



Parametrización de espectros RVS mediante técnicas de IA

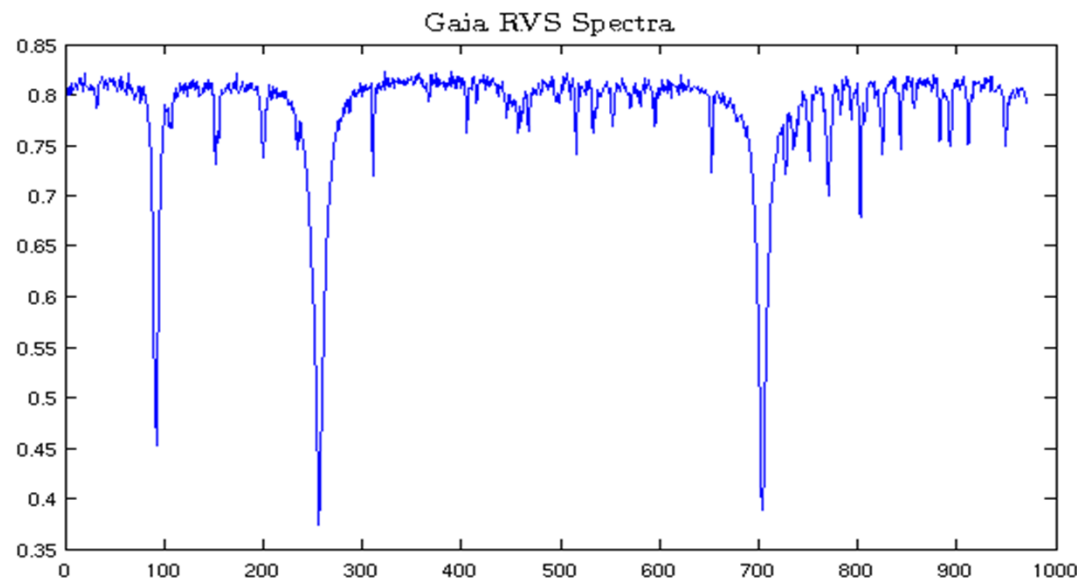
**D.Ordóñez, D.Fustes , C.Dafonte, M.Manteiga
y B.Arcay**

Introducción

- GGG (Galician Group for Gaia): Grupo constituyente del CU8 en el DPAC. Involucrado en tareas de clasificación y parametrización mediante técnicas de Inteligencia Artificial
- Trabajo con simulaciones del dispositivo RVS:
 - Estimación de parámetros:
 - Temperaturas efectivas
 - Gravedades superficiales
 - Abundancia de hierro con respecto al hidrógeno (Fe/H)
 - Abundancias de elementos ligeros frente al hierro (α/Fe)

Simulaciones del instrumento RVS

- RVS: 8470-8740 Å. Estrellas FKG
- Librería compilada por A. Recio, P. de Laverny et al. en 2006 y completada por C.Allende
- Alta resolución (HR): $6 < V < 11$. Espectros con 971 píxeles (longitudes de onda). $R=11500$
- Baja Resolución (LR): $11 < V < 17$. Espectros con 315 píxeles. $R=5500$
- Diversos niveles SNR: 5, 10, 50, 200, ..., ∞ $R=$



Configuración del entrenamiento

- Arquitectura
 - Red *FeedForward* con 3 capas
 - $\text{Max}(n/2, 200)$ neuronas en la capa oculta
- *Online training*. Adaptación de los pesos patrón a patrón
- Adaptación del *learning rate* durante las iteraciones
- 5000 iteraciones de entrenamiento. Utilización de *early stopping* para buscar el punto de generalización (dilema bias/varianza)
- Una red por parámetro

