

Predicción e incertidumbre en sistemas de energía renovable

Maria Luisa Daza Torres

Matemáticas para Energías Renovables y Resiliencia Climática

September 9, 2026



¿Por qué necesitamos predicciones?

La generación renovable depende de variables meteorológicas que cambian en el tiempo y que no conocemos exactamente de antemano.

- Solar: irradiancia, nubosidad, temperatura.
- Eólica: velocidad del viento, densidad del aire, temperatura.
- La demanda eléctrica también cambia en el tiempo.

Por tanto, la generación disponible dentro de unas horas o al día siguiente no se conoce exactamente en el presente.

El papel de los pronósticos

Los pronósticos son esenciales para:

- planificar la operación del mercado eléctrico;
- balancear la oferta y la demanda de energía;
- prevenir sobrecargas en elementos críticos de la red;
- tomar decisiones operativas considerando el riesgo.

Predecir no es el objetivo final: la predicción es información para tomar decisiones.

Decisiones en una mini-grid

Consideremos una mini-grid que abastece una demanda D_t mediante generación renovable, almacenamiento y generación de respaldo.

En cada tiempo t , debe mantenerse el balance:

$$\text{Generación renovable} + \text{Almacenamiento} + \text{Respaldo} \geq D_t.$$

Sin embargo, la generación futura

$$G_{t+1}, G_{t+2}, \dots$$

no se conoce exactamente al tomar las decisiones.

Esto afecta decisiones como:

- dimensionar la capacidad de generación y almacenamiento;
- cargar o descargar la batería;
- activar la generación de respaldo.

Fuentes de incertidumbre

En una mini-grid existen diferentes fuentes de incertidumbre:

Generación renovable	Demanda	Disponibilidad del sistema
----------------------	---------	----------------------------

En esta sesión nos concentraremos en la incertidumbre asociada con la generación:

Solar

Eólica

Buscaremos entender y predecir su evolución en el tiempo y cuantificar la incertidumbre de los pronósticos.

Variabilidad de la generación renovable

La producción depende de las condiciones atmosféricas y de las características del sistema de generación.

Factores externos

Solar

- Irradiancia solar.
- Nubosidad.
- Temperatura.

Eólica

- Velocidad y dirección del viento.
- Temperatura.
- Presión y humedad.

Factores internos

- Capacidad instalada.
- Características de paneles y turbinas.
- Mantenimiento e indisponibilidad.
- Envejecimiento de los equipos.
- Incorporación o retiro de capacidad.

Predicción como problema matemático

Supongamos que hemos observado una serie temporal hasta el tiempo t :

$$y_1, y_2, \dots, y_t.$$

Queremos predecir sus valores futuros:

$$y_{t+1}, y_{t+2}, \dots, y_{t+h}$$

donde h es el **horizonte de predicción**.

En nuestro caso, y_t puede representar:

- potencia fotovoltaica generada en el tiempo t ;
- potencia eólica generada en el tiempo t .

Horizontes de predicción

El horizonte de predicción depende de la decisión que queremos tomar.

Horizonte	Aplicación
Minutos–horas	Control y almacenamiento
Horas–1 día	Operación de la mini-grid
Días	Planificación operativa
Meses–años	Dimensionamiento e inversión

Diferentes decisiones requieren diferentes horizontes y diferentes tipos de predicción.

Predicción puntual y probabilística

Una **predicción puntual** proporciona un único valor:

$$\hat{y}_{t+h} = 120 \text{ kW.}$$

Una **predicción probabilística** describe los diferentes valores posibles y sus probabilidades:

$$p(y_{t+h} \mid y_{1:t}).$$

Por ejemplo,

$$P(95 < y_{t+h} < 145 \text{ kW} \mid y_{1:t}) = 0.90.$$

Para tomar decisiones importa cuánto esperamos generar y también la incertidumbre asociada con esa estimación.

¿Por qué importa la incertidumbre?

Imaginemos dos pronósticos para mañana:

A: 100 kW

B: 100 kW

Tienen la misma predicción puntual.

Pero:

$A : 95 - 105 \text{ kW}$

$B : 40 - 160 \text{ kW}$

Las decisiones de almacenamiento y respaldo deberían ser muy diferentes.

Métodos de predicción

Los métodos para predecir generación solar y eólica pueden organizarse en tres grandes grupos:

- Modelos físicos:

Meteorología $\xrightarrow{\text{modelo físico}}$ Potencia

Utilizan las relaciones físicas que gobiernan el sistema.

- Modelos estadísticos / machine learning:

Datos históricos $\xrightarrow{\text{modelo}}$ Potencia futura

Aprenden relaciones directamente de los datos.

- Modelos híbridos

Física + Datos

Combinan conocimiento físico con aprendizaje estadístico.

Modelos físicos

Para energía solar:

$$P_t^{PV} = f(\text{irradiancia, temperatura, características del panel, } \dots)$$

Para energía eólica:

$$P_t^W = f(\text{velocidad, densidad, características de la turbina, } \dots)$$

Ventaja: incorporan conocimiento del proceso físico.

Limitación: para predecir la potencia futura necesitamos conocer o predecir las variables meteorológicas futuras.

La incertidumbre meteorológica se propaga hacia la predicción de generación.

Utilizan observaciones históricas para aprender la dinámica temporal:

$$y_t = f(y_{t-1}, y_{t-2}, \dots, x_t) + \epsilon_t.$$

- Regresión.
- Modelos autorregresivos.
- ARIMA.
- SARIMA.
- Modelos de espacio de estados.
- Machine learning.
- Redes neuronales.

Modelos híbridos

No tenemos que elegir necesariamente entre física y estadística.
Podemos combinar:

$$\underbrace{\text{Modelo físico}}_{\text{conocimiento}} + \underbrace{\text{Modelo estadístico}}_{\text{datos}}$$

Por ejemplo:

$$P_t^{obs} = P_t^{physical} + \epsilon_t,$$

donde ϵ_t representa el error o las dinámicas que el modelo físico no captura.

