

Optimización No-paramétrica de Bandas de Confianza para Generación Eólica al corto plazo

Dr. Ing. Claudio Risso (crisso@fing.edu.uy)

Dr. Gustavo Guerberoff (gguerber@fing.edu.uy)

Seminario ANII-FSE_110454

23 de mayo de 2018

Generación Eólica

La energía eólica representa hoy más del 30% de la potencia nacional instalada, y más del 35% de la energía consumida.

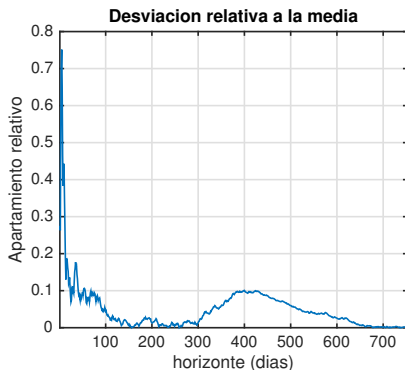
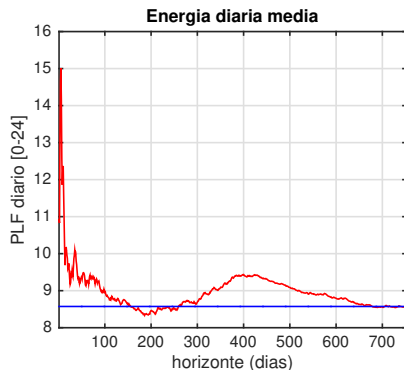
Tipo de Fuente	Cantidad de Unidades	Potencia Instalada (MW)	Energía Total 2017 (GWh)
Biomasa	12	200	900
Eólica	37	1.437	4.400
Solar	17	230	200
Hidroeléctrica	4	1.538	6.200
Ciclo Combinado	1	550	100
Otras Térmicas	4	604	800
Sub-Totales	76	4.559	12.600
Interc. Argen.	2	2.000	-1.200
Interc. Brasil	2	360	0
Totales	82	6.919	11.400

Tabla: [Fuente: Base a programación estacional de ADME].

Como sucede con otras renovables no-convencionales la eólica: no es controlable, es intermitente e incierta, especialmente en horizontes de pocos días. Lo anterior constituye un problema para el despacho al corto plazo.

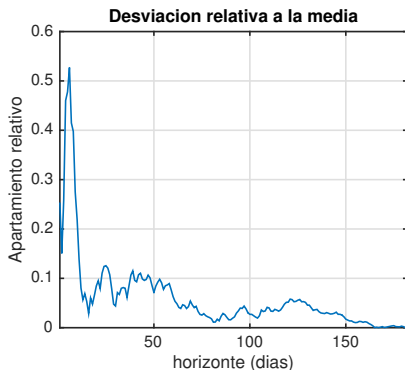
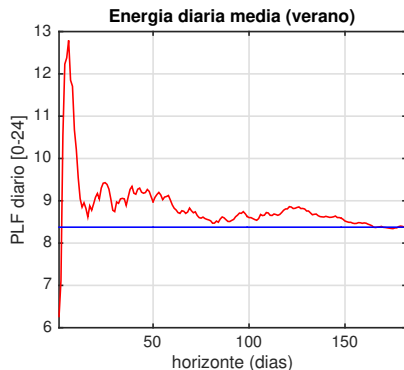
Regularidad Estacional

- ▶ La desviación temporal de la eólica respecto a la media histórica (datos 2014 a 2016) es acotada
- ▶ El rebote intermedio se explica por la estacionalidad



Regularidad Estacional

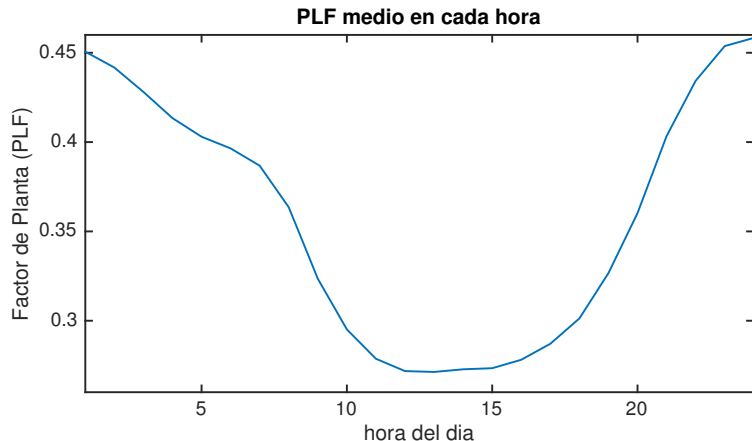
- ▶ La desviación temporal de la eólica respecto a la media histórica (datos 2014 a 2016) es acotada
- ▶ Desaparece al limitarse a una estación



En el último gráfico sólo se han usado días veraniegos.