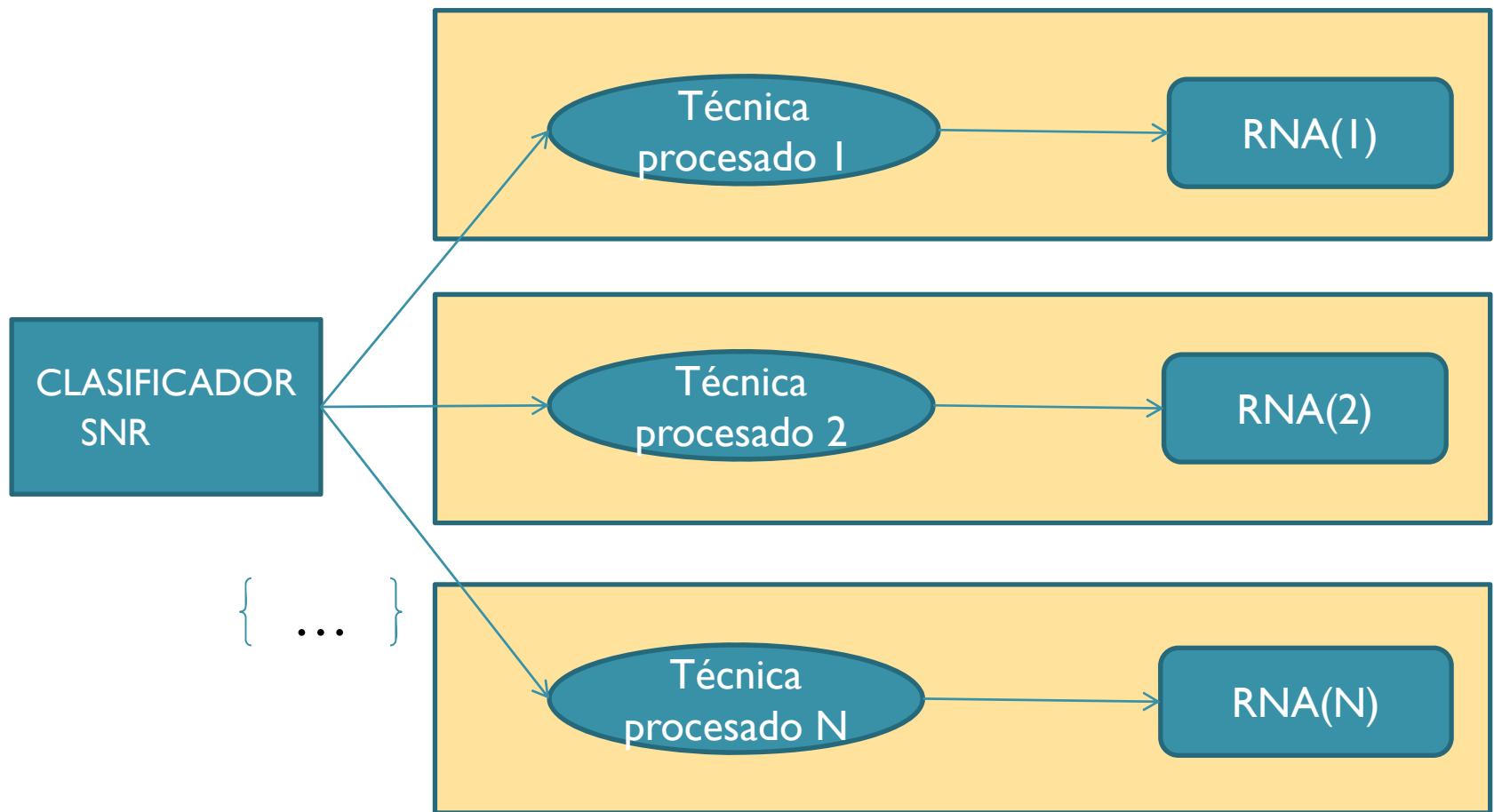
Efecto del ruido en el entrenamiento

- Comparativa de modelos: Se calcula el error medio absoluto (de todo el conjunto de test)
- Mejores resultados aplicando el mismo nivel de ruido tanto al conjunto de entrenamiento como al de test

Teff

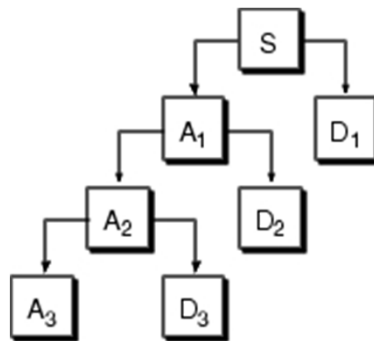
SNR Train/Test	10	25	50	75	100	500	∞
10	366	323	344	353	359	374	376
25	456	225	188	191	196	209	211
50	579	247	155	147	147	160	163
75	566	264	161	139	130	132	134
100	662	308	178	142	131	132	135
500	494	274	175	146	128	100	98
∞	515	288	184	147	128	90	86

