

Estudio estadístico de datos aerobiológicos de Catalunya.*

Trabajo de fin de grado.

Grado de Estadística Aplicada UAB.

Alumno: Fco Javier Mateos Garcia

Tutora: Rosario Delgado **

2 de Septiembre 2013

*Quiero agradecer a la Dra. Jordina Belmonte, y a la Dra. Concepción De Linares de la Unitat de Botànica i del ICTA (Institut de Ciència i Tecnologia Ambientals, UAB), la oportunidad de tener acceso a sus datos y poder analizarlos. Sus consejos y las cuestiones que plantearon han sido la guía y el principal motivo de este trabajo de fin de grado.

**Quiero agradecer a la Dra. Rosario Delgado su imprescindible labor como tutora de este trabajo.

Índice

1. Introducción	5
2. Análisis del registro anual del polen (API).	8
2.1. Análisis descriptivo.	8
2.1.1. Normalidad de las series.	8
2.1.2. Análisis descriptivo de 3 taxones.	9
2.1.3. Distribución por estaciones.	12
2.2. Correlaciones.	18
2.2.1. Mismo taxón entre estaciones.	18
2.2.2. Entre taxones en una misma estación.	23
2.3. Tendencias significativas.	27
2.4. Comparativa de estaciones.	32
2.4.1. Introducción.	32
2.4.2. Vector de porcentajes anuales de taxones.	34
2.4.3. Vector con serie histórica de porcentajes por taxón.	37
3. Análisis del calendario polínico.	40
3.1. Introducción.	40
3.2. Análisis descriptivo.	40
3.2.1. Gráficos.	41
3.2.2. Medianas, rango y MAD. Test de Friedman.	43
3.3. Correlaciones.	48
3.4. Tendencias.	51
3.5. Comparativa de estaciones.	53
4. Conclusiones.	57
Anexos	60

Índice de figuras

1.	Boxplot API taxón <i>Pinus</i> por estaciones.	9
2.	Boxplot API taxón <i>Poaceae</i> por estaciones.	10
3.	Boxplot API taxón <i>Urticaceae</i> por estaciones.	11
4.	Distribución taxones en Barcelona (BCN).	13
5.	Distribución taxones en Bellaterra (BTU).	13
6.	Distribución taxones en Girona (GIC).	14
7.	Distribución taxones en Lleida (LLE).	14
8.	Distribución taxones en Manresa (MAN).	15
9.	Distribución taxones en Tarragona (TAU).	15
10.	Distribución taxones en Tortosa (TOR).	16
11.	Distribución taxones en Vielha (TOR).	16
12.	Distribución taxones en Catalunya (datos agregados).	17
13.	Series del taxón <i>Pinus</i> segun estación	19
14.	Series del taxón <i>Poaceae</i> segun estación	20
15.	Series del taxón <i>Urticaceae</i> según estación	21
16.	Series del taxón <i>Cupressaceae</i> según estación	22
17.	Series del taxón <i>Chenopodiaceae/amaranthaceae</i> según estación	23
18.	Serie API taxón <i>Platanus</i> en MAN (1996:2012).	30
19.	Serie API taxón <i>Platanus</i> en BCN (1994:2012).	30
20.	Serie API taxón <i>Chenop.</i> en BTU (1994:2012).	31
21.	Serie API taxón <i>Ulmus</i> en MAN (1996:2012).	31
22.	Calendario polínico del taxón <i>Pinus</i> en BTU 2012	41
23.	Calendario polínico del taxón <i>Urticaceae</i> en BTU 2012	42
24.	Calendario polínico del taxón <i>Poaceae</i> en BTU 2012	42
25.	Días inicio de floración del taxón <i>Olea</i> según estación 1996:2012.	49
26.	Días inicio de floración del taxón <i>Quercusperenne</i> según estación 1996:2012.	50

1. Introducción

La aerobiología es una rama de la biología que estudia las partículas orgánicas que son transportadas por el aire. El polen es una de esas partículas. Actualmente su estudio, que corresponde a la disciplina de la botánica denominada palinología, se ha incrementado debido, entre otras causas, a su gran incidencia en la salud, ya que provoca ciertas alergias, y a su valor como indicador de cambios de vegetación en el pasado -paleoecología-, y de posibles cambios en el futuro.

En el seno de la UAB, el Instituto de Ciencia y Tecnología Ambientales (ICTA) incorpora un grupo de investigación en palinología (PIA)¹, que gestiona la Red Aerobiológica de Catalunya (XAC)². Esta red dispone de ocho estaciones de recogida de muestras repartidas por el área geográfica catalana (véase Cuadro 2). Cada una de estas estaciones aerobiológicas dispone de un captador esporo-polínico de tipo Hirst, que funciona continuamente y que proporciona una muestra diaria. Cuando las muestras llegan al laboratorio de palinología de la UAB, son analizadas al microscopio óptico con el objetivo de contar y determinar los pólenes y las esporas de hongos que contienen. Los datos se publican y se divultan con la finalidad de informar del estado actual y de la predicción de futuro de los niveles de pólenes alergógenos, que son los que más interesan a la población en general (video explicativo de la Televisión Pública Catalana , TV3: “Quant i quin polen hi ha?”, visualizable a la web de la UAB³).

Entre los estudios realizados por el grupo de investigación del PIA, coordinado por la Doctora Jordina Belmonte, destacamos los recientes trabajos realizados por los investigadores Fernandez-Llamazares⁴ et al y De Linares⁵ et al, en los que se analiza el cambio de polinización en diversas regiones de Catalunya y España, y su posible relación con el cambio climático. En estos trabajos los análisis estadísticos son parte relevante de su investigación y la base para las conclusiones que se obtienen.

La motivación de este trabajo de fin de grado de Estadística Aplicada es la de intentar ampliar estos estudios allí donde no incidieron, buscando una descripción más explicativa de los datos, sus posibles relaciones, las tendencias y cualquier otro aspecto que podamos detectar bajo un enfoque puramente estadístico. Nuestros métodos serán tanto descriptivos como inferenciales, algunos paramétricos y en general no paramétricos, debido a la falta de normalidad de las series de datos analizadas. Los diferentes tests de normalidad, valorados en el siguiente capítulo, así lo refrendan.

En el análisis inferencial hemos intentando utilizar los métodos estadísticos que consideramos idóneos. Algunos de ellos ya han sido utilizados en otros estudios referentes al polen, pero creemos que otros se usan por primera vez en este contexto, principalmente en el análisis de uno de los objetivos, que es evaluar si es posible prescindir de alguna estación, en un futuro, manteniendo el nivel de información, es decir, si existen estaciones que pudieran estar dando una información redundante.

El conjunto de las propiedades de un polen es característico para cada planta y hace

posible identificar con más o menos precisión de qué taxón procede el polen. Es necesario el uso de la palabra taxón (que designa cualquier unidad de determinación dentro de un sistema jerárquico de categorías) porque no siempre puede identificarse de qué especie procede el polen; en bastantes casos la precisión llega sólo al nivel de género (es decir, a un grupo de especies), familia (es decir, a un grupo de géneros), o incluso a un grupo de familias o categorías superiores. Los datos de que disponemos son los correspondientes al recuento de los taxones más comunes y claramente diferenciables, procedentes de varias especies vegetales muy presentes en la geografía catalana. El Cuadro 1 muestra los nombres de esos taxones, el nombre común y tipo de la especie y el porcentaje del taxón en los registros acumulados de toda Catalunya. Estos recuentos vienen diferenciados para las 8 estaciones catalanas. No todas disponen del mismo número de datos, sin embargo, y en el Cuadro 2 podemos ver el periodo de años con datos disponibles para cada estación.

Principalmente trabajaremos con el API -annual pollen index- de cada taxón y para cada estación. El API es la suma anual de las estimaciones de las medias diarias de cada taxón, calculadas tras el análisis de las muestras del captador esporo-polínico, con una función dependiente del número de partículas contadas al microscopio, de la superficie de muestra analizada y del volumen de aire aspirado por el captador. Estas medias diarias valoran la cantidad de granos de polen por metro cúbico de aire de cada taxón para cada día. El API es la suma anual de esos valores diarios. También dispondremos de estos datos diarios con el fin de poder analizar el calendario polínico y sus cambios según año, taxón y estación, en el capítulo 3.

Para todos los análisis hemos usado el programa estadístico de código libre R⁶. Usaremos diferentes niveles de significación. Los más habituales, 0.05 y 0.01, para los análisis de tendencias y correlaciones, mientras que también usaremos el nivel de 0.1 para ciertos tests.

	Taxón	Nombre.común.especie	Tipo.de.planta	Porcentaje
1	Alnus	Aliso / Vern	Árbol	0.78
2	Betula	Abedul / Bedoll	Árbol	0.78
3	Castanea	Castaño/ Castanyer	Árbol	0.64
4	Cupressaceae	Ciprés/ Xiprer	Árbol	23.77
5	Fagus	Haya / Faig	Árbol	0.14
6	Fraxinus	Fresno / Freixe	Árbol	2.26
7	Olea	Olivo / Olivera	Árbol	5.21
8	Pinus	Pino / Pi	Árbol	15.38
9	Platanus	Plátano / Plataner	Árbol	13.68
10	Quercus Caducifolis	Roble / Roure	Árbol	4.64
11	Quercus Perennifolis	Encina / Alzina	Árbol	12.17
12	Ulmus	Olmo / Om	Árbol	0.46
13	Corylus	Avellano / Avellaner	Arbusto	1.89
14	Pistacia	Lentisco / Llentisc	Arbusto	0.32
15	Ambrosia	Ambrosia / Ambrosia	Herbácea	0.01
16	Artemesia	Artemisia / Donzell	Herbácea	0.52
17	Chenopodiaceae Amaranthaceae	Ceñigos / Blet	Herbácea	2.77
18	Plantago	Llantén / Plantatge	Herbácea	2.84
19	PoaceaeTotal	Gramíneas / Gramíñies	Herbácea	4.61
20	PoligonàciesTotals	Ruibarbo / Ruibarbre	Herbácea	0.25
21	Urticàcies	Ortiga /Ortiga	Herbácea	6.88

Cuadro 1: Taxones

	Código	Estación	Años	Alt_mts.	T_Media_C.	Clima_meteo.cat.
1	BCN	Barcelona	1994-2012	93	16.4	Mediterráneo-litoral central
2	BTU	Bellaterra	1994-2012	245	15.2	Mediterráneo-prelitoral central
3	GIC	Girona	1996-2012	98	15.0	Mediterráneo-prelitoral norte
4	LLE	Lleida	1996-2012	202	15.1	Mediterráneo-continental seco
5	MAN	Manresa	1996-2012	291	13.6	Mediterráneo-continental subhúmedo
6	TAU	Tarragona	1996-2012	44	15.8	Mediterráneo-litoral sud
7	VIE	Vielha	2004-2012	974	9.9	Oceánico
8	TOR	Tortosa	2006-2012	14	16.8	Mediterráneo-litoral sud

Cuadro 2: Estaciones

2. Análisis del registro anual del polen (API).

2.1. Análisis descriptivo.

En el Cuadro 1 vemos que los 21 taxones referenciados tienen una incidencia muy diferente en el registro general. Hay taxones con mucha presencia y otros que apenas la tienen, lo cual no indica que no sean importantes de analizar, ya sea por su valor en el paisaje como por su incidencia en las alergias. Además, esta presencia varía entre estaciones, es decir, hay taxones que no están presentes en ciertas estaciones, pero que en otras son los más comunes. También hemos de mencionar el caso específico del taxón *Quercus*. En el registro disponemos de dos taxones diferenciados, uno para el árbol perenne -encina- y otro para el árbol caduco -roble. Por interés del grupo de investigación del PIA, también se analizará, en algún apartado, la suma de estos dos taxones en un registro común denominado *QuercusTotal*.

La principal característica de las series analizadas es su gran variabilidad. Existe algún año con registros muy elevados de API, seguido por otro con registros muy bajos. Múltiples factores influyen en una mayor o menor polinización. El clima es, tal vez, el factor más influyente, pero no el único. Las lluvias antes de la floración suelen aumentar los registros de polen⁷, mientras que si se producen durante la polinización, disminuye la presencia del polen en el aire.

2.1.1. Normalidad de las series.

Para poder analizar las tendencias y correlaciones de los diferentes valores del API de los taxones, creemos necesario estudiar primero si los datos tienen una distribución normal. Tenemos 22 taxones(21+*QuercusTotal*) en 8 estaciones. Serán 176 series temporales. En BCN y BTU tienen 19 datos, mientras que en TOR apenas 7 -ver Cuadro 2. Un análisis gráfico de muchas de las series indica una distribución no normal para la gran mayoría de ellas. Decidimos realizar 4 diferentes test de normalidad implementados en R a las 176 series. En el Cuadro 3 vemos los porcentajes de los tests que superan los pvalores de 0.05 y 0.1, es decir, que no rechazamos la distribución normal de los datos. Hemos decidido incluir estos dos niveles de significación debido al pequeño número de datos en cada serie y por tratarse de un estudio global de normalidad de las series. Las pocas observaciones en la estación de Tortosa -7 datos-, impiden poder realizar algunos de esos tests. Los resultados nos indican que sobre una tercera parte de las series se rechaza la hipótesis nula de normalidad de los datos. En consecuencia optamos por utilizar, principalmente, métodos no paramétricos sin descartar el uso de algun método paramétrico en algún análisis concreto.

Test	pvalor>0.5	pvalor>0.1
Shapiro-Wilk	76 %	66 %
Lilliefors	62 %	53 %
Anderson-Darling	69 %	61 %
Cramer-vonMises	72 %	65 %

Cuadro 3: Porcentaje de tests que NO rechazan norm.

2.1.2. Análisis descriptivo de 3 taxones.

Como ejemplo graficamos los boxplots, por estaciones, de los datos de 3 de los taxones más comunes en todas ellas. Adjuntamos los pvalores de los tests de normalidad para esas series graficadas, y los estadísticos más importantes para comprobar la gran variabilidad de los datos, su distribución asímetrica, los problemas de curtosis y la diferencia de homogeneidad entre estaciones con un coeficiente de variación muy diferente entre ellas. Todo ello para constatar la necesidad de realizar tests no parámetricos para confirmar relaciones y tendencias.

Pinus. El taxón Pinus destaca por su gran presencia en todas las estaciones. Su estudio interesa por su valor en el paisaje silvestre y por ser un buen indicador del cambio climático. Es el taxón que presenta mayor normalidad en las series analizadas. No obstante vemos que existen problemas de asimetría y curtosis, presencia de outliers y poca homogeneidad en el coeficiente de variación.

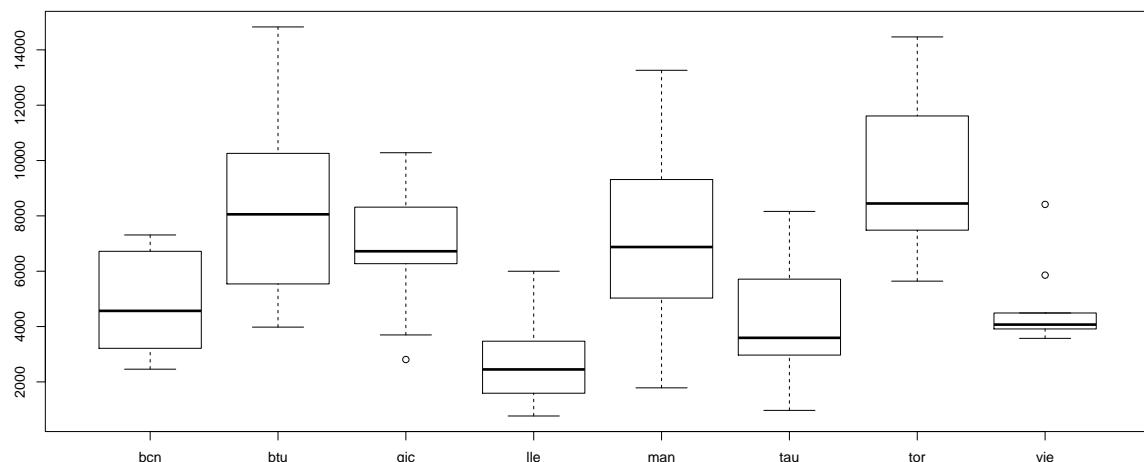


Figura 1: Boxplot API taxón Pinus por estaciones.

	test	bcn	btu	gic	lle	man	tau	tor	vie
1	Shapiro-Wilk	0.03*	0.46	0.66	0.25	0.99	0.43	0.16	0.001**
2	Lilliefors	0.25	0.66	0.70	0.41	0.90	0.20	0.15	0.01*
3	Anderson-Darling	0.04*	0.60	0.51	0.30	0.99	0.33		0.001**
4	Cramer-vonMises	0.09	0.79	0.44	0.30	0.99	0.29		0.001**

Cuadro 4: Pvalores test de normal. Pinus*estación

	estación	media	mínimo	máximo	desv_standard	coefvar	asimetría	curtosis
1	bcn	4849	2460	7310	1815	37.43	0.10	1.32
2	btu	8188	3980	14829	3012	36.78	0.23	2.17
3	gic	6982	2808	10282	1961	28.09	-0.32	2.48
4	lle	2813	768	5999	1551	55.16	0.59	2.19
5	man	7041	1785	13262	3187	45.26	0.11	1.96
6	tau	4243	971	8161	2017	47.55	0.44	2.12
7	tor	4686	3574	8415	1558	33.25	1.46	3.75
8	vie	9535	5641	14468	3430	35.98	0.48	1.35

Cuadro 5: Estadísticos taxón Pinus*estación

Poaceae. El taxón PoaceaeTotal representa los pólenes producidos por algunas gramíneas. En este caso su importancia no es sólo paisajística, ya que es uno de los pólenes que produce más alergias. En este taxón aumentan las series con tests de normalidad en los que se rechaza la hipótesis nula. También destaca la presencia de outliers, los problemas de asimetría y los valores altos de curtosis.

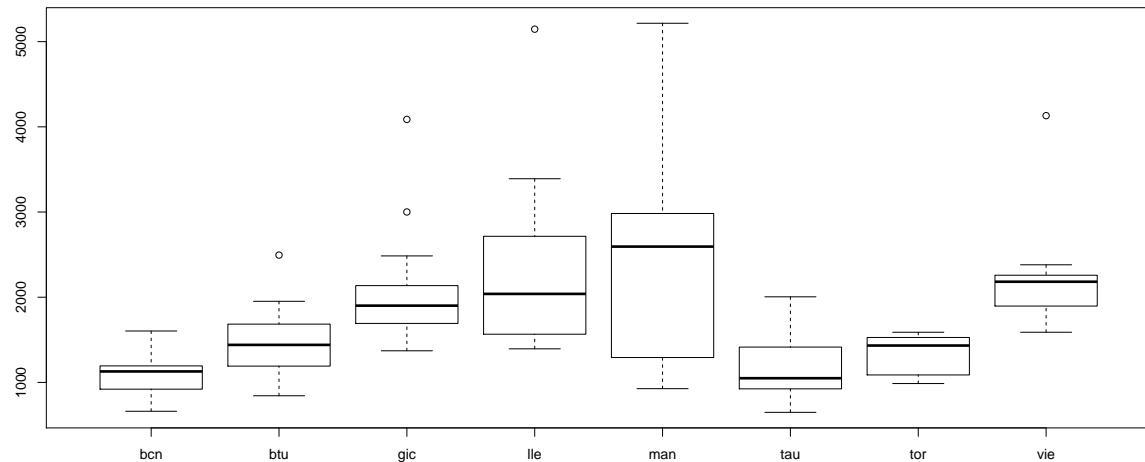


Figura 2: Boxplot API taxón Poaceae por estaciones.

	test	bcn	btu	gic	lle	man	tau	tor	vie
1	Shapiro-Wilk	0.64	0.55	0.001**	0.001**	0.21	0.27	0.13	0.001**
2	Lilliefors	0.29	0.73	0.01*	0.07	0.68	0.20	0.25	0.001**
3	Anderson-Darling	0.51	0.71	0.001**	0.01*	0.31	0.31		0.01*
4	Cramer-vonMises	0.45	0.80	0.001**	0.02*	0.44	0.28		0.01*

Cuadro 6: Pvalores test de normal. Poaceae*estación

	estación	media	mínimo	máximo	desv_standard	coefvar	asimetría	curtosis
1	bcn	1107	662	1604	253	22.89	0.26	2.26
2	btu	1462	844	2495	405	27.72	0.63	3.04
3	gic	2054	1371	4087	654	31.84	1.80	5.92
4	lle	2305	1394	5146	950	41.23	1.54	5.02
5	man	2415	927	5216	1169	48.41	0.51	2.64
6	tau	1203	649	2005	423	35.15	0.39	1.76
7	tor	2282	1589	4131	746	32.70	1.53	4.34
8	vie	1320	987	1589	258	19.54	-0.24	0.97

Cuadro 7: Estadísticos taxón Poaceae*estación

Urticaceae. Por último, el taxón Urticaceae es importante por ser una planta ruderal, que suele aparecer en hábitats alterados por el hombre -escombreras, márgenes de caminos, etc.-, y también por provocar alergias. En sus series se repiten los mismos problemas ya mencionados anteriormente, falta de normalidad en la mayoría de ellas, presencia de outliers y problemas de asimetría y curtosis.

