

ADVERTIMENT. L'accés als continguts d'aquesta tesi doctoral i la seva utilització ha de respectar els drets de la persona autora. Pot ser utilitzada per a consulta o estudi personal, així com en activitats o materials d'investigació i docència en els termes establerts a l'art. 32 del Text Refós de la Llei de Propietat Intel·lectual (RDL 1/1996). Per altres utilitzacions es requereix l'autorització prèvia i expressa de la persona autora. En qualsevol cas, en la utilització dels seus continguts caldrà indicar de forma clara el nom i cognoms de la persona autora i el títol de la tesi doctoral. No s'autoritza la seva reproducció o altres formes d'explotació efectuades amb finalitats de lucre ni la seva comunicació pública des d'un lloc aliè al servei TDX. Tampoc s'autoritza la presentació del seu contingut en una finestra o marc aliè a TDX (framing). Aquesta reserva de drets afecta tant als continguts de la tesi com als seus resums i índexs.

ADVERTENCIA. El acceso a los contenidos de esta tesis doctoral y su utilización debe respetar los derechos de la persona autora. Puede ser utilizada para consulta o estudio personal, así como en actividades o materiales de investigación y docencia en los términos establecidos en el art. 32 del Texto Refundido de la Ley de Propiedad Intelectual (RDL 1/1996). Para otros usos se requiere la autorización previa y expresa de la persona autora. En cualquier caso, en la utilización de sus contenidos se deberá indicar de forma clara el nombre y apellidos de la persona autora y el título de la tesis doctoral. No se autoriza su reproducción u otras formas de explotación efectuadas con fines lucrativos ni su comunicación pública desde un sitio ajeno al servicio TDR. Tampoco se autoriza la presentación de su contenido en una ventana o marco ajeno a TDR (framing). Esta reserva de derechos afecta tanto al contenido de la tesis como a sus resúmenes e índices.

WARNING. Access to the contents of this doctoral thesis and its use must respect the rights of the author. It can be used for reference or private study, as well as research and learning activities or materials in the terms established by the 32nd article of the Spanish Consolidated Copyright Act (RDL 1/1996). Express and previous authorization of the author is required for any other uses. In any case, when using its content, full name of the author and title of the thesis must be clearly indicated. Reproduction or other forms of for profit use or public communication from outside TDX service is not allowed. Presentation of its content in a window or frame external to TDX (framing) is not authorized either. These rights affect both the content of the thesis and its abstracts and indexes.

Universitat Politècnica de Catalunya
Departament d'Estadística i Investigació Operativa

Tesi Doctoral

Anàlisi de Dades Discretes:
Freqüència de Freqüències i Dades Multinomial

Presentada per: Xavier Puig i Oriol
Director: Josep Ginebra i Molins

octubre de 2009

Capítol 13

Extensions

Aquesta segona part de la tesi ha abordat el problema de la identificació de patrons de vot i de l'estudi de la seva distribució espacial. A continuació llistem alguns problemes en aquesta àrea en els que pensem seguir treballant.

1. Volem estendre el model cluster jeràrquic per al cas de tot Catalunya, tenint en compte la dependència que existeix entre els percentatges de vot i la grandària d'electors del districte. Això és necessari perquè quan implementem el Model 6 a aquestes dades trobem que hi ha una forta correlació entre el valor esperat a posteriori de la mesura de discrepància D_{2i}^2 i el nombre d'electors. Aquest fet indica que els districtes no són intercanviables i cal modificar el model, ja que si no l'estimació dels perfils serà errònia al no estar ponderada per la grandària dels districtes. Una possible manera de modelar l'efecte de la grandària dels districtes serà introduint-la com a covariable en un model Multinomial logístic (Leonard, 1994), però això complica molt la interpretació dels models i l'anàlisi simultani de dues eleccions consecutives com es proposa al punt 4 d'aquesta secció.
2. A sèries temporals i a estadística espacial un model no s'hauria de donar per bo fins que la part no explicada no presenti un patró aleatori independent i idènticament distribuït. A l'analitzar les dades electorals de Barcelona es posa de manifest que els models cluster que fem servir no eliminen del tot la dependència espacial i cal plantejar-se la possibilitat de fer servir models que contemplin aquesta dependència espacial. En cas que es considerés necessari s'estudiaria la manera d'incloure-la en la formulació del model i es buscarien maneres d'interpretar-la, potser definint un error aleatori amb estructura jeràrquica, tot adaptant la idea de Besag, York i Mollie (1991). No obstant la dificultat més immediata serà especificar a priori

quines àrees són veïnes. En el nostre cas podem definir que dos districtes són veïns si són adjacents però també podem definir com a veïns tots aquells districtes que pertanyen a un mateix municipi, alhora que podem definir també com a veïns d'un districte tots els districtes dels municipis adjacents.

3. Explorar, tal i com ha suggerit un revisor, la possibilitat d'escollir entre el Model 3, no jeràrquic, i el Model 6.1, jeràrquic, formulant un test d'hipòtesis sobre el grau d'heterogeneïtat denotat per τ , ja que en el límit quan $1/\tau$ tendeix a 0 el model jeràrquic esdevé no jeràrquic. De la mateixa manera es podria verificar si sobren clusters comparant si algunes de les w_r podrien ser zero.
4. Implementar el model jeràrquic d'un i dos clusters per a les dades del Tirant lo Blanc, on cada cluster correspondria a un autor. Estem formulant un model per avaluar la inclusió de la correlació temporal considerant que la probabilitat de que un capítol pertanyi a un cluster depengui en part d'aquesta probabilitat pels capítols anterior i posterior, ja que es raonable suposar que dos capítols consecutius tenen més probabilitat d'haver estat escrits per un mateix autor que dos capítols escollits a l'atzar.
5. Modelar alhora els resultats de dues eleccions consecutives introduint matrius de transició entre categories d'eleccions consecutives, amb l'objectiu d'estimar la magnitud i la direcció del transvasament de vots entre categories a dins de cada un dels clusters identificats. Com més intercanviable siguin les àrees pertanyents a cadascun dels clusters considerats més s'assemblarà el comportament global capturat a través d'aquestes matrius amb el comportament a nivell individual de cada elector.

Apèndix A

Cadenes de les simulacions obtingudes amb el WinBugs

És molt important dedicar atenció a la convergència de les cadenes resultants de la simulació MCMC. S'ha de verificar si s'ha assolit la convergència de les cadenes o bé es necessari augmentar el nombre de simulacions d'escalfament. Per avaluar la convergència de les cadenes s'ha combinat la inspecció visual de les traces mostrals amb el càlcul de l'autocorrelació mostral i la mesura diagnòstic proposada per Gelman i Rubin (1992). També és important la inspecció visual de les cadenes en els models de mixtura per detectar l'existència de problemes d'identificabilitat.

En aquest apèndix a mode il·lustratiu presentem el procés d'avaluació de la convergència de les cadenes per al cas concret del model de tres clusters jeràrquic, $M_{3,J}$, de l'any 1999.

És habitual no utilitzar una sola cadena si no dues o més cadenes en paral·lel amb valors inicials allunyats en diferents direccions de la vertadera distribució a posteriori, llavors les cadenes poden ser dibuixades en un mateix gràfic i veure si existeix un punt en que les cadenes s'arriben a superposar. El problema és que desconeixem si les cadenes utilitzades estan allunyades en diferents direccions de la vertadera distribució a posteriori ja que aquesta a priori és desconeguda.

El procés d'avaluació de la convergència no és possible automatitzar-lo, cosa que acaba requerint un judici subjectiu a través de la inspecció visual. Altrament la inspecció visual permet detectar fàcilment alguns problemes en la convergència poden ser l'existència de modes, la falta d'identificabilitat dels models, l'existència d'autocorrelació, el no assoliment del procés estacionari, etc.

Existeixen algunes regles de diagnostic d'assoliment de la convergència que es poden consultar a Cowles i Carlin (1996) i Mengersen *et al.* (1999). Entre els diagnòstics més populars, i que hem utilitzat, es pot destacar el proposat per Gelman i Rubin (1992) que es basa en comparar la variabilitat de dins les cadenes amb la variabilitat entre cadenes per a un paràmetre donat.

Una altra eina diagnòstica, que també hem utilitzat, i que ajuda a valorar la convergència és explorar l'autocorrelació entre simulacions consecutives, l'existència d'autocorrelació d'una banda alenteix la convergència i de l'altre la inferència és menys precisa que un altre simulació de grandària similar sense autocorrelació. L'autocorrelació no és en sí un problema però quan passa és necessari augmentar el nombre de simulacions, també davant la presència d'autocorrelació s'acostuma a prendre simulacions equiespaiades de manera que disminueixi l'autocorrelació. Quan es té un gran nombre de paràmetres és molt útil per no haver de guardar un nombre massa gran de simulacions.

En problemes d'un gran nombre de paràmetres, amb existència d'autocorrelació i possibles problemes d'identificabilitat una estratègia per millora la convergència és partir de valors inicials basats en les dades tal i com proposen Gelman *et al.* (2004).

La inspecció visual de les cadenes com a estratègia per avaluar la convergència és una pràctica habitual (Kass, 1998), però aquesta anàlisi es pot completar amb altres mesures i gràfics d'autocorrelació. En aquest línia Cowles i Carlin (1996) recomana la combinació de diferents estratègies que combinades amb el sentit comú i l'experiència esdevenen la a millor estratègia per al complexa problema del diagnòstic de la convergència. En tot cas cal emfasitzar que tot i que hi han resultats que garanteixen la convergència de les cadenes en l'infinit amb una mostra finita de MCMC no és possible assegurar mai que sigui representativa de la distribució estacionària.

Donada la dificultat d'assolir la convergència en els models jeràrquics, i especialment ho hem comprovats amb simulacions preliminars per aquests models, s'han assignat valors inicials per les cadenes tant valors empírics com valors obtinguts del model no jeràrquic afegint una perturbació aleatòria, i els valors inicials per τ han estat valors molt grans per una cadena i petits per l'altre, de manera que s'ha establert el període d'escalfament fins que les respectives cadenes no han mostrejat en el mateix espai. Per a cadascun dels models hem fet córrer dues cadenes MCMC en paral·lel amb diferents valors inicials.

Primer mostrem el comportament de les cadenes de les 5000 primeres iteracions, i després mostrem les cadenes utilitzades per a la inferència un cop descartades les primeres 50000 simulacions que s'han considerat necessàries per assolir la convergència.

A.1 Simulacions d'escalfament

A les Figures A.1 i A.2 s'hi representen les traces mostrals per les 5000 primeres simulacions de cadascuna de les dues cadenes per als principals paràmetres d'interès on s'observa que en aquestes 5000 primeres simulacions les cadenes encara no s'han barrejat de forma satisfactòria. La Figura A.3 presenta les traces mostrals del període d'escalfament per les primeres 50 variables latents on s'observa com en les primeres 5000 primeres iteracions hi ha una gran variabilitat en totes elles per després estabilitzar-se. Aquesta variabilitat es manifesta especialment en la cadena amb valors inicials petits per τ que equival a partir d'un grau alt d'heterogeneïtat.

Les Figures A.4 i A.5 presenten la funció d'autocorrelació i s'observa una forta correlació per algunes μ 's i una molt forta correlació a les τ . L'existència d'autocorrelació és indicatiu de convergència lenta amb el que caldrà utilitzar cadenes llargues i/o prendre simulacions equiespaiades.

Per aquest cas concret no s'han observat les cadenes barrejades de forma satisfactòria per a tots els paràmetres fins després de les 20000 primeres simulacions. No obstant en altres models el període d'escalfament necessari ha estat superior, i és per aquest motiu s'ha decidit escollir una valor únic, de forma que per a tots el models s'ha decidit descartar les primeres 50000 iteracions de cada cadena com a escalfament previ a la convergència.

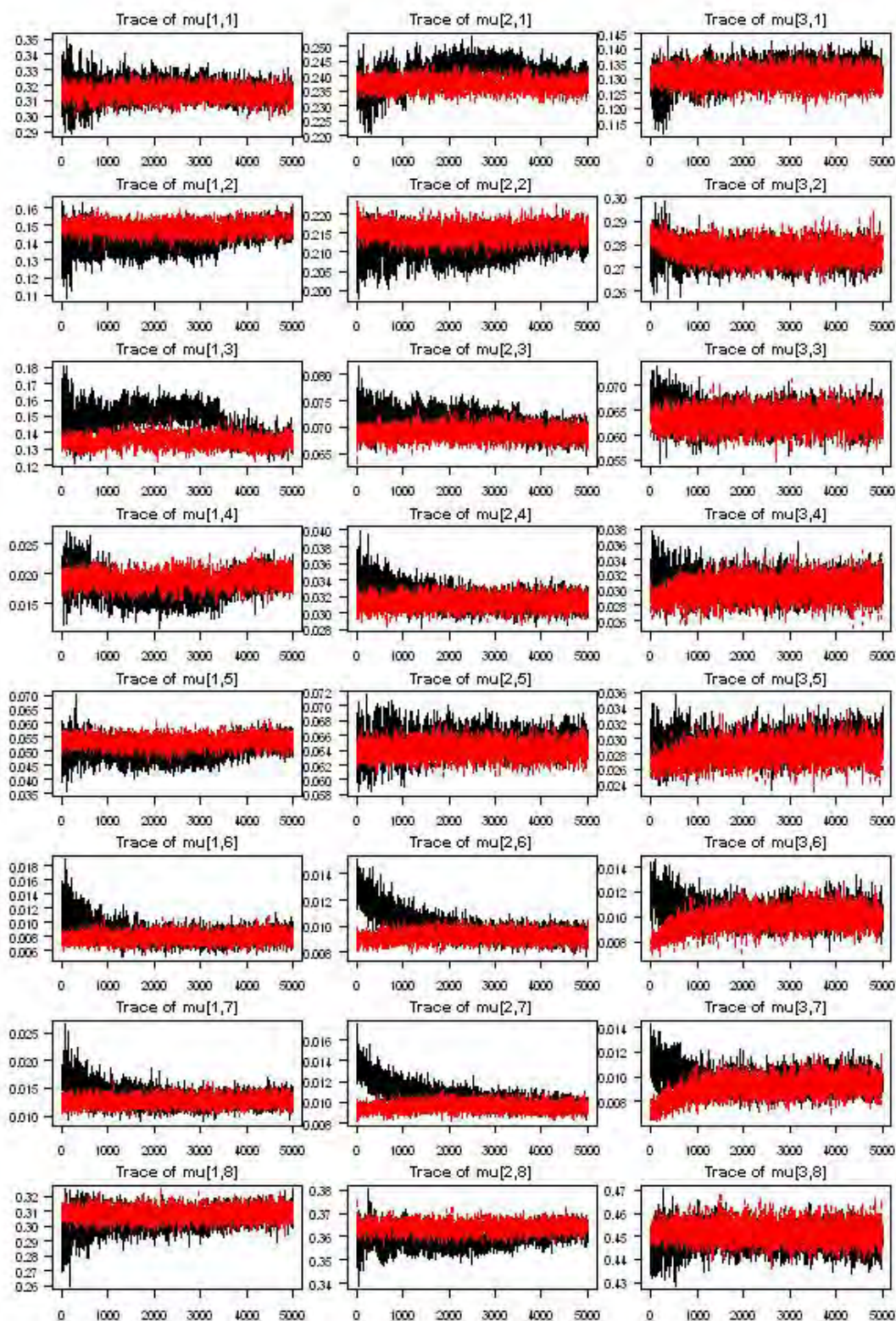


Figura A.1: Traces mostrals per per les 5000 primeres simulacions per als paràmetres μ_{rj} , per $j = 1 \dots, 8$ i per $r = 1, 2, 3$.

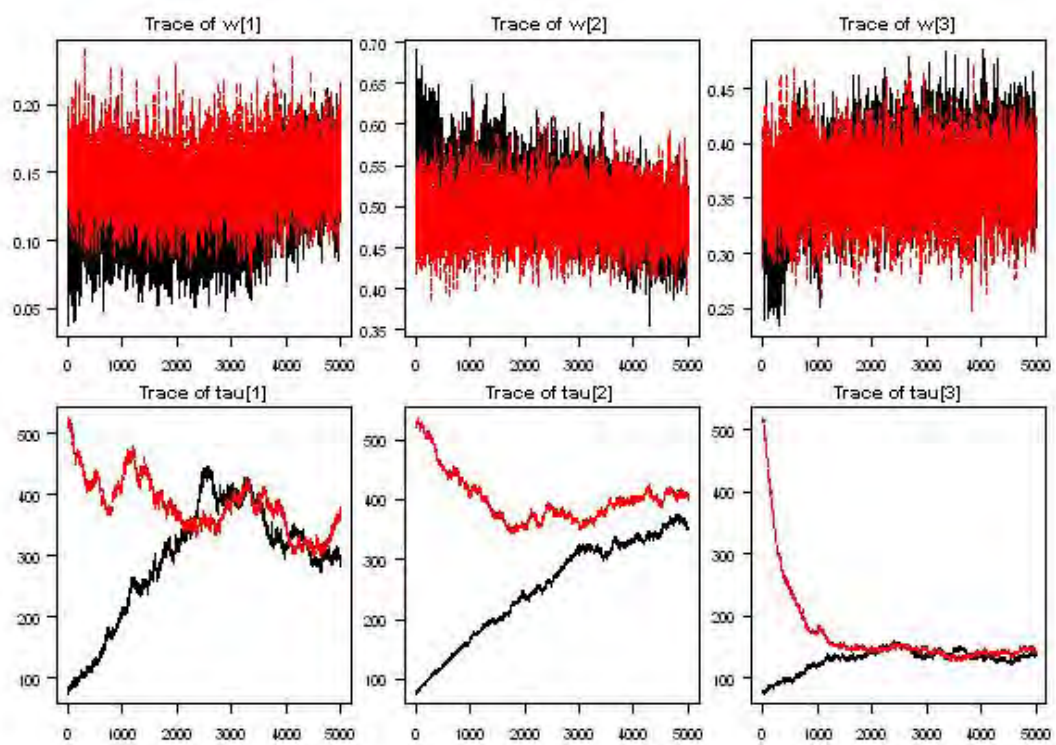


Figura A.2: Traces mostrals per les 5000 primeres simulacions per als paràmetres τ_r i ω_r , per $r = 1, 2, 3$.

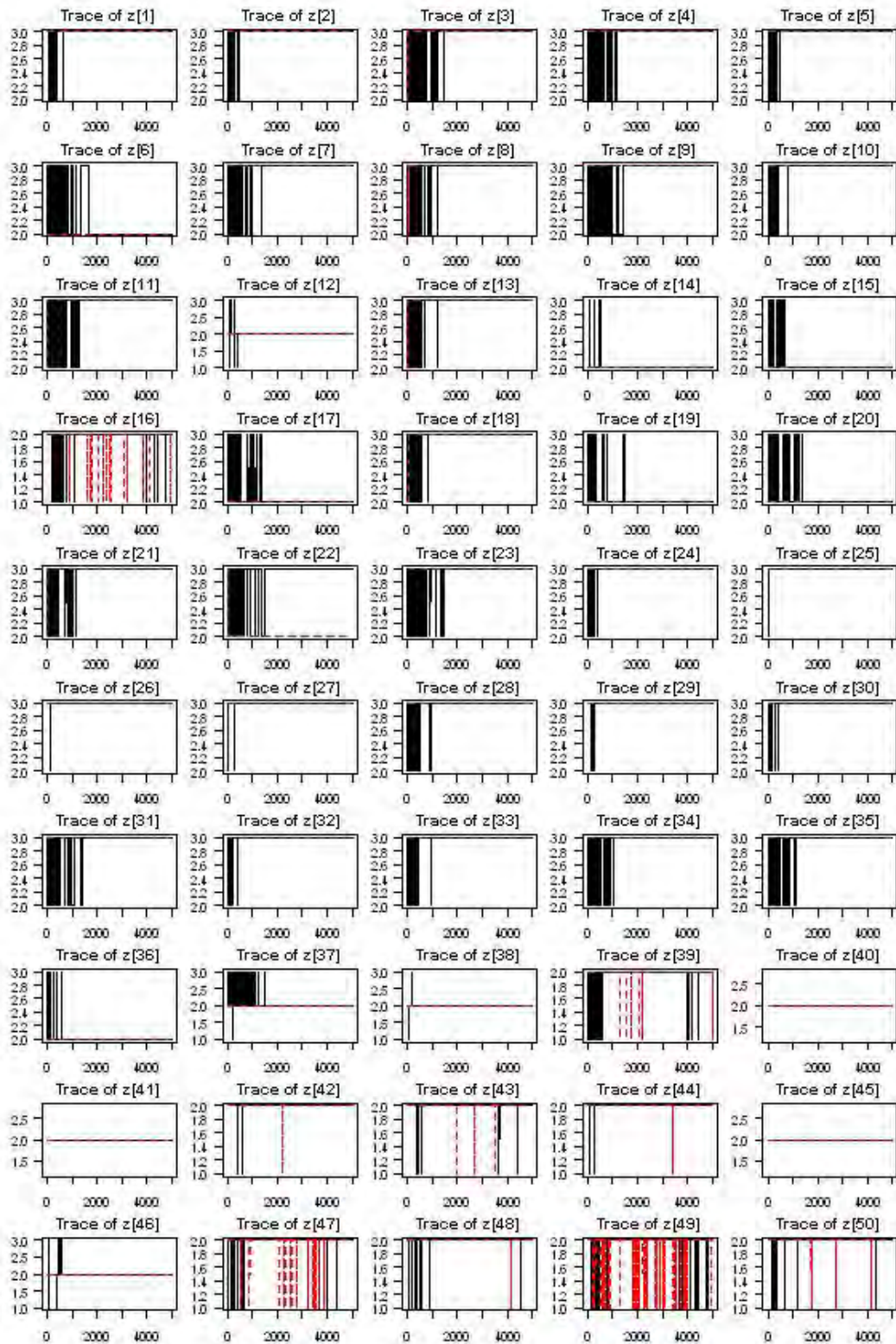


Figura A.3: Traces mostrals per a les primeres 50 variables latents ζ_r , per $r = 1, \dots, 50$, per les primeres 5000 simulacions.

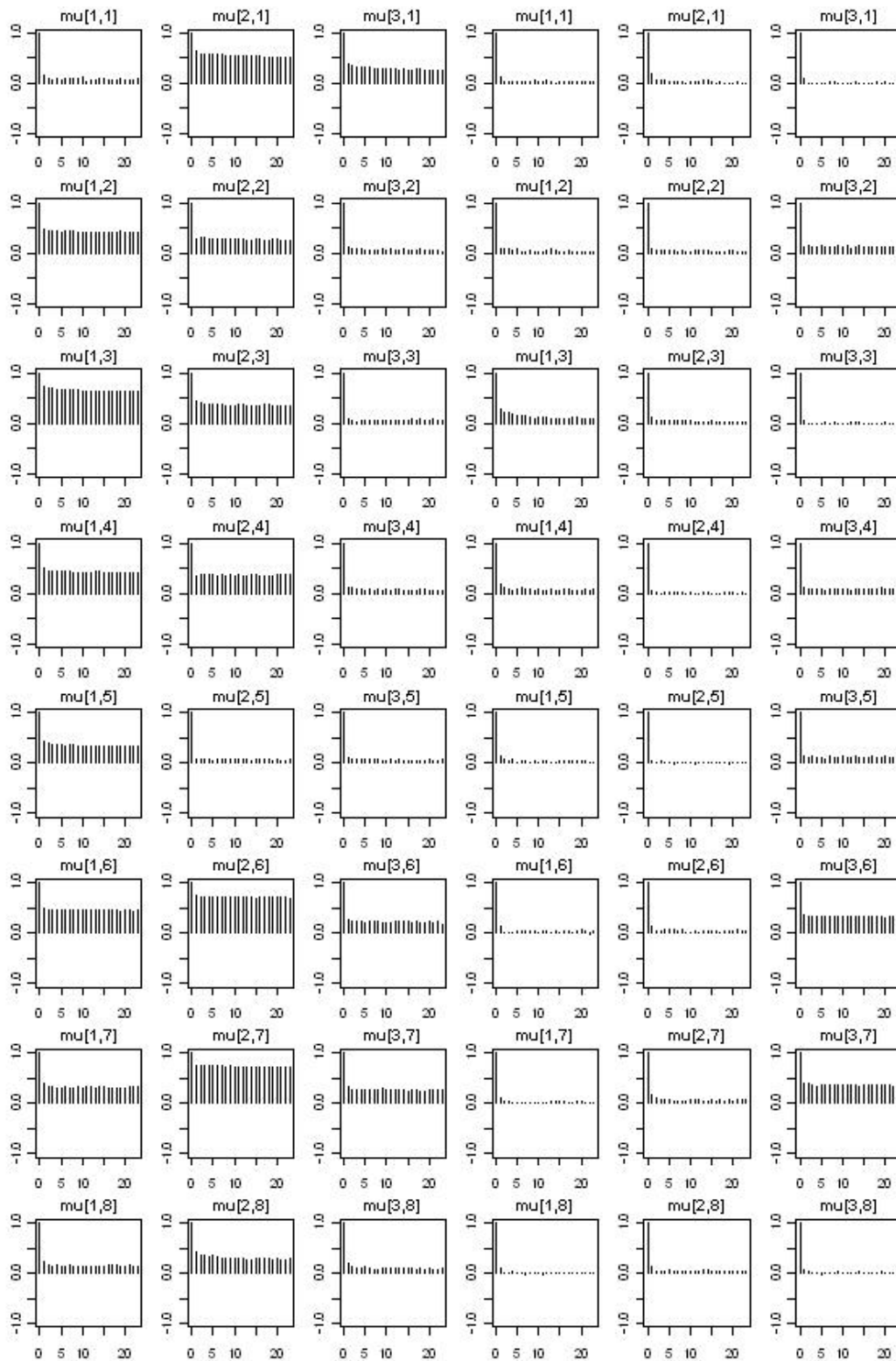


Figura A.4: Autocorrelació mostrada per als paràmetres μ_{rj} , per $j = 1 \dots, 8$ i per $r = 1, 2, 3$, per les 5000 primeres simulacions. Les tres primeres columnes corresponen a la primera cadena i les tres darreres columnes corresponen a la segona cadena.

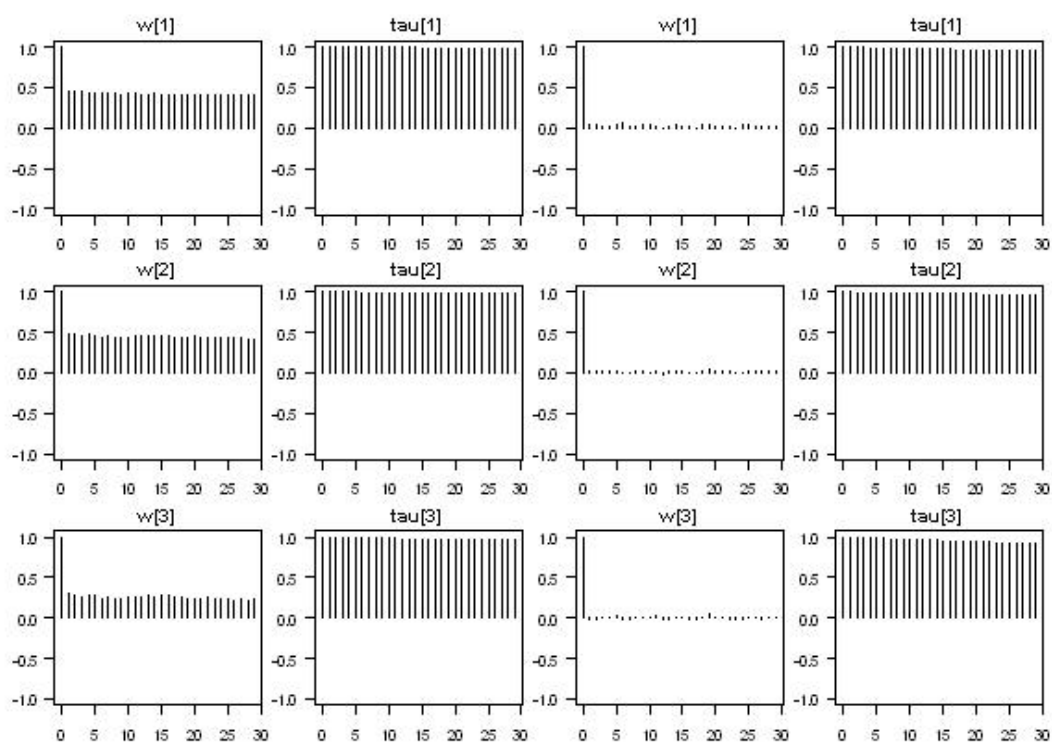


Figura A.5: Autocorrelació mostral per als paràmetres τ_r i ω_r , per per $r = 1, 2, 3$, per les 5000 primeres simulacions. Les dues primeres columnes corresponen a la primera cadena i les dues darreres columnes corresponen a la segona cadena.

A.2 Simulacions monitoritzades

Un cop descartades les primeres 50000 iteracions de cada cadena, de les simulacions següents se n'han retingut una de cada 100, i la mostra final per a les anàlisis de la distribució a posteriori ha estat de 3000 simulacions, 1500 de cada cadena.

El temps de computació de PC per aquest model i aquest any, incloent les simulacions d'escalfament, ha estat de 47 hores i 24 minuts, els temps augmenten amb les dades de l'any 2006, ja que hi ha una categoria més, i amb el model de quatre clusters, de manera que per exemple la simulació de les cadenes del model M_{4J} per l'any 2006 ha requerit prop de 57 hores.

A les Figures A.6 i A.7 s'hi presenten les traces mostrals monitoritzades per a la inferència de les dues cadenes per als diferents paràmetres 'interès on s'observa com les cadenes s'han barrejat satisfactòriament. La Figura A.8 presenta les traces mostrals per les primeres 50 variables latents on s'observa un patró satisfactori excepte per alguna ζ_i , concretament en el gràfic ζ_9 i ζ_{23} . Per aquest model M_{3J} amb les dades de l'any 1999 trobem que en total 7 zrp que en una cadena pertanyen a un cluster i en una altra en un altre cluster, aquest fenomen s'observa amb igual mesura en els anys 1995 i 1999 tant pel model M_{3J} com pel model M_{4J} , i pràcticament no s'observa ni en els altres models ni en els altres anys.

Les Figures A.9 i A.10 presenten la funció d'autocorrelació pels diferents paràmetres, i s'observa com al monitoritzar una simulació de cada 100 s'ha eliminat l'autocorrelació de les μ i es redueix la de les τ .

La mesura per avaluar la convergència proposada per Gelman i Rubin (1992) es basa en la comparació de les variàncies a dintre les cadenes i entre les cadenes, de forma similar a l'anàlisi de la variància. La mesura apareix a la literatura com \hat{R} , valors substancialment majors que 1 indiquen falta de convergència, en general valors menors que 1.2 es consideren satisfactoris.