Estrategia de selección de entradas/dominio



Dominios de entrada

- FFT: Transformada Rápida de Fourier
 - Se aplica una ventana Hamming
- DWT: Transformada Wavelet Discreta.
 - Análisis multinivel.
 - Wavelet madre: Daubechies 5.



$$\begin{aligned} S &= A_1 + D_1 \\ &= A_2 + D_2 + D_1 \\ &= A_3 + D_3 + D_2 + D_1 \end{aligned}$$

Resultados RVS con alta resolución

SNR= ∞	Teff	Log g	[Fe/H]	[α /Fe]
FFT	36	0,07	0,06	0,03
Wavelength	72	0,15	0,12	0,07
A1	81	0,16	0,13	0,07
A2	85	0,18	0,15	0,08
A3	93	0,21	0,17	0,08
A4	108	0,22	0,21	0,09
A5	115	0,26	0,24	0,11
D1	166	0,28	0,19	0,08
D2	159	0,34	0,19	0,07
D3	139	0,27	0,17	0,07
D4	105	0,23	0,16	0,07
D5	118	0,24	0,23	0,09

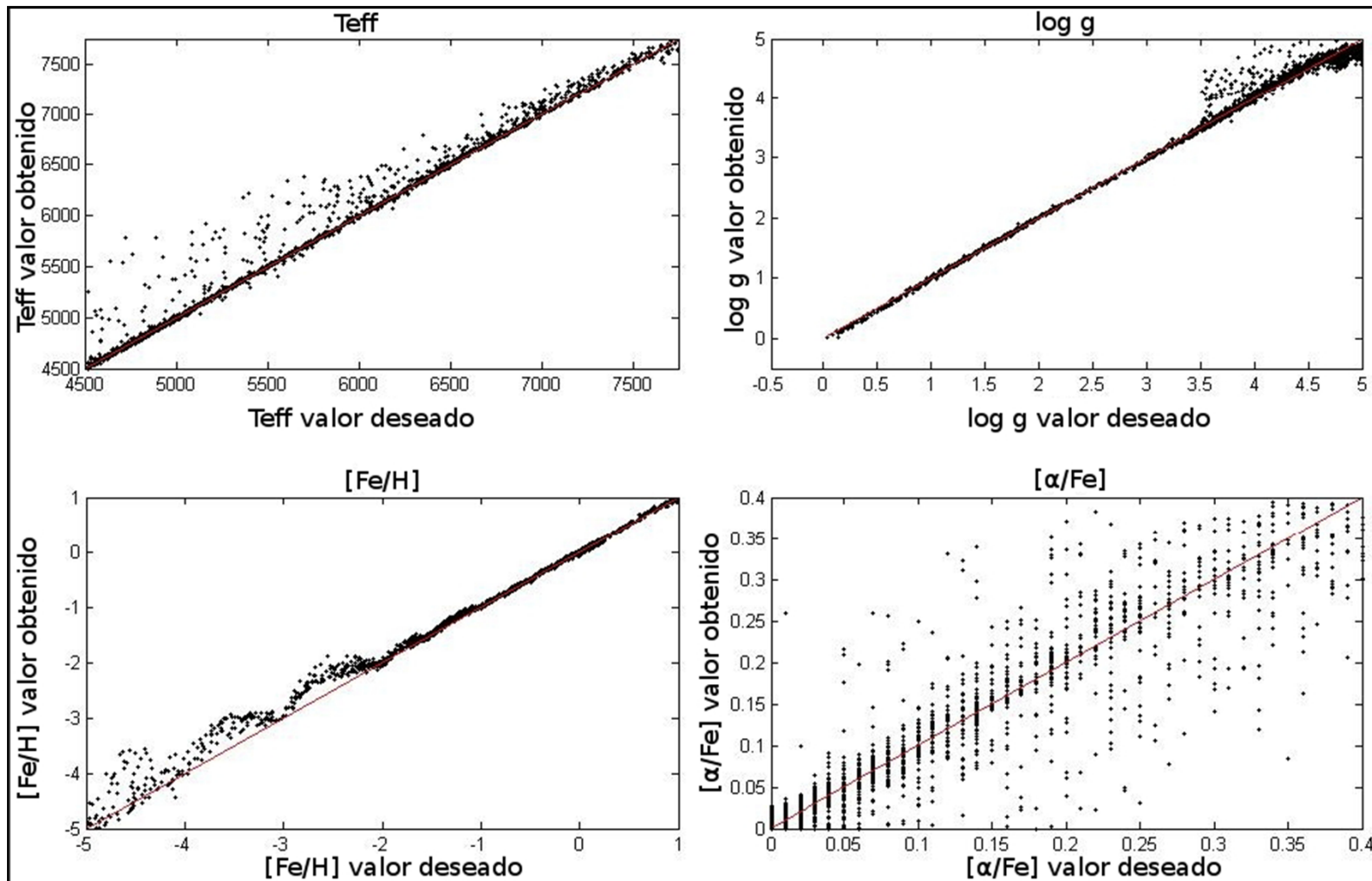
