

REPRESENTACIÓ I CLÚSTERING QUALITATIU  
DE CORBES DE GLUCOSA

CARLES ESCRIG ROYO

TUTORS

LLEDÓ MUSEROS

ISMAEL SANZ



UNIVERSITAT  
JAUME·I

Treball Fi de Màster

Màster en Sistemes Intel·ligents  
Escola Superior de Tecnologia i Ciències Experimentals  
Universitat Jaume I

Novembre 2014 – Castelló de la Plana



## ABSTRACT

---

*Diabetes Mellitus* is a disease that affects more than 8% of the world population, and this number is expected to continue growing. It is primarily characterized by the presence of high levels of blood glucose, either due to the absence of insulin or that it does not take effect. There are many microvascular and macrovascular complications associated with it that can lead to kidney failure, blindness, ulcers and heart attacks, with high mortality.

The traditional therapy is to maintain manual control (blood analysis) of glycemic levels and intake of exogenous insulin. The continuous advancement has led us to be closer to get a technological substitute simulating the endocrine function of the pancreas: the "artificial pancreas". It consists of a continuous glucose monitoring system, an insulin infusion pump and a control algorithm. At the moment, the predictive capabilities of the monitoring system and the integration of the whole system are the weakest points. Food intake and increased subsequent glycemic level is one of the most critical moments. Improving the knowledge of the effects caused by a mixed meal would advance the predictive reliability of the monitoring systems, which are a fundamental part of the artificial pancreas.

In this research we study patterns of absorption of glucose, provided by the project ClosedLoop4Meal (UdG, UPV). We make a selection of relevant features, and evaluate different ways to discretize them. The resulting qualitative description is used to explore the meals' behavior using clustering techniques, and to compare the suitability and performance of the qualitative approach versus the purely quantitative one. Lastly, we present a description in natural language of the models of glucose absorption and suggest some improvements to the model designed.

## RESUMEN

---

La *Diabetes Mellitus* es una enfermedad que afecta a más de un 8 % de la población mundial, y se prevé que este número continúe creciendo. Está caracterizada primordialmente por la presencia de altos niveles de glucosa en la sangre, bien debido a la ausencia de insulina o bien a que no tiene efecto. Existen múltiples complicaciones microvasculares y macrovasculares asociadas, que pueden conducir a insuficiencia renal, ceguera, úlceras e infarto de miocardio, con una alta mortalidad.

El terapia tradicional consiste en mantener un control manual (análisis de sangre) del nivel glucémico, y el tratamiento con insulina exógena. El avance continuo nos ha llevado a estar más cerca de obtener un sustituto tecnológico que simule la función endocrina del páncreas: el "páncreas artificial". Este está constituido por un sistema de monitorización continua de glucosa, una bomba de infusión continua de insulina y un algoritmo de control. Por ahora, las capacidades predictivas del sistema de monitorización y la integración de todo el conjunto, son los puntos más débiles. La ingestión de alimentos y el aumento del nivel glucémico posterior es uno de los momentos más críticos. Mejorar el conocimiento de los efectos que provoca una comida, permitiría avanzar en la fiabilidad predictiva de los sistemas de monitorización, que son una parte fundamental del páncreas artificial.

En el presente trabajo se estudian modelos de absorción de glucosa, proporcionados por el proyecto ClosedLoop4Meal (UdG, UPV). Se hace una selección de las características relevantes, y se evalúan diferentes maneras de discretizarlas. Se utiliza la descripción cualitativa resultante para explorar mediante técnicas de clustering el comportamiento de las comidas, y para comparar la idoneidad y rendimiento del modelo cualitativo frente el puramente cuantitativo. Por último también se hace una descripción en lenguaje natural de los modelos de absorción de glucosa, y se sugieren mejoras en el modelo diseñado.

## RESUM

---

La *Diabetis Mellitus* és una malaltia que afecta a més d'un 8% de la població mundial, i es preveu que aquest nombre continue creixent. Està caracteritzada primordialment per la presència d'alts nivells de glucosa a la sang, bé degut a l'absència d'insulina o bé a que no té efecte. Existeixen múltiples complicacions microvasculars i macrovasculars associades, que poden conduir a insuficiència renal, ceguesa, úlceres i infart de miocardi, amb una alta mortalitat.

La teràpia tradicional consisteix en mantenir un control manual (anàlisi de sang) del nivell glucèmic, i el tractament amb insulina exògena. L'avanç continu ens ha portat a estar més a prop d'obtenir un substitut tecnològic que simule la funció endocrina del pàncrees: el "pàncrees artificial". Aquest està constituït per un sistema de monitorització continua de glucosa, una bomba d'infusió continua d'insulina i un algorisme de control. Per ara, les capacitats predictives del sistema de monitorització i la integració de tot el conjunt, són els punts més febles. La ingestió d'aliments i l'augment del nivell glucèmic posterior és un dels moments més crítics. Millorar el coneixement dels efectes que provoca una menjada, permetria avançar en la fiabilitat predictiva dels sistemes de monitorització, que són una part fonamental del pàncrees artificial.

En el present treball s'estudien models d'absorció de glucosa, proporcionats pel projecte ClosedLoop4Meal (UdG, UPV). Es fa una selecció de les característiques rellevants, i s'avaluen diferents maneres de discretitzar-les. S'utilitza la descripció qualitativa resultant per explorar mitjançant tècniques de clústering el comportament dels menjars, i per comparar la idoneïtat i rendiment del model qualitatiu front el purament quantitatiu. Per últim també es fa una descripció en llenguatge natural dels models d'absorció de glucosa, i es suggereixen millores en el model dissenyat.



*I don't want to believe. I want to know.*

— Carl Sagan

## AGRAÏMENTS

---

El primer agraïment va adreçat a Lledó Museros i Ismael Sanz, tutors d'aquest Treball Fi de Màster, per haver-me oferit aquest tema d'investigació amb aplicació directa sobre problemàtiques humanes reals. També per la gran paciència mostrada durant tot el procés de realització del mateix, els consells i l'atenta revisió dels materials, fins i tot quan els terminis de temps eren més que justs. He de donar gràcies també a tots els seus col·legues, especialment els responsables del projecte ClosedLoop4Meals i Yenny Leal, perquè el meu treball es construeix per complet sobre el seu.

A Carmela, per la seua continua companyia, paciència, comprensió, amor, i sobretot per compartir il·lusions. A Trini i Carlos, els meus pares, per haver intentat fer de mi una persona *de profit*. A tota la nostra família, en general, per no ser res i ser tant al mateix temps. A Empar, la meua mestra, i Paco, perquè el seu compromís és una referència i una continua inspiració.

Al Departament d'Enginyeria i Ciència dels Computadors, al Departament de Llenguatges i Sistemes Informàtics, i totes les bones professores i professors (de molts departaments) que he tingut durant els meus anys de formació acadèmica a la Universitat Jaume I. A Rafael Berlanga, coordinador del Màster en Sistemes Intel·ligents, pel seu suport.

A Richard Stallman, per mostrar al món que *sí se pot*. A Douglas Crockford, per exposar el millor de JavaScript i despertar en mi una profunda estima pels llenguatges de programació. A Manuel de Pedrolo, perquè llegint-lo em sento part d'un país. Finalment, a Carl Sagan, perquè gràcies a ell puc veure i gaudir de tota la bellesa que ens envolta.





# ÍNDIX

---

<b>i</b>	<b>INTRODUCCIÓ</b>	<b>1</b>
1	INTRODUCCIÓ	3
1.1	Motivació	3
1.2	Objectius	4
1.3	Planificació	5
1.4	Materials	7
1.5	Estructura del treball	8
2	INFORMACIÓ FONAMENTAL	9
2.1	Visió de conjunt sobre la DM	9
2.2	Raonament qualitatiu	14
2.3	Clústering i mesures de qualitat	15
2.3.1	Clústering jeràrquic	16
2.3.2	Mesures de qualitat	17
<b>ii</b>	<b>DESENVOLUPAMENT</b>	<b>19</b>
3	DISSENY I DESENVOLUPAMENT	21
3.1	Pre-processament	21
3.1.1	L'escenari inicial	21
3.1.2	Formulació d'un model	22
3.2	Clústering	30
3.2.1	Similitud entre períodes	30
3.2.2	Similitud entre corbes	32
3.2.3	Mètode de clústering	33
3.3	Validació	33
4	EXPERIMENTACIÓ I DISCUSSIÓ	35
4.1	Resultats	35
4.1.1	Observació de la qualitat	35
4.1.2	Comparació amb altres clústers «de referència»	46
4.1.3	Descripció en llenguatge natural	50
4.2	Discussió	58
<b>iii</b>	<b>CONCLUSIONS I TREBALL FUTUR</b>	<b>61</b>
5	CONCLUSIONS I TREBALL FUTUR	63
5.1	Conclusions	63
5.2	Treball futur	64
5.2.1	Treball futur immediat	64
5.2.2	Treball futur en segon terme	65
<b>iv</b>	<b>APÈNDIX</b>	<b>67</b>
A	APROXIMACIONS ALTRENAITIVES	69
A.1	Aproximació estadística	69
A.1.1	Model	69

A.1.2	Mètode	70	
A.1.3	Resultats	70	
A.2	Aproximació relativa	75	
A.2.1	Model	75	
A.2.2	Mètode	76	
A.2.3	Resultats	76	
B	FIGURES COMPLETES DELS RESULTATS COMENTATS		85
	BIBLIOGRAFIA	107	

## ÍNDIX DE FIGURES

---

Figura 1	Gantt de la planificació	7
Figura 2	Quant a la insulina	10
Figura 3	Persones amb diabetis al món	11
Figura 4	Bucle de control de glucosa	12
Figura 5	Exemple d'un dendrograma	16
Figura 6	Exemple de corbes de glucosa, insulina i $Ra$	22
Figura 7	Exemple dels períodes d'una corba	24
Figura 8	Mecanisme d'absorció de períodes	25
Figure 9	Histogrames de $Ra$	26
Figure 10	Histogrames de $\Delta Ra$	28
Figure 11	Histogrames de la durada dels períodes	30
Figura 12	Etiquetes del clústering DI (mostra)	37
Figura 13	Etiquetes del clústering agregat (mostra)	39
Figura 14	Etiquetes del clústering agregat, 2a versió (mostra)	40
Figura 15	Grups del clústering DI+Silueta (resum)	43
Figura 16	Grups del 14è clústering DI+Silueta (resum)	44
Figura 17	Grups del clústering de referència (resum)	47
Figura 18	Grups del clústering elegit (resum)	49
Figura 19	Grups del clústering aprox. estadística (1)	71
Figura 20	Grups del clústering aprox. estadística (2)	72
Figura 21	Etiquetes del clústering aprox. estadística (1)	73
Figura 22	Etiquetes del clústering aprox. estadística (2)	74
Figura 23	Grups del clústering aprox. relativa (1)	78
Figura 24	Grups del clústering aprox. relativa (2)	79
Figura 25	Etiquetes del clústering aprox. relativa (1)	80
Figura 26	Etiquetes del clústering aprox. relativa (2)	81
Figura 27	Grups del clústering aprox. relativa alt.	82
Figura 28	Etiquetes del clústering aprox. relativa alt. (1)	83
Figura 29	Etiquetes del clústering aprox. relativa alt. (2)	84
Figura 30	Etiquetes del clústering DI (1)	86
Figura 31	Etiquetes del clústering DI (2)	87
Figura 32	Etiquetes del clústering agregat (1)	88
Figura 33	Etiquetes del clústering agregat (2)	89
Figura 34	Etiquetes del clústering agregat, 2a versió (1)	90
Figura 35	Etiquetes del clústering agregat, 2a versió (2)	91
Figura 36	Grups del clústering DI+Silueta (1)	92
Figura 37	Grups del clústering DI+Silueta (2)	93
Figura 38	Etiquetes del clústering DI+Silueta (1)	94
Figura 39	Etiquetes del clústering DI+Silueta (2)	95
Figura 40	Grups del 14è clústering DI+Silueta (1)	96

Figura 41	Grups del 14è clústering DI+Silueta (2)	97
Figura 42	Etiquetes del 14è clústering DI+Silueta (1)	98
Figura 43	Etiquetes del 14è clústering DI+Silueta (2)	99
Figura 44	Grups del clústering de referència (1)	100
Figura 45	Grups del clústering de referència (2)	101
Figura 46	Grups del clústering elegit (1)	102
Figura 47	Grups del clústering elegit (2)	103
Figura 48	Etiquetes del clústering elegit (1)	104
Figura 49	Etiquetes del clústering elegit (2)	105

## ÍNDIX DE TAULES

---

Taula 1	Planificació del projecte	6
Taula 2	Matriu de dissimilitud amb veïnatge conceptual	31
Taula 3	Matriu de dissimilitud amb distància entre intervals	32
Taula 4	Equivalències dels períodes	33
Taula 5	Els 10 millors resultats	36
Taula 6	Resultats per Connectivitat	36
Taula 7	Resultats per DI	37
Taula 8	Els 10 millors resultats, 2a versió	38
Taula 9	Els 20 millors resultats, 3a versió	39
Taula 10	Resultats per DI, 2a versió	41
Taula 11	Resultats per Silueta	41
Taula 12	Els 15 millors resultats per DI+Silueta	42
Taula 13	Resultats per VI amb la «referència»	46
Taula 14	Taula de modes d'un clústering	59
Taula 15	Validació aprox. relativa	76

## ACRONYMS

---

AP	pàncrees artificial
AUC	àrea sota la corba
BG	glucosa en plasma
CGMS	Sistemes de Monitorització Continua de Glucosa
CHO	carbohidrats

CV	Coeficient de Variació de Pearson
DI	Índex de Dunn
DM	Diabetis Mellitus
DM-1	Diabetis Mellitus Tipus 1
DM-2	Diabetis Mellitus Tipus 2
DMG	Diabetis Gestacional
IDE	entorn integrat de desenvolupament
IDF	Federació Internacional de Diabetis
SMBG	Sistema d'Auto-Monitorització de Glucosa
PAM	Partitioning Around Medoids
VI	variació d'informació



Part I

INTRODUCCIÓ





## INTRODUCCIÓ

---

### 1.1 MOTIVACIÓ

La glucosa és un monosacàrid que les cèl·lules utilitzen com a font d'energia i com a intermediari metabòlic. El seu nivell de concentració en el flux sanguini és control·lat de manera precisa per dues hormones produïdes al pàncrees: el glucagó i la insulina. La primera eleva el nivell de glucosa en sang, i la segona el redueix. La funció més important de la insulina és fer que les cèl·lules del fetge, dels músculs i del teixit adipós agafen i emmagatzemen la glucosa de la sang (provinent, per exemple, dels aliments digerits).

La Diabetis Mellitus (DM) és una malaltia metabòlica caracteritzada primordialment per la presència d'alts nivells de glucosa a la sang (hiperglucèmia) [19]. Existeixen múltiples complicacions microvasculars i macrovasculars associades amb la DM. Les complicacions microvasculars inclouen la nefropatia, que porta a la insuficiència renal, la retinopatia, que porta a la ceguesa i la neuropatia, que porta a úlceres o amputació del peu. Entre les macrovasculars trobem l'infart de miocardi i els accidents macrovasculars, que tenen associada un alta mortalitat [13].

Segons els càlculs més recents de la Federació Internacional de Diabetis (IDF) 8,3% dels adults –382 milions de persones– tenen diabetis, i el nombre de persones que la pateixen s'incrementarà en més de 592 milions de persones en menys de 25 anys. El 80% del total d'afectades viuen en països d'ingressos mitjans i baixos, on l'epidèmia s'accelera. A l'Estat espanyol es calcula que hi ha cap a 4 milions de casos de diabetis, representant una prevalença de 10,83% [1].

Per a cobrir les deficiències d'insulina al cos, els pacients es tracten mitjançant teràpia amb insulina. Tanmateix, aquesta teràpia intensiva queda limitada pel risc de sofrir hipoglucèmies. La insulina s'administra mitjançant múltiples injeccions diàries o mitjançant infusió continua d'insulina subcutània. Abans de cada menjada, els pacients mesuren el seu nivell de glucosa en plasma amb un Sistema d'Auto-Monitorització de Glucosa (SMBG), per tal d'ajustar la dosi d'insulina en relació a la quantitat de carbohidrats (CHO) que es pretenen ingerir.

Amb les poques mostres diàries que subministren els SMBG no hi ha prou per monitoritzar tota la variabilitat glucèmica ni, per descomptat, detectar excursions crítiques –hiperglucèmies postprandials<sup>1</sup> i hipoglucèmies per excés en la dosi d'insulina–. Per aquesta raó han sorgit els Sistemes de Monitorització Continua de Glucosa (CGMS),

---

<sup>1</sup> Que ocorre després de menjar.

que intenten estimar la glucosa en plasma (BG) de manera continuada (d'entre 1 a 5 minuts).

Encara no hi ha cap CGMS acceptat com a substitut dels tradicionals SMBG, degut a les imprecisions en les estimacions de glucosa. Tot i això, el seu ràpid desenvolupament ha impulsat la ideació de l'anomenat pàncrees artificial (AP), un sistema de control de llaç tancat on s'integra un CGMS, un algorisme de control i una bomba d'infusió continua d'insulina.

El projecte ClosedLoop4Meal<sup>2</sup>, desenvolupat conjuntament per la Universitat de Girona i la Universitat Politècnica de València, té com a objectiu «el desenvolupament de noves estratègies segures i eficients per al control de la glucosa postprandial als pacients amb DM-1». La primera fita que estableix el projecte per aconseguir-ho és: «Augmentar la comprensió de l'efecte glicèmic d'un menjar mesclat. Estudiar models d'absorció de glucosa després de les menjades i incorporar-los a nous esquemes de control anticipatiu».

## 1.2 OBJECTIUS

Partint de models d'absorció de glucosa postprandial del projecte ClosedLoop4Meal, l'objectiu general d'aquest treball és proporcionar una descripció qualitativa dels mateixos, i a partir d'aquesta descripció obtindre una classificació dels models. La descripció qualitativa serà fàcilment transformable a llenguatge humà, per tal de facilitar el raonament i la comprensió de l'efecte glicèmic dels menjars. La descripció d'aquests models mitjançant l'abstracció de les dades discretes, podria ajudar a reconèixer respostes similars en els diferents menjars, i per tant representaria un avanç en la informació disponible per a les estimacions de glucosa en plasma.

Com a objectius específics es poden definir:

- Estudi bibliogràfic sobre models qualitatius, treballs relacionats amb el control de la glucosa, sobre clústering i sobre com mesurar la qualitat dels clústers obtinguts.
- Estudi del llenguatge de programació R, amb el qual es desenvoluparà el projecte, així com de totes les biblioteques o extensions necessàries.
- Desenvolupar almenys una teoria per descriure una corba de glucosa de manera qualitativa.
- Obtenir diversos clústerings amb les corbes descrites.
- Comparar els clústerings per determinar quin és el millor, i/o les característiques més rellevants del mateixos.

<sup>2</sup> <http://closedloop4meal.org>

En cas de que la web del projecte deixi d'estar disponible en línia, es pot consultar una còpia a l'Internet Archive (6 de setembre de 2014).

- Descriure en llenguatge natural les característiques de cada classe del millor clúster determinat.

### 1.3 PLANIFICACIÓ

Es pot veure un resum de les tasques previstes i les hores estimades a la Taula 1 a la pàgina següent. Alguns valors de la columna «Hores Final», incorporen una titlla (~) indicant que el valor és aproximat. A la Figura 1 a la pàgina 7 hi ha un diagrama de Gantt, amb la planificació temporal estimada per dedicar 5 hores per dia. La durada total del projecte ha estat manifestament superior a les 12 setmanes que s'hi planifiquen perquè, en primer lloc s'han dedicat una mitjana diària d'hores menor (aproximadament 2 h/dia), i en segon lloc s'han produït desviacions imprevistes en algunes tasques, que a continuació comentem.

R és un llenguatge que té una sintaxi relativament similar a C, per tant té un nivell d'entrada molt baix per a la gran majoria de persones que tenen algun coneixement de programació<sup>3</sup>. Havíem planificat unes 20 hores per adquirir suficients coneixements per a poder llegir dades, analitzar-les i generar alguns gràfics. Tanmateix les exigències de l'estudi han anat en augment a mesura que avançava el projecte. La fase de desenvolupament ha requerit un coneixement més profund –de l'estimat– de les estructures de dades essencials d'R. A més de la sintaxi i les estructures de control, els tipus de dades bàsics són l'altra «matèria primera» necessària per poder programar eficientment. D'altra banda, tot i que el llenguatge incorpora molts mètodes gràfics d'ús habitual en el camp estadístic, la generació de gràfics més personalitzats també ha exigut més treball del previst. Aquest últim aspecte, afecta a l'augment d'hores en les tasques «Estudi R» i «Memòria».

Tal volta la tasca «disseny de la solució» retroalimentada per la fase «Testing» s'ha dilatat un pèl més del planificat, per provar diverses opcions per descriure les corbes fins aconseguir una descripció qualitativa que fos mínimament acceptable. Però sobretot perquè cap al final del projecte hem intentant aconseguir un clústering més significatiu –encara que no era un objectiu explícit– que mostres correlacions entre els grups i els macronutrients. Això també s'ha inclòs en aquestes tasques, on ens hem entretingut –de manera completament voluntària– implementant diferents «exploracions» mixtes numèriques i qualitatives. Els resultats d'aquestes exploracions no formen part del nucli del projecte i s'han incorporat a l'Apèndix A.

<sup>3</sup> Entenem que C ha estat un dels llenguatges de més èxit i difusió en tots els àmbits, i especialment en estudis d'informàtica segueix sent un estàndard sòlid d'ensenyament.

N	TASCA	H. EST.	FINAL	OBJECTIU
1	Estudi DM	30	30	Estudiar la malaltia, estat de la qüestió, i tota la informació necessària per a poder decidir què és rellevant per al problema.
2	Estudi models qualitatis	30	30	Estudiar estat de la qüestió, tècniques de discretització, ferramentes disponibles, tècniques de classificació qualitatives, etc.
3	Estudi R	20	~30	Adquirir suficient coneixement bàsic del llenguatge per poder desenvolupar un script amb una mínima qualitat i reusable.
4	Disseny de la solució	20	~30	Aplicar els coneixements obtinguts en les tasques 1 i 2 per obtenir un model qualitatiu: intervals, rangs, valors, etc.
5	Implementació	80	80	Desenvolupar un script R que tinga com a entrada els models d'absorció de glucosa, com a eixida una descripció qualitativa dels mateixos, i permeta fer classificació qualitativa.
6	Testing	50	~80	Contrastat els resultats, mesurar la qualitat, millorar paràmetres, etc.
7	Memòria	50	80	Redactar i maquetar la memòria del treball.
8	Presentació	20	20	Preparar la presentació per a la defensa del projecte i realitzar-la.
TOTAL		300	380	

Taula 1: Planificació del projecte.

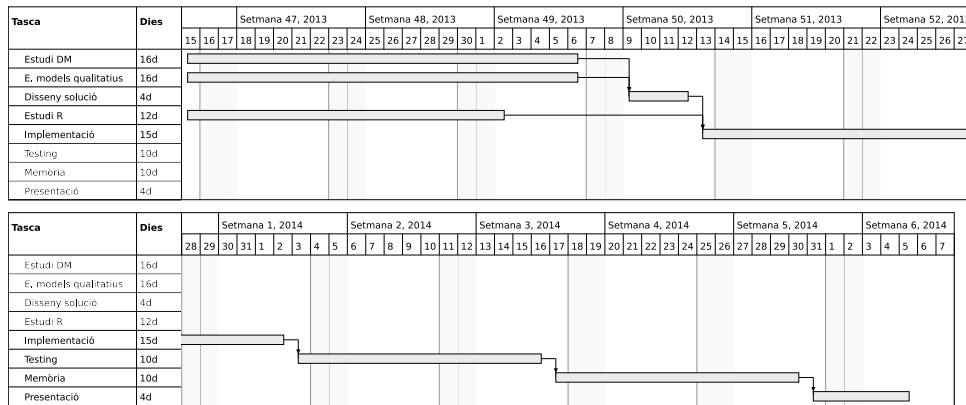


Figura 1: Diagrama de Gantt amb la planificació estimada de les tasques per a un volum de treball de 5 hores per dia.

#### 1.4 MATERIALS

Tots els treballs relacionats amb aquest projecte s'han realitzat en un ordinador Lenovo Thinkpad X200, amb una CPU Intel® Core™2 Duo P8600 @ 2.40GHz, 4GB de memòria RAM, i sobre un sistema operatiu GNU/Linux (Arch Linux per arquitectures x86 de 64 bits).

El programari essencial que hem usat ha estat R, mitjançant l'entorn integrat de desenvolupament (IDE) RStudio<sup>4</sup>, en diverses versions. Les últimes on s'ha executat el codi del projecte són la versió 0.98.1062 d'RStudio i la versió 3.1.2 d'R. A continuació es llisten les biblioteques R usades que no formen part de la distribució estàndard:

- `rootSolve`, per trobar les arrels d'equacions no lineals.
- `clValid`, per a la validació dels resultats dels clústerings.
- `RankAggreg`, per l'agregació amb pes de llistes parcialment ordenades.
- `mcclust`, per la variació d'informació (VI) entre clústers.
- `gridBase` i `gridExtra`, per manipulació gràfica (*viewports* i *taules*).
- `psych`, per generar estadístics descriptius bàsics (similar a `summary`, però més complet).

Ha estat molt útil la documentació i manuals de projecte R, especialment el manual "*An Introduction to R*".