Cuantificar la Incertidumbre en las predicciones

Los modelos de predicción contienen cantidades que no conocemos exactamente:

$$\theta = (\theta_1, \dots, \theta_p).$$

Por ejemplo:

- coeficientes de regresión;
- parámetros autorregresivos;
- parámetros físicos;
- estados no observados.

Cuantificar la incertidumbre en las predicciones

Una estimación puntual proporciona un único valor,

$$\hat{\theta},$$

En un enfoque probabilístico, describimos nuestra incertidumbre sobre los parámetros mediante una distribución:

$$p(\theta \mid y).$$

Esta incertidumbre se propaga a las predicciones futuras:

$$\boxed{p(\theta \mid y) \longrightarrow p(y_{t+h} \mid y)}$$

Tres fuentes de incertidumbre

- 1 Parámetros: no conocemos exactamente los coeficientes del modelo.
- 2 Variabilidad futura: incluso con parámetros conocidos, el clima cambia.
- 3 Especificación: ningún modelo representa perfectamente el proceso físico.

La inferencia bayesiana representa muy bien las dos primeras. La tercera debe estudiarse comparando modelos y verificando predicciones.

Inferencia Bayesiana

Supongamos que nuestro modelo depende de parámetros desconocidos

$$\theta = (\theta_1, \dots, \theta_p).$$

Antes de observar los datos, representamos nuestro conocimiento sobre estos parámetros mediante una distribución

$$p(\theta).$$

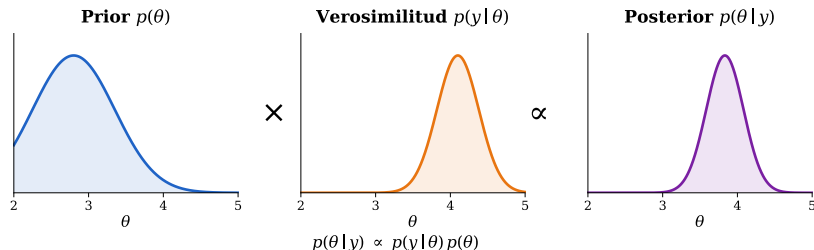
Después observamos datos y , cuya distribución depende de los parámetros:

$$p(y \mid \theta).$$

El teorema de Bayes permite actualizar nuestro conocimiento:

$$p(\theta \mid y) = \frac{p(y \mid \theta)p(\theta)}{p(y)}$$

Prior, verosimilitud y posterior



- **Prior** representa nuestro conocimiento sobre θ antes de observar los nuevos datos.
- **Verosimilitud** relaciona los datos observados con los posibles valores de θ .
- **Posterior** combina la información previa con la información contenida en los datos.

¿Qué obtenemos de la distribución posterior?

La inferencia Bayesiana proporciona una distribución para los parámetros desconocidos:

$$p(\theta \mid y).$$

A partir de ella podemos obtener:

- estimaciones puntuales, por ejemplo

$$\mathbb{E}[\theta \mid y];$$

- intervalos de probabilidad;
- probabilidades sobre los parámetros;
- predicciones de observaciones futuras.

De la inferencia a la predicción

Después de observar $y_{1:t}$, tenemos información sobre los parámetros:

$$p(\theta \mid y_{1:t}).$$

Pero nuestro objetivo es predecir una observación futura y_{t+1} .

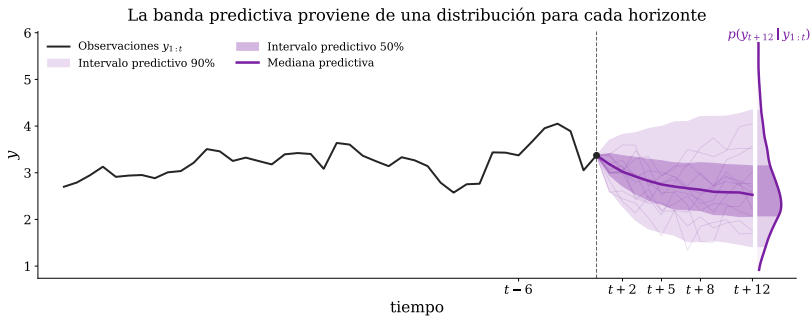
La distribución predictiva posterior es

$$p(y_{t+1} \mid y_{1:t}) = \int p(y_{t+1} \mid \theta, y_{1:t}) p(\theta \mid y_{1:t}) d\theta.$$

Esta expresión combina:

$$\underbrace{p(y_{t+1} \mid \theta, y_{1:t})}_{\text{variabilidad futura}} \quad y \quad \underbrace{p(\theta \mid y_{1:t})}_{\text{incertidumbre en los parámetros}} \quad .$$

Interpretación de la distribución predictiva



La predicción no es solamente un valor $\hat{y}_{t+1}$, sino una distribución:

$$p(y_{t+1} | y_{1:t}).$$

Esta distribución describe los posibles valores futuros y la incertidumbre asociada a ellos.

Una predicción probabilística describe qué puede ocurrir y con qué incertidumbre.