Resultados RVS con alta resolución

SNR=200	Teff	Log g	[Fe/H]	[α /Fe]
FFT	175	0,28	0,19	0,09
Wavelength	80	0,16	0,12	0,07
A1	106	0,17	0,16	0,08
A2	91	0,19	0,15	0,08
A3	99	0,24	0,17	0,09
A4	111	0,22	0,21	0,09
A5	130	0,27	0,25	0,11
D1	260	0,56	0,27	0,12
D2	208	0,40	0,20	0,09
D3	188	0,34	0,19	0,09
D4	148	0,29	0,18	0,09
D5	145	0,29	0,24	0,10

Resultados RVS con alta resolución

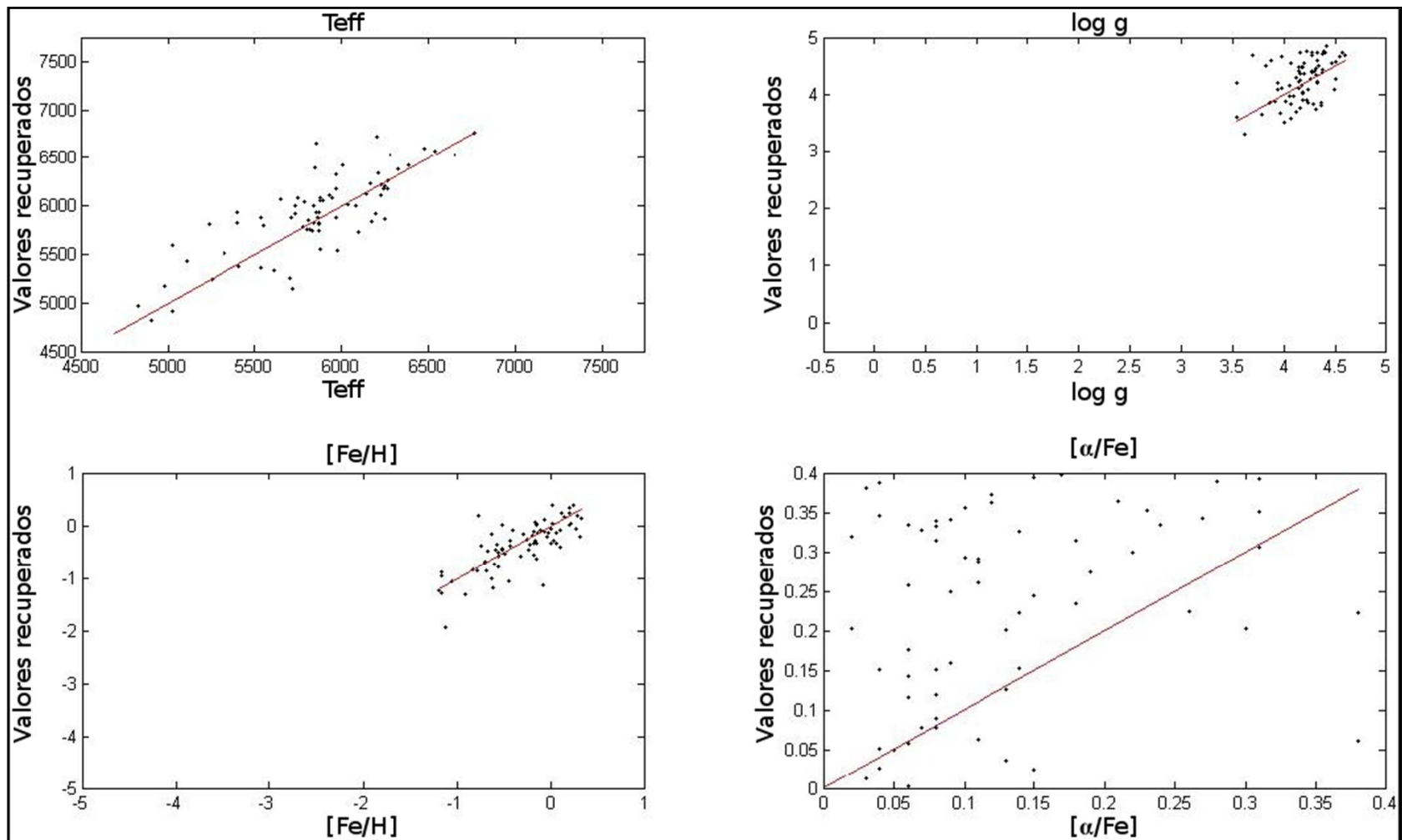
SNR=10	Teff	Log g	[Fe/H]	[α /Fe]
FFT	739	1,31	1,07	0,19
Wavelength	297	0,59	0,37	0,17
A1	289	0,57	0,36	0,16
A2	272	0,55	0,35	0,16
A3	268	0,57	0,35	0,17
A4	284	0,61	0,41	0,17
A5	328	0,68	0,44	0,18
D1	835	1,62	1,25	0,23
D2	613	1,26	0,71	0,18
D3	554	1,08	0,62	0,18
D4	497	0,91	0,58	0,18
D5	437	0,76	0,57	0,18

