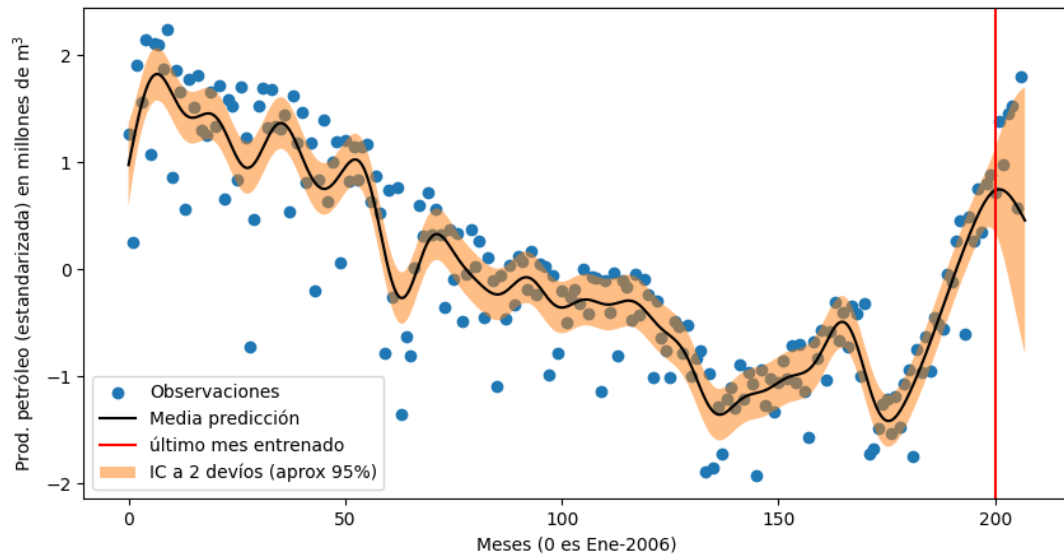


Figura 7. Predicción de los últimos 6 meses de la producción total mensual de petróleo.

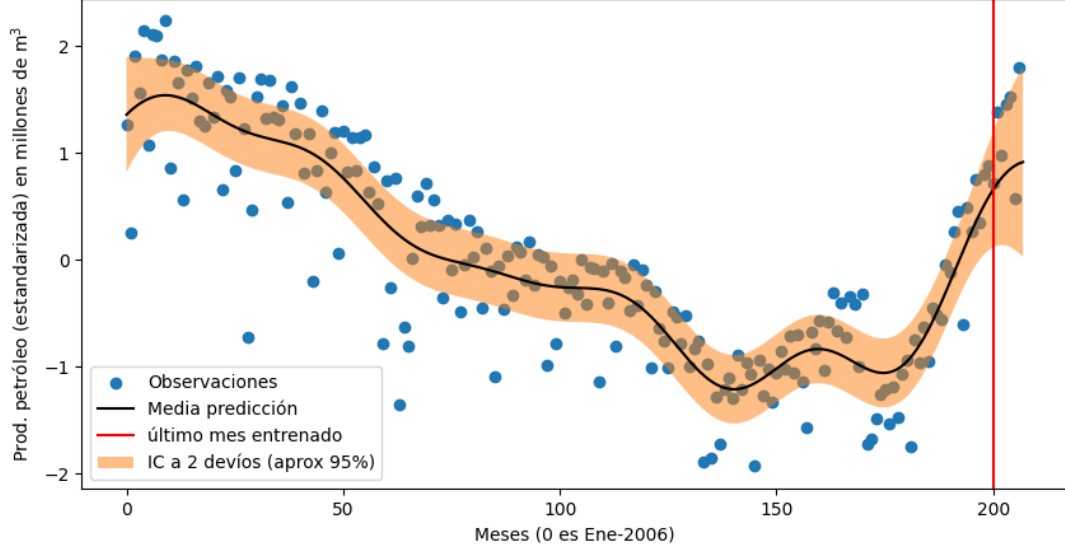


Figura 8. Predicción de los últimos 6 meses de la producción total mensual de petróleo.

4.2. Limitaciones y mejoras

Resulta evidente, aunque no deja de ser importante mencionarlo, que una limitación general de los \mathcal{GP} es el modelado mediante gaussianas multivariadas. Menos evidente resulta la limitación de utilizar un solo kernel, en particular un kernel RBF como se utilizó para producir los resultados de la sección 3.1, pues resulta casi imposible capturar tendencias de diferentes períodos temporales: largo plazo, corto plazo, estacionalidades, ruido, periodicidades, etc. En este sentido el rediseño del kernel, combinando varios, puede conllevar a mejoras en la predicción (ver por ejemplo [7]).

Relacionado con lo anterior, la ventana de tiempo hacia atrás utilizada como observaciones dato, e incluso también la ventana de tiempo hacia adelante utilizada para predecir, son cuestiones que influyen a la hora tanto del diseño del kernel como de la optimización de sus parámetros, pues tienen influencia en las tendencias temporales que nuestro \mathcal{GP} deberá capturar.

Transversalmente a todo el modelado, las covariables utilizadas bien pueden no ser únicamente el paso del tiempo, y por lo tanto una posible mejora puede provenir de que el insumo del \mathcal{GP} sean más de un feature. Incluso la predicción del \mathcal{GP} bien puede ser un vector y ya no un simple escalar, y en el caso particular de la predicción de la producción de petróleo y gas, puede implicar esto una mejora pues se sabe que existe una correlación en la producción de ambos hidrocarburos. Para una descripción algo más técnica de \mathcal{GP} multi output y de cómo se hace naturalmente uso de la correlación entre los outputs, se puede consultar la sección 8.5.3 de [2].