Aquesta memòria ha estat redactada i maquetada amb LyX versió 2.1.1, usant la plantilla `classicthesis`<sup>5</sup> d'André Miede, inspirada

<sup>4</sup> <http://www.rstudio.com>

<sup>5</sup> `classicthesis` està disponible per a L<sup>A</sup>T<sub>E</sub>X i LyX:  
<http://code.google.com/p/classicthesis/>

per l'influent llibre sobre tipografia "*The Elements of Typographic Style*" de Robert Bringhurst.

Finalment, també hem fet ús del programa de gràfics vectorials Inkscape (versió 0.48) per manipular algunes de les imatges aquí presentades. Les diapositives de la presentació del projecte s'han realitzat amb aquest mateix programa i l'extensió Sozi<sup>6</sup>.

## 1.5 ESTRUCTURA DEL TREBALL

El Capítol 2 fa un repàs de la informació fonamental que cal conèixer per abordar el problema i complir els objectius que ens hem marcat. En la Secció 2.1 donem una visió de conjunt de la DM i dels sistemes de monitorització continua de glucosa. En la Secció 2.2 se sintetitzen els fonaments i passos a seguir per fer una descripció qualitativa d'un problema. Finalment, en la Secció 2.3 s'explica breument el procés de clústering i les mesures de qualitat que l'acompanyen.

El Capítol 3 descriu el procés de disseny d'una solució i com s'aplicarà. A la Secció 3.1 es fa un estudi de les característiques rellevants i de la discretització de les mateixes. A la Secció 3.2 es parla de la comparació entre corbes per a fer el clústering.

El Capítol 4 explica els resultats obtinguts. Primer observem la qualitat dels mateixos, i descrivim el procés de descart d'opcions invàlides. Després comparem els resultats amb el clústering de referència, i a la Secció 4.1.3 es fa una descripció en detall dels clústers obtinguts. Per últim, afegim algunes reflexions sobre els resultats.

Al Capítol 5 hi ha les conclusions obtingudes, i el possible treball futur a realitzar.

A l'Apèndix A a la pàgina 69 s'inclouen algunes de les solucions alternatives que hem explorat durant el projecte.

Al llarg d'aquesta memòria es comenten molts resultats que només van acompanyats amb figures parcials, per raons de maquetació. L'Apèndix B a la pàgina 85 recull la versió completa de totes aqueixes figures.

---

<sup>6</sup> <http://sozi.baierouge.fr>

## INFORMACIÓ FONAMENTAL

---

### 2.1 VISIÓ DE CONJUNT SOBRE LA DM

La **DM** és una malaltia crònica que consisteix en la incapacitat de produir insulina, o d'utilitzar-la eficaçment. La insulina és una hormona reguladora secretada per les cèl·lules *beta* dels Illots de Langerhans al pàncrees, que estimula el flux de transportadors de glucosa, permetent-ne la difusió a les cèl·lules a través de la membrana cel·lular al múscul i teixit adipós, així com l'emmagatzemament en forma de glucògen al fetge (Figura 2a a la pàgina següent). Tant l'absència d'insulina com la incapacitat d'usar-la amb normalitat tenen com a conseqüència l'aparició d'un estat fisiològic anomenat hiperglucèmia, que s'associa amb una major vulnerabilitat davant d'infeccions i disfunció multi-orgànica.

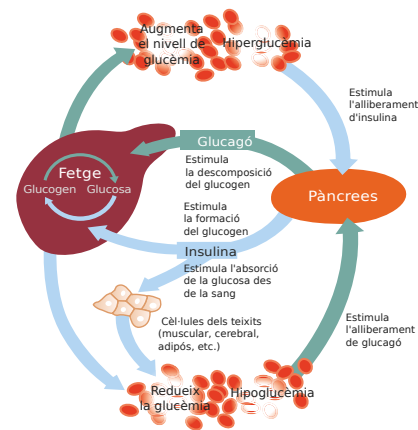
Hi ha tres tipus principals de diabetis:

- Diabetis Mellitus Tipus 1 (**DM-1**): Insuficient capacitat del pàncrees per produir l'hormona insulina, amb la conseqüent inhabilitat de les cel·lules per absorbir la glucosa en sang. És el resultat d'un procés autoimmune –d'inici sobtat–, en què el sistema de defensa del cos ataca les cèl·lules beta productores d'insulina. Se'n desconeix la causa. Sense insulina diària, una persona amb **DM-1** mor.
- Diabetis Mellitus Tipus 2 (**DM-2**): Consisteix en la incapacitat per a usar la insulina amb eficàcia. És la més comuna de les tres, però pot tardar molts anys en diagnosticar-se, quan l'efecte acumulatiu de l'excés de glucosa en sang ha desenvolupat complicacions visibles.
- Diabetis Gestacional (**DMG**): Desenvolupament d'una resistència a la insulina –similar a **DM-2**– durant l'embaràs. Sol ocórrer tard (pels volts de la setmana 24) quan el fetus ja està ben format, i per tant no és tan greu com els casos en què la mare té **DM-1** o **DM-2**.

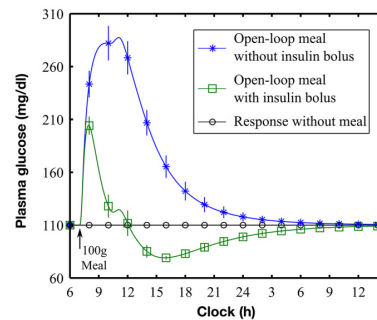
Per a una persona sana, els nivells de glucosa considerats normals estan entre 70–140 mg/dL (normoglucèmia), i el de glucosa basal<sup>1</sup> és de 90 mg/dL. Un pacient diabètic té un nivell de glucosa basal més alt, entre 110–140 mg/dL, i degut a activitats físiques, ingestió d'aliments, etc. és molt fàcil que aquest nivell augmente. L'objectiu

---

<sup>1</sup> Nivell de glucèmia després del dejuni nocturn.



(a) Producció i acció de la insulina. Imatge d'IDF [1].



(b) Resposta després d'una menjada amb (verd) i sense (blau) un bolus òptim. El bolus ha estat calculat per a mantenir el nivell mínim de  $BG \geq 70$  mg/dL. Imatge de Laxminarayan et al. [12]

Figura 2: Quant a la insulina.

dels pacients amb DM és controlar i mantenir el nivell de glucosa en el rang acceptable 70–180 mg/dL [13]. En la Figura 2b es pot apreciar com, després d'una menjada, s'ha mantingut el nivell de glucosa dintre del rang acceptable administrant un bolus d'insulina òptim. La base per als tractaments d'ajustament és el mesurament de la glucosa, per tant quan més precís i freqüent siga aquest, millor serà el control.

Segons les últimes dades de l'IDF s'estima el nombre de persones diabètiques en 382 milions, un 8,3% dels adults del món. Un 80% de les afectades viuen en països d'ingressos mitjans o baixos. Es preveu un augment mundial del 55% per al 2035, fins els 592 milions de persones (Figura 3 a la pàgina següent). A l'Estat espanyol hi ha 3,8 milions d'adults afectats (10,8%) [1]. Si bé les xifres actuals són alarmants, més encara ho és la perspectiva de creixement en la prevalença de la DM. Els costos de l'atenció sanitària d'una persona amb DM dupliquen els costos d'una persona sense la malaltia (més d'un 10% de la despesa sanitària a Europa).

### 2.1.0.1 Tractament per a la DM-1

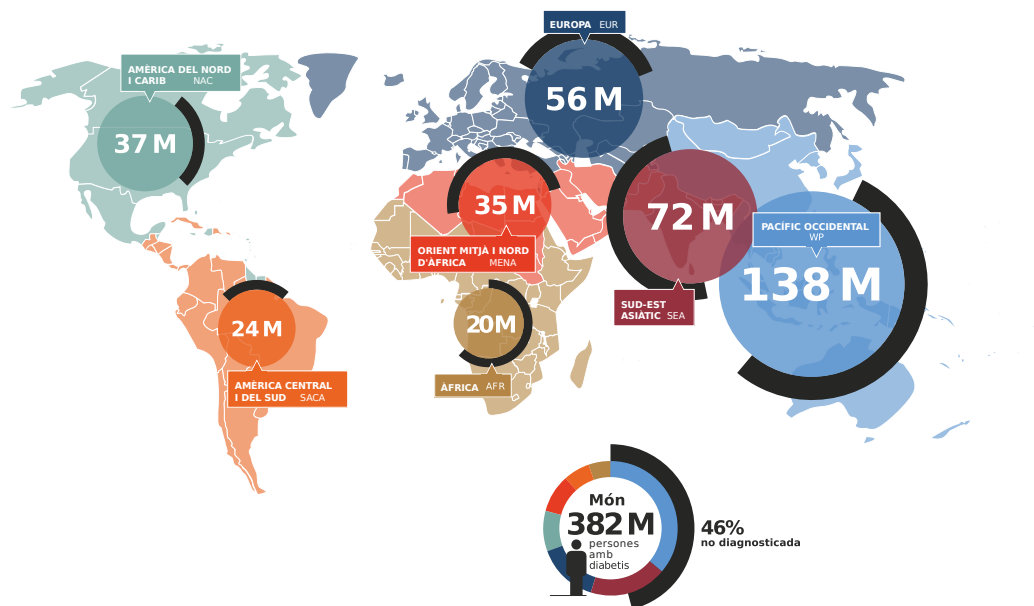
El tractament consisteix en compensar els baixos nivells d'insulina endògena administrant infusions d'insulina exògena (*insulinoteràpia*). La combinació d'insulina basal<sup>2</sup> i bolus<sup>3</sup> emula el comportament del pàncrees. Per calcular la dosi necessària del bolus, cal mesurar el

2 Controla el nivell de glucosa entre les menjades i durant la nit. Generalment es dona una o dues vegades al dia.

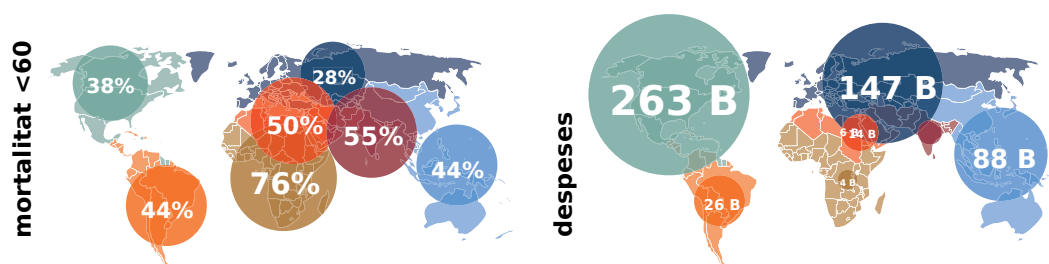
3 Subministrada abans de les menjades per tal de controlar l'augment de glucosa després del menjar. Són d'acció ràpida.



**Regions de l'IDF i projeccions del nombre de persones amb diabetis (20-79 anys), 2013 i 2035**



REGIÓ DE L'IDF	2013 MILIONS	2035 MILIONS	AUGMENT %
● Àfrica	19.8	41.4	<b>109%</b>
● Orient Mitjà i Nord d'Àfrica	34.6	67.9	<b>96%</b>
● Sud-Est Asiàtic	72.1	123	<b>71%</b>
● Amèrica Central i del Sud	24.1	38.5	<b>60%</b>
● Pacífic Occidental	138.2	201.8	<b>46%</b>
● Amèrica del Nord i Carib	36.7	50.4	<b>37%</b>
● Europa	56.3	68.9	<b>22%</b>
<b>Món</b>	<b>381.8</b>	<b>591.9</b>	<b>55%</b>



Proporció de morts atribuïdes a la diabetis en gent menor de 60 anys, 2013

Despeses sanitàries (USD) per diabetis (20-79 anys), 2013

Figura 3: Nombre de persones amb diabetis al món. Imatge d'IDF [1].

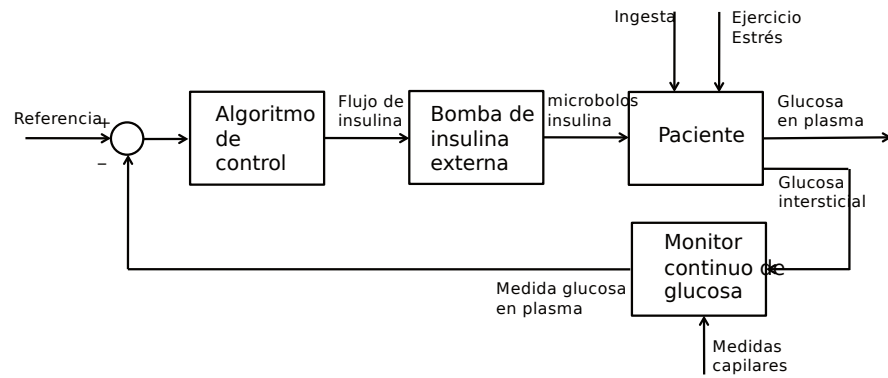


Figura 4: Bucle de control de la glucosa usat amb la monitorització contínua de la glucosa i infusió d'insulina subcutània. Imatge de Bondía et al. [3].

nivell de glucosa preprandial i tenir en compte la quantitat de CHO a ingerir. Hi ha uns quants factors que determinen la freqüència i quantitat de les dosis: pes del pacient, activitat física, sensibilitat a la insulina, etc. El subministrament es fa per via subcutània (unes quantes injeccions al dia), o mitjançant una bomba d'infusió contínua. En qualsevol cas, l'efecte més perjudicial de la insulinoteràpia és l'augment d'episodis hipoglucèmics.

#### 2.1.0.2 Estat de la investigació actual

Actualment hi ha dos enfocaments diferents que intenten millorar el tractament de la DM-1: un biològic –ús de cèl·lules mare per a substituir a les *beta*– i un tecnològic.

La solució tecnològica més completa consisteix en la creació de l'AP, que simula la funció endocrina de l'òrgan. L'AP està constituït per un CGMS acoplat a un dispositiu de subministrament d'insulina i un algorisme que proporciona la dosi correcta d'insulina en el moment adequat sense intervenció humana. D'entre les diferents combinacions possibles per desenvolupar l'AP, la més prometedora és la que usa monitorització contínua de glucosa subcutània (s.c) i infusió d'insulina subcutània (s.c) [3], de caràcter «mínimament invasiu» (Figura 4).

Alguns dels problemes o reptes oberts per al desenvolupament de l'AP són:

- La infusió subcutània d'insulina té un temps d'absorció més lent –a diferència de la fisiològica– que produeix una sobre-actuació del controlador, provocant una hipoglucèmia degut a l'acumulació d'insulina.
- Hi ha molts errors en l'estimació per a compensar les menjades (sobre-actuació, hipoglucèmia).
- És necessària més robustesa entre pacients, ja que presenten gran variabilitat de resposta.

- Fer front a les grans perturbacions: composició dels menjars, estrès i exercici físic.
- Integrar estratègies que garanteixen la seguretat en les bombes d'infusió d'insulina, així com la detecció d'errades en els dispositius.

### 2.1.0.3 *Monitorització Continua de la Glucosa*

A principis de la dècada de 1980 apareixen els primers glucòmetres d'ús domèstic (SMBG). Aquests mesuren el nivell glucèmic mitjançant l'extracció d'una xicoteta mostra de sang per punció capilar, arterial o venosa, que es col·loca en una tira reactiva per analitzar-se. El resultat és molt precís i directe, però el mètode és invasiu i dolorós. Sol realitzar-se quatre vegades al dia, una abans de cada menjada i una abans d'anar a dormir. Els SMBG són fonamentals perquè els malalts puguen mantenir el control glucèmic, però la seua funcionalitat no incorpora un registre històric i per tant, no tenen capacitat predictiva.

Els CGMS van un pas més enllà: funcionen amb un sensor que s'instal·la sobre la pell i mesura la concentració de glucosa freqüentment (entre 1 i 5 minuts). El sensor electroquímic s'insereix via subcutània amb una agulla, i queda ubicat en la capa de grassa que hi ha sota la pell, on habitualment no molesta. Degut a això, les lectures de BG que proporcionen són estimacions a partir de medicions en el fluid intersticial<sup>4</sup>. Tenen una unitat de visualització de dades, que proporciona informació de la variabilitat de la glucosa durant el dia. Així doncs, també tenen el potencial per ser predictius, però actualment ningun té una precisió analítica i fiabilitat comparables als SMBG—especialment en el rang hipoglucèmic.

Segons la revisió feta per Leal [13], entre les tècniques plantejades per predir esdeveniments hipo/hiperglucèmics hi ha:

- Mètodes de projecció. Generen alertes quan la tendència actual suggereix la probabilitat d'un esdeveniment hipoglucèmic a curt termini.
- Models d'entrada–sortida. Generen alertes en base a prediccions calculades a partir de concentracions prèvies, utilitzant models de sèries temporals.

Aquests enfocaments usen models amb estructura fixa i mínima complexitat, però amb paràmetres variables en el temps, que en cada instant es reajusten en base a la nova mostra. Més recentment s'han proposat:

<sup>4</sup> El fluid intersticial o fluid tissular es el líquid contingut a l'interstici, o espai entre las cèl·lules.

- Tècniques d'estimacions òptimes. Prediccions basades en el filtre de Kalman i un model no paramètric estocàstic considerant tres diferents horitzons de predicció.
- Ús de models de xarxes neuronals artificials.

#### 2.1.0.4 Limitacions

La principal dificultat per a produir estimacions fiables és la gran variabilitat de resposta metabòlica inter-pacients i intra-pacients. Aquesta variabilitat fa necessària la recollida d'una gran quantitat de dades, obtingudes de pacients amb característiques heterogènies, per tal de poder dissenyar algorismes fiables. La recollida de dades és difícil perquè els assaigs clínics són costosos, tant en temps com en diners, i habitualment es donen conflictes ètics pel treball amb éssers humans. A més, la disponibilitat de dades públiques és escassa i limitada.

## 2.2 RAONAMENT QUALITATIU

Segons Forbus [9], el modelatge qualitatiu té a veure amb la representació i el raonament simbòlic –similar a l'humà– sobre aspectes continus d'entitats i sistemes. Raonar directament amb models qualitius pot capturar patrons de comportament importants, produint automàticament descripcions que estan més aprop d'allò que la gent en diu «comprensió» del comportament d'un sistema.

Hi ha tres principis fonamentals que regeixen el modelatge qualitatiu:

1. Discretització: Les propietats contínues d'un sistema s'han de representar de forma limitada. Així, obtenim unes entitats o símbols amb els quals podem raonar de manera abstracta.
2. Rellevància: La discretització elegida per a una representació qualitativa està limitada tant per la naturalesa del sistema, com pel raonament que s'ha de fer amb ell. Es a dir, depenent de la tasca que s'haja d'acomplir, la descripció qualitativa d'un mateix sistema pot variar.
3. Ambigüitat: L'abstracció amb la què es treballa té com a contrapartida l'ambigüitat en les respostes. Per això, els models qualitius són un complement ideal dels models matemàtics. El model qualitatiu pot servir per a raonar a alt nivell i emmarcar el problema, per aplicar posteriorment un model matemàtic concret.

A continuació veurem una panoràmica del procés de raonament qualitatiu.

#### 2.2.0.5 Formulació del model

El primer pas consisteix en fer un model del sistema o situació. La descripció inicial s'anomena *escenari*. Els tipus d'entitats i fenòmens que es poden donar es representen com a *fragments del model*, normalment desats a una biblioteca anomenada *teoria del domini*. A partir de fragments rellevants es munta un model per a l'escenari mitjançant un procés anomenat *formulació del model*. S'utilitzen tant els continguts de l'escenari com les restriccions imposades per la tasca per a la qual s'està construint el model.

#### 2.2.0.6 Elaboració d'un estat qualitatiu

El model consisteix en un conjunt de fragments, que representen propietats i relacions que poden ser veritat o no en cada moment. Els fragments que hi ha actius, defineixen el conjunt d'entitats i els seus paràmetres. Habitualment, els fragments actius només imposen limitacions parcials sobre els valors dels paràmetres. Es defineix un *estat qualitatiu* amb el conjunt de fragments actius i els valors qualitius per a tots els paràmetres del sistema.

#### 2.2.0.7 Simulació qualitativa

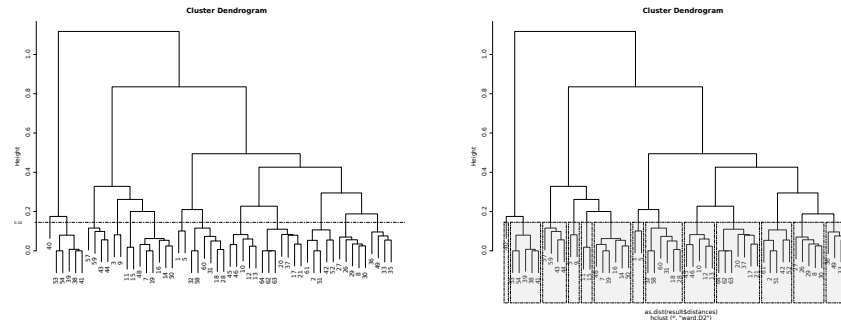
La simulació qualitativa identifica quins estats poden succeir a l'actual. Si s'aplica recursivament es poden obtenir tots els estats que poden aparèixer des d'un estat inicial concret. Per a models simples això es pot traduir en complexitat polinòmica, però és exponencial respecte al nombre de constituents de l'estat. Alguns estats poden durar per sempre, i el nombre d'estats qualitius també pot ser infinit, per això cal tenir molta cura a l'hora de dissenyar el sistema de raonament.

### 2.3 CLÚSTERING I MESURES DE QUALITAT

El clústering és una tècnica multidisciplinària que assigna objectes en grups similars, de manera que es maximitze la similitud intra-clúster alhora que es minimitze la similitud inter-clúster. És una de les aproximacions més usades avui en dia per a trobar patrons desconeguts o inesperats en grans quantitats de dades espacials-temporals. L'anàlisi dels clústers dona respostes a «quins objectes?», «on?» i «quan?», i ens suggereix investigar el «per què?» [16].

Hi ha una gran varietat de algorismes de clústering basats en diferents models. Destaquem les categories que ens resulten més rellevants de la classificació oferida per Wikipedia [18]:

- Models basats en la connectivitat (p. ex. el clústering jeràrquic [15]).
- Models basats en particions (p. ex. *k*-means [14]).



- (a) El dendrograma conté la informació de totes les particions possibles. Depenent de la distància per on es talla, s'obté un nombre d'agrupacions.
- (b) A l'exemple de dalt, tallant a una distància de 0.14 obtenim 13 clústers.

Figura 5: Exemple d'un dendrograma.

- Models basats en la densitat (p. ex. DBSCAN [7]).

En aquest treball usarem principalment el primer model perquè té una característica que ens resultarà molt avantatjosa: és més fàcil usar qualsevol mesura de distància (només necessita una matriu de distàncies). Després veurem per què això ens ajuda. Els desavantatges més destacats són la poca tolerància als *outliers*, i la complexitat  $\mathcal{O}(n^3)$ . En general aquests models estan considerats obsolets, però per a iniciar l'exploració d'unes poques dades –com en el nostre cas– és suficient.

### 2.3.1 Clústering jeràrquic

Els models basats en la connectivitat construeixen una jerarquia de clústers usant una d'aquestes dues estratègies:

- Agregativa: cada observació comença en el seu propi clúster, i es van fusionant parelles de clústers a mesura que es puja en la jerarquia.
- Divisiva: totes les observacions comencen en un clúster, i s'apliquen divisions binàries recursives a mesura que baixem en la jerarquia.

La jerarquia queda representada en un dendrograma. L'eix de les  $y$  estableix les distàncies on els clústers es fusionen, mentre les observacions se situen a l'eix de les  $x$ . A diferents distàncies es formen diferents clústers. Per tant el dendrograma conté totes les possibles particions del conjunt de dades. A la Figura 5 es pot veure un exemple.

El que realment diferencia entre els diferents models basats en connectivitat és el *criteri d'enllaç*, que especifica quins clústers s'agre-

guen/divideixen en cada nivell. Entre les opcions més populars s'hi troben [15]:

- *single-linkage*: la distància mínima entre els objectes.
- *complete-linkage*: la distància màxima entre els objectes.
- *average-linkage* o *UPGMA*: la distància mitjana entre els objectes.
- *Ward*: el decreixement en la variància per al clúster que es fusiona.

### 2.3.2 Mesures de qualitat

Degut a la gran quantitat de variabilitat que pot resultar del procés de selecció de característiques rellevants, normalització, selecció d'una funció de dissimilitud, etc. avaluar els resultats mitjançant «inspecció visual» pot ésser altament subjectiu i perillós. Podem acabar reforçant les nostres assumpcions prèvies i ignorant resultats inesperats –el contrari del que es vol amb un procés de classificació no-supervisat. Per això és necessari incorporar un procés de validació més objectiu.

Hi ha dos grans grups de tècniques de validació: externes i internes. Les externes comprenen totes aquelles que avaluen els resultats d'un clústering basant-se en conèixer previament les classes correctes. Com és normal, això resulta útil per dissenyar i comparar algorismes de clústering sobre dades etiquetades, que sabem a quina classe pertanyen. Si desconexem les classes o aquestes són dubtoses, hem d'usar mesures de validació interna, que basen la qualitat únicament en informació intrínseca a les dades. En concret, intenten mesurar com de bé una partició concreta correspon amb l'estructura «natural» de les dades. Handl et al. [10] resumeix els següents tipus de mesures de qualitat internes que ens resulten rellevants:

1. Homogeneïtat: mesuren la compactesa o homogeneïtat del clúster; el representant més popular és la variància intra-clúster, però hi ha moltes variants.
2. Connectivitat: mesuren el grau de densitat local i l'agrupament dels elements amb els seus veïns més pròxims en l'espai de les dades.
3. Separació: mesuren el grau de separació entre clústers individuals.
4. Combinacions: Molts mètodes combinen la mesures dels tipus anteriors. En especial és popular la combinació dels tipus 1 i 3, perquè mostren comportaments oposats: mentre la homogeneïtat intra-clúster millora amb el nombre de clústers, la distància entre clústers es deteriora. En són exemples ben coneguts l'Índex de Dunn (DI) [6] i la Silueta [17].