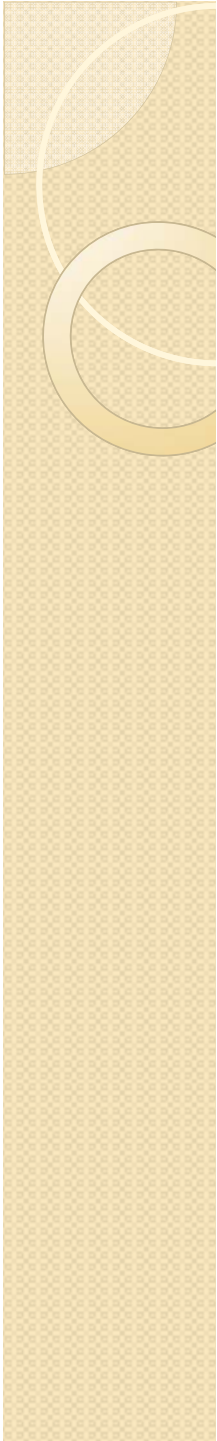
Resultados RVS con alta resolución



Resultados con espectros observacionales

- Librería Cenarro et al.





Combinación de componentes Wavelet

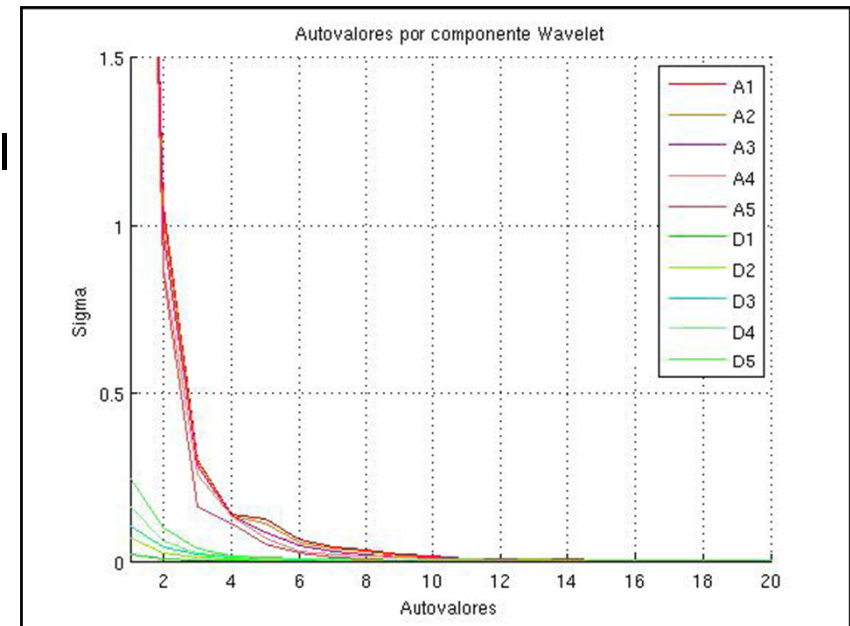
- Las componentes individuales wavelet de conllevan pérdida de información (bien sea ruido o señal)
- La combinación de varias de ellas puede mejorar los resultados con respecto a componentes simples
- Métodos de selección de componentes:
 - Análisis de componentes principales (PCA)
 - Algoritmos genéticos (GA)