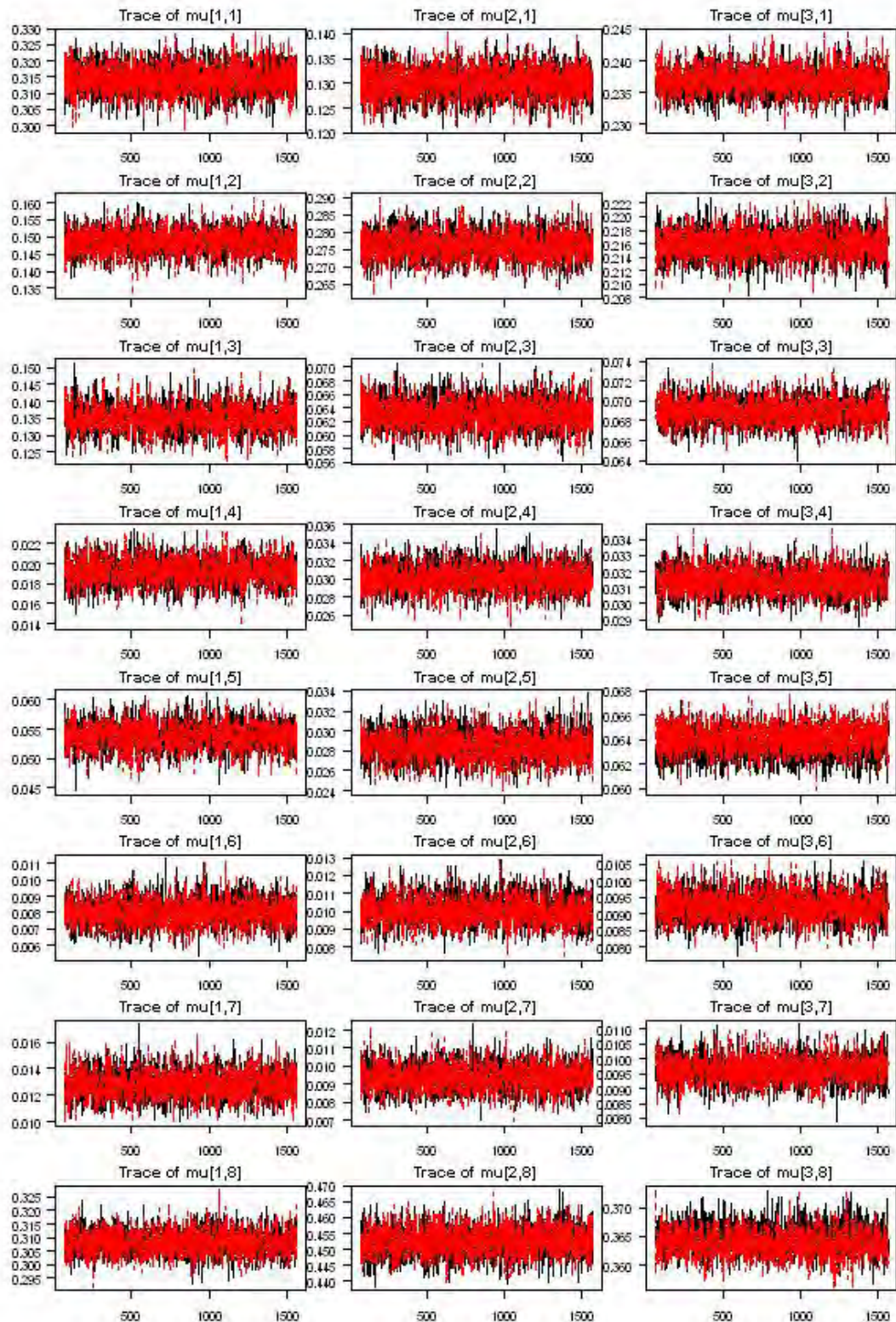


Figura A.6: Traces mostrals per als paràmetres μ_{rj} , per $j = 1 \dots, 8$ i per $r = 1, 2, 3$, per les 1500 simulacions de cada una de les cadenes obtingudes a partir de monitoritzar-ne una de cada 100 d'un total de 150000 simulacions per a cada cadena. El període d'esclafament havia estat de 50000 simulacions en cada cadena.

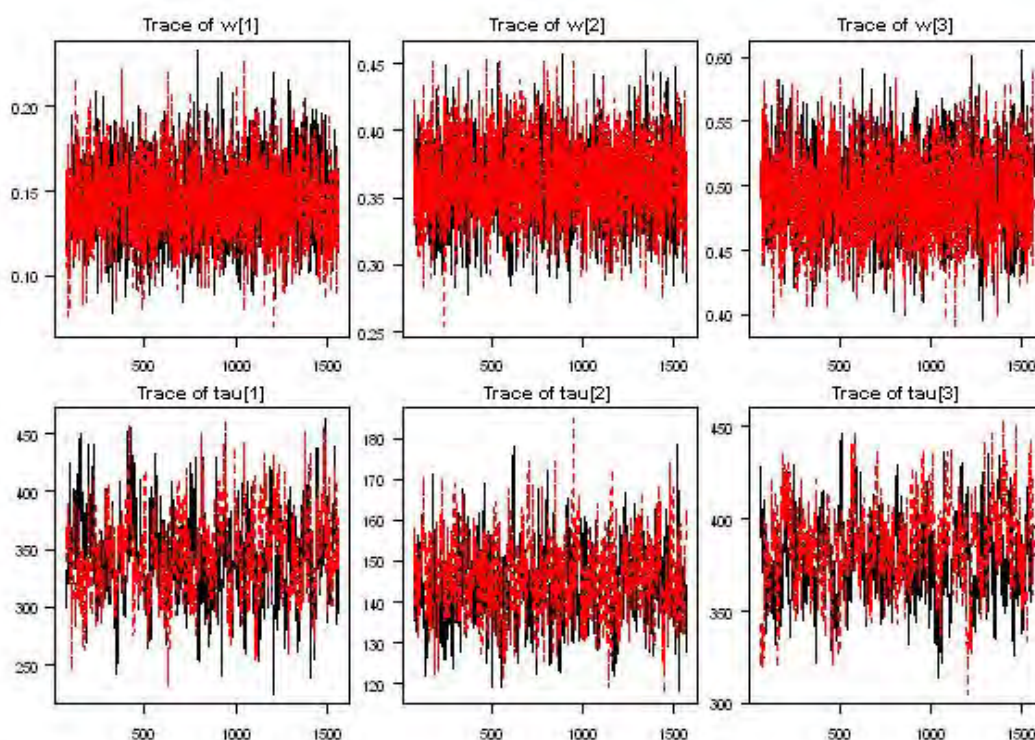


Figura A.7: Traces mostrals per als paràmetres τ_r i ω_r , per $r = 1, 2, 3$, per les 1500 simulacions de cada una de les cadenes obtingudes a partir de monitoritzar-ne una de cada 100 d'un total de 150000 simulacions per a cada cadena. El període d'esclafament havia estat de 50000 simulacions en cada cadena.

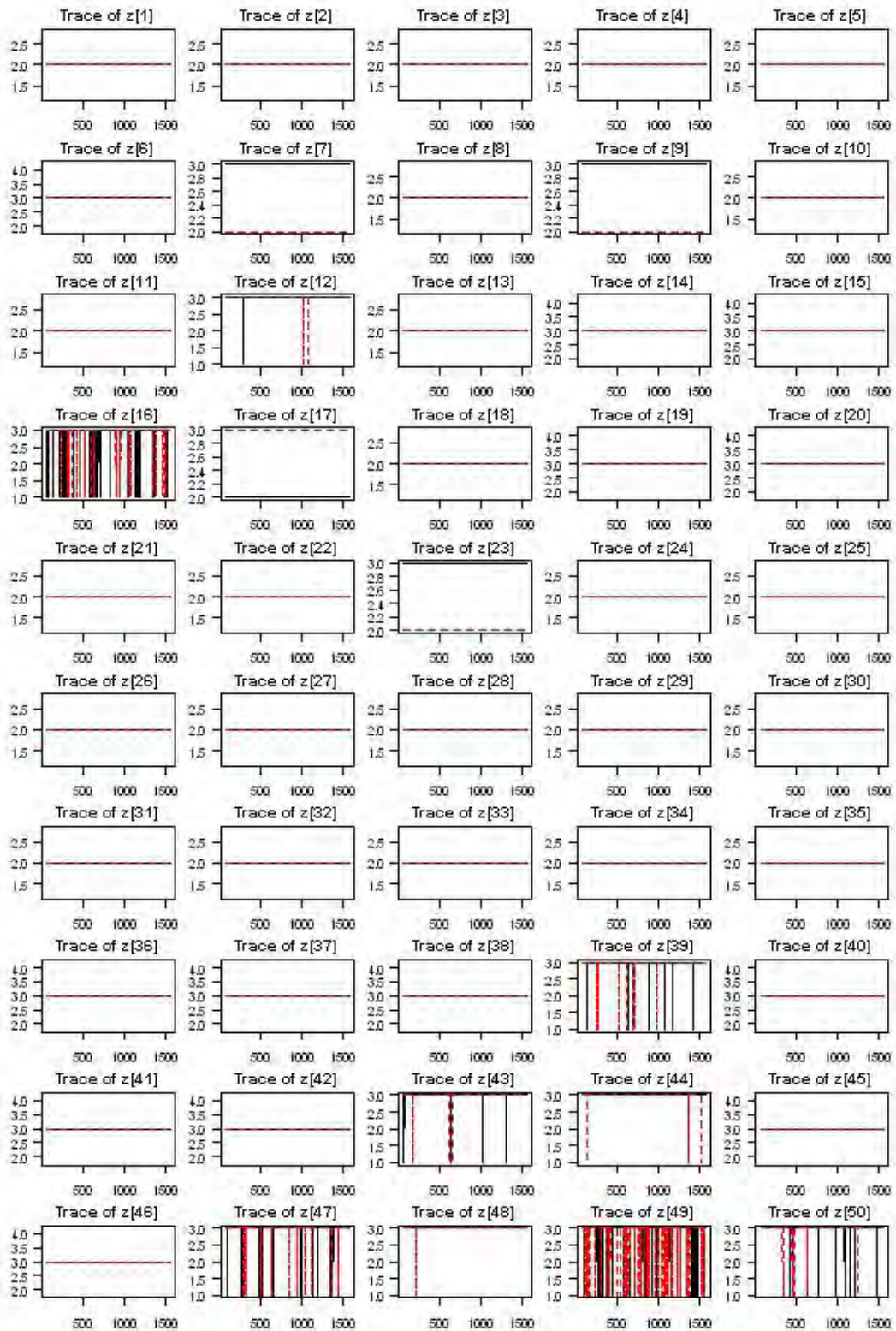


Figura A.8: Traces mostrals per a les primeres 50 variables latents ζ_r , per $r = 1, \dots, 50$, per les 1500 simulacions de obtingudes a partir de monitoritzar-ne una de cada 100 d'un total de 150000 simulacions per a cada cadena. El període d'esclafament havia estat de 50000 simulacions en cada cadena.

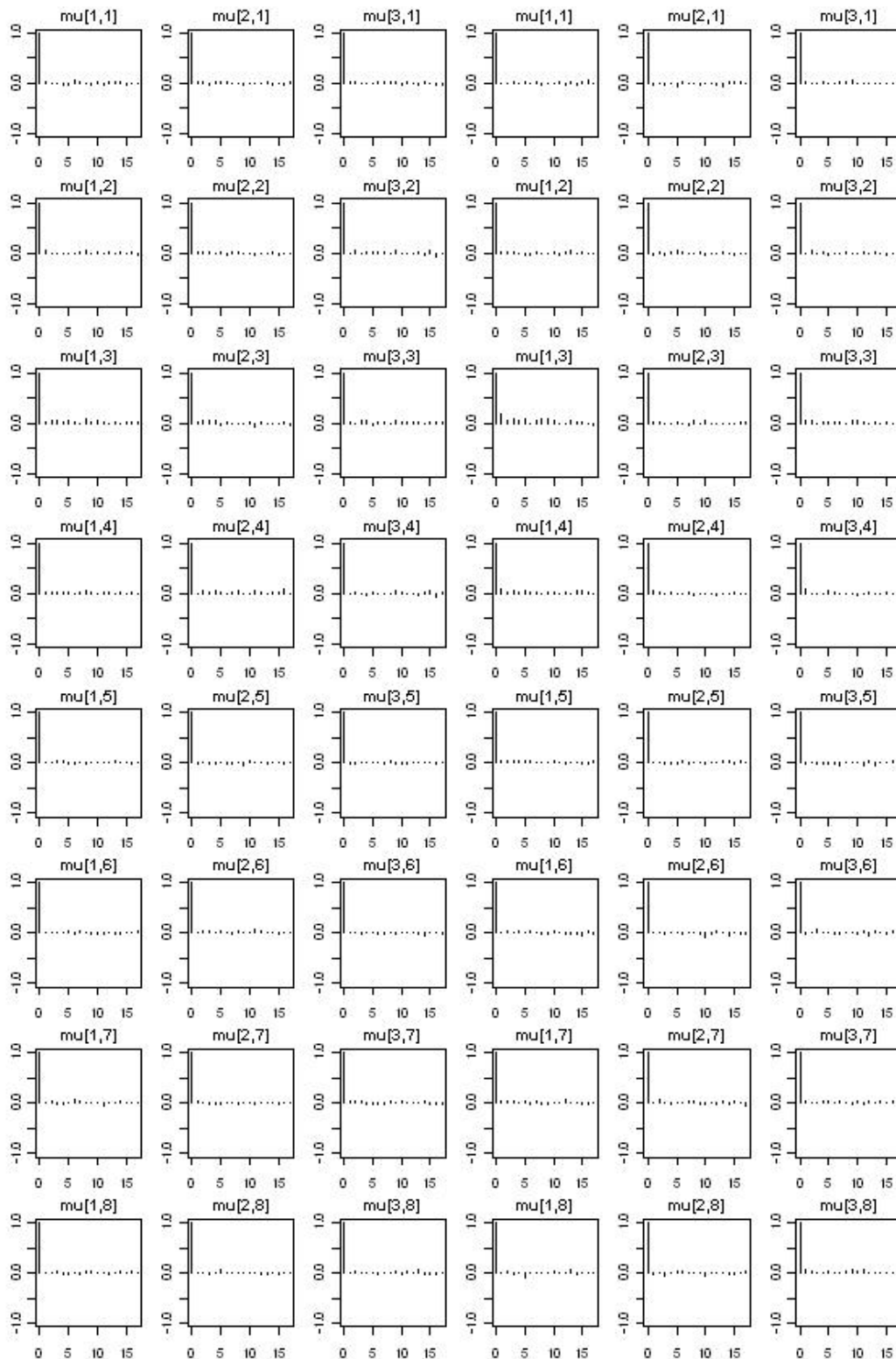


Figura A.9: Autocorrelació mostrada per als paràmetres μ_{rj} , per $j = 1, \dots, 8$ i per $r = 1, 2, 3$, per les 1500 simulacions de cada una de les cadenes obtingudes a partir de monitoritzar-ne una de cada 100 d'un total de 150000 simulacions per a cada cadena. Les tres primeres columnes corresponen a la primera cadena i les tres darreres columnes corresponen a la segona cadena.

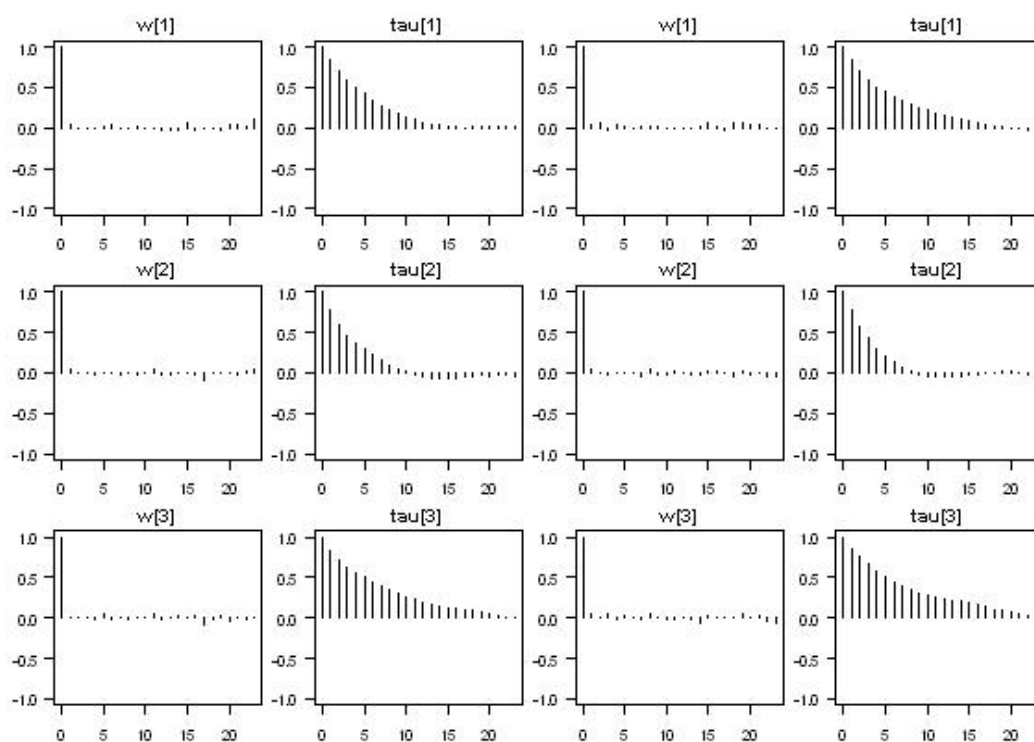


Figura A.10: Autocorrelació mostral per als paràmetres τ_r i ω_r , per per $r = 1, 2, 3$, per les 1500 simulacions de cada una de les cadenes obtingudes a partir de monitoritzar-ne una de cada 100 d'un total de 150000 simulacions per a cada cadena. Les dues primeres columnes corresponen a la primera cadena i les dues darreres columnes corresponen a la segona cadena.

	cluster 1	cluster 2	cluster 3
mu[r,1]	1.00	1.01	1.03
mu[r,2]	1.00	1.00	1.00
mu[r,3]	1.00	1.00	1.00
mu[r,4]	1.00	1.00	1.00
mu[r,5]	1.00	1.02	1.15
mu[r,6]	1.00	1.01	1.00
mu[r,7]	1.00	1.00	1.00
mu[r,8]	1.00	1.01	1.11
w[r]	1.00	1.01	1.01
tau[r]	1.00	1.03	1.06

Taula A.1: Mesura potencial de reducció de la convergència, \widehat{R} , proposada per Gelman i Rubin (1992) aplicada a les simulacions del model de tres clusters de l'any 1999, obtingudes a partir de monitoritzar una simulació de cada 100 d'un total de 150000 simulacions per a cada cadena. El període d'esclafament havia estat de 50000 simulacions en cada cadena.

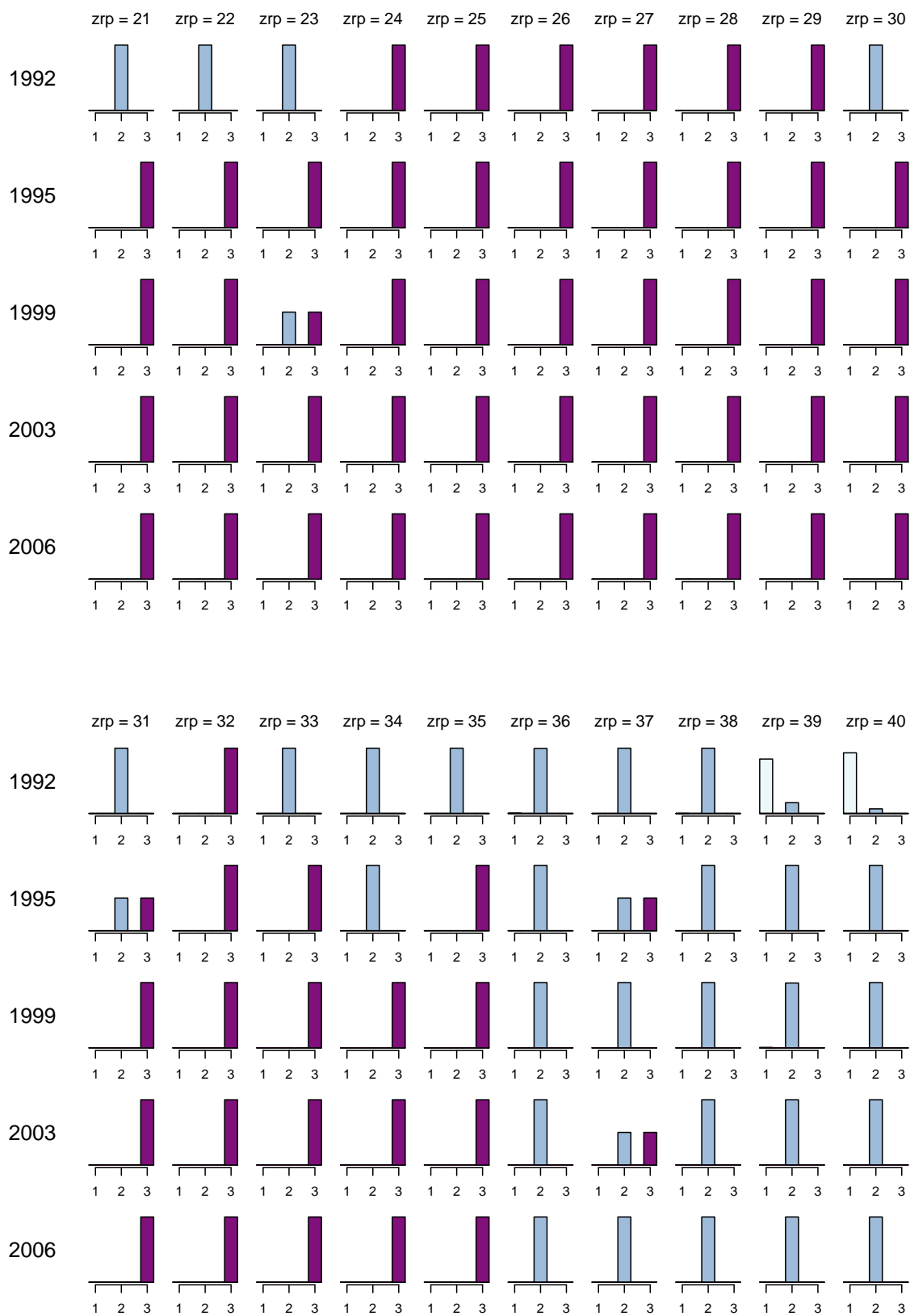
Apèndix B

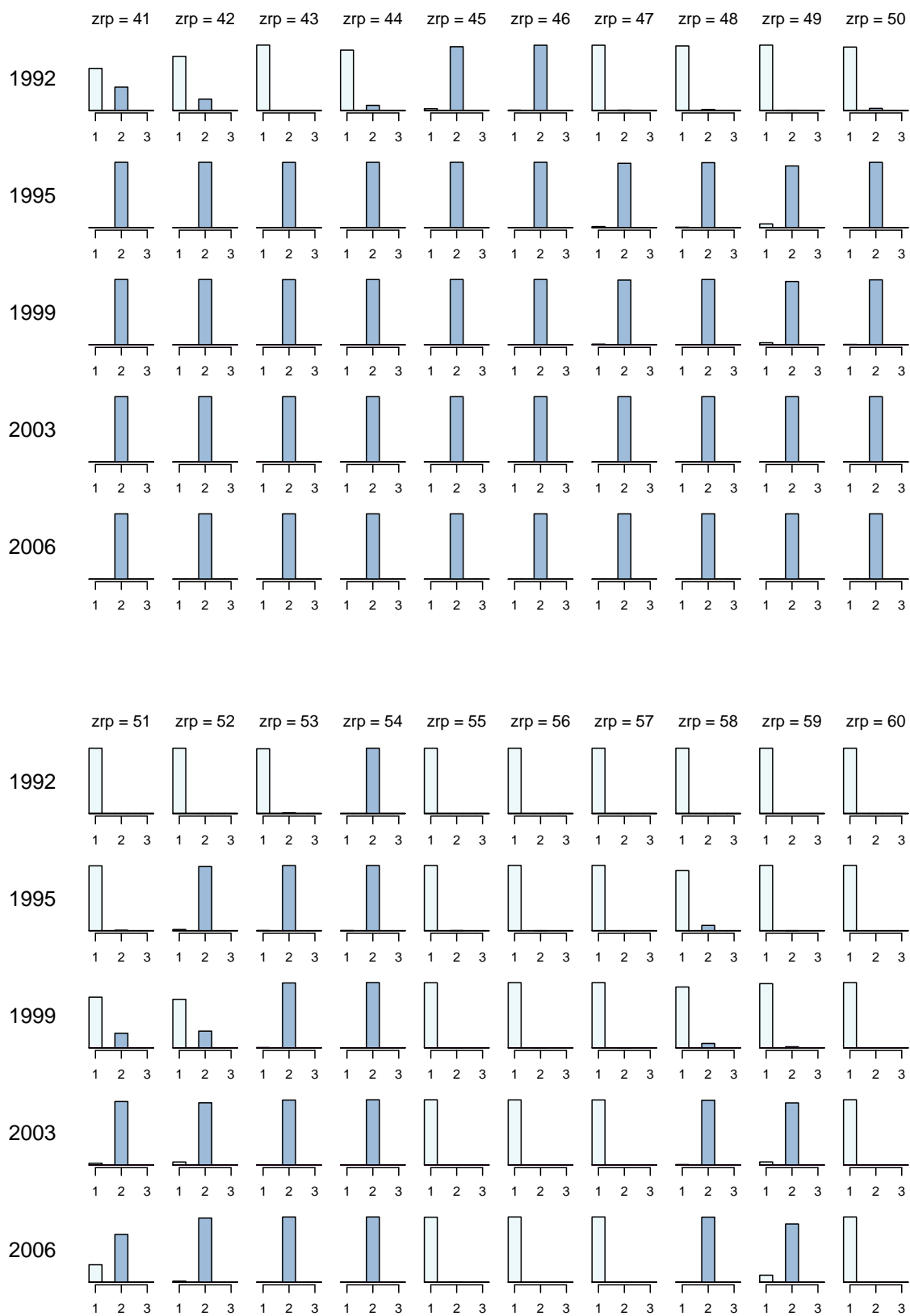
Distribució a posteriori de ζ_i per als models M_{3J} i M_{4J}

En aquest apèndix presentem de forma gràfica la distribució a posteriori de cada zrp de pertànyer a cadascun dels clusters, $\pi(\zeta_i|y)$, per a cada any, sota el model de tres clusters, M_{3J} , i sota el model de quatre clusters, M_{4J} . La representació gràfica és una alternativa a la presentació de resultats en taules que en aquest cas permet visualitzar més bé l'evolució de cada zrp al llarg de les cinc eleccions estudiades. A la primera secció d'aquest apèndix presentem els resultats per al model M_{3J} i a la segona secció els resultats per al model M_{4J} .