5. Estabilitat: Es basen en modificar el conjunt de dades original i recalculer el clústering. La consistència dels resultats proporciona una estimació de la importància dels clústers obtinguts. Requereixen tenir accés a l'algorisme de clústering.

Les mesures de qualitat internes es poden usar per seleccionar el nombre adequat de clústers d'un conjunt de dades. Això requereix recalculer el clústering per a un rang determinat de clústers. Si l'algorisme de clústering i la mesura de validació interna són adequats per al conjunt de dades concret, el millor nombre de clústers destacarà en la corba de mesura.

El clústering és un procés complicat i interactiu, i no hi ha receptes per triar el mètode i els paràmetres adequats per a les nostres dades. El mateix es pot aplicar a la validació. Per ajudar, el paquet d'R `clValid` [4] ens permet aplicar diferents mètodes de clústering amb diferent nombre de clústers i obtenir mesures de qualitat «internes» i d'«estabilitat» de manera sistemàtica.



Part II

DESENVOLUPAMENT



## DISSENY I DESENVOLUPAMENT

---

Per efectuar l'anàlisi de les corbes de glucosa, procedirem en tres passos:

1. Pre-processament: Selecció de característiques rellevants i discretització.
2. Clústering: Selecció d'una funció de dissimilitud, configuració i aplicació d'un o diversos mètodes de clústering.
3. Validació: Avaluació de les particions resultants. La qualitat de les particions s'usa per a revisar les eleccions fetes en els passos anteriors.

### 3.1 PRE-PROCESSAMENT

L'objectiu de fer una descripció qualitativa de les corbes de la glucosa postprandial, és fer grups de comportament similar per observar quins són els factors comuns en els menjars que provoquen aquests comportaments.

#### 3.1.1 *L'escenari inicial*

Tant per a la construcció del pàncrees artificial, com per a la gestió de la diabetis i altres estats patològics relacionats, comprendre el **ritme d'aparició de glucosa** ( $Ra$ ) després de la ingestió d'un menjar mesclat podria ser de gran ajuda. El criteri de referència per estimar el  $Ra$  és l'ús d'un protocol de glucosa oral amb multi-traçador, però aquesta tècnica és complexa i poc convenient per a estudis grans [11]. Per això moltes de les dades a estudiar provenen d'estimacions, que poden ser basades en models mínims de glucosa-insulina o altres mètodes. La unitat per al  $Ra$  sol ser mg/kg/min, i mesura l'aparició de glucosa respecte d'una mesura basal prèvia (p. ex. 110 mg/dL). A la figura 6 a la pàgina següent es pot veure la relació de la glucosa i la insulina en plasma aparegudes després d'un menjar.

Partim d'una base de dades de 60 menjars mesclats, recollits de la literatura existent (1994–2008). De cadascun coneixem: els ingredients, la composició –en macronutrients: CHO, grasses, proteïnes i fibra–, l'energia (kcal) i finalment, la resposta  $Ra$  estimada que provoca en pacients amb DM-2. El  $Ra$  per a cada menjar és un vector de 84 valors reals que corresponen a 84 «observacions» preses cada 5 minuts. Es a dir, que per a cada menjar tenim el  $Ra$  dels 420 minuts

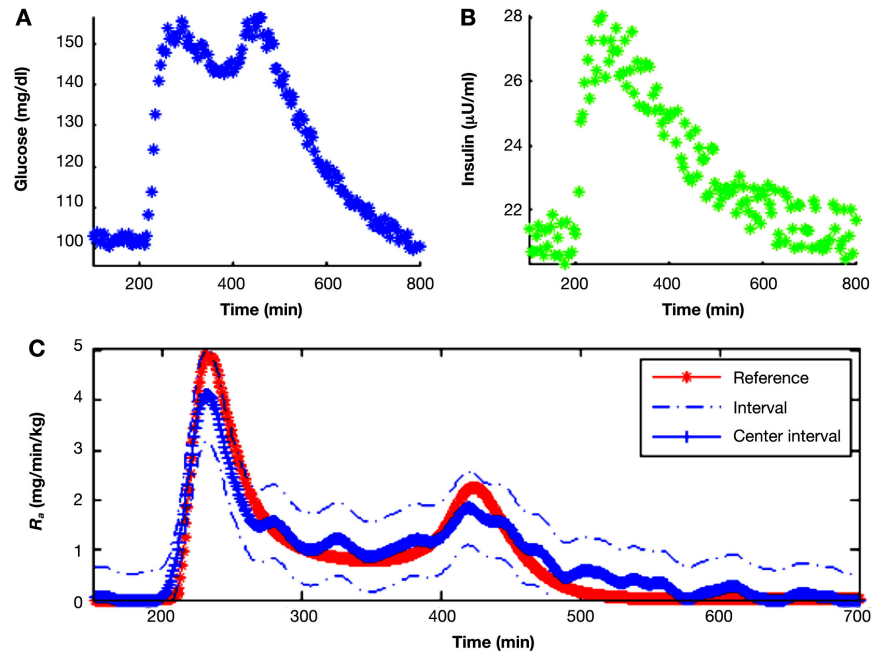


Figura 6: Estimació (blau) i referència (roig) de l'interval  $R_a$  (C), glucosa en plasma (A), i insulina en plasma (B). Imatge d' Herrero et al. [11].

següents a la seua ingestió –similar a la línia roja del gràfic C de la Figura 6.

A més d'aquestes dades, també comptem amb els següents coneixements que ens han transmès els experts:

- Els bolus d'insulina, encara que aquesta siga del tipus més ràpid, tarden més en fer efecte que la insulina endògena. Això vol dir que per a evitar hiperglucèmies, és necessari predir-les almenys amb mitja hora d'antelació. Si això no es fa bé, no sols no la evitem sinó que segurament provocarem una hipoglucèmia.
- Els pics de magnitud en el  $R_a$  són molt importants perquè representen els estats potencialment més perillosos per al pacient i són el principal esdeveniment que volem evitar.
- Un cop s'ha superat el 80% de l'àrea sota la corba (AUC) del  $R_a$ , ja no hi ha perill.

### 3.1.2 Formulació d'un model

Partim d'una variable explícita, la magnitud del  $R_a$ , i d'una implícita, el temps. A grans trets, el més important és la magnitud que assoleix el  $R_a$  i el temps que dura. Per suposat, la variabilitat amb que ho fa també compta. Com que volem obtenir detalls del comportament de les corbes, proporcionar únicament aquests tres paràmetres per al

total de la corba, representaria un nivell d'abstracció excessiu. Serà necessari descriure aquests paràmetres dintre de *subdivisions*.

D'altra banda, tenim registrades 84 observacions del *Ra* cada 5 minuts, però això no vol dir que totes les corbes duren 420 min. Quan a partir d'una observació  $o_n$  amb valor proper a zero, totes les següents  $\{o_{n+1}, \dots, o_{84}\}$  són iguals o menors, aleshores podem considerar que  $o_n$  és el final d'una corba de *duració total*  $= n \times 5$  min. Encara podem acotar més la duració de la corba, perquè l'escenari especifica que a partir del 80% d'*AUC* el *Ra* és negligible. Així podem dir que una corba té una *duració efectiva*  $= i \times 5$  min, on  $o_i$  és la *primera* observació que compleix la propietat<sup>1</sup>:

$$\left\{ \exists i \mid \frac{o_i + \sum_{j=1}^{i-1} o_j}{\sum_{j=1}^{84} o_j} \geq 0.8 \right\}$$

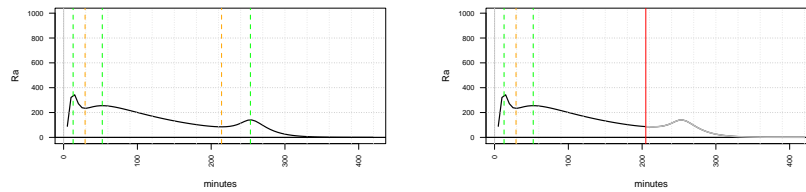
Els valors del *Ra* són sempre positius  $[0, +\infty)$ . En cas de donar-se el valor zero, simplement voldria dir que no apareix glucosa per damunt del nivell basal de referència, cosa que no té perquè implicar un estat hipoglucèmic. Així mateix, si bé una baixada paulatina en el *Ra* indica una situació positiva –s'està fent ús de la glucosa i la insulina va degradant-se–, una *baixada pronunciada* en el *Ra* podria indicar risc d'hipoglucèmies sense que puguem saber-ho. Per consegüent, amb les dades de partida de què disposem, tant els augments com les disminucions en la variable *Ra* són rellevants. Això ho capturem en un paràmetre que anomenem *C*, simbolitzant el *canvi* en *Ra*.

Fins ara hem parlat d'un paràmetre explícit *magnitud* de *Ra* (diguem-li *M*) i un altre paràmetre *canvi* en *Ra* (*C*) que s'hauran de definir per a cada *subdivisió* o període temporal. Passem a veure quants d'aquests períodes hi ha dintre de la corba. Tenim dos opcions per a elegir-ne la quantitat: (1) dividim la corba en períodes de duració *t* constant de tal manera que resulten  $n = \frac{i \times 5}{t}$  períodes, o (2) dividim la corba en períodes variables. La primera opció té l'avantatge de resultar més còmoda a l'hora de fer les divisions i comparar diferents corbes, tanmateix presenta uns inconvenients severs:

- Com fem per seleccionar una durada *t* de manera que els períodes siguin adequats en totes les corbes? Per exemple, per a corbes xicotetes –duració efectiva de 120 min– una granularitat de  $t = 20$  min és adequada, però continua sent-ho per a corbes de 420 min?
- Una de les raons a favor de la divisió en períodes temporals constants, és la facilitat per a comparar entre corbes ja que totes tindrien el mateix nombre de períodes. Però aleshores com podem parametritzar el fet que cada corba tinga una durada efectiva diferent? Hauríem d'afegir un nou atribut que indiqués si la corba ha finalitzat o no.

<sup>1</sup> Amb  $o_i$  s'assoleix el 80% d'*AUC*.

*Una hipoglucèmia no es veuria en un gràfic que només representa el ritme d'aparició de glucosa.*



- (a) Exemple de com es fan les subdivisions de la corba en períodes usant els pics (verd) i les valls (taronja) relatives. El menjar mostrat (n. 9) té 3 períodes de creixement i 3 de decreixement.
- (b) El mateix menjar queda reduït a 4 períodes perquè descartem la informació més enllà del 80% d'AUC.

Figura 7: Exemples de com es talla una corba en períodes.

- El criteri per a decidir els límits d'un període no està associat amb els valors que pren  $Ra$  en aqueix període, i per tant tenim una altíssima probabilitat de trobar-nos amb valors  $Ra$  que creixen o decreixen dintre d'un mateix període. Evidentment això dificulta la selecció d'un valor qualitatiu que caracteritzi el període.

D'altra banda, amb la segona opció –temps variable– obtindríem una quantitat diferent de períodes per a cada corba, dificultant-ne la comparativa. Tanmateix, aquest obstacle es pot salvar, i hi ha diversos avantatges:

- El límit dels períodes s'origina a partir del  $Ra$ , capturant blocs de comportament similar que seran més fàcils de descriure qualitativament.
- No és necessari parametritzar explícitament la duració efectiva de la corba. Hi haurà  $n$  períodes de duració variable  $t$  només fins arribar al 80% d'AUC, i la resta d'informació queda descartada.

Considerem que els arguments per dividir les corbes en períodes de temps variables són consistents amb els principis del modelatge qualitatiu, i que l'única raó per no fer-ho així seria la dificultat tècnica que pot presentar.

Com a resultat de l'argumentat anteriorment, cada corba de  $Ra$  de glucosa postprandial associada amb un menjar quedarà representada per  $n$  períodes que comprenen la informació rellevant de la mateixa. Cadascun d'aquests períodes contindrà tres paràmetres: *magnitud*  $Ra$  ( $M$ ), *canvi en*  $Ra$  ( $C$ ) i durada temporal ( $D$ ).

La selecció dels períodes es farà fixant-nos en en els canvis de sentit del  $Ra$ , de manera que la frontera entre períodes sempre serà un màxim o un mínim relatiu en la variable  $Ra$  (Figura 7). En la següent sec-

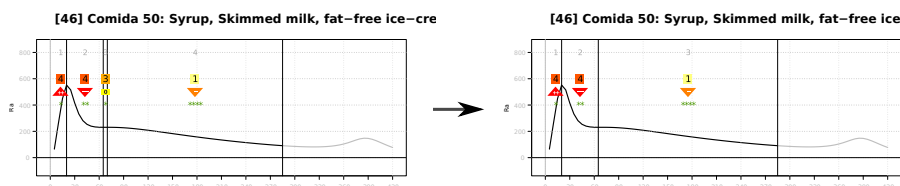


Figura 8: Mecanisme d'absorció de períodes.

ció parlarem de quins valors podran tenir aquests paràmetres, però abans cal comentar un detall més.

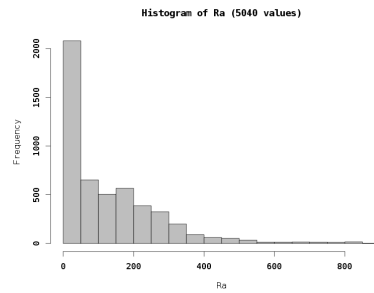
**ERROR EN LA SUBDIVISIÓ** Degut a algun error del programari a l'hora d'aproximar la funció de la corba i la seua derivada, en algunes ocasions la subdivisió en períodes no donava el resultat esperat. El problema al que ens referim es pot observar a la Figura 8 en la corba de l'esquerra, on es veu que s'han detectat de més un parell pic/vall posterior al segon període. Degut a això, durant la implementació hem hagut de incorporar un «mecanisme d'absorció de períodes». Aquest mecanisme consisteix en que els períodes de creixement zero (assignats a l'etiqueta central en els intervals de C) poden quedar absorbits per un període anterior o posterior que tinga un valor conceptualment veí (el veïnatge conceptual s'explica a la Secció 3.2.1.1).

#### 3.1.2.1 Selecció d'etiquetes per a la magnitud Ra

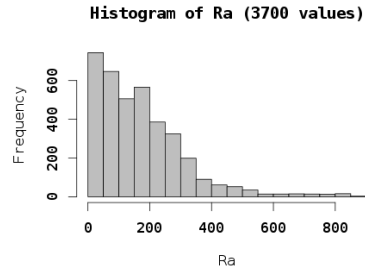
Sabem que l'objectiu dels pacients amb DM és controlar i mantenir el nivell de glucosa en un rang 70–180 mg/dL, tanmateix el Ra no indica exactament això, i per tant no ens pot servir com a referència per a establir rangs de nivell. Lamentablement tampoc tenim cap altra informació més enllà de les pròpies dades.

Els valors de la mostra estan en el rang  $[0, 861]$  i presenten un histograma com el de la Figura 9a a la pàgina següent. Degut a que moltes corbes no duren els 420 min complets, hi ha un alt nombre de valors propers a zero. Si només tenim en compte la *duració total* –com hem descrit en l'apartat anterior– desapareixen tots els valors propers a zero, i passem de 5040 a 3700 observacions (Figura 9b). Si a més, només tenim en compte la *duració efectiva* de les corbes, passem a tenir 2094 observacions en el rang  $[2, 861]$  (Figura 9c). Ara sí, veiem que la major freqüència de valors ha passat a estar prop del valor 200. En última instància no volem etiquetar les observacions directament sinó períodes d'observacions, caracteritzats per algun estadístic del valor –generalment la mitjana aritmètica. En altres paraules, volem discretitzar el període.

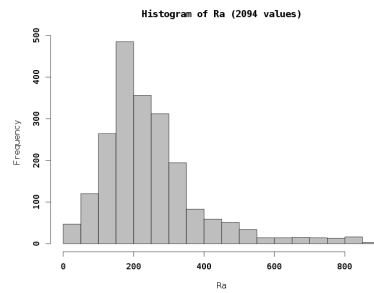
A partir dels valors mitjans dels períodes, en el rang  $[72.6, 748.7]$ , hem de decidir quants intervals farem i quines etiquetes els assignarem. Per exemple, per fer-ho ràpid, inspirats per l'histograma de la Figura 9d a la pàgina següent, podríem decidir fer 4 intervals amb els



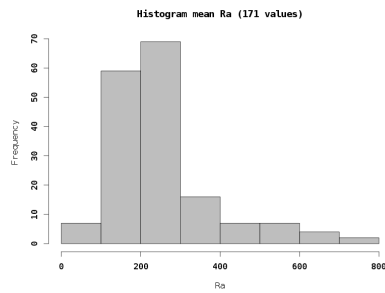
(a) Histograma (Sturges) de tots els valors de Ra.



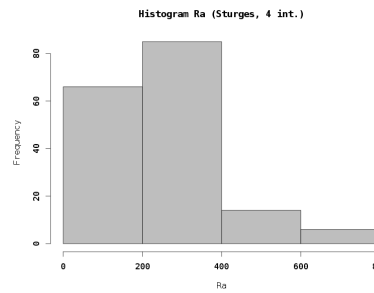
(b) Histograma (Sturges) dels valors de Ra, sostraint els valors que no estan dintre de la duració total.



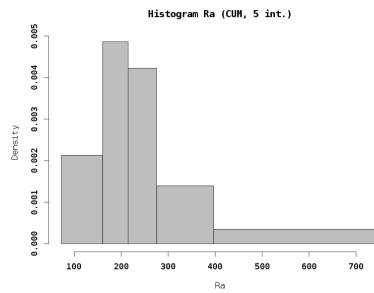
(c) Histograma (Sturges) dels valors de Ra, sostraint els valors que no estan dintre de la duració efectiva.



(d) Histograma (Sturges) dels valors mitjans de Ra en cada període; s'han obtingut 8 intervals.



(e) Histograma (Sturges) amb 4 intervals dels valors mitjans de Ra.



(f) Histograma (CUM) amb 5 intervals dels valors mitjans de Ra.

Figure 9: Histogrames dels valors de Ra.



límits [0, 200, 400, 600, 800] i les etiquetes de magnitud corresponents [«baixa», «mitjana», «alta», «gegant»]. Un aspecte negatiu d'establir els intervals de manera equitativa és que al tenir una major densitat de freqüència en la zona [100, 300], el comportament de la variable en aqueixa zona queda «difuminat». Com es veu a la Figura 9e a la pàgina anterior, el 88% de la mostra cau dintre de les magnituds «baixa» i «mitjana».

*Aquesta és la notació usada en R, on els intervals són l'espai que queda entre els valors del vector.*

Si elegim una menor amplitud d'intervals en les zones amb més freqüència, obtindrem més detall en el comportament de la variable. Existeix un mètode alternatiu per agrupar la variable en base a la minimització de la mesura de les desviacions típiques dintre de cada interval anomenat CUM, formalitzat per Abril i Ruíz [2]. Amb el mètode CUM aplicat per a 5 intervals obtenim els límits: [72.6, 160.5, 214.6, 275.4, 397.1, 748.7] (vegeu Figura 9f), de manera que tots els intervals obtenen entre 12% i 25% de la mostra. La contrapartida d'aquest mètode és que «atenua» el grup de pics més elevats, ficant en el mateix interval els valors entre [397.1, 748.7]. Per ara no tenim manera de saber si això és acceptable mèdicament, així que tindrem en compte els dos enfocaments a l'hora de realitzar els experiments. Observant els resultats del mètode CUM per a 3 intervals ([72.6, 194.3, 295.7, 748.7]) i per a 4 ([72.6, 174, 248.4, 343, 748.7]), ens decantem per elegir 5 intervals perquè amb els altres ens sembla que es produeix una excessiva «atenuació» dels valors elevats.

En resum, a l'hora de fer l'experimentació, provarem a fer 3, 4 i 5 intervals d' $M$  amb el mètode CUM. Finalment, a mode d'exemple, les etiquetes per al paràmetre  $M$  amb 5 intervals queden establerts així:

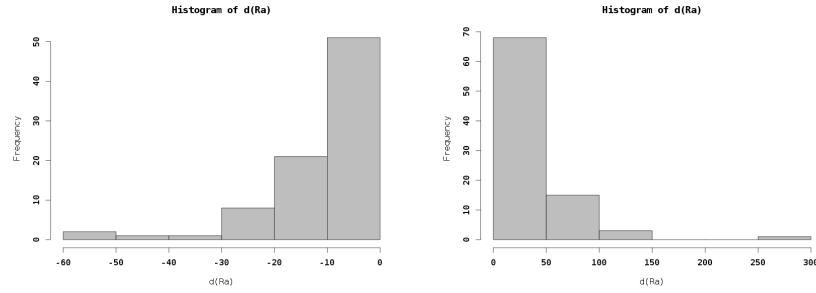
$$C_{INT} = \{[0, 160.5], (160.5, 214.6], (214.6, 275.4], (275.4, 397.1], (397.1, 748.7]\}$$

$$C_{LAB} = \{baixa(1), mitjanaInferior(2), mitjanaSuperior(3), alta(4), gegant(5)\}$$

Cada etiqueta del vector  $C_{LAB}$  està en cursiva i va seguida entre parèntesis del símbol que la representa. Ho farem així per a totes les etiquetes que definim.

### 3.1.2.2 Selecció d'etiquetes per al canvi en $Ra$

El canvi ( $\Delta Ra$ ) pot ser positiu o negatiu,  $Ra$  creix o decreix. Només pren el valor zero en un moment instantani que usem per a dividir la corba en períodes o finestres, per tant aquests moments ja queden caracteritzats implícitament en el model. Tanmateix, tenir únicament dues etiquetes per definir si un període és d'un tipus «+» o d'un altre «-», encara que resulta força intuïtiu, pot ser no siga suficient perquè hi ha molta diferència semàntica entre un creixement moderat (amb una pendent menor a 1,5) o un creixement pronunciat (p. ex. amb



- (a) Histograma (Sturges) de tots els períodes on la mitjana de la derivada primera de Ra té un valor negatiu.
- (b) Histograma (Sturges) de tots els períodes on la mitjana de la derivada primera de Ra té un valor positiu.

Figure 10: Histogrames dels valors mitjans de la derivada primera de Ra per a tots els períodes.

pendent de més de 10). El primer representa un augment de magnitud dilatat que permet més temps de reacció (p. ex. una insulina d'acció més lenta), mentre el segon és una situació extremadament rellevant que requereix una actuació immediata (idealment s'hauria de preveure amb antelació). Per això és necessari establir més etiquetes.

Al igual que en el cas de la magnitud, la freqüència dels valors de canvi s'acumula en un rang concret:  $[-10, 50]$  (vegeu les Figures 10a i 10b). De nou tenim la possibilitat d'eleger el tall d'interval: a mà, equitativament o amb CUM. L'aspecte de les dades és el següent:

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
-58.440	-7.582	1.336	10.150	16.090	289.700

A diferència de amb la magnitud, en aquest cas no tenim cap inconvenient en «atenuar» els valors extrems precisament perquè són exagerats (una pendent de 50 i una de 200 són igualment grandíssimes). A més, la derivada és el resultat d'una aproximació i aquests valors extrems poden ser deguts a un error en la mateixa. A priori no tenim ninguna raó per preferir un mètode de tall o un altre.

**ALTERNATIVA AMB PICS/VALLS** Si considerem que els pics i les valls de magnitud són els punts més rellevants de les corbes, aleshores podria ser adequat eleger els intervals de la variable  $C$  tenint en compte els valors de la derivada primera al voltant d'ells. Com és evident, hauríem de seleccionar un nombre suficient d'observacions prèvies i posteriors al pic/vall ja que en el propi pic/vall el valor de canvi és zero. Considerant les 3 observacions prèvies i posteriors, tindrem en compte una finestra temporal de 30 min al voltant del pic (a no ser que excedim la duració de la corba o es produeixen solapaments amb un altre pic/vall). En aquest cas no usem cap estadístic, recollim directament els 6 valors  $\Delta Ra$  al voltant del pic/vall. Després

usem tots els valors recollits per seleccionar els intervals, per exemple podríem dir que el 75% dels valors més elevats (a partir del primer quartil) defineixen els intervals de C de més creixement.

Caldrà esperar a la implementació, observar els resultats i mesurar-ne la qualitat. En resum provarem les següents opcions:

1. Pics/valls (pks) – Partint de tots els valors recollits com s’ha descrit en el paràgraf anterior, diguem-ne  $A$ , obtenim dos subconjunts  $N$  i  $P$  tal que:

$$\forall x\{x \in N \rightarrow x \in A \wedge x < 0\}$$

$$\forall x\{x \in P \rightarrow x \in A \wedge x > 0\}$$

Aleshores definim 5 intervals així:

$$C_{INT} = \{[\min(N), Q_1(N)], (Q_1(N), Q_3(N)], \\ (Q_3(N), Q_1(P)], (Q_1(P), Q_3(P)], (Q_3(P), \max(P))\}$$

2. CUM (CUM) – Sobre tots els valors  $mean(\Delta Ra)$  de tot els períodes, usem el mètode CUM per a fer de 2 a 7 intervals.