Combinación de componentes Wavelet

SNR=10	Teff	Log g	[Fe/H]	[α /Fe]
A1A2	278	0,55	0,35	0,16
A1A3	277	0,55	0,35	0,16
A1A4	279	0,55	0,35	0,16
A1D1	370	0,74	0,50	0,19
A2A3	266	0,55	0,35	0,16
A2A4	267	0,56	0,35	0,16
A2A5	264	0,56	0,35	0,16
A2D5	264	0,55	0,35	0,16
A3A4	271	0,55	0,37	0,16
A3A5	273	0,55	0,35	0,16
A3D3	278	0,59	0,36	0,17
A3D4	279	0,58	0,36	0,17

Combinación de componentes Wavelet (PCA)

- Calculamos los autovalores para cada aproximación A_i y detalle D_i
- Seleccionamos el número de componentes principales necesario para explicar un % de variabilidad adecuado (bien sea el codo de la curva o el 99% de variabilidad)
- Número de componentes:
 - Codo: 50
 - 99% : 100



Combinación de componentes Wavelet (PCA)

SNR	CODO	99%
∞	61	39
200	107	101
10	865	895

Teff

SNR	CODO	99%
∞	0.12	0.07
200	0.19	0.19
10	1,55	1,87

Logg

SNR	CODO	99%
∞	0.09	0.07
200	0.12	0.12
10	0.75	1,26

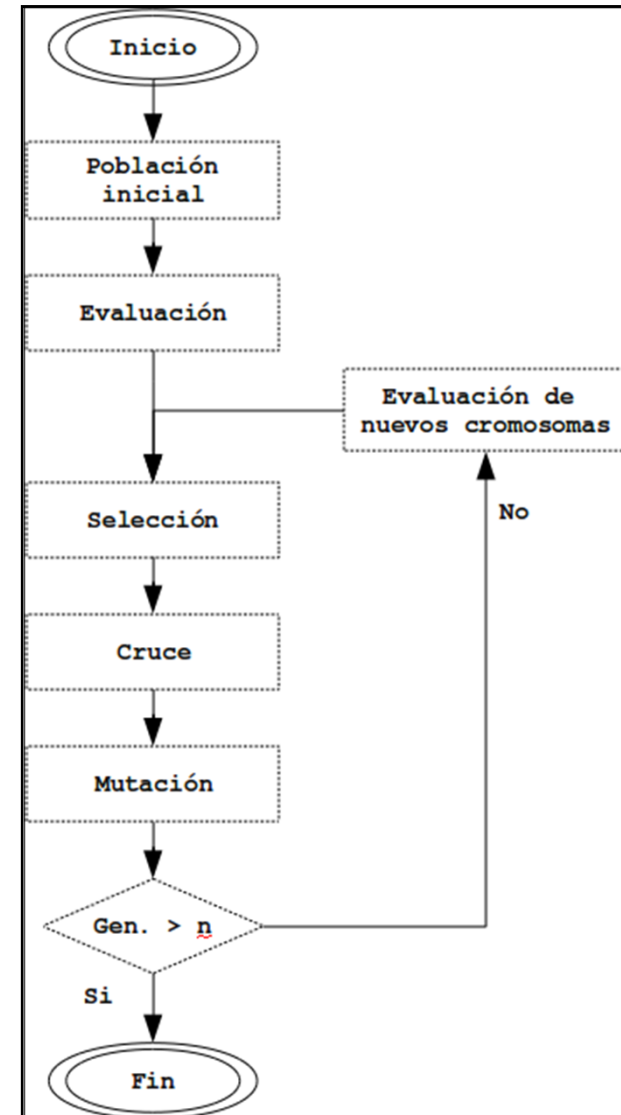
Fe/H

SNR	CODO	99%
∞	0.05	0,03
200	0.07	0.07
10	0.24	0.25

α /Fe

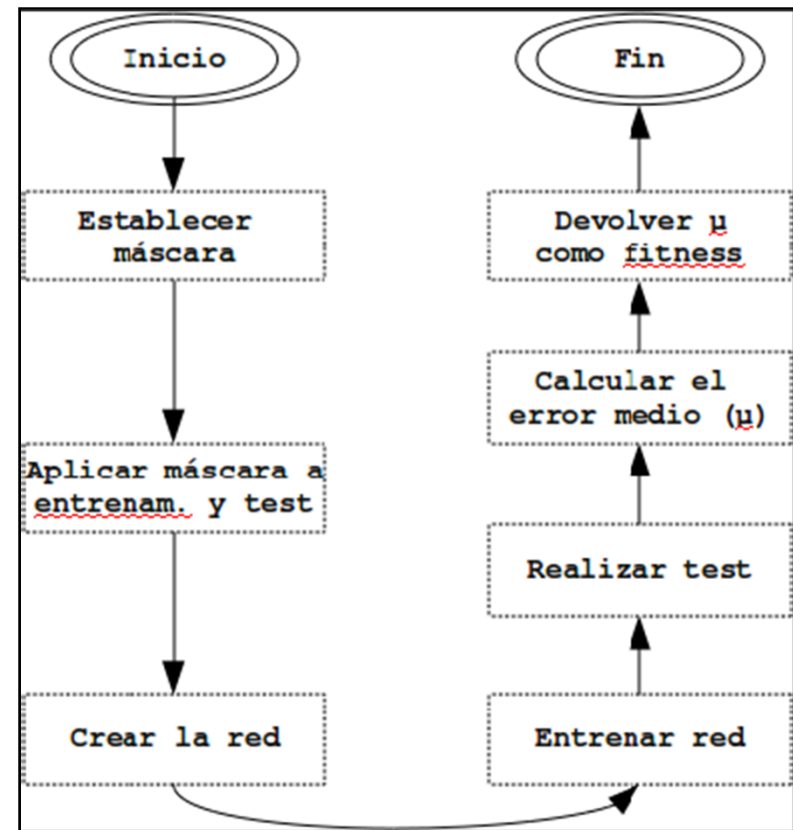
Combinación de componentes Wavelet (GA)

- Algoritmos Genéticos: Potentes técnicas de optimización de parámetros. Basados en la naturaleza.
- Selección de componentes en RVS: Se crea una máscara de 0s y 1s. Los 1s representan los píxeles seleccionados.



Combinación de componentes Wavelet (GA)

- **Selección:** Algoritmo de la ruleta. Mejor que torneo
- **Cruce:** Las tiras de 1s y 0s de los distintos cromosomas se mezclan
- **Mutación:** Un 5% de 1s son susceptibles de cambiar a 0 (y viceversa)
- **Evaluación:** Mediante la RNA (ver diagrama). Tan sólo entrenamiento preliminar (100 iteraciones), debido a problemas computacionales