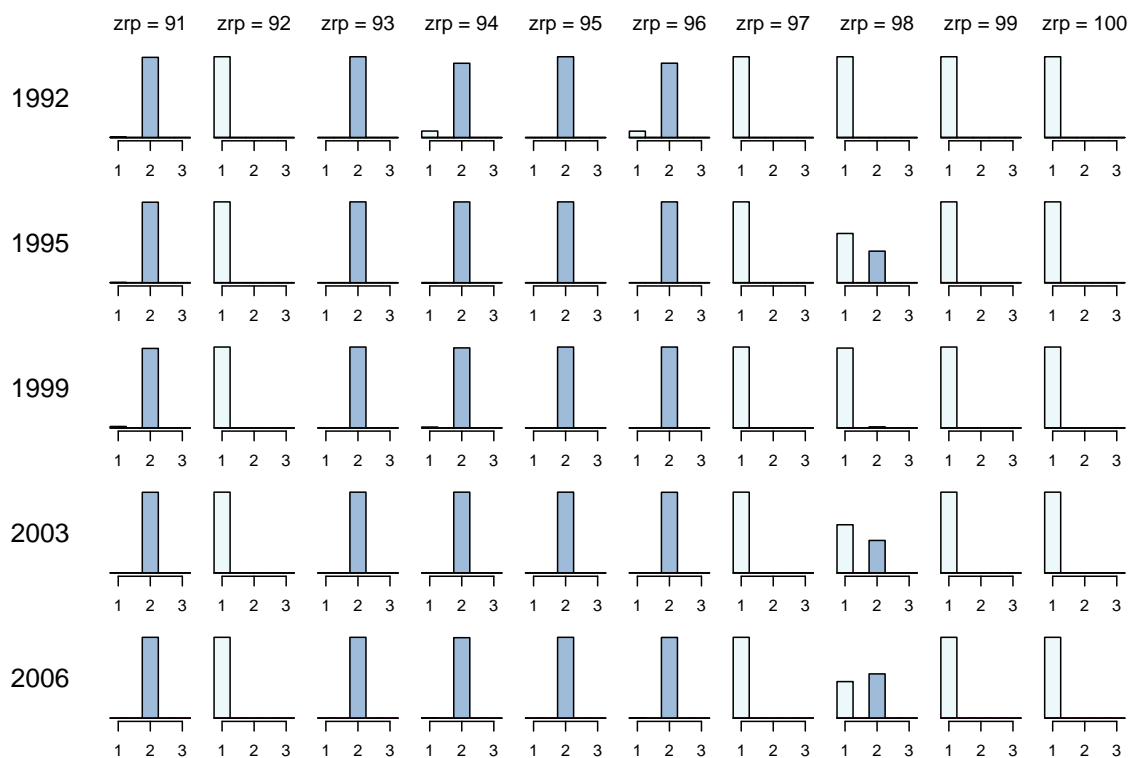
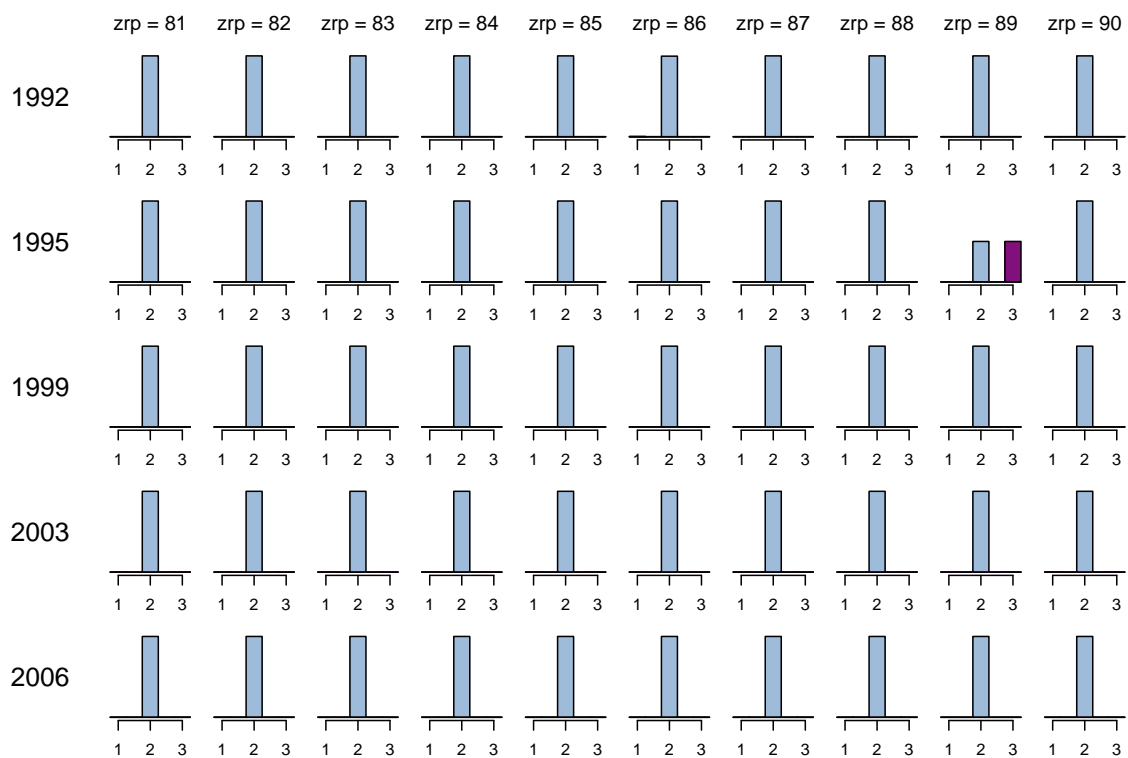
B.1 Distribució a posteriori de ζ_i per al model M_{3J}

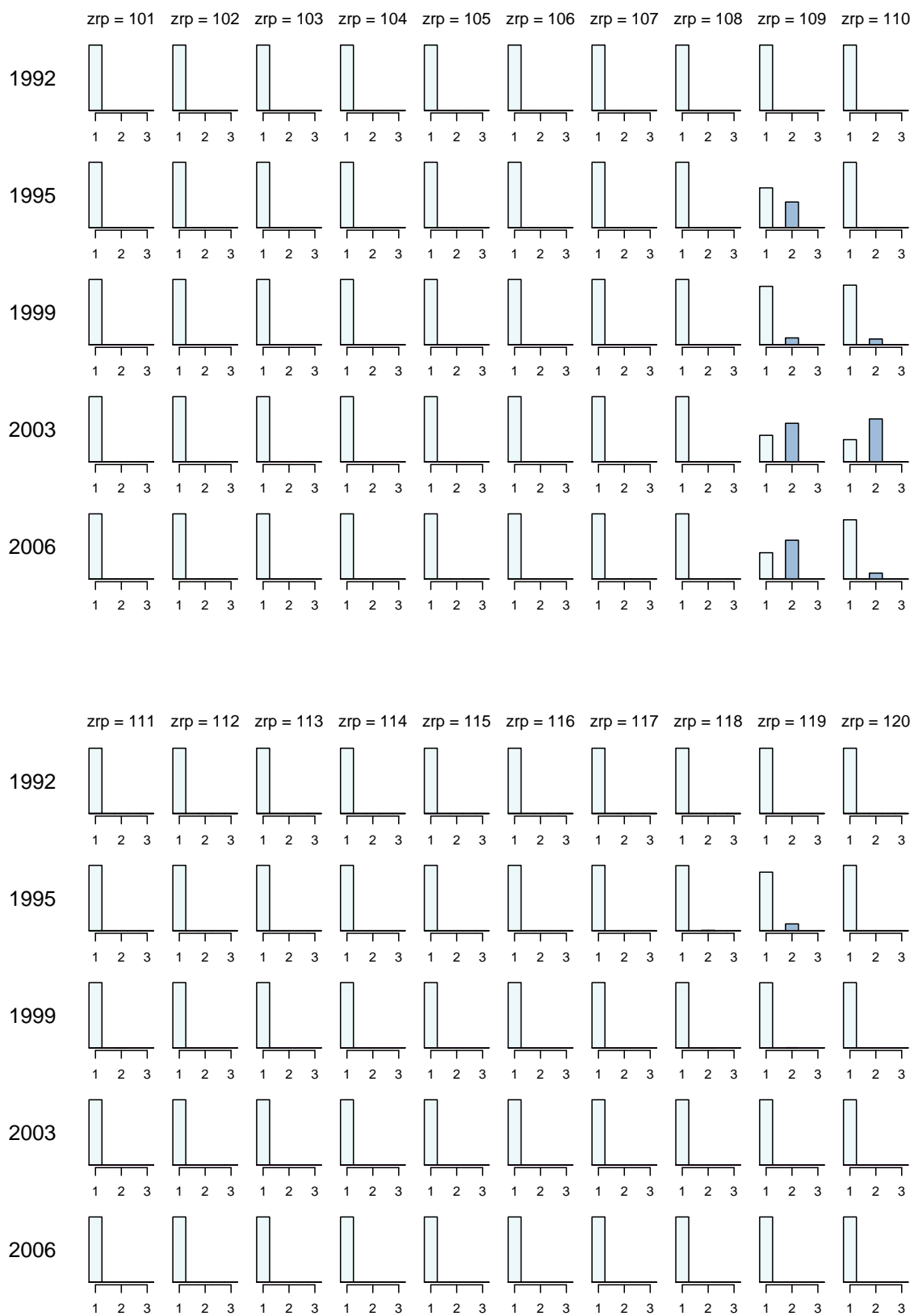


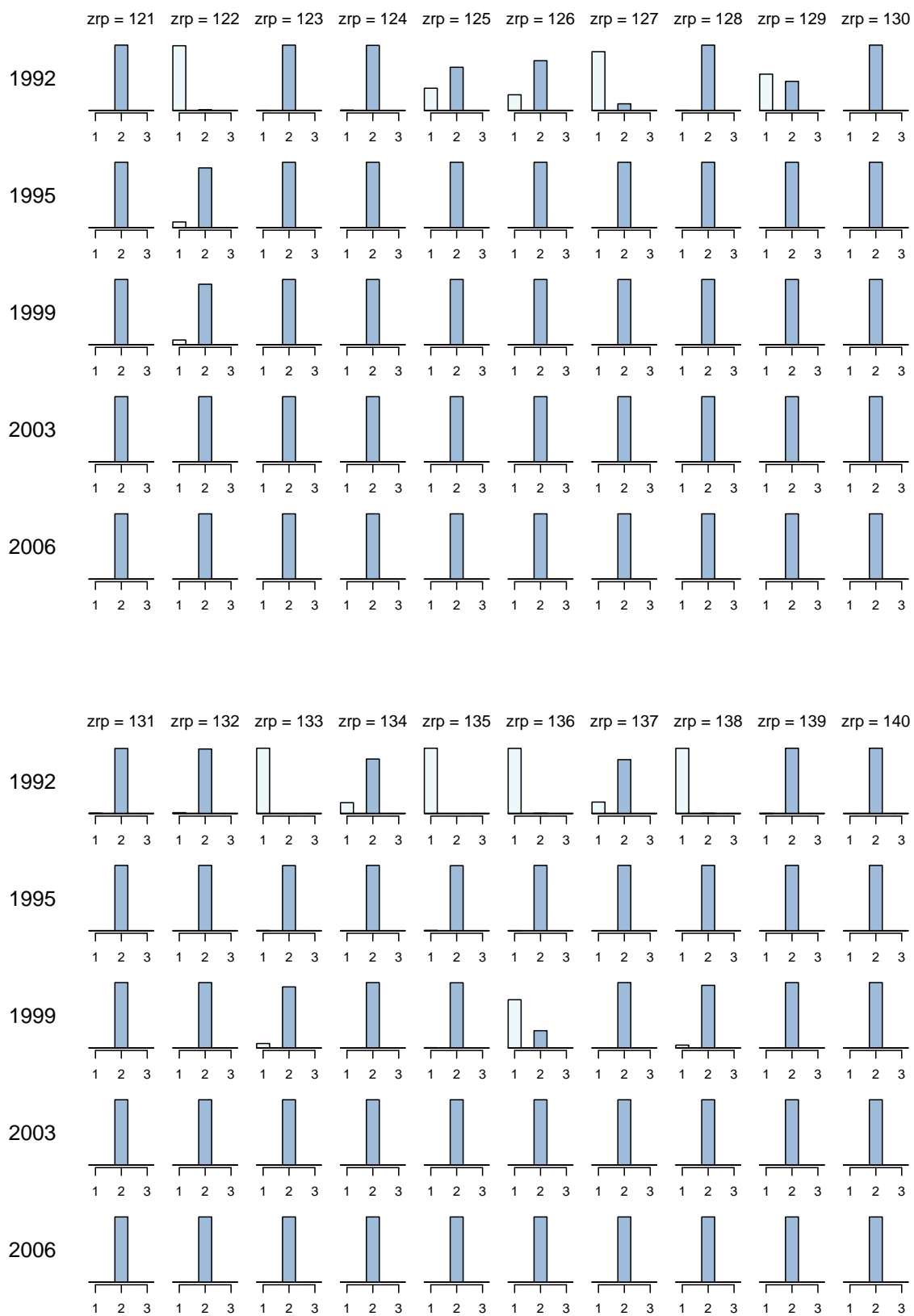


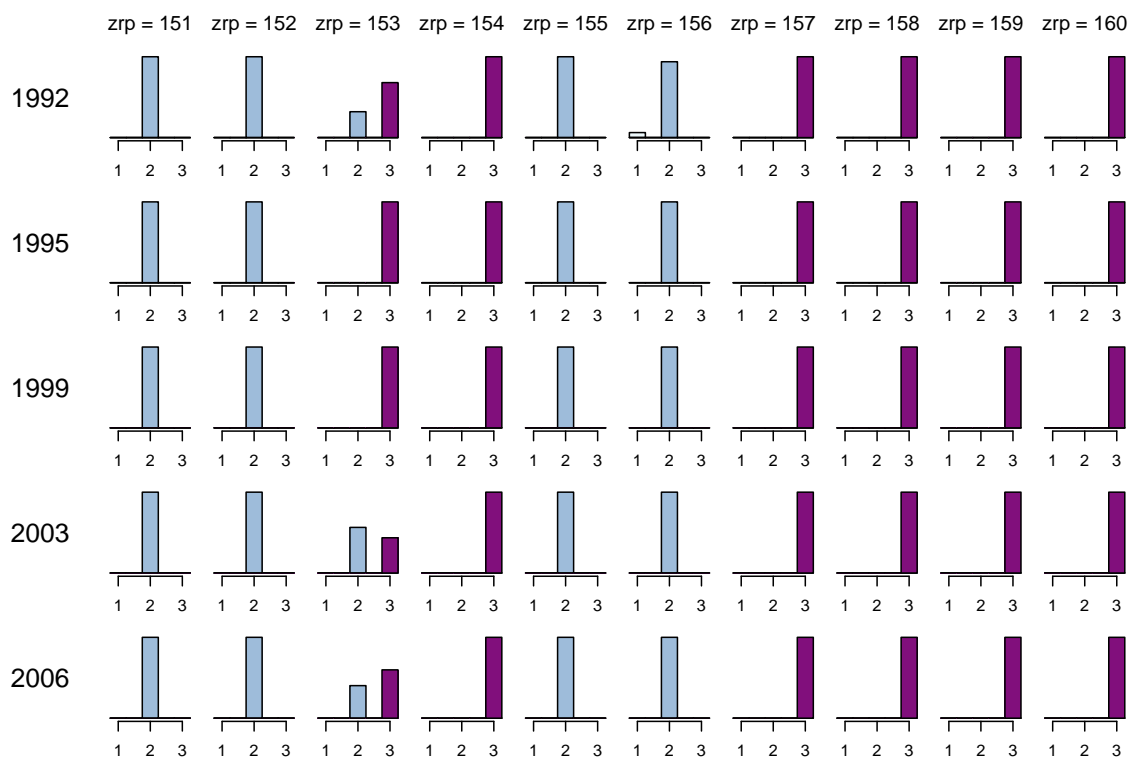
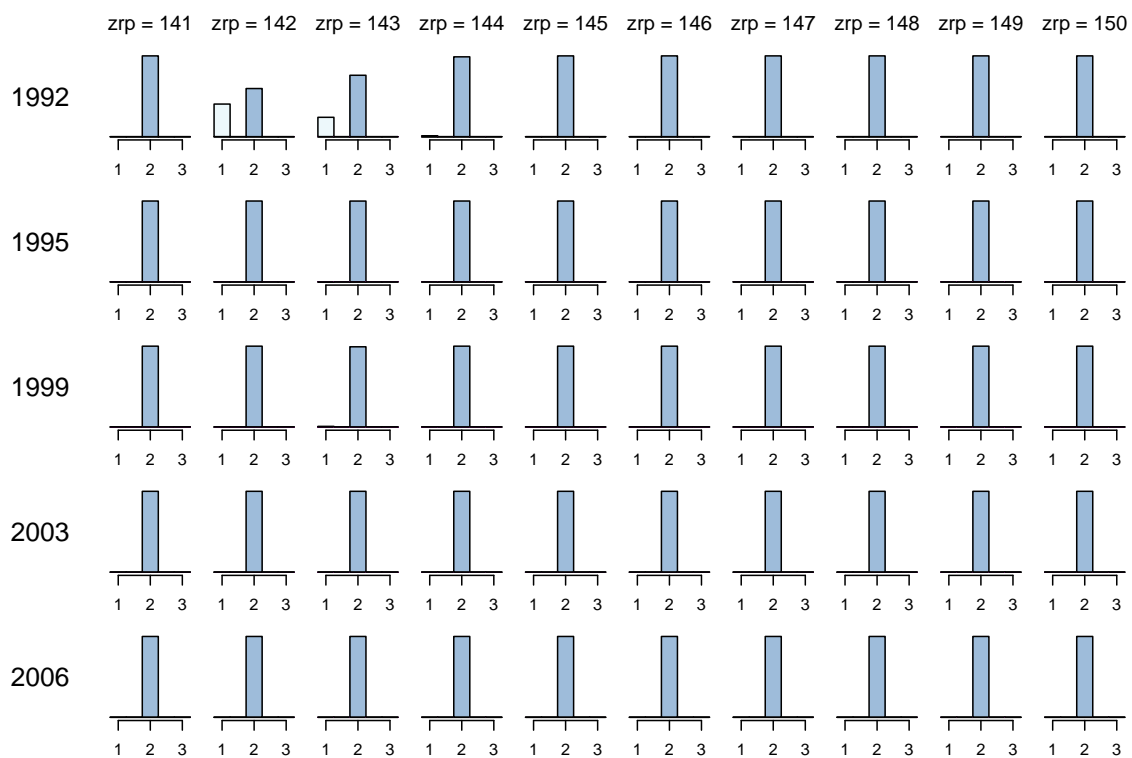




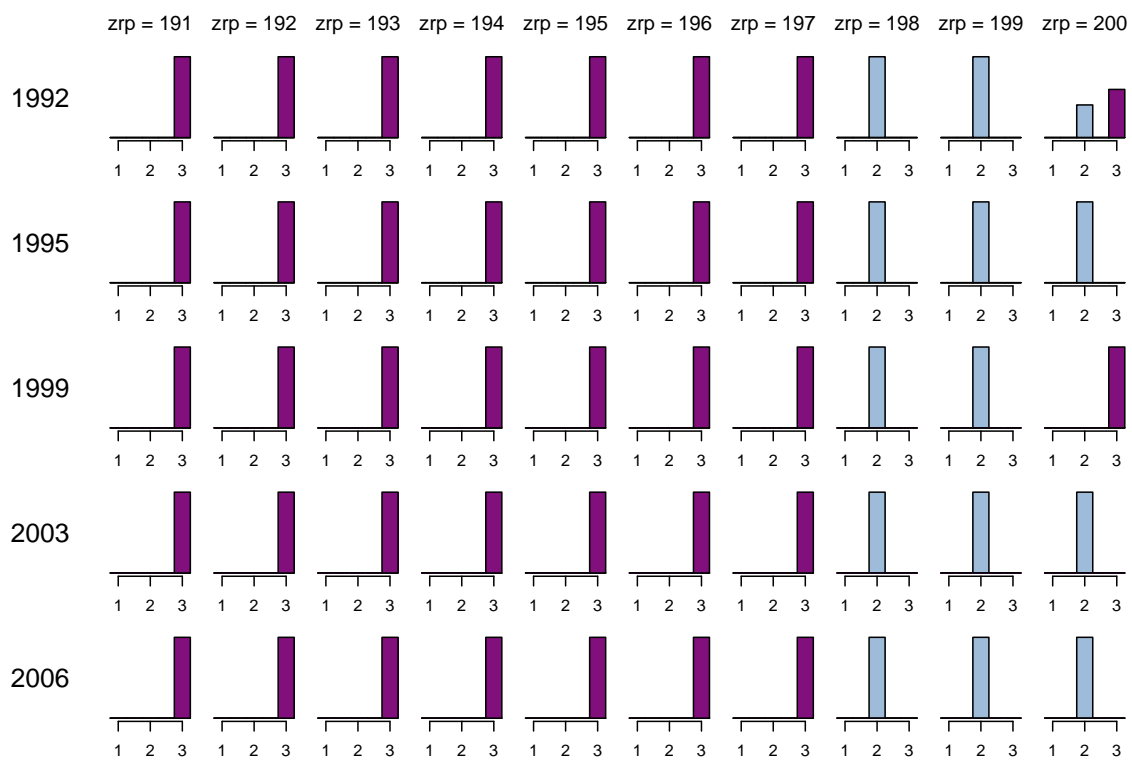
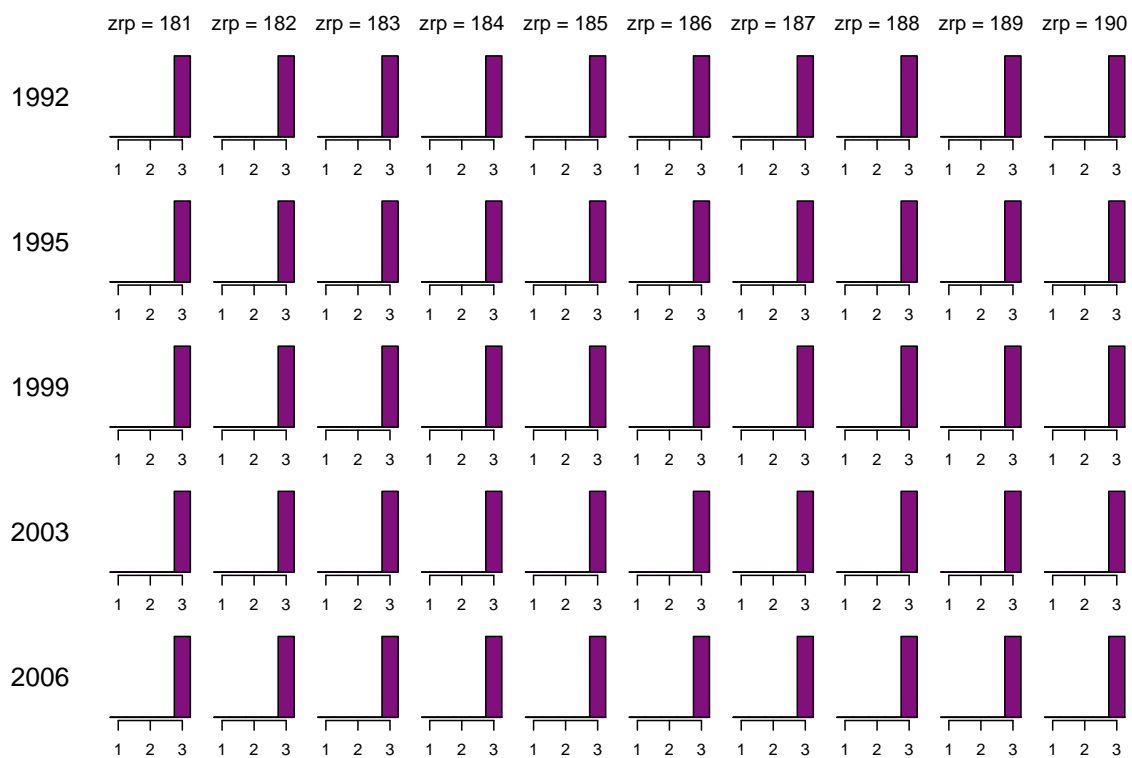