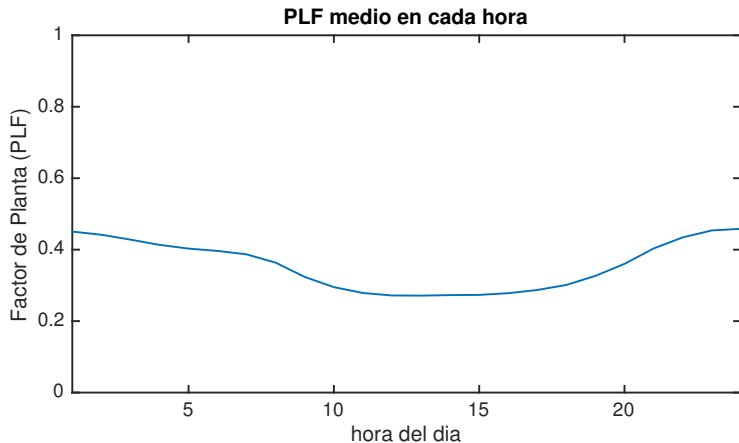
Comportamiento Diario

- ▶ El promedio diario tiene la siguiente característica
- ▶ Se aprecia que es nocturno (contrario a la solar)



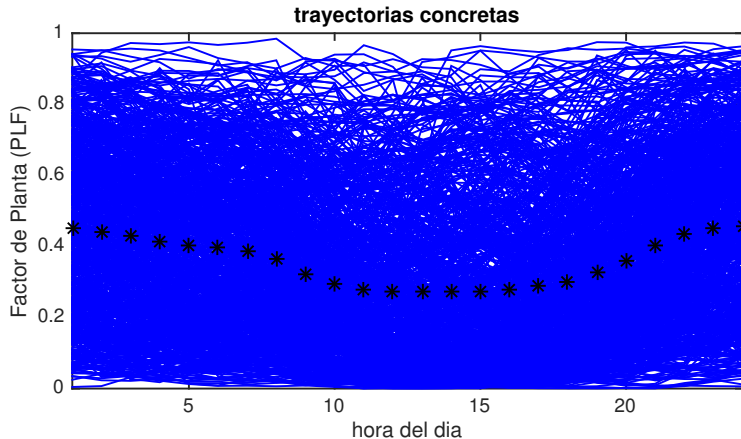
Comportamiento Diario

- ▶ El promedio diario tiene la siguiente característica
- ▶ Y además uniforme a lo largo del día



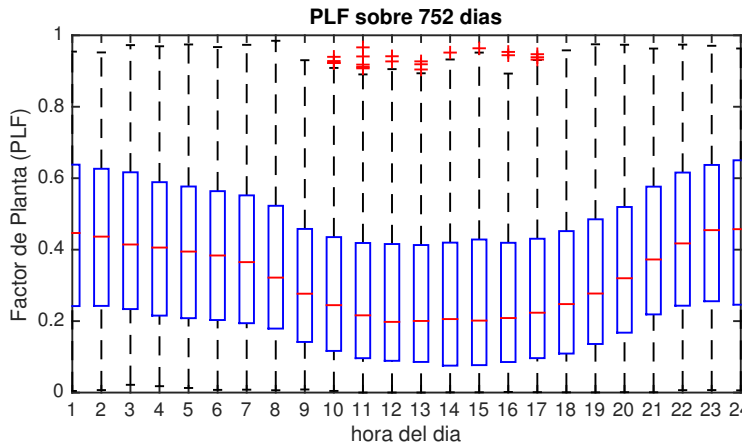
Comportamiento Diario

- ▶ El promedio diario tiene la siguiente característica
- ▶ Pero la verdad es que el proceso es altamente volátil



Comportamiento Diario

- ▶ El promedio diario tiene la siguiente característica
- ▶ Pero la verdad es que el proceso es altamente volátil



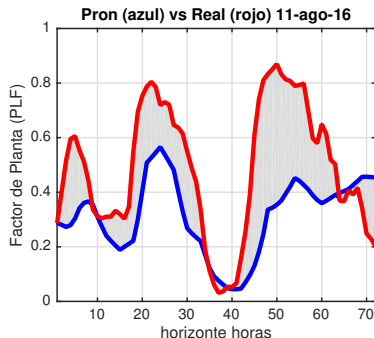
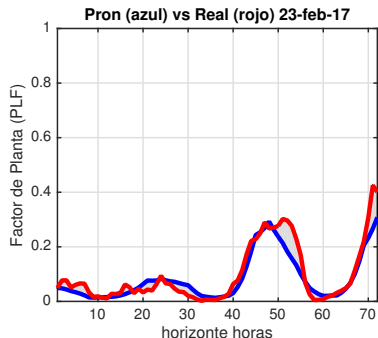
Usar valores medios u otra información puramente estadística usual expone a errores importantes.

Pronósticos de Eólica

- ▶ En horizontes de tiempo de pocos días, los pronósticos de eólica funcionan mejor que la descriptiva estadística clásica
- ▶ Dentro de los pronósticos, destacan los basados en modelos numérico atmosféricos
- ▶ Se han usado los históricos de pronósticos de Garrad Hassan (abr/16-mar/17) y Metológica (ene/16-mar/17)
- ▶ Ambos disponibles en la WEB de UTE
- ▶ También se desarrolló uno propio basado en técnicas estadísticas (PSF), pero mostró pobre performance
- ▶ Los pronósticos de potencia se usan como insumo de un modelo de predespacho estocástico, donde la segunda magnitud en relevancia es la diferencia de energía
- ▶ Así que durante el proyecto se ha usado la norma 1 ($\| \cdot \|_1$) como medida del error

Performance de los Pronósticos

- ▶ El error medio (PLF, 72hs ahead) para GH es $\mu_{err} = 6.80$
- ▶ Mientras que la desviación es $\sigma_{err}^2 = 6.15$