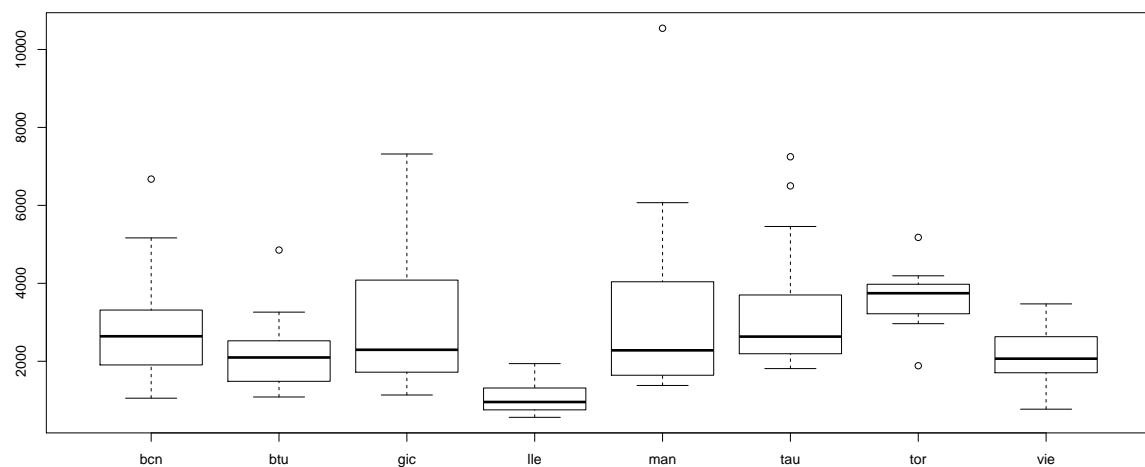


Figura 3: Boxplot API taxón Urticaceae por estaciones.

	test	bcn	btu	gic	lle	man	tau	tor	vie
1	Shapiro-Wilk	0.06	0.01*	0.01*	0.20	0.001**	0.001**	0.91	0.93
2	Lilliefors	0.09	0.4	0.01*	0.59	0.01*	0.09	0.82	0.68
3	Anderson-Darling	0.11	0.04*	0.001**	0.27	0.001**	0.001**		0.78
4	Cramer-vonMises	0.12	0.12	0.001**	0.31	0.001**	0.01*		0.77

Cuadro 8: Pvalores test de normal. Urticaceae*estación

	estación	media	mínimo	máximo	desv_standard	coefvar	asimetría	curtosis
1	bcn	2834	1053	6674	1405	49.56	1.09	3.74
2	btu	2134	1084	4854	883	41.37	1.44	5.16
3	gic	3084	1135	7318	1962	63.61	0.96	2.42
4	lle	1037	561	1940	373	36.00	0.69	2.61
5	man	3201	1379	10544	2329	72.77	1.84	6.06
6	tau	3308	1812	7247	1651	49.90	1.15	3.07
7	tor	2144	771	3472	779	36.33	-0.03	2.10
8	vie	3600	1886	5179	1020	28.35	-0.15	1.99

Cuadro 9: Estadísticos taxón Urticaceae*estación

2.1.3. Distribución por estaciones.

En los siguientes gráficos de Pareto podemos ver la distribución de los taxones en las diferentes estaciones. Únicamente se grafican los taxones cuyos porcentajes en la estación superan el 1% del registro acumulado. Los demás taxones se agrupan en una sola categoría (resto).

Cada estación tiene su particularidad, aunque es evidente que las de mayor proximidad tendrán una distribución de taxones similar. En todas las estaciones -excepto en Vielha- los taxones Cupressaceae, Pinus, Quercus, Platanus, Olea, Urticaceae y PoaceaeTotal son los más comunes. Vielha es la estación con la distribución de taxones más diferenciada a las demás. La vegetación de su entorno, por altitud y clima, es muy diferente al resto de estaciones.

Barcelona En Bcn destaca el taxón *Platanus*. Generalmente procede del Plátano de sombra, árbol muy común en las calles y parques de Bcn, aunque ya veremos que a lo largo de los años ha ido variando su presencia. Los demás taxones con porcentaje superior al 1 % son los más comunes en Catalunya: Cupressaceae, *Pinus*, *Quercus*, Urticaceae, *Olea*...

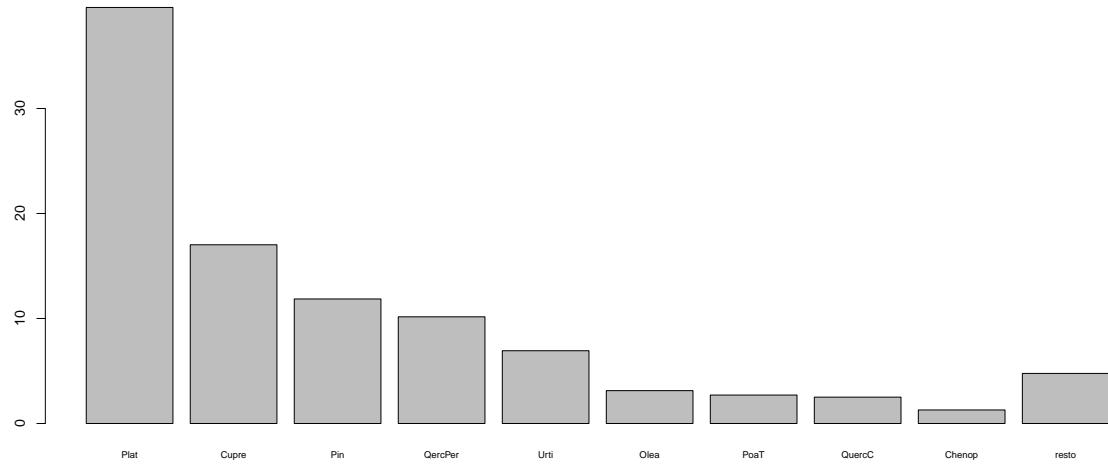


Figura 4: Distribución taxones en Barcelona (BCN).

Bellaterra. En la estación de Bellaterra destacan el *Pinus*, Cupress.., y la presencia de las plantas herbáceas más comunes, Poaceae, *Chenop*...

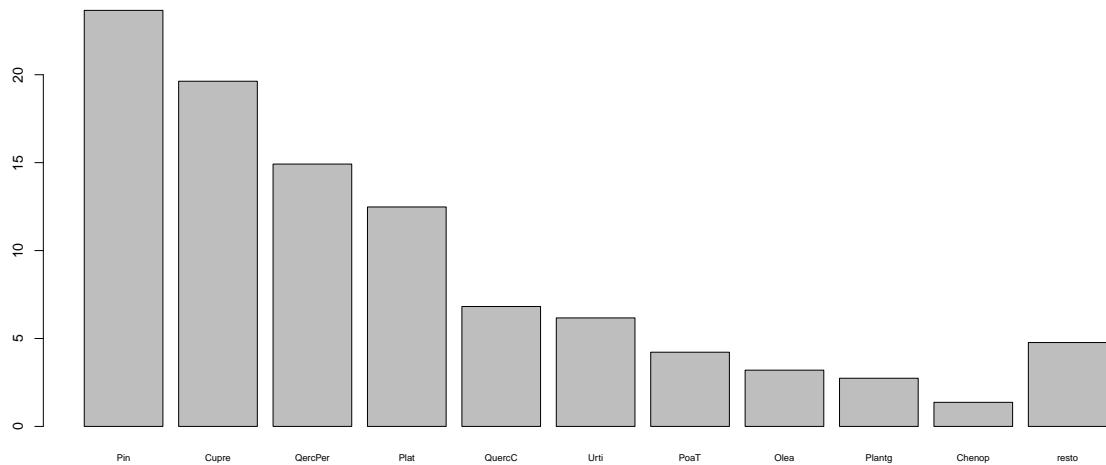


Figura 5: Distribución taxones en Bellaterra (BTU).

Girona. En Girona destacan los Quercus. Aparte de los taxones más comunes, aparecen también el Fraxinus, Castanea, Alnus y Corylus.

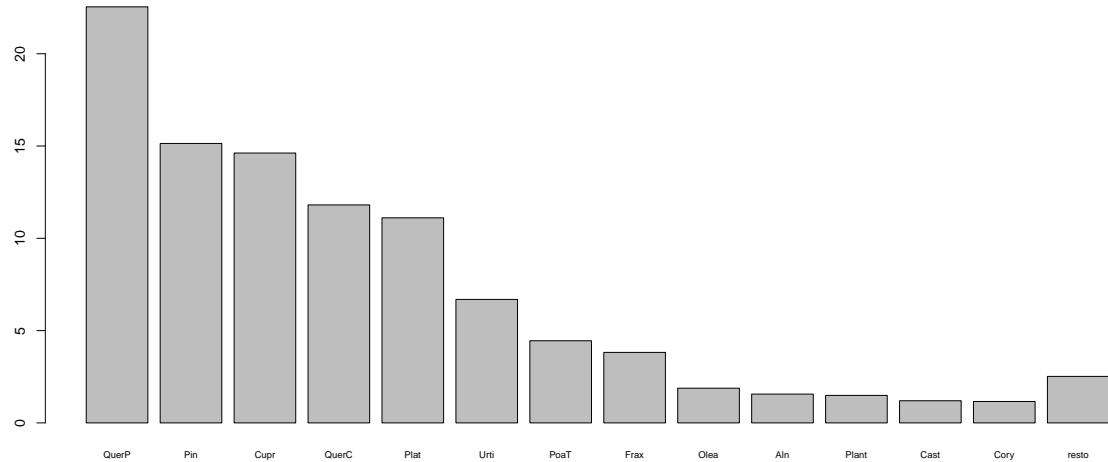


Figura 6: Distribución taxones en Girona (GIC).

Lleida. En Lleida destacan los taxones Cupressaceae y Chenopodiaceae-Amaranthaceae.

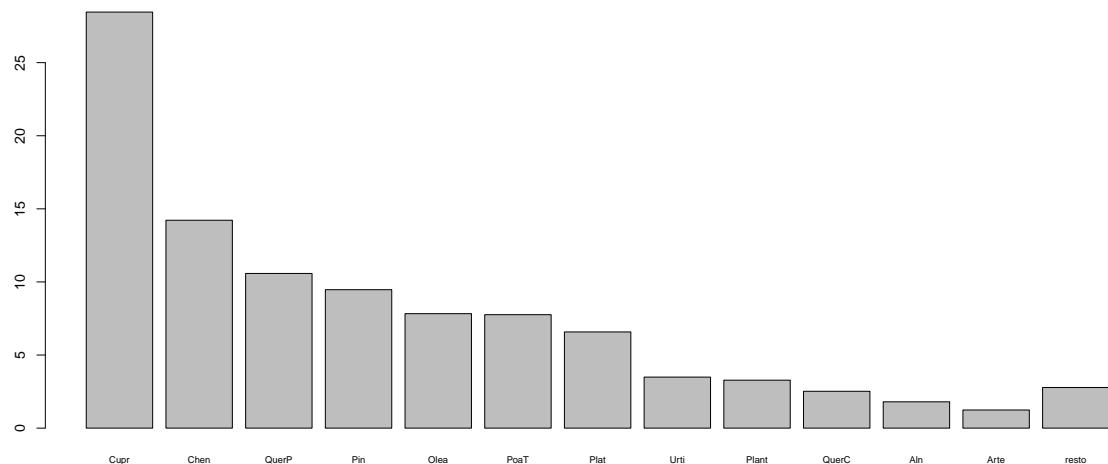


Figura 7: Distribución taxones en Lleida (LLE).

Manresa. En Manresa destacan los taxones Crupessaceae, Pinus, etc. Como particularidad destaca la fuerte presencia del Plantago.

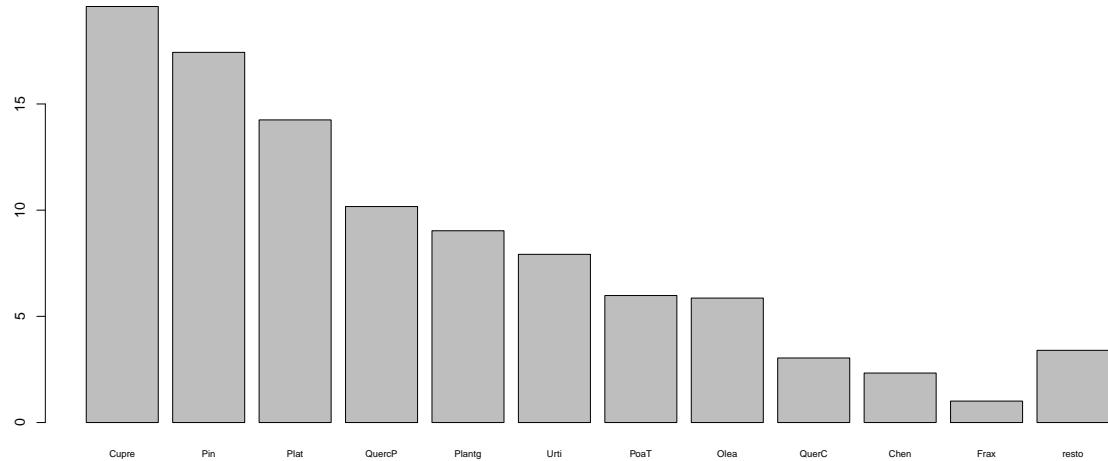


Figura 8: Distribución taxones en Manresa (MAN).

Tarragona. En Tarragona destacan el taxón Crupessaceae y la fuerte presencia del taxón Olea. En Tarragona, como en Lleida y Tortosa, es importante el cultivo de este árbol para la producción de aceite de oliva.

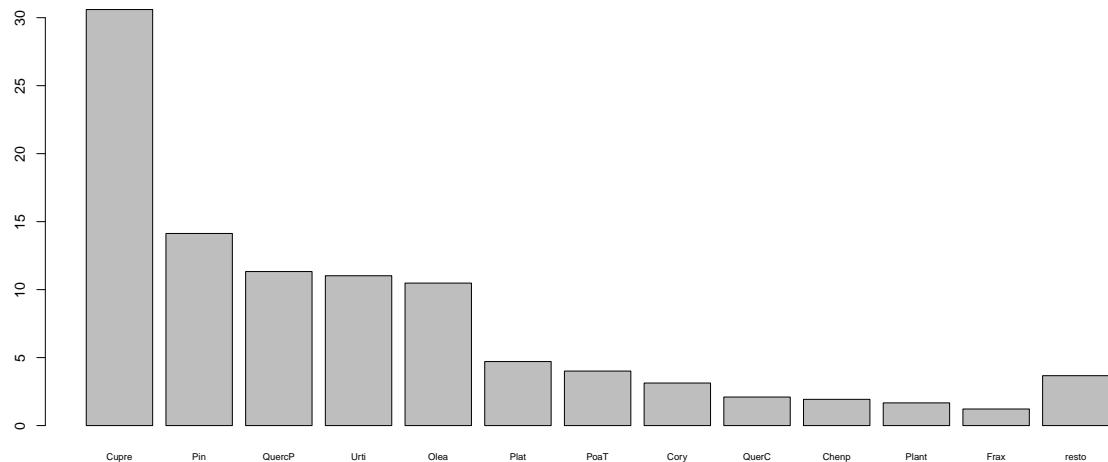


Figura 9: Distribución taxones en Tarragona (TAU).

Tortosa. En Tortosa-Roquetas hay menos taxones con porcentaje mayor al 1 %. Destaca el taxón Cupressaceae, con más del 50 % del registro total.

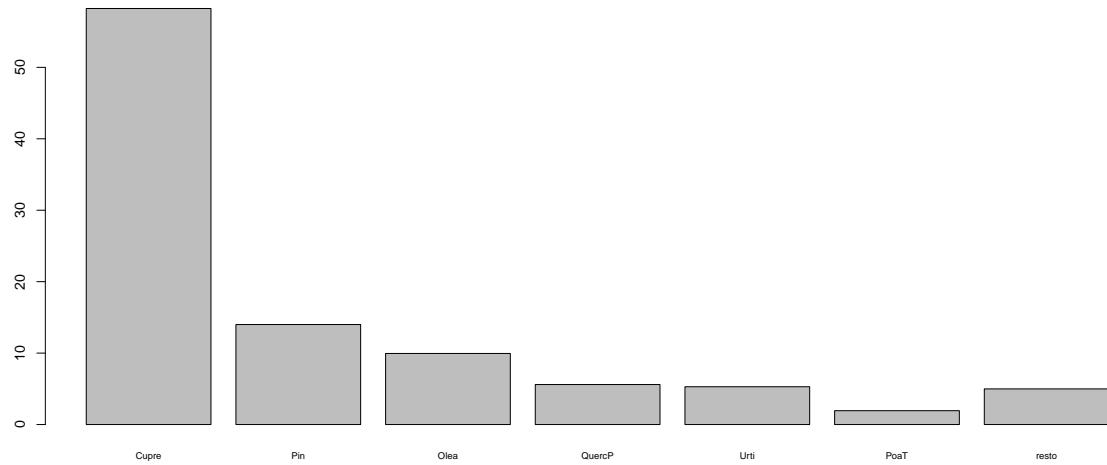


Figura 10: Distribución taxones en Tortosa (TOR).

Vielha. Ya hemos comentado la gran diferencia de vegetación en Vielha con respecto al resto de Catalunya. Entre los taxones más comunes en esta estación aparecen el *Corylus*, *Fraxinus*, *Betula* y *Fagus*, con muy poca presencia en el resto de estaciones.

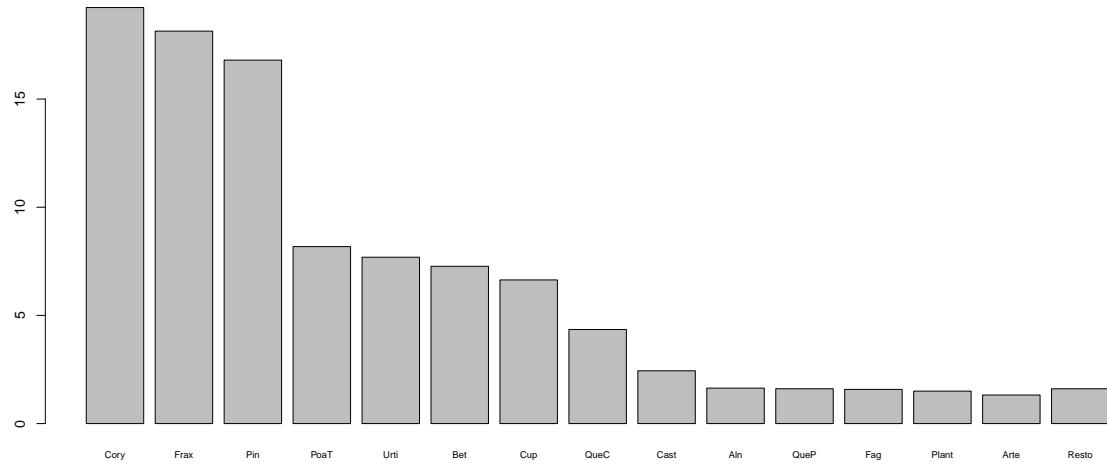


Figura 11: Distribución taxones en Vielha (TOR).

Catalunya. Por último graficamos la distribución de los principales taxones, agregada para toda Catalunya. El taxón Cupressaceae es el que tiene mayor porcentaje en Catalunya, no tanto por la presencia física de los cipreses, sino por ser una especie muy polinizadora, tanto en cantidad de polen como en duración de su floración. Detrás aparece el Pinus, que es un árbol muy presente en toda Catalunya. Destaca también el Platanus -árbol habitual en zonas urbanizadas-, los Quercus -roble y encina-, que son árboles autóctonos de nuestros bosques , el olivo -Olea- con fuerte presencia ornamental y agrícola, y por último las plantas herbáceas más comunes: Urticaceae, Poaceae, Plantago, Chenop.