Combinación de componentes Wavelet (GA)

- **Ventajas:**

- Selección especializada por cada parámetro
- Dependiente del problema, no de la entrada. Selección no lineal de entradas (al contrario que PCA)
- Rápida ejecución en producción, una vez encontrada la máscara

- **Inconvenientes**

- Enorme cantidad de cómputo necesaria para calcular la máscara
 - Implementación en supercomputador (OpenMp, MPI).
 - Implementación en Grid (Hadoop). Colaboración con el CESGA
- No contempla combinaciones lineales/no lineales de componentes (aunque podría plantearse)

- **Desarrollo futuro:** Testar otras SNR, aumentar iteraciones de la red, aumentar el tamaño de la población ...

SNR=200	Teff	Log g	[Fe/H]	[α /Fe]
Wavelet (A3D3D2D1)	106	0.16	0.12	0.06
Wavelength	111	0.17	0.11	0.06

Conclusiones

- Las RNA son un método viable para hacer regresión no lineal de espectros RVS con el objetivo de estimar sus parámetros
- La representación en otros dominios de los espectros puede mejorar la estimación de parámetros bajo distintas condiciones de ruido
- Las técnicas de selección de entradas(PCA) pueden reducir la complejidad de los modelos, mejorando su generalización con alta SNR
- Los algoritmos genéticos permiten la selección de regiones de interés a la hora de estimar cada parámetro, aunque requieren de computación de altas prestaciones

SNR	Dominio	Teff	Log g	[Fe/H]	[α /Fe]
∞	FFT	36	0,07	0,06	0,03
200	Wavelength	72	0,15	0,12	0,07
10	A2D5	264	0,55	0,35	0,16



¡¡¡GRACIAS POR SU ATENCIÓN!!!

- ¿Alguna pregunta/comentario?

- Contacto: dfustes@udc.es