Aproximación de la distribución predictiva

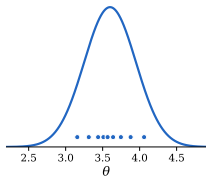
La integral predictiva no siempre puede calcularse de forma analítica. Si disponemos de muestras de la posterior,

$$\theta^{(1)}, \dots, \theta^{(M)} \sim p(\theta \mid y_{1:t}),$$

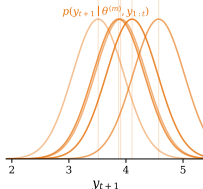
podemos aproximarla mediante Monte Carlo:

$$p(y_{t+1} \mid y_{1:t}) \approx \frac{1}{M} \sum_{m=1}^M p(y_{t+1} \mid \theta^{(m)}, y_{1:t}).$$

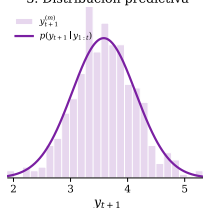
1. Muestras posteriores $\theta^{(m)}$



2. Predictivas condicionales



3. Distribución predictiva



Una simulación $y_{t+1}^{(m)} \sim p(y_{t+1} \mid \theta^{(m)}, y_{1:t})$ por cada muestra posterior

Simulación predictiva

En lugar de construir explícitamente la mezcla anterior, podemos generar directamente valores futuros.

Para cada muestra posterior

$$\theta^{(m)} \sim p(\theta \mid y_{1:t}),$$

simulamos

$$y_{t+1}^{(m)} \sim p(y_{t+1} \mid \theta^{(m)}, y_{1:t}).$$

Entonces

$$y_{t+1}^{(1)}, \dots, y_{t+1}^{(M)}$$

son muestras de la distribución predictiva posterior

$$p(y_{t+1} \mid y_{1:t}).$$

Cálculo mediante simulación

El procedimiento completo puede resumirse como

$$y_{1:t} \longrightarrow p(\theta \mid y_{1:t}) \longrightarrow \theta^{(m)} \longrightarrow y_{t+1}^{(m)}.$$

- Cuando la posterior tiene una forma cerrada, podemos trabajar directamente con ella.
- Cuando no la tiene, podemos obtener muestras mediante métodos como MCMC:

$$\theta^{(1)}, \dots, \theta^{(M)} \sim p(\theta \mid y_{1:t}).$$

Existen implementaciones en software como Stan, JAGS, twalk y otros paquetes de inferencia Bayesiana.

Ejemplo: un primer modelo bayesiano

Comencemos con el modelo más sencillo. Supongamos que las observaciones fluctúan alrededor de una media constante:

$$y_t = \mu + \varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma^2).$$

Equivalentemente,

$$y_t \mid \mu, \sigma^2 \sim N(\mu, \sigma^2), \quad t = 1, \dots, T.$$

Suponemos inicialmente que σ^2 es conocida y que, dado μ , las observaciones son independientes.

Asignamos una distribución prior a la media desconocida:

$$\mu \sim N(\mu_0, \tau_0^2).$$

Actualizando nuestro conocimiento sobre μ

Tenemos

$$y_t \mid \mu \sim N(\mu, \sigma^2), \quad \mu \sim N(\mu_0, \tau_0^2).$$

Después de observar $y_1, \dots, y_T$, aplicamos Bayes:

$$p(\mu \mid y_{1:T}) \propto p(y_{1:T} \mid \mu) p(\mu).$$

En este caso, la posterior también es Normal:

$$\mu \mid y_{1:T} \sim N(\mu_T, \tau_T^2)$$

donde

$$\tau_T^2 = \left(\frac{1}{\tau_0^2} + \frac{T}{\sigma^2} \right)^{-1},$$
$$\mu_T = \tau_T^2 \left(\frac{\mu_0}{\tau_0^2} + \frac{T \bar{y}}{\sigma^2} \right).$$

La media posterior como promedio ponderado

La media posterior puede escribirse como

$$\mu_T = \frac{\frac{1}{\tau_0^2}\mu_0 + \frac{T}{\sigma^2}\bar{y}}{\frac{1}{\tau_0^2} + \frac{T}{\sigma^2}}.$$

Es decir,

$$\mu_T = w_{\text{prior}}\mu_0 + w_{\text{datos}}\bar{y},$$

donde

$$w_{\text{prior}} = \frac{1/\tau_0^2}{1/\tau_0^2 + T/\sigma^2}, \quad w_{\text{datos}} = \frac{T/\sigma^2}{1/\tau_0^2 + T/\sigma^2},$$

y

$$w_{\text{prior}} + w_{\text{datos}} = 1.$$

La fuente con mayor precisión recibe mayor peso.

De la inferencia a la predicción

Con los datos hemos aprendido sobre μ :

$$p(\mu \mid y_{1:T}).$$

Ahora queremos predecir una nueva observación y_{T+1} :

$$p(y_{T+1} \mid y_{1:T}) = \int p(y_{T+1} \mid \mu) p(\mu \mid y_{1:T}) d\mu.$$

En este ejemplo la integral puede calcularse exactamente:

$$y_{T+1} \mid y_{1:T} \sim N\left(\mu_T, \sigma^2 + \tau_T^2\right).$$

Por tanto, la incertidumbre predictiva contiene

$$\underbrace{\sigma^2}_{\text{variabilidad futura}} + \underbrace{\tau_T^2}_{\text{incertidumbre sobre } \mu}.$$

Caso de estudio: Cuernavaca

Consideraremos producción fotovoltaica horaria durante 2019.

- Ubicación: 18.9331° N, 99.2600° O.
- Fuente: Renewables.ninja con reanálisis MERRA-2.
- Sistema de referencia: 1 kWp fijo.
- Inclinación: 19° , orientación sur.
- Pérdidas: 10%.

La variable

$$y_t = \text{potencia producida por kWp}$$

representa la generación fotovoltaica en la hora t .

Caso de estudio: generación solar

Antes de elegir un modelo, observemos la estructura de los datos.