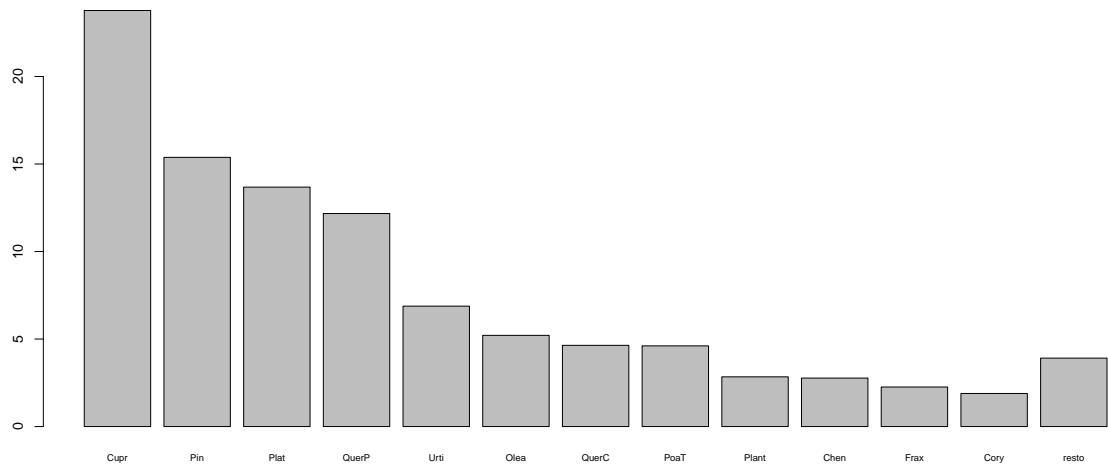


Figura 12: Distribución taxones en Catalunya (datos agregados).

2.2. Correlaciones.

El interés en estudiar las relaciones entre el API de los diferentes taxones radica en comprobar que, a pesar de la gran variabilidad, muchos de ellos parecen tener unas tendencias similares a lo largo de los años. Se trataría de analizar, por una parte, si un mismo taxón responde de forma similar en las diferentes estaciones, y, por otra parte, si hay algunos grupos de taxones con una respuesta similar, en alguna estación, a lo largo de los años registrados. Para analizar estas relaciones decidimos usar el coeficiente de correlación de Spearman. Como ya vimos, no podemos asumir normalidad para muchas de las series. No obstante, también se valoró el coeficiente de Pearson y los resultados, en cuanto a significación de la relación, fueron muy similares a los obtenidos con el coeficiente de Spearman.

Dividiremos el análisis de la correlación en dos secciones. En la primera analizamos la relación de un único taxón entre estaciones y comprobaremos que no todos los taxones tienen un comportamiento similar en las diferentes estaciones. En la segunda sección se analizarán las relaciones de todos los taxones entre sí en una misma estación y veremos que hay relaciones muy significativas entre algunos de ellos.

2.2.1. Mismo taxón entre estaciones.

La mayoría de taxones tienen una fuerte correlación entre estaciones. Parece lógico pensar que la polinización de cada especie tenga una respuesta similar a lo largo de los años. Por eso, si analizamos las series de un mismo taxón según las estaciones, encontramos tendencias similares. Pero esto no ocurre para todos los taxones.

Antes de ver algunos ejemplos hemos de comentar que se han analizado las correlaciones de Spearman en dos versiones, una para las correlaciones entre las 6 primeras estaciones -comparten datos 1996:2012-, y otra para las otras 2 estaciones -comparten datos 2006:2012. En los cuadros correspondientes aparece el valor del coeficiente de correlación de Spearman en las celdas superiores. En las celdas inferiores aparece el pvalor correspondiente, si significativo(<0.05) o cercano a la significación (<0.1), calculado según los datos 1996:2012 para las 6 primeras estaciones y 2006:2012 para las correlaciones con Vie y Tor.

Pinus. Las correlaciones son todas positivas y la mayoría significativas, sobre todo para las 6 primeras estaciones, donde el número de datos es mayor.

En el gráfico conjunto de las series del taxón Pinus podemos ver un comportamiento similar de los API para todas las estaciones.

	Pinus	BCN	BTU	GIC	LLE	MAN	TAU	VIE	TOR
1	BCN		0.72	0.62	0.63	0.69	0.62	0.64	0.35
2	BTU	0.001**		0.68	0.66	0.62	0.88	0.50	0.67
3	GIC	0.009**	0.003**		0.63	0.77	0.73	0.53	0.35
4	LLE	0.008**	0.004**	0.007**		0.87	0.82	0.78	0.67
5	MAN	0.003**	0.009**	0.001**	0.001**		0.76	0.82	0.50
6	TAU	0.009**	0.001**	0.001**	0.001**	0.001**		0.60	0.78
7	VIE				0.048*	0.034*			0.75
8	TOR						0.06		

Cuadro 10: Coef. correlación Spearman y pvalor Pinus*estación

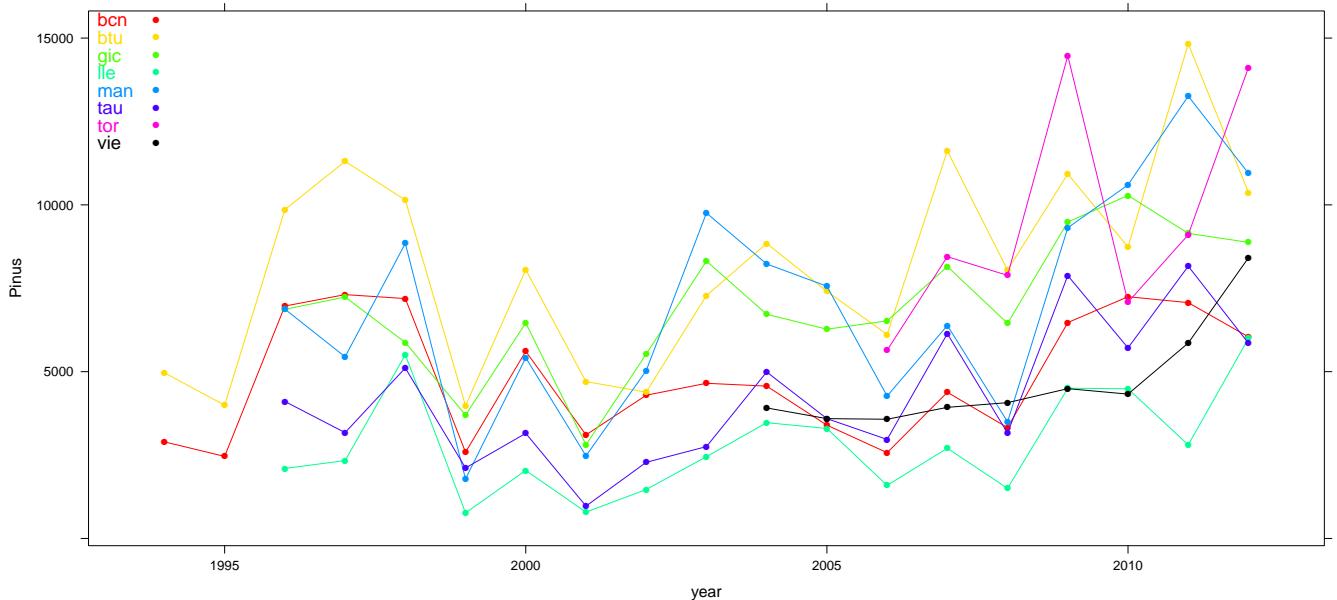


Figura 13: Series del taxón Pinus segun estación

Poaceae. Este taxón no tiene tantas correlaciones significativas. Destaca el valor negativo en los coeficientes entre Vie y la mayoría de las estaciones. Aún así en el gráfico se puede apreciar un comportamiento similar en algunos años, donde destaca el cambio entre 2004 -valores máximos para varias est.- y 2005-valores bajos.

	Poaceae	BCN	BTU	GIC	LLE	MAN	TAU	VIE	TOR
1	BCN		0.58	0.27	0.28	0.32	-0.01	-0.67	0.21
2	BTU	0.016*		0.45	0.20	0.13	0.33	-0.25	0.07
3	GIC		0.07		0.47	0.24	0.60	-0.39	0.60
4	LLE			0.057		0.54	0.47	-0.39	0.25
5	MAN				0.027*		0.43	-0.67	0.00
6	TAU				0.012*	0.061	0.08		-0.42
7	VIE								0.42
8	TOR								

Cuadro 11: Coef. correlación Spearman y pvalor Poaceae*estación

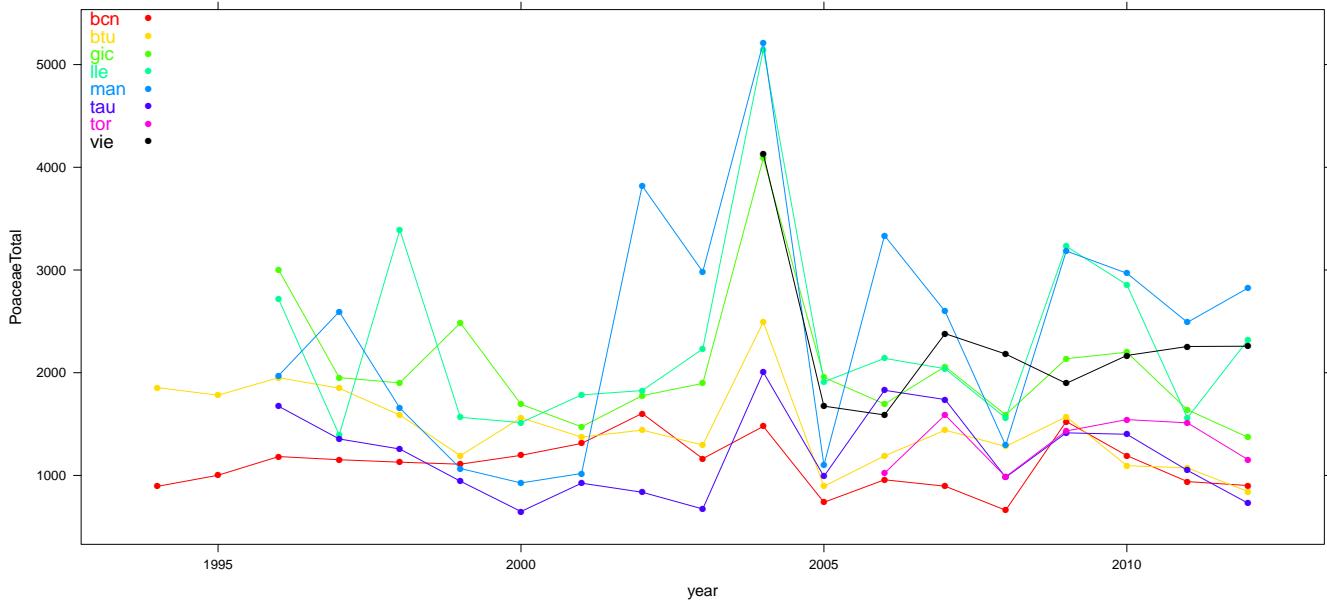


Figura 14: Series del taxón Poaceae segun estación

Urticaceae. En este taxón sí existen fuertes correlaciones entre estaciones. Las series de su API también tienen un comportamiento muy similar. En la subsección donde estudiamos las tendencias veremos que en bcn, btu, gic y man han disminuido significativamente los registros del API anual de este taxón. En el gráfico conjunto ya se aprecia esta disminución a partir de 1996.

	Urticacea	BCN	BTU	GIC	LLE	MAN	TAU	VIE	TOR
1	BCN		0.82	0.84	0.30	0.81	0.48	0.67	0.89
2	BTU	0.001**		0.78	0.63	0.70	0.76	0.92	0.96
3	GIC	0.001**	0.001**		0.32	0.81	0.55	0.78	0.96
4	LLE		0.008**			0.40	0.65	0.67	0.89
5	MAN	0.001**	0.002**	0.001**			0.53	0.28	0.46
6	TAU	0.049*	0.001**	0.022*	0.005**	0.029*		0.50	0.64
7	VIE		0.007**	0.048*					0.85
8	TOR	0.012*	0.003**	0.003**	0.012*			0.024*	

Cuadro 12: Coef. correlación Spearman y pvalor Urticaceae*estación

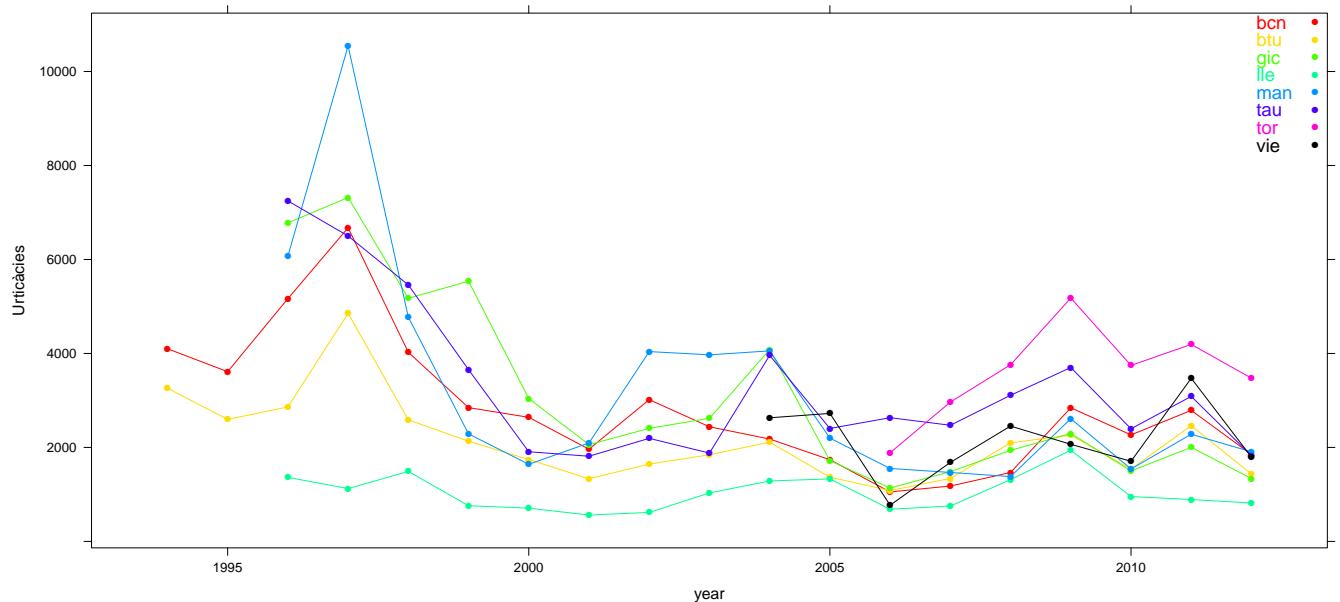


Figura 15: Series del taxón Urticaceae según estación

Resto de taxones. La mayoría tienen fuertes correlaciones, no obstante los taxones Cupre., Chenop., Frax., Olea, Plantago y Platanus son los menos correlacionados. En la mayoría de estos taxones su fuerte presencia en alguna estación, como Chenop en Lleida, Cupre. en Tor., o Platanus en Bcn, combinada con su menor presencia en las demás, afecta al resultado de los coeficientes de correlación de Spearman. Aunque, gráficamente, se intuyen tendencias similares, debemos admitir que la mayoría de taxones no han tenido un comportamiento similar a lo largo de estos últimos años en las distintas estaciones. Las particularidades de cada una de ellas son significativas.

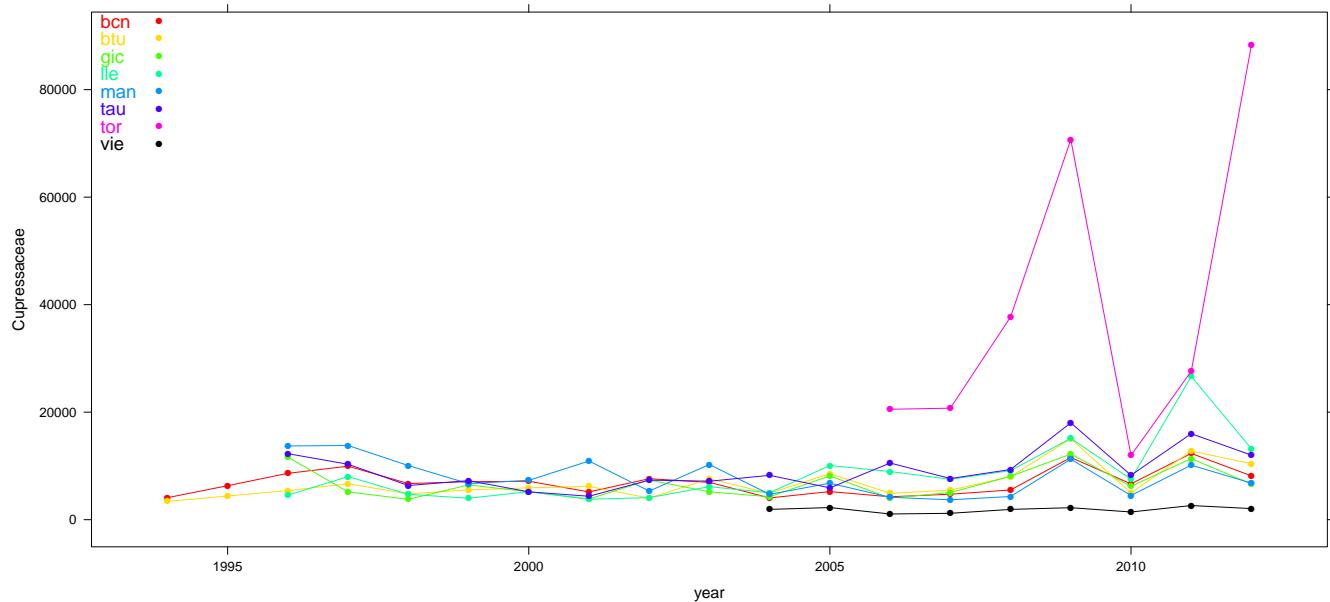


Figura 16: Series del taxón Cupressaceae según estación

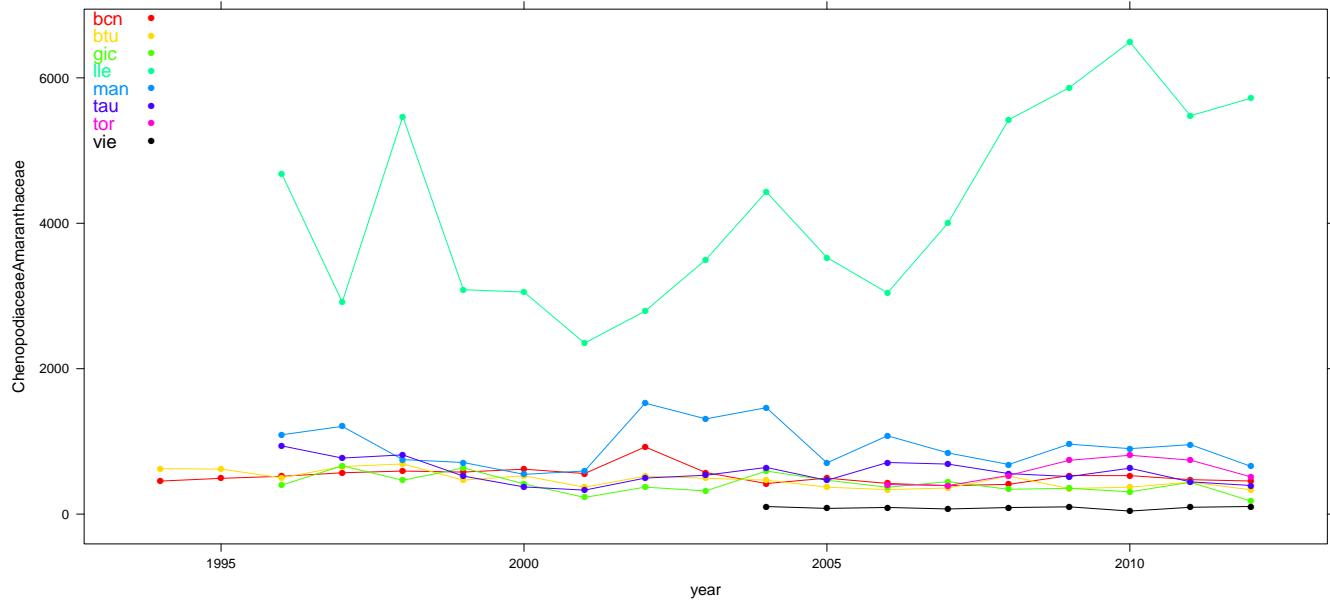


Figura 17: Series del taxón Chenopodiaceae/amaranthaceae según estación

2.2.2. Entre taxones en una misma estación.

Analizar todas las correlaciones de los 21 taxones en todas las estaciones sería materia para un trabajo específico, por lo tanto vamos a intentar resumir los resultados e intentar sacar conclusiones globales. La primera conclusión es que sí existen correlaciones entre los taxones. En el Cuadro 13 vemos el porcentaje de correlaciones significativas según el nivel de significación sea 0.1 o 0.05. Vemos que los porcentajes bajan para Vie y Tor, pero es debido al pequeño número de observaciones en ambas estaciones. Hemos preferido usar 0.1 y 0.05 como niveles de significación por tener pocas observaciones y por interesarnos una perspectiva global de las posibles correlaciones, y no tanto la significación específica de alguna en concreto.

	n	est	pvalor<0.1	pvalor<0.05
	19	BCN	28 %	18 %
	19	BTU	28 %	21 %
	17	GIC	21 %	17 %
	17	LLE	24 %	18 %
	17	MAN	37 %	29 %
	17	TAU	29 %	20 %
	9	VIE	14 %	7 %
	7	TOR	11 %	5 %

Cuadro 13: Porcentajes de corr.Spearman significativas

En el Cuadro 14 vemos los coeficientes de correlación de Spearman, en las celdas superiores, y su pvalor correspondiente, en celdas inferiores, pero sólo si este pvalor es menor que 0.1, para los 21 taxones y sólo en la estación de Manresa, que es la estación con mayor número de correlaciones significativas.