3. Manual (man) – Selecció manual amb 2 a 7 intervals definits així:

$$C_{INT} = \{[\min(\Delta Ra), 0], (0, \max(\Delta Ra))\}$$

$$C_{LAB} = \{decreix(-), creix(+)\}$$

$$C_{INT} = \{[\min(\Delta Ra), -2], (-2, 2], (2, \max(\Delta Ra))\}$$

$$C_{LAB} = \{decreix(-), estable(=), creix(+)\}$$

$$C_{INT} = \{[\min(\Delta Ra), -10], (-10, 0], (0, 10], (10, \max(\Delta Ra))\}$$

$$C_{LAB} = \{decreixExtrem(--), decreix(-), creix(+), creixExtrem(++)\}$$

$$C_{INT} = \{[\min(\Delta Ra), -15], (-15, -1], (-1, 1], (1, 15], (15, \max(\Delta Ra))\}$$

$$C_{LAB} = \{decreixExtrem(--), decreix(-), estable(=), creix(+), \\ creixExtrem(++)\}$$

$$C_{INT} = \{[\min(\Delta Ra), -15], (-15, -1], (-1, 0], (0, 1], (1, 15], \\ (15, \max(\Delta Ra))\}$$

$$C_{LAB} = \{decreixExtrem(---), decreixFort(--), \\ decreix(-), creix(+), creixFort(++), creixExtrem(+++)\}$$

$$C_{INT} = \{[\min(\Delta Ra), -20], (-20, -15], (-15, -1], (-1, 1], (1, 15], \\ (15, 20], (20, \max(\Delta Ra))\}$$

$$C_{LAB} = \{decreixExtrem(---), decreixFort(--), decreix(-), \\ estable(=), creix(+), creixFort(++), creixExtrem(+++)\}$$

### 3.1.2.3 Selecció d’etiquetes per a la durada dels períodes

Afortunadament aquest paràmetre és molt intuïtiu, perquè està «en escala humana» i ens resultarà més fàcil avaluar-lo. Una panoràmica de la variable:

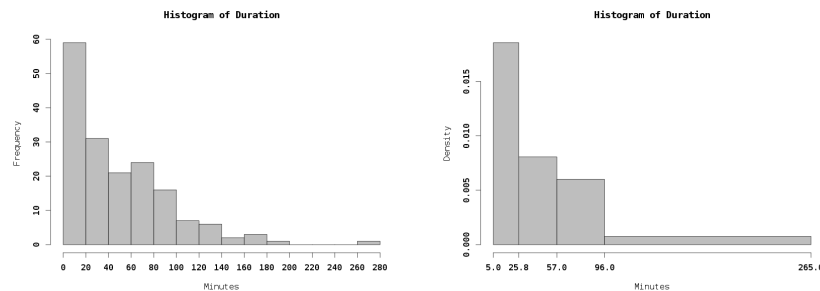
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
5.00	17.50	40.00	52.69	80.00	265.00

Amb aquestes dades i l'histograma de la Figura 11a, si volguéssim fer 4 intervals a mà, podríem triar p. ex: un període curt fins 20 min, un mitjà fins a 40/50, un llarg fins a 90/100, i un llarguíssim per a la resta. Clar que hi ha molta diferència entre un període de 120 min o un de 260 min, però no sembla desenraonat pensar que períodes tan llargs deguen tenir causes comuns. Davant de l'arbitrarietat de triar els períodes a mà, provem el mètode CUM per 3 intervals ([5, 36, 80, 265]) i 4 intervals ([5, 26, 57, 96, 265]). En la primera opció, el període de temps que etiquetaríem com a «curt» no ho sembla tant, a priori la segona opció (Figura 11b) s'aproxima més al que hauríem decidit a mà. Tot i això, quan fem l'experimentació provarem les dues opcions. A mode d'exemple, les etiquetes i els intervals de  $D$  podrien ser:

*Un major nombre d'intervals per a  $D$  afegiria complexitat a l'hora de transformar i combinar períodes, com es despren de la Secció 3.2.2.1.*

$$D_{INT} = \{[5, 26], (26, 57], (57, 96], (96, 265]\}$$

$$D_{LAB} = \{\text{curt}(\cdot), \text{normal}(\cdot), \text{llarg}(\cdot), \text{llarguíssim}(\cdot)\}$$



(a) Histograma (Freedman-Diaconis) de la durada de tots els períodes; s'han obtingut 14 intervals. (b) Histograma (CUM) amb 4 intervals.

Figure 11: Histogrames de la durada dels períodes.

## 3.2 CLÚSTERING

Per a poder agrupar les corbes, hem de ser capaços de comparar-les. Per tal de comparar-les cal definir una mesura entre elles. Amb aquesta fi, primer caldrà obtenir una mesura de similitud entre els períodes que les componen.

### 3.2.1 Similitud entre períodes

Per comparar dos períodes, cal establir una mesura de similitud entre els paràmetres que els defineixen: magnitud ( $M$ ), canvi ( $C$ ) i duració

M	«baixa»	«mitjana-inf.»	«mitjana-sup.»	«alta»	«gegant»
«baixa»	0	1	2	3	4
«mitjana-inf.»	1	0	1	2	3
«mitjana-sup.»	2	1	0	1	2
«alta»	3	2	1	0	1
«gegant»	4	3	2	1	0

Taula 2: Matriu de dissimilitud usant «veïnatge conceptual» entre les etiquetes (Secció 3.1.2.1) del paràmetre  $M$ .

del període ( $D$ ). En aquest sentit contemplem les 2 possibilitats proposades per Falomir et al. [8]: veïnatge conceptual i distància entre intervals.

### 3.2.1.1 Veïnatge conceptual

Les relacions de veïnatge conceptual s'estableixen entre les etiquetes dels paràmetres qualitius. Podem imaginar un graf on els vèrtexs són les etiquetes d'un paràmetre i les arestes uneixen parelles d'etiquetes contigües. Aleshores la distància entre etiquetes és igual al cost del camí òptim entre els vèrtexs corresponents del graf. Per exemple, en el cas de la magnitud de  $Ra$  ( $M$ ) que hem definit a l'apartat 3.1.2.1, obtindríem la Taula 2.

Com els paràmetres magnitud ( $M$ ), canvi ( $C$ ) i duració ( $D$ ) estan definits en intervals continus, les matrius de dissimilitud es calcularien amb la forma general (1) on  $X \in \{M, C, D\}$ .

$$ds(X_i, X_j) = |i-j| \text{ amb } i, j = 1 \dots K_X \quad (1)$$

Així la similitud entre dos períodes  $P_A$  i  $P_B$  quedaria establerta com:

$$Sim(P_A, P_B) = 1 - \sum_{i \in \{M, C, D\}} w_i \frac{ds(i)}{Ds(i)} \quad (2)$$

On  $ds(i)$  denota la dissimilitud entre el paràmetre  $i$  dels períodes  $P_A$  i  $P_B$ ;  $Ds(i)$  la màxima dissimilitud possible entre paràmetres  $i$ . A més,  $w_i$  denota el pes que se li vol donar a un paràmetre, complint-se la propietat  $w_M + w_C + w_D = 1$  i  $w_i > 0$  per a tots els paràmetres.

### 3.2.1.2 Distància entre intervals

Si els intervals que hem definit no són equidistants, tenim la possibilitat d'enriquir el càlcul de dissimilitud afegint-hi l'amplitud dels intervals. Donat un interval  $I_1 = (a_1, b_1)$  definim  $c_1 = (a_1 + b_1)/2$  (centre) i  $r_1 = (b_1 - a_1)/2$  (radi), aleshores establim la distància amb un segon interval  $I_2 = (a_2, b_2)$  de centre i radi anàlegs com:

M	«baixa»	«mitjana-inf.»	«mitjana-sup.»	«alta»	«gegant»
«baixa»	0	73.0	129.2	220.4	475
«mitjana-inf.»	73.0	0	57.6	152.5	413.1
«mitjana-sup.»	129.2	57.6	0	96.2	358.7
«alta»	220.4	152.5	96.2	0	263.1
«gegant»	475	413.1	358.7	263.1	0

Taula 3: Matriu de dissimilitud usant «distància entre intervals» entre les etiquetes (Secció 3.1.2.1) del paràmetre M.

$$d(I_1, I_2) = \sqrt{(c_2 - c_1)^2 + (r_2 - r_1)^2}$$

Fent servir aquesta distància construiríem les matrius de dissimilitud per als nostres tres paràmetres  $M$ ,  $C$  i  $D$  (vegeu la Taula 3 com a exemple). La similitud entre dos períodes  $P_A$  i  $P_B$  es calcularia amb la fórmula (2) exactament igual que per al veïnatge conceptual.

### 3.2.2 Similitud entre corbes

Si les corbes  $A$  i  $B$  tenen la mateixa quantitat  $n$  de períodes, la comparació serà directa: compararem el 1r període de  $A$  amb el 1r de  $B$  obtenint  $Sim(P_{A_1}, P_{B_1})$ , el 2n de  $A$  amb el 2n de  $B$  obtenint  $Sim(P_{A_2}, P_{B_2})$ , i així amb tots els períodes, de manera que la similitud entre les corbes  $A$  i  $B$  serà igual a la mitjana de les similituds entre tots els seus períodes:

$$Sim(A, B) = \frac{\sum_{i=1}^n Sim(P_{A_i}, P_{B_i})}{n}$$

Ara bé, quan ens trobem amb dues corbes que tenen diferent nombre de períodes no podem fer una comparació directa. Haurem d'intentar convertir les corbes d'alguna manera per a fer-les compatibles.

#### 3.2.2.1 Corbes amb diferent nombre de períodes

Afortunadament tenim una manera intuïtiva de fer-ho, explotant el fet que els períodes tenen una característica temporal i aplicant-hi els principis del «veïnatge».

Primer assignarem un nombre natural ascendent a les etiquetes de  $D$ , de manera que  $D_{LAB} = \{curt, normal, llarg, llarguíssim\}$  equival a  $D_{NUM} = \{1, 2, 3, 4\}$  i usarem aquestes últimes. Aleshores donat un període qualsevol  $P$  definit com a  $\{M_P, C_P, D_P\}$  el podem subdividir en  $n$  períodes equivalents  $P', P'', \dots, P^n$  on el sumatori  $\sum_{i=1}^n D_{P_i}$  ha de ser igual al valor de  $D$  del període  $P$  original. Per exemple, un període amb  $D_P(3)$  pot donar lloc a:

- 2 nous períodes amb:  $D_{P'}(2), D_{P''}(1)$

$D_p$	Possibles descomposicions
[D(4)]	[D(1),D(3)], [D(3),D(1)], [D(1),D(1),D(2)], [D(1),D(2),D(1)], [D(2),D(1),D(1)], [D(1),D(1),D(1),D(1)]
[D(3)]	[D(1),D(2)], [D(2),D(1)], [D(1),D(1),D(1)]
[D(2)]	[D(1),D(1)]
[D(1)]	--

Taula 4: Descomposició de períodes basada en l'etiqueta temporal  $D$ .

- 2 nous períodes amb:  $D_{p'}(1), D_{p''}(2)$
- 3 nous períodes amb:  $D_{p'}(1), D_{p''}(1), D_{p'''}(1)$

On l'ordre importa, perquè les corbes estan definides com a períodes temporals consecutius que es compararan 1 a 1. La Taula 4 mostra un exemple de la descomposició dels 4 possibles valors de  $D$ . En totes les subdivisions, els paràmetres  $M$  i  $C$  mantenen el valor del període original.

Per a dues corbes  $A$  i  $B$  que tenen diferent nombre de períodes i sent  $A$  la que té menor quantitat, generarem totes les possibles descomposicions de tots els períodes de  $A$  i les combinarem. Així obtindrem  $n$  corbes alternatives a  $A$ , de les quals només algunes seran compatibles amb  $B$  (tindran el mateix nombre de períodes). Finalment definirem la similitud entre les corbes  $A$  i  $B$ , com la menor de les similituds entre  $B$  i totes les alternatives a  $A$  compatibles amb  $B$ .

### 3.2.3 Mètode de clústering

Per ara ens limitarem a usar un algorisme jeràrquic de clústering<sup>2</sup> perquè la nostra funció de distància és un poc especial –si les dues corbes no tenen el mateix nombre de característiques, les transforma. Fins on sabem, en la majoria de paquets R els mètodes de clústering habituals ( $k$ -means, p. ex.) esperen rebre com a entrada una matriu *observacions*  $\times$  *variables*, i en el nostre cas això és impossible.

Calcularem la similitud, com s'ha descrit en l'apartat anterior, entre totes les corbes. Amb això obtenim una taula de distàncies que ens permet fer el clústering i obtenir un dendrograma. A partir d'aquest obtindrem tots els clústering possibles dintre d'un rang  $k$  acceptable, per exemple de 5 a 13 grups.

*Una corba A pot transformar-se de diferent manera, depenent de amb quina corba B la comparem.*

## 3.3 VALIDACIÓ

Un cop hem obtingut el clústering, aplicarem 3 mètodes de validació:

<sup>2</sup> Mètode `hclust` del paquet R estàndard `stats`, usant `'Ward.D2'`

1. **Connectivitat**<sup>3</sup>: mesura si les observacions «veïnes» s'han classificat en el mateix grup. Dona com a resultat un valor entre 0 i  $\infty$ , que hauria de ser minimitzat.
2. **DI**<sup>4</sup>: mesura la compactesa dels clústers i la bona separació entre ells. Dona com a resultat un valor entre 0 i  $\infty$ , que hauria de ser maximitzat.
3. **Silueta**<sup>5</sup>: mesura com de bé encaixen les observacions dintre dels seus clústers. Dona un resultat en el rang  $[-1, 1]$ , que hauria de ser maximitzat.

El paquet `clValid` de R permet automatitzar el procés de validació per a molts mètodes de clústering i moltes mesures de validació, però degut a la manera en què calculem la distància entre corbes no podem usar-lo. Per fer-ho caldria modificar-lo, i això excediria el temps planificat per al projecte. Tanmateix, utilitzarem la tècnica que s'usa en el paquet per agregar les puntuacions de les diferents mesures de validació. L'agregació<sup>6</sup> serveix per a reconciliar les classificacions i produir una llista global, que usant totes les mesures de validació alhora determina el guanyador en general [4].

---

3 Mètode `connectivity` del paquet R `clValid`

4 Mètode `dunn` del paquet R `clValid`

5 Mètode `silhouette` del paquet R `cluster`

6 Paquet R `RankAggreg`



## EXPERIMENTACIÓ I DISCUSSIÓ

---

### 4.1 RESULTATS

Hem combinat una gran quantitat d'opcions per configurar diferents clústerings i mesurar-ne la qualitat. En total han resultat 624 combinacions, i per a cadascuna s'ha aplicat un algorisme jeràrquic per obtenir de [5, 13] clústers. Un total de 5616 possibilitats, cadascuna amb les seues mesures de qualitat.

Les opcions de configuració que s'han combinat són:

- **M**: El nombre d'interval·s per a  $M$  (mitjana de  $Ra$  del període).  
Valors: [3, 4, 5]
- **C**: El nombre d'interval·s per a  $C$  (mitjana de  $\Delta Ra$  del període).  
Valors: [2, 3, 4, 5, 6, 7]
- **D**: El nombre d'interval·s per a  $D$  (*duració* en min del període).  
Valors: [3, 4]
- **C me**: El mètode per elegir els interval·s de  $C$ .  
Valors: ['man', 'pks', 'CUM']
- **Diss**: El mètode per calcular la taula de dissimilitud entre corbes (distància entre interval·s o veïnatge conceptual).  
Valors: ['ID', 'CN']
- **w**: El pes dels paràmetres  $C$ ,  $M$  i  $D$ .  
Valors: [{0.35, 0.35, 0.3}, {0.33, 0.33, 0.33}, {0.3, 0.3, 0.4}, {0.25, 0.25, 0.5}]
- **k**: El nombre de clústers a obtenir.  
Valors: [5, 6, 7, 8, 9, 10, 11, 12, 13]

*Quan s'usa el mètode dels pics/valls (pks), l'opció C queda limitada al valor 5.*

#### 4.1.1 Observació de la qualitat

Els 10 millors clústerings d'acord amb la Connectivitat (Taula 6 a la pàgina següent) tenen tots  $k=5$ , el menor possible. Això és un símptoma d'esbiaixament en la mesura. En observar tots els resultats de la Connectivitat, veiem que hi ha una correlació entre l'augment de  $k$  i l'empitjorament en la puntuació, fins a tal punt que la resta d'opcions acaben important poc. Una hipòtesi que explicaria aquest comportament és que la selecció de característiques rellevants ha provocat que totes les observacions estiguen molt a prop. La Connectivitat puntua

*La Connectivitat produeix valors més grans per a clústerings de pitjor qualitat.*

	k	M	C	D	C me.	Diss.	$\{w_C, w_M, w_D\}$	Conn.	DI	Sil.
1	13	3	3	3	CUM	CN	{0.35,0.35,0.30}	49.154	0.857	0.813
2	13	3	3	3	CUM	CN	{0.33,0.33,0.33}	51.829	1	0.806
3	13	3	3	3	CUM	CN	{0.25,0.25,0.50}	50.083	0.5	0.867
4	12	3	3	3	CUM	CN	{0.33,0.33,0.33}	46.003	1	0.767
5	5	5	3	4	CUM	ID	{0.25,0.25,0.50}	10.652	0.244	0.663
6	5	3	3	4	CUM	ID	{0.25,0.25,0.50}	11.62	0.173	0.641
7	5	5	3	4	CUM	ID	{0.35,0.35,0.30}	11.786	0.388	0.637
8	5	5	3	4	CUM	ID	{0.33,0.33,0.33}	11.786	0.359	0.635
9	13	3	3	3	CUM	CN	{0.30,0.30,0.40}	50.133	0.5	0.852
10	5	5	3	3	CUM	ID	{0.25,0.25,0.50}	12.746	0.487	0.714

Taula 5: Els 10 millors resultats agregant les tres mesures de qualitat interna.

	k	M	C	D	CUM	Diss.	$\{w_C, w_M, w_D\}$	Conn.
1	5	5	3	4	CUM	ID	{0.25,0.25,0.50}	10.652
2	5	3	3	4	CUM	ID	{0.25,0.25,0.50}	11.62
3	5	5	3	4	CUM	ID	{0.35,0.35,0.30}	11.786
4	5	5	3	4	CUM	ID	{0.33,0.33,0.33}	11.786
5	5	5	3	3	CUM	ID	{0.25,0.25,0.50}	12.746
...								
5613	13	5	4	3	man	CN	{0.35,0.35,0.30}	74.038
5614	13	5	6	3	man	CN	{0.35,0.35,0.30}	74.442
5615	13	3	5	4	CUM	CN	{0.30,0.30,0.40}	74.678
5616	13	5	4	4	man	CN	{0.25,0.25,0.50}	75.476

Taula 6: Els millors resultats segons la Connectivitat.

positivament les observacions «veïnes» dintre d’una mateixa partició. Si el modelatge ha fet que totes estiguen «a prop», aleshores amb menys clústers millora la puntuació. Donada aquesta situació, la mesura no ens és de cap ajuda per a triar el nombre de clústers adequat.

Pel que fa al **DI** (Taula 7 a la pàgina següent) a continuació mostrem les estadístiques dels 50 millors resultats:

*Al contrari que amb la Connectivitat, per al DI i la Silueta un valor més alt significa millor qualitat del clústering.*

k		M		C		D		C me.		Diss.		w	
Min.	: 5.0	Min.	:3.00	Min.	:3.00	Min.	:3.00	Min.	:2.00	Min.	:1.00	Min.	:1.00
1st Qu.	: 8.0	1st Qu.	:3.00	1st Qu.	:3.00	1st Qu.	:3.00	1st Qu.	:3.00	1st Qu.	:1.00	1st Qu.	:2.00
Median	:11.0	Median	:3.00	Median	:3.00	Median	:3.00	Median	:3.00	Median	:1.00	Median	:3.00
Mean	:10.4	Mean	:3.52	Mean	:3.04	Mean	:3.32	Mean	:2.98	Mean	:1.38	Mean	:2.98
3rd Qu.	:12.0	3rd Qu.	:4.00	3rd Qu.	:3.00	3rd Qu.	:4.00	3rd Qu.	:3.00	3rd Qu.	:2.00	3rd Qu.	:4.00
Max.	:13.0	Max.	:5.00	Max.	:5.00	Max.	:4.00	Max.	:3.00	Max.	:2.00	Max.	:4.00

-----

C me: { 1='man', 2='pks', 3='CUM' }  
 Diss: { 1='ID', 2='CN' }  
 w : { 1=(0.35, 0.35, 0.3), 2=(0.33, 0.33, 0.33), 3=(0.3, 0.3, 0.4), 4=(0.25, 0.25, 0.5) }

Si ens fixem en el valor mitjà, veiem que destaca tant el predomini de C me='CUM' i C=3, que resulta sospitós. Una ullada ràpida a la des-

	k	M	C	D	CUM	Diss.	$\{w_M, w_C, w_D\}$	DI
1	13	3	3	3	CUM	CN	$\{0.33, 0.33, 0.33\}$	1
2	12	3	3	3	CUM	CN	$\{0.33, 0.33, 0.33\}$	1
3	13	3	3	3	CUM	CN	$\{0.35, 0.35, 0.30\}$	0.857
4	12	3	3	3	CUM	CN	$\{0.35, 0.35, 0.30\}$	0.857
5	7	5	3	3	CUM	ID	$\{0.25, 0.25, 0.50\}$	0.68
...								
5613	5	3	5	3	pks	ID	$\{0.25, 0.25, 0.50\}$	0.05
5614	7	3	5	3	pks	ID	$\{0.30, 0.30, 0.40\}$	0.045
5615	6	3	5	3	pks	ID	$\{0.30, 0.30, 0.40\}$	0.045
5616	5	3	5	3	pks	ID	$\{0.30, 0.30, 0.40\}$	0.045

Taula 7: Els millors resultats segons el DI.

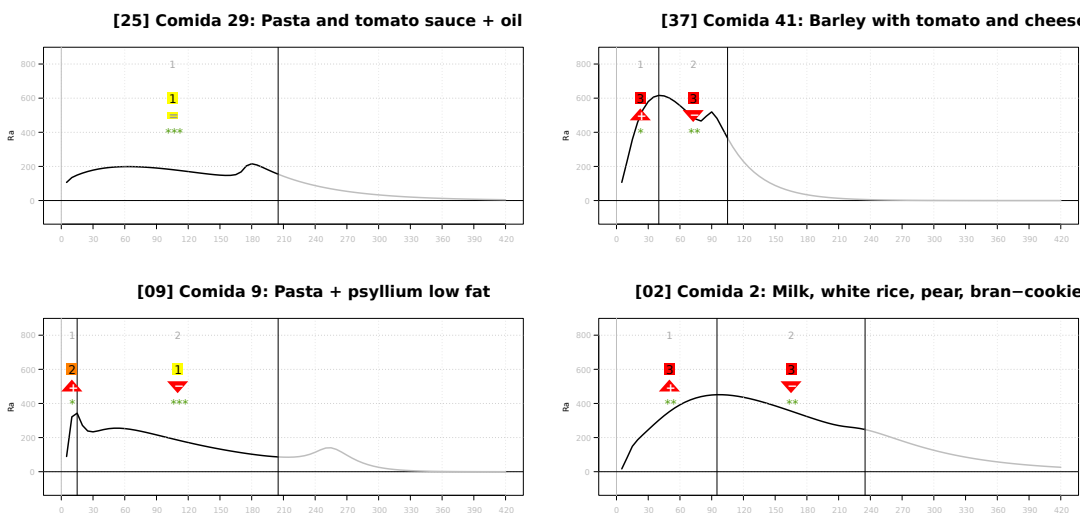


Figura 12: Mostra de 4 descripcions qualitatives extreptes del millor resultat segons el DI (1a fila de la Taula 7). En el menjar 29 queda molt clar que l'interval central *estable* (=) generat amb el mètode CUM és massa ampli, ja que inclou valors de creixement i decreixement moderats. Es pot veure la descripció qualitativa de tots els menjars generada per aquesta configuració a les Figures 30 i 31 a la pàgina 87 (Apèndix B).

	k	M	C	D	C me.	Diss.	$\{w_C, w_M, w_D\}$	Conn.	DI	Sil.
1	12	3	3	3	man	ID	$\{0.30, 0.30, 0.40\}$	50.289	0.44	0.55
2	9	5	3	3	man	ID	$\{0.30, 0.30, 0.40\}$	35.006	0.375	0.552
3	9	5	3	3	man	ID	$\{0.25, 0.25, 0.50\}$	34.864	0.303	0.552
4	5	5	3	3	man	ID	$\{0.25, 0.25, 0.50\}$	17.158	0.219	0.458
5	11	3	3	3	man	ID	$\{0.30, 0.30, 0.40\}$	47.634	0.44	0.537
6	12	5	5	3	pks	CN	$\{0.25, 0.25, 0.50\}$	56.167	0.5	0.495
7	5	3	6	3	man	ID	$\{0.35, 0.35, 0.30\}$	17.908	0.179	0.403
8	5	5	3	3	man	ID	$\{0.30, 0.30, 0.40\}$	18.137	0.235	0.422
9	7	3	3	3	man	ID	$\{0.25, 0.25, 0.50\}$	27.257	0.296	0.547
10	5	4	6	3	man	ID	$\{0.35, 0.35, 0.30\}$	18.175	0.198	0.383

Taula 8: Els 10 millors resultats, excloent CUM &amp;&amp; C&lt;=3.

cripció qualitativa generada per aquestes opcions (vegeu p. ex. els menjars 29 i 41 a la Figura 12 a la pàgina anterior) ho explica ràpidament: CUM amb 3 intervals ha generat un interval central *estable* (=) massa ampli:  $C_{INT} = \{[-58.4, -2.7], (-2.7, 21.6], (21.6, 289.7]\}$

Això és molt negatiu perquè estem perdent informació que considerem essencial per descriure la corba. Amb el mètode CUM haurem d'usar almenys 4 intervals per elegir C, a partir d'ara ometrem dels resultats les combinacions que no compleixen això (passem de 5616 a 4752 combinacions).