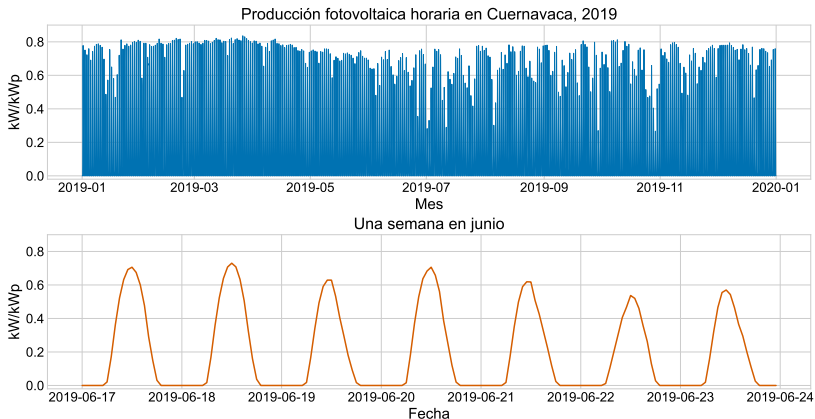


Figure 1: Arriba: generación horaria durante 2019. Abajo: una semana representativa. La producción sigue un ciclo diario y es nula durante la noche.

Estructura temporal de la generación

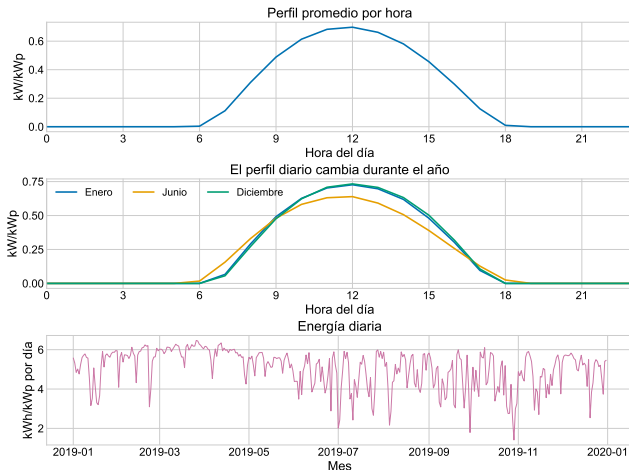


Figure 2: Izquierda: perfil medio por hora. Centro: perfiles medios en enero, junio y diciembre. Derecha: generación diaria acumulada, en kWh/kWp.

Relación con observaciones pasadas

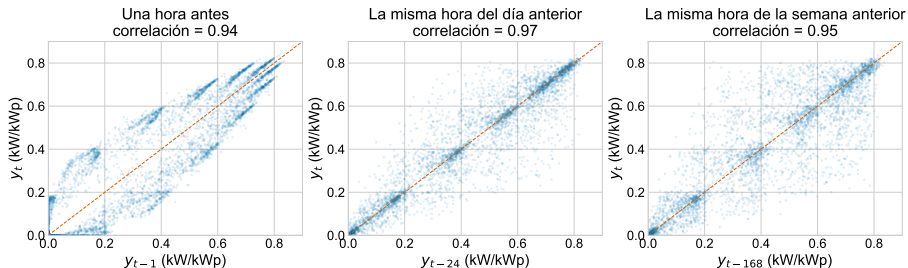


Figure 3: Cada punto representa una hora. Comparamos la producción actual con la de una hora, un día y una semana antes. Las relaciones observadas pueden reflejar el ciclo solar, no necesariamente memoria residual.

y_{t-1}	:	dependencia de corto plazo,
y_{t-24}	:	ciclo diario,
y_{t-168}	:	ciclo semanal.

¿Es suficiente una media constante?

Nuestro primer modelo supone

$$y_t = \mu + \varepsilon_t.$$

Por tanto,

$$\mathbb{E}[y_t \mid \mu] = \mu$$

para cualquier hora del día y cualquier época del año.

Pero la generación solar presenta una estructura temporal clara. [La media de la serie no es constante en el tiempo.](#)

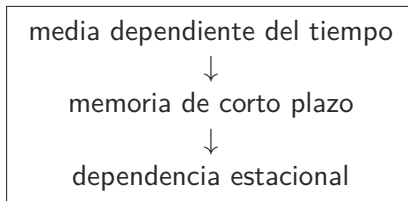
¿Qué estructura observamos?

Al observar la serie en diferentes escalas aparecen:

- ceros durante la noche;
- un ciclo diario fuerte;
- diferencias entre días;
- cambios a lo largo del año;
- persistencia entre horas cercanas.

Construcción progresiva del modelo

Partiremos de modelos sencillos e incorporaremos estructura solo cuando los datos lo justifiquen:



Usamos las primeras 8,592 horas para ajustar los modelos y reservamos las últimas 168 horas para evaluar los pronósticos.

De una media constante a una media temporal

Nuestro primer modelo era

$$y_t = \mu + \varepsilon_t.$$

Pero la media cambia con el tiempo.

Reemplazamos entonces

$$\mu \longrightarrow m_t$$

y obtenemos

$$y_t = m_t + \varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma^2).$$

Ahora

$$\mathbb{E}[y_t] = m_t.$$

No estamos construyendo un modelo diferente: estamos generalizando el anterior.

Representando el patrón periódico

Primero describimos la producción esperada según la hora y la época del año:

$$\begin{aligned} m_t = & \beta_0 + \beta_1 \sin(2\pi h_t/24) + \beta_2 \cos(2\pi h_t/24) \\ & + \beta_3 \sin(4\pi h_t/24) + \beta_4 \cos(4\pi h_t/24) \\ & + \beta_5 \sin(2\pi d_t/365) + \beta_6 \cos(2\pi d_t/365). \end{aligned}$$

donde

- h_t : hora del día;
- d_t : día del año;
- m_t : generación típica para esa posición del calendario.

Los términos diarios representan el ciclo día–noche y los términos anuales permiten que el patrón cambie durante el año.