Por último, en el Cuadro 15, hemos sumarizado el número de correlaciones significativas ($pvalor < 0.05$). En las celdas superiores están marcadas con 0 las celdas cuyos taxones NO están relacionados, significativamente, en ninguna estación. En las celdas inferiores está el número total de todas las correlaciones significativas, pero siempre que los taxones relacionados hayan tenido 3 o más en el total de las 8 estaciones.

Destacan las siguientes relaciones:

- Pinus con Quercus perenne.
- Cupressaceae con Pinus.
- La relación de las plantas herbáceas: Chenop., Plantago, Poaceae, Urticaceae, etc.
- Fagus con Betula y Corylus.
- Fraxinus con Quercus caduco y Ulmus.
- Artemisia con Ambrosia y Urticaceae.

taxón	Ahn	Amb	Arte	Bet	Cast	Chen	Cory	Cupr	Fag	Frax	Olea	Pin	Pist	Plantg	Plat	Poac	Poliq	Qead	Qper	Ulm	Urtic	
Ahn	0.09	0.17	0.55	0.22	0.40	0.01	0.49	0.51	0.14	-0.21	-0.08	0.08	0.00	0.37	0.19	0.21	0.26	0.19				
Amb	0.75	0.08	0.73	0.58	0.18	0.21	-0.05	-0.22	-0.15	-0.25	0.01	-0.59	0.18	0.35	0.2	-0.19	-0.27	0.69				
Arte	0.00**	0.20	0.81	0.60	0.20	0.47	-0.03	-0.04	-0.21	0.20	0.12	0.03	-0.42	0.26	0.34	0.4	0.19	-0.32	0.85			
Bet	0.02*		0.17	0.65	0.42	-0.17	0.70	0.42	0.52	-0.01	0.18	0.51	0.16	0.76	0.64	0.37	0.38	0.22	0.18			
Cast	0.00**	0.00**	0.54	0.22	0.41	0.15	0.12	-0.05	0.23	0.07	-0.05	-0.33	0.31	0.29	0.44	0.23	-0.09	0.79				
Chen	0.01*	0.01*	0.00*	0.02*	0.26	0.09	0.45	0.33	0.41	0.19	0.1	0.54	-0.04	0.78	0.73	0.72	0.34	0.01	0.59			
Cory		0.09			-0.07	0.46	0.36	0.33	0.05	0.21	0.39	0.18	0.30	0.39	0.04	0.39	0.43	0.21				
Cupr	0.06					-0.46	-0.25	-0.26	0.22	0.31	-0.51	-0.50	-0.24	0.03	0.32	-0.11	-0.33	0.66				
Fag	0.04*		0.00**		0.06	0.06	0.00**	0.76	0.52	-0.06	-0.08	0.49	0.39	0.63	0.35	0.36	0.43	0.40	-0.04			
Frax	0.03*		0.09				0.03*	0.00**	0.67	0.37	0.23	0.49	0.66	0.60	0.24	0.57	0.73	0.63	-0.04			
Olea		0.03*							0.51	0.48	0.65	0.72	0.71	0.47	0.52	0.64	0.71	-0.22				
Pin								0.03*		0.67	0.29	0.54	0.37	0.32	0.43	0.72	0.41	0.14				
Pist									0.05*		0.00**	0.30	0.45	0.14	0.23	0.38	0.57	0.42	0.11			
Plantg									0.00**		0.02*	0.07	0.00**	0.66	0.68	0.37	0.42	0.60	0.56	-0.18		
Plat	0.01*	0.09	0.04*		0.02*	0.03*	0.04*	0.00**	0.01*	0.00**	0.00**	0.09	0.42	0.03	0.26	0.73	0.73	0.73	-0.55			
Poac		0.00**	0.00**		0.00**	0.00**	0.00**	0.01*	0.00**				0.68	0.61	0.68	0.32	0.18					
Poliq		0.00**	0.00**		0.00**	0.00**	0.00**	0.00**	0.00**				0.00**	0.00**	0.36	0.34	0.14	0.48				
Qead		0.08		0.00**					0.01*	0.03*	0.08	0.09	0.01*	0.00**	0.00**	0.51	0.21	0.41				
Qper									0.08	0.00**	0.00**	0.00**	0.09	0.02*	0.00**	0.03*	0.50	0.07				
Ulm	0.00**	0.00**		0.00**	0.01*	0.08							0.02*	0.02*	0.04*	0.04*		-0.35				
Urtic															0.05							

Cuadro 14: Coef. correlación Spearman y pvalor en Manresa

N_cor	Aln	Amb	Arte	Bet	Cast	Chen	Cory	Cupr	Fag	Frax	Olea	Pin	Pist	Plantg	Plat	Poac	Polig	Qcad	Qper	Ulm	Urtic
Aln	0	0									0	0	0	0	0	0			0		
Amb		0									0		0								
Arte	5!	0			0				0	0											
Bet			0						0		0	0	0						0	0	
Cast									0	0	0	0							0		
Chen	3	4																			
Cory											0	0	0	0		0	0	0	0		
Cupr		3																			
Fag			5!			4					0			0							
Frax	3						3	3								0					
Olea									5!		3								0		
Pin										4							0				
Pist																					
Plantg		4			6!		3														
Plat	3										3						0				
Poac		3		4	6!						3			6!							
Polig						5!								3	4		0				
Qcad										4	3					4				0	
Qper											7!!		3	3	3						0
Ulm								4		4						3	4				
Urtic	3	6!		3	6!		3														

Cuadro 15: Número de correlaciones significativas entre taxones (por estaciones).

2.3. Tendencias significativas.

El análisis de la tendencia del API anual de los diferentes taxones podría indicarnos, entre otros:

- El aumento o la reducción de la especie en el territorio.
- Los efectos del cambio climático en la polinización⁸.
- Un posible aumento de casos de alergias en la población sensible.

Ya comentamos en la introducción el artículo del investigador Fernandez-Llamazares⁴ et al, en el cual se estudia la influencia del cambio climático en las tendencias positivas del API anual de varios taxones en Catalunya. En nuestro estudio podemos confirmar sus resultados con una observación más, correspondiente al año 2012, y ampliarlos, al analizar más taxones. No obstante, como estadísticos, no nos corresponde analizar el motivo o la causa de estas tendencias, sino simplemente constatar que son tendencias estadísticamente significativas.

Para analizar estas tendencias hemos preferido realizar varios tests, paramétricos y no paramétricos, para todas las series de cada taxón en cada estación:

- Test de Mann-Kendall estandarizado.
- Test Rho de Spearman estandarizado.
- Regresión lineal.
- Estimador de Theil-Sen para la pendiente en caso de significación en Mann-Kendall o Rho de Spearman.

En el Cuadro 16 podemos ver las tendencias significativas. Estas pueden ser positivas o negativas, es decir, que el API del taxón tiende a aumentar o disminuir en nuestras series. También hemos analizado las series de la proporción de cada taxón en cada estación. El motivo es poder comparar si esa tendencia significativa de mayor o menor API anual pudiera ser producto de una mayor o menor presencia de la especie en la estación. En este caso únicamente indicamos si el test de Man-Kendall es significativo para la serie anual de la proporción del taxón en la estación y, entre parentesis, el signo del estimador Theil-Sen para esas series de proporciones anuales significativas (última columna del Cuadro 16).

Podemos observar que en todos los taxones con tendencia negativa (estimador Theil-Sen negativo), la tendencia de su proporción en la estación también lo es, mientras que en la gran mayoría de taxones con tendencia positiva, su serie de proporción en la estación

no lo es. Esto puede indicar que la tendencia a producir más polen no tiene por qué estar relacionada con una mayor presencia de la especie en el territorio, mientras que lo contrario sí ocurre, es decir, una menor polinización sí parece estar influenciada por la disminución de la presencia de la especie en el territorio.

Antes de analizar algunas tendencias concretas, destacaremos las tendencias negativas de la mayoría de las plantas herbáceas, debido al aumento de la construcción en los años de bonanza económica -no olvidemos que las estaciones están situadas en un entorno urbano-, el aumento de los *Quercus* -especie autóctona en nuestros bosques que va recuperando el terreno perdido tras las replantaciones de otras especies no autóctonas-, y el aumento de los árboles ornamentales como *Olea*, *Ulmus*, *Platanus*, *Fraxinus* y *Cupressaceae*.

taxón	estación	MK.2sided.	Reg.lineal	Sperman.rho	Theil.Sen.est.	MK.prop.sig
artem	gic	0.01	0.02	0.02	-4.65	si (-)
artem	lle	0.007**	0.001**	0.005**	-28.22	si (-)
artem	man	No sig	0.007**	0.04	-10.85	si (-)
chenop	btu	0.001**	0.001**	0.001**	-15.97	si (-)
chenop	gic	0.04	0.05	0.05	-12.27	si (-)
chenop	lle	0.02	0.02	0.01	185.00	no
corylus	gic	No sig	No sig	0.04	26.95	no
cupres	btu	0.006**	0.004**	0.007**	312.74	no
cupres	lle	0.002**	0.003**	0.001**	533.36	si (+)
fraxi	bcn	0.03	0.02	0.02	10.54	no
fraxi	tau	0.001**	0.001**	0.001**	22.42	si (+)
fraxi	man	0.009**	0.009**	0.008**	18.26	no
olea	bcn	No sig	0.02	0.03	43.99	si (+)
olea	btu	0.02	0.02	0.02	33.52	no
olea	tau	No sig	0.03	0.03	139.62	si (+)
olea	man	0.004**	0.004**	0.001**	225.09	si (+)
pinus	gic	0.03	0.006**	0.01	244.06	no
pinus	tau	0.023	0.01	0.02	240.62	si (+)
pinus	vie	0.005**	0.01	0.001**	278.95	no
platan	bcn	No sig	0.01	0.02	-709.80	si (-)
platan	btu	0.02	0.004**	0.008**	131.01	no
platan	gic	0.006**	0.02	0.002**	351.71	si (+)
platan	man	0.0001**	0.0001**	0.0001**	772.05	si (+)
poacea	btu	0.001**	0.009**	0.001**	-47.47	si (-)
polig	gic	0.02	0.008**	0.02	-5.22	si (-)
querCad	bcn	0.0078**	0.01	0.003**	56.83	si (+)
querCad	btu	0.003**	0.004**	0.004**	112.80	no
querCad	vie	0.04	0.05	0.03	161.35	no
querPer	btu	0.02	0.01	0.01	198.80	no
querPer	gic	0.03	0.03	0.04	397.88	no
querPer	tau	0.02	0.02	0.02	140.35	no
querPer	man	0.02	0.01	0.006**	241.08	no
querTot	btu	0.004**	0.002**	0.004**	296.36	no
querTot	gic	No sig	0.04	0.03	381.65	no
querTot	tau	No sig	0.03	0.03	129.08	no
querTot	man	0.04	0.02	0.01	278.38	no
querTot	vie	0.02	0.01	0.01	169.66	no
ulmus	btu	0.001**	0.001**	0.002**	16.89	no
ulmus	man	0.02	0.001**	0.0001**	278.38	si (+)
ulmus	tor	0.04	0.007**	0.02	479.33	no
urtic	bcn	0.0021**	0.002**	0.001**	-177.98	si (-)
urtic	btu	0.02	0.02	0.01	-75.46	si (-)
urtic	gic	0.001**	0.0001**	0.0001**	-296.00	si (-)
urtic	man	0.009**	0.007**	0.007**	-204.40	si (-)

Cuadro 16: Tendencias significativas según taxon y estación.

En los siguientes gráficos podemos ver las series con tendencia más acusada.

Platanus. Este taxón ha tenido una tendencia creciente muy clara en Manresa, mientras que en Barcelona tuvo una fuerte bajada en 2004-2005, en esos años se tuvo que actuar contra una plaga de un insecto, podando y eliminando muchos árboles justo antes de la floración. No obstante, podemos ver que en BCN la tendencia de los últimos registros es a aumentar el API.

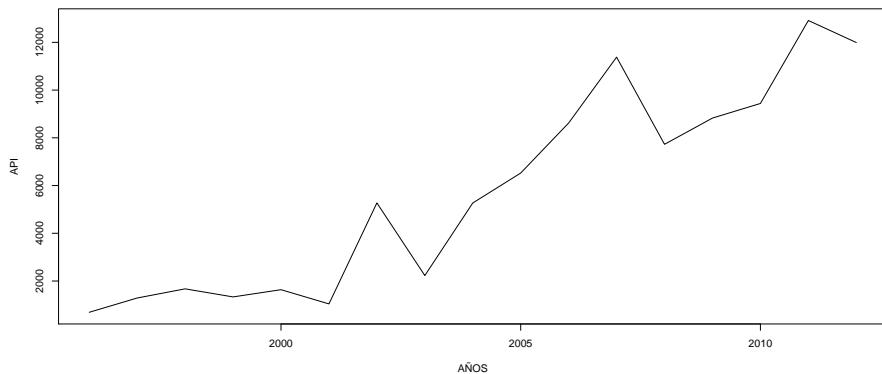


Figura 18: Serie API taxón Platanus en MAN (1996.2012).

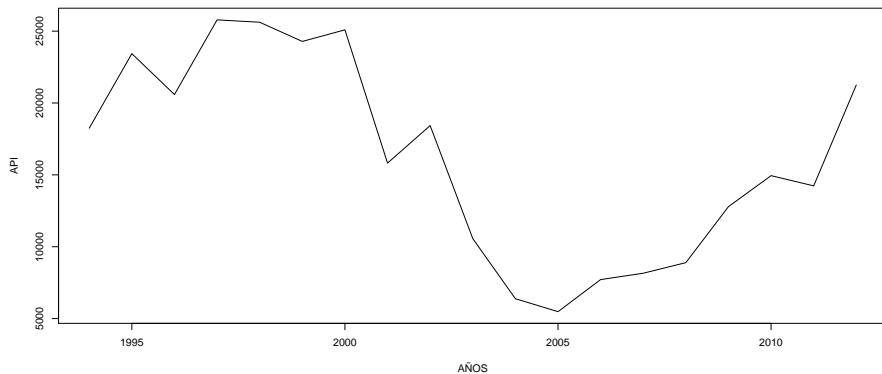


Figura 19: Serie API taxón Platanus en BCN (1994:2012).

Chenop. En el gráfico vemos la serie de Chenop. en Bellaterra. Esta planta herbácea ha perdido presencia por la intervención humana en el paisaje de la zona, ya que la urbanización y el ajardinado han aumentado en Bellaterra.

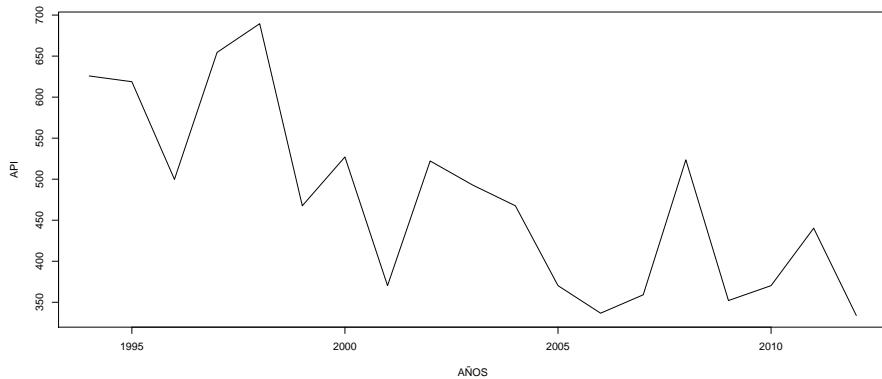


Figura 20: Serie API taxón Chenop. en BTU (1994:2012).

Ulmus. Una de las tendencias crecientes más acusada es la del taxón Ulmus en Manresa. No hemos podido encontrar los posibles motivos de este crecimiento, aunque ya hemos comentado su uso como planta ornamental urbana. En el gráfico se aprecia un fuerte aumento en 2004 que parece tener continuidad en años posteriores.

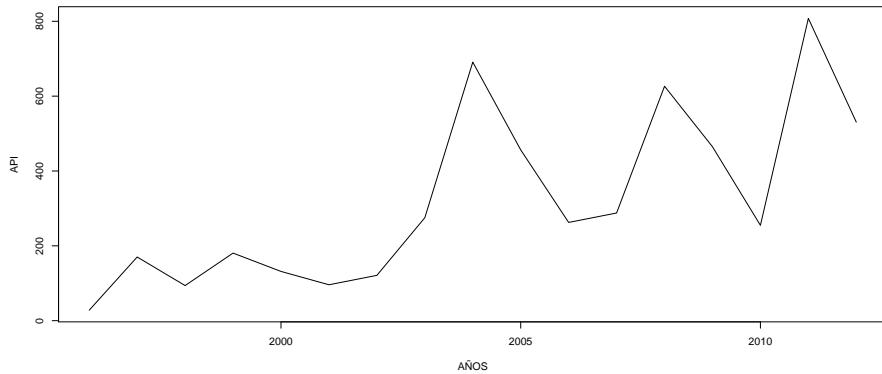


Figura 21: Serie API taxón Ulmus en MAN (1996:2012).

2.4. Comparativa de estaciones.

2.4.1. Introducción.

El grupo de investigación en palinología (PIA)¹ mostró interés en comparar la información que recibe de las diferentes estaciones. Previendo la necesidad, u obligación, de recortar gastos y tener que cerrar alguna estación, necesitaba analizar si existían estaciones que no aportasen información significativamente diferente a otras, buscando las candidatas idóneas para un posible cierre, o para centrar sus recursos en las más representativas del conjunto de 8 estaciones.

En el análisis descriptivo ya hemos encontrado rasgos que permiten intuir cierta similitud entre estaciones. Hemos visto que algunos de los taxones están bastante correlacionados en muchas de las estaciones y que los taxones más comunes se repiten en muchas de ellas. Pero también hemos encontrado grandes diferencias. La estación de Vielha no se puede comparar con ninguna otra. Sus registros contienen información sobre taxones que apenas aparecen en las otras estaciones. En esas otras 7 estaciones existen particularidades propias de cada una, y en una de ellas, Tortosa, sólo disponemos de series con 7 años (2006:2012).

Por todo ello hemos decidido centrar nuestro análisis en las 6 estaciones de las que disponemos de más datos conjuntos y son más homogéneas. En ellas podemos comparar las 21 series temporales de los 21 taxones durante estos últimos 17 años (1996:2012).