Amb la retirada de les opcions problemàtiques, els resultats segons puntuació agregada han canviat substancialment (Taula 8 *versus* l'anterior Taula 5 a la pàgina 36). En visualitzar la descripció qualitativa del millor resultat (Figura 13 a la pàgina següent) veiem que de nou passa alguna cosa estranya amb els 3 intervals de C: només s'aprecien 2 intervals (o creix, o decreix). Això pot ser a causa del mecanisme d'absorció de períodes (explicat a la pàgina 25), o a causa de que la selecció manual de l'interval central *estable* (=) inclou pocs valors. Per ara, tot i que la descripció no és errònia *a priori*, preferirem evitar les combinacions amb menys de 4 intervals (de qualsevol mètode) per a C, ja que no es comporten com teníem previst.

Després d'un segon filtratge passem a tenir 3888 combinacions d'opcions. De nou la llista de millors puntuacions (Taula 9 a la pàgina següent) ha canviat substancialment. La descripció qualitativa de la millor combinació (Figura 14 a la pàgina 40) mostra com estem tornant a perdre un interval de C, però ara almenys ens assegurem que no hi ha una absorció excessiva ja que amb  $C > 4$  el primer i últim interval ( $--, ++$ ) no es poden fusionar amb el central (=). Per tant, podem considerar que aquesta descripció és suficient i adequada, almenys pel que fa a C, i no tenim raons per descartar més paràmetres.

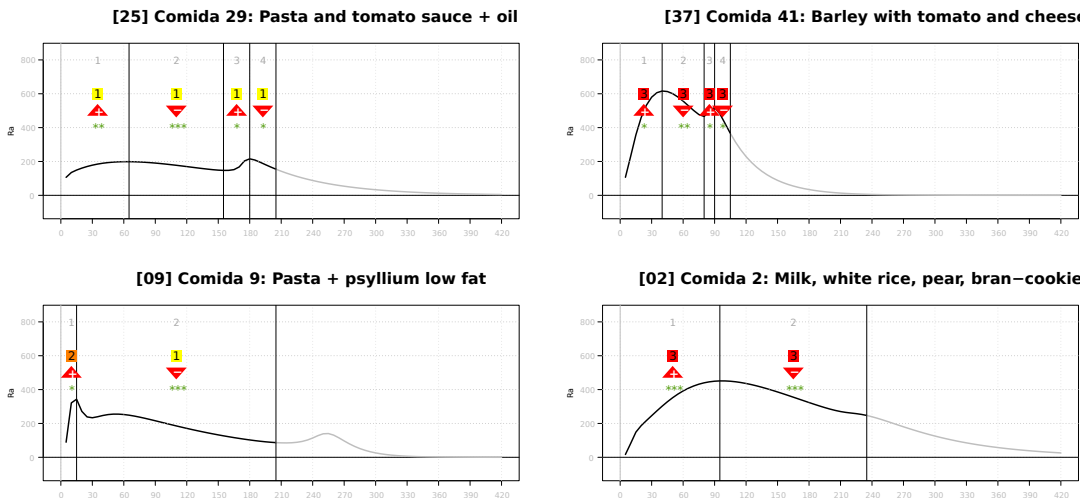


Figura 13: Mostra de 4 descripcions qualitatives extremes del millor resultat segons la puntuació agregada (1a fila de la Taula 8 a la pàgina anterior). Com es pot comprovar a la descripció qualitativa completa (Figures 32 i 33 a la pàgina 89, Apèndix B) encara que la configuració és  $C=3$ , de facto només hi ha dos intervals de  $C$ : *creix*(+), *decreix*(-).

	k	M	C	D	C me.	Diss.	$\{w_C, w_M, w_D\}$	Conn.	DI	Sil.
1	12	5	5	3	pks	CN	{0.25, 0.25, 0.50}	56.167	0.5	0.495
2	5	3	6	3	man	ID	{0.35, 0.35, 0.30}	17.908	0.179	0.403
3	13	3	5	3	pks	ID	{0.25, 0.25, 0.50}	61.429	0.136	0.538
4	5	4	6	3	man	ID	{0.35, 0.35, 0.30}	18.175	0.198	0.383
5	7	3	5	4	pks	ID	{0.35, 0.35, 0.30}	27.794	0.255	0.53
6	5	3	5	3	pks	CN	{0.35, 0.35, 0.30}	18.927	0.211	0.423
7	5	3	5	4	pks	CN	{0.30, 0.30, 0.40}	19.154	0.193	0.421
8	5	3	5	4	man	ID	{0.33, 0.33, 0.33}	19.467	0.284	0.43
9	5	3	5	4	man	ID	{0.35, 0.35, 0.30}	19.523	0.306	0.455
10	13	5	5	4	CUM	CN	{0.35, 0.35, 0.30}	70.064	0.467	0.356
11	11	3	5	3	pks	ID	{0.25, 0.25, 0.50}	57.374	0.102	0.525
12	13	5	6	4	CUM	CN	{0.25, 0.25, 0.50}	68.955	0.442	0.35
13	12	3	5	3	pks	ID	{0.25, 0.25, 0.50}	59.208	0.136	0.516
14	13	3	4	3	CUM	ID	{0.25, 0.25, 0.50}	54.612	0.172	0.514
15	5	3	5	4	pks	CN	{0.35, 0.35, 0.30}	19.591	0.224	0.431
16	10	3	5	3	pks	ID	{0.25, 0.25, 0.50}	54.473	0.102	0.511
17	5	5	6	4	man	ID	{0.33, 0.33, 0.33}	19.698	0.217	0.367
18	13	3	4	3	CUM	ID	{0.30, 0.30, 0.40}	55.319	0.2	0.511
19	12	3	4	3	CUM	ID	{0.25, 0.25, 0.50}	52.512	0.172	0.506
20	10	5	5	3	pks	ID	{0.25, 0.25, 0.50}	50.196	0.157	0.509

Taula 9: Els 20 millors resultats, exclouent  $C \leq 3$ .

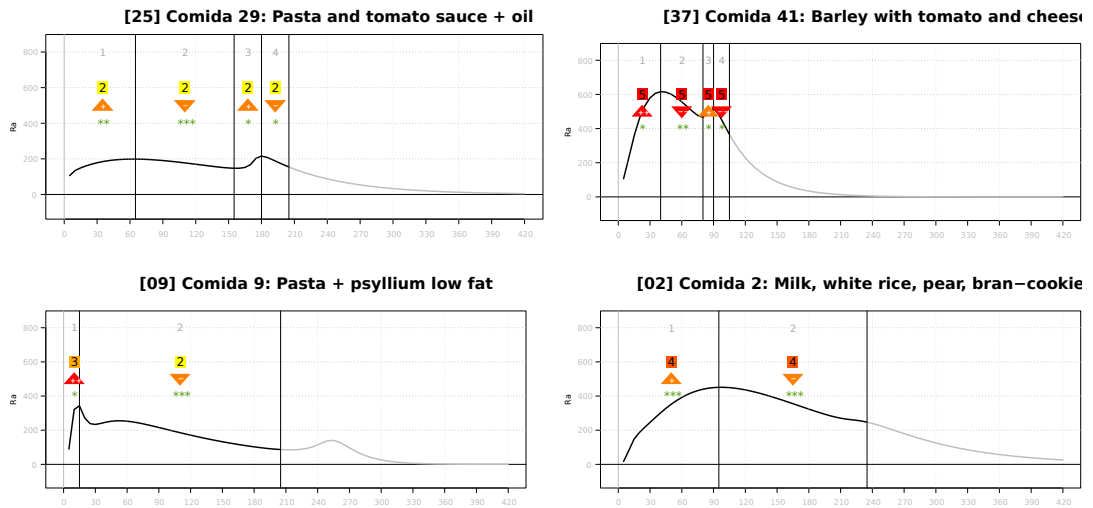


Figura 14: Mostra de 4 descripcions qualitatives extremes del millor resultat segons la puntuació agregada excloent  $C \leq 3$  (1a fila de la Taula 9 a la pàgina anterior). Es pot veure la descripció qualitativa de tots els menjars generada per aquesta configuració a les Figures 34 i 35 a la pàgina 91 (Apèndix B).

Degut a l'esbiaixament tan accentuat que hem observat en la Connectivitat, creiem que es millor centrar-se només en els resultats amb **DI** i Silueta. Respecte al **DI** (Taula 10 a la pàgina següent) podem observar una certa tendència inversa a la què hem vist en la Connectivitat: la puntuació es degrada a menor nombre de clústers. La Silueta també té aquesta tendència amb excepcions, per exemple en el 2n millor resultat (Taula 11 a la pàgina següent).

Les tendències oposades en algunes opcions indiquen que la mesura està esbiaixada respecte a aqueixa opció, i per tant no la podem donar per bona. Per exemple en el càlcul de dissimilitud, **DI** prefereix veïnatge conceptual mentre la Silueta prefereix distància entre intervals: quina és millor? No ho podem dir. En canvi, quan ambdós mesures coincideixen en assenyalar certes opcions com a millors, ho acceptem (amb certa cautela) com un bon indicatiu. Per exemple, les dues mesures es decanten pel pes  $\{0.25, 0.25, 0.50\}$  i  $C=5$ , la resta no és concloent.

Si creuem les dues mesures i observem les millors puntuacions (en blau a la Taula 12 a la pàgina 42), es confirmen uns millors resultats amb un pes de  $\{0.25, 0.25, 0.50\}$  i  $C=5$ , amb menys seguretat també destaquen el mètode dels pics/valls per elegir els intervals de  $C$  i la dissimilitud mitjançant veïnatge conceptual. El valor  $k$  continua essent elevat, podria deure's a la inexistència de «clústers naturals» en el conjunt de dades, però sembla molt més probable que es dega a l'elecció de les característiques rellevants per a la discretització. De les 4 combinacions destacades, incloem les imatges del clústering generat per la 1a combinació (Figura 15 a la pàgina 43) i la 14a combinació (Figura 16 a la pàgina 44).

	k	M	C	D	CUM	Diss.	$\{w_C, w_M, w_D\}$	DI
1	12	5	5	3	pks	CN	{0.25, 0.25, 0.50}	0.5
2	13	5	5	4	CUM	CN	{0.35, 0.35, 0.30}	0.467
3	13	5	6	4	CUM	CN	{0.25, 0.25, 0.50}	0.442
4	13	4	5	3	pks	CN	{0.25, 0.25, 0.50}	0.429
5	13	5	5	4	CUM	CN	{0.33, 0.33, 0.33}	0.429
6	13	3	5	4	man	CN	{0.25, 0.25, 0.50}	0.424
7	11	3	5	4	pks	CN	{0.25, 0.25, 0.50}	0.421
8	12	5	5	4	CUM	CN	{0.25, 0.25, 0.50}	0.421
9	13	5	5	4	CUM	CN	{0.25, 0.25, 0.50}	0.421
10	12	5	5	4	CUM	CN	{0.35, 0.35, 0.30}	0.42
	...							
3886	7	3	5	3	pks	ID	{0.30, 0.30, 0.40}	0.045
3887	6	3	5	3	pks	ID	{0.30, 0.30, 0.40}	0.045
3888	5	3	5	3	pks	ID	{0.30, 0.30, 0.40}	0.045

Taula 10: Els millors resultats segons DI excloent  $C \leq 3$ .

	k	M	C	D	CUM	Diss.	$\{w_C, w_M, w_D\}$	Sil.
1	13	3	5	3	pks	ID	{0.25, 0.25, 0.50}	0.538
2	7	3	5	4	pks	ID	{0.35, 0.35, 0.30}	0.53
3	11	3	5	3	pks	ID	{0.25, 0.25, 0.50}	0.525
4	12	3	5	3	pks	ID	{0.25, 0.25, 0.50}	0.516
5	13	3	4	3	CUM	ID	{0.25, 0.25, 0.50}	0.514
6	10	3	5	3	pks	ID	{0.25, 0.25, 0.50}	0.511
7	13	3	4	3	CUM	ID	{0.30, 0.30, 0.40}	0.511
8	10	5	5	3	pks	ID	{0.25, 0.25, 0.50}	0.509
9	12	3	4	3	CUM	ID	{0.25, 0.25, 0.50}	0.506
10	13	4	5	3	pks	ID	{0.30, 0.30, 0.40}	0.505
	...							
3886	5	5	4	4	man	CN	{0.35, 0.35, 0.30}	0.218
3887	6	5	4	3	man	CN	{0.35, 0.35, 0.30}	0.218
3888	5	5	5	3	CUM	CN	{0.33, 0.33, 0.33}	0.216

Taula 11: Els millors resultats segons la Silueta excloent  $C \leq 3$ .

	k	M	C	D	C me.	Diss.	$\{w_C, w_M, w_D\}$	DI	Sil.
1	12	5	5	3	pks	CN	{0.25, 0.25, 0.50}	0.5	0.495
2	13	3	5	3	pks	ID	{0.25, 0.25, 0.50}	0.136	0.538
3	7	3	5	4	pks	ID	{0.35, 0.35, 0.30}	0.255	0.53
4	11	3	5	3	pks	ID	{0.25, 0.25, 0.50}	0.102	0.525
5	13	5	5	4	CUM	CN	{0.35, 0.35, 0.30}	0.467	0.356
6	12	3	5	3	pks	ID	{0.25, 0.25, 0.50}	0.136	0.516
7	13	3	4	3	CUM	ID	{0.25, 0.25, 0.50}	0.172	0.514
8	10	3	5	3	pks	ID	{0.25, 0.25, 0.50}	0.102	0.511
9	13	5	6	4	CUM	CN	{0.25, 0.25, 0.50}	0.442	0.35
10	13	4	5	3	pks	CN	{0.25, 0.25, 0.50}	0.429	0.474
11	13	3	4	3	CUM	ID	{0.30, 0.30, 0.40}	0.2	0.511
12	13	5	5	4	CUM	CN	{0.33, 0.33, 0.33}	0.429	0.356
13	10	5	5	3	pks	ID	{0.25, 0.25, 0.50}	0.157	0.509
14	13	3	5	4	man	CN	{0.25, 0.25, 0.50}	0.424	0.436
15	11	3	5	4	pks	CN	{0.25, 0.25, 0.50}	0.421	0.431

Taula 12: Els 15 millors resultats segons DI i Silueta. Destaquen les files 1, 10, 14 i 15. Es pot veure el clústering de la 1a combinació a la Figura 15 a la pàgina següent, i de la 14a combinació a la Figura 16 a la pàgina 44.



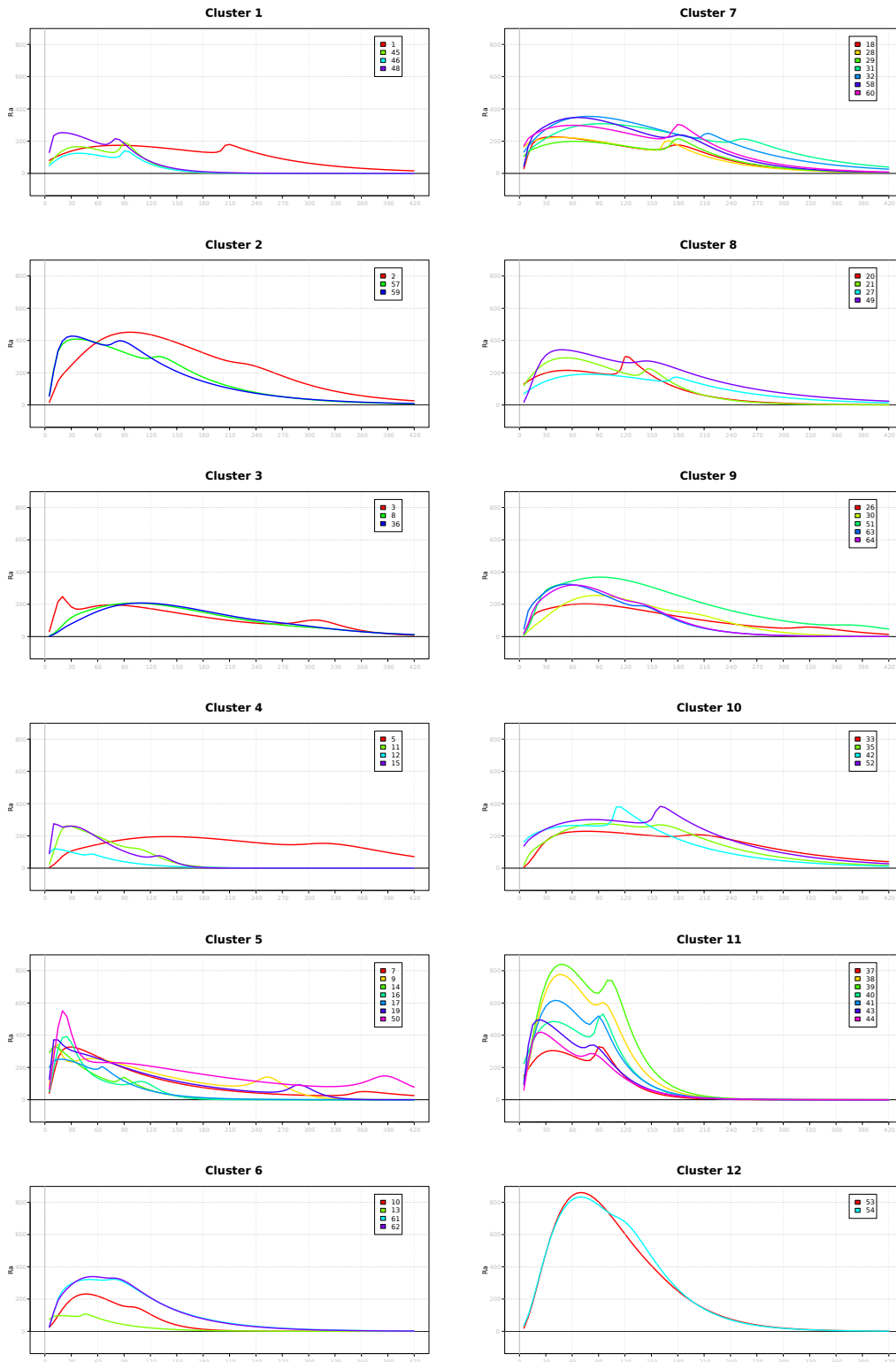


Figura 15: Clústering generat pel millor resultat segons DI i Silueta (1a fila de la Taula 12 a la pàgina anterior). Es pot veure més informació sobre cada grup a les Figures 36 i 37 a la pàgina 93, i la descripció qualitativa de cada corba a les Figures 38 i 39 a la pàgina 95 (Apèndix B).

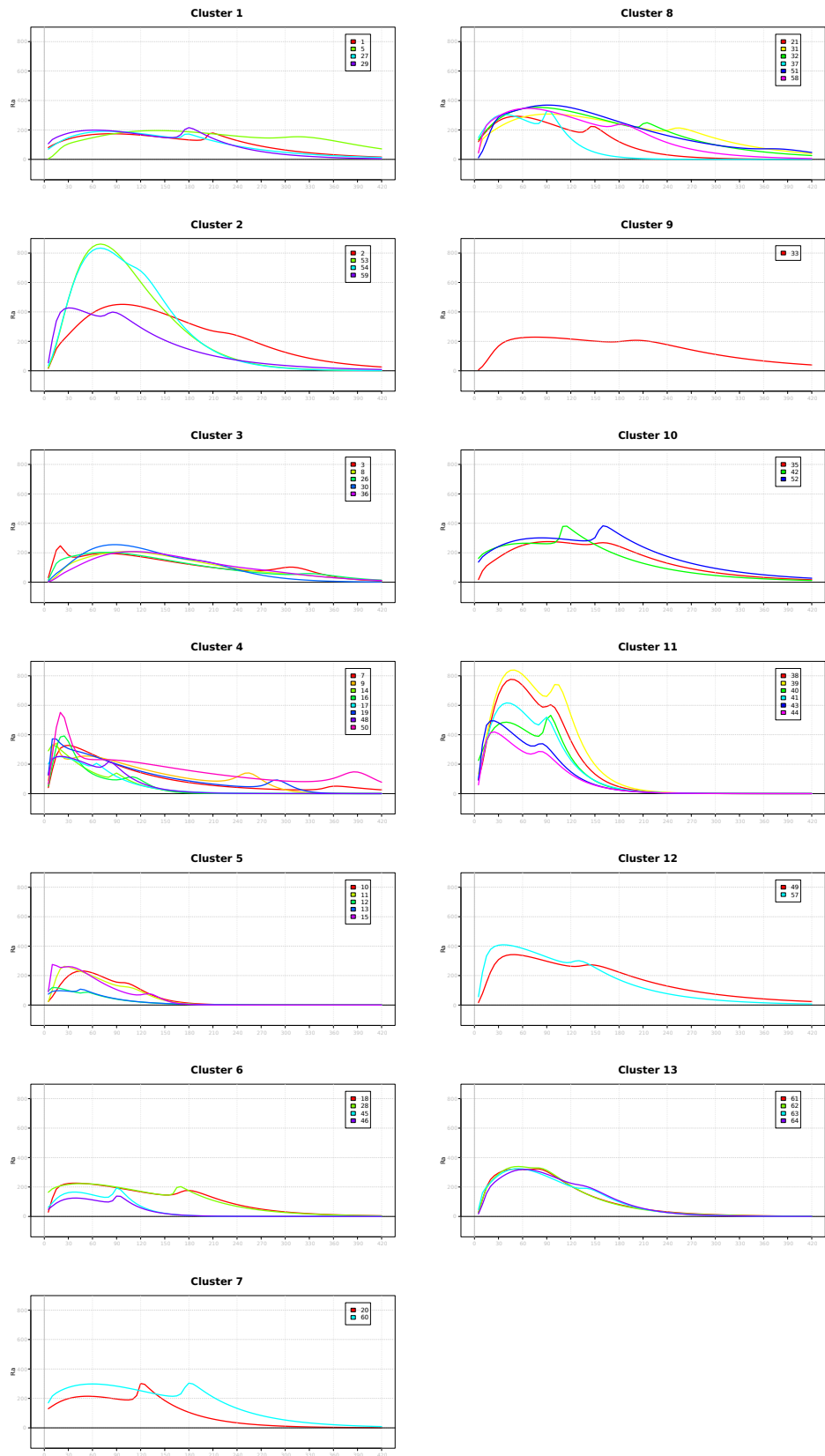


Figura 16: Clústering generat pel 14è millor resultat segons *DI* i Silueta (14a fila de la Taula 12 a la pàgina 42). Es pot veure més informació sobre cada grup a les Figures 40 i 41 a la pàgina 97, i la descripció qualitativa de cada corba a les Figures 42 i 43 a la pàgina 99 (Apèndix B).

Estudiant la 1a combinació es veu que alguns grups similars s'haurien de fusionar (p. ex. clúster 7 i 8), com era previsible tenint un valor  $k$  elevat. Entre les particions generades, s'observen algunes corbes que aparentment no encaixen al grup assignat:

- Els menjars 1 (clúster 1) i 5 (clúster 4) són un exemple de descripció qualitativa defectuosa. Com es pot apreciar a la Figura 38 a la pàgina 94 (Apèndix B), la baixa magnitud d'aquestes corbes hauria provocat que es passessen per alt períodes de creixement/decreixement molt subtils.
- El menjar 2 (clúster 2) és una altre exemple, però en aquest cas el perfil de la corba no és tant diferent al del grup. En canvi la quantitat de períodes sí és molt diferent.

Pel que fa al primer problema, com es pot veure a la Figura 42 a la pàgina 98 (Apèndix B), aquest no es reproduïx en la 14a combinació perquè el mètode per a seleccionar els intervals de  $C$  és manual. Respecte al segon problema, pareix que serà impossible evitar-lo, donat que no hem establert diferents mètodes per unificar el nombre de períodes entre corbes. Com estem veient, el clústering de la 14a combinació ofereix un millor aspecte pel que fa a la coherència dels grups. Una excepció n'és el clúster 2, on s'evidencia que  $M=3$  és insuficient.

