

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

G. Sebastián Pedersen

sebasped@gmail.com

Inicio: Lun 27 de Marzo de 2023. Última actualización: 1 de mayo de 2023

## Índice

<b>1. Dataset, pregunta y objetivo</b>	<b>1</b>
1.1. Conjunto de datos . . . . .	1
1.2. Pregunta y objetivos . . . . .	1
1.3. Técnicas . . . . .	2
<b>Referencias</b>	<b>2</b>

## 1. Dataset, pregunta y objetivo

### 1.1. Conjunto de datos

Los datos elegidos son de producción de petróleo y gas mensual desde 2006 publicada por la Secretaría de Energía de la Nación. Los mismos contienen el detalle mensual de la producción de petróleo, gas y agua, discriminada principalmente por pozo, yacimiento, cuenca, concesión y provincia, entre otras cosas. Los datos son abiertos y accesibles desde:

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

La elección de este conjunto de datos en parte es motivada porque el autor trabajó en la industria petrolera durante algunos años, y por lo tanto posee cierto conocimiento y familiaridad con la industria y el negocio.

### 1.2. Pregunta y objetivos

Las preguntas u objetivos son dos, una de carácter metodológico y otro de carácter práctico, y ambas hacen uso de una técnica en particular, con lo cual resulta adecuado mencionarla desde un principio: los procesos gaussianos (ver [1, 2, 3, 4]).

Por un lado hay una pregunta u objetivo metodológico sobre los procesos gaussianos que involucrará ahondar en ellos y entenderlos, darles un marco teórico adecuado e investigar sus potencialidades y limitaciones, al menos desde un punto de vista teórico. Con esto se espera tener una idea metodológica adecuada de los procesos gaussianos, tanto de su formulación teórica como de sus potenciales aplicaciones. La relevancia de este objetivo radica en que al no ser los procesos gaussianos parte de la currícula del posgrado, es una oportunidad para extrapolar conocimientos y habilidades adquiridos en otros temas y volcarlos al estudio de una técnica no antes estudiada.

Por otro lado se investigará en concreto las bondades de los procesos gaussianos para realizar predicciones sobre series temporales, tomando como ejemplo práctico en particular el dataset mencionado en 1.1. La relevancia de esta pregunta radica en que intentar predecir la producción de hidrocarburos

en la Argentina es de suma importancia en tanto y en cuanto es una materia prima indispensable para cubrir las necesidades energéticas del país, y por lo tanto tener una estimación de su producción tiene consecuencias en varias área estratégicas, a saber: planeamiento energético interno, exportación como commodity, seguimiento del desarrollo de la explotación de recursos naturales, planeamiento de necesidades en infraestructura (gasoductos, maquinaria, etc.) entre otras.

### 1.3. Técnicas

Como ya se mencionó en 1.2 la técnica principal serán los procesos gaussianos. Como uno de los objetivos es estudiar metodológicamente a los procesos gaussianos, y el otro es investigar la potencialidad de los mismos para realizar predicciones sobre series temporales, resulta adecuado emplear otras técnicas a modo de benchmark para poder comparar tanto resultados predictivos prácticos, así como diferencias y similitudes entre los marcos teóricos, y poder entonces entender mejor desde los dos puntos de vista por qué la técnica es más o menos efectiva, al menos como comparación relativa contra otras. La propuesta de técnicas a emplear es:

- Modelos naive: mean y drift, por ejemplo. (ver [5] sec. 5.2 )
- Modelos GAM (modelos aditivos generalizados): STL (Seasonal and Trend descomp usando Loess, ver [5] secs. 3.6 y 12.1), Prophet (ver [5] sec. 12.2).
- Procesos gaussianos (ver [1, 2, 3, 4]).

## 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](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](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.