Formulación como regresión lineal

Aunque utilizamos senos y cosenos, el modelo es lineal en los parámetros:

$$m_t = \mathbf{x}_t^\top \boldsymbol{\beta}.$$

Por tanto,

$$y_t = \mathbf{x}_t^\top \boldsymbol{\beta} + \varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma^2).$$

Para todas las observaciones,

$$\mathbf{y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2 I).$$

Inferencia bayesiana del patrón periódico

El modelo es

$$\mathbf{y} \mid \beta, \sigma^2 \sim N(X\beta, \sigma^2 I).$$

Usamos el prior conjugado

$$\beta \mid \sigma^2 \sim N(\mathbf{m}_0, \sigma^2 V_0), \quad \sigma^2 \sim \text{InvGamma}(a_0, b_0).$$

La posterior conserva la misma forma:

$$\beta \mid \sigma^2, \mathbf{y} \sim N(\mathbf{m}_n, \sigma^2 V_n), \quad \sigma^2 \mid \mathbf{y} \sim \text{InvGamma}(a_n, b_n),$$

donde

$$V_n = (V_0^{-1} + X^\top X)^{-1}, \quad \mathbf{m}_n = V_n (V_0^{-1} \mathbf{m}_0 + X^\top \mathbf{y}).$$

Distribución predictiva

Para un nuevo tiempo t_* , el modelo es

$$y_* = \mathbf{x}_*^\top \boldsymbol{\beta} + \varepsilon_*, \quad \varepsilon_* \sim N(0, \sigma^2).$$

Como $\boldsymbol{\beta}$ es desconocido, consideramos todos sus posibles valores de acuerdo con su distribución posterior:

$$p(y_* | \mathbf{y}, \sigma^2) = \int p(y_* | \boldsymbol{\beta}, \sigma^2) p(\boldsymbol{\beta} | \mathbf{y}, \sigma^2) d\boldsymbol{\beta}.$$

Si

$$\boldsymbol{\beta} | \mathbf{y}, \sigma^2 \sim N(\mathbf{m}_n, \sigma^2 V_n),$$

entonces

$$y_* | \mathbf{y}, \sigma^2 \sim N\left(\mathbf{x}_*^\top \mathbf{m}_n, \sigma^2 \left[1 + \mathbf{x}_*^\top V_n \mathbf{x}_*\right]\right).$$

A priori para nuestro ejemplo

El vector de regresores tiene $p = 7$ componentes: intercepto, dos armónicos diarios y un armónico anual.

Usamos

$$\beta \mid \sigma^2 \sim N(\mathbf{m}_0, \sigma^2 V_0), \quad \sigma^2 \sim \text{InvGamma}(a_0, b_0),$$

con los valores

$$\mathbf{m}_0 = \mathbf{0}_7, \quad V_0 = 10I_7, \quad a_0 = 2, \quad b_0 = 0.05.$$

- $\mathbf{m}_0 = \mathbf{0}$: antes de ver datos, no favorecemos coeficientes positivos ni negativos.
- $V_0 = 10I_7$: prior amplio para los coeficientes, en la escala de σ^2 .
- $a_0 = 2$ y $b_0 = 0.05$: prior débilmente informativo para la variabilidad, compatible con una serie acotada aproximadamente entre 0 y 1 kW/kWp.

Evaluación de las predicciones

Para evaluar la predicción comparamos los valores observados y_t con las predicciones puntuales $\hat{y}_t$.

Error absoluto medio (MAE)

$$\text{MAE} = \frac{1}{n} \sum_{t=1}^n |y_t - \hat{y}_t|$$

Mide el tamaño promedio de los errores.

Raíz del error cuadrático medio (RMSE)

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{t=1}^n (y_t - \hat{y}_t)^2}$$

Penaliza más fuertemente los errores grandes.

En ambas métricas, valores menores indican mejores predicciones.

Desempeño del patrón periódico

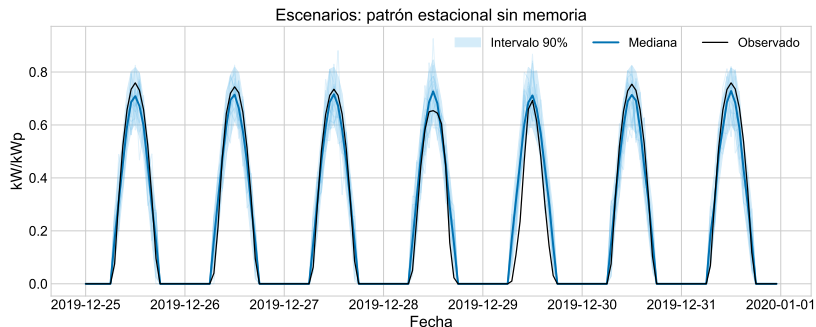
Modelo	MAE	RMSE	Cobertura 90%
Semana anterior	0.0349	0.0816	–
Patrón periódico Bayesiano	0.0307	0.0550	93.5%

El patrón diario y anual explica una parte importante de la serie.
Sin embargo, el modelo supone errores temporalmente independientes:

$$\varepsilon_t \perp \varepsilon_{t-1}.$$

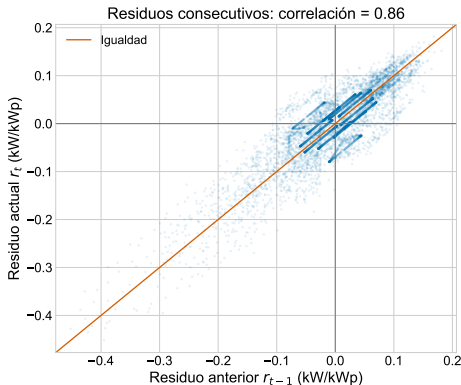
Las desviaciones respecto al patrón pueden conservar estructura temporal.

Pronóstico con patrón estacional bayesiano



La línea negra muestra la generación observada; la línea azul, la mediana de la distribución predictiva; y la banda azul clara, el intervalo predictivo central al 90%. El modelo reproduce el ciclo diario y la amplitud estacional, pero no ajusta por completo las desviaciones puntuales respecto al patrón esperado.