En resum, podem extraure les següents conclusions:

- El fet d'obtenir una millor puntuació en les mesures de validació interna no es tradueix en generar particions amb més bon aspecte o millor encaix.
- S'ha d'anar molt en compte amb com s'estableixen els intervals de les variables, especialment la combinació del mètode CUM i pocs intervals pot conduir a resultats inesperats i descripcions de poca qualitat. Són recomanables  $M=5$  i  $C \geq 5$ .
- Cal repensar el mecanisme per fer equivalències entre les corbes per penalitzar més la transformació/partició de períodes.
- La 14a combinació segons DI i Silueta ens sembla el millor clústering generat fins ara, encara que no és l'únic vàlid, com veurem en la següent secció. Els *outliers* del grup 2, que és el més destacat negativament, són una conseqüència directa de la pèrdua d'informació que sempre es produeix en el procés de modelatge qualitatiu. Recordem de nou les opcions de configuració d'aquesta combinació:

- 3 intervals d' $M$ :  $\{[72.6, 194.3], (194.3, 295.7], (295.7, 748.7]\}$
- 5 intervals de  $C$ :  $\{-104.9, -15], (-15, -1], (-1, 1], (1, 15], (15, 516.5]\}$
- 4 intervals de  $D$ :  $\{[5, 26], (26, 57], (57, 96], (96, 265]\}$

	k	M	C	D	C me.	Diss.	$\{w_C, w_M, w_D\}$	VI (gi)	DI	Sil.
1	11	5	5	4	CUM	CN	{0.30, 0.30, 0.40}	1.787	0.278	0.344
2	13	5	5	4	CUM	CN	{0.30, 0.30, 0.40}	1.854	0.32	0.361
3	10	5	5	4	CUM	CN	{0.30, 0.30, 0.40}	1.854	0.278	0.333
4	11	5	5	4	pks	CN	{0.30, 0.30, 0.40}	1.882	0.319	0.387
5	12	5	5	4	CUM	CN	{0.30, 0.30, 0.40}	1.882	0.278	0.353
6	13	5	7	3	CUM	CN	{0.35, 0.35, 0.30}	1.883	0.297	0.412
7	12	5	5	4	CUM	CN	{0.35, 0.35, 0.30}	1.892	0.42	0.346
8	12	5	5	4	CUM	CN	{0.33, 0.33, 0.33}	1.892	0.409	0.348
9	8	5	5	4	CUM	CN	{0.30, 0.30, 0.40}	1.898	0.267	0.327
	...									
19	13	5	5	4	CUM	CN	{0.35, 0.35, 0.30}	1.956	0.467	0.356
20	13	5	5	4	CUM	CN	{0.33, 0.33, 0.33}	1.956	0.429	0.356

Taula 13: Fragment dels 20 millors resultats d'acord amb la similitud amb el clústering de referència del projecte ClosedLoop4Meal.

- Selecció manual dels intervals de C.
- Dissimilitud amb veïnatge conceptual.
- Pes dels paràmetres: {0.25, 0.25, 0.50}

#### 4.1.2 Comparació amb altres clústers «de referència»

*Encara que usem el terme clústering «de referència», no és per a res un «gold standard», si fos així aquest projecte no tindria gaire sentit.*

Per tal de recollir més informació sobre les opcions de configuració, mesurarem la VI<sup>1</sup> entre els clústerings obtinguts i el clústering de 16 grups que acompanyava les dades originals del projecte ClosedLoop4Meal (Figura 17 a la pàgina següent). Pensant que el clústering original conté característiques «desitjables», és lògic considerar valuoses les opcions que ens porten fins particions properes.

Dels resultats (Taula 13) destaquen les opcions: pes {0.30, 0.30, 0.40}, M=5, C=5, D=4, C me= 'CUM' i veïnatge conceptual per a la dissimilitud. Això és compatible amb les opcions destacades en els apartats anteriors (recordem que només havíem pogut assegurar un predomini de C=5 i un pes més elevat per a D).

El clústering de referència aconsegueix que totes les corbes d'un mateix clúster encaixen en un perfil mínim, tanmateix té dues parts negatives. Per un costat té un nombre de grups elevat (13) per a les mostres que hi ha (57 menjars); això generalment és un mal indicatiu. Per un altre costat, no té en compte «el moviment» de les corbes. En optimitzar el perfil, encaixa magnituds i temps similars però pot

<sup>1</sup> Mètrica per mesurar la distància entre 2 particions d'elements.  
[http://en.wikipedia.org/wiki/Variation\\_of\\_information](http://en.wikipedia.org/wiki/Variation_of_information)  
 S'ha usat el mètode `vi.dist` del paquet R `mcclust`.

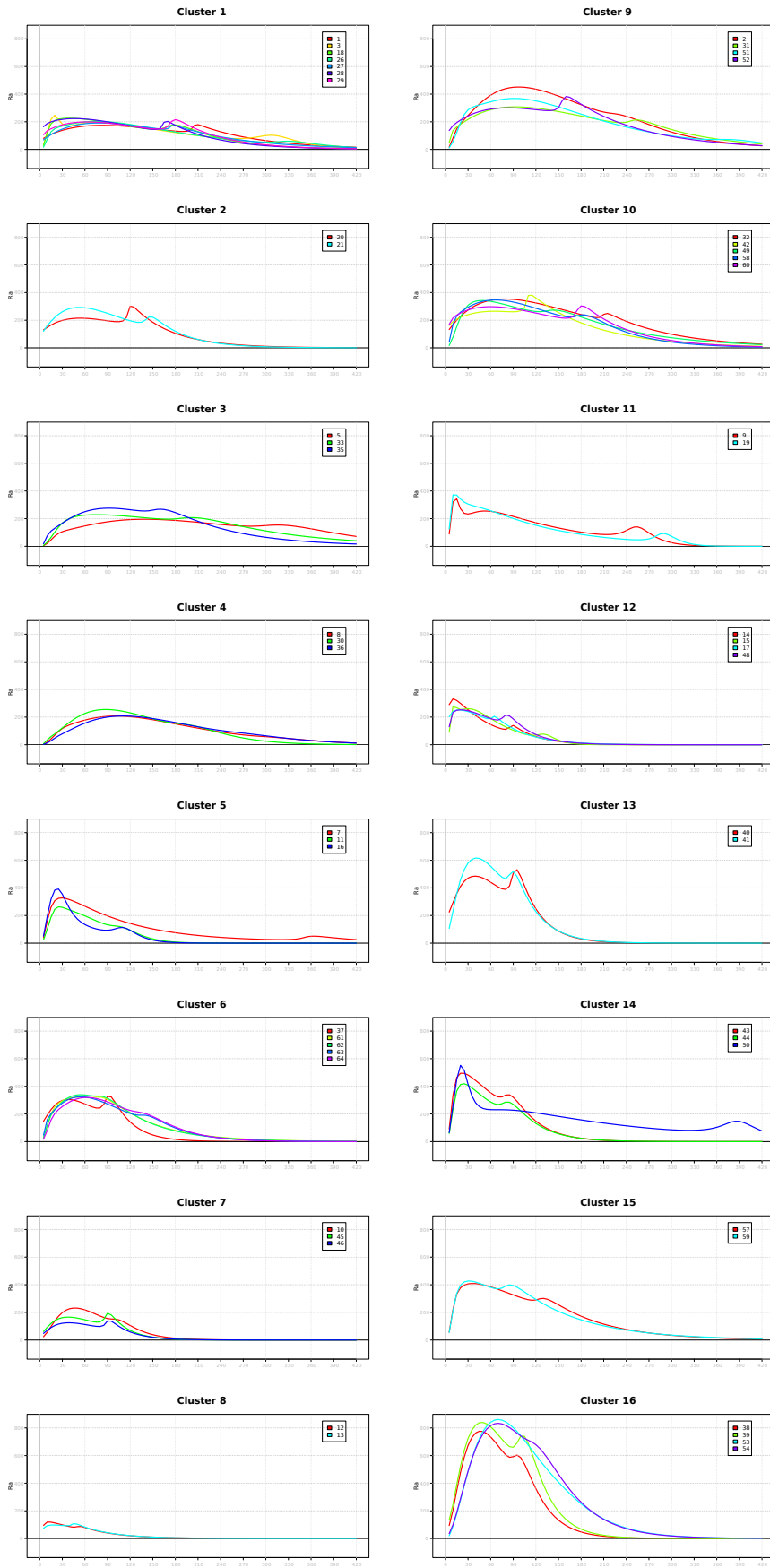


Figura 17: Clústering original del projecte ClosedLoop4Meals. Es pot veure la descripció completa dels grups a les Figures 44 i 45 a la pàgina 101 (Appendix B).

*El clústring de referència només conté 57 dels 60 menjars, per poder mesurar-ne la distància adequadament hem fet les particions amb aquestes mateixes 57.*

passar per alt algunes formes característiques que van més enllà d'aquestes variables més explícites. Per exemple, el menjar 3 del clúster 1 presenta un perfil similar al del grup, però té un pic inicial, una característica molt diferent a la resta del grup.

En lloc de mostrar quin dels nostres resultats s'assembla més al de referència, mostrarem un que s'hi assemble i al mateix temps gaudisca d'unes bones puntuacions de qualitat interna. Podríem elegir qualsevol de les combinacions que hi ha en blau a la Taula 13, però estudiarem la 19a per ser la que té una puntuació DI+Silueta més elevada.

Aquesta combinació presenta uns resultats bons en general (Figura 18 a la pàgina següent), però destaquen negativament el clúster 2 i el 8. Cal observar la descripció qualitativa dels membres (Figura 48 i 49 a la pàgina 105) per esbrinar la raó. En el clúster 2 trobem un fenomen negatiu que segurament es reproduirà arreu perquè rau en la pròpia base del model qualitatiu triat: ha primat el temps total dels períodes i la magnitud, per sobre de la quantitat de períodes i els seus valors del canvi. Això indica que és necessari revisar el model per especificar un comportament predible davant una circumstància així. En el clúster 8 trobem un fenomen que no sabem si serà positiu o negatiu: corbes que tenen una forma molt similar però una escala temporal substancialment diferent. És el cas dels menjars 17 i 48 *versus* 21 o 31. Tal volta hi ha factors comuns que determinen eixa forma i ens interessaria detectar-la. Tal volta no és així, i ens hem de cenyir a trobar perfils de comportament similars pel que fa a la magnitud i els temps.

No podem afirmar que cap dels clústerings trobats siga millor que el de referència, tanmateix sí podem assegurar que la nostra aproximació permet capturar el comportament de les corbes de manera més concreta i coherent. Hem vist que en alguns casos hi ha resultats ambigus (clúster 2) que són conseqüència inherent del raonament qualitatiu, però en altres dubtem de la idoneïtat dels resultats (clúster 8) perquè ens falta informació d'experts. Encara així, considerem que aquest últim clústering analitzat pot ser un bon candidat per extreure una descripció en llenguatge natural, ja que incorpora tots els avantatges: grups propers als de referència i coherents, puntuació alta en les mesures de qualitat interna i generat a partir d'una de les nostres descripcions qualitatives (ens dona accés a la descripció qualitativa). Finalment, recordem quines opcions han generat aquest clústering:

- 5 intervals d'M: {[61.2, 157.4], (157.4, 219.3], (219.3, 274.3], (274.3, 398.1], (398.1, 748.7]}
- 5 intervals de C: {[−58.4, −8.7], (−8.7, −0.008], (−0.008, 0.047], (0.047, 26.6], (26.6, 289.7]}
- 4 intervals de D: {[1, 5.2], (5.2, 11.4], (11.4, 19.2], (19.2, 53]}

*Fem notar que dels 5 intervals de C, el central és nul. Així que a la pràctica només hi ha 4 intervals de C, com es pot comprovar a la descripció a la pàgina 104.*

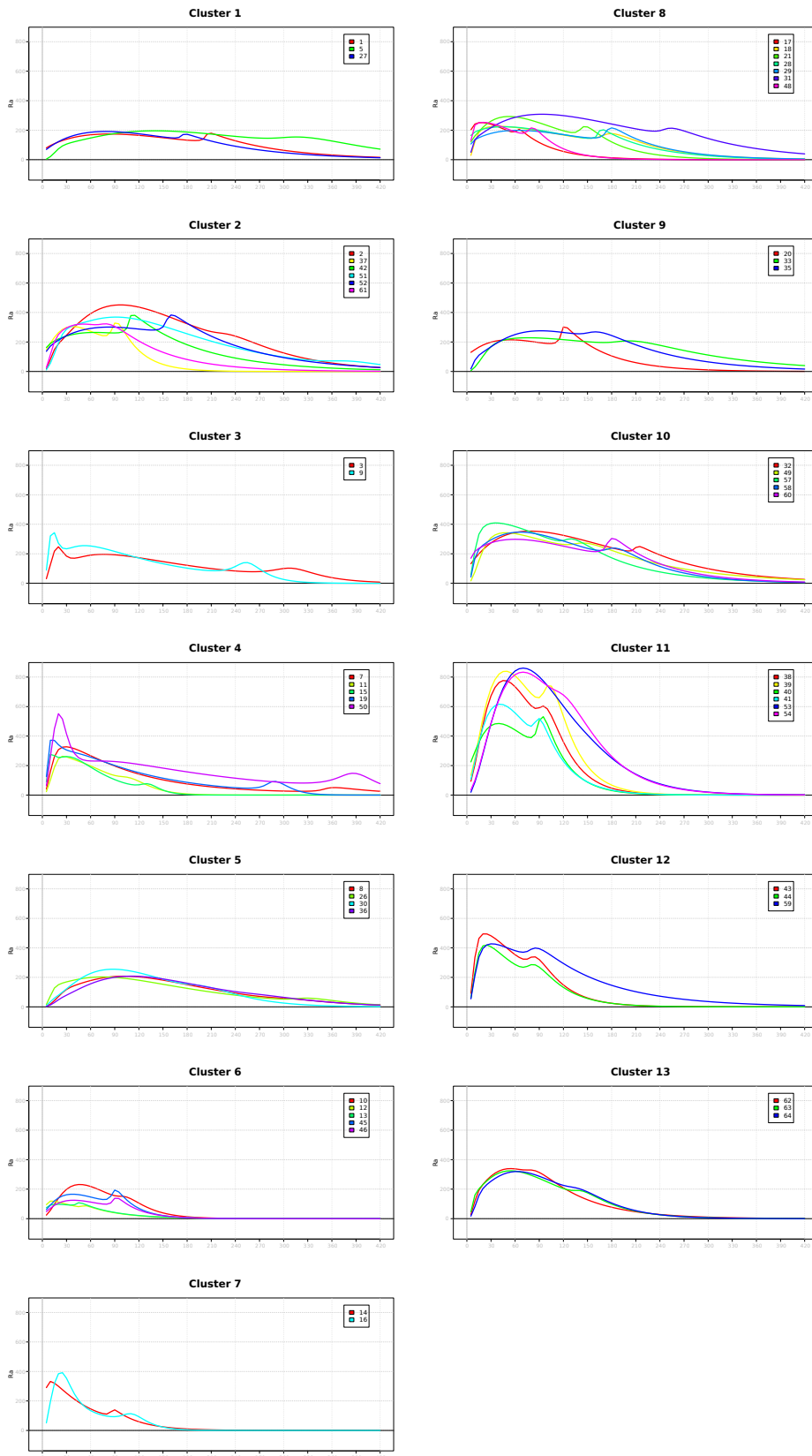


Figura 18: Clústering proper a l'original i de relativa qualitat (fila 19 de la Taula 13 a la pàgina 46). Es pot veure més informació sobre cada grup a les Figures 46 i 47 a la pàgina 103, i la descripció qualitativa de cada corba a les Figures 48 i 49 a la pàgina 105 (Apèndix B).

- Selecció dels intervals de  $C$  amb el mètode CUM.
- Dissimilitud amb veïnatge conceptual.
- Pes dels paràmetres:  $\{0.35, 0.35, 0.30\}$

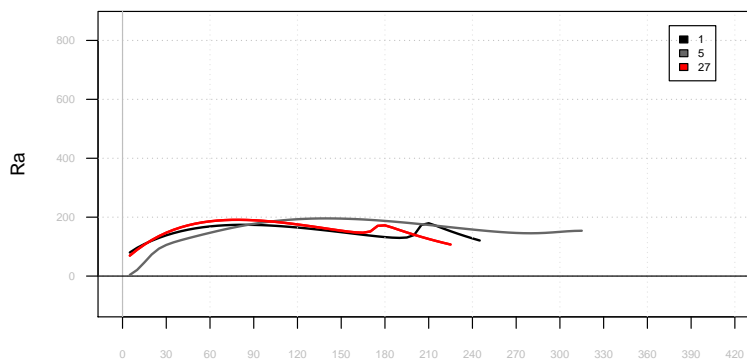
#### 4.1.3 Descripció en llenguatge natural

A continuació descriurem els grups de l'últim clústering que hem vist (Figura 18 a la pàgina anterior). Recordem que la configuració és la hem detallat al final de l'anterior secció (també es pot veure a la fila 19 de la Taula 13 a la pàgina 46).

Per a cada clúster, enumerem els seus membres i el nombre de períodes. Per a tots els grups que tenen més de 2 membres incorporem la moda (valor més freqüent) de totes les etiquetes. A partir de l'observació de la representació gràfica dels valors qualitius de les corbes, i dels estadístics, hem extret una descripció (similar al que es podria programar ràpidament). Per últim es mostra un gràfic amb totes les corbes del clúster, marcant en roig la medoide.

##### CLÚSTER 1

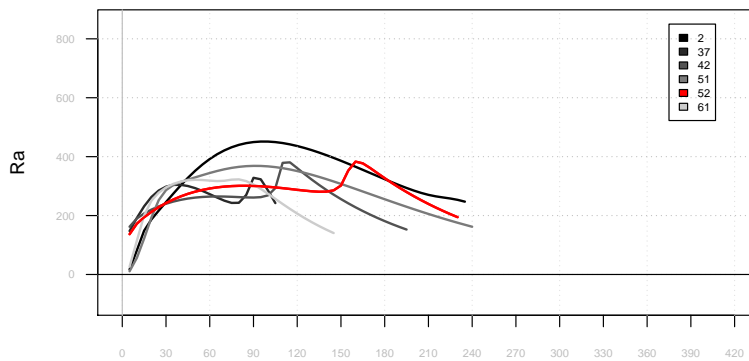
- 3 membres: "1", "5", "27"
- 4 períodes
- Moda: {  
["+", "1", "\*\*\*"],  
["-", "2", "\*\*\*\*\*"],  
["+", "1", "\*"],  
["-", "1", "\*\*"]}
- Descripció: *Comença amb un període llarg on creix suaument fins a magnitud mitjana-inferior, i torna a decreixer amb suavitat durant un període llarguíssim. Cap al final creix breument, i acaba amb un període de durada normal on decreix fins magnitud baixa.*
- Observacions: La corba es caracteritza per la suavitat de les corbes i unes magnituds baixes.





## CLÚSTER 2

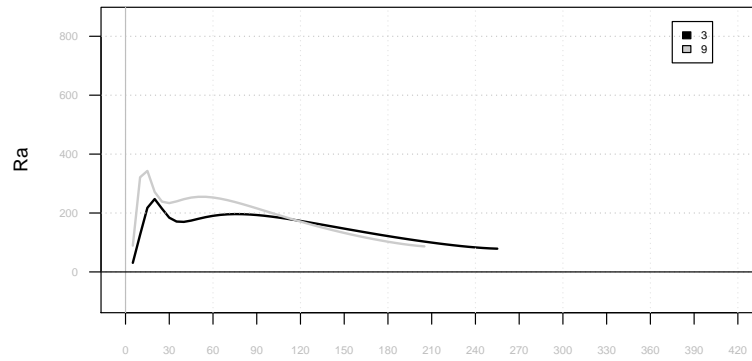
- 6 membres: "2", "37", "42", "51", "52", "61"
- 4 períodes
- Moda: {  
["+", "3", "\*\*\*"],  
["-", "4", "\*"],  
["+", "4", "\*"],  
["--", "4", "\*\*\*"]}
- Descripció: *Comença amb un període llarg on creix suaument fins magnitud alta, i després decreix breument. De meitat endavant hi ha un creixement breu (pic), i finalitza amb un període llarg de decreixement pronunciat.*
- Observacions: Es caracteritza per corbes relativament suaus i unes magnituds altes. Els menjars "2" i "51" no encaixen en la descripció (mireu el clúster 5).



## CLÚSTER 3

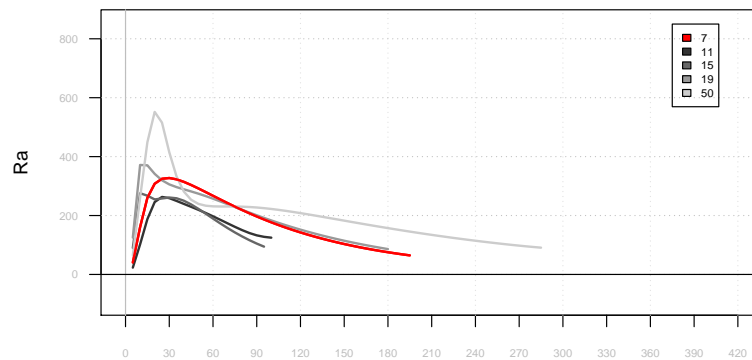
- 2 membres: "3", "9"
- 4 períodes
- Descripció: *Comença amb un període curt de creixement pronunciat fins a magnitud mitjana-superior i un període curt amb decreixement pronunciat fins magnitud mitjana-inferior (pic inicial). Després continua amb un període normal de creixement lleuger, i finalment es dona un període llarguíssim de decreixement fins arribar a magnitud baixa.*

- Observacions: La corba es caracteritza per un pic inicial pronunciat.



#### CLÚSTER 4

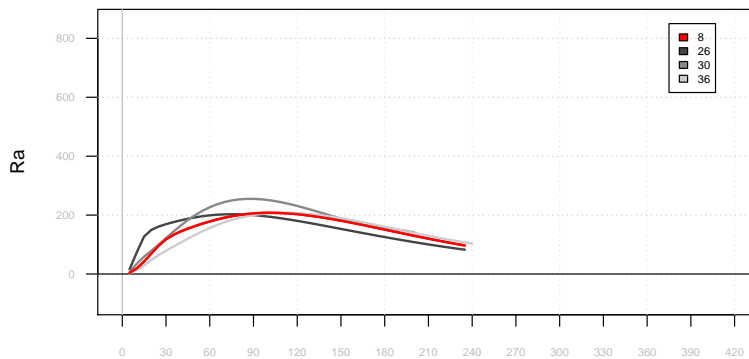
- 5 membres: "7", "11", "15", "19", "50"
- 2 períodes
- Moda: {  
["++", "3", "\*"],  
["-", "2", "\*\*\*\*\*"]}
- Descripció: Comença amb un període curt de creixement pronunciat fins magnitud mitjana-superior (pic inicial), i continua amb un període llarguíssim on decreix fins magnitud mitjana-inferior.
- Observacions: La corba es caracteritza per un pic inicial pronunciat. El menjar "50" té una forma molt semblant al clúster 3 però amb molta més magnitud.



#### CLÚSTER 5

- 4 membres: "8", "26", "30", "36"
- 2 períodes

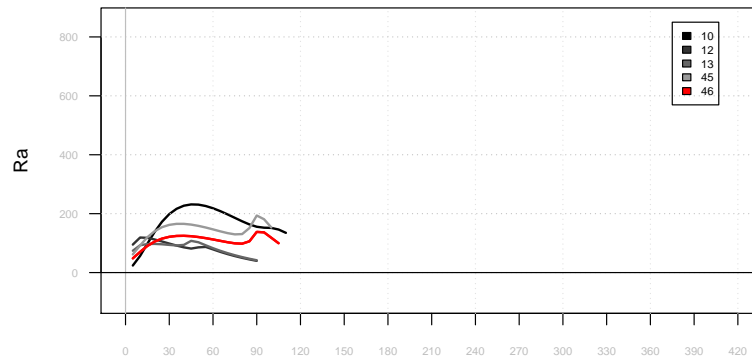
- Moda: {  
["+", "1", "\*\*\*"],  
["-", "2", "\*\*\*\*\*"]}
- Descripció: *Comença amb un període llarg de creixement suau fins magnitud mitjana-inferior, i acaba amb un període llarguíssim on decreix fins magnitud baixa. El pic es produeix a meitat de la corba i és suau.*
- Observacions: Els menjars "2" i "51" del clúster 2 tenen una forma similar a aquest clúster. Però no es deuen haver classificat ací per tenir una magnitud molt més alta.



#### CLÚSTER 6

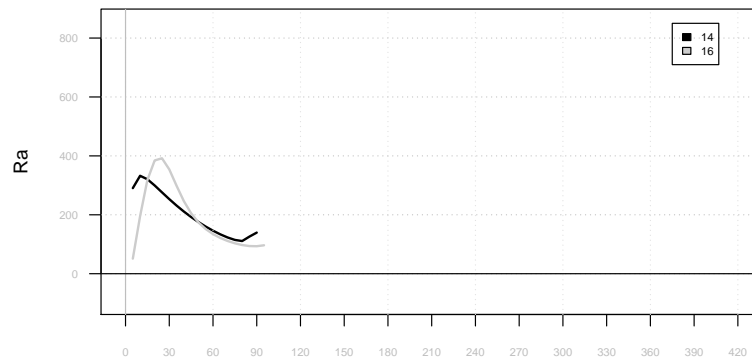
- 5 membres: "10", "12", "13", "45", "46"
- 4 períodes
- Moda: {  
["+", "1", "\*\*\*"],  
["-", "1", "\*\*\*"],  
["+", "1", "\*"],  
["-", "1", "\*"]}
- Descripció: *Comença amb un període normal on creix suaument fins magnitud baixa, i després decreix breument. De meitat endavant hi ha un creixement breu (pic), i finalitza amb un període curt de decreixement.*
- Observacions: El menjar "10" no encaixa, té una forma similar al clúster 5, però en una escala temporal molt més comprimida. Per la resta, la descripció recorda al clúster 2, però en una escala

de magnitud menor.



### CLÚSTER 7

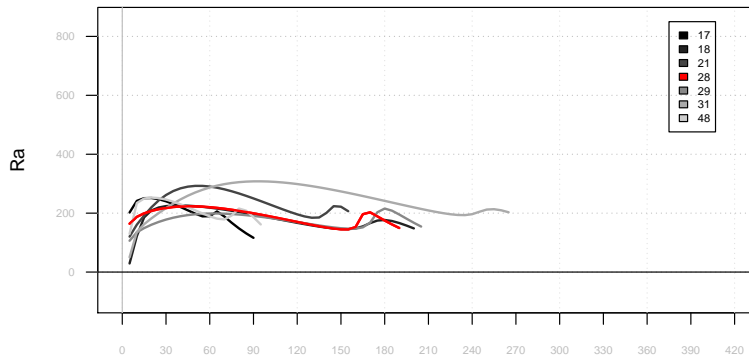
- 2 membres: "14", "16"
- 3 períodes
- Descripció: *Comença amb un període curt de creixement pronunciat fins magnitud alta (pic inicial), i continua amb un període llarg de decreixement pronunciat fins magnitud baixa. Finalment hi ha un repunt curt.*
- Observacions: La corba es caracteritza per un pic inicial pronunciat.