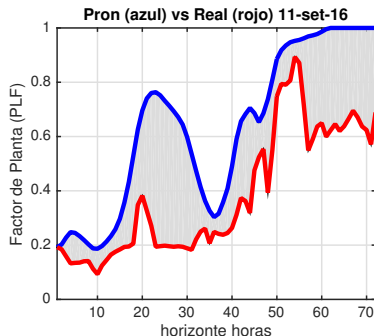
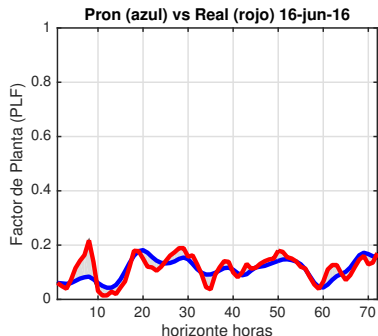


Se muestran además el mejor y el peor pronóstico (error absoluto) entre los datos disponibles.

En el peor día el error es 14.71, un 57% del PLF medio (0.357).

Performance de los Pronósticos

- ▶ El error medio (PLF, 72hs ahead) para MT es $\mu_{err} = 5.99$
- ▶ Mientras que la desviación es $\sigma_{err}^2 = 4.88$

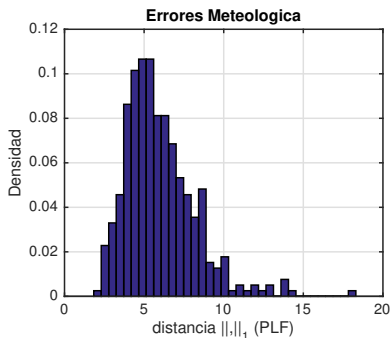
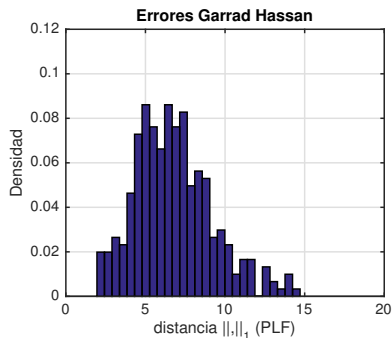


Aunque mejor en media y desviación, Meteorológica registra el peor de los peores casos.

En ese día el error fue 18.31, un 71% del PLF medio.

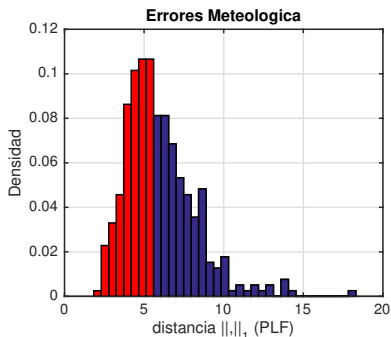
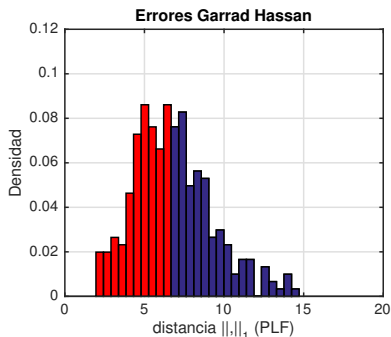
Performance de los Pronósticos

En lo que hace a la distribución de esos errores, se reitera la mayor regularidad de Meteorológica.



Performance de los Pronósticos

En lo que hace a la distribución de esos errores, se reitera la mayor regularidad de Meteorológica.



Las áreas sombreadas en rojo concentran el 50% de la probabilidad en ambos casos. De usar uno de estos pronósticos se debe esperar un apartamiento respecto al valor esperado del factor de la planta instalada de entre 23.3% y 26.5%.

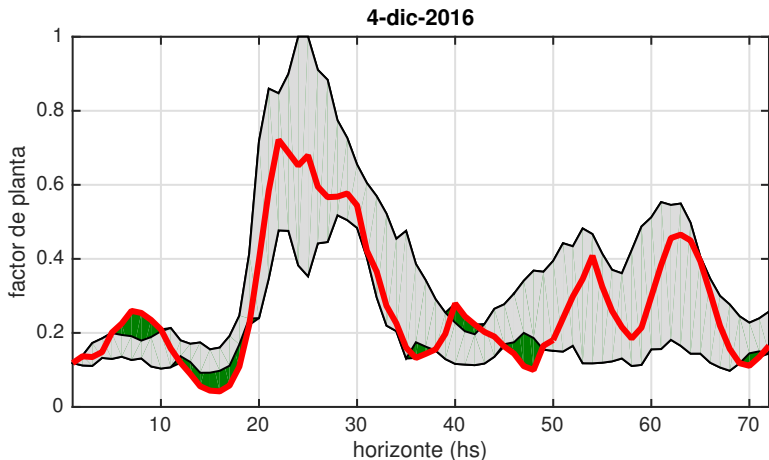
Coexistiendo regularmente con muchos valores bastante peores.