Dependencia en los residuos



Primero ajustamos el patrón periódico y calculamos

$$r_t = y_t - \hat{m}_t.$$

Cada punto compara (r_{t-1}, r_t) . La nube sigue la diagonal y la correlación es 0.86: quedan desviaciones parecidas en horas consecutivas.

Cautela: cerca de cero aparecen bandas porque muchas horas nocturnas tienen producción cero. Parte de esa estructura aún refleja el ciclo solar y la aproximación periódica.

La gráfica motiva evaluar un AR(1), pero no demuestra por sí sola que este mejorará el pronóstico.

Modelo AR(1) para los residuos

Escribimos

$$y_t = m_t + r_t$$

y modelamos las desviaciones mediante

$$r_t = \phi r_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma^2).$$

Equivalentemente,

$$y_t = m_t + \phi(y_{t-1} - m_{t-1}) + \varepsilon_t.$$

El parámetro ϕ mide la persistencia de las desviaciones respecto al patrón esperado.

Interpretación de la persistencia

En

$$r_t = \phi r_{t-1} + \varepsilon_t,$$

- $\phi \approx 0$: poca persistencia;
- $0 < \phi < 1$: las desviaciones persisten y se atenúan;
- ϕ cercano a 1: persistencia fuerte.

Por ejemplo, si la generación está por debajo del patrón por nubosidad persistente, puede continuar por debajo durante las horas siguientes.

Inferencia Bayesiana con dependencia temporal

Los parámetros son

$$\theta = (\beta, \phi, \sigma).$$

Por ejemplo,

$$\beta_j \sim N(0, 1),$$

$$\phi \sim N(0, 0.5^2) \mathbb{I}(-1 < \phi < 1),$$

$$\sigma \sim \text{HalfNormal}(0.5).$$

La posterior es

$$p(\beta, \phi, \sigma \mid y) \propto p(y \mid \beta, \phi, \sigma) p(\beta) p(\phi) p(\sigma).$$

Esta posterior conjunta no tiene una forma cerrada sencilla.

Cuando no podemos calcular la posterior exactamente

En los primeros modelos obtuvimos directamente

$$p(\mu \mid y) \quad y \quad p(\beta \mid y).$$

Al aumentar la complejidad del modelo, podemos recurrir a métodos computacionales para obtener muestras:

$$\theta^{(1)}, \dots, \theta^{(M)} \sim p(\theta \mid y).$$

Por ejemplo, mediante métodos de Monte Carlo vía cadenas de Markov (MCMC).

La inferencia Bayesiana no cambia; cambia la forma de calcular la posterior.

Predicción mediante muestras posteriores

Para cada muestra

$$\theta^{(m)} = (\beta^{(m)}, \phi^{(m)}, \sigma^{(m)}),$$

generamos

$$y_{T+1}^{(m)} = m_{T+1}^{(m)} + \phi^{(m)} \left(y_T - m_T^{(m)} \right) + \varepsilon_{T+1}^{(m)}.$$

con

$$\varepsilon_{T+1}^{(m)} \sim N \left(0, (\sigma^{(m)})^2 \right).$$

Repitiendo el procedimiento obtenemos muestras de

$$p(y_{T+1} \mid y_{1:T}).$$

Trayectorias predictivas

Para un horizonte h , propagamos el modelo recursivamente:

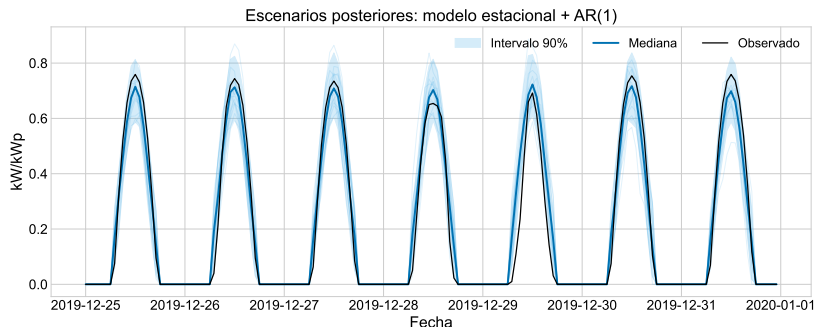
$$y_{T+1}^{(m)} \rightarrow y_{T+2}^{(m)} \rightarrow \cdots \rightarrow y_{T+h}^{(m)}.$$

Cada muestra posterior genera una posible trayectoria:

$$\mathbf{y}_{T+1:T+h}^{(m)} = \left(y_{T+1}^{(m)}, \dots, y_{T+h}^{(m)} \right).$$

La predicción probabilística produce escenarios futuros, no solamente una curva promedio.

Pronóstico con memoria residual



Pronóstico de la última semana de 2019 con patrón estacional y residual AR(1). La línea negra muestra la producción observada, la línea naranja la mediana predictiva y la banda el intervalo predictivo al 90%.

Efecto de la memoria residual

Modelo	MAE	RMSE	Cobertura 90%
Patrón periódico Bayesiano	0.0307	0.0550	93.5%
Patrón periódico + AR(1)	0.0308	0.0552	94.1%

En esta semana de prueba, añadir memoria residual de corto plazo no mejora el error puntual.

La complejidad adicional debe validarse, no suponerse beneficiosa.

Modelos autorregresivos de mayor orden

Un AR(1) utiliza un solo rezago:

$$y_t = c + \phi_1 y_{t-1} + \varepsilon_t.$$

Podemos incorporar varios valores anteriores:

$$y_t = c + \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + \varepsilon_t.$$

Esto define un modelo

$$\text{AR}(p).$$

El parámetro p indica cuántos rezagos de la serie participan en el modelo.