### CLÚSTER 8

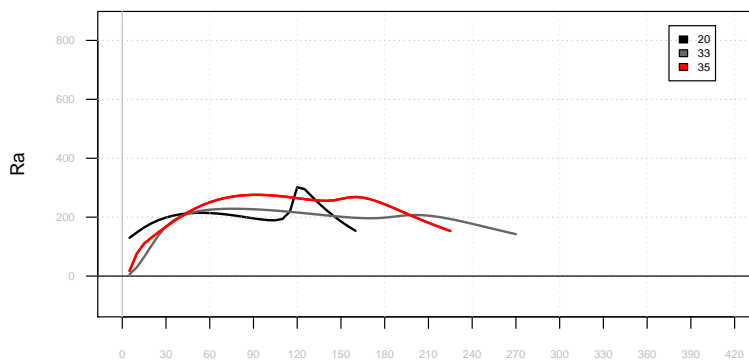
- 7 membres: "17", "18", "21", "28", "29", "31", "48"
- 4 períodes
- Moda: {  
 ["+", "2", "\*\*\*"],  
 ["-", "2", "\*\*\*\*\*"],  
 ["+", "2", "\*"],  
 ["--", "2", "\*"]}

- *Descripció: Comença amb un període normal de creixement suau fins magnitud mitjana-superior, seguit d'un període llarguíssim de decreixement moderat. Després hi ha un període curt de creixement, i finalment un període curt de decreixement pronunciat fins magnitud mitjana-inferior. No es produeixen pics destacats.*
- *Observacions: En aquest cas podem veure que els menjars "17" i "48", tot i tenir una escala temporal més comprimida, s'han classificat ací perquè comparteixen magnitud i canvi amb el grup.*



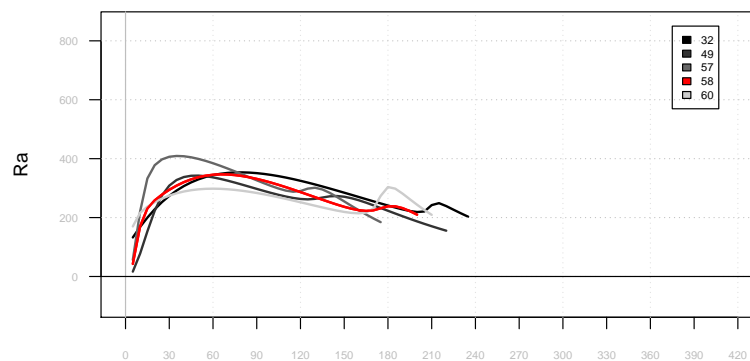
#### CLÚSTER 9

- 3 membres: "20", "33", "35"
- 4 períodes
- Moda: {  
 ["+", "2", "\*\*\*"],  
 ["-", "2", "\*\*\*"],  
 ["+", "3", "\*"],  
 ["-", "2", "\*\*\*"]}
- *Descripció: Comença amb un període llarg de creixement fins magnitud mitjana-inferior, seguit d'un període normal de decreixement moderat. Després hi ha un curt creixement fins magnitud mitjana-superior, i finalment un període llarg de decreixement moderat.*



## CLÚSTER 10

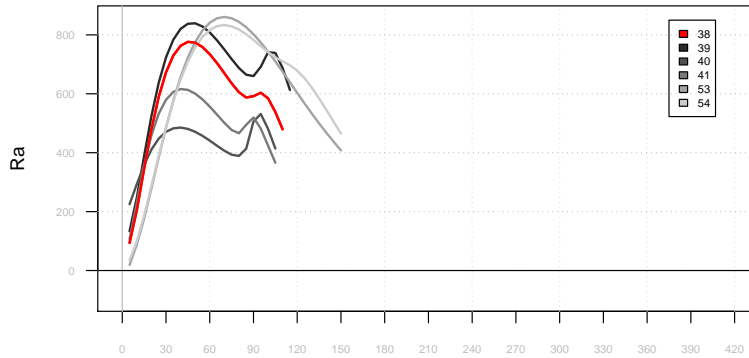
- 5 membres: "32", "49", "57", "58", "60"
- 4 períodes
- Moda: {  
["++", "4", "\*\*"],  
["-", "4", "\*\*\*\*\*"],  
["+", "3", "\*"],  
["--", "3", "\*"]}
- Descripció: *Comença amb un període normal de creixement pronunciat fins a magnitud alta (pic prompte). Continua amb un període llarguíssim de decreixement lleuger. Un xicotet repunt precedeix un període final de duració normal i decreixement pronunciat fins magnitud mitjana-superior.*
- Observacions: Es caracteritza per un pujada inicial relativament pronunciada i destacable manteniment de la magnitud durant un llarg període de decreixement lleuger.



## CLÚSTER 11

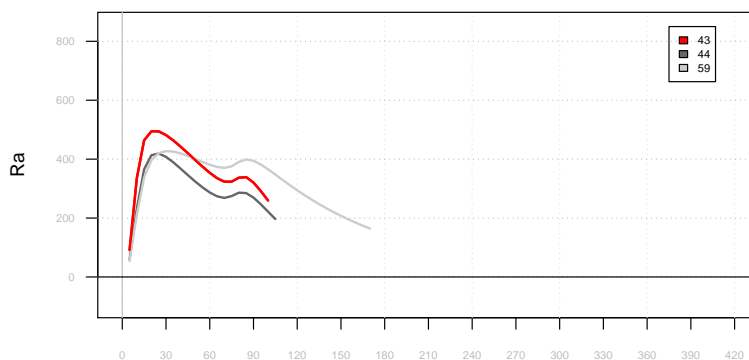
- 6 membres: "38", "39", "40", "41", "53", "54"
- 4 períodes
- Moda: {  
["++", "5", "\*\*"],  
["--", "5", "\*\*"],  
["+", "5", "\*"],  
["--", "5", "\*"]}
- Descripció: *Comença amb un període normal de creixement pronunciat fins a magnitud gegant. Segueix un període normal de decreixement pronunciat. Finalment un xicotet repunt precedeix un període final curt de decreixement pronunciat.*

- Observacions: La forma podria dir-se que es semblant a la dels clústers 2 i 6, però amb una magnitud moltíssim més gran.



### CLÚSTER 12

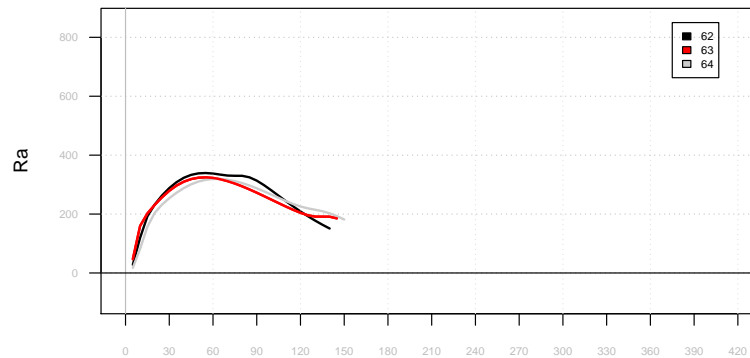
- 3 membres: "43", "44", "59"
- 4 períodes
- Moda: {  
["++", "4", "\*"],  
["--", "5", "\*\*"],  
["+", "4", "\*"],  
["--", "4", "\*"]}
- Descripció: Comença amb un període curt de creixement pronunciat fins magnitud alta (pic). Segueix un període normal de decreixement pronunciat. Finalment un xicotet repunt precedeix un període final curt de decreixement pronunciat des de magnitud alta.
- Observacions: Té bastant similitud amb el clúster 11, però amb menys magnitud.



### CLÚSTER 13

- 3 membres: "62", "63", "64"

- 2 períodes
- Moda: {  
["++", "3", "\*\*\*"],  
["-", "3", "\*\*\*"]}
- Descripció: *Comença amb un període normal de creixement pronunciat fins magnitud mitjana-superior. Finalitza amb un període llarg de decreixement des de magnitud mitjana-superior.*
- Observacions: Per què el menjar "10" no està ací? Sembla degut a una qüestió de magnituds.



## 4.2 DISCUSSIÓ

Recordem que durant la fase d'experimentació s'han generat diversos clústerings que hem considerat bons. S'han estudiat el 1r i 14è de la Taula 12 a la pàgina 42, però també eren bons el 10è i 15è. Després hem buscat quins clústering eren més propers al de referència, i d'entre ells també hem assenyalat candidats (files 4, 6, 7, 8 i 19 de la Taula 13 a la pàgina 46). Finalment ens hem decidit a analitzar amb més detall els grups del 19è de la Taula 13, per comptar amb tots els avantatges que estàvem buscant: grups similars al de referència, bones puntuacions de qualitat interna i descripció qualitativa adequada. Els següents comentaris es refereixen a aqueix clústering.

Com demostra el cas del clúster 2, la forma d'unificar el nombre d'interval·ls entre corbes pot fer que hi hagen corbes classificades incorrectament. La transformació dels menjars 2 i 51 (de 2 períodes a 4), no s'hauria de produir o en tot cas s'hauria de penalitzar perquè desvirtua la descripció qualitativa. Per això pensem que és necessari replantejar-se la manera de trobar la distància entre corbes amb diferent nombre de períodes.

Degut també als menjars 2 i 51, ens sorgeix un altra pregunta: volem que corbes amb la mateixa «forma», però d'una magnitud molt diferent, es classifiquen juntes? De ser així, p. ex. incloent-les al clúster 5, també hauríem de replantejar el model per a què no pena-



	Període 1			Període 2			Període 3			Període 4		
	C	M	T	C	M	T	C	M	T	C	M	T
Clúster 1	+	1	***	-	2	****	+	1	*	-	1	**
Clúster 2	+	3	***	-	4	*	+	4	*	--	4	***
Clúster 4	++	3	*	-	2	****						
Clúster 5	+	1	***	-	2	****						
Clúster 6	+	1	**	-	1	**	+	1	*	-	1	*
Clúster 8	+	2	**	-	2	****	+	2	*	--	2	*
Clúster 9	+	2	***	-	2	**	+	3	*	-	2	***
Clúster 10	++	4	**	-	4	****	+	3	*	--	3	*
Clúster 11	++	5	**	--	5	**	+	5	*	--	5	*
Clúster 12	++	4	*	--	5	**	+	4	*	--	4	*
Clúster 13	++	3	**	-	3	***						

Taula 14: Taula de modes del clústering analitzat.

litze la diferència en magnitud. En un cas extrem, fins i tot es podria eliminar el paràmetre.

Quan observem en conjunt les descripcions obtingudes (Taula 14), veiem una constant continua: creix-decreix creix-decreix. Inclús els clústers 4, 5 i 13, de només 2 períodes, tenen aquest segon repunt. Tanmateix, o bé es massa tímid per a ser detectat com a un pic, o bé cau fora del 80% d'AUC i no es té en compte. A totes llums aquest sembla un fet important, que s'hauria d'explotar en reconsiderar el disseny del model. Per exemple, pressuposant l'existència d'aquestes dues muntanyes, la discretització es podria centrar en mesurar les seues característiques.

Pel que fa a la composició dels menjars (Figures 46 i 47 a la pàgina 103, Apèndix B), el resultats obtinguts no aporten cap informació nova. Com ja s'observava en el clústering de referència, i es confirma en els nostres, els menjars amb poc percentatge de grassa, acostumen a iniciar-se de manera pronunciada (clúster 4, 12). Inversament, un alt percentatge de grassa «aixafa» la corba, de manera que es perllonga l'absorció de la glucosa (clúster 5, 9). Això no es compleix sempre, per tant ha d'haver altres factors que hi influeixin.

Un desavantatge molt gran de les dades de que disposem, és que la qualitat pel que fa a la informació nutricional és molt baixa. No sabem si això es deu a la dificultat d'estimar la composició de menjars mesclats, a una mala qualitat en el disseny de les proves per obtenir la informació nutricional, o a un descuit. Per un costat hi ha molts valors en blanc, dels quals no podem assumir un valor zero. Vege's p. ex. el cas dels menjars 28i 29 (pasta amb salsa de tomata), que per força han de tenir algun percentatge de grassa i proteïna. Per un al-

*En les taules estadístiques que mostren el % de macronutrients, hem omès els menjars que no tenien un valor numèric vàlid per a CHO, Grasa i Proteïna.*

tre costat, només comptar amb una «panoràmica» dels ingredients no és suficient. Els CHO és una categoria d'aliments que engloba sucres simples (fructosa, galactosa), sucres dobles (lactosa, maltosa, sacaro-sa) i carbohidrats «complexes» (abundants en midó). Sabem que la velocitat d'absorció és molt diferent depenent del tipus de CHO [5]. Si aquesta situació també es extrapolable a altres macronutrients (p. ex. hi ha diferent resposta amb grassa satura o insaturada), aleshores gairebé és segur que no podrem veure cap correlació entre el comportament de les corbes i la informació nutricional que ara tenim.

Part III

CONCLUSIONS I TREBALL FUTUR



## CONCLUSIONS I TREBALL FUTUR

---

### 5.1 CONCLUSIONS

En aquest treball hem fet un repàs de la **DM**, una malaltia que es caracteritza per la presència d'alts nivells de glucosa a la sang. Els tractaments clàssics han estat i continuen sent fonamentals per a mantenir la salut dels pacients amb **DM-1**, però són incòmodes, invasius i difícils de gestionar per als malalts. S'ha vist que el pàncrees artificial (**AP**) és una solució tecnològica que pretén simular la funció endocrina del pàncrees, però encara és necessari millorar les seues capacitats de monitorització i predicció de **BG**. El present treball s'ha centrat en proporcionar una descripció qualitativa dels models d'absorció de glucosa després de les menjades. D'aquesta manera pretenem augmentar la comprensió de l'efecte glicèmic d'un menjar mesclat, que posteriorment podrà ser incorporat a nous esquemes de control anticipatiu de l'**AP**.

S'ha implementat un script R configurable i reutilitzable, que ha servit per generar tots els resultats mostrats en aquesta memòria. El programa és una part important del treball perquè en ell s'acumulen la majoria d'hores realitzades. A més a més, pot servir per reproduir els resultats presentats i de punt de partida per provar noves combinacions, paràmetres, intervals, etc.

S'ha vist una panoràmica del procés de modelització qualitativa, així com dels diferents algorismes de clústering i de les mesures de validació que podem usar per verificar la qualitat dels resultats. Hem fet un estudi inicial de les característiques rellevants que caracteritzen una corba d'aparició de glucosa postprandial i definit unes quantes opcions per a etiquetar aquestes característiques, generant diverses possibilitats per a un model qualitatiu. Posteriorment s'ha fet un descart d'opcions mitjançant tècniques de validació interna de clústering. Aquest procés ha donat com a resultat una teoria per a descriure una corba de glucosa.

Els clústerings obtinguts s'han comparat entre sí i amb el clústering de referència, per veure les diferències i determinar-ne el millor. Mentrestant, hem analitzat les deficiències del nostre model i suggerit millores allà on es podia. On no, ens hem limitat a fer les preguntes que ens portarien cap a altres models més sòlids.

Finalment hem estudiat amb més detall el clústering que hem considerat millor (Figura 18 a la pàgina 49). Les opcions que s'han usat per configurar aquest clústering han estat: 5 intervals de magnitud (*M*); 5 intervals de canvi (*C*), 4 a la pràctica; 4 intervals de duració

(*D*); selecció dels intervals de *C* amb el mètode CUM; dissimilitud amb veïnatge conceptual; pes {0.35, 0.35, 0.3}. Les característiques d'aquest clústering són:

- Bona similitud amb el clústering de referència (distància VI: 1.956), amb grups coherents.
- Bones puntuacions de validació interna (DI: 0.467; Silueta: 0.356).
- Una descripció qualitativa acceptable, generada amb les opcions que s'acaben de comentar a dalt.

Hem vist que la possibilitat de construir una descripció en llenguatge natural de la corba, es fàcilment realitzable. Es pot enriquir més enllà d'una simple enumeració de les etiquetes perquè podem deduir coses a partir de la seqüència temporal.

Considerem que el més important està en la construcció d'un model qualitatiu que realment capture les característiques rellevants. Per ara la descripció qualitativa és «paral·lela» a la quantitativa i es limita bàsicament a «traduir» els valors en etiquetes, però encara així ja té avantatges davant la quantitativa, com la incorporació de semàntica en la descripció. Si es treballés més estretament amb experts es podrien resoldre molts dels dubtes que ens han sorgit durant el desenvolupament, i prendre decisions més concretes respecte al model, centrant-nos en aconseguir un millor encaix en la ràtio d'aspecte, en el perfil o en la forma/pics més enllà de la magnitud, etc.

Sabem que la clau per a aconseguir una bona descripció, està en la comprensió de les corbes, comprensió que en certa mesura s'ha eixamplat amb el present treball. A base de fer descripcions qualitatives roïnes, anem observant errades en el model i comprenent millor en quines característiques ens hem de fixar. Hem deixat obertes diverses preguntes, i amb les respostes necessàries es podria seguir avançant ràpidament. Esperem que els resultats presentats siguin suficientment engrescadors com per a encoratjar en el projecte ClosedLoop4Meal la possibilitat de continuar amb l'anàlisi qualitatiu de les dades. Ens sembla que aquest pot ajudar al descobriment i codificació de coneixements tàcits, adquirits amb molts anys d'experiència mèdica.

## 5.2 TREBALL FUTUR

### 5.2.1 *Treball futur immediat*

Per poder seguir avançant en la línia d'aquest treball, seria necessari aplicar les següents mesures immediatament.

### 5.2.1.1 *Canvis en el model*

- Com ja s'ha comentat durant la discussió, és necessari repensar la manera d'unificar corbes que tenen diferents nombre de períodes per a què no es violen propietats que considerem fonamentals en la descripció.
- En lloc de fer les divisions dels períodes basant-nos en els pics/-valls, en alguns casos sembla que seria més convenient fixar-nos en altres coses, com ara bé el ritme de canvi o si s'ha arribat a una magnitud relativa concreta. Per exemple, fixem-nos en el menjar 51: té un creixement fort fins el minut 30, moment a partir del qual es produeix una «meseta» i el pic no arriba fins el minut 90. Ara mateix tot això queda obviat en dos períodes creix-decreix, quan en realitat té un aspecte més similar a creix-estable-decreix.
- Degut a que el 90% de les corbes tenen dos períodes principals creix-decreix creix-decreix, resultaria molt interessant investigar si en considerar tota la corba (sense descartar a partir del 80% d'AUC) podem capturar millor les reaccions similars.

### 5.2.1.2 *Augment de les ferramentes*

- Provar amb altres mètodes de clústering, i sobretot amb altres mètodes de validació interna, en concret els d'estabilitat. D'aquesta manera podríem detectar amb més fiabilitat els esbiaixaments de les combinacions model-algorisme-opcions-validació.

### 5.2.1.3 *Augment de la informació*

- És necessari tenir més informació pel que fa als menjars. A ser possible tenir més mostres, però sobretot seria ideal que incorporaren informació nutricional desglosada.
- És fonamental establir una via de comunicació directa amb experts mèdics per poder decidir quan anem en la bona direcció.

## 5.2.2 *Treball futur en segon terme*

Si s'avancés en els aspectes anteriors, almenys en els aspectes fonamentals que ens portaren a millorar les descripcions qualitatives, aleshores tindria sentit fer un programa que generés les descripcions en llenguatge natural. Després s'haurien de validar aquestes descripcions, presentant-les a un expert per veure si es capaç de deduir-ne la forma de la corba o d'identificar el menjar que produeix la reacció descrita.

En cas de que descobrirem alguna correlació entre la informació nutricional i la forma de les corbes, seria desitjable fer un arbre de decisió per veure'n la consistència.



Part IV

APÈNDIX



## APROXIMACIONS ALTRENAITIVES

---

Aquest apèndix recull mètodes alternatius per caracteritzar les corbes. Després d'haver invertit una gran quantitat d'hores fent aquestes proves, ens ha semblat adequat incloure una breu panoràmica de les idees i resultats que hi ha darrere de cadascun d'ells.

### A.1 APROXIMACIÓ ESTADÍSTICA

La idea darrere d'aquesta aproximació és caracteritzar les corbes amb la menor quantitat d'informació possible, per produir un clústering similar al de referència. Es a dir, que capture «la forma» absoluta de les corbes, de manera que el perfil de les corbes siga similar.

#### A.1.1 *Model*

Utilitzem tres valors per a definir la corba:

1.  $t$ : El percentatge que representa la *duració efectiva* de la corba respecte del temps observat (420 min). Caracteritza la temporalitat de les corbes de forma absoluta de manera que corbes llargues i curtes mai acabaran juntes. Els valors possibles van de 0 a 1.
2.  $sd+$ : La desviació típica de la corba (només per als valors dintre de la *duració total* de la corba). Caracteritza la variació de les corbes de manera absoluta, es a dir que en certa manera també captura la magnitud de la corba i evitarà que corbes que de magnituds molt diferents acabin juntes. S'ha normalitzat per deixar-la en el rang  $[0, 1]$ .
3.  $CV$ : El Coeficient de Variació de Pearson ( $CV$ ) de la corba (només per als valors dintre de la *duració efectiva* de la corba). Mesura la variació de la corba «respecte a si mateixa», es a dir que és un valor relatiu que evitarà que corbes suaus es classifiquen amb corbes espadades. Aquesta mesura no s'ha normalitzat, però els seus valors poques vegades superen 1 i quan ho fan és per una bona raó, així que hem preferit que tingués «més pes» si es dona eixa circumstància.

### A.1.2 Mètode

En aquesta ocasió sí hem pogut fer servir el paquet R `clValid` per provar amb diferents algorismes de clústering, perquè totes les corbes tenen el mateix nombre d'atributs i es pot generar una matriu  $60 \times 3$ . La millor puntuació l'hem obtinguda amb Partitioning Around Medoids (PAM) i  $k = 11$ .

*No hem especificat res, així que `clValid` fa servir la distància euclídia entre els atributs.*

### A.1.3 Resultats

Els resultats de qualitat interna són una Silueta de 0.38, un DI de 0.25 i una Connectivitat de 59.3. Paral·lelament, la VI amb el clústering de referència és de 1.88.

Vegeu el clústering resultant a les Figures 19 i 20 a la pàgina 72, i el detall sobre els tres valors característics de cada corba a les Figures 21 i 22 a la pàgina 74. La validació és prou bona, de la Connectivitat ja hem vist que no podem fer gaire cas. La similitud amb el clústering de referència és major que la del clústering elegit per a la descripció final (Secció 4.1.3).

Entre els punts forts d'aquesta aproximació hi ha: (1) la rapidesa i senzillesa, i (2) els bons resultats, on fins i tot observem afinitats en els percentatges de grassa dels grups generats. La part negativa és que no té res a veure amb una aproximació qualitativa, i seria molt difícil generar una descripció en llenguatge natural més enllà de: corba curta-llarga, molta-poca magnitud i molta-poca variació. Encara així, ens pot servir de punt de partida per dissenyar un model qualitatiu nou inspirat en aquests paràmetres. Per exemple, basant-nos en el fet que el 99% de les corbes tenen dos pics o muntanyes –és una de les observacions que hem conclòs en el treball–, podem establir que hi haurà 2 períodes per a totes les corbes i descriure'ls amb paràmetres qualitatius mesclats (els de l'estudi, i els d'aquesta aproximació). Així obtindríem un model amb tots els avantatges, ja que a més a més evitariem la problemàtica de la unificació entre corbes amb diferent nombre de períodes.