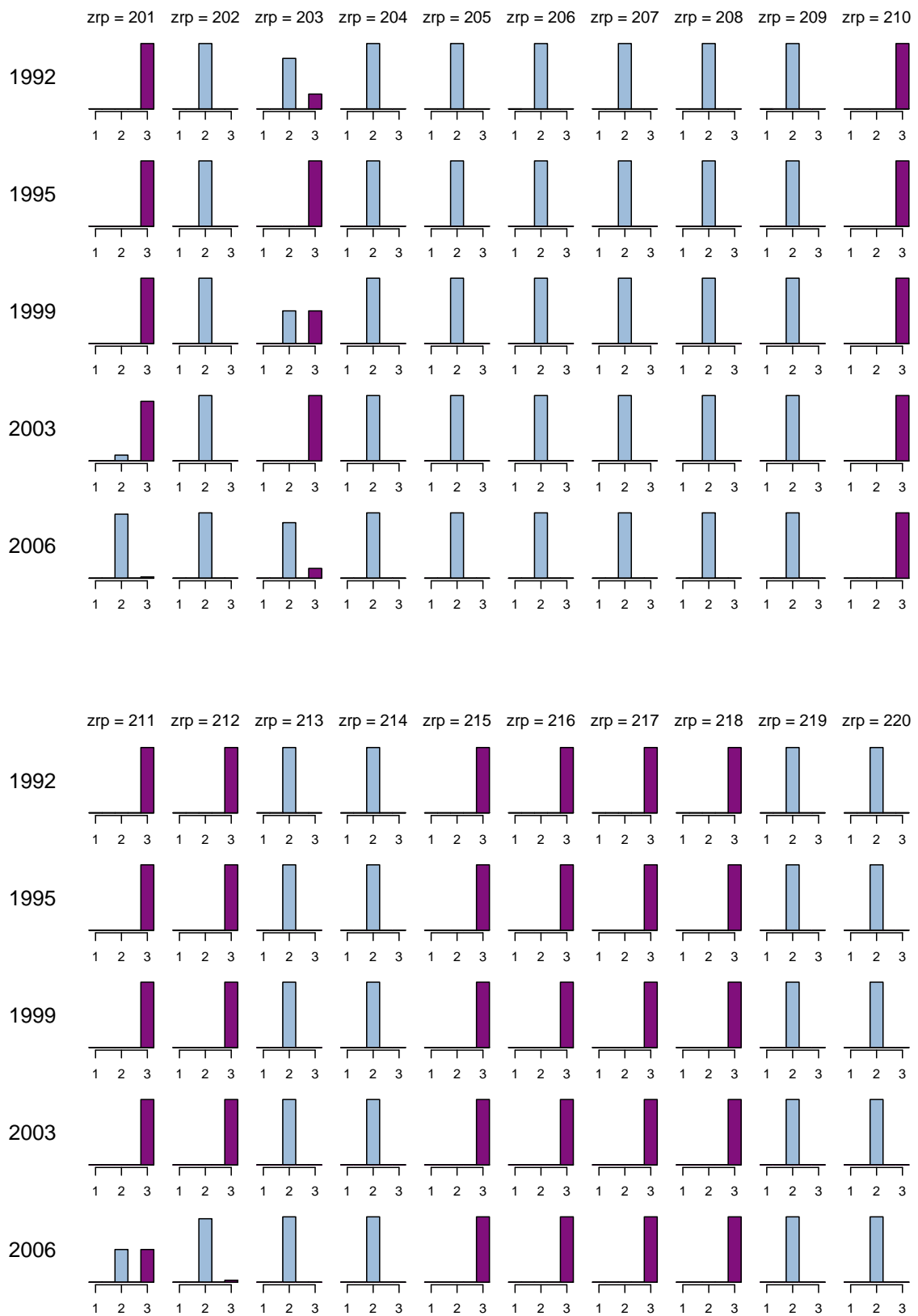


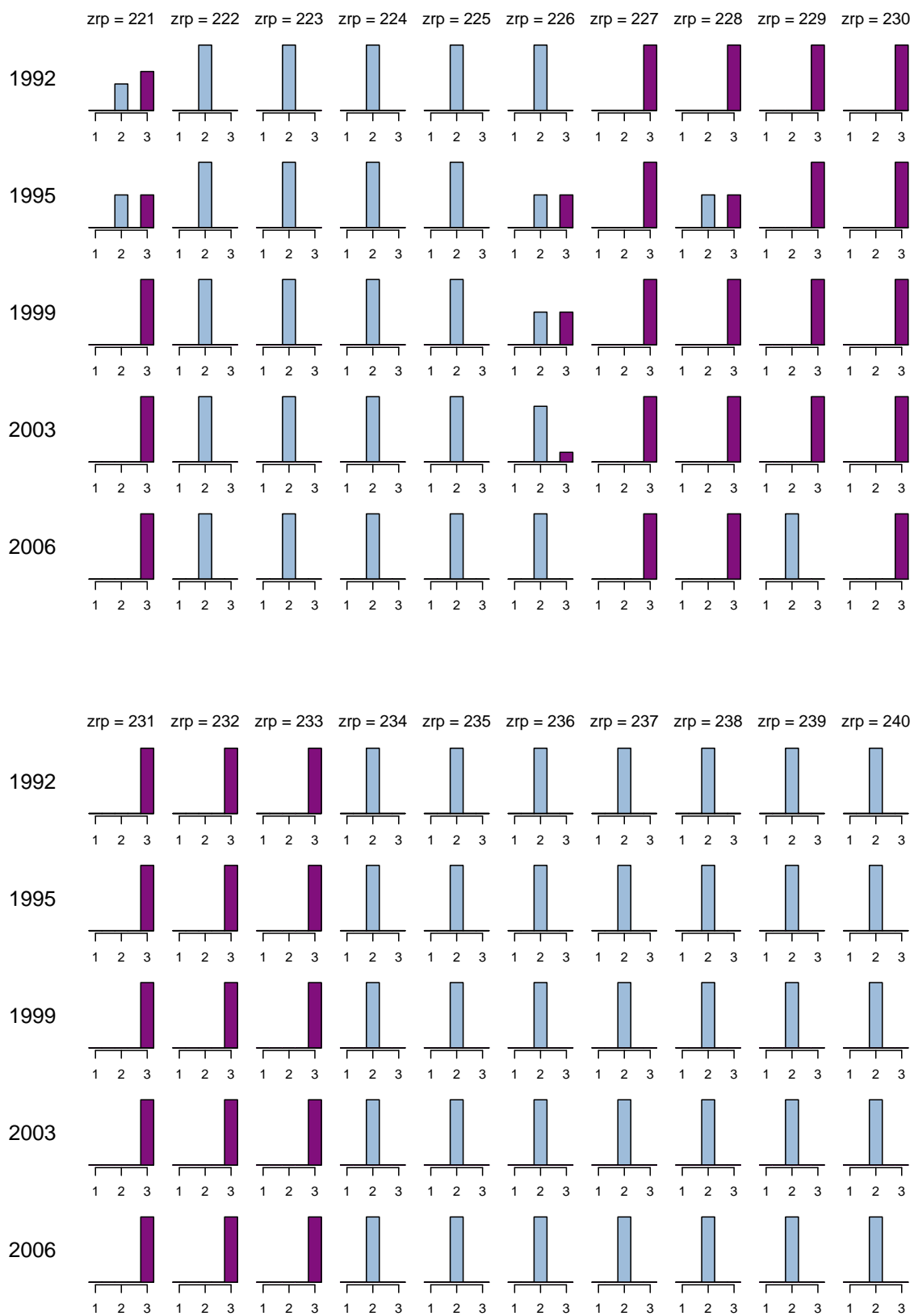


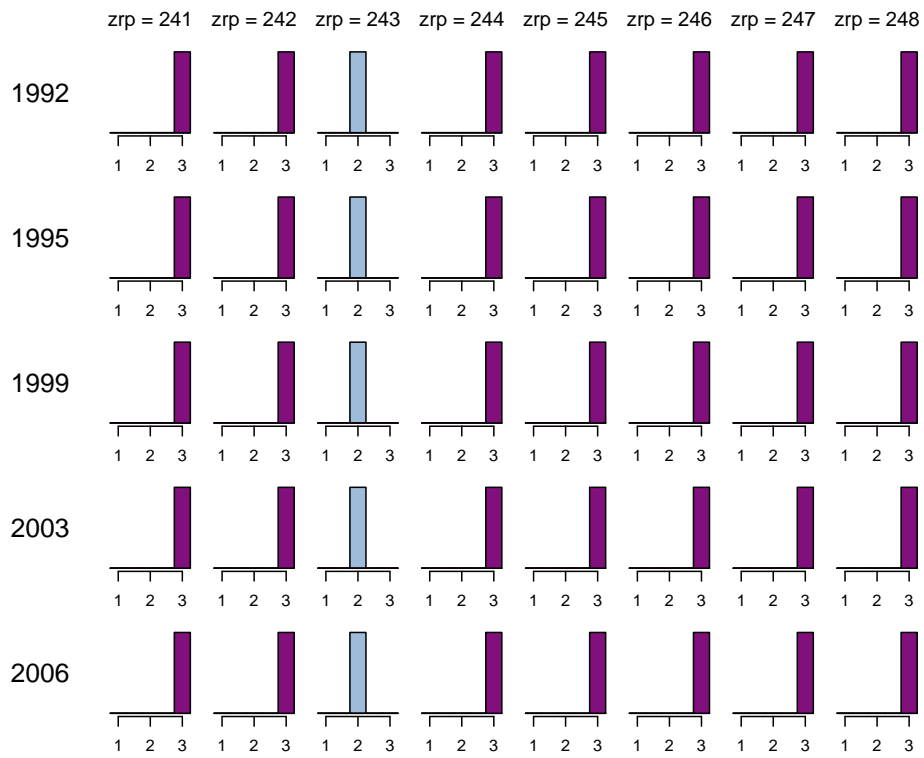




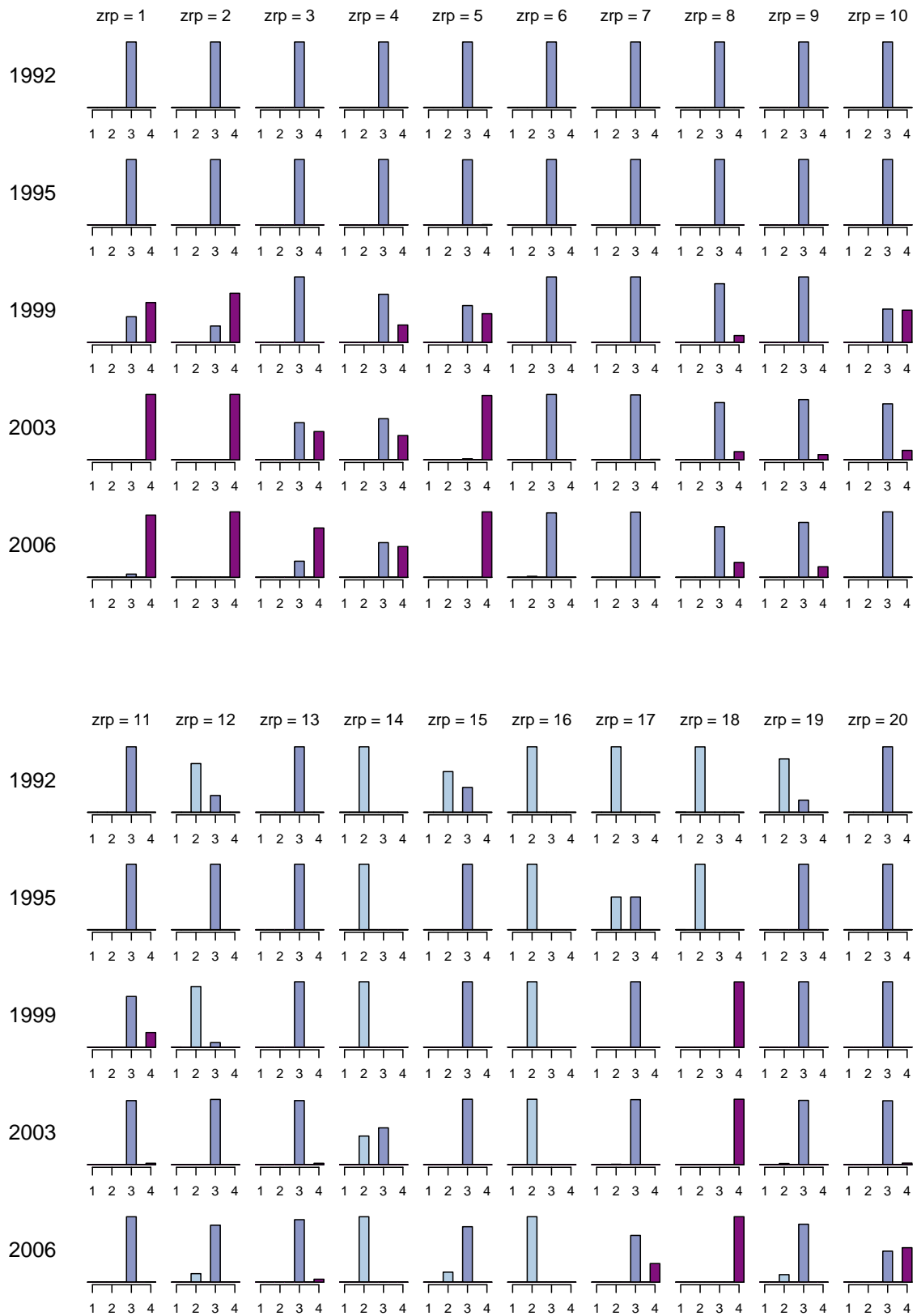


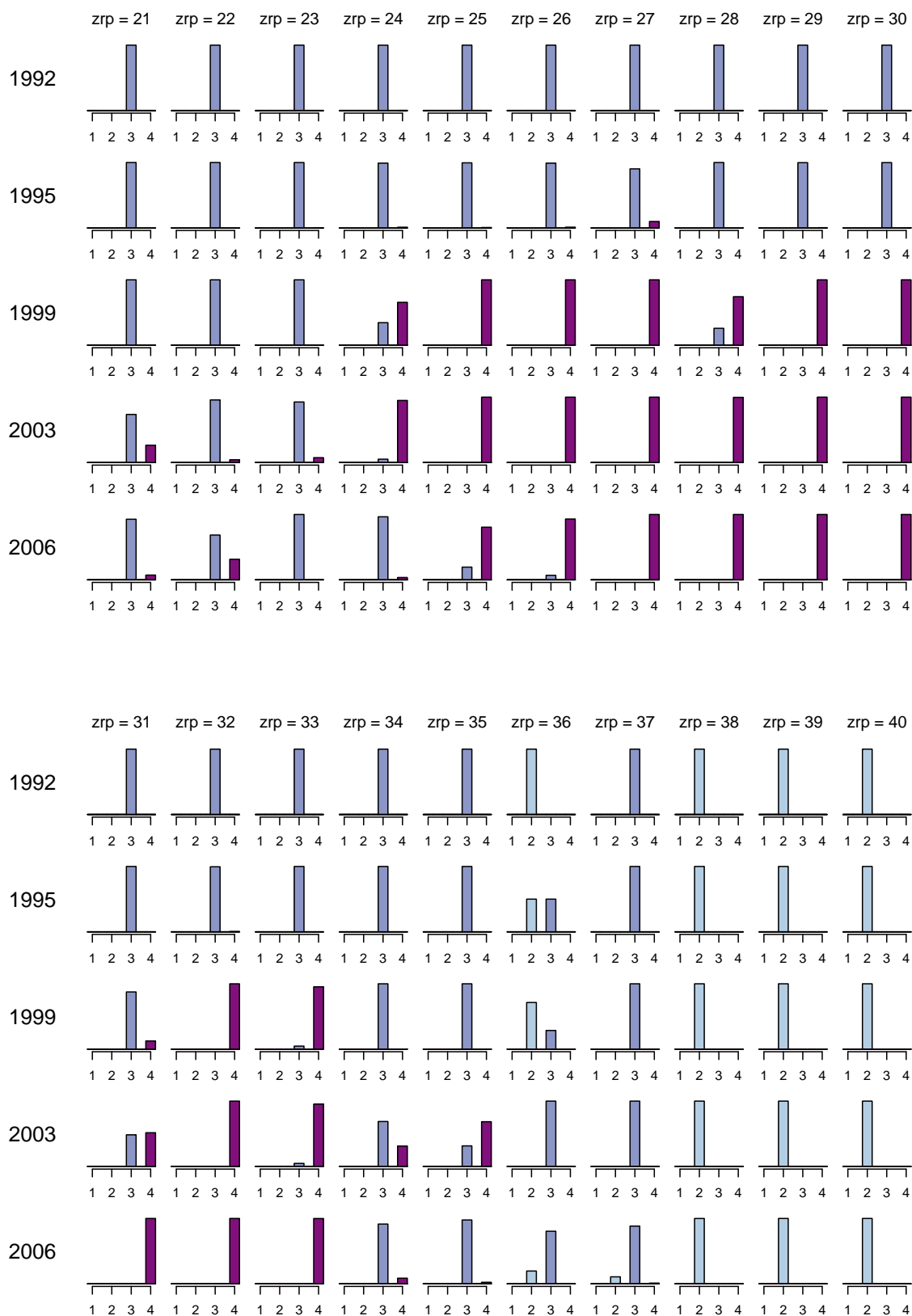


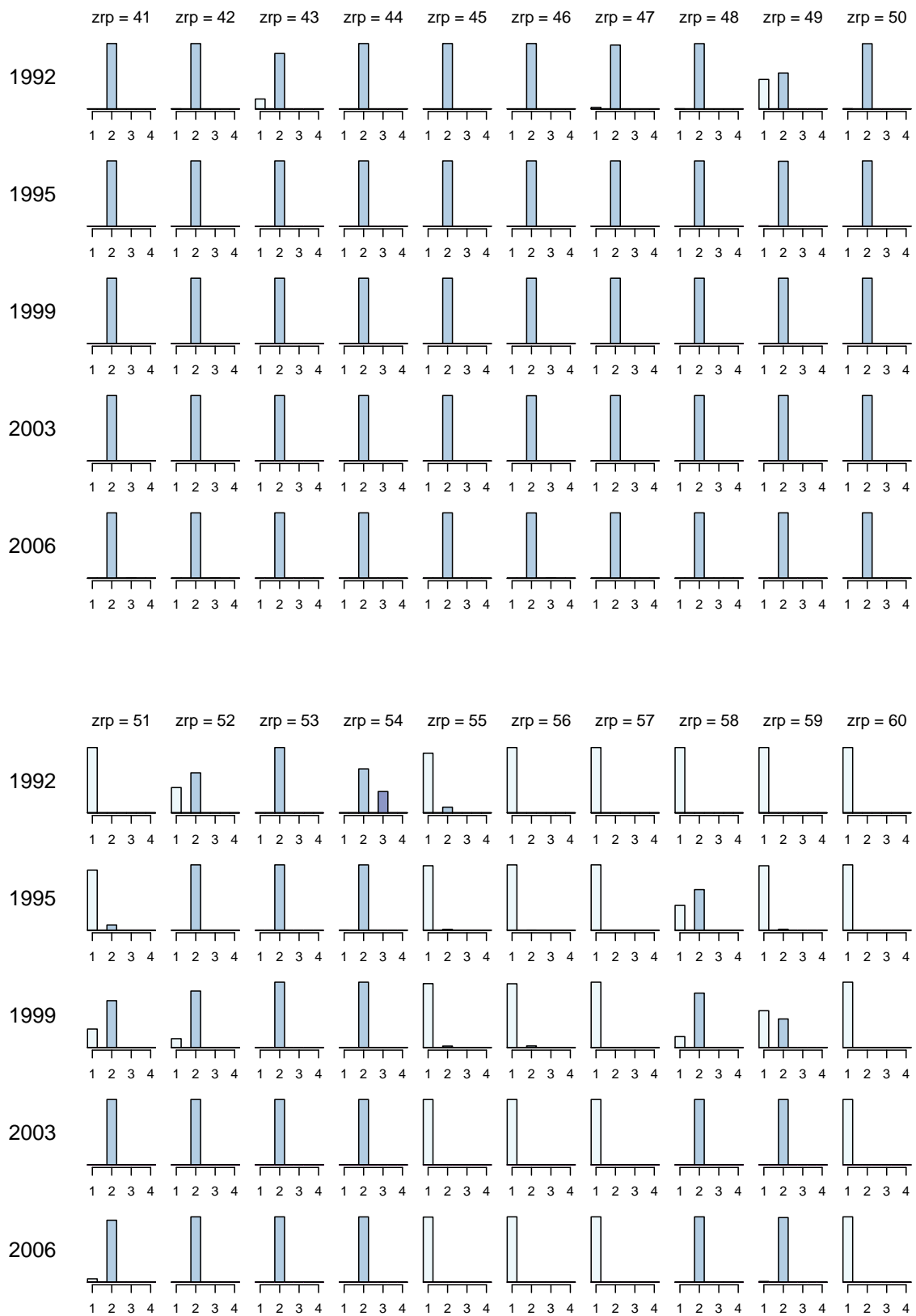




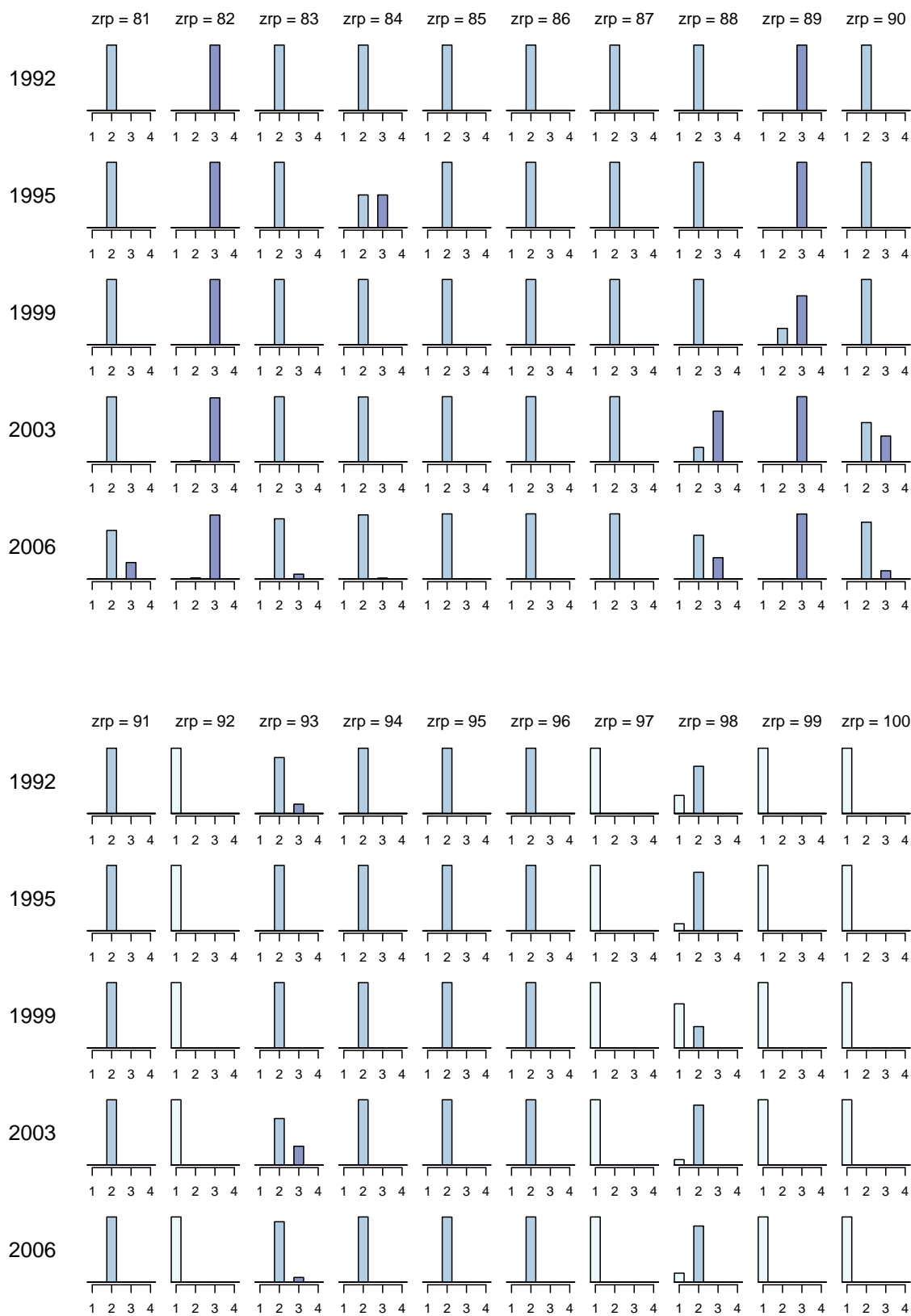
B.2 Distribució a posteriori de ζ_i per al model M_{4J}

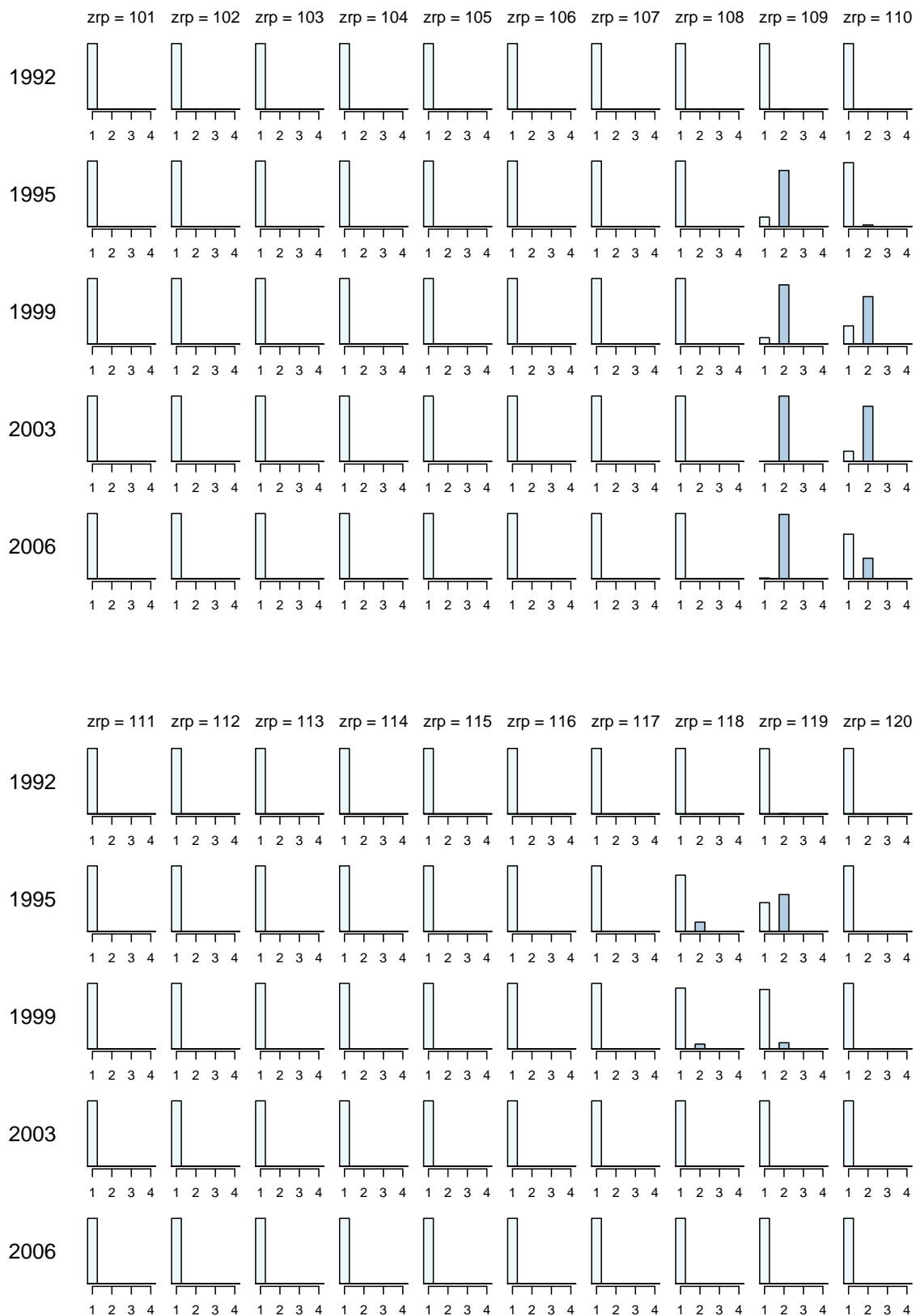


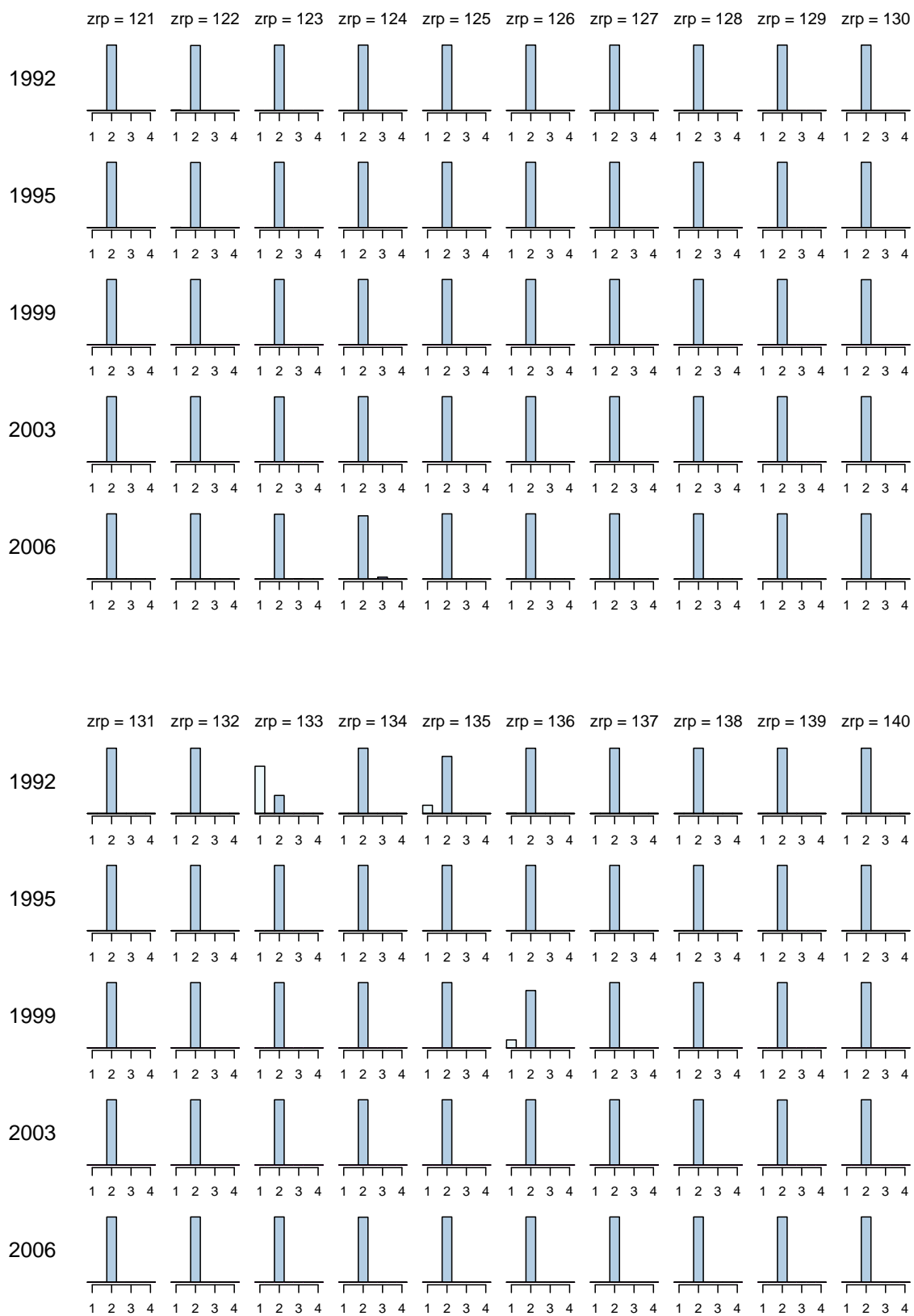




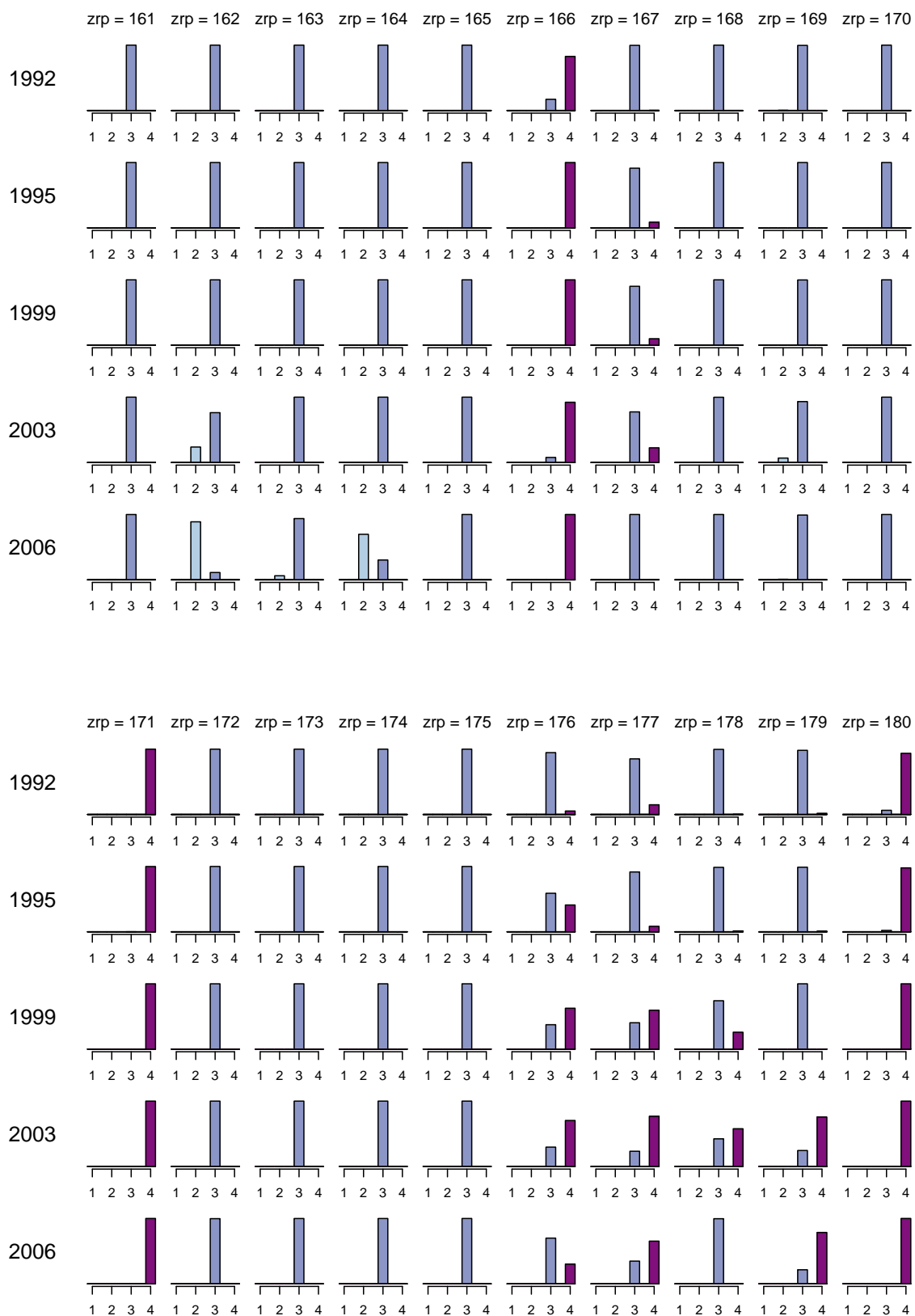




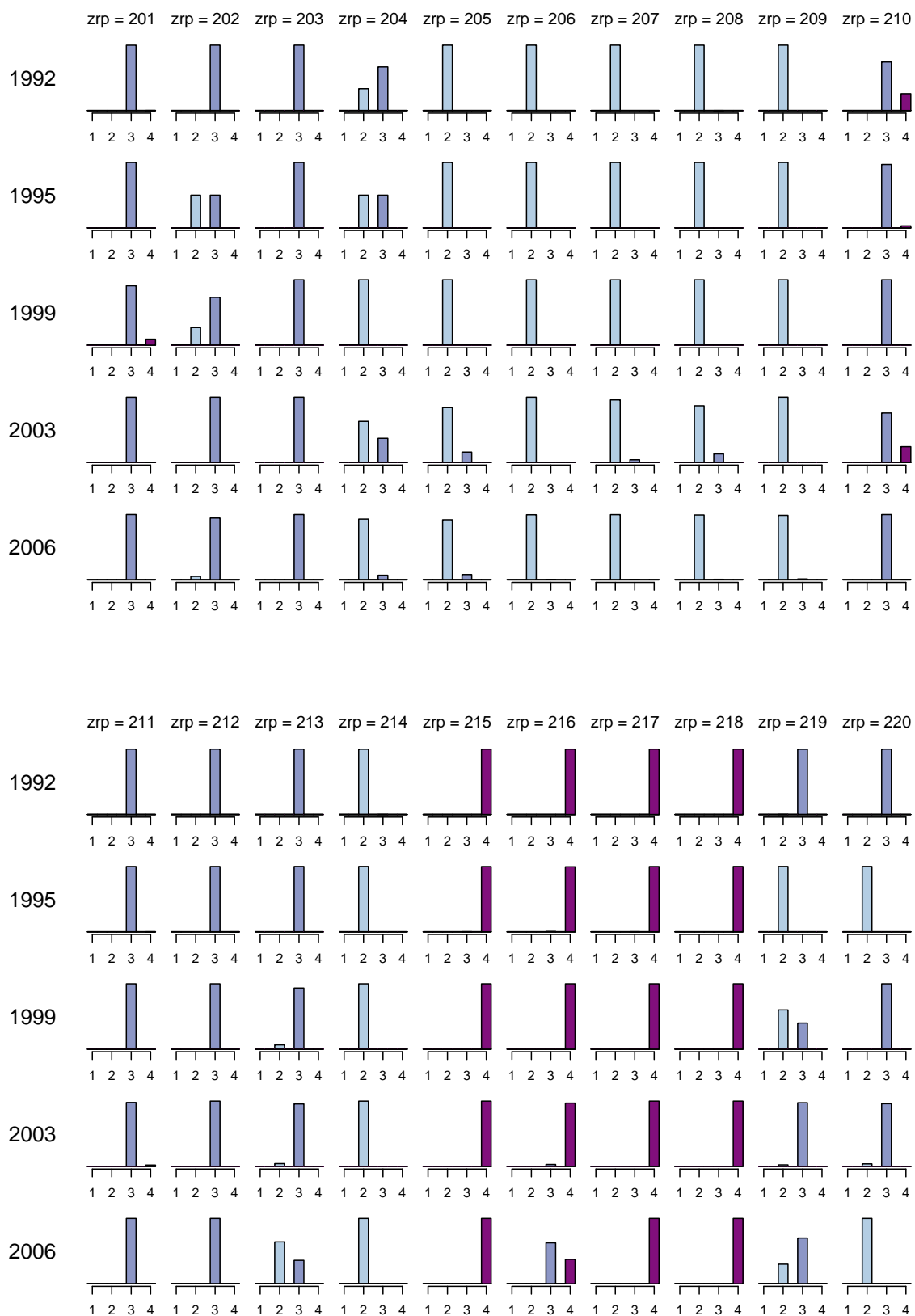


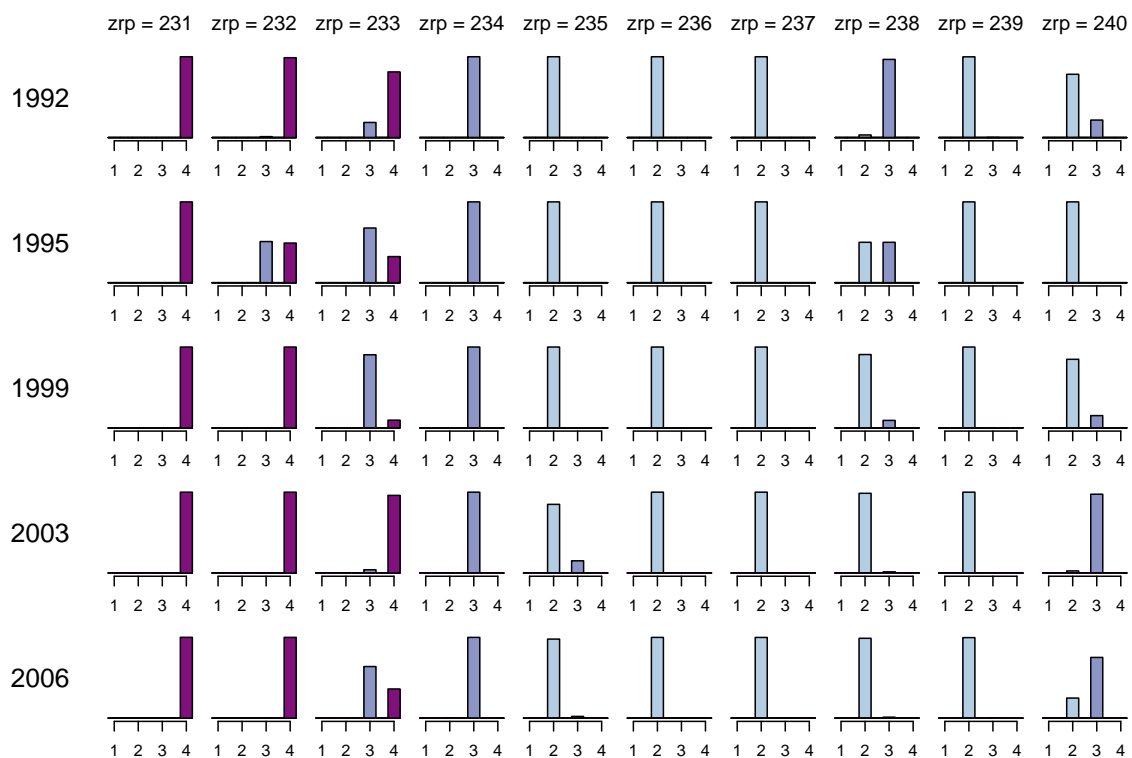
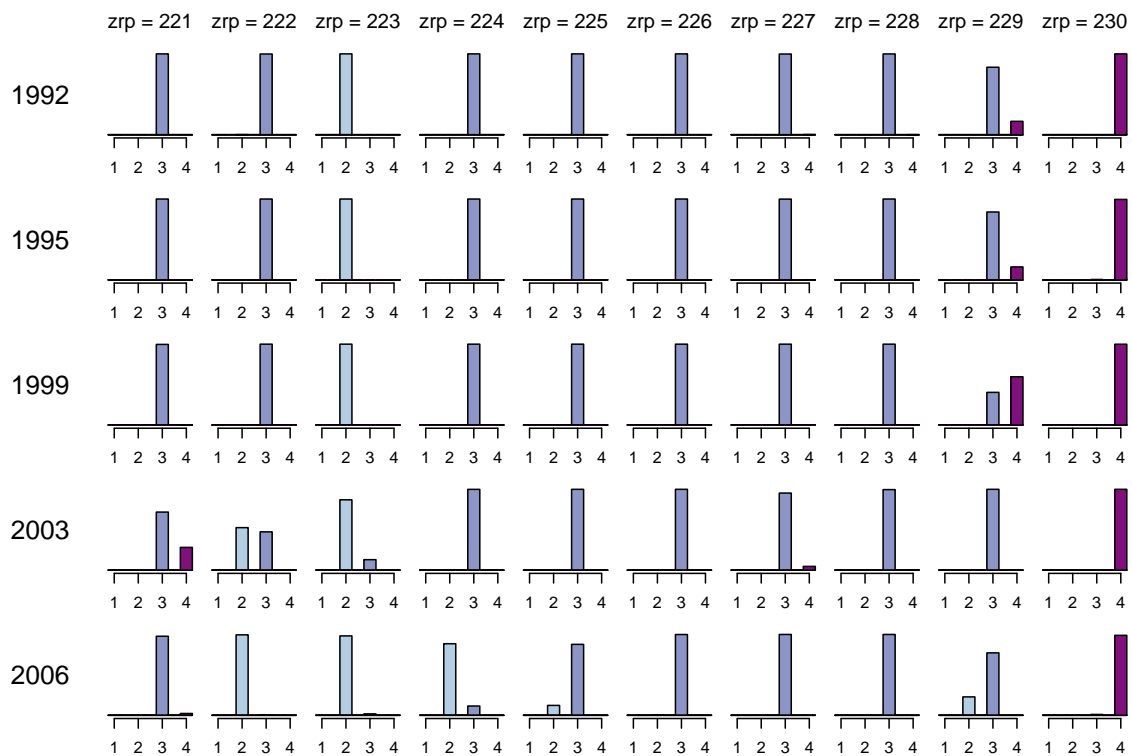