Innovaciones pasadas

Otra fuente de información son las innovaciones, es decir, aquello que el modelo no anticipó.

Si

$$\hat{y}_{t-1}$$

era la predicción y observamos y_{t-1} , definimos

$$\varepsilon_{t-1} = y_{t-1} - \hat{y}_{t-1}.$$

Un modelo MA(1) incorpora esta innovación:

$$y_t = \mu + \varepsilon_t + \theta \varepsilon_{t-1}.$$

AR utiliza valores pasados; MA utiliza innovaciones pasadas.

Modelos ARMA

Combinando valores e innovaciones pasadas obtenemos

$$\text{ARMA}(p, q).$$

Por ejemplo,

$$\text{ARMA}(1, 1) : \quad y_t = c + \phi y_{t-1} + \varepsilon_t + \theta \varepsilon_{t-1}.$$

Aquí

p : número de rezagos de la serie,

q : número de rezagos de las innovaciones.

Modelación de cambios

En algunas series es más conveniente modelar los cambios que los niveles originales.

La primera diferencia es

$$\Delta y_t = y_t - y_{t-1}.$$

Por ejemplo, si

$$y_{t-1} = 0.60, \quad y_t = 0.68,$$

entonces

$$\Delta y_t = 0.08.$$

La diferenciación puede eliminar estructuras persistentes y producir una serie más adecuada para un modelo ARMA.

Modelos ARIMA

Si aplicamos diferenciación y modelamos la serie transformada mediante ARMA, obtenemos

$$\text{ARIMA}(p, d, q).$$

El parámetro d indica el número de diferencias ordinarias.
Por ejemplo,

$$\text{ARIMA}(1, 1, 1)$$

utiliza,

$$z_t = y_t - y_{t-1}$$

y modela

$$z_t = \phi z_{t-1} + \varepsilon_t + \theta \varepsilon_{t-1}.$$

Es decir,

$$y_t - y_{t-1} = \phi(y_{t-1} - y_{t-2}) + \varepsilon_t + \theta \varepsilon_{t-1}.$$

Una escala temporal adicional

En los datos horarios de generación solar existe un ciclo diario:

$$s = 24.$$

Además de la relación entre

$$y_t \text{ y } y_{t-1},$$

puede existir una relación importante entre

$$y_t \text{ y } y_{t-24}.$$

El segundo término corresponde a la misma hora del día anterior.

Diferenciación ordinaria y estacional

La diferencia ordinaria compara periodos consecutivos:

$$\Delta y_t = y_t - y_{t-1}.$$

Su orden se representa mediante d .

La diferencia estacional compara posiciones equivalentes de dos ciclos consecutivos:

$$\Delta_{24} y_t = y_t - y_{t-24}.$$

Su orden se representa mediante D .

d y D representan operaciones diferentes.

Diferencias entre ciclos diarios

Definimos

$$z_t = y_t - y_{t-24}.$$

Por ejemplo,

$$\underbrace{y_t}_{\text{hoy, 12:00}} - \underbrace{y_{t-24}}_{\text{ayer, 12:00}} = \text{cambio entre días.}$$

La transformación elimina parte del patrón que se repite cada 24 horas y concentra el modelo en las diferencias entre ciclos.

Una vez predicho z_t ,

$$\hat{y}_t = y_{t-24} + \hat{z}_t.$$

Modelamos cómo cambia el ciclo de un día al siguiente.

Modelo SARIMA

SARIMA combina componentes ordinarias y estacionales:

$$\text{SARIMA}(p, d, q)(P, D, Q)_s.$$

	No estacional	Estacional
Autorregresión	p	P
Diferenciación	d	D
Media móvil	q	Q

Para nuestros datos,

$$s = 24.$$

Componentes estacionales

Para $s = 24$,

- **P : autorregresión estacional**

incorpora dependencia a escala de ciclos completos;

$$t - 24, \quad t - 48, \dots$$

- **D : diferenciación estacional**

para $D = 1$,

$$z_t = y_t - y_{t-24};$$

- **Q : media móvil estacional**

incorpora innovaciones de ciclos anteriores;

$$\varepsilon_{t-24}, \quad \varepsilon_{t-48}, \dots$$

Patrón periódico y dependencia estacional

Patrón periódico

Los armónicos estiman

$$m_t = \text{comportamiento esperado}$$

según la hora y la época del año. Representan lo que típicamente ocurre en esa posición del calendario.

SARIMA

La componente estacional utiliza la relación dinámica entre ciclos:

$$y_t - y_{t-24}$$

Representa cómo cambia la serie respecto a ciclos recientes.

Los dos modelos representan periodicidad, pero utilizan información temporal diferente.

Modelo utilizado

Para la serie solar utilizamos

$$\text{SARIMA}(1, 0, 1)(1, 1, 1)_{24}.$$

Por tanto,

$$p = 1, \quad d = 0, \quad q = 1,$$

$$P = 1, \quad D = 1, \quad Q = 1, \quad s = 24.$$

En particular,

$$d = 0$$

indica que no utilizamos la diferencia ordinaria $y_t - y_{t-1}$, mientras que

$$D = 1$$

indica que utilizamos

$$z_t = y_t - y_{t-24}.$$

Dinámica de la serie diferenciada

Como

$$z_t = y_t - y_{t-24},$$

la componente no estacional incorpora

$$\underbrace{z_{t-1}}_{\text{un periodo atrás}}, \quad \underbrace{\varepsilon_{t-1}}_{\text{innovación reciente}},$$

mientras que la componente estacional incorpora

$$\underbrace{z_{t-24}}_{\text{un ciclo atrás}}, \quad \underbrace{\varepsilon_{t-24}}_{\text{innovación estacional}}.$$

El modelo combina dependencia de corto plazo y dependencia diaria.