Estas 6 estaciones - BCN, BTU, GIC, LLE, MAN, TAU- comparten el mismo clima mediterráneo (ver Cuadro 2), con ciertos matices diferenciales. La temperatura media es muy similar y la altura tampoco difiere en exceso. Es lógico pensar que el hábitat ha de ser, cuanto menos, bastante similar. Aún así reiteramos que cada una tiene alguna particularidad que la hace algo diferente a las demás. Por ese motivo hemos descartado compararlas de dos en dos. La mayoría de estudios publicados referentes a esta cuestión, comparan sólo dos estaciones y en un periodo de tiempo corto, poniendo más énfasis en la comparación del registro diario⁹. Nosotros basaremos nuestro análisis en una comparación conjunta. Para ello crearemos una estación ficticia (Catalunya) agregando todos los datos de las 6 estaciones. Esta estación ficticia recogerá toda la información sobre el polen en Catalunya -excepto Vielha y Tortosa, que son estaciones en el límite del territorio. Comparando la información de esa estación ficticia con las 6 estaciones reales, trataremos de encontrar la estación cuya información podría representar mejor la información agregada que se recogería en esa estación conjunta.

Para conocer qué estación puede ser la más representativa de las 6 comparadas, hemos decidido no usar los registros del API anual de cada estación, sino el porcentaje que ese API anual representa en el agregado para todos los taxones. La población de cada especie es muy diferente en la zona de influencia de cada estación. No estamos interesados, a la hora de encontrar una estación representativa, en conocer los valores exactos de un taxón

concreto en alguna estación, sino en tener una muestra de valores que pueda explicar, hasta cierto punto, el comportamiento de la polinización en toda Catalunya -excepto Vielha y Tortosa.

Nuestros datos a comparar serán los porcentajes anuales de cada taxón. Tenemos esos datos para las 6 estaciones reales en matrices A_j , $j = 1, \dots, 6$ (BCN,BTU,GIC,LLE,MAN,TAU) con 17 vectores anuales (filas, 1996:2012), cada uno de ellos con 21 valores (columnas, taxones) cuya suma es 100. Esta sería la matriz tipo:

$$\begin{pmatrix} \%_{1,1996,1,alnus} & \cdots & \%_{1,1996,11,olea} & \cdots & \%_{1,1996,21,urtica} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \%_{8,2003,1,alnus} & \cdots & \%_{8,2003,11,olea} & \cdots & \%_{8,2003,21,urtica} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \%_{17,2012,1,alnus} & \cdots & \%_{17,2012,11,olea} & \cdots & \%_{17,2012,21,urtica} \end{pmatrix}$$

También hemos calculado esa matriz para la estación ficticia conjunta- matriz B - que recordemos agrega la suma de los diferentes API de las 6 estaciones. En el Cuadro 17 podemos ver la matriz con los datos de esta estación ficticia conjunta. Las filas representan el porcentaje anual del API de cada taxón dentro de la suma de API total recogido en el año. Todas las filas suman 100.

year	alnus	ambrosia	artem	betula	castan	chenop	corylus	cypress	fagus	fraxin	olea	pinus	pistac	planta	platanus	poaceae	poligo	querad	querper	ulmus	urtica
1 1996	0.5	0.1	1.0	0.1	0.9	3.4	0.2	23.8	0.0	1.2	2.9	15.6	0.1	2.3	12.4	5.3	0.3	5.3	11.8	0.2	12.5
2 1997	1.1	0.0	1.2	1.0	0.3	2.6	0.9	20.8	0.0	1.7	3.3	14.2	0.4	2.7	15.0	4.0	0.3	4.9	10.7	0.4	14.3
3 1998	0.6	0.0	0.7	0.1	0.6	3.9	0.5	16.3	0.0	1.1	2.9	19.1	0.2	2.2	16.6	4.9	0.4	3.5	15.8	0.1	10.5
4 1999	0.9	0.0	0.7	0.4	1.3	3.2	2.6	19.9	0.1	2.6	5.3	8.0	0.3	2.1	20.3	4.5	0.2	5.0	12.1	1.0	9.3
5 2000	0.6	0.0	0.4	0.0	0.6	3.1	0.7	20.1	0.0	1.6	2.9	17.1	0.2	1.8	19.7	4.2	0.3	5.8	13.7	0.5	6.5
6 2001	1.1	0.0	0.5	0.5	0.4	3.2	0.8	24.7	0.0	1.5	6.4	10.7	0.3	1.4	17.8	5.7	0.2	5.7	11.4	0.4	7.1
7 2002	0.7	0.0	0.6	0.6	0.6	3.4	1.2	18.2	0.1	1.3	2.8	11.7	0.3	5.8	16.9	5.8	0.2	6.2	16.0	0.4	7.1
8 2003	0.5	0.0	0.4	0.5	0.6	3.3	0.6	21.4	0.0	0.9	7.0	17.4	0.4	3.9	10.9	5.1	0.3	5.0	14.6	0.3	6.8
9 2004	1.4	0.1	0.7	1.2	1.3	3.6	1.6	14.2	0.4	1.6	4.6	16.8	0.2	5.2	9.0	9.3	0.4	5.6	13.9	0.7	8.1
10 2005	0.5	0.0	0.4	0.4	0.7	3.4	0.6	25.1	0.0	1.0	6.9	17.7	0.5	2.6	12.7	4.3	0.3	5.0	11.1	0.8	6.0
11 2006	1.3	0.0	0.3	0.6	0.2	3.1	1.6	19.2	0.1	2.2	6.7	12.6	0.1	3.3	15.9	5.8	0.3	7.1	14.4	0.8	4.2
12 2007	0.5	0.0	0.3	0.8	0.3	3.1	1.8	15.8	0.0	1.2	6.2	18.3	0.2	5.3	16.4	5.0	0.3	2.9	17.0	0.5	4.0
13 2008	1.6	0.0	0.5	0.3	0.4	4.0	1.2	22.2	0.0	2.5	4.2	13.0	0.2	4.9	19.4	3.7	0.2	5.7	9.2	0.9	5.7
14 2009	0.6	0.0	0.2	0.4	0.5	2.7	0.9	26.2	0.1	1.4	7.7	15.3	0.4	2.2	11.9	4.1	0.2	4.9	15.1	0.4	4.9
15 2010	0.7	0.0	0.5	0.4	0.3	3.7	0.5	15.4	0.0	1.4	6.7	19.0	0.1	4.0	15.1	4.7	0.2	6.4	16.2	0.5	4.1
16 2011	0.6	0.0	0.2	0.2	0.4	2.6	1.0	27.9	0.0	1.5	3.6	17.3	0.5	3.4	15.5	2.7	0.1	5.5	12.0	0.7	4.2
17 2012	0.9	0.0	0.3	0.1	0.4	2.7	0.6	20.2	0.0	1.3	5.6	17.0	0.2	1.4	19.1	3.2	0.1	4.4	18.6	0.5	3.2

Cuadro 17: Matriz B de porcentajes por taxón, para cada año, en la estación ficticia conjunta

Para evitar que el peso de alguna estación real en la conjunta fuera excesivo y provocara resultados muy favorables para esa supuesta estación, también hemos creado otras 6 estaciones ficticias restringidas a la suma de los datos de las 5 estaciones diferentes a la comparada -matrices C_j , es decir, que si comparamos BCN con esta estación ficticia

restringida, en ella la suma de los API se limitará a las otras 5 estaciones, excluyendo BCN. El motivo es reafirmar la representatividad de la estación aún cuando sus valores no se incluyan en la estación ficticia, y conseguir mayor independencia entre los vectores anuales.

2.4.2. Vector de porcentajes anuales de taxones.

Distancia entre proporciones anuales. Nuestro estadístico comparativo será la distancia de Mahalanobis. Con ella pretendemos determinar la similitud entre dos variables aleatorias multidimensionales. En R existe una función¹⁰ que calcula el cuadrado de esa distancia. Con ella obtendremos la distancia de cada vector anual de porcentajes de cada estación real y el correspondiente vector de la estación ficticia -considerado como vector de medias-, usando como matriz de covariancias la de la matriz de los 17 vectores de esta estación ficticia. El análisis se repetirá con la estación ficticia restringida -suma de todos los valores excepto el de la estación comparada. La fórmula usada por la función Mahalanobis de R es:

$$D_i^2(\vec{x}_i, \vec{\mu}_i) = (\vec{x}_i - \vec{\mu}_i)^T \Sigma^{-1} (\vec{x}_i - \vec{\mu}_i)$$

donde \vec{x}_i es el vector anual de porcentajes en la estación comparada, $i = 1996, \dots, 2012$, $\vec{\mu}_i$ es el correspondiente vector en la estación ficticia, considerado como media de la distribución, y Σ la matriz de covariancias de los vectores columna (taxones) de la matriz de la estación ficticia.

Para cada estación $j = 1, \dots, 6$ realizamos la comparación, $A_j versus B$, y obtendremos un vector \vec{v}_j con las 17 distancias anuales. Son los 6 vectores columna del Cuadro 21. La idea es conocer cuál es la estación que tiene menores distancias, que, teóricamente, será la que mejor representaría, en porcentajes anuales, la supuesta estación Catalunya. El criterio es que la estación j más representativa es la que tiene menor la norma del vector \vec{v}_j . Debemos concretar este criterio fijando la norma que vamos a usar, que puede ser cualquiera de las tres que proponemos a continuación.

Para $\vec{v}_j = (v_{j,1}, \dots, v_{j,17})$, consideramos las siguientes normas:

- Norma ℓ_1 :

$$\|\vec{v}_j\|_1 = \sum_{n=1}^{17} |v_{j,n}|$$

- Norma ℓ_2 o norma euclídea:

$$\|\vec{v}_j\|_2 = \sqrt{\sum_{n=1}^{17} v_{j,n}^2}$$

- Norma ℓ_∞ :

$$\|\vec{v}_j\|_\infty = \max_{1 \leq n \leq 17} |v_{j,n}|$$

En el Cuadro 18 podemos ver las normas de los vectores \vec{v}_j para las comparaciones entre cada estación real A_j y la ficticia conjunta B (tipo .a) , y para las comparaciones entre cada estación real A_j y la conjunta restringida correspondiente C_j (tipo .b).

Los resultados nos indican que la estación de Bellaterra es la que tiene más proximidad, ya que sus valores son los menores para todas las normas calculadas.

	BCN	BTU	GIC	LLE	MAN	TAU
norma 1.a	123828	37843	53953	75835	78851	81689
norma euclídea.a	37153	10999	15081	32420	26610	21599
norma infin.a	19261	5106	6399	29045	19539	9593
norma 1.b	183063	48827	75555	88087	86486	113860
norma euclídea.b	51808	14014	20691	33360	30330	29509
norma infin.b	24434	6401	8249	28094	23365	12010

Cuadro 18: Comparativa de estaciones. Normas de los vectores distancia de Mahalanobis

Test χ^2 de bondad de ajuste. También decidimos realizar un test de bondad de ajuste, basado en el test ji-cuadrado de Pearson. En este caso hemos limitado los vectores anuales de porcentajes de cada taxón a los taxones más comunes en la estación ficticia conjunta, con un valor de ese porcentaje anual superior o muy cercano al 5 %. Estos taxones son: Chenop, Cupressacea, Olea, Pinus, Plantago, Platanus, PoaceaeTotal, Quercus caduco, Quercus perenne y Urticaceae. El resto de taxones han sido incluidos en una categoría conjunta que suma todos los valores de sus porcentajes en uno único conjunto. La idea es tomar este vector de 11 porcentajes como la distribución teórica de las 11 categorías -10 taxones más comunes, más el resto- en Catalunya. Sobre esta distribución teórica, para cada año 1996:2012, analizamos si los porcentajes reales de esas mismas 11 categorías, en cada estación real comparada, han seguido esa distribución teórica en los diferentes años. Este análisis tambien se realizó con los mismos dos criterios anteriores, usando la matriz de la suma total y la de la suma restringida.

La función de R con la que hemos calculado el pvalor de los cuadros 19 y 20 ha sido chisq.test¹¹, que permite analizar si el vector anual de porcentajes de cada estación real sigue la distribución del vector anual conjunto, en este caso en proporciones. En el Cuadro 19 están únicamente los pvalores $\geq 0,01$ conseguidos para cada año y estación, tomando como distribución teórica la de la matriz conjunta total, mientras que en el Cuadro 20 la matriz teórica es la restringida -recordemos que es la suma total menos los valores de la estación testada.

Los resultados nos indican que la estación de Bellaterra es la que obtiene mayor número de pvalores a favor de la hipótesis nula, que supone que el vector muestral sigue la

distribución teórica. También comprobamos que hay años -2005 y 2009- donde parece existir mayor similitud de respuesta entre las estaciones.

	Año	BCN	BTU	GIC	MAN	TAU
1	1996					
2	1997		0.092	0.028	0.012	0.014
3	1998		0.518		0.013	
4	1999		0.397	0.013	0.010	
5	2000		0.358			
6	2001		0.067			
7	2002		0.896	0.020		
8	2003		0.814		0.033	
9	2004		0.499		0.019	
10	2005	0.224	0.487	0.106	0.438	
11	2006	0.020	0.383			
12	2007	0.030	0.112	0.112		
13	2008	0.023	0.137			
14	2009	0.022	0.684	0.031	0.504	0.056
15	2010		0.637			
16	2011	0.055	0.194	0.012		
17	2012		0.391		0.339	

Cuadro 19: Pvalores($>=0.01$) del test de bondad de ajuste. Matriz total

Comparativa de resultados entre distancias y bondad de ajuste. Todo parece indicar que la estación de Bellaterra es la más adecuada para poder representar, con mayor fiabilidad, todo el polen que se recoge en Catalunya -exceptuando Vielha y Tortosa. Los resultados de ambos análisis comparativos parecen haber coincidido en ese aspecto. Pero debemos analizar ciertas coincidencias y, también, ciertas discrepancias entre ambos análisis.

En el Cuadro 21 podemos ver los 6 vectores (columnas) de las distancias de Mahalanobis consideradas entre las matrices A_j y la matriz conjunta ficticia B . Si analizamos esas distancias vemos que BTU no consigue la menor puntuación en todos los años. De hecho solo lo consigue en 7 de ellos, mientras que LLE lo consigue en 6, y, sin embargo, LLE no consigue ningún pvalor $\geq 0,01$ en los tests de bondad de ajuste anuales. Creemos que esto es debido a la fuerte presencia de los taxones Cupressaceae y Chenop en LLE, con valores de API muy extremos en algunos años, destacando el valor de 2011, año con excepcional presencia de polen de ciprés en Lleida. Con esto pretendemos indicar que la fuerte variabilidad de los API anuales para algunos taxones en algunas estaciones influye mucho en algunos resultados de los análisis previos. Tal vez los mejores resultados de BTU se deban a una menor variabilidad de sus registros del API, y, consecuentemente de los porcentajes, e, incluso, al hecho de ser la estación más centrada en el territorio.

	Año	BCN	BTU	MAN
1	1996			0.035
2	1997			
3	1998		0.204	
4	1999		0.106	
5	2000		0.079	
6	2001			
7	2002		0.751	
8	2003		0.589	0.137
9	2004		0.194	
10	2005	0.041	0.180	
11	2006		0.107	
12	2007			0.307
13	2008			
14	2009		0.400	
15	2010		0.396	0.121
16	2011		0.020	0.035
17	2012		0.075	

Cuadro 20: Pvalores(≥ 0.01) del test de bondad de ajuste. Matriz restringida.

Las coincidencias entre los análisis son evidentes en los años 2005 y 2009. En esos años es donde más pvalores ≥ 0.01 se consiguen en el test de bondad de ajuste. Además, si consideramos para cada año 1996:2012 el vector fila de las distancias de Mahalanobis, dadas en el cuadro 21, y consideramos, por ejemplo, la norma ℓ_1 resulta que los vectores con menor valor de esa norma corresponden a esos años 2005 y 2009.

2.4.3. Vector con serie histórica de porcentajes por taxón.

En este apartado queremos ser críticos con nuestros resultados. Parece evidente que la estación de Bellaterra es la más representativa a la hora de poder ajustar la distribución conjunta de los diferentes taxones para toda Catalunya (excepto Vielha y Tortosa), pero ¿lo es también para cada taxón individual?. Para responder la pregunta anterior compararemos los vectores de la serie histórica de los porcentajes de cada taxón en cada estación, con su correspondiente en la estación ficticia. Estos vectores corresponderían a las 21 columnas de las matrices A_j y B . En este caso no podemos realizar un test de bondad de ajuste, ni tampoco tiene mucho sentido calcular la distancia entre vectores. Para realizar este análisis usaremos el test no paramétrico de los rangos con signo de Wilcoxon, considerando que los dos vectores comparados son datos emparejados que comparten año. Nuestro objetivo es determinar, para cada taxón, si su presencia proporcional es la misma, a lo largo de los 17 años, en cada estación real y la ficticia. Pretendemos saber si alguna estación concreta representa mejor un mayor número de

	BCN	BTU	GIC	LLE	MAN	TAU
1996	8398	4729	1890	1409	10691	9593
1997	9282	2601	3094	2936	7461	6674
1998	13751	4385	5331	2651	4096	2897
1999	19261	143	1667	1925	5609	3738
2000	15864	1790	3670	1417	5378	2665
2001	12867	1059	1924	3283	19539	4309
2002	8045	653	1146	756	2900	3018
2003	3046	1586	6399	1160	2629	7614
2004	1167	3018	4692	3280	3564	2993
2005	1230	2886	311	5649	455	2925
2006	4158	769	4277	7550	2371	4750
2007	2795	5106	4255	1955	4536	3026
2008	5667	670	3365	7399	2608	5734
2009	1760	2093	756	3299	226	7233
2010	4319	3419	4648	870	3220	2644
2011	3288	1911	988	29045	1005	7142
2012	8930	1024	5541	1250	2564	4735

Cuadro 21: Distancias de Mahalanobis anuales entre estación real y ficticia conjunta.

taxones y, a su vez, si algún taxón tiende a estar mejor representado que otros. Debemos constatar que tambien se realizó este test comparando las matrices A_j con las restingidas C_j , pero los resultados fueron prácticamente idénticos a los conseguidos con la ficticia conjunta, matriz B , por lo cual solo hemos valorado los de esta comparación.

Test de los rangos con signo de Wilcoxon. La función de R usada para implementar el test es `wilcox.test`¹². En este caso el test implementado es de dos colas con datos emparejados. Los resultados están en el Cuadro 22.

Vemos que hay taxones mejor representados que otros. Destacan Ambrosia y Fagus, que son taxones con muy poca presencia en las 6 estaciones analizadas. Si centramos la atención en los taxones más comunes, anteriormente citados, vemos que su representación solo es buena en alguna estación. Parece evidente que, a pesar de que BTU demostró ser la estación más representativa en conjunto, hemos de ser cautos y admitir que perdemos cierta representación particular de algunos taxones. No obstante, en este análisis también es BTU la que consigue mejores resultados, al tener el mayor número de taxones con pvalor a favor de la hipótesis nula de similares porcentajes anuales comparados con la estación conjunta.

	taxón	BCN	BTU	GIC	LLE	MAN	LLE1
1	alnus						
2	ambrosia	0.870	0.030	0.610	0.830	0.020	0.110
3	artem	0.010	0.090			0.020	
4	betula	0.410	0.520				0.170
5	castan	0.370					0.510
6	chenop						
7	corylus			0.060			
8	cupress		0.330			0.710	
9	fagus	0.030	0.010		0.180		0.030
10	fraxin						0.020
11	olea					0.750	
12	pinus			0.640		0.040	
13	pistac		0.040		0.220		
14	planta		0.130		0.580		
15	platanus			0.010		0.260	
16	poaceae		0.010	0.150			
17	poligo			0.310			0.830
18	quercad		0.010				
19	querper		0.060				
20	ulmus					0.030	
21	urtica	0.550		0.460		0.050	

Cuadro 22: Pvalores ≥ 0.01 tests rangos con signos de Wilcoxon series de taxón.

3. Análisis del calendario polínico.

3.1. Introducción.

Con los taxones no sólo nos interesan los valores del recuento diario del polen, sino también su calendario polínico, que está determinado por la floración de la especie o especies que lo produce. Normalmente los valores más importantes de ese calendario son el comienzo, el final, el día, o primer día, con el máximo recuento del polen del taxón (día pico), y la duración de la floración (día final - día comienzo).

El análisis de ese calendario polínico es tal vez más importante para los taxones que producen alergias. Es al inicio de la floración, y en los días con más concentración de polen del taxón alergógeno, cuando más afecta a las personas con problemas de alergia. Los taxones considerados alergógenos son : Betula, Cupressaceae, Fraxinus, Olea, Platanus, Corylus, Artemisia, Chenopodiaceae amaranthaceae, Poaceae, Plantago, Polygonaceae, y Urticaceae.