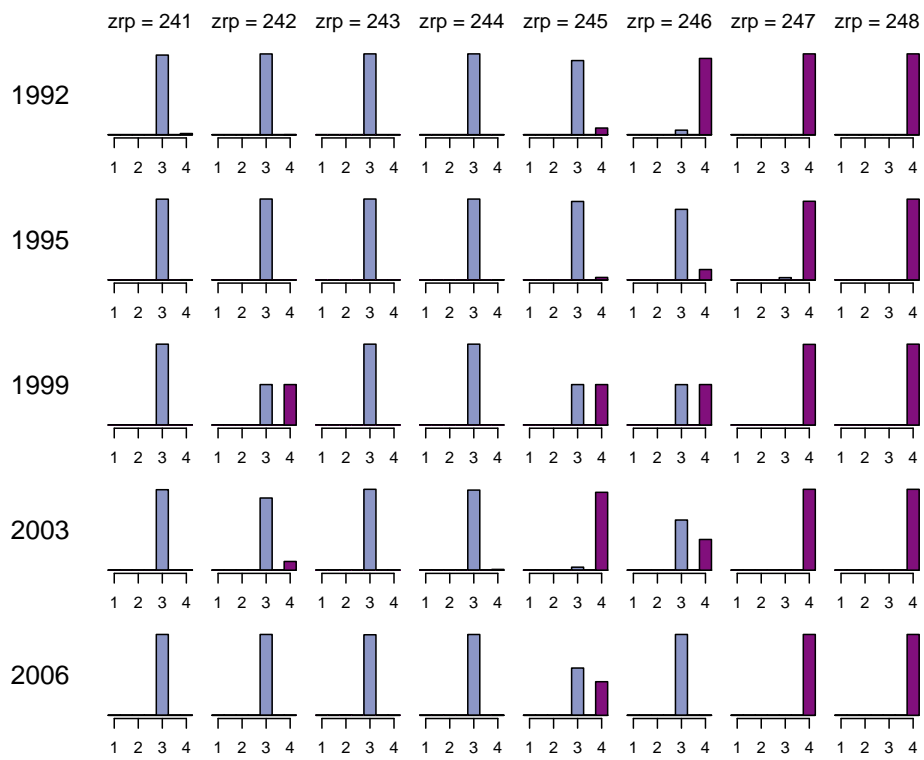












Apèndix C

Anàlisi provisional dels resultats a tot Catalunya

En aquest apèndix presentem resultats preliminars de l'anàlisi cluster de les dades per tot Catalunya basat en els models no jeràrquics presentats al Capítol 7. Això inclourà l'anàlisi per les cinc eleccions basat en models d'un dos, tres i cinc clusters i la presentació de diferents aspectes de la validació d'aquests models, que indiquen que aquests models no capturen prou bé alguns aspectes importants de les dades. Els models que apareixeran en aquest apèndix es denotaran per:

M_1 : model multinomial homogeni d'un sol cluster,

M_2 : model multinomial de dos clusters,

$M_{2 \times 2}$: model que separa els districtes en funció de si el nombre d'electors del municipi al que pertànyen és major o menor de 5000 electors, i assumeix un model com el M_2 per cada conjunt de districtes,

M_3 : model multinomial de tres clusters, i

M_5 : model multinomial de cinc clusters.

També es reflexiona perquè els models jeràrquic que capturaven bé els aspectes fonamentals dels resultats a Barcelona ciutat, no funcionen quan s'analitzen els resultats a nivell de tot Catalunya.

C.1 Anàlisi basada en el model no jeràrquic M_2

A la Figura C.1 trobem per a cada any els mapes de la classificació dels districtes a cada cluster resultat d'ajustar el model de dos clusters, M_2 . És molt notable l'agregació espacial que presenten els clusters, concentrant-se els districtes del cluster 2 en zones de l'àrea metropolitana de Barcelona i de la seva influència, la Vall d'Aran i Terres de l'Ebre. Malgrat que a nivell de territori predominen els districtes del cluster 1, al voltant d'un 80% dels districtes, aquests corresponen generalment als municipis més petits i representen poc més del 50% del total d'electors.

A la Figura C.2 es mostra la classificació a nivell de zrp de Barcelona on també és molt notable l'agregació espacial que presenten els clusters, de manera que en la part central i a ponent predomina el cluster 1 i a la resta el cluster 2. I a la Figura C.3 es presenten la classificació dels districtes de les tres capitals de província, Girona, Lleida, Tarragona, així com Sabadell.

A la Figura C.4 per a cada any trobem representades, mitjançant diagrames de caixes, simulacions de les distribucions a posteriori dels perfils per a cada cluster, així com simulacions de la distribució a posteriori de $\log \frac{\theta_{2j}}{\theta_{1j}}$, $\pi(\log \frac{\theta_{2j}}{\theta_{1j}} | y)$ per $j = 1, \dots, k$, que permet copsar els partits que estan més (o menys) representats en un cluster respecte l'altre, i finalment la columna de la dreta d'aquesta figura hi ha la distribució a posteriori de la proporció de districtes del cluster 1. De forma complementària a la Taula C.1 hi ha els valors esperats dels perfils així com del tamany relatiu dels districtes dels clusters per a cada any. La proporció de districtes pertanyents al cluster 1 manifesta una tendència a disminuir, essent el valor esperat de la proporció de districtes pertanyents al cluster 1 per a l'any 2006 del 78% mentre que l'any 1992 era del 83%.

Si bé al mapa de classificació de la Figura C.1 hi costen de veure, grans canvis pel que fa a la distribució geogràfica dels dos clusters, la distribució dels perfils han variat considerablement al llarg dels diferents comicis. Mirant la distribució de $\pi(\log \frac{\theta_{2j}}{\theta_{1j}} | y)$ a la Figura C.4, destaca que l'any 1992 CIU, PPC, ERC i $b+n$ estaven notablement més representats en el cluster 1, especialment CIU i ERC, mentre que en el cluster 2 ho estaven el PSC, ICV, *altres* i l'abstenció. I l'any 2006 tant el PPC com ICV es troben més representats en el cluster oposat al que ho estaven l'any 1992, i aquest canvi ha estat gradual.

A la Figura C.5 observem per a cada partit, així com per a l'abstenció, l'evolució dels perfils de cada cluster i l'ombrejat emfasitza la distància entre els dos clusters. S'observa com CIU, tot i estar sempre més representat en el cluster 1 la distància entre ells ha

	cluster	CIU	PSC	PPC	ICV	ERC	Cs	altres	b+n	abs	$E[\omega y]$
1992	1	0.33	0.12	0.04	0.03	0.06	-	0.02	0.01	0.39	0.83
	2	0.14	0.20	0.03	0.05	0.02	-	0.03	0.01	0.53	
1995	1	0.33	0.13	0.09	0.04	0.08	-	0.00	0.01	0.32	0.83
	2	0.17	0.20	0.08	0.08	0.03	-	0.01	0.01	0.42	
1999	1	0.30	0.18	0.06	0.02	0.07	-	0.01	0.01	0.36	0.80
	2	0.15	0.27	0.06	0.03	0.03	-	0.01	0.01	0.45	
2003	1	0.26	0.16	0.07	0.04	0.14	-	0.01	0.01	0.32	0.79
	2	0.13	0.24	0.08	0.05	0.06	-	0.01	0.01	0.42	
2006	1	0.23	0.13	0.06	0.06	0.10	0.02	0.01	0.02	0.38	0.78
	2	0.12	0.18	0.06	0.05	0.05	0.02	0.01	0.01	0.49	

Taula C.1: Esperança a posteriori dels perfils de probabilitat per al model M_2 , $E[\theta_{1j}|y]$ i $E[\theta_{2j}|y]$ per $j = 1, \dots, k$, i el tamany relatiu de cada cluster, $E[\omega|y]$, per a cada una de les cinc eleccions.

disminuït lleugerament, quelcom semblant passa en el PSC especialment els darrers anys però amb la variant que està més representat en el cluster 2. És interessant el cas de ICV en que els primers anys estava més representat al cluster 2, i el darrer en el cluster 1 i el PPC a qui li succeeix el mateix canviant però les etiquetes dels clusters. Per la seva banda a les tres primeres eleccions ERC mostra perfils dels grups equidistants mentre aquests es distancien l'any 2003.

Amb l'objectiu de validar el model representem a la Figura C.8 el valor esperat a posteriori de la mesura de discrepància D_{2i}^2 a nivell de zrp, $E[D_{2i}^2|y]$, en funció del nombre d'electors del districte i observem una clara correlació positiva. A la dreta de la mateixa figura hi representem $E[D_{2i}^2|y]$ per a cada districte versus el nombre d'electors del municipi al que pertanyen.

Mitjançant simulacions de la predictiva a posteriori per D_{2i}^2 donat el model que estem ajustant trobem que aquests haurien de prendre valors al voltant de 8, i al 2006 al voltant de 9, fet que coincideix amb el nombre de perfils menys ú, $k - 1$. El fet d'observar valors de $E[D_{2i}^2|y]$ tant grans, assenyala mancances en el model utilitzat en el sentit que hi ha més variabilitat de l'esperada pel model. Això en un model de regressió clàssic portaria a pensar que falten covariables, però en el nostre cas on l'objectiu no és identificar les covariables importants hem de reformular el model per arribar-ne a un que capturi aquesta sobredispersió per tal de que la inferència que fem sigui correcta, ja sigui per exemple definint més de dos clusters i/o definint un model amb més variabilitat intrínseca via models jeràrquics i/o modelant el fet que les dades tinguin una posició a l'espai mitjançant un model amb dependència espacial. Els valors alts de D_{2i}^2 així com

la seva correlació amb el nombre d'electors posen de manifest mancances en el model.

Donat que la validació indica que el model no és bo cal reformular o estendre el model, en aquest sentit a continuació presentem resultats de diferents models alternatius per a l'any 2003.

C.2 Anàlisi basada en els models no jeràrquics: $M_{2 \times 2}$, M_3 i M_5

Presentem primer part dels resultats del model $M_{2 \times 2}$ fruit d'estratificar l'anàlisi per als municipis de menys de 5000 electors i els de més, i ajustar a cada estrat el model M_2 , els resums dels perfils per a l'any 2003 es presenten a la Taula C.2 on s'observa com el cluster 2 dels petits és molt similar al cluster 1 dels grans. Aquest fet podria indicar que en lloc de dos cluster n'hi ha 3. D'altra banda també el cluster 2 dels grans és similar al cluster 2 del model sense estratificar que observàvem a la Taula C.1.

A la Taula C.3 es mostra l'esperança a posteriori per als perfils dels models de 3 i 5 clusters, M_3 i M_5 , així com la proporció a posteriori de districtes a cada cluster. Primer de tot cal comentar que a l'executar les cadenes per al model de 3 clusters competeixen dues estructures és a dir la meitat de les vegades el model convergeix cap a l'estructura que hem anomenat M_{3a} i l'altre a la M_{3b} .

Aquest fet és indicatiu de que ambdues estructures són altament versemblants. La principal diferència radica en que zrp dels districtes de Sarrià-Sant Gervasi i Les Corts les inclou al cluster 2 o al 1 variant lleugerament els perfils. En canvi en el model amb 5 clusters, M_5 , aquestes zrp aconsegueixen tenir un cluster per a ells sols (el cluster 2). En els perfils dels diferents clusters, tant per al model M_3 com per al M_5 s'observa una gradació en la que si ordenem els clusters en ordre decreixent per la magnitud del perfil de CIU, tal i com ho està a la Taula C.3, llavors ERC també apareix raonablement ordenada en ordre decreixent, mentre que PSC i l'abstenció apareixen en ordre creixent. Aquesta gradació de perfils també apareix ordenada a l'espai, tal i com indica la Figura C.9, en el sentit de que el canvi d'un cluster a un altre es fa gradualment seguint aquest ordre a mesura que et desplaces pel territori.

		cluster	CIU	PSC	PPC	ICV	ERC	altres	b+n	abs	$E[\omega y]$
2003	petits	1	0.34	0.11	0.04	0.03	0.19	0.01	0.01	0.27	0.72
		2	0.24	0.18	0.06	0.03	0.12	0.01	0.01	0.34	
	grans	1	0.23	0.17	0.08	0.05	0.13	0.01	0.01	0.33	0.63
		2	0.12	0.24	0.08	0.05	0.06	0.01	0.01	0.43	

Taula C.2: Esperança a posteriori dels perfils de probabilitat i del tamany relatiu de cada cluster pel model $M_{2 \times 2}$, quan s’analitza separant els districtes de pobles de menys de 5000 habitants de districtes de pobles de més de 5000 habitants, per l’any 2003.

C.3 Comparació de models no jeràrquics

La Figura C.10 presenta per a cada un dels models considerats l’esperança a posteriori de D_{2i}^2 , $E[D_{2i}^2|y]$, on s’observa com per a tots ells es manté la dependència amb el nombre d’electors del districte i lo allunyats que estan del valor esperat en cas de ser bo el model, que hauria de ser al voltant de $\approx k - 1$. Al mateix gràfic, a la dreta, es presenta un altre aspecte de la validació del model fruit de representar en un histograma els valors obtinguts de $p_B^{D_{2i}^2}$, i cap dels models és capaç de capturar la dispersió de les observacions. A la Figura C.11 hi trobem representats amb histogrames la distribució dels $p_B^{D_{2i}^2}$ estratificada per opció de vot per al model de dos clusters amb les dades de l’any 2003. Aquest mateix patró es repeteix per a tots els models i ressalta el fet que totes les categories presenten major variabilitat que la que és capaç de capturar el model, si bé pel cas de la categoria d’*altres* i blancs i nuls la desviació de la uniformitat és més moderada.

A la Taula C.4 es mostra per a cadascun dels diferents models mesures resum de la bondat d’ajust, $E[D_1(y, \theta)|y]$ i $E[D_2(y, \theta)|y]$. Al passar d’un a dos clusters reduïm dràsticament el valor esperat a posteriori de les mesures de discrepància. A l’augmentar el nombre de clusters a més de dos també reduïm, tot i que menys dràsticament, els valors de les mesures de discrepància. Mitjançant simulacions de les predictives a posteriori ens adonem que si el model fos bo, llavors $E[D_2(y, \theta)|y] \approx (k - 1) * n = 7 * 1447 = 10129$ que està molt lluny dels valors obtinguts. El fet que la deviança es redueixi tant brutalment per cada cluster afegit, essent aquesta reducció molt superior al nombre de paràmetres afegits, fa que de ben segur que si fóssim capaços de calcular una mesura com ara el DIC ens portaria a escollir un model amb a un nombre molt elevat de clusters.

Finalment dir que hem calculat el *p-valor* bayesià associat a $D_2(y, \theta)$, $p_B^{D_2}$, definit al capítol 8, i en tots els models el seu valor calculat a estat 0, indicant que les dades presenten una variabilitat molt superior de la que són capaços de de capturar els models utilitzats.

Model	cluster	CIU	PSC	PPC	ICV	ERC	altres	b+n	abs	$E[\omega_r y]$
M_{3a}	1	0.31	0.13	0.05	0.03	0.17	0.01	0.01	0.29	0.52
	2	0.21	0.18	0.08	0.05	0.11	0.01	0.01	0.35	0.33
	3	0.11	0.25	0.08	0.05	0.05	0.01	0.01	0.43	0.15
M_{3b}	1	0.30	0.13	0.07	0.04	0.16	0.01	0.01	0.29	0.59
	2	0.20	0.19	0.08	0.05	0.11	0.01	0.01	0.36	0.27
	3	0.11	0.25	0.08	0.05	0.05	0.01	0.01	0.43	0.14
M_5	1	0.32	0.12	0.04	0.03	0.18	0.01	0.01	0.28	0.48
	2	0.28	0.14	0.14	0.04	0.11	0.01	0.01	0.27	0.06
	3	0.22	0.18	0.07	0.05	0.13	0.01	0.01	0.34	0.25
	4	0.15	0.22	0.08	0.05	0.08	0.01	0.01	0.40	0.14
	5	0.09	0.27	0.08	0.05	0.04	0.01	0.01	0.45	0.08

Taula C.3: Esperança a posteriori dels perfils per a cada cluster, $E[\theta_{r,j}|y]$, i la grandària relativa de cada cluster, $E[\omega_r|y]$, per als models no jeràrquics M_3 i M_5 per a les dades de tot Catalunya de l'any 2003.

Model	$E[D_1(y, \theta) y]$	$E[D_2(y, \theta) y]$
M_1	609831.2	552949.2
M_2	339342.0	289024.4
M_{2x2}	292843.5	241067.8
M_{3a}	272948.4	220505.9
M_{3b}	273756.4	224013.3
M_5	215807.1	159245.0

Taula C.4: Mesures de bondat d'ajust per als diferents models no jeràrquics considerats per les dades de tot Catalunya de l'any 2003.

C.4 Extensions

Per tal d'estendre el model a fi i efecte que capturi prou be les característiques de les dades més importants proposem considerar que el procés generador de les dades no es correspon a una *Multinomial* si no per exemple a una distribució amb més variabilitat com ara la *Multinomial – Dirichlet*, és a dir formular un Model jeràrquic com el Model 6.2. No obstant la forta correlació observada entre el valor esperat a posteriori de la mesura de discrepància D_{2i}^2 i el nombre d'electors assenyalen que a diferència del que passa a Barcelona ciutat, a nivell de Catalunya els districtes no són intercanviables i a la pràctica l'estimació dels perfils no estarà ponderada per la grandària dels districtes. Una possible manera de modelar l'efecte de la grandària dels districtes serà introduint-la com a covariable en un model multinomial logístic (Leonard, 1994). També ens plan-

tejem definir un error aleatori amb estructura espacial, adaptant la idea del model de Besag, York i Mollie (1991). Posteriorment es podria calcular l'índex *Moran* i analitzar l'existència de correlació espacial en la part no explicada pels diferents models, tal i com hem fet per al cas de Barcelona.

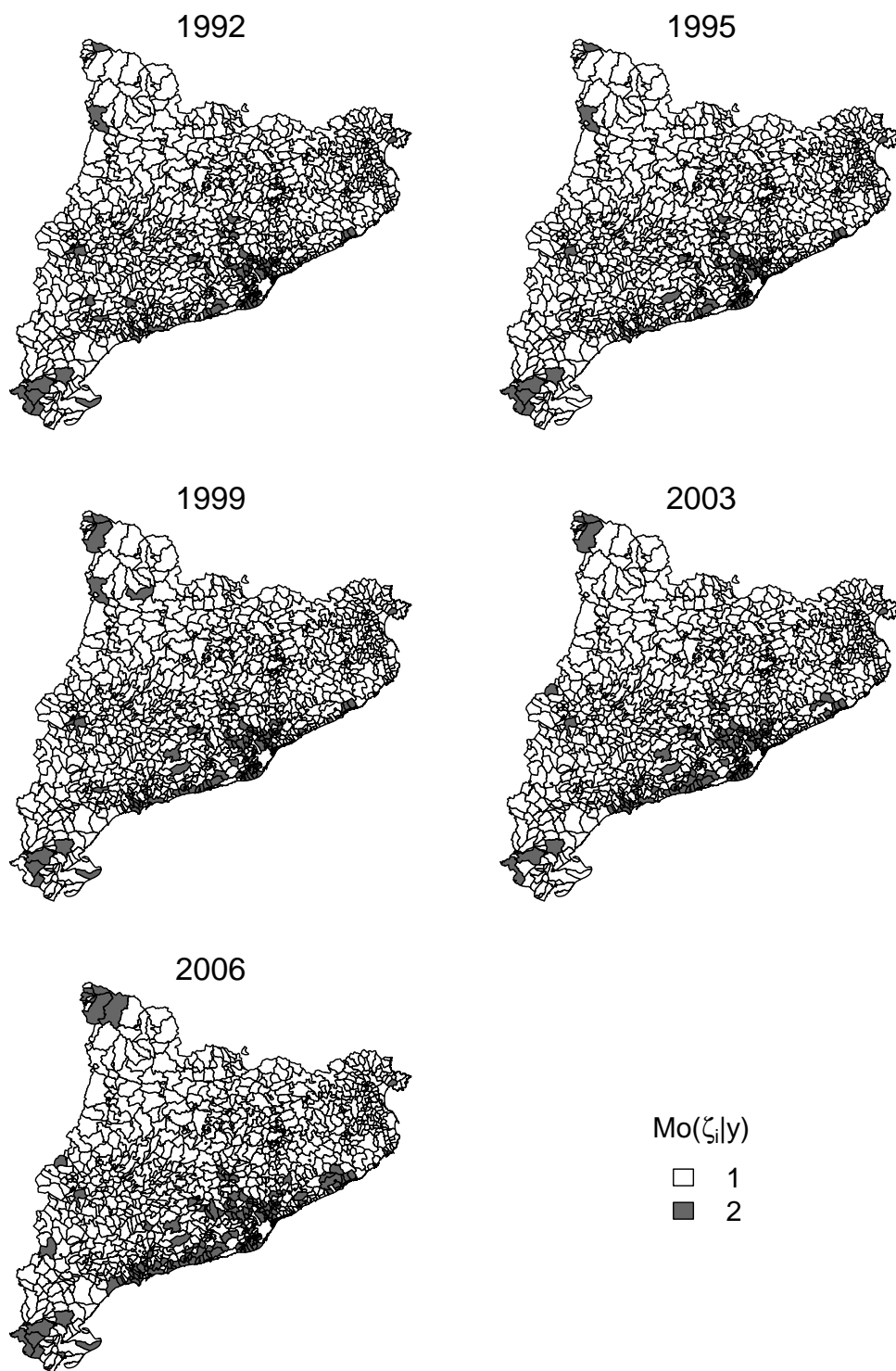


Figura C.1: Classificació dels districtes a cada un dels dos clusters per al model no jeràrquic M_2 , basada en la moda a posteriori de les variables latents ζ_i per als resultats a les eleccions al Parlament de Catalunya del 1992 al 2006.

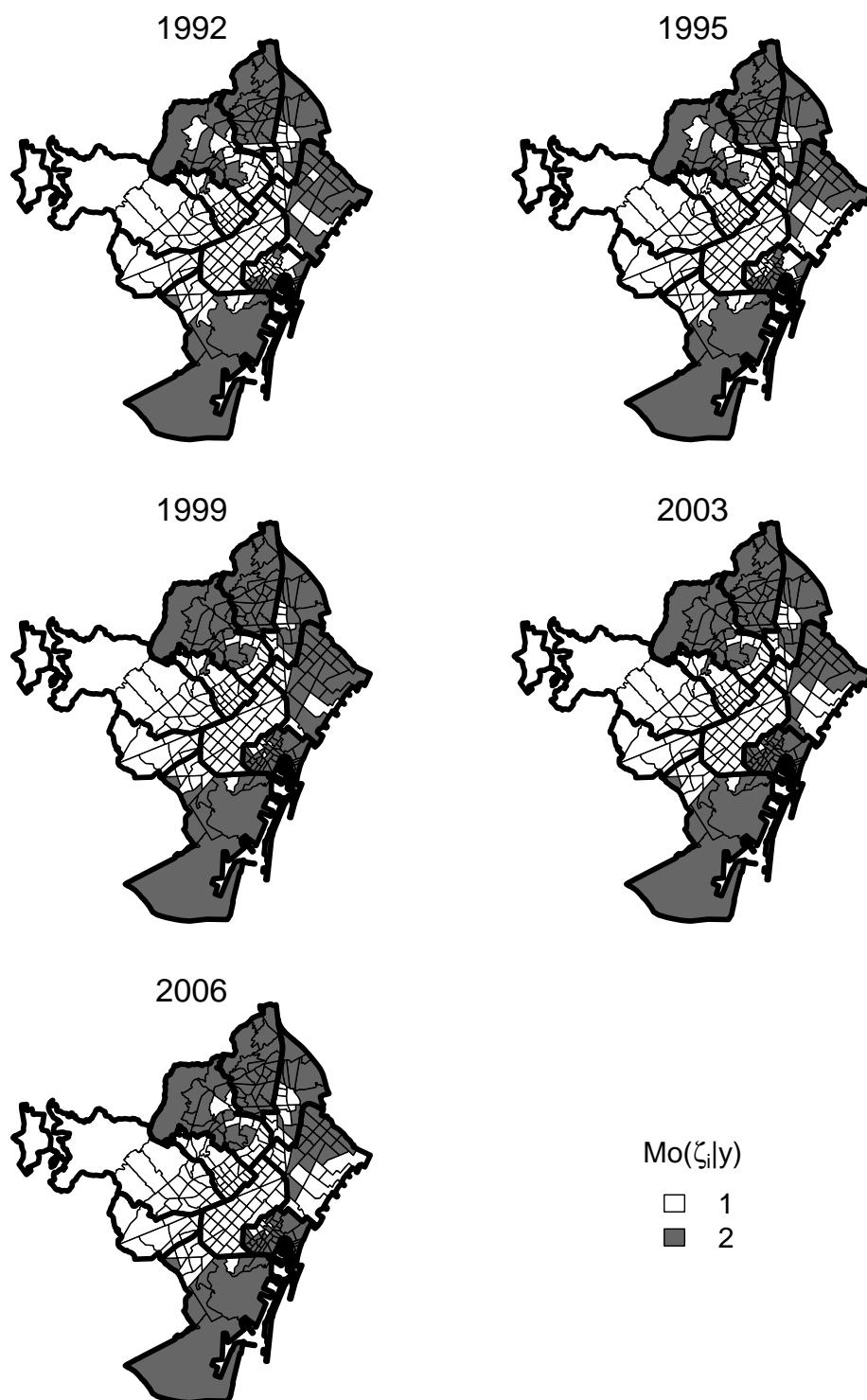


Figura C.2: Classificació dels districtes a cada un dels dos clusters per al model no jeràrquic M_2 basada en la moda a posteriori de les variables latents ζ_i per als resultats a les eleccions del 1992 al 2006 a la ciutat de Barcelona.