Uso de Bandas de Confianza

- ▶ En el corto plazo la energía eólica es intrínsecamente incierta
- ▶ Hay que tomar como un hecho convivir con esa incertidumbre
- ▶ En este proyecto se apuntó a desarrollar un modelo de optimización estocástica
- ▶ Que planifica el despacho energético al corto plazo sobre un conjunto de trayectorias que concentran la mayor parte de la probabilidad del proceso
- ▶ Y que por tanto es robusto ante desviaciones del proceso real (potencia) respecto a los pronósticos
- ▶ A su vez, se busca integrar formas de compensación (demanda gestionable, etc.), cuyo cumplimiento hace más a la energía en el período que a las potencias puntuales
- ▶ De ahí el uso de la norma 1 como medida del error

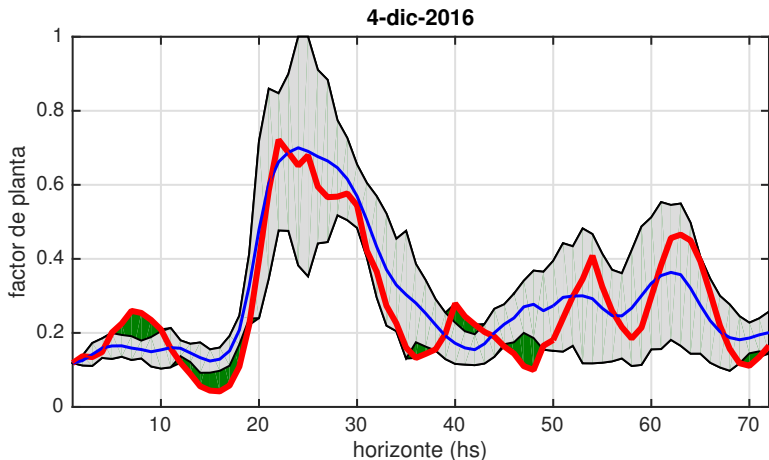
Uso de Bandas de Confianza

En lugar de pronósticos se usan bandas de confianza, calculadas para que la energía off-band (verde) esperada esté acotada.



Uso de Bandas de Confianza

En lugar de pronósticos se usan bandas de confianza, calculadas para que la energía off-band (verde) esperada esté acotada.



Se puede pensar que el centroide de cada banda es el pronóstico de banda simétrica más confiable.

¿Qué se busca?

Se buscan las desviaciones óptimas x_t , para construir la banda $[(1 - x_t)p_t^d, (1 + x_t)p_t^d]_{t=1, \dots, T}$ de menor ancho esperado.

$$\left\{ \begin{array}{l} \min \sum_{t=1}^T \hat{w}_t x_t \\ p_t^d x_t - y_d + z_t^d \geq |w_t^d - p_t^d| - 1, \quad 1 \leq t \leq T, d \in D \\ \sum_{t=1}^T z_t^d \leq T(\theta + 1 - y_d), \quad d \in D \\ \sum_{d \in D} y_d \geq \lambda D, \\ y_d \in \{0, 1\}, 0 \leq x_t, z_t^d \leq 1. \end{array} \right.$$

¿Qué se busca?

Se buscan las desviaciones óptimas x_t , para construir la banda $[(1 - x_t)p_t^d, (1 + x_t)p_t^d]_{t=1, \dots, T}$ de menor ancho esperado.

$$\left\{ \begin{array}{l} \min \sum_{t=1}^T \hat{w}_t x_t \\ p_t^d x_t - y_d + z_t^d \geq |w_t^d - p_t^d| - 1, \quad 1 \leq t \leq T, d \in D \\ \sum_{t=1}^T z_t^d \leq T(\theta + 1 - y_d), \quad d \in D \\ \sum_{d \in D} y_d \geq \lambda D, \\ y_d \in \{0, 1\}, 0 \leq x_t, z_t^d \leq 1. \end{array} \right.$$

Cuya energía off band z_t^d en cada día d no supere un umbral θ .
Sobre una muestra de entrenamiento D .

¿Qué se busca?

Se buscan las desviaciones óptimas x_t , para construir la banda $[(1 - x_t)p_t^d, (1 + x_t)p_t^d]_{t=1, \dots, T}$ de menor ancho esperado.

$$\left\{ \begin{array}{l} \min \sum_{t=1}^T \hat{w}_t x_t \\ p_t^d x_t - y_d + z_t^d \geq |w_t^d - p_t^d| - 1, \quad 1 \leq t \leq T, d \in D \\ \sum_{t=1}^T z_t^d \leq T(\theta + 1 - y_d), \quad d \in D \\ \sum_{d \in D} y_d \geq \lambda D, \\ y_d \in \{0, 1\}, 0 \leq x_t, z_t^d \leq 1. \end{array} \right.$$

Cuya energía off band z_t^d en cada día d no supere un umbral θ .
Sobre una muestra de entrenamiento D , de la que hay que usar al menos una fracción λ de días.

¿Qué se busca?