En esta sección nos limitaremos a analizar únicamente los taxones alergógenos más presentes en las 6 estaciones con registros desde 1996 hasta 2012, pero también incluiremos los taxones Pinus y Quercus, este último en sus dos especies diferenciadas, caduca y perenne. Describiremos sus calendarios a lo largo de esos años, sus posibles correlaciones, sus tendencias y, por último, los compararemos por estaciones para poder ampliar la comparativa realizada en el punto anterior. También nos inclinamos por usar técnicas no paramétricas al no poder asumir la distribución normal de los datos.

Debemos aclarar que no existe un criterio único para realizar el cálculo de los valores del calendario polínico de un taxón. Nuestra intención, a la hora de usar un criterio concreto, es poder tener una herramienta única que nos permita analizar la relación entre los calendarios polínicos, calculados bajo un mismo criterio, de los distintos taxones y en las estaciones de nuestros datos. Pero esta herramienta ha de considerarse únicamente en ese aspecto. Un biólogo podría usar los resultados de nuestro análisis para valorar la bondad de ajuste del criterio utilizado a la hora de explicar la floración de las especies que producen el taxón.

3.2. Análisis descriptivo.

Para nuestro análisis, hemos optado por el criterio que asigna el día de inicio de la floración, para un taxón concreto, aquel en que la suma acumulada de la concentración diaria registrada hasta ese día supera, por primera vez, el 2,5 % del API total anual acumulado. Para evaluar el final de la floración se toma el primer día con la suma acumulada que supera el 97,5 % del API total¹³. Para evaluar el día pico se toma el día con valores máximos registrados dentro del cómputo anual (el primer día, en caso de

empate). El año considerado es el natural, que va del 1 de enero al 31 de diciembre. Es obvio que estos calendarios sólo se pueden elaborar a posteriori, cuando ya tenemos todo el registro anual.

3.2.1. Gráficos.

En los gráficos siguientes podemos ver el calendario polínico del pasado año 2012 en Bellaterra para 3 taxones. El color rojo correspondería al periodo de floración de la especie en cuestión (95 % del registro acumulado del taxón). La línea punteada corresponde al día pico.

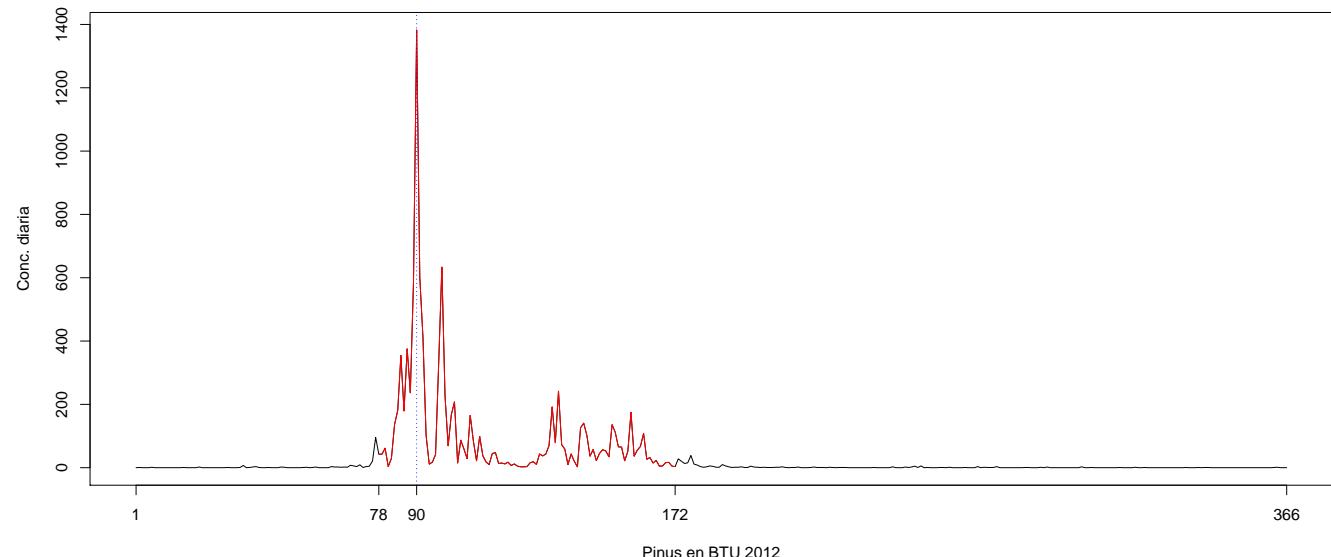


Figura 22: Calendario polínico del taxón *Pinus* en BTU 2012

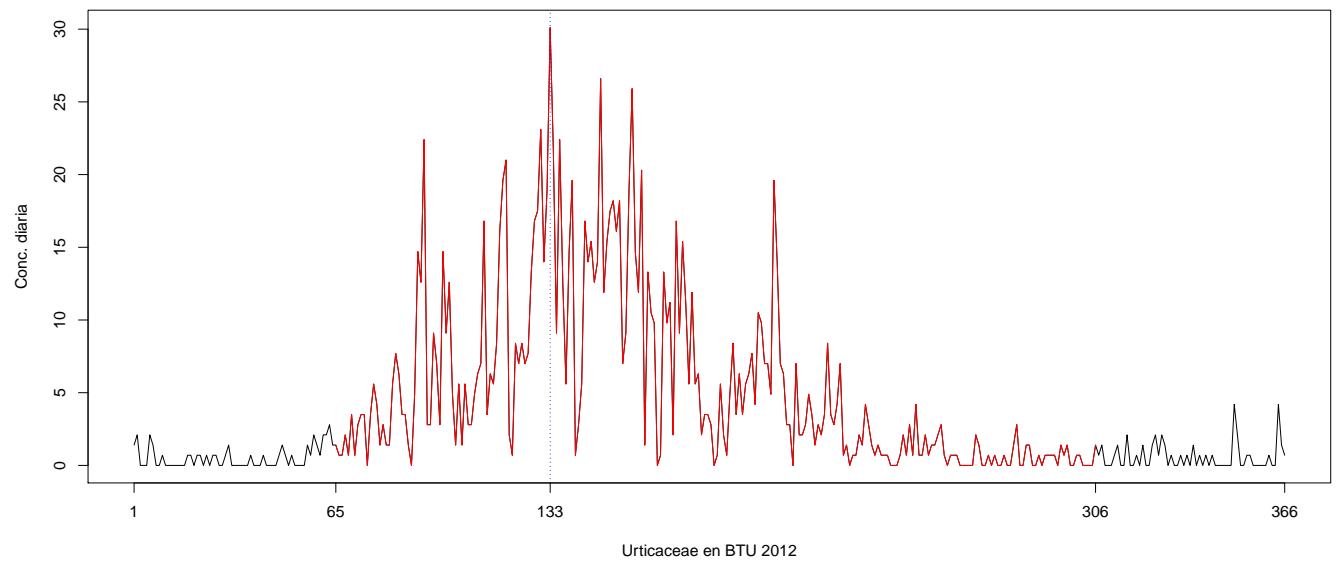


Figura 23: Calendario polínico del taxón Urticaceae en BTU 2012

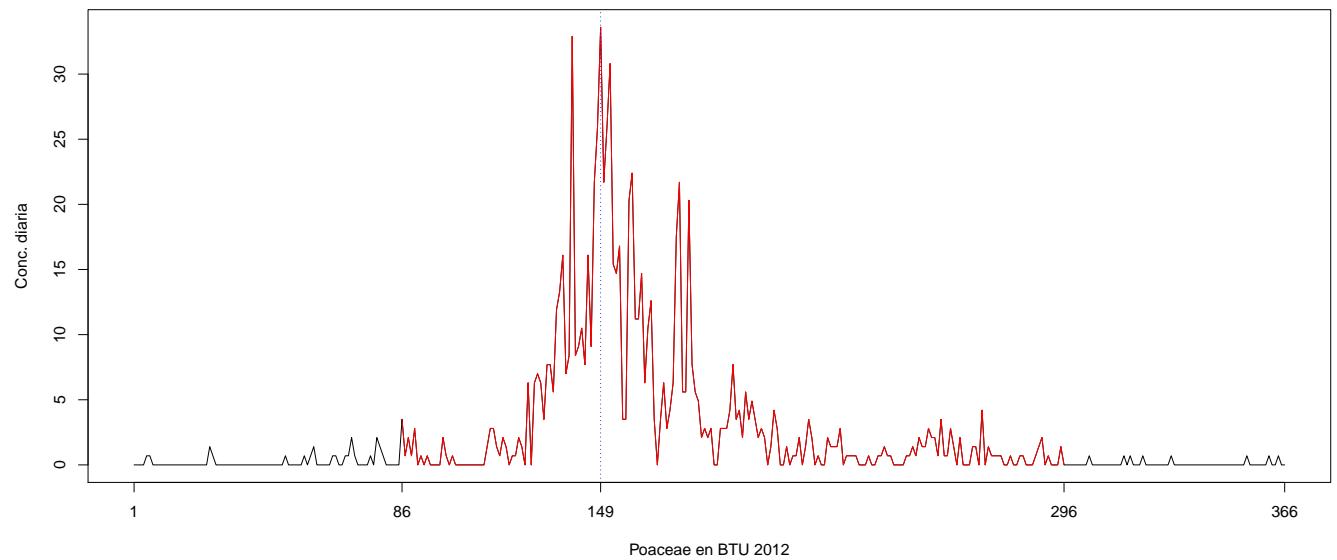


Figura 24: Calendario polínico del taxón Poaceae en BTU 2012

3.2.2. Medianas, rango y MAD. Test de Friedman.

Para analizar los calendarios de los diferentes taxones en las 6 estaciones, con datos 1996:2012, hemos optado por usar las medianas, el rango y el valor de la desviación absoluta de la mediana (MAD), que es la mediana de los valores absolutos de las diferencias entre los datos y su mediana. Todo ello para los cuatro vectores con los valores calculados para las series temporales (comienzo, final, día pico y duración, para cada uno de los 17 años del período), correspondientes a los 10 taxones más importantes en esas estaciones, donde 7 son alérgenos. Nuestro interés radica en comprobar que, considerando un mismo taxón, no parece existir excesiva diferencia entre las diferentes estaciones. Por ello hemos realizado el test de Friedman, test no paramétrico equivalente a una prueba ANOVA para medidas no independientes. En nuestro caso consideramos que el factor bloque son los años, mientras el factor grupo -tratamiento- son las estaciones. Nuestra hipótesis nula es la NO diferencia de calendario polínico de un mismo taxón entre estaciones, a lo largo de los años analizados (según valor testado).

Comienzo de la floración. En el Cuadro 23 podemos ver las medianas para el comienzo de las floraciones. Puntualicemos que ese valor nos indica el día del año natural. Como ejemplo, una mediana de valor 100 correspondería al 5 de abril. En relación a los años bisiestos, hemos tenido en cuenta el valor del 29 de Febrero, cuando existe, pero no hemos considerado necesario realizar ningun ajuste especial para evitar el sesgo en los cálculos, al considerarlo no importante en nuestros análisis.

Aunque la primera impresión es que los datos no parecen muy diferentes entre las estaciones, que es la relación que más nos interesa, en los tests de Friedman para los diferentes taxones sólo hemos podido no rechazar la hipótesis nula con el taxón Cupressaceae ($p\text{valor}=0.052$). En el test de los demás taxones el $p\text{valor}$ ha sido <0.001 . Parece evidente que, excepto para el taxón Cupressaceae, existen grandes diferencias entre los valores de inicio del calendario polínico de los taxones entre algunas estaciones durante los años analizados. No obstante, BTU, LLE Y MAN parecen ser las estaciones con resultados más similares y podemos intuir que es en BCN y TAU donde antes comienzan las floraciones.

En los Cuadros 24 y 25 vemos los valores del MAD y el rango de los vectores con los días de inicio de floración de cada año. Con ello pretendemos analizar la variabilidad del inicio de floración de cada taxón en estos últimos 17 años. Usamos la desviación absoluta de la mediana (MAD) por ser una medida más robusta que el rango, al no estar tan influenciada por los valores extremos. Pero también incluimos el rango para constatar el hecho de que el inicio de la floración ha llegado a variar en más de dos meses en algún taxón (Chenop), y en más de un mes para la mayoría. Respecto a los valores del MAD intuimos que hay taxones con mayor varibilidad en el inicio de floración, principalmente las herbáceas (Chenop, Poaceae), y se podría considerar que algunas especies de árboles (Olea, Platanus y Cupressaceae), tienen una menor variabilidad.

	taxón	bcn	btu	gic	lle	man	tau
1	ChenopodiaceaeAmaranthaceae	100	104	122	121	115	104
2	Cupressaceae	22	23	19	26	26	19
3	Olea	124	128	129	130	134	128
4	Pinus	65	69	71	76	74	66
5	Plantago	95	107	112	110	115	94
6	Platanus	74	75	80	80	81	75
7	PoaceaeTotal	84	90	103	82	79	75
8	QuercusCaducifolis	94	92	103	97	95	96
9	QuercusPerennifolis	113	114	116	111	114	103
10	Urticaceae	40	57	73	68	69	25

Cuadro 23: Medianas del día inicio de floración según taxón y estación -1996:2012.

		bcn	btu	gic	lle	man	tau
ChenopodiaceaeAmaranthaceae		9	9	11	6	17	8
Cupressaceae		2	8	4	4	8	4
Olea		5	4	5	6	7	3
Pinus		6	4	6	7	4	6
Plantago		6	4	4	8	6	4
Platanus		5	6	5	5	6	6
PoaceaeTotal		9	8	7	11	7	13
QuercusCaducifolis		6	4	6	6	9	6
QuercusPerennifolis		9	10	4	3	8	9
Urticaceae		9	8	9	8	5	5

Cuadro 24: MAD de los vectores con el inicio de floración según taxón y estación - 1996:2012.

		bcn	btu	gic	lle	man	tau
ChenopodiaceaeAmaranthaceae		79	63	63	48	69	76
Cupressaceae		27	40	35	38	39	40
Olea		26	39	56	30	39	37
Pinus		40	36	53	47	38	44
Plantago		29	26	44	31	39	42
Platanus		25	27	26	25	30	21
PoaceaeTotal		54	41	55	45	38	49
QuercusCaducifolis		29	31	32	40	35	43
QuercusPerennifolis		40	48	46	49	45	43
Urticaceae		56	41	54	45	49	39

Cuadro 25: Rangos de los vectores con el inicio de floración según taxón y estación -1996:2012.

Final de la floración. Analizando las medianas de los finales de floración para los años considerados, en Cuadro 26, vemos que las diferencias parecen ahora mayores que para el inicio. En los test de Friedman, al igual que anteriormente, sólo no rechazamos la hipótesis nula para el taxón Cupressaceae ($p\text{valor}=0.75$). En el resto de los tests el $p\text{valor}$ ha sido <0.001 . Hay diferencias significativas entre los valores de final de floración de algunas estaciones para la mayoría de taxones.

	taxón	bcn	btu	gic	lle	man	tau
1	ChenopodiaceaeAmaranthaceae	298	300	304	278	297	312
2	Cupressaceae	326	325	327	314	327	326
3	Olea	183	182	184	180	174	178
4	Pinus	181	171	183	187	176	179
5	Plantago	261	252	261	252	241	249
6	Platanus	118	114	119	113	115	120
7	PoaceaeTotal	277	271	258	275	270	300
8	QuercusCaducifolis	169	151	160	166	155	168
9	QuercusPerennifolis	192	174	174	171	172	176
10	Urticaceae	313	294	292	300	307	334

Cuadro 26: Medianas del día final de floración según taxón y estación -1996:2012.

Considerando los rangos y los valores MAD del final de floración constatamos que los taxones que tenían más variabilidad en el inicio de floración son los que parecen tener menor variabilidad en el final (Chenop y Poaceae). Por contra, uno de los taxones con menor variabilidad en el inicio (Cupressaceae), es el que mayor variabilidad tiene en el final. También destaca la variabilidad en BCN y TAU para la mayoría de taxones. En todos estos casos deberíamos considerar la conveniencia del método del 97,5% del API anual acumulado para asignar el final de la floración.

	bcn	btu	gic	lle	man	tau
ChenopodiaceaeAmaranthaceae	9	7	6	4	7	8
Cupressaceae	16	13	27	32	16	21
Olea	9	6	11	8	10	10
Pinus	5	5	10	9	12	7
Plantago	10	15	8	7	15	20
Platanus	13	4	7	3	4	11
PoaceaeTotal	10	7	6	5	12	10
QuercusCaducifolis	14	5	7	11	7	14
QuercusPerennifolis	18	11	12	10	10	12
Urticaceae	8	12	11	10	14	9

Cuadro 27: MAD de los vectores con el final de floración según taxón y estación - 1996:2012.

	bcn	btu	gic	lle	man	tau
Chenopodiaceae	33	37	34	25	31	45
Amaranthaceae	207	213	245	266	222	226
Cupressaceae	50	40	68	68	63	84
Olea	40	43	48	104	37	27
Pinus	64	84	58	54	85	125
Plantago	151	20	41	66	44	156
Platanus	41	39	51	24	44	71
Poaceae	161	45	48	77	33	167
Total	141	99	121	58	126	142
Quercus Caducifolis	90	63	60	40	52	60
Quercus Perennifolis						
Urticaceae						

Cuadro 28: Rango de los vectores con el final de floración según taxón y estación - 1996:2012.

Día pico. Analizando las medianas, podemos observar valores más cercanos, que en el inicio y fin, entre las diferentes estaciones, para un mismo taxón. Los test de Friedman no rechazan la hipótesis nula para los taxones Cupressaceae ($p_v=0.98$), Poaceae ($p_v=0.05$) y Chenop (pv=0.22). Para el resto de tests obtenemos pvalores <0.001 .

	taxón	bcn	btu	gic	lle	man	tau
1	Chenopodiaceae	240	240	252	240	248	241
2	Amaranthaceae	51	55	58	60	56	55
3	Cupressaceae	142	143	144	150	150	141
4	Olea	90	81	134	99	100	76
5	Pinus	148	142	158	147	152	136
6	Plantago	80	89	92	87	92	82
7	Platanus	148	153	152	142	150	151
8	Poaceae	130	107	130	127	124	122
9	Total	140	141	145	137	139	135
10	Quercus Caducifolis	133	153	154	143	147	134
	Quercus Perennifolis						
	Urticaceae						

Cuadro 29: Medianas del día pico de floración según taxón y estación -1996:2012.

Por contra, la variabilidad de los vectores con el valor del día pico es grande. Destacan los valores de los taxones Chenop y Urticaceae, tanto en rangos como en los valores del MAD.

	bcn	btu	gic	lle	man	tau
Chenopodiaceae	15	15	12	2	8	32
Amaranthaceae						
Cupressaceae	10	7	5	6	11	8
Olea	6	5	7	5	5	3
Pinus	16	5	16	7	9	9
Plantago	8	6	10	6	8	9
Platanus	4	5	6	4	5	6
Poaceae	7	4	6	7	3	8
Total						
Quercus	14	6	11	11	10	9
Caducifolis						
Quercus	6	7	5	5	6	11
Perennifolis						
Urticaceae	33	20	17	29	15	13

Cuadro 30: MAD de los vectores con el día pico de la floración según taxón y estación 1996:2012.

	bcn	btu	gic	lle	man	tau
Chenopodiaceae	131	144	155	114	40	149
Amaranthaceae						
Cupressaceae	50	42	46	48	39	59
Olea	40	37	47	40	29	34
Pinus	91	38	75	98	80	107
Plantago	77	40	53	46	67	52
Platanus	29	29	30	20	28	25
Poaceae	51	27	31	123	39	33
Total						
Quercus	49	59	66	54	50	46
Caducifolis						
Quercus	40	44	57	43	59	55
Perennifolis						
Urticaceae	123	85	119	146	77	77

Cuadro 31: Rango de los vectores con el día pico de la floración según taxón y estación 1996:2012.

Conclusiones. Hemos visto que existe una gran variabilidad en los diferentes valores del calendario polínico para la gran mayoría de taxones y estaciones. No hemos analizado la duración por ser una función de los valores del inicio y fin, pero es lógico que su variabilidad también sea grande, como lo es de hecho. No parece existir similitud entre todas las estaciones para un mismo taxón. Las particularidades que cada una de ellas ya tenía en los valores anuales de los diferentes API, vuelven a aparecer en los valores de los diferentes calendarios polínicos, aunque parece que sí existe cierta similitud entre algunas de ellas.