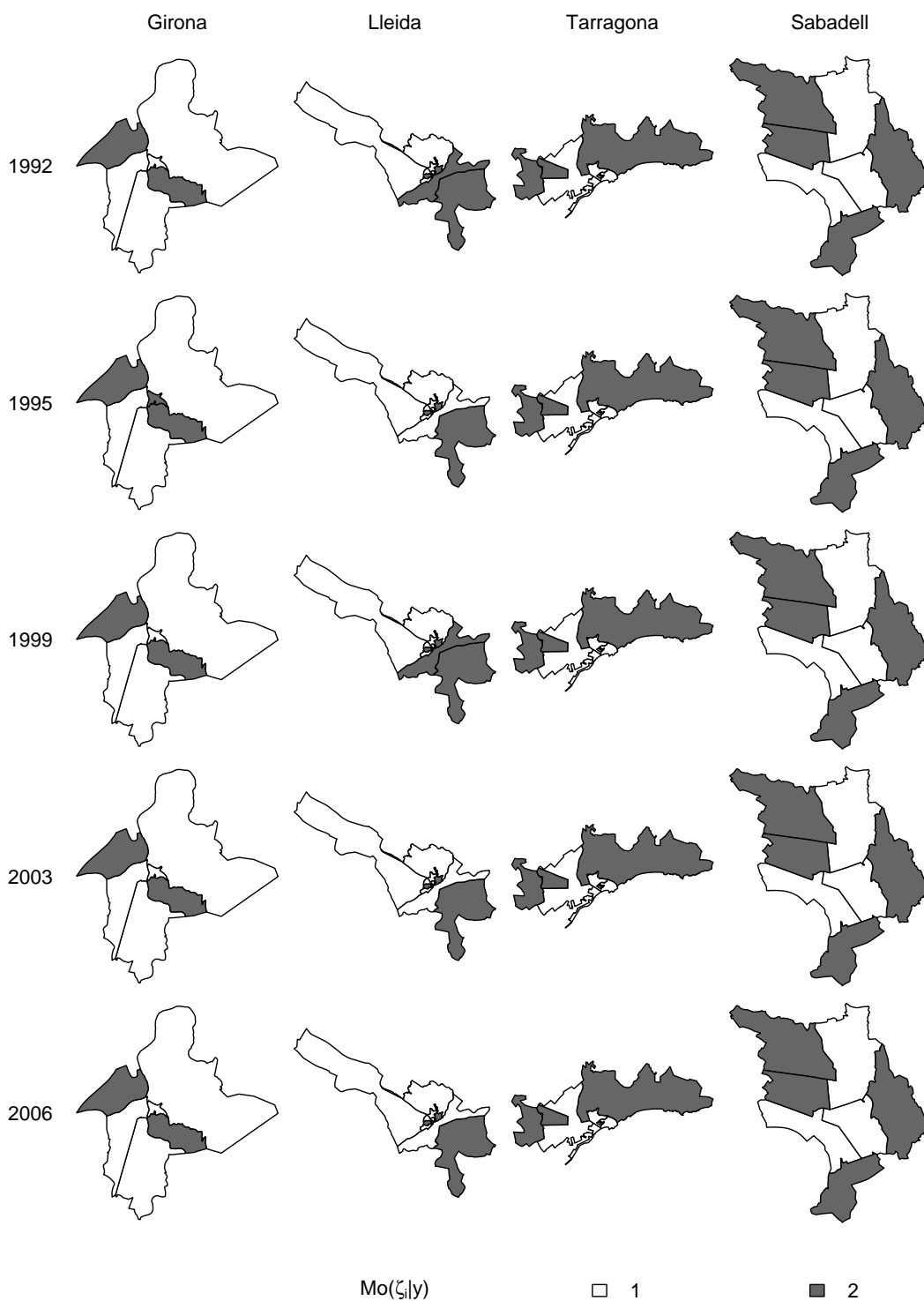


Figura C.3: Classificació dels districtes de les ciutats de Girona, Lleida, Tarragona i Sabadell a cada un dels dos clusters per al model no jeràrquic M_2 .

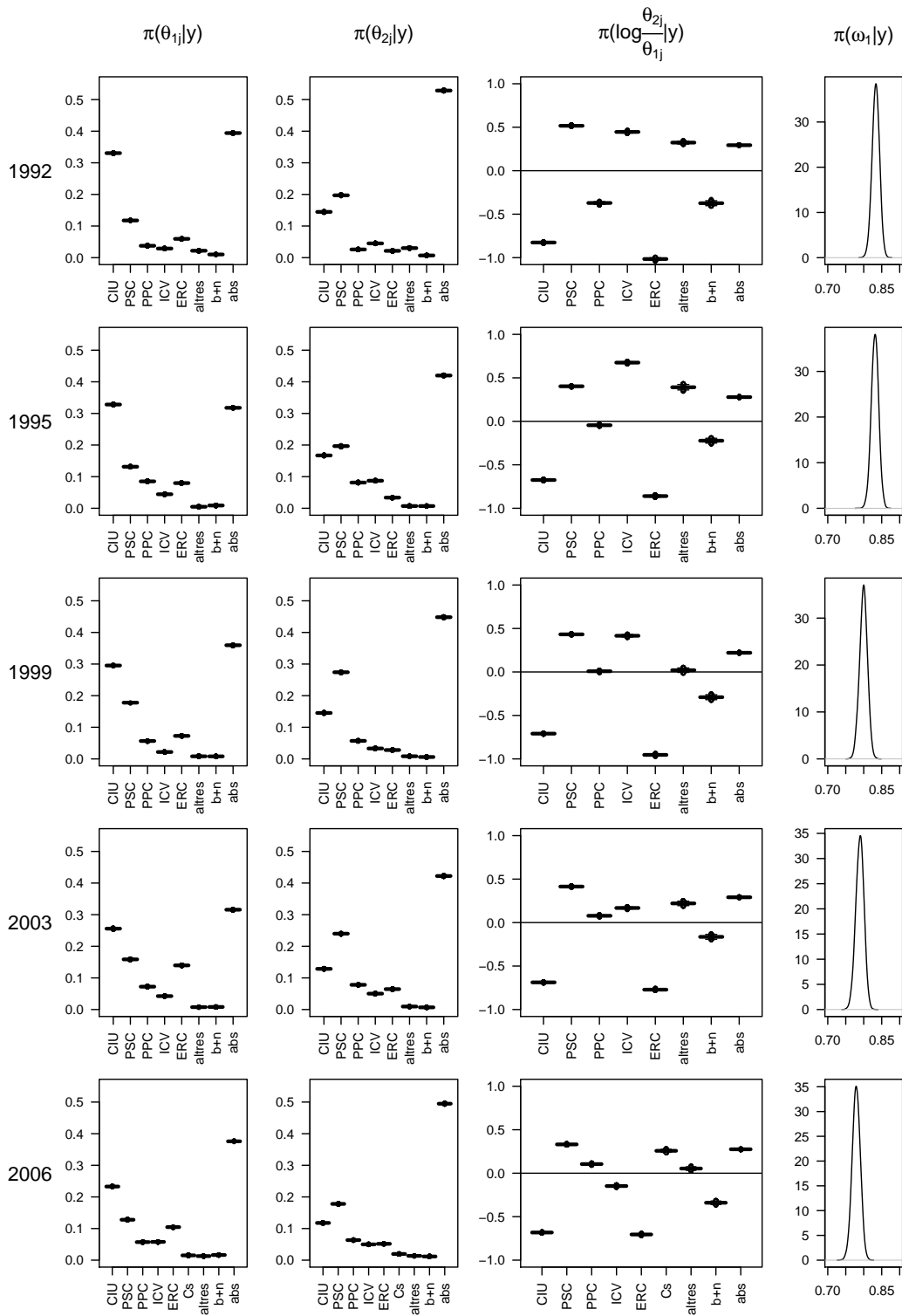


Figura C.4: Distribució a posteriori dels perfils de cada cluster, $\pi(\theta_{1j}|y)$ i $\pi(\theta_{2j}|y)$, del logaritme del quocient dels perfils, $\pi(\log \frac{\theta_{2j}}{\theta_{1j}}|y)$ per $j = 1, \dots, k$ i de la proporció de districtes del cluster 1, $\pi(\omega_1|y)$, per als resultats de totes les eleccions a Catalunya utilitzant el model no jeràrquic M_2 .

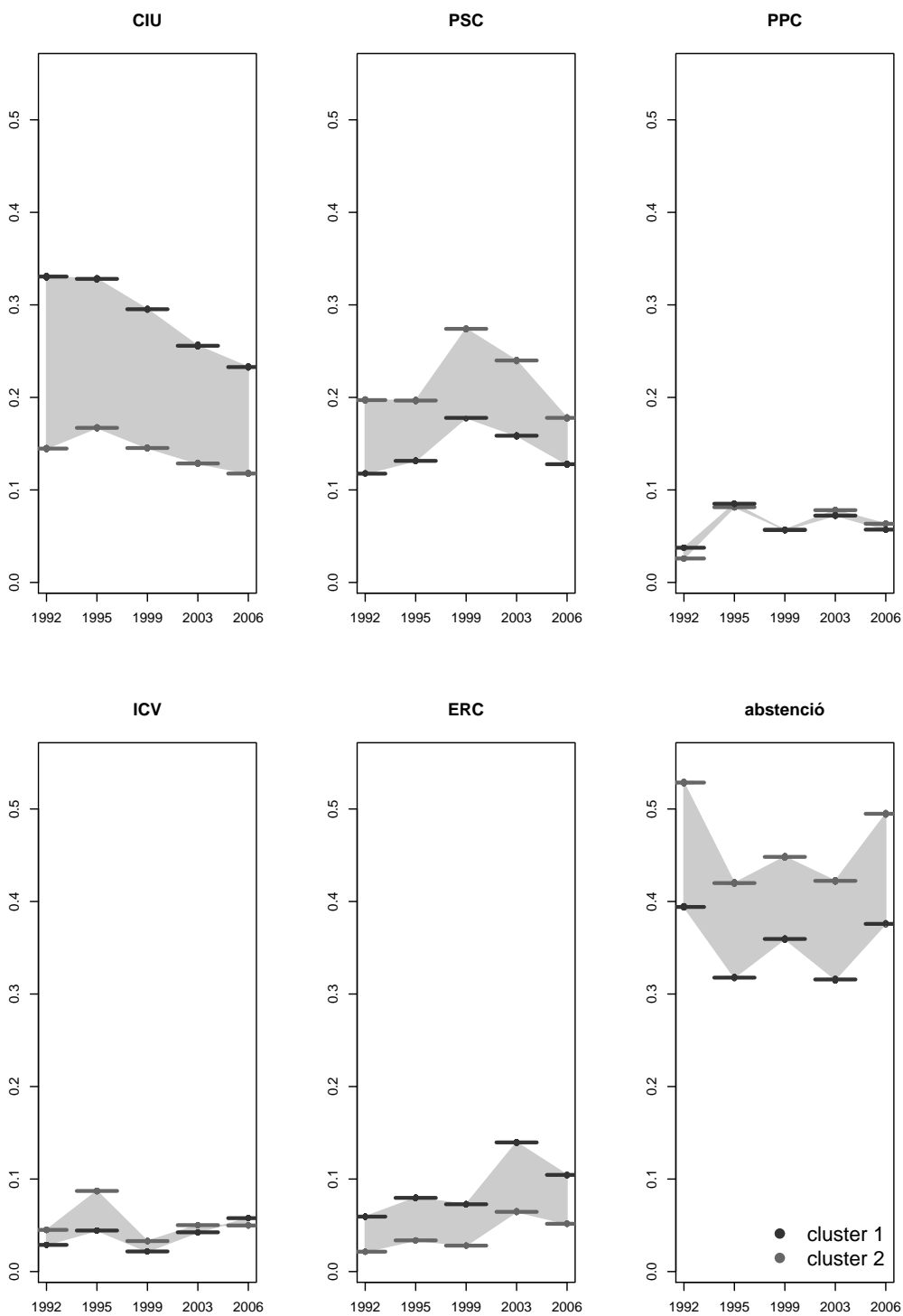


Figura C.5: Evolució estratificada per partit de les distribucions a posteriori dels perfils i de la seva distància entre els clusters, utilitzant el model no jeràrquic M_2 per les dades de tot Catalunya.

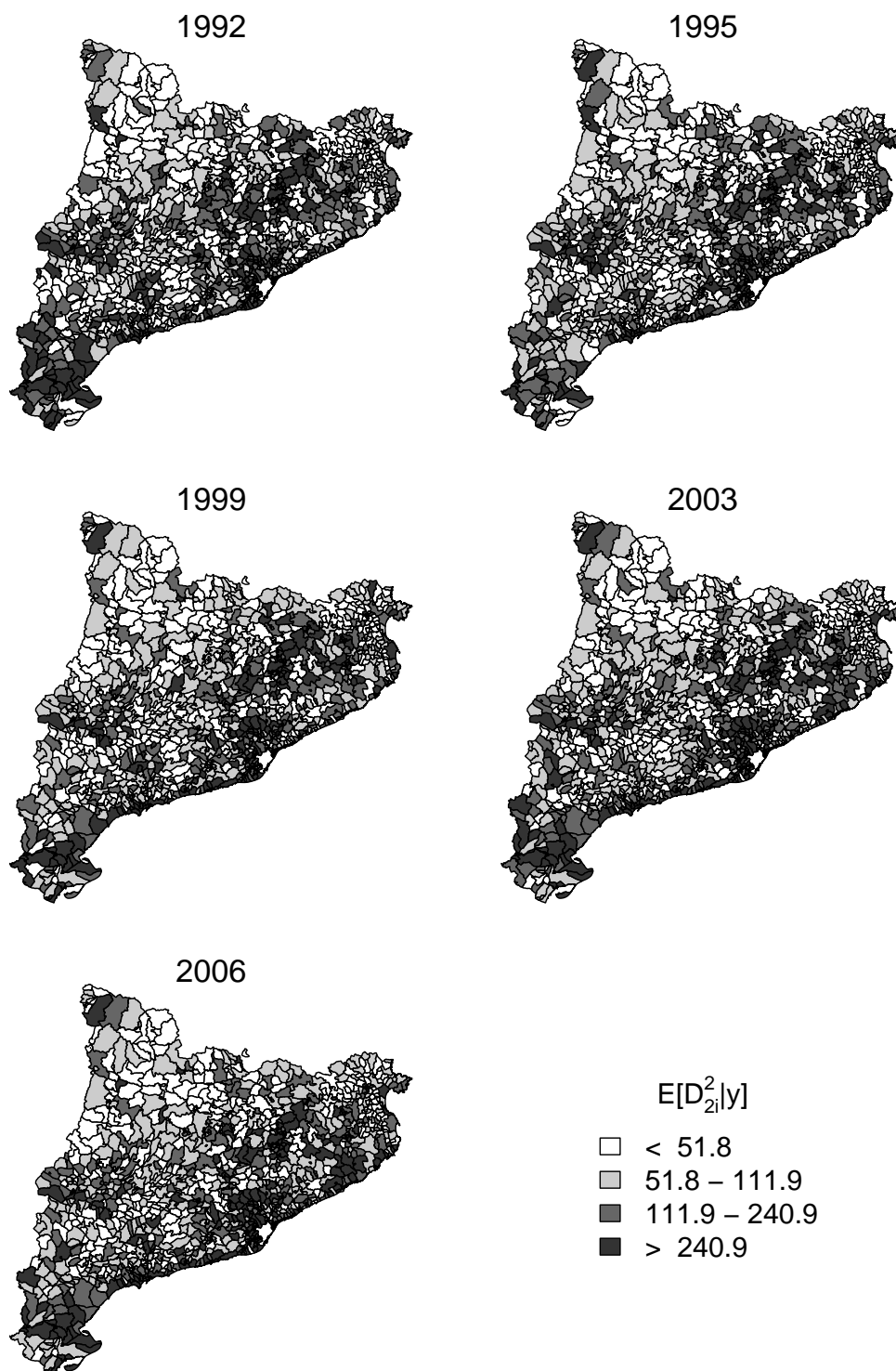


Figura C.6: Distribució espacial de l'esperança a posteriori de D_{2i}^2 sota el model M_2 pels resultats de les cinc darreres eleccions al Parlament de Catalunya. Per a la representació s'han categoritzat els valors prenent com a punt de tall les quartiles de l'any 1992.

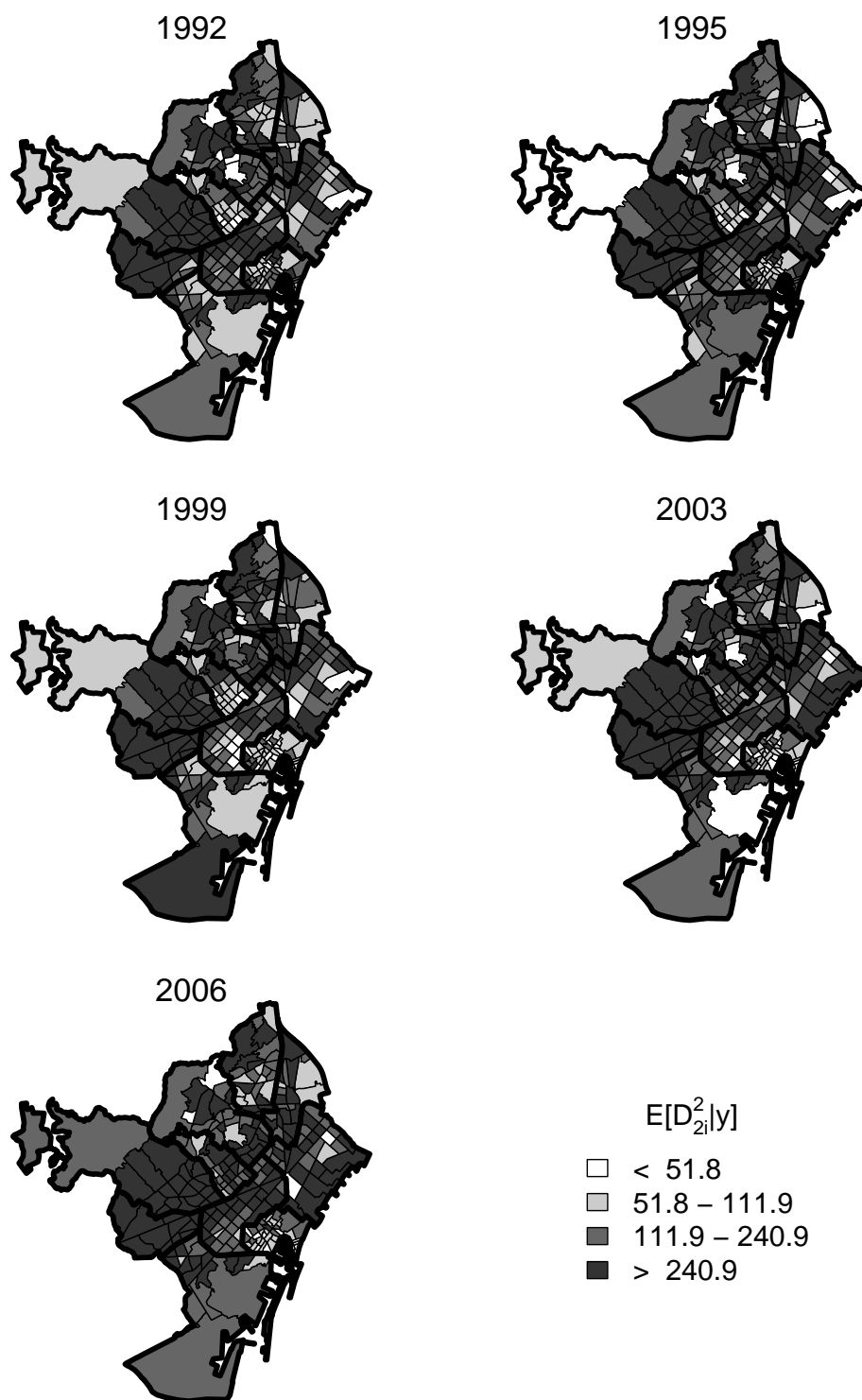


Figura C.7: Distribució espacial de l'esperança a posteriori de D_{2i}^2 , sota el model M_2 pels resultats de les cinc darreres eleccions al Parlament de Catalunya a la ciutat de Barcelona. Per a la representació s'han categoritzat els valors prenent com a punt de tall les quartiles de l'any 1992 del global de Catalunya.

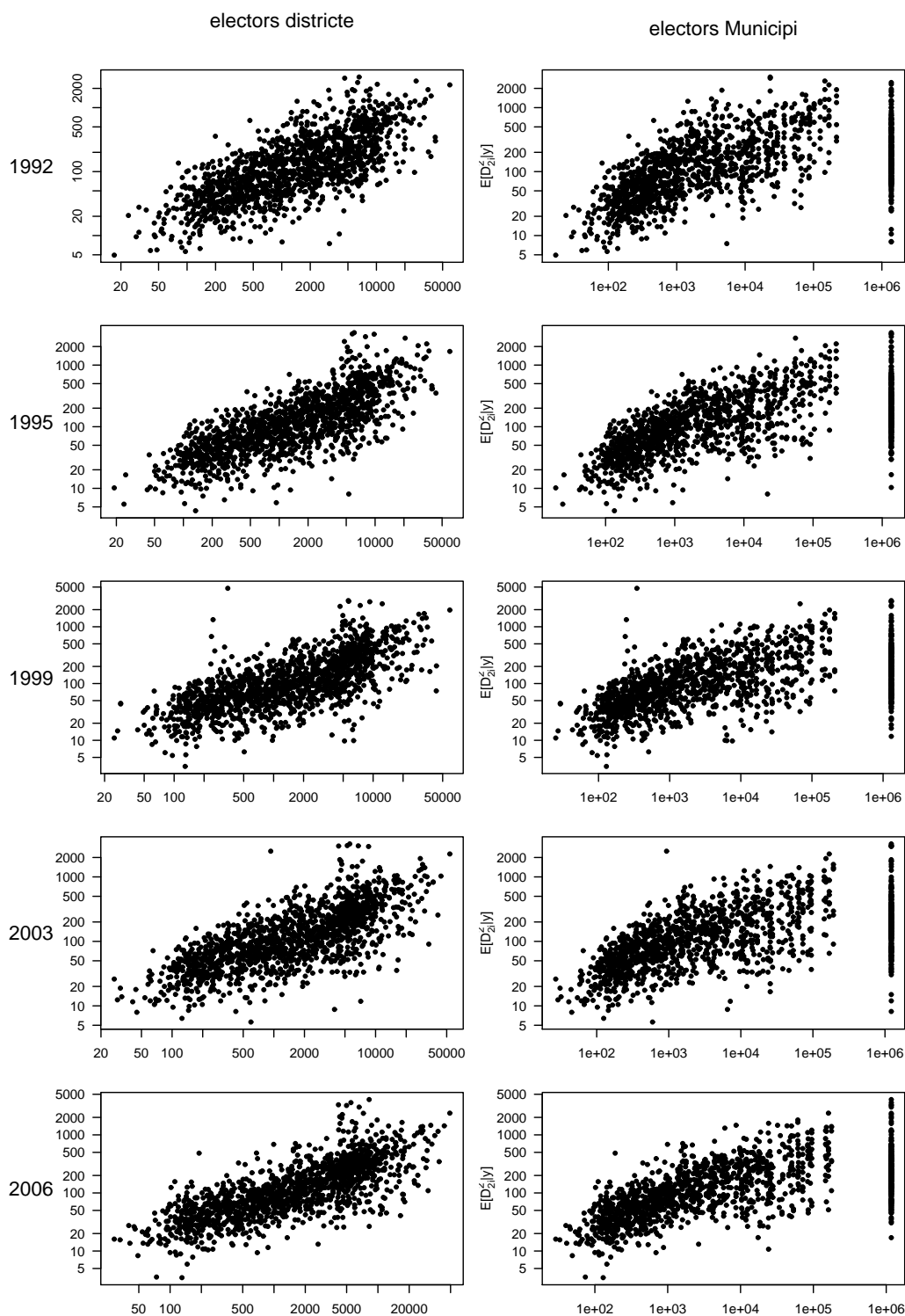


Figura C.8: Esperança a posteriori de D_{2i}^2 , $E[D_{2i}^2|y]$, en funció del nombre d'electors del districte i del nombre d'electors del municipi al que pertanyen, en escala logarítmica, utilitzant el model no jeràrquic M_2 per les dades de tot Catalunya.

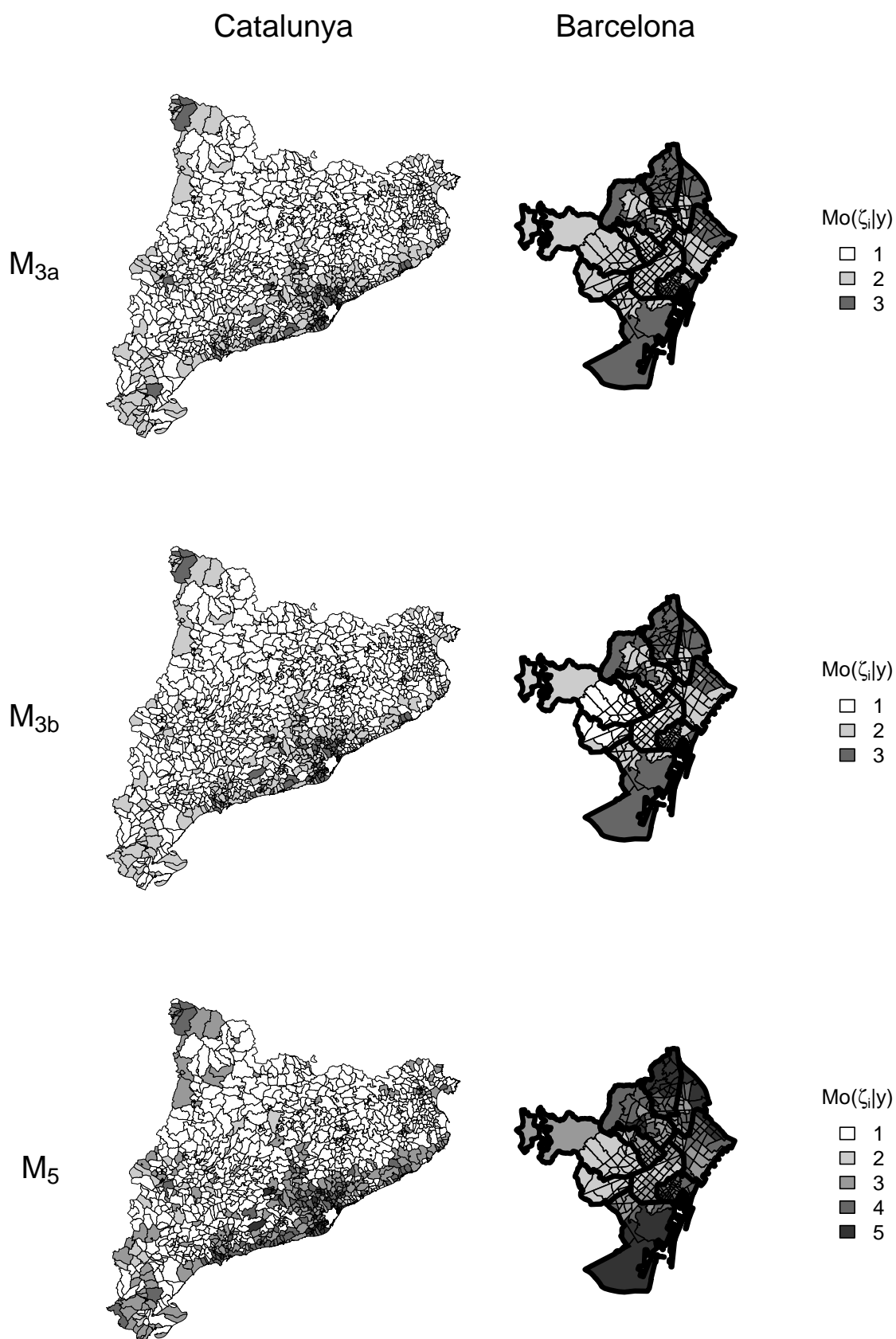


Figura C.9: Classificació dels districtes utilitzant com a criteri la moda a posteriori de les variables latents ζ_i , $Mo(\zeta_i|y)$, pels models no jeràrquics M_{3a} , M_{3b} i M_5 per les eleccions de 2003.

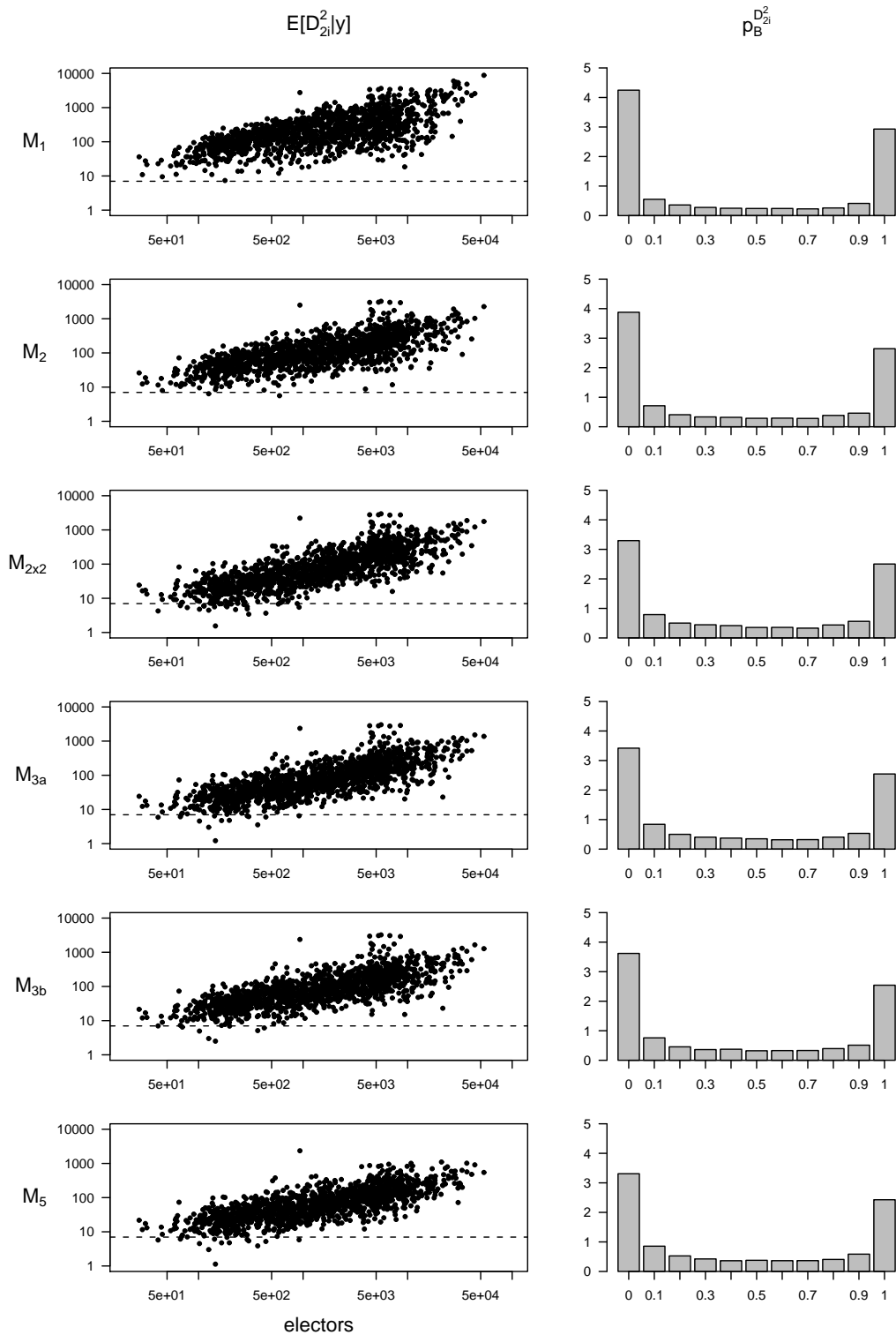


Figura C.10: A l'esquerra l'esperança a posteriori de D_{2i}^2 , $E[D_{2i}^2|y]$, en funció del nombre d'electors del districte, en escala logarítmica, i amb línia discontinua el valor esperat de la a posteriori $E[\widetilde{D}_{2i}^2|y]$, i a la dreta la distribució de $p_B^{D_{2i}^2}$ per a les dades de tot Catalunya del 2003.