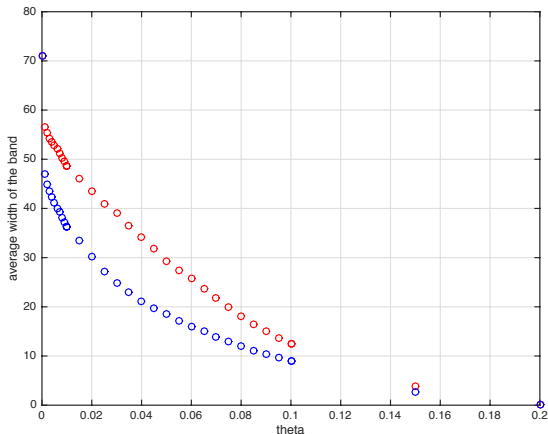
Se buscan las desviaciones óptimas x_t , para construir la banda $[(1 - x_t)p_t^d, (1 + x_t)p_t^d]_{t=1, \dots, T}$ de menor ancho esperado.

$$\left\{ \begin{array}{l} \min \sum_{t=1}^T \hat{w}_t x_t \\ p_t^d x_t - y_d + z_t^d \geq |w_t^d - p_t^d| - 1, \quad 1 \leq t \leq T, d \in D \\ \sum_{t=1}^T z_t^d \leq T(\theta + 1 - y_d), \quad d \in D \\ \sum_{d \in D} y_d \geq \lambda D, \\ y_d \in \{0, 1\}, 0 \leq x_t, z_t^d \leq 1. \end{array} \right.$$

Lo último implementa una segmentación automática de *outliers* cuando $\lambda < 1$. Si $\lambda = 1$ entonces se debe usar toda la muestra y el problema es de optimización continua.

El ancho de banda medio como métrica de calidad

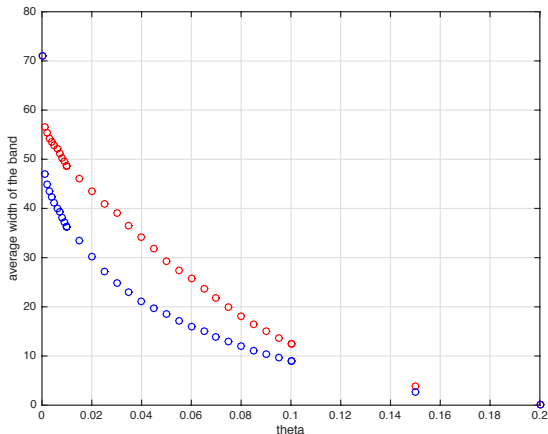
Podemos pensar que el ancho de banda óptimo para un conjunto de pronósticos es una medida de su calidad.



Al entrenar sobre una muestra al azar de 123 en común (se cuenta con 319 días comunes), y usando $\lambda = 1$ se obtuvo lo anterior.

El ancho de banda medio como métrica de calidad

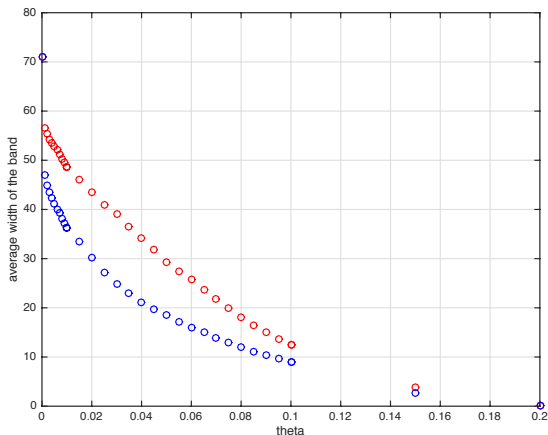
Podemos pensar que el ancho de banda óptimo para un conjunto de pronósticos es una medida de su calidad.



Se destaca que: i) Para tener error 0 hay que cubrir toda el área posible ($PLF_{72hs} = 71$).

El ancho de banda medio como métrica de calidad

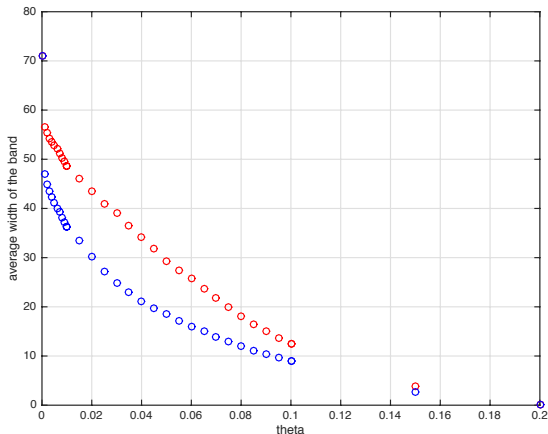
Podemos pensar que el ancho de banda óptimo para un conjunto de pronósticos es una medida de su calidad.



Se destaca que: ii) No se necesita banda si se está dispuesto a aceptar hasta un 20% de error en el PLF a 72 horas.

El ancho de banda medio como métrica de calidad

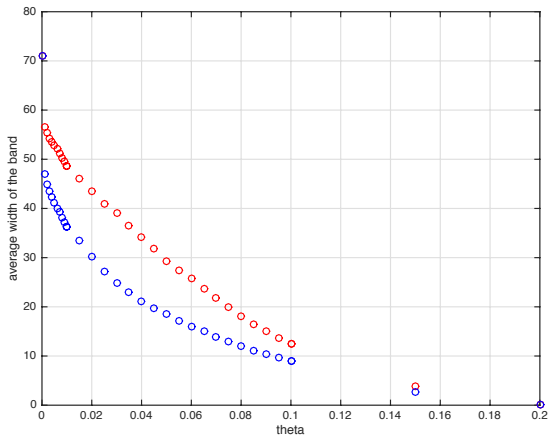
Podemos pensar que el ancho de banda óptimo para un conjunto de pronósticos es una medida de su calidad.