Desde el punto de vista de conocimiento de dominio del problema, resulta natural experimentar prediciendo por zonas de explotación de hidrocarburos, y no solamente sumando la producción total, pues es sabido que cada zona puede poseer sus propios tiempos y volúmenes de producción. Más aún discriminar por antigüedad de pozos y/o yacimientos también es de relevancia experimentar, pues es sabido que el volumen de producción puede estar correlacionado a la antigüedad del pozo y/o del yacimiento.

5. Conclusiones

La figura 8 es un ejemplo de lo razonable que es utilizar a los \mathcal{GP} para realizar predicciones sobre series temporales. Al mismo tiempo, y junto con las figuras 7 y 6, también es una muestra de que el diseño del kernel, tanto en naturaleza como en optimización de sus hiperparámetros, resulta clave si se desea utilizar la técnica como herramienta que asista a la predicción sobre series temporales de producción de petróleo y gas.

Como grandes ventajas de los \mathcal{GP} mencionamos su fácil entrenamiento y su natural medida de incertidumbre sobre las predicciones, lo que facilita la elección de una métrica de performance de la técnica.

Otra ventaja es la facilidad tanto para interpolar como extrapolar a partir de los datos de entrenamiento, siendo tanto el costo como el cálculo exactamente el mismo en ambos casos.

De los resultados y análisis presentados en la sección 4.1 concluimos que, incluso utilizando un diseño de kernel mónico RBF, se puede observar la flexibilidad a la hora de querer capturar en mayor o menor medida la tendencia a largo o corto plazo de la serie de tiempo.

Como conclusión vemos que los \mathcal{GP} son de utilidad para predecir sobre series temporales de producción de petróleo y gas, como también se menciona en [6], en donde el diseño del kernel y su entrenamiento resultan factores relevantes sino claves a la hora de adaptar la técnica a los datos en particular utilizados en este trabajo, y en donde la flexibilidad de adaptación del kernel a los datos y la medida de incertidumbre en la predicción dada naturalmente por los \mathcal{GP} , distinguen a la técnica de otras más clásicas.

Ideas que todavía no puse en ningún lado

- Relacionar los modelos aditivos de series de tiempo clásicos: tendencia largo plazo + estacionalidad1 + estacionalidad2 + ... + irregularidades + ruido, con la aditividad de kernels. Intuición de esto: cada kernel modela un término del clásico.

Idea general de esto: el \mathcal{GP} es el análogo probabilístico al modelo clásico.

- Hablar más largo y tendido sobre los diferentes kernels populares, y sobre el diseño de kernels en general.
- Hablar sobre cómo esa inversa de la K en la media de la posterior para predecir, se puede interpretar como un leverage.

Idea general de esto: hablar más sobre la interpretación de la influencia de los datos de entrenamiento en la posterior para predecir. Se puede ver [7].

- Comentar que el ruido en la predicción posterior se puede interpretar como una especie de regularizador (por supuesto influenciado por el diseño del kernel, y con influencia en determinados hiperparámetros).

Referencias

- [1] Robert B. Gramacy, *Gaussian process modeling, design and optimization for the applied sciences*, <https://bookdown.org/rbg/surrogates/> 2023-03-05.
- [2] Felipe Tobar, *Aprendizaje de máquinas. Cap. 8: procesos gaussianos*, https://raw.githubusercontent.com/GAMES-UCHile/Curso-Aprendizaje-de-Maquinas/master/notas_de_clase.pdf 2021.
- [3] Carl E. Rasmussen; Christopher K. I. Williams, *Gaussian Processes for Machine Learning*, <http://gaussianprocess.org/gpml/> MIT Press, 2006.
- [4] Kevin P. Murphy, *Machine Learning: A Probabilistic Perspective. Ch. 15: Gaussian processes*, http://noiselab.ucsd.edu/ECE228/Murphy_Machine_Learning.pdf MIT Press, 2012.
- [5] Rob J. Hyndman; George Athanasopoulos, *Forecasting: Principles and Practice*, <https://otexts.com/fpp3/> 3ra. ed., 2021.
- [6] Osaro, Etinosa; Okorie, Vivian; Aloroyo, Sonia, *Exploring the Usefulness of Gaussian Process Regression for the Prediction of Oil, Water and Gas Production Rates*, Abril 2023, 14. 10.35248/2157-7463.23.14.506. https://www.researchgate.net/profile/Etinosa-Osaro/publication/370099861_

Exploring_the_Usefulness_of_Gaussian_Process_Regression_for_the_Prediction_
of_Oil_Water_and_Gas_Production_Rates/links/643ed71a39aa471a5248f03a/
Exploring-the-Usefulness-of-Gaussian-Process-Regression-for-the-Prediction-of-Oil-Water-and-Gas-Production.
pdf

- [7] David K. Duvenaud, *Automatic Model Construction with Gaussian Processes*,
<https://www.cs.toronto.edu/~duvenaud/thesis.pdf>
PhD thesis, 2014, University of Cambridge.
- [8] Código fuente utilizado para el trabajo,
https://github.com/sebasped/tti_gp/blob/main/tt1_2023cuat1.ipynb