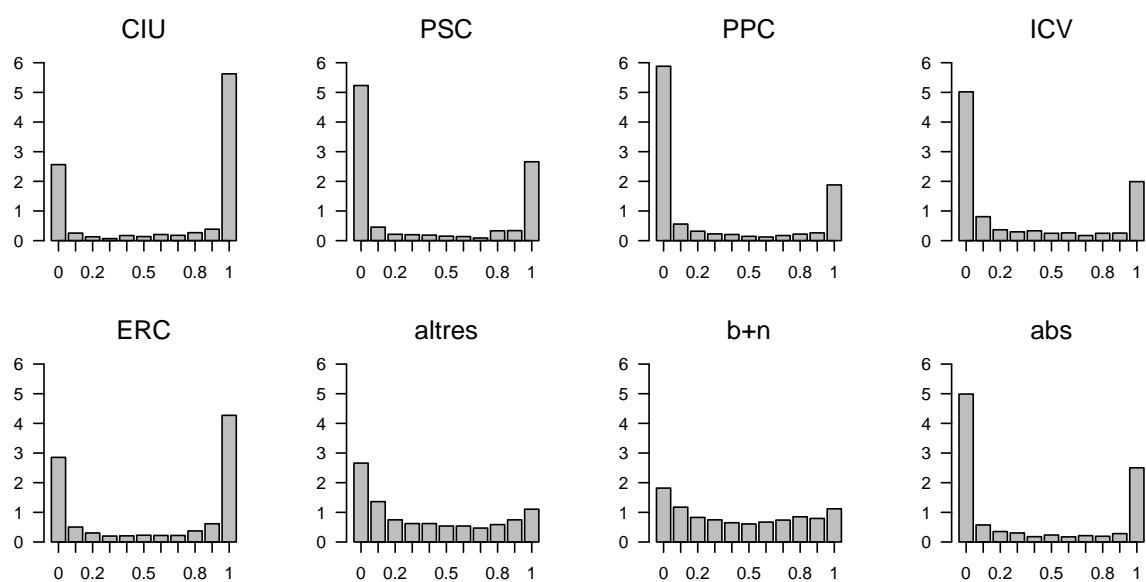


Figura C.11: Distribució de $p_B^{D_{2i}^2}$ per a cada categoria per al model no jeràrquic de 2 clusters, M_2 , Catalunya 2003.

Bibliografia

- Aitchison J (1986). *The Statistical Analysis of Compositional Data*. London: Chapman & Hall.
- Allenby GM, Rossi PE (1999). Marketing models of consumer heterogeneity. *Journal of Econometrics*, 89, 57-78.
- Baker SG (1995). The multinomial-poisson transformation. *The Statistician*, 43, 495-504.
- Banerjee S, Carlin BP, Gelfand AE (2004). *Hierarchical Modelling and Analysis for Spatial Data*. Chapman and Hall/CRC.
- Banfield JD, Raftery AE (1993). Model based Gaussian and non-Gaussian clustering. *Biometrics*, 49, 803-821.
- Bayarri MJ, Berger JO (2000). P-values for composite null models. *J. Amer. Statist. Assoc.*, 95, 1127-1142.
- Bayarri MJ, Castellanos ME (2007). Bayesian Checking of the Second Levels of Hierarchical Models. *Statistical Science*, 22, 322-343.
- Berger J (1985). *Statistical Decision Theory and Bayesian analysis*. Springer.
- Bernardo JM (1979). Reference posteriori distributions for Batesian inference (with discussion), *Journal of the Royal Statistical Society Series B*, 41, 113-147.
- Bernardo JM, Smith AFM (1994). *Bayesian Theory*. Wiley.
- Bernardo JM, Girón J. (1988), A Bayesian approach to cluster analysis. *Qüestió*, 12, 97-112.
- Besag J, York JC, Mollié A (1991). Bayesian image restoration, with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics*, 43, 1-20.

- Bivand RS, Pebesma EJ, Gómez-Rubio V (2008). *Applied Spatial Data Analysis with R (Use R)*. Springer.
- Box GEP (1980). Sampling and Bayes inference in scientific modelling and robustness (with discussion). *Journal of the Royal Statistical Society Series A*, 143, 383-430.
- Brooks SP, Giudici P, Roberts GO (2003). Efficient construction of reversible jump MCMC proposal distributions (with discussion). *Journal of the Royal Statistical Society B*, 65, 3-55.
- Carlin BP, Louis TA (2000). *Bayes and Empirical Bayes Methods for Data Analysis*. 2nd ed. Boca Raton, FL; Chapman and Hall/CRC Press.
- Casella G, George EI (1992). Explaining the gibbs sampler. *The American Statistician*, 46, 167-174.
- Celeux G, Forbes F, Robert CP, Titterton DM (2006). Deviance Information Criteria for Missing Data Models. *Bayesian Analysis*, 4, 651-674.
- Chang K, Siddarth S, Winberg CB (1999). The impact of heterogeneity in purchase timing and price responsiveness on estimates of sticker shock effects. *Marketing Science*, 18, 178-192
- Chib S, Greenberg E (1995). Understanding the Metropolis-Hasting Algorithm, 49, 327-335.
- Congdon P (2006). *Bayesian Statistical Modelling* (2nd edition). Wiley.
- Congdon P (2005). *Bayesian Models for Categorical Data*. Wiley.
- Congdon P (2003). *Applied Bayesian Modelling*. Wiley.
- Cook RD (1977). Detection of influential observations in linear Regression, *Technometrics*, 19, 15-19.
- Cook S, Gelman A, Rubin DB (2006). Validation of software for Bayesian models using posteriori quantiles. *Journal of Computational and Graphical Statistics*, 15, 675-692.
- Cowles MK, Carlin BP (1996). Markov chain Monte Carlo convergence diagnostics: A comparative review. *J. Amer. Statist. Assoc*, 91, 883-904.
- Cressie NAC (1993). *Spatial for Spatial Data*. 2nd ed. New York: Wiley.
- Dellaportas P, Papageorgiou I. (2004). Multivariate mixtures of normals with unknown number of components. Working paper, Athens University of Economics and Business.

- Deloiero M, Robert CP (2002). Discussion on the Paper by Spiegelhalter, Best, Carlin and van der Linde. *Roy. Statist. Soc., Ser. B*, 64, 627.
- Dempster AP, Laird NM, Rubin DB (1977). Maximum likelihood form incomplete data va the EM alhorithm (with discussion), *Journal of the Royal Statistics Society Ser. B*, 39, 1-38.
- DiCiccio TJ, Dass R, Raftery A, Wasserman L (1977). Computing Bayes Factor by combining simulation and asymptotic approximations. *Journal of the American Statistical Association*, 92, 903-915.
- Diebolt J, Robert CP (1994). Estimation of finite mixture distributions through Bayesian sampling. *Journal of the Royal Statistical Society Ser. B*, 56, 363-375.
- Geisser S (1980). In Discussion of Box (1980).
- Gelfand A (2000). Gibbs Sampling. *Journal of the American Statistical Association*, 95, 1300-1304.
- Gelfand AE, Ghosh SK (1998). Model choice: a minimum posterior predictive loss approach. *Biometrika*, 85, 1-11.
- Gelman A (2003). A Bayesian formulation of exploratory data analysis and goodness-of-fit teting. *Internat. Statist. Rev.* 71, 369-382.
- Gelman A, Meng X, Stern H (1996). Posterior predictive assessment of model fitness via realized discrepancies., *Statistica Sinica*, 6, 733-807.
- Gelman A, Carlin JC, Stern H, Rubin DB (2004). *Bayesian Data Analysis*, second ed. Chapman and Hall, Neww York.
- Gelman A, Carlin JC, Stern H, Rubin DB (1995). *Bayesian Data Analysis*, first ed. Chapman and Hall, Neww York.
- Gemlan A, Rubin DB (1992). Inference from iterative simulation using miltiple sequences (with discussion). *Statistical Science*, 7, 457-511.
- Gilks WR, Best N, Tan KKC (1995). Adaptive rejection Metropolis sampling within Gibbs sampling. *Applied Statistics*, 44, 455-472.
- Giron J, Ginebra J, Riba A. (2005). Bayesian analysis of a Multinomial Sequence and Homogeneity of Literary Style, *The American Statistician*, 59, 19-30.
- Good I (1976). On the application of symmetric dirichlet distributions and their mixtures to contingency tables, *The Annals of Statistics*, 4, 1159-1189.

- Good I, Crook JF (1976). The robustness and sensitivity of the mixed-Dirichlet Bayesian Test for independence in contingency tables, *The Annals of Statistics*, 15, 670–693.
- Graffelman J, Morales-Camarena J (2008). Graphical tests for Hardy-Weinberg equilibrium based on the ternary plot. *Human Heredity*, 65, 77-84.
- Greenacre M (2007). *Correspondence Analysis in Practice* (Second Edition). Chapman & Hall/CRC, London.
- Hjort NL, Dahl FA, Steinbakk GH (2006). Post-processing posterior predictive p-values. *J. Amer. Statist. Assoc.*, 101, 1157-1174.
- Jasra CC, Homes CC, Stephens DA (2005). Markov chain Monte carlo methods and the label switching problem in Bayesian mixture modelling. *Statistical Science*, 20, 50-67.
- Johnson N, Kotz S, Kemp A (1993). *Univariate discrete distributions*. New York, Wiley.
- Johnson W, Geisser S (1983). A predictive view of the detection and characterization of influential observations in regression analysis, *Journal of the American Statistical Association*, 78, 137-144.
- Kass RE, Carlin BP, Gelman A, Neal R (1998). Markov chain Monte Carlo in practice: a roundtable discussion. *American Statistician*, 52, 93-100.
- Kass RE, Raftery AE (1995). Bayes factors. *Journal of the American Statistical Association*, 90, 773-795.
- Lavine M, West M (1992). A Bayesian method for classification and discrimination. *Canadian Journal of Statistics*, 20, 451-461.
- Lee PM (2001). *Bayesian Statistics. An Introduction. 3rd Ed.* Arnold.
- Lefkimmiatis S, Maragos P, Papandreou G (2009). Bayesian Inference on Multiscale Models for Poisson Intensity Estimation: Applications to Photon-Limited Image Denoising. To appear in *The IEEE Transactions on Image Processing*.
- Lenk P, DeSarbo W (2000). Bayesian inference for finite mixtures of generalized linear models with random effects. *Psychometrika*, 65, 93-119.
- Leonard T (1972). Bayesian methods for binomial data. *Biometrika*, 59, 581-589.

- Leonard T, Hsu J (1994). *The Bayesian analysis of categorical data - a selective review*, in Aspects of Uncertainty: a Tribute to DV Lindeley, P Freeman and A Simith (eds). Wiley.
- Leonard T, Hsu J (1994). *Bayesian Methods*. Cambridge U. Press.
- Mengersen KL, Robert CP, Guihenneuc-Jouyaux C (1999). MCMC convergence diagnostics: A review (with discussion). In *Batesian Statics 6*, eds Bernardo JM, Berger JO, Dawid AP, Smith AFM. Oxford: Oxford University Press, 415-440.
- McLachlan GJ, Basford K (1988). *Mixture Models: Inference and Applications to Clustering*. New York: Marcel Dekker.
- Moran PAP. (1950). Notes on continuous stochastic phenomena, *Biometrika*, 37, 17-23.
- Nandram B (1998). A Bayesian analysis of the tree-stage hierarchical multinomial model, *Journal of Statistical Computation and Simulation*, 61, 97-126.
- Petit LI (1986). Diagnostics in Bayesian model choice, *The Statistician*, 35, 183-190.
- Rao JNK (2003). *Small Area Estimation*, second ed. John Wiley and Sons, Inc., Hoboken, New Jersey.
- Richardson S, Green PJ (1997). On Bayesian analysis of mixtures with an unknown number of components. *Journal of the Royal Statistical Society Series B*, 59, 731-792.
- Richardson S (2002). Discussion on the Paper by Spiegelhalter, Best, Carlin and van der Linde. *Roy. Statist. Soc., Ser. B*, 64, 627.
- Ripley BD (1981). *Spatial Statistics*. Wiley.
- Robert C (2001). *The Bayesian Choice: From Decision Theoretical Foundations, to Computational Implementation*. Springer.
- Robert CP, Casella G (1999). *Monte Carlo Statistical Methods*. New York: Springer-Verlag.
- Roberts HV (1965). Probabilistic prediction. *Journal of the American Statistical Association*, 60, 55-62.
- Robins JM, van der Vaart A, Ventura V (2000). Asymptotic distribution of p values in composite null models (with discussion). *J Amer. Statist. Assoc.* 95, 11473-1156.
- Rossi PE, Allenby GM, McCulloch R (2005). *Bayesian Statistics and Marketing*. Wiley.

- Rubin DB (1984). Bayesianly justifiable and relevant frequency calculations for the applied statistician. *Annals of Statistics*, 12, 1151-1172.
- Stern HS, Sinharay S (2005). Bayesian Model Checking and Model Diagnostics, Shuckers M (1998). A Hierarchical Bayesian approach for analyzing a polychotomous response from a cluster sample. Proceeding of the ASA Section on Survey Research Methods, American Statistical Association, Alexandria, VA.
- Smith A, Gelfand A (1992). Bayesian Statistics Without Tears: A Sampling-Resampling Perspective. *American Statistical Association*, 46, 84-88.
- Spiegelhalter KF, Best N, Carlin BP, van der Linde A (2002). Bayesian measures of model complexity and fit (with discussion). *J. Roy. Statist. Soc., Ser. B*, 64, 583-639.
- Spiegelhalter KJ, Thomas Q, Best N, Lunn K (2003). *WinBugs: Bayesian inference Using Gibbs Sampling. Manual version 1.4*, Medical Research Council Biostatistics Unit., Cambridge, (<http://www.mrc-bsu.cam.ac.uk/bugs/>).
- Steinbakk GH, Storvik GO (2009). Posteriori Predictive p-values in Bayesian Hierarchical Models. *Scandinavian Journal of Statistics*, 36, 320-336.
- Wasserman L (2004). *All of statistics. A concise course in statistical Inference*. Springer-Verlag. New York.

Índex de taules

2.1	Part of the word frequency counts in the 425 chapters of <i>Tirant lo Blanc</i> with more than 200 words, obtained from Riba (2002).	8
2.2	Part of the word frequency counts of the nouns in Macaulay's essay on Bacon, and of all the words in <i>Alice in Wonderland</i> , of <i>Through the Looking Glass</i> , of the <i>War of the Worlds</i> , of <i>Max Havelaar</i> and of a Turkish archeology text.	8
2.3	Observed and expected word frequency counts and the corresponding Pearson residuals, $(v_{r:n} - v_n \hat{p}_{r:n}) / \sqrt{v_n \hat{p}_{r:n}}$, for Chapter 5 of <i>Tirant lo Blanc</i> under the extended and under the unextended truncated IG-Poisson models.	21
3.1	Part of the word frequency count sets of the nouns in Macaulay's essay on Bacon, considered in Sichel (1975), and of all the words in a Turkish archeology text, in <i>Alice in Wonderland</i> and in <i>Through the Looking Glass</i> by Lewis Carroll, in <i>The Hound of the Baskervilles</i> by Doyle, in <i>The War of the Worlds</i> by Wells, and in <i>Max Havelaar</i> by Dekker, which are all considered in Baayen (2001). Note that $v_n = \sum_{r=1}^n v_{r:n}$ and $n = \sum_{r=1}^n r v_{r:n}$.	28
3.2	Maximum likelihood estimate of (b, c, g) , maximum value of the log-likelihood function, value of the $X^2(\hat{b}, \hat{c}, \hat{g})$ goodness of fit test statistic and its corresponding p-value, all for the truncated IG-Poisson and GIG-Poisson models and for the IG-Truncated Poisson and GIG-Truncated Poisson models. Between brackets, the degrees of freedom associated to X^2	36
3.3	Estimates of the size and of two measures of the diversity of the vocabulary of the authors of the texts in Table 3.1, based on the maximum likelihood estimates of the mixing distribution of the truncated GIG-Poisson model.	41

3.4	Estimates of $v(\psi')$, of two measures and of the diversity in the observed part of the vocabulary for the texts in Table 3.1, based on the maximum likelihood estimate of the mixing distribution of the GIG-Truncated Poisson model, and sample versions of these measures.	45
4.1	Word frequency count data for the nouns in Macaulay's essay on Bacon, considered in Sichel (1975), and for all the words in a Turkish archeology text, in <i>Alice in Wonderland</i> and in <i>Through the Looking Glass</i> by Carroll, in <i>The Hound of the Baskervilles</i> by Doyle, in <i>The War of the Worlds</i> by Wells, and in <i>Max Havelaar</i> by Dekker, all considered in Baayen (2001). Note that r_{max} denotes the frequency of the most frequent word.	50
4.2	Maximum likelihood estimate of (b, c, g) for the truncated GIG-Poisson model, value of the χ^2 goodness of fit test statistic, and a lower bound for the corresponding p-value.	52
4.3	Measures of the diversity within (π_1, \dots, π_v) , their expected value assuming that the π_i are identically distributed as $\psi(\pi)$ and that $v = 1/E_\psi[\pi]$, $D(\psi)$, and their maximum value when they are restricted for populations with a given total number of classes v . $D(\psi)$ is the ratio between a measure of the lack of variability of π distributed as $\psi(\pi)$ and $E_\psi[\pi]$. . .	62
4.4	Measures of the evenness of (π_1, \dots, π_v) associated to the diversity measures in Table 4.3 and (4.8), and their expected value when the π_i are identically distributed as $\psi(\pi)$. To compute $E_{GA}(\psi)$ one assumes that the π_i are both independent as well as identically distributed.	63
4.5	Estimates of measures of the diversity of the vocabulary of the authors of the texts in Table 4.1, based on the mle of the mixing distribution of the truncated GIG-Poisson model.	69
4.6	Estimates of measures of the evenness of the vocabulary of the authors of the texts in Table 4.1, based on the mle of the mixing distribution of the truncated GIG-Poisson model.	69
6.1	Districtes als que pertanyen les 248 zrp de Barcelona ciutat i nombre de zrp pertanyents a cadascun dels districtes.	89

- 6.2 Part de la taula dels resultats a les eleccions al Parlament de Catalunya de l'any 2003 per a la ciutat de Barcelona. N és el nombre total d'electors, abs és l'abstenció, $b+n$ són el total de vots blancs i nuls, $altres$ és l'agregació de tots els partits amb menys d'1% del global de vots, i la resta de categories respon a les sigles dels respectius partits. 89
- 6.3 Nombre de vots a la ciutat de Barcelona per a cadascuna de les opcions i el nombre d'electors (N) per a cadascuna de les darreres cinc eleccions al parlament de Catalunya. Les dades presentades a la taula han estat prèviament processades de manera que a l'any 1999 als vots de ICV se li han sumat els de EUiA. 89
- 6.4 Percentatge de vots a la ciutat de Barcelona per a cadascuna de les opcions per a cadascuna de les darreres cinc eleccions al parlament de Catalunya. Les dades presentades a la taula han estat prèviament processades de manera que a l'any 1999 als vots de ICV se li han sumat els de EUiA. 90
- 6.5 Part de la taula dels resultats a les eleccions al Parlament de Catalunya de l'any 2003. N és el nombre total d'electors, abs : és l'abstenció, $b+n$: són el total de vots blancs i nuls, $altres$ és l'agregació de tots els partits amb menys d'1% del global de vots, i la resta de categories respon a les sigles dels respectius partits. 102
- 6.6 Nombre de vots per a cadascuna de les opcions i el nombre d'electors (N) 103
- 6.7 Percentatge de vots a per a cadascuna de les opcions per a cadascuna de les darreres cinc eleccions al parlament de Catalunya. Les dades presentades a la taula han estat prèviament processades de manera que a l'any 1999 als vots de ICV se li han sumat els de EUiA així com la part imputada de vots fora de la circumscripció de Barcelona que han estat restats del PSC. 103
- 9.1 Districtes de Barcelona ciutat. 152
- 9.2 Esperança a posteriori de les mesures de discrepància D_1 i D_2 , i els p valors bayesians associats, $p_B^{D_1}$ i $p_B^{D_2}$ per els resultats a les eleccions al Parlament de Catalunya del 2003 a la ciutat de Barcelona. 156

- 9.3 Esperança a posteriori de la mesura de discrepància D_2 , i el p valor bayesià associats, $p_B^{D_2}$, per els resultats a les eleccions al Parlament de Catalunya del 2003 a la ciutat de Barcelona. 174
- 9.4 Índex de Moran i p valor associat obtingut fent servir el test de permutació partint de $E[D_{2i}^2|y]$ per a cadascun dels cinc models considerats. 175
- 10.1 Esperança a posteriori dels perfils de probabilitat per als clusters de cada model, que en el cas dels models no jeràrquics correspon a $E[\theta_j|y]$ mentre que en el cas del model jeràrquic correspon a $E[\mu_j|y]$ per $j = 1, \dots, k$. Per a cada model també es dona $E[\omega|y]$, que indica el tamany relatiu de cada cluster, i pel cas jeràrquic també es dona $E[\tau|y]$ que indica el grau d'heterogeneïtat del cluster. 200
- 11.1 p valor associats a l'índex de Moran, calculats a partir de $E[D_{2i}^2|y]$, per als models M_{2J} , M_{3J} i M_{4J} per a cadascuna de les cinc eleccions al Parlament. 212
- 12.1 Esperança a posteriori els perfils de la probabilitat esperada per a cada cluster sota el model M_{3J} , $E[\mu_{rj}|y]$ per $r = 1, 2, 3$ i per $j = 1, \dots, 8$ o 9 , del tamany relatiu de cada cluster, $E[\omega_r|y]$, i del grau d'heterogeneïtat del cluster, $E[\tau_r|y]$, i el percentatge de vots de cada opció per a cada una de les cinc eleccions considerades. 230
- 12.2 Canvis en la classificació de les zrp a cada cluster entre dues eleccions consecutives. Per exemple $1 \rightarrow 2$ indica el nombre de zrp que canvien del cluster 1 al cluster 2 d'una elecció a la següent. 230
- 12.3 Esperança a posteriori els perfils de la probabilitat esperada per a cada cluster sota el model M_{4J} , $E[\mu_{rj}|y]$ per $r = 1, 2, 3, 4$ i per $j = 1, \dots, 8$ o 9 , del tamany relatiu de cada cluster, $E[\omega_r|y]$, i del grau d'heterogeneïtat del cluster, $E[\tau_r|y]$, i el percentatge de vots de cada opció per a cada una de les cinc eleccions considerades. 241
- 12.4 Canvis en la classificació de les zrp a cada cluster entre dues eleccions consecutives. Per exemple $1 \rightarrow 2$ indica el nombre de zrp que canvien del cluster 1 al cluster 2 d'una elecció a la següent. 242

-
- A.1 Mesura potencial de reducció de la convergència, \widehat{R} , porposada per Gelman i Rubin (1992) aplicada a les simulacions del model de tres clusters de l'any 1999, obtingudes a partir de monitoritzar una simulació de cada 100 d'un total de 150000 simulacions per a cada cadena. El període d'esclafament havia estat de 50000 simulacions en cada cadena. 269
- C.1 Esperança a posteriori dels perfils de probabilitat per al model M_2 , $E[\theta_{1j}|y]$ i $E[\theta_{2j}|y]$ per $j = 1, \dots, k$, i el tamany relatiu de cada cluster, $E[\omega|y]$, per a cada una de les cinc eleccions. 303
- C.2 Esperança a posteriori dels perfils de probabilitat i del tamany relatiu de cada cluster pel model $M_{2 \times 2}$, quan s'analitza separant els districtes de pobles de menys de 5000 habitants de districtes de pobles de més de 5000 habitants, per l'any 2003. 305
- C.3 Esperança a posteriori dels perfils per a cada cluster, $E[\theta_{rj}|y]$, i la grandària relativa de cada cluster, $E[\omega_r|y]$, per als models no jeràrquics M_3 i M_5 per a les dades de tot Catalunya de l'any 2003. 306
- C.4 Mesures de bondat d'ajust per als diferents models no jeràrquics considerats per les dades de tot Catalunya de l'any 2003. 306

Índex de figures

2.1	Contour plots for the probability function of the extended truncated IG-Poisson(α, θ) model, $p_{r:n}^{etigp}(\alpha, \theta)$, computed recursively through (2.7), (2.8) and (2.9) for $r = 1, \dots, 10, 15, 20$	14
2.2	Probability mass function of the extended truncated IG-Poisson model, $p_{r:n}^{etigp}(\alpha, \theta)$, computed recursively through (2.7), (2.8) and (2.9) for sixteen pairs of values of (α, θ)	15
2.3	Contour plots for the log-likelihood function under the extended truncated IG-Poisson(α, θ) model, with $(\alpha, \theta) \in (-1, \infty) \times (0, 1)$, for the word frequency count data of Chapters 1 to 9 of <i>Tirant lo Blanc</i> partially presented in Table 2.1.	17
2.4	Relationship between the estimates of (α, θ) and n , between the estimate for θ under the extended model and the one under the unextended model, and between the χ^2 goodness of fit test p -value of the extended truncated IG-Poisson model and n . The data are the word frequency count sets of the 425 chapters of <i>Tirant lo Blanc</i> with more than 200 words.	18
2.5	Base 10 logarithm of the ratio between the maximum of the likelihoods of the extended truncated IG-Poisson(α, θ) model (with $\alpha > -1$) and of the unextended model (with $\alpha \geq 0$), difference between the χ^2 goodness of fit test statistics of the unextended and of the extended models, and relation between the χ^2 goodness of fit test p -values of both models, all for the 425 chapters of <i>Tirant lo Blanc</i> with more than 200 words.	20

2.6	Contour plots for the log-likelihood function for the extended truncated IG-Poisson model for the word frequency count of the nouns in Macaulay's essay on Bacon, and of all the words in <i>Alice in Wonderland</i> , in <i>Through the Looking Glass</i> , in <i>War of the Worlds</i> , in <i>Max Havelaar</i> and in a Turkish archeology text, all partially presented in Table 2.2.	22
3.1	Contour plots of the probability function of the truncated GIG-Poisson model, $p_{r:n}^{tsich3}(b, c, g)$, when $r = 1, \dots, 5$ and $n = 25000$. The left hand side panels correspond to the contour plots for the truncated IG-Poisson model, $p_{r:n}^{tsich2}(b, c)$	31
3.2	Contour plots of the probability function of the GIG-Truncated Poisson model, $p_{r:n}^{gigtp}(b, c, g)$, when $r = 1, \dots, 5$ and $n = 25000$. The left hand side panels correspond to the contour plots for the IG-Truncated Poisson model, $p_{r:n}^{igtp}(b, c)$	33
3.3	Contour plots for the log-likelihood function of the zero truncated Sichel (GIG-Poisson(b, c, g)) model, in (3.3), near its maximum for the word frequency count data sets of four of the seven texts in Table 3.1.	37
3.4	Pearson residuals, $e_{r:n}(\hat{b}, \hat{c}, \hat{g})$, under the truncated IG-Poisson model, on the left hand side, and under the truncated GIG-Poisson model, on the right hand side.	38
3.5	Maximum likelihood estimate of the mixing density functions, $\psi(\pi)$, under the truncated GIG-Poisson model (on the left), next to the one of $\psi'(\pi')$ under the GIG-Truncated Poisson model (on the right), for the texts in Table 3.1. The ones on the left estimate the density of the word frequencies of the theoretical vocabulary of the author; the smaller its expectation the larger v and the smaller its variability the more diverse the vocabulary. The ones on the right estimate the word frequencies of the vocabulary observed in the text.	40
3.6	Contour plots for the log-likelihood function of the GIG-Truncated Poisson model near its maximum for the word frequency count data sets of four of the texts in Table 3.1.	43

3.7	Pearson residuals, $e_{r:n}(\hat{b}, \hat{c}, \hat{b})$, under the IG-Truncated Poisson model, on the left hand side panels, and under the GIG-Truncated Poisson model, on the right hand side panels.	44
4.1	Maximum likelihood estimates of the mixing densities of the zero truncated GIG-Poisson model for the texts in Table 4.1. They serve as estimates of the density of the word frequencies of the vocabulary of the author. The smaller $E_{\psi}[\pi]$ the larger the total number of words in the vocabulary of the author, v . For a given v , the smaller $Var_{\psi}[\pi]$ the more even and diverse that vocabulary.	53
4.2	Contour plots of $\log_{10} v(\psi)$, $D_{GS}(\psi)$, $D_{SQ}(\psi)$, $D_E(\psi)$ and $D_{r=1}(\psi)$ when the distribution of word frequencies is the GIG(b, c, g). Observe that all these measures except the last one are non-increasing functions of b, c and g	67
4.3	Contour plots of $E_{GA}(\psi)$, $E_{GS}(\psi)$, $E_{SQ}(\psi)$, $E_E(\psi)$ and $E_{r=1}(\psi)$ as a function of (b, c, g) when the distribution of word frequencies is the GIG(b, c, g). The left hand side panels correspond to the case where $\psi(\pi)$ is the IG(b, c) distribution. Observe that all these measures are non-decreasing with b and non-increasing with c and when $g < -0.5$ with g	68
6.1	Divisió territorial utilitzada per a les anàlisi. A dalt, Catalunya a nivell de districte. A baix, Barcelona a nivell de zrp i, en un contorn més gruixut, de districte.	88
6.2	Distribució percentual dels vots a les eleccions al Parlament de Catalunya a nivell de les zrp de la ciutat de Barcelona entre les diferents categories per a cadascun dels anys. El punt blau és el percentatge marginal de cada categoria.	94
6.3	Evolució de la distribució percentual de vot a les eleccions al Parlament de Catalunya a nivell de zrp de la ciutat de Barcelona per a cadascuna de les opcions de vot. El punt blau és el percentatge marginal de cada categoria.	95
6.4	Distribució del nombre d'electors de les zrp de la ciutat de Barcelona el 2006.	96