Se destaca que: iii) En general (para errores intermedios) el pronóstico de Meteorológica (azul) requiere bandas más angostas para lograr el mismo error que Garrad Hassan (rojo).

El ancho de banda medio como métrica de calidad

Podemos pensar que el ancho de banda óptimo para un conjunto de pronósticos es una medida de su calidad.



Se destaca que: iv) Pero al final ninguno alcanza los niveles de calidad a los que aspirábamos para realizar un pre-despacho eficiente y confiable.

Análisis de los *atypicals*

- ▶ Al bajar λ el problema se vuelve combinatorio y por tanto intensivo en términos de cómputo
- ▶ Como se usa el subconjunto de entrenamiento de mejor calidad, el ancho esperado de las bandas se reduce
- ▶ Se omiten los *malos pronósticos* de la muestra de entrenamiento, por lo que se construye una banda optimista
- ▶ Y al evaluar la calidad de la banda sobre la muestra de testing, el resultado tiende a alejarse del target θ
- ▶ Clasificamos como *atypical days* a los días de la muestra de testing donde la energía off-band supera al target
- ▶ Y los modelamos como una variable aleatoria en $\{0, 1\}$, donde marcamos con 1 a los *atypicals*

Análisis de los *atypicals*

La siguiente tabla resume la *performance* sobre la muestra de testing de las bandas asociadas a los distintos pronósticos para distintos λ 's. En todos los casos se usó $\theta=3.5\%$ del PLF.

λ	%atypical	%BW	t(sec)	%atypical	%BW	t(sec)	%atypical	%BW	t(sec)
1.00	6.64%	29.6%	< 1	5.61%	47.6%	< 1	5.44%	86.5%	< 1
0.95	14.8%	22.2%	117	12.2%	37.2%	129	7.97%	78.6%	77
0.90	20.4%	19.2%	138	16.3%	34.4%	428	11.2%	71.7%	66
0.85	27.0%	17.4%	265	22.5%	30.7%	272	13.4%	67.2%	292
0.80	32.1%	16.3%	664	25.5%	27.4%	1518	16.3%	63.3%	421
0.75	36.7%	15.0%	1296	37.2%	24.2%	1097	20.7%	58.1%	657
0.70	45.4%	13.7%	2020	45.4%	21.7%	1950	28.3%	53.5%	842
0.65	49.0%	12.3%	1225	49.0%	18.2%	1480	28.6%	49.9%	1468
0.60	54.6%	11.5%	1390	52.6%	16.5%	1560	37.0%	47.5%	2425
Meteorológica			Garrad Hassan			PSF44			

Tabla: Resultados de las distintas bandas sobre la muestra de testing.

Se buscaba mantener la energía media off-band bajo el 10% de la generación eólica media, esto es: $\theta \leq 0.1 \cdot 0.357 \approx 0.035$ (3.5%).

Análisis de los *atypicals*

La siguiente tabla resume la *performance* sobre la muestra de testing de las bandas asociadas a los distintos pronósticos para distintos λ 's. En todos los casos se usó $\theta=3.5\%$ del PLF.

λ	%atypical	%BW	t(sec)	%atypical	%BW	t(sec)	%atypical	%BW	t(sec)
1.00	6.64%	29.6%	< 1	5.61%	47.6%	< 1	5.44%	86.5%	< 1
0.95	14.8%	22.2%	117	12.2%	37.2%	129	7.97%	78.6%	77
0.90	20.4%	19.2%	138	16.3%	34.4%	428	11.2%	71.7%	66
0.85	27.0%	17.4%	265	22.5%	30.7%	272	13.4%	67.2%	292
0.80	32.1%	16.3%	664	25.5%	27.4%	1518	16.3%	63.3%	421
0.75	36.7%	15.0%	1296	37.2%	24.2%	1097	20.7%	58.1%	657
0.70	45.4%	13.7%	2020	45.4%	21.7%	1950	28.3%	53.5%	842
0.65	49.0%	12.3%	1225	49.0%	18.2%	1480	28.6%	49.9%	1468
0.60	54.6%	11.5%	1390	52.6%	16.5%	1560	37.0%	47.5%	2425
Meteorológica			Garrad Hassan			PSF44			

Tabla: Resultados de las distintas bandas sobre la muestra de testing.

También se buscaba que las violaciones a esa meta (*atypicals*) no superaran el 10%.

Análisis de los *atypicals*

La siguiente tabla resume la *performance* sobre la muestra de testing de las bandas asociadas a los distintos pronósticos para distintos λ 's. En todos los casos se usó $\theta=3.5\%$ del PLF.

λ	%atypical	%BW	t(sec)	%atypical	%BW	t(sec)	%atypical	%BW	t(sec)
1.00	6.64%	29.6%	< 1	5.61%	47.6%	< 1	5.44%	86.5%	< 1
0.95	14.8%	22.2%	117	12.2%	37.2%	129	7.97%	78.6%	77
0.90	20.4%	19.2%	138	16.3%	34.4%	428	11.2%	71.7%	66
0.85	27.0%	17.4%	265	22.5%	30.7%	272	13.4%	67.2%	292
0.80	32.1%	16.3%	664	25.5%	27.4%	1518	16.3%	63.3%	421
0.75	36.7%	15.0%	1296	37.2%	24.2%	1097	20.7%	58.1%	657
0.70	45.4%	13.7%	2020	45.4%	21.7%	1950	28.3%	53.5%	842
0.65	49.0%	12.3%	1225	49.0%	18.2%	1480	28.6%	49.9%	1468
0.60	54.6%	11.5%	1390	52.6%	16.5%	1560	37.0%	47.5%	2425
Meteorológica			Garrad Hassan			PSF44			

Tabla: Resultados de las distintas bandas sobre la muestra de testing.