3.3. Correlaciones.

Entre API y duración de la floración. Podríamos pensar que una duración mayor de la floración podría estar asociada a un mayor registro anual, sin embargo veremos que esto no es cierto.

Para analizar este punto hemos estudiado la correlación de Spearman entre el vector con la duración del calendario polínico anual, en días, de cada taxón (1996:2012), con el vector correspondiente al API anual registrado para ese taxón. En el Cuadro 32 vemos esos coeficientes de correlación de Spearman, ρ . La gran mayoría son negativos, lo que nos indica que en los años con una mayor cantidad de API registrado no ha habido una mayor duración de la floración, sino más bien lo contrario. Podríamos pensar que un mayor registro del API se debe a floraciones “explosivas”, producto de ciertas condiciones clímaticas, más que a floraciones largas en el tiempo. En cualquier caso, podemos destacar el taxón *Quercus perenne*, con relación significativa en todas las estaciones, y la estación de Lleida, con mayor número de relaciones significativas.

	taxón	bcn	btu	gic	lle	man	tau
1	ChenopodiaceaeAmaranthaceae	-0.58*	-0.13	-0.54*	-0.44	-0.62**	-0.26
2	Cupressaceae	-0.27	-0.20	-0.47	-0.72**	-0.30	-0.37
3	Olea	-0.64**	-0.44	-0.30	-0.80**	-0.86**	-0.44
4	Pinus	-0.12	-0.23	-0.47	-0.53*	-0.57*	-0.31
5	Plantago	-0.11	0.08	0.10	-0.41	-0.47	0.06
6	Platanus	-0.42	-0.29	-0.01	-0.64**	-0.61**	-0.60*
7	PoaceaeTotal	-0.14	-0.01	0.00	-0.61**	-0.32	-0.14
8	QuercusCaducifolis	-0.53*	-0.45	-0.48*	-0.13	-0.36	-0.10
9	QuercusPerennifolis	-0.66**	-0.57*	-0.73**	-0.65**	-0.80**	-0.48*
10	Urticaceae	-0.50	-0.14	-0.30	-0.32	-0.22	-0.66**

Cuadro 32: Coef. de correlación de Spearman entre API anual y duración de la floración.

Entre estaciones para los valores del calendario polínico. Ahora analizaremos si existe relación para los valores del calendario polínico de un taxón, entre las diferentes estaciones. Para ello calculamos el coeficiente de correlación de Spearman entre los 6 vectores con los 17 datos para cada uno de los 10 taxones y de los 4 valores considerados (inicio, fin, día pico y duración). En este caso, dado el gran número de correlaciones analizadas, nos limitaremos a comentar los aspectos que consideramos más relevantes de esas relaciones.

Los taxones donde la relación es más significativa son: *Pinus*, *Platanus*, *Olea* y *Quercus perenne*. Los que menor relación presentan son *Poaceae*, *Chenop* y *Cupressaceae*. Las relaciones más significativas se producen en los inicios, mientras que en el día pico apenas existen relaciones significativas. Esto nos indica que, a pesar de no tener valores similares

a lo largo de los años, como mostró el test de Friedman, sí hay taxones con una tendencia muy similar. También destaca el taxón Cupressaceae, que, a pesar de ser el único taxón con valores similares en las 6 estaciones, según el test de Friedman, no en todas las estaciones sus valores están correlacionados.

En los gráficos siguientes podemos ver la similitud entre los días de inicio de los taxones Olea y Quercus perenne en todas las estaciones a lo largo de los años analizados 1996:2012. En ambos taxones todos los coeficientes de correlación de los vectores con los valores del inicio entre estaciones son muy significativos. Por contra, vemos el mismo gráfico para el taxón Poaceae, que apenas tiene correlaciones significativas entre estaciones.

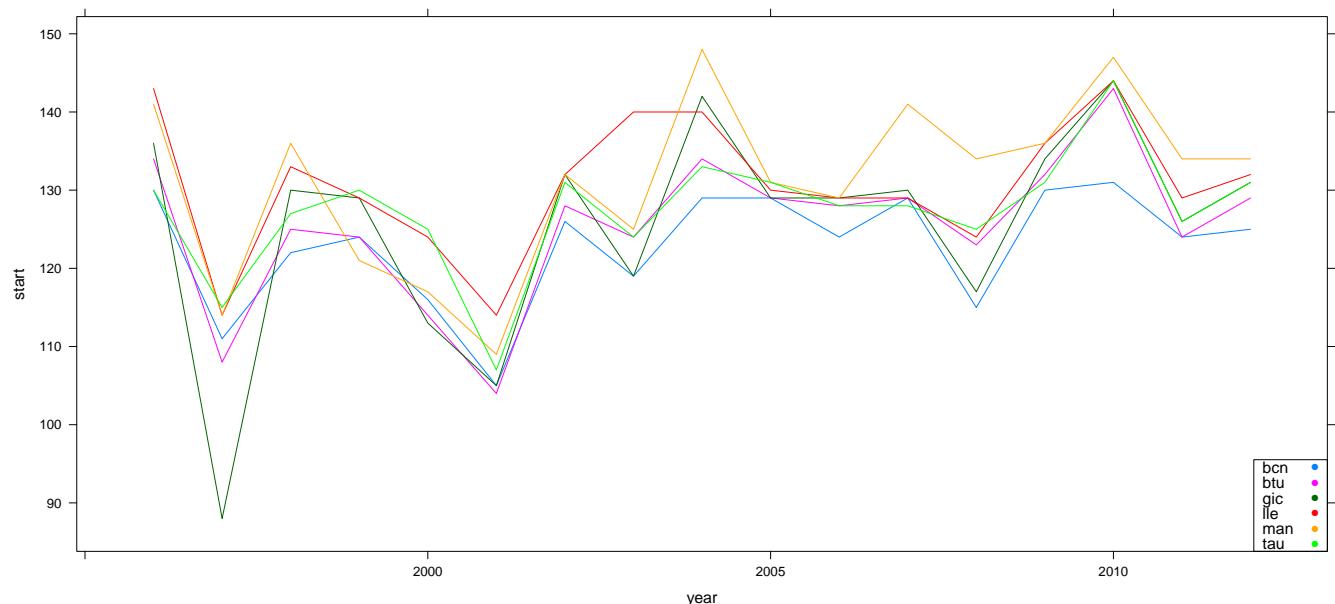


Figura 25: Días inicio de floración del taxón Olea según estación 1996:2012.

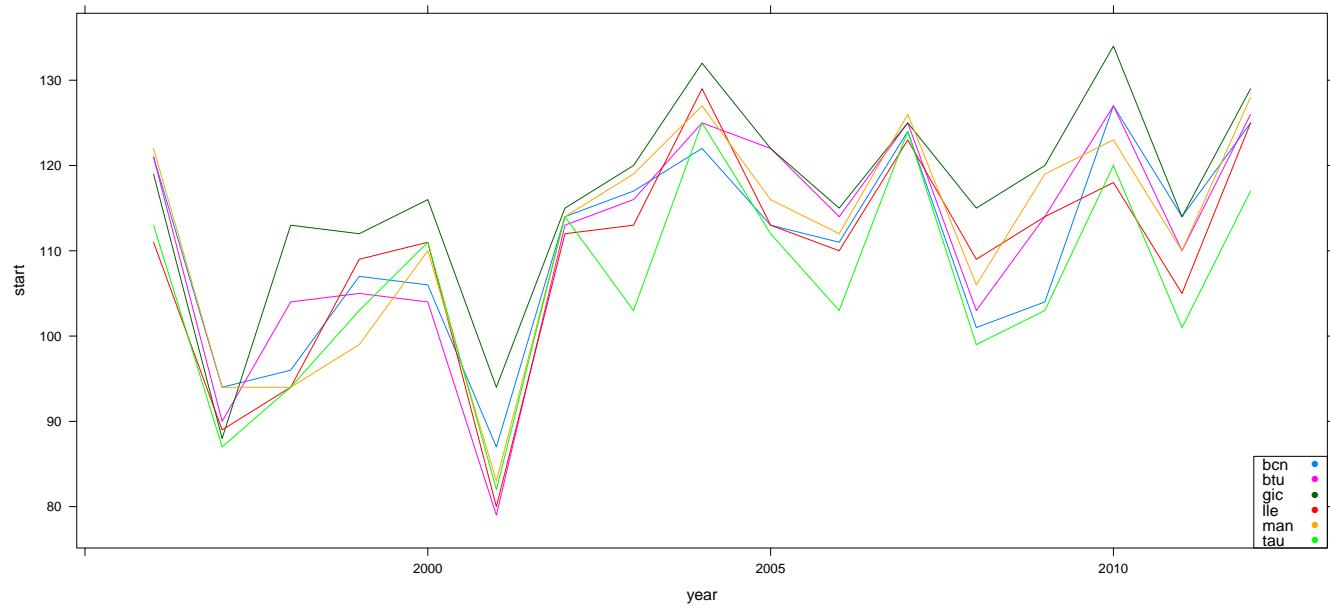


Figura 26: Días inicio de floración del taxón *Quercusperenne* según estación 1996:2012.

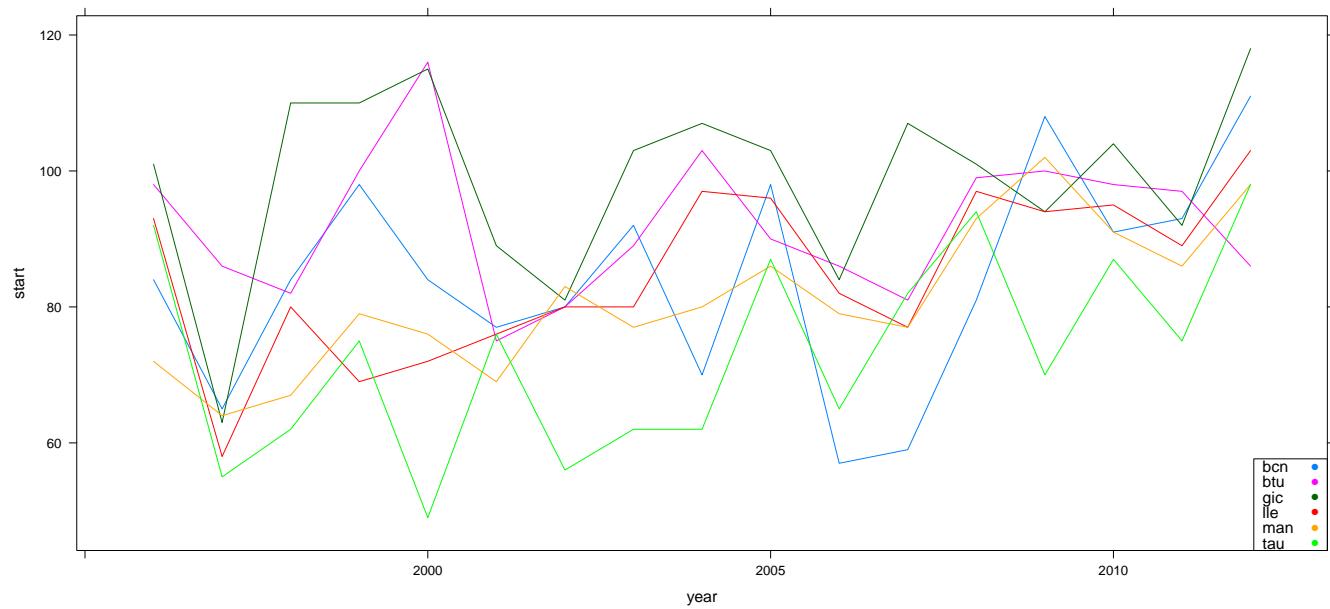


Figura 27: Días inicio de floración del taxón *Poaceae* según estación 1996:2012.

3.4. Tendencias.

Para analizar las tendencias de los valores del calendario polínico, en estos 17 últimos años, de los diferentes taxones en las diferentes estaciones, hemos optado por usar el test de MannKendall estandarizado, ya utilizado en otro punto anterior. En el Cuadro 33 podemos ver las tendencias significativas resultantes y el signo de la estimación de la pendiente, en su caso.

Encontramos tendencias significativas para los valores de inicio y duración de la floración. Todas estas tendencias son de signo positivo para el inicio, pero de signo negativo para la duración. Podemos afirmar que algunos taxones, en algunas estaciones concretas, presentan una tendencia significativa, en estos últimos años, de atrasar el inicio de la floración y reducir su duración. Deberíamos estudiar la posible relación de este resultado con el cambio climático, pero eso requeriría un estudio más completo.

	taxón	start	pvalor.tend.	length	pvalor.tend.1
1	poaceae	lle	<0.01** (+)	lle	<0.05 *(-)
2	poaceae	man	<0.01** (+)	man	<0.01** (-)
3	cupressaceae			gic	<0.05 *(-)
4	olea			btu	<0.05 *(-)
5	olea			gic	<0.05 *(-)
6	olea			man	<0.01** (-)
7	platanus	man	<0.05*(+)	man	<0.01** (-)
8	plantago			btu	<0.05 *(-)
9	plantago	gic	<0.05*(+)	gic	<0.01** (-)
10	plantago	man	<0.05*(+)	man	<0.05 *(-)
11	pinus	bcn	<0.05*(+)		
12	pinus			gic	<0.05 *(-)
13	pinus	lle	<0.05*(+)	lle	<0.05 *(-)
14	pinus	man	<0.05*(+)	man	<0.05 *(-)
15	quercus perenne	btu	<0.05*(+)	btu	<0.05 *(-)
16	quercus perenne	gic	<0.05*(+)	gic	<0.01** (-)
17	quercus perenne	lle	<0.05*(+)	lle	<0.01** (-)
18	quercus perenne	man	<0.05*(+)	man	<0.01** (-)
19	quercus perenne			tau	<0.05 *(-)

Cuadro 33: Tendencias significativas del inicio y duración de la floración.

En el gráfico siguiente podemos ver las series de inicio y duración del taxón Poaceae en Manresa. Se aprecia claramente la tendencia positiva del inicio y la tendencia negativa de la duración de la floración.

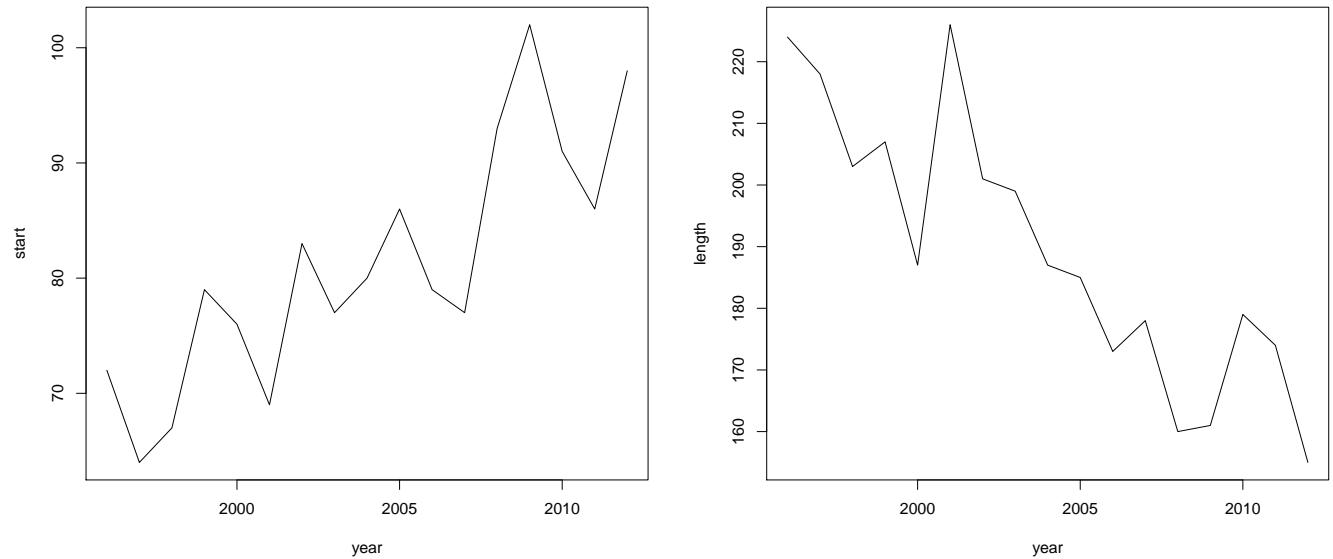


Figura 28: Inicio y duración del calendario polínico del taxón Poaceae en Manresa 1996:2012.

3.5. Comparativa de estaciones.

De la misma manera que hemos analizado, en un punto anterior, la estación más representativa para la distribución de la proporción del API de cada taxón de toda Catalunya, también podemos analizar la representatividad de las diferentes estaciones para un único hipotético calendario polínico, de cada taxón, en el conjunto del territorio.

Nuestro objetivo de encontrar la estación más representativa, de ese hipotético calendario para toda Catalunya, necesita primero definir cual sería ese calendario. Hemos optado por definirlo como el correspondiente al de la estación ficticia conjunta ya creada para análisis anteriores. Es decir, sumarizamos los valores de concentración diaria de las 6 estaciones para cada uno de los 10 taxones analizados. Sobre estos datos diarios calculamos los valores del calendario polínico correspondiente a cada año y taxón con el mismo criterio usado anteriormente para las estaciones reales.

Para comparar las estaciones hemos calculado algunas distancias entre la matriz conjunta y la real de cada estación. Todo ello para los 10 taxones que analizamos en este apartado. En los siguientes puntos se explican los cálculos de esas distancias y los resultados, ya que hemos realizado un análisis conjunto y otro por valor (inicio, final, etc.). En el Cuadro 34 podemos ver una esas matrices, correspondiente a los valores calculados para el taxón Olea en BTU.

	start	end	length	peakday
1996	134	184	50	148
1997	108	163	55	118
1998	125	194	69	151
1999	124	188	64	138
2000	114	203	89	151
2001	104	182	78	133
2002	128	188	60	142
2003	124	188	64	125
2004	134	185	51	155
2005	129	177	48	143
2006	128	164	36	140
2007	129	179	50	140
2008	123	184	61	144
2009	132	175	43	147
2010	143	175	32	150
2011	124	168	44	136
2012	129	173	44	148

Cuadro 34: Matriz con los valores calendario polínico del taxón Olea en BTU

Entre días de inicio de floración. Queremos conocer las diferencias entre el inicio de floración de cada taxón, en cada estación real, comparada con la floración conjunta. En este caso calculamos la diferencia entre los vectores que indican ese valor en la estación real y el vector correspondiente en la estación ficticia conjunta, que corresponden a las primeras columnas de las matrices comparadas. Sobre cada vector diferencia calculamos su norma euclídea ℓ_2 . Ese valor es el que aparece en el Cuadro 35, para cada taxón y estación. Si comparamos estos nuevos vectores columna, que contienen la norma euclídea, usando la norma ℓ_1 , calculada en la última fila del cuadro, vemos que BTU es la estación con menores distancias. Comparando por taxones vemos que las herbáceas tienden a tener inicios de floración más diferenciados entre estaciones.

	taxón	bcn	btu	gic	lle	man	tau
1	ChenopodiaceaeAmaranthaceae	95	53	72	71	69	75
2	Cupressaceae	23	21	29	34	25	25
3	Olea	22	9	23	21	24	14
4	Pinus	27	11	28	45	22	34
5	Plantago	43	13	30	27	61	49
6	Platanus	16	8	16	11	19	16
7	PoaceaeTotal	42	49	68	28	32	67
8	QuercusCaducifolis	30	24	26	19	22	27
9	QuercusPerennifolis	15	13	27	14	14	24
10	Urticaceae	65	27	75	56	65	126
11	Suma vectores columna	378	228	394	326	353	457

Cuadro 35: Distancias euclídeas de los vectores diferencia día inicio.

Entre días pico de floración. En este punto realizamos el mismo cálculo anterior, pero en este caso con los valores del día pico (columna 4 de cada matriz). Como podemos ver en el Cuadro 36 los mejores resultados son para MAN.