-
- 6.5 Percentatge de vot en funció del nombre d'electors (N) de les zrp de la ciutat de Barcelona per a cadascuna de les opcions de vot a les eleccions al Parlament de Catalunya del 2006. El nombre d'electors s'ha representat en escala logarítmica. 97
- 6.6 Distribució espacial del percentatge de vot a les Eleccions al Parlament de Catalunya per CIU, PSC, PPC, ERC i l'abstenció a nivell de zrp a Barcelona ciutat per als anys 1992 i 2006 . S'han categoritzat els percentatges de cada partit segons les quartiles de 1992. 98
- 6.7 Anàlisi de correspondències per a les dades de Barcelona ciutat a les cinc darreres eleccions del Parlament de Catalunya. En les etiquetes dels eixos, entre parèntesi, hi trobem la inèrcia explicada per les respectives components. 99
- 6.8 Anàlisi de correspondències per Barcelona ciutat a les eleccions del Parlament de Catalunya del 2006, presentant per separat les zrp de cada districte. 100
- 6.9 Valor que prenen quatre estadístics a nivell de zrp de Barcelona ciutat a les eleccions al Parlament de Catalunya del 2006. 101
- 6.10 Representació ternària de la distribució de vot de (PSC, CIU, PPC), i (PSC, CIU, ERC) a nivell de zrp de Barcelona ciutat a les eleccions al Parlament de Catalunya del 2006. 101
- 6.11 Distribució percentual dels vots a les eleccions al Parlament de Catalunya a nivell de districte, i en el cas de Barcelona de zrp, entre les diferents categories per a cadascun dels anys. El punt blau és el percentatge marginal de cada categoria. 106
- 6.12 Evolució de la distribució percentual de vot a les eleccions al Parlament de Catalunya a nivell de districte, i en el cas de Barcelona de zrp, per a cadascuna de les opcions de vot. El punt blau és el percentatge marginal de cada categoria. 107
- 6.13 Percentatge de vot en funció del nombre d'electors (N) del districte, i en el cas de Barcelona zrp, per a cadascuna de les opcions de vot per a les eleccions al Parlament de Catalunya del 2006. El nombre d'electors s'ha representat en escala logarítmica. 108

-
- 6.14 Distribució del nombre d'electors per districtes, i en el cas de Barcelona zrp, pel 2006 a Catalunya. 109
- 6.15 Nombre d'electors per municipis a Catalunya els anys 1992 i 2006. 109
- 6.16 Distribució espacial del percentatge de vot a les Eleccions al Parlament de Catalunya per CIU, PSC, PPC, ERC i l'abstenció a nivell de districtes per als anys 1992 i 2006. Barcelona està representada a la Figura 6.6. Els percentatges de cada partit s'han categoritzat segons les quartiles de 1992. 110
- 6.17 Anàlisi de correspondències dels diferents comicis al Parlament de Catalunya. Els punts representen els districtes i , en el cas de Barcelona, les zrp. En les etiquetes dels eixos entre parèntesis hi trobem el percentatge d'inèrcia explicada per les respectives components. 111
- 9.1 El primer gràfic representa els valors observats per $D_{ai}(y_i)$ a les zrp de cada districte a les eleccions al Parlament de Catalunya del 2003 a Barcelona ciutat i les quatre files següents corresponen a quatre rèpliques d'aquest estadístic simulades a partir de la predictiva a posteriori de cada model no jeràrquic considerat. 158
- 9.2 El primer gràfic representa els valors observats per $D_{bi}(y_i)$ a les zrp de cada districte a les eleccions al Parlament de Catalunya del 2003 a Barcelona ciutat i les quatre files següents corresponen a quatre rèpliques d'aquest estadístic simulades a partir de la predictiva a posteriori de cada model no jeràrquic considerat. 159
- 9.3 El primer gràfic representa els valors observats per $D_{ci}(y_i)$ a les zrp de cada districte a les eleccions al Parlament de Catalunya del 2003 a Barcelona ciutat i les quatre files següents corresponen a quatre rèpliques d'aquest estadístic simulades a partir de la predictiva a posteriori de cada model no jeràrquic considerat. 160
- 9.4 El primer gràfic representa els valors observats per $D_{di}(y_i)$ a les zrp de cada districte a les eleccions al Parlament de Catalunya del 2003 a Barcelona ciutat i les quatre files següents corresponen a quatre rèpliques d'aquest estadístic simulades a partir de la predictiva a posteriori de cada model no jeràrquic considerat. 161

- 9.5 La fila superior presenta els valors observats per $D_{ai}(y_i)$, $D_{bi}(y_i)$, $D_{ci}(y_i)$ i $D_{di}(y_i)$ a les eleccions al Parlament de Catalunya del 2003 a les zrp de Barcelona per districtes, i les altres files presenten una rèplica de les dades obtingudes a partir de la predictiva a posteriori de cada model no jeràrquic considerat. 162
- 9.6 Validació gràfica dels diferents models a partir de representacions ternàries dels perfils de tres columnes de la Taula 6.2. El primer gràfic correspon als perfils observats a les eleccions al Parlament de Catalunya del 2003 a Barcelona ciutat i les altres files de corresponen a rèpliques de les dades obtingudes simulant de la predictiva a posteriori de cada un dels cinc models no jeràrquics considerats. 163
- 9.7 Validació gràfica dels diferents models a partir de representacions ternàries dels perfils de tres columnes de la Taula 6.2. El primer gràfic correspon als perfils observats a les eleccions al Parlament de Catalunya del 2003 a Barcelona ciutat i les altres files de corresponen a rèpliques de les dades obtingudes simulant de la predictiva a posteriori de cada un dels cinc models no jeràrquics considerats. 164
- 9.8 Validació gràfica dels diferents models a partir de representacions ternàries dels perfils resultants d'agregar columnes de la Taula 6.2. El primer gràfic correspon als perfils observats a les eleccions al Parlament de Catalunya del 2003 a Barcelona ciutat i les altres files corresponen a rèpliques de les dades obtingudes simulant de la predictiva a posteriori de cada un dels cinc models no jeràrquics considerats. 165
- 9.9 Distribucions a posteriori de les mesures de discrepància D_1 i D_2 per als cinc models no jeràrquics considerats. 166
- 9.10 Esperança a posteriori de D_{2i}^2 , $E[D_{2i}^2|y]$, per a cada zrp en funció del seu nombre d'electors per cada model no jeràrquic considerat. 167
- 9.11 Distribució espacial de l'esperança a posteriori de D_{2i}^2 per a cada un dels cinc models no jeràrquics considerats per els resultats a les eleccions al Parlament de Catalunya del 2003 a Barcelona ciutat. 168

- 9.12 El primer gràfic representa els valors observats per $D_{ai}(y_i)$ a les zrp de cada districte a les eleccions al Parlament de Catalunya del 2003 a Barcelona ciutat i les quatre files següents corresponen a quatre rèpliques d'aquest estadístic simulades a partir de la predictiva a posteriori de cada model jeràrquic considerat. 176
- 9.13 El primer gràfic representa els valors observats per $D_{bi}(y_i)$ a les zrp de cada districte a les eleccions al Parlament de Catalunya del 2003 a Barcelona ciutat i les quatre files següents corresponen a quatre rèpliques d'aquest estadístic simulades a partir de la predictiva a posteriori de cada model jeràrquic considerat. 177
- 9.14 El primer gràfic representa els valors observats per $D_{ci}(y_i)$ a les zrp de cada districte a les eleccions al Parlament de Catalunya del 2003 a Barcelona ciutat i les quatre files següents corresponen a quatre rèpliques d'aquest estadístic simulades a partir de la predictiva a posteriori de cada model jeràrquic considerat. 178
- 9.15 El primer gràfic representa els valors observats per $D_{di}(y_i)$ a les zrp de cada districte a les eleccions al Parlament de Catalunya del 2003 a Barcelona ciutat i les quatre files següents corresponen a quatre rèpliques d'aquest estadístic simulades a partir de la predictiva a posteriori de cada model jeràrquic considerat. 179
- 9.16 La fila superior presenta els valors observats per $D_{ai}(y_i)$, $D_{bi}(y_i)$, $D_{ci}(y_i)$ i $D_{di}(y_i)$ a les eleccions al Parlament de Catalunya del 2003 a les zrp de Barcelona per districtes, i les altres files presenten una rèplica de les dades obtingudes a partir de la predictiva a posteriori de cada model jeràrquic considerat. 180
- 9.17 Distribució dels p valors, p_B , de les zrp estratificant per districte associats als estadístics $D_{ai}(y_i)$, $D_{bi}(y_i)$, $D_{ci}(y_i)$ i $D_{di}(y_i)$ a les eleccions al parlament de Catalunya del 2003 de Barcelona. 181
- 9.18 Els punts representen els valors observats per la mediana (med), el primer quartil (Q1), el tercer quartil (Q3) i el rang interquartílic (IQR) de la distribució de l'estadístic $D_{ai}(y_i) = \log(y_{i,CIU}/y_{i,PSC})$ a cada un dels districtes, i els diagrames de caixa representen la distribució predictiva a posteriori d'aquest estadístic per cada districte sota cada model jeràrquic considerat. 182

- 9.19 Els punts representen els valors observats per la mitjana, el mínim (min), el màxim (max) i el logaritme de la variància de la distribució de l'estadístic $D_{ai}(y_i) = \log(y_{i,CIU}/y_{i,PSC})$ a cada un dels districtes, i els diagrames de caixa representen la distribució predictiva a posteriori d'aquest estadístic per cada districte sota cada model jeràrquic considerat. 183
- 9.20 Els punts representen els valors observats per la mediana (med), el primer quartil (Q1), el tercer quartil (Q3) i el rang interquartílic (IQR) de la distribució de l'estadístic $D_{bi}(y_i) = \log(y_{i,CIU+PPC}/y_{i,PSC+ERC+ICV})$ a cada un dels districte, i els diagrames de caixa representen la distribució predictiva a posteriori d'aquest estadístic per cada districte sota cada model jeràrquic considerat. 184
- 9.21 Els punts representen els valors observats per la mitjana, el mínim (min), el màxim (max) i el logaritme de la variància de la distribució de l'estadístic $D_{bi}(y_i) = \log(y_{i,CIU+PPC}/y_{i,PSC+ERC+ICV})$ a cada un dels districtes, i els diagrames de caixa representen la distribució predictiva a posteriori d'aquest estadístic per cada districte sota cada model jeràrquic considerat. 185
- 9.22 Els punts representen els valors observats per la mediana (med), el primer quartil (Q1), el tercer quartil (Q3) i el rang interquartílic (IQR) de la distribució de l'estadístic $D_{ci}(y_i) = \log(y_{i,CIU+ERC}/y_{i,PSC+PPC+ICV})$ a cada un dels districtes, i els diagrames de caixa representen la distribució predictiva a posteriori d'aquest estadístic per cada districte sota cada model jeràrquic considerat. 186
- 9.23 Els punts representen els valors observats per la mitjana, el mínim (min), el màxim (max) i el logaritme de la variància de la distribució de l'estadístic $D_{ci}(y_i) = \log(y_{i,CIU+ERC}/y_{i,PSC+PPC+ICV})$ a cada un dels districtes, i els diagrames de caixa representen la distribució predictiva a posteriori d'aquest estadístic per cada districte sota cada model jeràrquic considerat. 187
- 9.24 Els punts representen els valors observats per la mediana (med), el primer quartil (Q1), el tercer quartil (Q3) i el rang interquartílic (IQR) de la distribució de l'estadístic $D_{di}(y_i) = \log(y_{i,abs}/N_i)$ a cada un dels districtes, i els diagrames de caixa representen la distribució predictiva a posteriori d'aquest estadístic per cada districte sota cada model jeràrquic considerat. 188

- 9.25 Els punts representen els valors observats per la mitjana, el mínim (min), el màxim (max) i el logaritme de la variància de la distribució de l'estadístic $D_{di}(y_i) = \log(y_{i,abs}/N_i)$ a cada un dels districtes, i els diagrames de caixa representen la distribució predictiva a posteriori d'aquest estadístic per cada districte sota cada model jeràrquic considerat. 189
- 9.26 Validació gràfica dels diferents models a partir de representacions ternàries dels perfils de tres columnes de la Taula 6.2. El primer gràfic correspon als perfils observats a les eleccions al Parlament de Catalunya del 2003 a Barcelona ciutat i les altres files de corresponen a rèpliques de les dades obtingudes simulant de la predictiva a posteriori de cada un dels cinc models jeràrquics considerats. 190
- 9.27 Validació gràfica dels diferents models a partir de representacions ternàries dels perfils de tres columnes de la Taula 6.2. El primer gràfic correspon als perfils observats a les eleccions al Parlament de Catalunya del 2003 a Barcelona ciutat i les altres files de corresponen a rèpliques de les dades obtingudes simulant de la predictiva a posteriori de cada un dels cinc models jeràrquics considerats. 191
- 9.28 Validació gràfica dels diferents models a partir de representacions ternàries dels perfils resultants d'agregar columnes de la Taula 6.2. El primer gràfic correspon als perfils observats a les eleccions al Parlament de Catalunya del 2003 a Barcelona ciutat i les altres files corresponen a rèpliques de les dades obtingudes simulant de la predictiva a posteriori de cada un dels cinc models jeràrquics considerats. 192
- 9.29 Distribucions a posteriori de la mesura de discrepància D_2 per als cinc models jeràrquics considerats. 193
- 9.30 Esperança a posteriori de D_{2i}^2 , $E[D_{2i}^2|y]$, per a cada zrp en funció del seu nombre d'electors per cada model jeràrquic considerat. Els eixos estan en escala logarítmica. 194
- 9.31 Distribució espacial de l'esperança a posteriori de D_{2i}^2 per a cada un dels cinc models jeràrquics considerats per els resultats a les eleccions al Parlament de Catalunya del 2003 a Barcelona ciutat. 195

9.32	Índex de Moran, calculat a partir de $E[D_{2i}^2 y]$, la seva distribució de referència sota el test de permutacions, i els correlogrames per als cinc models jeràrquics considerats.	196
10.1	Distribució a posteriori de la probabilitat de pertànyer a cada un dels clusters, $\pi(\omega_r y)$ per $r = 1, \dots, s$, per als models cluster no jeràrquics considerats.	201
10.2	Distribució a posteriori de la probabilitat de pertànyer a cada un dels clusters, $\pi(\omega_r y)$ per $r = 1, \dots, s$, per als models cluster jeràrquics considerats.	201
10.3	Distribució a posteriori de τ , $\pi(\tau_r y)$ per $r = 1, \dots, s$, per als cinc models jeràrquics considerats.	202
10.4	Classificació de les zrp en cada un dels clusters utilitzant com a criteri la moda a posteriori de les variables latents ζ_i , $Mo(\zeta_i y)$, sota tots els models cluster considerats.	203
10.5	Distribució a posteriori de les components del perfil de cada cluster per al model M_3 , $\pi(\theta_{rj} y)$ per $j = 1, \dots, 8$. Els tres clusters estan ordenats per ordre decreixent de percentatge de vot a CIU i creixent al PSC.	204
10.6	Distribució a posteriori de $\log(\theta_{rj}/\theta_{r'j})$ per al model M_3 per $r > r'$ i $j = 1, \dots, 8$	204
10.7	Distribució a posteriori de les components del perfil de cada cluster per al model M_{3J} , $\pi(\mu_{rj} y)$ per $j = 1, \dots, 8$. Els tres clusters estan ordenats per ordre decreixent de percentatge de vot a CIU i creixent al PSC.	205
10.8	Distribució a posteriori de $\log(\mu_{rj}/\mu_{r'j})$ per al model M_{3J} per $r > r'$ i $j = 1, \dots, 8$	205
10.9	Distribució a posteriori de les components del perfil de cada cluster per al model M_4 , $\pi(\theta_{rj} y)$ per $j = 1, \dots, 8$. Els quatre clusters estan ordenats per ordre decreixent de percentatge de vot a CIU i creixent al PSC.	206
10.10	Distribució a posteriori de $\log(\theta_{rj}/\theta_{r'j})$ per al model M_4 per $r > r'$ i $j = 1, \dots, 8$	206

- 10.11 Distribució a posteriori de les components del perfil de cada cluster per al model M_{4j} , $\pi(\mu_{rj}|y)$ per $j = 1, \dots, 8$. Els quatre clusters estan ordenats per ordre decreixent de percentatge de vot a CIU i creixent al PSC. 207
- 10.12 Distribució a posteriori de $\log(\mu_{rj}/\mu_{r'j})$ per al model M_{4j} per $r > r'$ i $j = 1, \dots, 8$ 207
- 11.1 La columna esquerra presenta els valors observats per $D_{ai}(y_i)$ a les eleccions al Parlament de Catalunya del 1992 al 2006 a nivell de zrp als 10 districtes de Barcelona, i les altres columnes de gràfic presenten una rèplica de les dades obtingudes a partir de la predictiva a posteriori dels models M_{2J} , M_{3J} i M_{4J} 213
- 11.2 La columna esquerra presenta els valors observats per $D_{bi}(y_i)$ a les eleccions al Parlament de Catalunya del 1992 al 2006 a nivell de zrp als 10 districtes de Barcelona, i les altres columnes de gràfic presenten una rèplica de les dades obtingudes a partir de la predictiva a posteriori dels models M_{2J} , M_{3J} i M_{4J} 214
- 11.3 La columna esquerra presenta els valors observats per $D_{ci}(y_i)$ a les eleccions al Parlament de Catalunya del 1992 al 2006 a nivell de zrp als 10 districtes de Barcelona, i les altres columnes de gràfic presenten una rèplica de les dades obtingudes a partir de la predictiva a posteriori dels models M_{2J} , M_{3J} i M_{4J} 215
- 11.4 La columna esquerra presenta els valors observats per $D_{di}(y_i)$ a les eleccions al Parlament de Catalunya del 1992 al 2006 a nivell de zrp als 10 districtes de Barcelona, i les altres columnes de gràfic presenten una rèplica de les dades obtingudes a partir de la predictiva a posteriori dels models M_{2J} , M_{3J} i M_{4J} 216
- 11.5 Els punts representen el valor observat de la mediana de l'estadístic $D_{ai}(y_i) = \log(y_{i,CIU}/y_{i,PSC})$ a les eleccions al Parlament de Catalunya del 1992 al 2006 a nivell de zrp als 10 districtes de Barcelona, i els diagrames de caixes representen les respectives distribucions predictiva a posteriori per cada model jeràrquic considerat. 217

- 11.6 Els punts representen el valor observat de la mediana de l'estadístic $D_{bi}(y_i) = \log(y_{i,CIU+PPC}/y_{i,PSC+ERC+ICV})$ a les eleccions al Parlament de Catalunya del 1992 al 2006 a nivell de zrp als 10 districtes de Barcelona, i els diagrames de caixes representen les respectives distribucions predictiva a posteriori per cada model jeràrquic considerat. 218
- 11.7 Els punts representen el valor observat de la mediana de l'estadístic $D_{ci}(y_i) = \log(y_{i,CIU+ERC}/y_{i,PSC+PPC+ICV})$ a les eleccions al Parlament de Catalunya del 1992 al 2006 a nivell de zrp als 10 districtes de Barcelona, i els diagrames de caixes representen les respectives distribucions predictiva a posteriori per cada model jeràrquic considerat. 219
- 11.8 Els punts representen el valor observat de la mediana de l'estadístic $D_{di}(y_i) = \log(y_{i,abs}/N_i)$ a les eleccions al Parlament de Catalunya de 1992 a 2006 a nivell de zrp als 10 districtes de Barcelona, i els diagrames de caixes representen les respectives distribucions predictiva a posteriori per cada model jeràrquic considerat. 220
- 11.9 La columna esquerra presenta la representació ternària dels valors observats dels perfils de tres columnes de la taula original a les eleccions al Parlament de Catalunya del 1992 al 2006 a nivell de zrp de Barcelona, i les altres columnes de gràfic presenten una rèplica de les dades obtingudes a partir de la predictiva a posteriori dels models M_{2J} , M_{3J} i M_{4J} 221
- 11.10 La columna esquerra presenta la representació ternària dels valors observats dels perfils de tres columnes de la taula original a les eleccions al Parlament de Catalunya del 1992 al 2006 a nivell de zrp de Barcelona, i les altres columnes de gràfic presenten una rèplica de les dades obtingudes a partir de la predictiva a posteriori dels models M_{2J} , M_{3J} i M_{4J} 222
- 11.11 Índex de Moran, calculat a partir de $E[D_{2i}^2|y]$ i la seva distribució de referència sota el test de permutacions per als models jeràrquics M_{2J} , M_{3J} i M_{4J} 223
- 11.12 Correlogrames fins al setè ordre, calculats utilitzant l'Índex de Moran a partir de $E[D_{2i}^2|y]$, per als models jeràrquics M_{2J} , M_{3J} i M_{4J} 224

12.1	Distribució a posteriori de la probabilitat de pertànyer a cada un dels clusters, $\pi(\omega_r y)$, per $r = 1, 2, 3$, sota el model M_{3J} pels resultats de cada una de les cinc darreres eleccions.	231
12.2	Distribució a posteriori del grau d'heterogeneïtat, $\pi(\tau_r y)$, per $r = 1, 2, 3$, sota el model M_{3J} pels resultats de cada una de les cinc darreres eleccions.	231
12.3	Classificació de les zrp utilitzant com a criteri la moda a posteriori de les variables latents ζ_i , $Mo(\zeta_i y)$, sota el model M_{3J} pels resultats de cada una de les cinc darreres eleccions.	232
12.4	Canvis en la classificació de les zrp als tres clusters entre dues eleccions consecutives, en funció de la direcció del canvi.	233
12.5	Distribució a posteriori dels components del perfil de probabilitat esperada per cada cluster sota el model M_{3J} , $\pi(\mu_{rj} y)$ per $r = 1, 2, 3$ i per $j = 1, \dots, 8$ o 9, pels resultats a cada una de les cinc darreres eleccions.	234
12.6	Distribució a posteriori de $\log(\mu_{rj}/\mu_{r'j})$ per $r > r'$ sota el model M_{3J} , pels resultats a cada una de les cinc darreres eleccions.	235
12.7	Representació ternària dels perfils resultants d'agregar columnes de la taula de dades observades. Cada zrp s'ha colorejat en funció al cluster al que han estat classificades seguint el criteri de la moda a posteriori sota el model M_{3J}	236
12.8	Representacions ternàries de (12.1), (12.2) i (12.3) sota el model M_{3J} . Cada punt es pot considerar com el valor més representatiu de cadascun dels tres clusters de la Figura 12.7.	237
12.9	Distribució a posteriori de la probabilitat de pertànyer a cada un dels clusters, $\pi(\omega_r y)$, per $r = 1, 2, 3, 4$, sota el model M_{4J} pels resultats de cada una de les cinc darreres eleccions.	242
12.10	Distribució a posteriori del grau d'heterogeneïtat, $\pi(\tau_r y)$, per $r = 1, 2, 3, 4$, sota el model M_{4J} pels resultats de cada una de les cinc darreres eleccions.	243
12.11	Classificació de les zrp utilitzant com a criteri la moda a posteriori de les variables latents ζ_i , $Mo(\zeta_i y)$, sota el model M_{4J} pels resultats de cada una de les cinc darreres eleccions.	244

- 12.12 Canvis en la classificació de les zrp als quatre clusters entre dues eleccions consecutives, en funció de la direcció del canvi. 245
- 12.13 Distribució a posteriori dels components del perfil de probabilitat esperada per cada cluster sota el model M_{4J} , $\pi(\mu_{rj}|y)$ per $r = 1, 2, 3, 4$ i per $j = 1, \dots, 8$ o 9 , pels resultats a cada una de les cinc darreres eleccions. . . . 246
- 12.14 Distribució a posteriori de $\log(\mu_{rj}/\mu_{r'j})$ per $r > r'$ sota el model M_{4J} , pels resultats a cada una de les cinc darreres eleccions. 247
- 12.15 Representació ternària dels perfils resultants d'agregar columnes de la taula de dades observades. Cada zrp s'ha colorejat en funció al cluster al que han estat classificades seguint el criteri de la moda a posteriori sota el model M_{4J} 248
- 12.16 Representacions ternàries de (12.1), (12.2) i (12.3) sota el model M_{4J} . Cada punt es pot considerar com el valor més representatiu de cadascun dels tres clusters de la Figura 12.15. 249
- 12.17 Distribució a posteriori de ζ_i , $\pi(\zeta_i|y)$, sota el model M_{3J} per una serie de 10 zrp escollides de forma que n'hi ha una per cada districte. 252
- 12.18 Distribució a posteriori de ζ_i , $\pi(\zeta_i|y)$, sota el model M_{4J} per una serie de 10 zrp escollides de forma que n'hi ha una per cada districte. 252
- A.1 Traces mostrals per per les 5000 primeres simulacions per als paràmetres μ_{rj} , per $j = 1 \dots, 8$ i per $r = 1, 2, 3$ 258
- A.2 Traces mostrals per les 5000 primeres simulacions per als paràmetres τ_r i ω_r , per $r = 1, 2, 3$ 259
- A.3 Traces mostrals per a les primeres 50 variables latents ζ_r , per $r = 1, \dots, 50$, per les primeres 5000 simulacions. 260
- A.4 Autocorrelació mostral per als paràmetres μ_{rj} , per $j = 1 \dots, 8$ i per $r = 1, 2, 3$, per les 5000 primeres simulacions. Les tres primeres columnes corresponen a la primera cadena i les tres darreres columnes corresponen a la segona cadena. 261

- A.5 Autocorrelació mostral per als paràmetres τ_r i ω_r , per per $r = 1, 2, 3$, per les 5000 primeres simulacions. Les dues primeres columnes corresponen a la primera cadena i les dues darreres columnes corresponen a la segona cadena. 262
- A.6 Traces mostrals per als paràmetres μ_{rj} , per $j = 1 \dots, 8$ i per $r = 1, 2, 3$, per les 1500 simulacions de cada una de les cadenes obtingudes a partir de monitoritzar-ne una de cada 100 d'un total de 150000 simulacions per a cada cadena. El període d'esclafament havia estat de 50000 simulacions en cada cadena. 264
- A.7 Traces mostrals per als paràmetres τ_r i ω_r , per $r = 1, 2, 3$, per les 1500 simulacions de cada una de les cadenes obtingudes a partir de monitoritzar-ne una de cada 100 d'un total de 150000 simulacions per a cada cadena. El període d'esclafament havia estat de 50000 simulacions en cada cadena. 265
- A.8 Traces mostrals per a les primeres 50 variables latents ζ_r , per $r = 1, \dots, 50$, per les 1500 simulacions de obtingudes a partir de monitoritzar-ne una de cada 100 d'un total de 150000 simulacions per a cada cadena. El període d'esclafament havia estat de 50000 simulacions en cada cadena. 266
- A.9 Autocorrelació mostral per als paràmetres μ_{rj} , per $j = 1 \dots, 8$ i per $r = 1, 2, 3$, per les 1500 simulacions de cada una de les cadenes obtingudes a partir de monitoritzar-ne una de cada 100 d'un total de 150000 simulacions per a cada cadena. Les tres primeres columnes corresponen a la primera cadena i les tres darreres columnes corresponen a la segona cadena. . . . 267
- A.10 Autocorrelació mostral per als paràmetres τ_r i ω_r , per per $r = 1, 2, 3$, per les 1500 simulacions de cada una de les cadenes obtingudes a partir de monitoritzar-ne una de cada 100 d'un total de 150000 simulacions per a cada cadena. Les dues primeres columnes corresponen a la primera cadena i les dues darreres columnes corresponen a la segona cadena. 268
- C.1 Classificació dels districtes a cada un dels dos clusters per al model no jeràrquic M_2 , basada en la moda a posteriori de les variables latents ζ_i per als resultats a les eleccions al Parlament de Catalunya del 1992 al 2006. 308

- C.2 Classificació dels districtes a cada un dels dos clusters per al model no jeràrquic M_2 basada en la moda a posteriori de les variables latents ζ_i per als resultats a les eleccions del 1992 al 2006 a la ciutat de Barcelona. 309
- C.3 Classificació dels districtes de les ciutats de Girona, Lleida, Tarragona i Sabadell a cada un dels dos clusters per al model no jeràrquic M_2 310
- C.4 Distribució a posteriori dels perfils de cada cluster, $\pi(\theta_{1j}|y)$ i $\pi(\theta_{2j}|y)$, del logaritme del quocient dels perfils, $\pi(\log \frac{\theta_{2j}}{\theta_{1j}}|y)$ per $j = 1, \dots, k$ i de la proporció de districtes del cluster 1, $\pi(\omega_1|y)$, per als resultats de totes les eleccions a Catalunya utilitzant el model no jeràrquic M_2 311
- C.5 Evolució estratificada per partit de les distribucions a posteriori dels perfils i de la seva distància entre els clusters, utilitzant el model no jeràrquic M_2 per les dades de tot Catalunya. 312
- C.6 Distribució espacial de l'esperança a posteriori de D_{2i}^2 sota el model M_2 pels resultats de les cinc darreres eleccions al Parlament de Catalunya. Per a la representació s'han categoritzat els valors prenent com a punt de tall les quartiles de l'any 1992. 313
- C.7 Distribució espacial de l'esperança a posteriori de D_{2i}^2 sota el model M_2 pels resultats de les cinc darreres eleccions al Parlament de Catalunya a la ciutat de Barcelona. Per a la representació s'han categoritzat els valors prenent com a punt de tall les quartiles de l'any 1992 del global de Catalunya. 314
- C.8 Esperança a posteriori de D_{2i}^2 , $E[D_{2i}^2|y]$, en funció del nombre d'electors del districte i del nombre d'electors del municipi al que pertanyen, en escala logarítmica, utilitzant el model no jeràrquic M_2 per les dades de tot Catalunya. 315
- C.9 Classificació dels districtes utilitzant com a criteri la moda a posteriori de les variables latents ζ_i , $Mo(\zeta_i|y)$, pels models no jeràrquics M_{3a} , M_{3b} i M_5 per les eleccions de 2003. 316
- C.10 A l'esquerra l'esperança a posteriori de D_{2i}^2 , $E[D_{2i}^2|y]$, en funció del nombre d'electors del districte, en escala logarítmica, i amb línia discontinua el valor esperat de la a posteriori $E[\widetilde{D}_{2i}^2|y]$, i a la dreta la distribució de $p_B^{D_{2i}^2}$ per a les dades de tot Catalunya del 2003. 317

C.11 Distribució de $p_B^{D_{2i}^2}$ per a cada categoria per al model no jeràrquic de 2 clusters, M_2 , Catalunya 2003.	318
--	-----