Finalmente, se requería una desviación máxima de potencia dentro de la banda que no superara el 10% de la potencia media total generada por el sistema, con lo que $\%BW \leq 20\%$

Análisis de los *atypicals*

La siguiente tabla resume la *performance* sobre la muestra de testing de las bandas asociadas a los distintos pronósticos para distintos λ 's. En todos los casos se usó $\theta=3.5\%$ del PLF.

λ	%atypical	%BW	t(sec)	%atypical	%BW	t(sec)	%atypical	%BW	t(sec)
1.00	6.64%	29.6%	< 1	5.61%	47.6%	< 1	5.44%	86.5%	< 1
0.95	14.8%	22.2%	117	12.2%	37.2%	129	7.97%	78.6%	77
0.90	20.4%	19.2%	138	16.3%	34.4%	428	11.2%	71.7%	66
0.85	27.0%	17.4%	265	22.5%	30.7%	272	13.4%	67.2%	292
0.80	32.1%	16.3%	664	25.5%	27.4%	1518	16.3%	63.3%	421
0.75	36.7%	15.0%	1296	37.2%	24.2%	1097	20.7%	58.1%	657
0.70	45.4%	13.7%	2020	45.4%	21.7%	1950	28.3%	53.5%	842
0.65	49.0%	12.3%	1225	49.0%	18.2%	1480	28.6%	49.9%	1468
0.60	54.6%	11.5%	1390	52.6%	16.5%	1560	37.0%	47.5%	2425
Meteorológica			Garrad Hassan			PSF44			

Tabla: Resultados de las distintas bandas sobre la muestra de testing.

¡¡Ninguno cumplía con los objetivos simultáneamente!!

Independencia de los *atypicals*

Los días atípicos entre pronósticos tienen una baja correlación cruzada, especialmente para λ 's altos, donde son más raros.

	MT	GH	PS
MT	1.0000	0.2621	-0.0177
GH	0.2621	1.0000	-0.0649
PS	-0.0177	-0.0649	1.0000

	MT	GH	PS
MT	1.0000	0.2913	-0.0710
GH	0.2913	1.0000	0.0241
PS	-0.0710	0.0241	1.0000

Tabla: Correlaciones entre atypicals [$\lambda = 0.6$ (izq) y $\lambda = 1$ (der)].

En particular, el pronóstico PSF44 es el único para el que aparecen correlaciones negativas.

Lo que suponemos responde a su independencia metodológica, o en las fuentes de información usadas.

Independencia de los *atypicals*

Los días atípicos entre pronósticos tienen una baja correlación cruzada, especialmente para λ 's altos, donde son más raros.

	MT	GH	PS
MT	1.0000	0.2621	-0.0177
GH	0.2621	1.0000	-0.0649
PS	-0.0177	-0.0649	1.0000

	MT	GH	PS
MT	1.0000	0.2913	-0.0710
GH	0.2913	1.0000	0.0241
PS	-0.0710	0.0241	1.0000

Tabla: Correlaciones entre *atypicals* [$\lambda = 0.6$ (izq) y $\lambda = 1$ (der)].

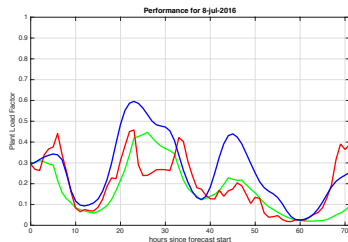
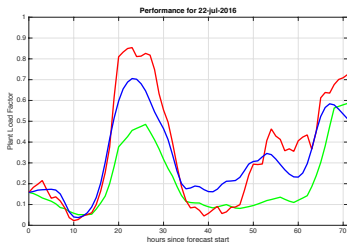
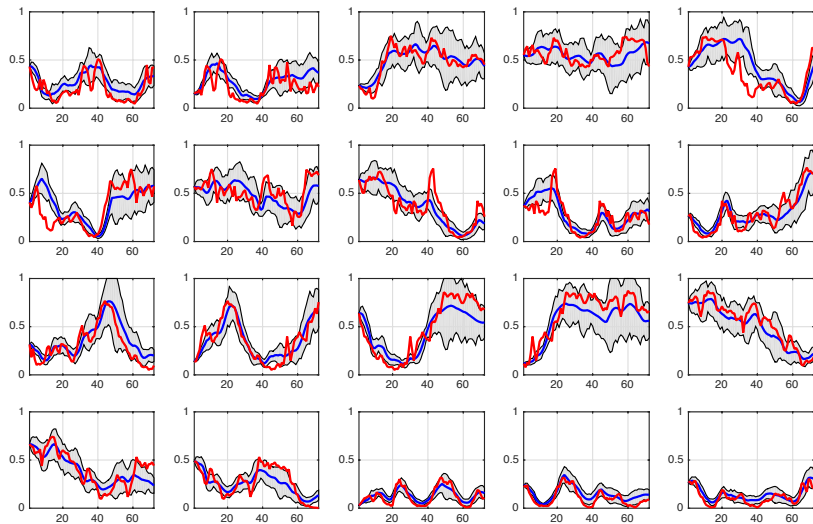


Figura: Curvas en días atípicos [rojo=potencia, azul=metlog, verde=ghassan].