Representación completa del SARIMA

Definimos el operador de rezago B por

$$By_t = y_{t-1}, \quad B^{24}y_t = y_{t-24}.$$

El modelo

$$\text{SARIMA}(1, 0, 1)(1, 1, 1)_{24}$$

puede escribirse como

$$(1 - \phi B)(1 - \Phi B^{24})(1 - B^{24})y_t = (1 + \theta B)(1 + \Theta B^{24})\varepsilon_t.$$

Los factores representan:

- $(1 - \phi B)$: AR de corto plazo,
- $(1 - \Phi B^{24})$: AR estacional,
- $(1 - B^{24})$: diferenciación estacional,
- $(1 + \theta B)$: MA de corto plazo,
- $(1 + \Theta B^{24})$: MA estacional.

Formulación Bayesiana del SARIMA

Los parámetros desconocidos son

$$\vartheta = (\phi, \theta, \Phi, \Theta, \sigma).$$

Asignamos distribuciones prior, por ejemplo,

$$\phi, \Phi \sim N(0, 0.5^2),$$

$$\theta, \Theta \sim N(0, 0.5^2),$$

$$\sigma \sim \text{HalfNormal}(0.5),$$

junto con las restricciones necesarias sobre los parámetros AR y MA.

Posterior del modelo SARIMA

Después de observar

$$\mathbf{y} = (y_1, \dots, y_T),$$

la posterior satisface

$$p(\vartheta \mid \mathbf{y}) \propto p(\mathbf{y} \mid \vartheta)p(\vartheta).$$

Es decir,

$$\begin{aligned} p(\phi, \theta, \Phi, \Theta, \sigma \mid \mathbf{y}) \\ \propto p(\mathbf{y} \mid \phi, \theta, \Phi, \Theta, \sigma)p(\phi)p(\theta)p(\Phi)p(\Theta)p(\sigma). \end{aligned}$$

A diferencia del ejemplo Normal–Normal, la posterior conjunta no tiene, en general, una forma cerrada sencilla.

Predicción Bayesiana con SARIMA

Obtenemos muestras de la posterior mediante MCMC:

$$\vartheta^{(1)}, \dots, \vartheta^{(M)} \sim p(\vartheta \mid \mathbf{y}).$$

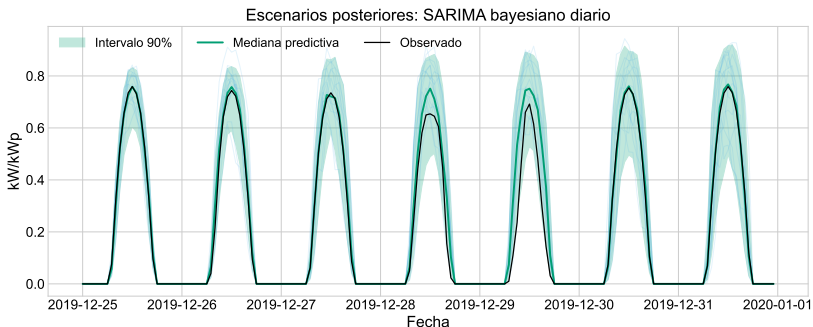
Para cada muestra posterior generamos una trayectoria futura:

$$\mathbf{y}_{T+1:T+h}^{(m)} \sim p(\mathbf{y}_{T+1:T+h} \mid \vartheta^{(m)}, \mathbf{y}).$$

El conjunto de trayectorias aproxima

$$p(\mathbf{y}_{T+1:T+h} \mid \mathbf{y}).$$

Pronóstico SARIMA Bayesiano



Pronóstico de la última semana de 2019 con SARIMA bayesiano de estacionalidad diaria. La línea negra muestra la producción observada, la línea verde la mediana predictiva y la banda el intervalo predictivo al 90%, obtenido a partir de trayectorias posteriores.

Comparación de los pronósticos

Modelo	MAE	RMSE	Cobertura 90%
Semana anterior	0.0349	0.0816	—
Patrón periódico Bayesiano	0.0307	0.0550	93.5%
Patrón periódico + AR(1)	0.0308	0.0552	94.1%
SARIMA Bayesiano	0.0232	0.0421	95.2%

En esta ventana de evaluación, el SARIMA produce el menor error de predicción.

Una diferencia importante es que incorpora explícitamente dependencia y diferenciación a escala diaria.

Construcción y evaluación de modelos

Los ejemplos muestran una progresión:

Patrón periódico $\rightarrow$ memoria local $\rightarrow$ dependencia estacional.

- El patrón diario y anual explica una parte importante de la generación solar.
- La memoria residual de corto plazo es una hipótesis que debe verificarse mediante predicción fuera de muestra.
- En esta ventana, SARIMA produjo mejores pronósticos.
- Un menor error puntual no es el único criterio: también importan la incertidumbre, la calibración y la coherencia de las trayectorias.

De los datos a la decisión

El proceso completo puede resumirse como

Datos $\longrightarrow$ Modelo $\longrightarrow$ Inferencia $\longrightarrow$ Predicción $\longrightarrow$ Decisión.

La inferencia Bayesiana permite mantener la incertidumbre a lo largo de este proceso:

$$p(\theta \mid y) \longrightarrow p(y_{\text{futuro}} \mid y).$$

Predecir no significa eliminar la incertidumbre, sino representarla para tomar mejores decisiones.

Gracias

Preguntas y comentarios

Proyectos, publicaciones y líneas de investigación.

Si te interesa colaborar o participar como estudiante, visita:

`mdazatorres.github.io`



Escanea el código para conocer más.

¡Gracias por su atención!

Te invitamos a conocer la
Maestría en Ciencias Matemáticas
del Instituto de Matemáticas, Unidad Cuernavaca.

También ofrecemos un
curso de preparación para el examen de ingreso.



Escanea el código QR para obtener más información.