Entre días final de floración. Seguimos efectuando el mismo análisis, pero en este caso para los valores del día final de la floración (columna 2 de las matrices). BTU y MAN son las estaciones con los mejores resultados. Ver Cuadro 37

Entre duración de floración. Realizamos el mismo análisis para los valores de la duración de la floración (columna 3 de las matrices). En este caso vuelve a ser BTU la estación con mejores resultados. Ver Cuadro 38

	taxón	bcn	btu	gic	lle	man	tau
1	ChenopodiaceaeAmaranthaceae	180	164	162	59	114	154
2	Cupressaceae	29	38	30	41	49	39
3	Olea	23	25	26	39	28	28
4	Pinus	116	107	120	89	77	108
5	Plantago	43	40	63	40	61	67
6	Platanus	28	14	23	17	27	22
7	PoaceaeTotal	36	26	24	98	35	33
8	QuercusCaducifolis	53	84	43	45	38	35
9	QuercusPerennifolis	19	25	29	42	29	40
10	Urticaceae	134	88	114	135	70	94
11	Suma vectores columna	661	611	634	605	528	620

Cuadro 36: Distancias euclídeas de los vectores diferencia día pico.

	taxón	bcn	btu	gic	lle	man	tau
1	ChenopodiaceaeAmaranthaceae	25	34	44	84	21	67
2	Cupressaceae	138	124	183	259	131	157
3	Olea	28	33	54	71	57	75
4	Pinus	24	49	34	90	27	37
5	Plantago	57	54	40	54	82	100
6	Platanus	238	90	66	80	88	136
7	PoaceaeTotal	45	29	79	40	40	101
8	QuercusCaducifolis	139	90	50	85	68	113
9	QuercusPerennifolis	125	61	91	100	76	132
10	Urticaceae	55	78	68	49	36	128
11	Suma vectores columna	874	642	709	912	626	1046

Cuadro 37: Distancias euclídeas de los vectores diferencia día final.

	taxón	bcn	btu	gic	lle	man	tau
1	ChenopodiaceaeAmaranthaceae	98	63	73	147	75	107
2	Cupressaceae	147	124	185	273	127	157
3	Olea	41	32	52	74	72	84
4	Pinus	38	55	33	97	35	48
5	Plantago	86	56	42	70	127	108
6	Platanus	249	87	73	85	98	140
7	PoaceaeTotal	64	66	133	52	52	149
8	QuercusCaducifolis	149	84	60	81	71	125
9	QuercusPerennifolis	129	65	100	99	74	145
10	Urticaceae	102	101	125	84	59	240
11	Suma vectores columna	1103	733	876	1062	790	1303

Cuadro 38: Distancias euclídeas de los vectores diferencia duración de la floración.

Distancia total. En este caso calcularemos la distancia total entre las matrices, utilizando la norma euclídea de cada matriz diferencia. Para ello calculamos la raíz cuadrada de la suma del cuadrado de todos los elementos de la matriz diferencia, que recordemos es la matriz de la estación real menos la matriz de la estación conjunta, todo ello para cada taxón. En el Cuadro 39 podemos ver los resultados de estos cálculos. En este caso, como es lógico, son BTU y MAN las estaciones con mejores resultados.

	taxón	bcn	btu	gic	lle	man	tau
1	ChenopodiaceaeAmaranthaceae	227	186	196	193	154	213
2	Cupressaceae	205	180	264	380	191	226
3	Olea	58	53	83	112	99	117
4	Pinus	127	130	132	166	91	129
5	Plantago	120	89	91	101	173	169
6	Platanus	346	125	102	119	135	197
7	PoaceaeTotal	97	91	170	121	81	195
8	QuercusCaducifolis	213	151	93	127	108	174
9	QuercusPerennifolis	181	94	141	148	111	201
10	Urticaceae	188	158	197	176	118	314
11	Suma vectores columna	1762	1257	1469	1643	1261	1935

Cuadro 39: Distancias euclídeas de las matrices diferencia de todos los valores.

Conclusiones. Bellaterra vuelve a ser la estación en conseguir los mejores resultados, en conjunto, en las diferentes comparaciones de los diferentes valores del calendario polínico de los taxones analizados, aunque su similitud con Manresa, en este aspecto, es evidente. En este caso también podría ser la estación idónea para representar, con menor desajuste, un calendario polínico conjunto para toda Catalunya, exceptuando de nuevo a Vielha y Tortosa.

4. Conclusiones.

Consideramos que este análisis de los datos aerobiológicos de Catalunya ha encontrado aspectos que merecerían un estudio más profundo. El hecho de no tener conocimientos sobre la biología de las especies y sus taxones nos hace ser cautos a la hora de sacar conclusiones que requerirían ser refrendadas por un botánico. Pero sí que podemos considerar los aspectos más significativos, estadísticamente, de nuestro análisis.

La primera conclusión hace referencia a que, pese a la gran variabilidad de nuestros datos, hemos encontrado relaciones importantes entre ellos. La polinización no parece algo trivial. Existen muchos factores que influyen no sólo en la producción de una mayor cantidad de polen, sino también en la recogida de éste por los captadores. Parecería lógico pensar que en un territorio no excesivamente grande como Catalunya, no deberían existir grandes diferencias en los registros de los diferentes taxones, pero existen. Destaca el caso de Vielha, con un registro muy diferenciado en el tipo de taxón recogido. Pero entre las 6 estaciones más homegéneas, también existen particularidades importantes, principalmente por la presencia de algunos taxones con un registro excepcional. No obstante hemos comprobado que la estación de Bellaterra podría ser un buen indicador del comportamiento conjunto de esas 6 estaciones. No tanto para la cantidad de granos de polen recogidos de cada taxón, pero sí para la distribución de los taxones dentro del registro total. Tal vez por su posición central en el territorio, o porque no tiene ciertos factores que hacen a otras estaciones tener más particularidades, como la cercanía al mar en el caso de Barcelona y Tarragona, o el clima más riguroso de LLeida. De cualquier modo parece haber quedado demostrado que Bellaterra debería ser la estación a potenciar si queremos estudiar el registro conjunto. Si nos interesan las particularidades, no debemos descartar ninguna estación y parece que su distribución actual en el territorio es óptima para registrar este aspecto.

Otra conclusión de este análisis hace referencia a las tendencias productivas de algunos taxones. Tal vez la poca cantidad de datos de las series temporales no sea la adecuada para sacar conclusiones definitivas al respecto, pero podemos considerar que los resultados obtenidos son suficientemente importantes para ser destacados. Hemos constatado que algunas especies han incrementado su polinización en los últimos años, y que ese incremento no ha sido debido ni a un aumento de la especie en el hábitat (no existe una tendencia positiva de su porcentaje en el registro total), ni a una mayor duración de la floración, sino más bien al contrario. Si este hecho de polinizar más en menos días se debe al cambio climático es un tema en el que nosotros no podemos entrar, pero que muchos trabajos referenciados así lo consideran. Por contra, en algunos taxones la tendencia decreciente, en la cantidad de polinización, sí parece estar asociada a una disminución de la especie en el territorio. Todo ello es importante para prevenir un posible aumento de los casos de alergia y para analizar los cambios de vegetación que se producen en el entorno. En este último aspecto creemos haber comprobado la influencia de la actuación humana en el paisaje, y consecuentemente en el registro polínico, aumentando las

especies ornamentales y disminuyendo las plantas ruderales.

Mención aparte requiere el calendario polínico. Nos ha sorprendido la diferencia de sus valores entre estaciones, considerando un mismo taxón, y el rango de estos valores. La variabilidad en los datos que proceden de la naturaleza es, hasta cierto punto, abrumadora. Estaciones con cierta proximidad geográfica no tienen un registro similar, cuando a priori parecería lógico que así fuera, aunque hemos constatado valores similares para algún taxón y tendencias similares para algunos de ellos.

Por otra parte queremos ser críticos con nuestro análisis. No hemos partido de hipótesis previas concretas. Hemos explorado los datos conjuntamente y no hemos analizado algunas particularidades. Hemos encontrado muchos caminos pero no hemos visto donde acaban la mayoría de ellos. Usando una metáfora podríamos decir que hemos estudiado el bosque pero no sus árboles. Por este motivo las conclusiones han de ser consideradas en conjunto. Un análisis específico de algún taxón podría dar resultados que no coincidan plenamente con los expuestos.

Referencias

¹<http://lap.uab.cat/aerobiologia/es/>

²<http://lap.uab.cat/aerobiologia/es/aboutus>

³http://www.uab.es/servlet/Satellite/videos/reproduccio-1192707516892.html?param1=50mitjanshistoricparam2=10Televisioparam5=1url_video=1345655425920

⁴Fernández-Llamazares, A., Belmonte, J., Delgado, R., De Linares, C., “A statistical approach to bioclimatic trend detection in the airborne pollen records of Catalonia (NE Spain)”, <http://link.springer.com/article/10.1007%2Fs00484-013-0632-4#page-1> International Journal of Biometeorology (2013).

⁵De Linares, C., Belmonte, J., Delgado, R., Días de la Guardia, C., Cariñanos, P., Alonso-Pérez, S., Cuevas, E., Tormo, R., “Pollination patterns and trends of airbone Pinus pollen (in Spain)”, preprint, 2013.

⁶<http://www.r-project.org/>

⁷C. Díaz de la Guardia, F. Alba, C. de Linares, D. Nieto-Lugilde, J. López Caballero, “Aerobiological and allergenic analysis of Cupressaceae pollen in Granada(Southern Spain)”

⁸Ziello C et al. (2012) Changes to Airborne Pollen Counts across Europe. PLoS ONE. doi: 10.1371/journal.pone.0034076.

⁹Trigo, M. et al, “A statistical approach to comparing the results from different aerobiological stations”, <http://www.uco.es/rea/publicaciones/publicaciones.html>

¹⁰<http://stat.ethz.ch/R-manual/R-patched/library/stats/html/mahalanobis.html>

¹¹<http://stat.ethz.ch/R-manual/R-patched/library/stats/html/chisq.test.html>

¹²<http://stat.ethz.ch/R-manual/R-patched/library/stats/html/wilcox.test.html>

¹³Docampo,S et al, “Risk of pollen allergy in Nerja (southern Spain): a pollen calendar”

<http://www.uco.es/rea/publicaciones/publicaciones.html>

Anexos

Test de bondad de ajuste (ji-cuadrado) de Pearson.[1]

El objetivo de este test no paramétrico es verificar si una muestra procede de una población con una determinada distribución de probabilidad, indicando en qué medida las diferencias existentes entre ambas se deben al azar. La hipótesis nula de este contraste considera que la muestra en cuestión sigue la distribución de probabilidad teórica.

Para medir la discrepancia entre la muestra observada y la distribución teórica calculamos el estadístico:

$$\chi^2 = \sum_{i=1}^n \frac{(O_i - E_i)^2}{E_i}$$

donde O_i son las frecuencias observadas y E_i las esperadas bajo la hipótesis nula. Siendo $i = 1, \dots, n$ el número de categorías mutuamente excluyentes.

Bajo la hipótesis nula, este estadístico sigue una distribución aproximada ji-cuadrado, si n es grande, con $n - 1$ grados de libertad, en caso de estar especificada la distribución teórica.

Dado que la distribución del estadístico es asintótica se utiliza, como regla de aproximación aceptable, la condición que todas las frecuencias esperadas sean mayores o iguales a 5.

Test de Friedman.[2]

El test de Friedman es una prueba no paramétrica equivalente a la prueba ANOVA para medidas repetidas. Está diseñada para comprobar si n elementos muestran diferencias cuando se les somete a k tratamientos distintos. La hipótesis nula es la igualdad de respuesta de los k tratamientos en los n elementos. La alternativa es que, al menos, alguno de los tratamientos tiene una respuesta diferente.

La información muestral se dispone en una matriz $\{x_{ij}\}_{n \times k}$ donde n representa los bloques y k representa los tratamientos. En nuestro caso particular n serían los años (17) y k serían las diferentes estaciones (6). Esta matriz original se transforma en otra donde se han ordenado las observaciones x_{ij} , dentro de cada bloque i asignándoles su correspondiente rango R_{ij} . La suma de estos rangos será constante para cada bloque e igual a $\frac{k(k+1)}{2}$, mientras que la suma de los rangos de cada tratamiento será diferente. El estadístico del contraste es:

$$\chi^2 = \frac{12}{nk(k+1)} \sum_{j=1}^k \bar{R}_{\cdot j}^2 - 3n(k+1)$$

donde n es el número de bloques, k el de tratamientos, y

$$\bar{R}_{\cdot j} = \sum_{i=1}^n R_{ij}$$

Asintóticamente puede considerarse que el estadístico tiene una distribución chi-cuadrado con $k - 1$ grados de libertad.

Test de Wilcoxon de rangos con signo.[3]

En nuestro análisis usamos esta prueba no paramétrica para contrastar la hipótesis nula de que las diferencias de n pares de observaciones $z_i = x_i - y_i$, de dos muestras relacionadas, tienen una distribución simétrica de mediana igual a 0, es decir, que no existen diferencias entre las poblaciones de las que provienen las muestras. Se suele utilizar como alternativa al test t de Student cuando no podemos asumir la distibución normal.

Los valores diferencia z_i se consideran independientes y, para verificar la hipótesis nula, se ordenan según su valor absoluto y se les asigna un rango R_i . El estadístico de la prueba es la suma de esos rangos, pero sólo en el caso de que el signo de la diferencia z_i sea positivo.

$$W^+ = \sum_{z_i > 0} R_i$$

La distribución del estadístico puede consultarse en unas tablas específicas para determinar si se acepta o no la hipótesis nula, pero para valores muestrales de $n > 15$ es posible recurrir a su distribución asintótica:

$$W^+ \xrightarrow{d} N\left(\frac{n(n+1)}{4}; \sqrt{\frac{n(n+1)(2n+1)}{24}}\right)$$

Test de Mann-Kendall.[4]

Este test no paramétrico evalua la tendencia monótona, ascendente o descendente, de una serie temporal. Se aplica, principalmente, en estudios de identificación de tendencias

en series de datos ambientales. Esta basado en el estadístico S de Mann Kendall

$$S = \sum_{k=1}^{n-1} \sum_{j=k+1}^n signo(x_j - x_k)$$

donde x_1, x_2, \dots, x_n son los valores de la serie.

La función $signo(x_j - x_k)$ con $j > k$ se define como:

$$signo(x_j - x_k) = \begin{cases} 1 & \text{si } (x_j - x_k) > 0 \\ 0 & \text{si } (x_j - x_k) = 0 \\ -1 & \text{si } (x_j - x_k) < 0 \end{cases}$$

Si el valor de este estadístico es un número positivo, los valores de la serie tomados más tarde tienden a ser más grandes que los tomados antes (tendencia ascendente), mientras que un valor negativo indicaría lo contrario. Para valorar la significación del estadístico existen unas tablas, pero si $n > 10$ y no hay empates, se puede aproximar por medio de una distribución normal con media 0 y variancia

$$\sigma^2 = \frac{1}{18}(n(n-1)(2n+5))$$

El estadístico S se puede estandarizar con la siguiente función:

$$Z_{MK} = \begin{cases} \frac{S-1}{\sigma} & \text{si } S > 0 \\ 0 & \text{si } S = 0 \\ \frac{S+1}{\sigma} & \text{si } S < 0 \end{cases}$$

Con este nuevo estadístico Z_{MK} y su distribución normal estandar, aproximada, se calcula la significación correspondiente, según se considere el test de una o dos colas, bajo la hipótesis nula de que no existe tendencia.

Test rho de Spearman.[5]

Este contraste no paramétrico valora la asociación entre dos variables aleatorias continuas. Pero también se usa para valorar la tendencia monótona de una serie temporal. En el caso de valorar la correlación de dos variables se desea verificar la hipótesis nula de no existencia de asociación entre ellas.

Para el cálculo del estadístico se ordenan de menor a mayor las observaciones de ambas variables y se asignan los rangos $R_i(x)$ y $R_i(y)$. Se consideran las diferencias $d_i = R_i(x) - R_i(y)$ y se calcula el estadístico

$$\rho = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n^3 - n}$$

Cuando $n > 10$ puede usarse la distribución asintótica calculando el estadístico

$$t = \frac{\rho\sqrt{n-2}}{\sqrt{(1-\rho^2)}}$$

que tiene, aproximadamente, una distribución t de Student con $n - 2$ grados de libertad.

En el caso de valorar la tendencia monótona de una serie se usan los mismos estadísticos anteriores, pero las diferencias d_i son, en este caso, $d_i = R_i(x) - i$, donde $R_i(x)$ equivale al rango de la observación x_i .

Estimador Theil-Sen.[6]

Este método no paramétrico estima la magnitud de una tendencia observada como significativa, previamente, en el test de Mann-Kendall o el de la rho de Spearman. En nuestro caso disponemos de unas observaciones y_1, \dots, y_n , de manera que el estimador es la mediana de los valores $(y_j - y_i)/(j - i)$ donde $i = 1, \dots, n$, $j = i + 1, \dots, n$.

Tests de normalidad.

Shapiro-Wilk.[7]

Este contraste se usa para verificar que una muestra de una variable aleatoria procede de una población con distribución normal. Se considera muy útil para muestras con pocas observaciones $n \leq 30$. La hipótesis nula es que la población tiene distribución normal. El estadístico de contraste es:

$$W = \frac{\sum_{i=1}^k a_{(n-i+1)}(x_{(n-i+1)} - x_{(i)})}{SD\sqrt{n-1}}$$

donde n es el número total de observaciones de la muestra, SD es su desviación estándar, $x(1), \dots, x(n)$ las observaciones ordenadas de menor a mayor valor, k el mayor número entero menor o igual a $n/2$ y $a_{(n-i+1)}$ son unos coeficientes que se encuentran tabulados en función de n .

No se rechaza la hipótesis nula, según el nivel de significación propuesto, si el valor del estadístico W es mayor que el valor crítico tabulado según n y el nivel de significación.

Lilliefors.[8]

Este contraste de normalidad es una adaptación del test de Kolmogorov-Smirnov para el caso en que desconocemos la media y la variancia de la distribución normal supuesta.

La hipótesis nula indica que la muestra testada proviene de una población que sigue una distribución normal. El estadístico de contraste es:

$$D = \max_{1 \leq i \leq n} \{ \max \{ |p(i) - \frac{i}{n}|, |p(i) - \frac{i-1}{n}| \} \}$$

donde $p(i) = \Phi([x_{(i)} - \bar{x}] / s)$, Φ la función de distribución normal estándar, \bar{x} y s son la media muestral y su desviación estándar. El valor crítico se obtiene de unas tablas según n y el nivel de significación.

Cramer-von Mises.[9]

Este test de normalidad, bajo la hipótesis nula de distribución normal de una muestra, también se basa en el análisis de la distribución empírica de dicha muestra. El estadístico de contraste es:

$$W = \frac{1}{12n} + \sum_{i=1}^n [p(i) - \frac{2i-1}{2n}]^2$$

El valor crítico del contraste esta tabulado según n y el nivel de significación.

Anderson-Darling.[10]

Este contraste de normalidad es una variación del test anterior. En este caso el estadístico de contraste es:

$$W = -n - \frac{1}{n} \sum_{i=1}^n [2i-1] [\ln(p(i)) + \ln(1-p(n-i+1))]$$

El valor crítico del contraste esta tabulado según n y el nivel de significación.

Referencias

- [1] Pearson, K. (1900). *On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can reasonably be supposed to have arisen from random sampling.* Philosophical Magazine, **50**, 157-172.
- [2] Friedman, M. (1937) *The use of ranks to avoid the assumption of normality implicit in the analysis of variance.* Journal of the American Statistical Association, **32**, 675-701.
- [3] Wilcoxon, F. (1945). *Individual comparisons by ranking methods.* Biometrics, **1**, 80-83.
- [4] Mann, H.B. (1945). *Nonparametric tests against trend.* Econometrica, **13**, 245-259.
- [5] Spearman, C. (1904). *The proof and measurement of association between two things.* Amer. J. Psychol., **15**, 72-101.
- [6] Sen, P. K.(1968). *Estimates of the regression coefficient based on Kendall's tau.* Journal of the American Statistical Association, **63**, 1379-1389.
- [7] Shapiro, S.S. and Wilk, M.B.(1965). *An Analysis of Variance Test for Normality (complete Samples).* Biometrics, **52**, 591-611.
- [8] Lilliefors, H.W. (1967). *On the Kolmogorov-Smirnov Test for Normality with Mean and Variace Unknown.* Journal of the American Statistical Association, **62**, 399-402.
- [9] Cramer, H. (1928) . *On the composition of elementary errors.* Skandinavisk Aktuarietidskrift, **11**, 141-180.
- [10] Anderson, T.W. and Darling, D.A. (1954). *A Test of Goodness of Fit.* Journal of the American Statistical Association, **49**, 765-769.