Combinación óptima de bandas

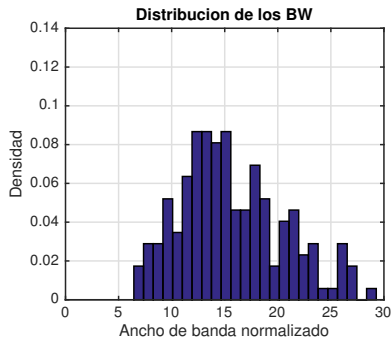
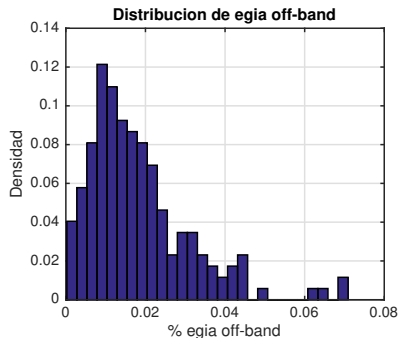
- ▶ Los promedios de bandas heredan los anchos de sus progenitores (no se gana con esto)
- ▶ Bajando el λ se consiguen bandas más angostas, pero se aumenta la fracción de atypicals
- ▶ En contrapartida, se esperaba que la combinación de bandas compensara los errores (redujera los atypicals)
- ▶ Se exploraron combinaciones entonces convexas del tipo:
$$bda_{cmp}(t) = \alpha bda_{mt}(t) + \beta bda_{gh}(t) + (1 - \alpha - \beta) bda_{ps}(t),$$
 usando distintas combinaciones de λ 's
- ▶ La combinación que mostró la mejor performance fue:
 $\alpha=0.66$, $\beta=0.25$, $\gamma=0.09$, para los λ 's: 0.85, 0.70 y 0.65
- ▶ Consiguiendo una banda de ancho medio 21.7% (cerca del 20%), y un 9.8% de atypicals para $\theta=0.035$

Días al azar en la muestra de testing



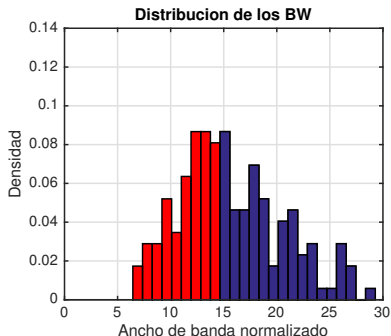
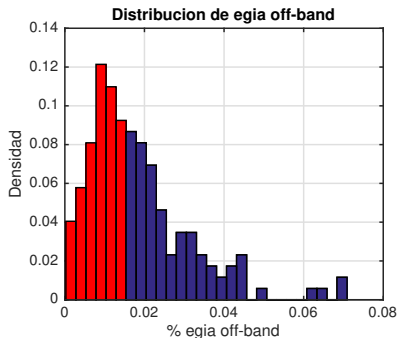
Distribución de Errores

En lo que hace a la distribución de energía off-band y anchos de banda, los resultados son estos:



Distribución de Errores

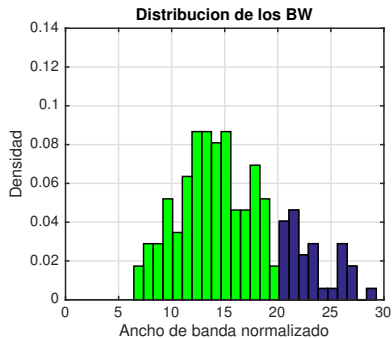
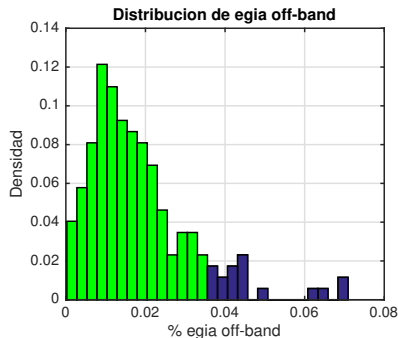
En lo que hace a la distribución de energía off-band y anchos de banda, los resultados son estos:



Las áreas sombreadas en rojo concentran el 50% de la probabilidad en ambos casos.

Distribución de Errores

En lo que hace a la distribución de energía off-band y anchos de banda, los resultados son estos:



Y las sombreadas en verde cumplen con las metas de confiabilidad a las que se apuntó (3.5% energía off-band y $\% \overline{BW} \leq 20\%$).

Conclusiones y Trabajo Futuro

- ▶ Utilizar la técnica para la Energía Solar
- ▶ Usar un modelo de bandas asimétricas
- ▶ Calibrar bandas estacionales (necesitamos más datos)
- ▶ Usar clusters para cubrir diferencias de magnitud a una misma hora (también se necesitan más datos)
- ▶ Mejoras el pronóstico estadístico (PSF44) ¡¡Se precisan al menos 10 veces más de datos!!
- ▶ Analizar propiedades teóricas de la técnica (complejidad, convergencia ante distribuciones conocidas)
- ▶ Diseñar heurísticas para poder aplicarla sobre grandes volúmenes de datos



¿Preguntas?