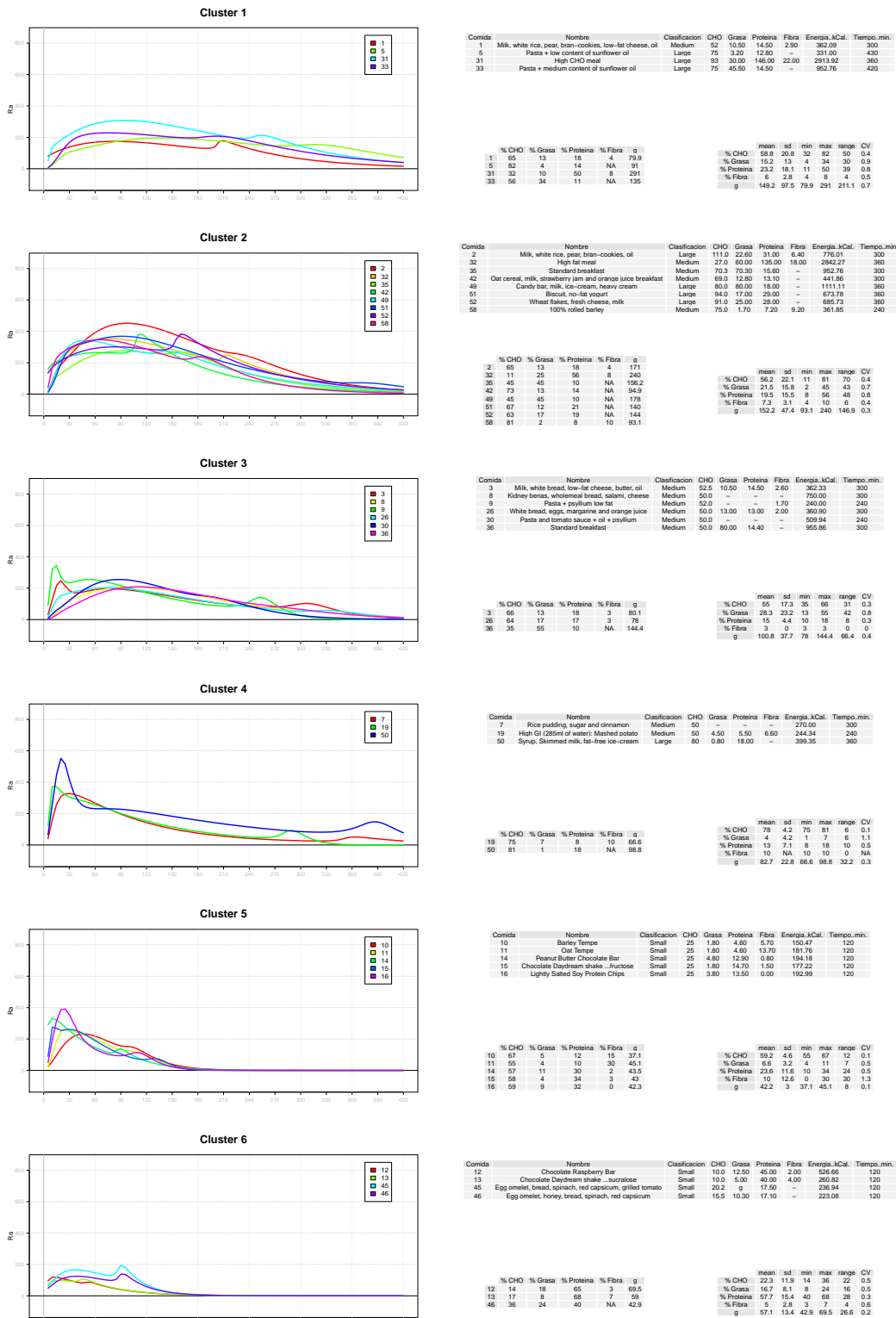
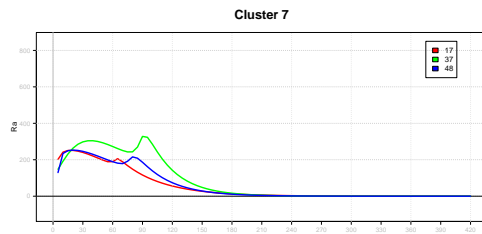
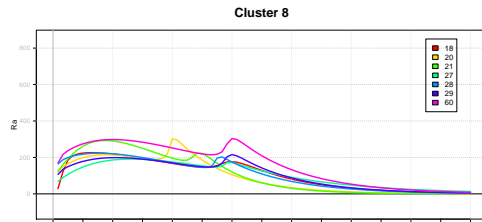


Figura 19: Clústering aconseguit amb l'aproximació estadística (part 1).



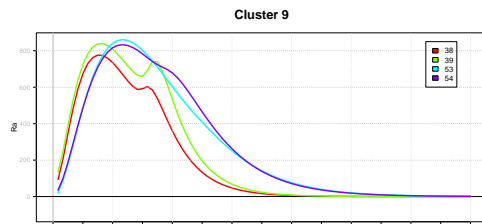
Comida	Nombre	Clasificación	CHO	Grasa	Proteína	Fibra	Energía_kCal	Tiempo_min
17	Soy Spaghetti	Small	25.0	1.10	10.60	0.80	151.91	120
37	Cheese omelet, bread and margarine	Small	38.0	24.00	24.00	-	466.94	120
48	Muffin, peanut butter	Small	27.3	15.40	4.00	-	265.12	120

% CHO	% Grasa	% Proteína	% Fibra	g	mean	sd	min	max	range	CV
17	67	3	28	2	37.5					
37	44	28	28	NA	86					
48	58	33	10	NA	47.2					
					66.3	11.8	44	87	23	0.2
					21.3	16.1	3	33	30	0.8
					22	10.4	10	28	18	0.5
					2	NA	2	0	NA	
					56.9	25.7	37.5	86	48.5	0.5



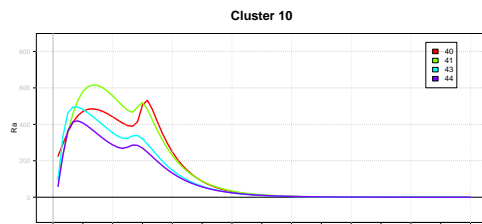
Comida	Nombre	Clasificación	CHO	Grasa	Proteína	Fibra	Energía_kCal	Tiempo_min
18	Low GI (250ml) of water; Boiled pearl barley	Medium	50	1.30	8.20	5.10	243.15	240
20	Baked potato, gelatin, turkey breast	Medium	45	18.00	15.00	-	-	180
21	Boiled rice, corn, turkey breast	Medium	50	18.00	15.00	-	-	180
27	Powdered nutritional supplement	Medium	50	12.00	14.00	0.00	366.87	300
28	Pasta and tomato sauce	Medium	50	-	-	-	240.04	240
29	Pasta and tomato sauce + oil	Medium	50	-	-	-	509.94	240
60	30% rolled barley	Medium	75	1.10	5.40	3.20	348.72	240

% CHO	% Grasa	% Proteína	% Fibra	g	mean	sd	min	max	range	CV
18	76	2	14	8	65.6					
20	58	23	19	NA	78					
21	60	22	18	NA	83					
27	66	16	18	0	76					
60	89	1	6	4	84.7					
					69.8	12.8	58	89	31	0.2
					12.8	10.7	1	23	22	0.8
					15	5.4	6	19	13	0.4
					4	4	0	8	8	1
					77.5	7.5	65.6	84.7	19.1	0.1



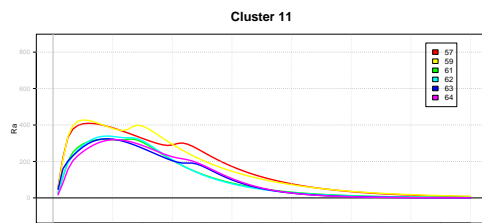
Comida	Nombre	Clasificación	CHO	Grasa	Proteína	Fibra	Energía_kCal	Tiempo_min
38	Spaghetti with tomato, cheese and lentils	Fast	67	9.00	25.00	-	527.85	120
39	Cornflakes, milk, bread and margarine	Fast	104	13.00	18.00	-	695.06	120
53	Wheat-biscuit, low-fat milk, whole-meal bread, fruit, poultry, vegetable	Large	123	29.00	17.00	-	673.78	180
54	Wheat-biscuit, low-fat milk, whole-meal bread, fruit, poultry, vegetable	Large	126	28.00	25.00	-	685.73	180

% CHO	% Grasa	% Proteína	% Fibra	g	mean	sd	min	max	range	CV
38	72	7	21	NA	121					
39	77	10	13	NA	135					
53	73	17	10	NA	169					
54	70	16	14	NA	179					
					73	9.9	70	77	7	0
					12.5	4.9	7	17	10	0.4
					14.5	4.7	10	21	11	0.3
					NA	NA	NA	NA	NA	NA
					151	27.5	121	179	58	0.2



Comida	Nombre	Clasificación	CHO	Grasa	Proteína	Fibra	Energía_kCal	Tiempo_min
40	Oatmeal, milk, bread and margarine	Fast	62	17.00	12.00	-	452.61	120
41	Barley with tomato and cheese	Fast	68	8.00	12.00	-	384.10	120
43	Fiber and fruit cereal, milk, rockmelon, pineapple	Fast	50	2.60	10.20	-	265.12	120
44	High fiber cereal, milk, strawberries, grapefruit	Fast	42	2.50	11.00	-	234.07	120

% CHO	% Grasa	% Proteína	% Fibra	g	mean	sd	min	max	range	CV
40	68	19	13	NA	91					
41	77	9	14	NA	88					
43	80	4	16	NA	62.8					
44	76	5	20	NA	55.5					
					75.2	5.1	68	80	12	0.1
					9.2	6.8	4	19	15	0.7
					15.8	3.1	13	20	7	0.2
					NA	NA	NA	NA	NA	NA
					74.3	17.8	55.5	91	35.5	0.2



Comida	Nombre	Clasificación	CHO	Grasa	Proteína	Fibra	Energía_kCal	Tiempo_min
57	White Rice	Medium	75	0.90	4.60	0.50	342.74	240
59	50% rolled barley	Medium	75	1.30	5.90	4.90	362.78	240
61	White Wheat Bread	Medium	50	3.00	8.40	3.10	266.79	180
62	Whole Kernel Rye Bread	Medium	50	2.60	7.40	12.80	258.91	180
63	B-glucan Rye Bread	Medium	50	2.40	10.50	17.10	270.85	180
64	Whole Meal Pasta	Medium	50	4.70	12.10	5.60	298.66	180

% CHO	% Grasa	% Proteína	% Fibra	g	mean	sd	min	max	range	CV
57	83	1	6	1	81					
59	86	1	7	6	87.1					
61	78	5	13	5	64.5					
62	69	4	10	18	72.8					
63	62	3	13	21	80					
64	69	6	17	8	72.4					
					76.2	11.7	62	93	31	0.2
					3.3	2.1	1	6	5	0.6
					11	4.1	6	17	11	0.4
					9.8	7.9	1	21	20	0.8
					76.3	8	64.5	87.1	22.6	0.1

Figura 20: Clústering aconseguit amb l'aproximació estadística (part 2).

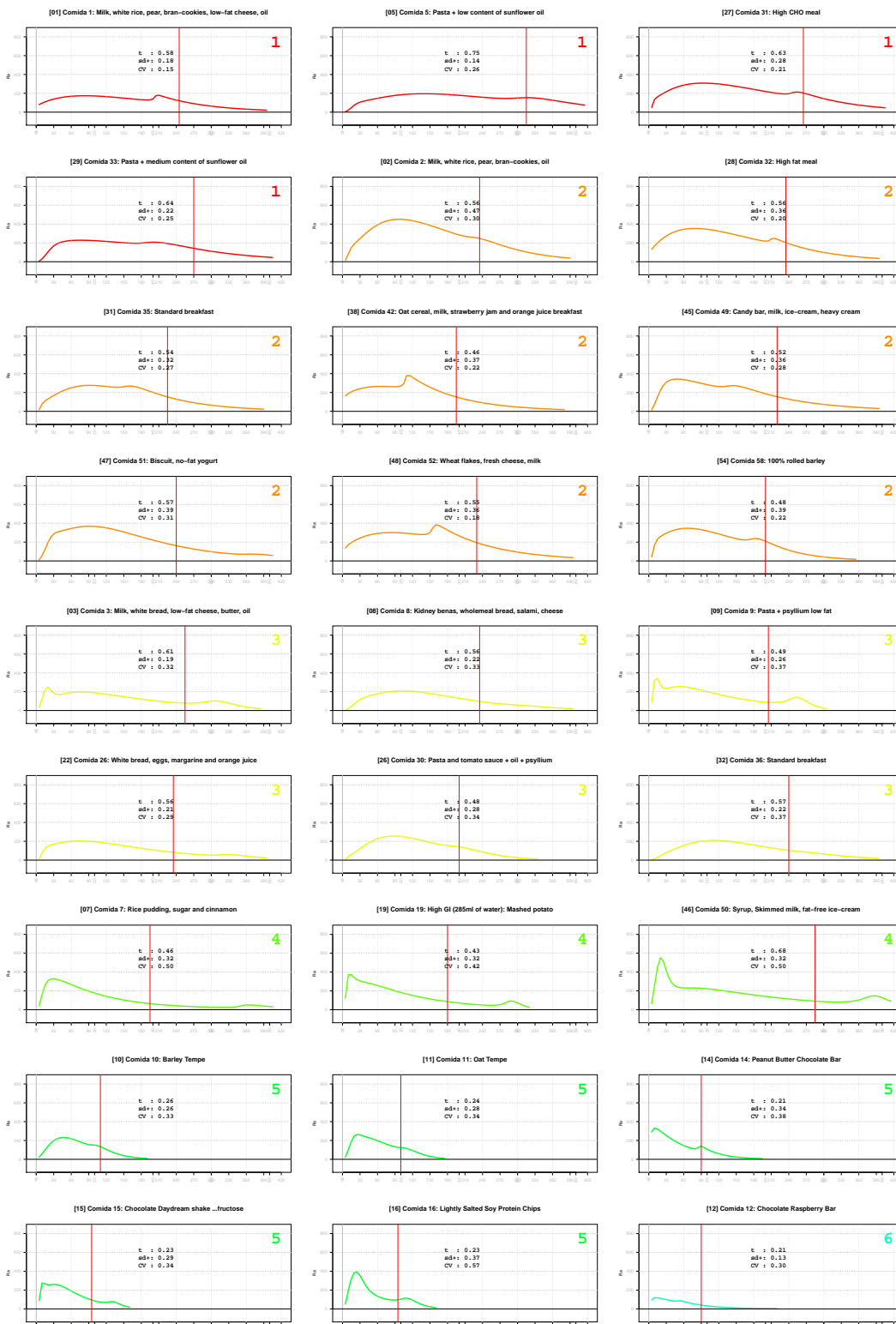


Figura 21: Descripció «estadística» de cada corba del clústering (part 1).

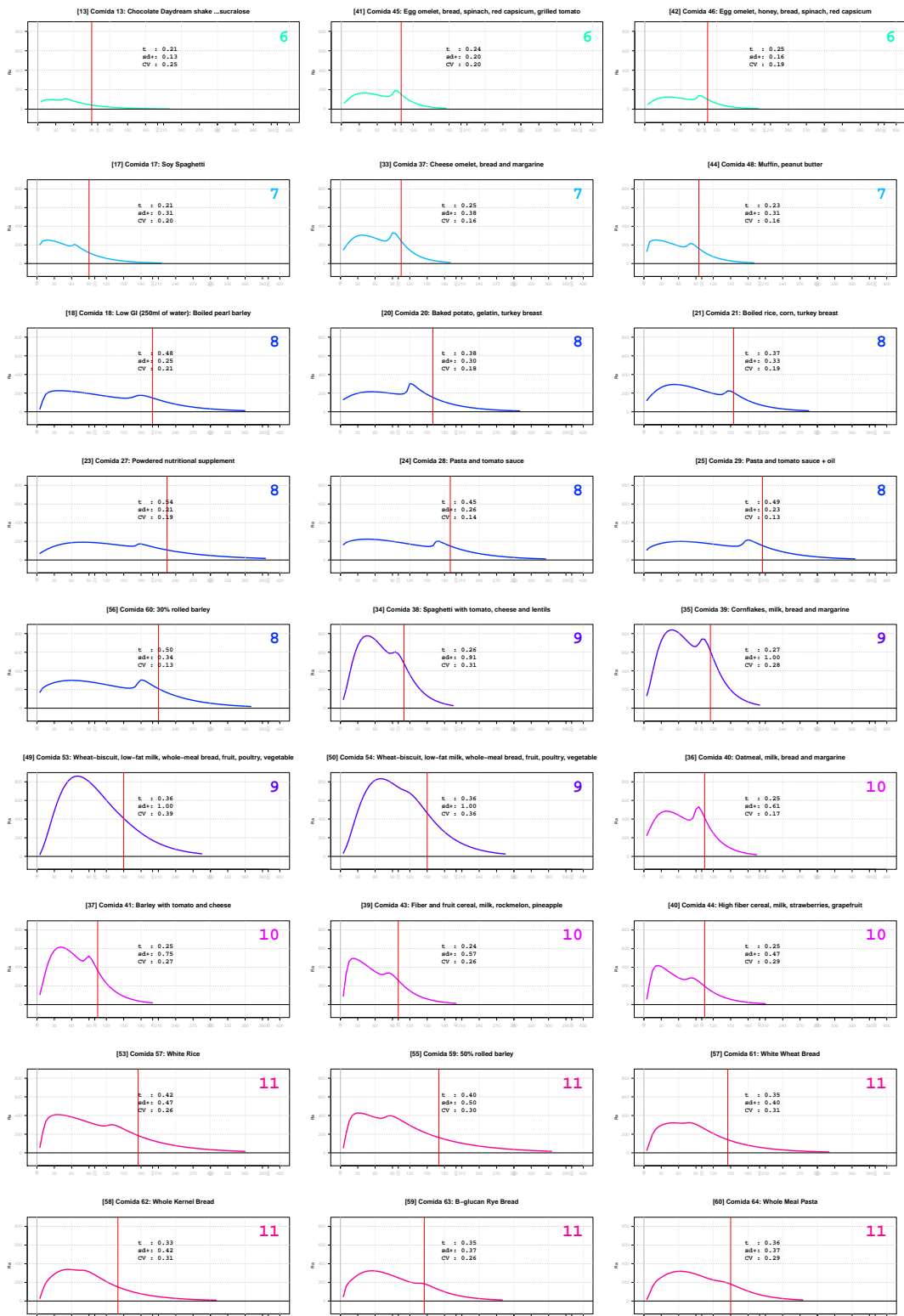


Figura 22: Descripció «estadística» de cada corba del clústering (part 2).



## A.2 APROXIMACIÓ RELATIVA

Quan dues corbes amb «forma» molt similar, queden separades per tenir magnituds molt diferents, inevitablement apareix la pregunta: la diferència en la magnitud és únicament resultat d'una diferència en la *quantitat* (g) de menjar? Per donar-hi resposta, amb aquesta aproximació hem intentant capturar característiques de la corba més enllà de les seues proporcions absolutes. Creiem que si es pot demostrar una relació de forma entre el mateix menjar subministrat en quantitats molt diferents, aquesta seria una informació rellevant a tenir en compte a l'hora de redissenyar el model qualitatiu.

## A.2.1 Model

Tenint en compte únicament el segment  $Ra$  dintre de la *duració efectiva* de la corba. Definim *middle* com el moment temporal en el que s'assoleix el 50% d'**AUC** del segment, i *peak* com el moment temporal en que s'assoleix la màxima magnitud de  $Ra$ . Aleshores:

$$peakRelevance = \frac{\max(Ra) - \text{mean}(Ra)}{\text{mean}(Ra)}$$

$$w_{max} = \begin{cases} peakRelevance & peakRelevance < 1 \\ 1 & peakRelevance > 1 \end{cases}$$

$$w_{middle} = 1 - w_{max}$$

$$massCenter = middle \cdot w_{middle} + peak \cdot w_{max}$$

1. Time: El percentatge que representa la *duració efectiva* de la corba respecte de la *duració total*. És una mesura relativa a la propia corba; els valors possibles van de 0 a 1.
2. Peak time: El percentatge que representa el *massCenter* respecte de la *duració efectiva*. És una mesura relativa a la propia corba; els valors possibles van de 0 a 1.
3. PA: L'«abruptesa» de la corba fins al moment «pic». Sobre l'àrea d'un suposat rectangle, on la base és el temps *middle* i l'altura la màxima magnitud observada fins a aquest temps, mesurem quanta d'aquesta àrea està ocupada realment per l'**AUC**. Formalment, definint  $F(x)$  com la funció que relaciona  $Ra$  i temps, i  $Ra'$  com el segment de valors  $Ra$  fins el moment *middle*, aleshores:

$$PA = 1 - \frac{\int_1^{middle} F(x)}{\max(Ra') \cdot middle}$$

Algorísme	k	Connectivity	DI	Silhouette
kmeans	12	65.001	0.262	0.273
kmeans	6	44.996	0.177	0.263

Taula 15: Validació dels resultats amb l'aproximació relativa.

4. *ratio*: L'aspecte de la corba, definit com la ràtio entre la *duració efectiva* i la mitjana de magnitud (influenciada pel «pic»):

$$ratio = \frac{duracioEfectiva}{mean(Ra) \cdot w_{max} + max(Ra) \cdot w_{middle}}$$

#### A.2.2 Mètode

Al igual que en l'aproximació anterior, hem pogut fer servir el paquet `R clValid` per provar amb diferents algorismes de clústering, perquè totes les corbes tenen el mateix nombre d'atributs i es pot generar una matriu  $60 \times 4$ . Encara que la millor puntuació l'hem obtinguda amb *k*-means i  $k = 12$  (Figura 23 i 24 a la pàgina 79; explicació d'atributs a les Figura 25 i 26 a la pàgina 81), considerem rellevant mostrar els resultats de *k*-means amb  $k = 6$  per ser la primera vegada que obtenim un resultat amb pocs grups i amb un cert sentit.

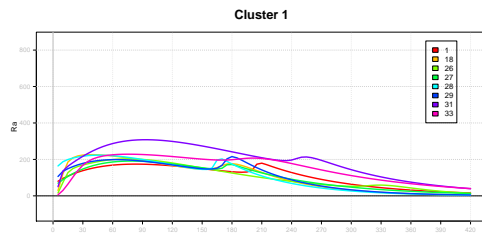
#### A.2.3 Resultats

Els resultats de qualitat interna es poden veure a la Taula 15. La VI amb el clústering de referència és de 2.36, hi ha més distància que amb qualsevol altre resultat, però això no es preocupant perquè realment estem buscant que emergisquen un altre tipus de grups.

Fixant-nos en el clústering de 6 grups (Figura 27 a la pàgina 82), a excepció del clúster 2, veiem que la resta s'han agrupat segons la forma a pesar de tenir proporcions de temps i magnitud absolutes molt diferents (p. ex. clústers 5 i 1). En el clúster 2 es veuen dos tipus de formes clarament diferenciades, el problema ha estat que l'atribut *ratio* no ha tingut prou «pes» per separar el grup, perquè els altres tres atributs són exactament iguals (vegeu el detall del grup 2 a la Figura 28 a la pàgina 83). Més enllà d'aquesta situació, en la resta de clústers, l'homogeneïtat en les proporcions de macronutrients dels grups (especialment clústers 1, 5 i 6) és un fet altament remarcable. Pensem que això demostra la validesa en essència d'aquesta aproximació, almenys com una via alternativa a explorar.

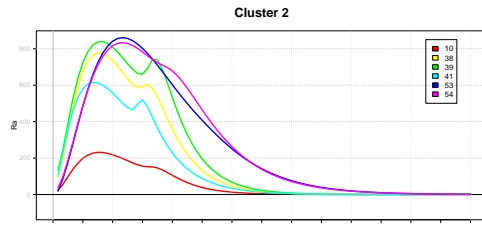
Pel que fa a la possibilitat de generar una descripció en llenguatge natural, no podria ser tan rica com en l'aproximació qualitativa, però sí molt més rica que l'aproximació estadística. Es podria descriure la duració relativa de la corba, quin és el punt relatiu més rellevant

pel que fa a l'aparició de glucosa, com d'abrupte és l'inici de la corba i l'aspecte apaïsat-vertical. Com les característiques tenen un fort component semàntic, la «qualitativització» d'aquesta aproximació seria molt directa, simplement generant intervals/etiquetes per a les mateixes.



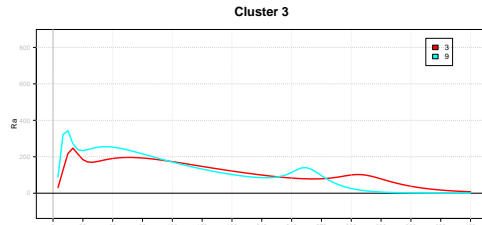
Comida	Nombre	Clasificación	CHO	Grasa	Proteína	Fibra	Energía_kCal	Tiempo_min
1	Milk, white rice, pasta, bran-cookies, low-fat cheese, oil	Medium	52	10.50	14.50	2.90	362.09	300
18	Low GI (250ml of water): Boiled pearl barley	Medium	50	1.30	8.20	5.10	243.15	240
26	White bread, eggs, margarine and orange juice	Medium	50	13.00	13.00	2.00	360.90	300
27	Powdered nutritional supplement	Medium	50	12.00	14.00	0.00	366.87	300
28	Pasta and tomato sauce	Medium	50	-	-	-	240.04	240
29	Pasta and tomato sauce + oil	Medium	50	-	-	-	509.84	240
31	High CHO meal	Large	93	30.00	146.00	22.00	2913.92	360
33	Pasta + medium content of sunflower oil	Large	75	45.50	14.50	-	952.78	420

% CHO	% Grasa	% Proteína	% Fibra	g	mean	sd	min	max	range	CV	
1	65	13	18	4	79.9						
18	76	2	14	8	65.6						
26	64	17	17	3	76						
27	66	16	18	0	76						
31	32	10	50	8	291						
33	56	34	11	NA	135						
					mean	sd	min	max	range	CV	
					% CHO	59.8	15.1	32	76	44	0.3
					% Grasa	15.3	10.6	2	34	32	0.7
					% Proteína	21.3	14.3	11	50	39	0.7
					% Fibra	4.6	3.4	0	8	8	0.7
					g	120.9	86.9	65.6	291	225.4	0.7



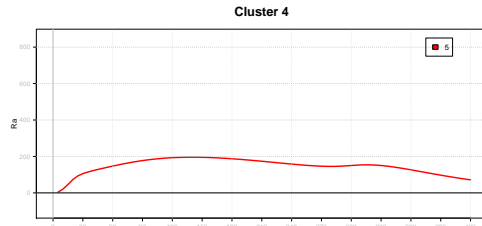
Comida	Nombre	Clasificación	CHO	Grasa	Proteína	Fibra	Energía_kCal	Tiempo_min
10	Barley Tempe	Small	25	1.90	4.60	5.70	150.47	120
38	Spaghetti with tomato, cheese and lentils	Fast	87	8.00	25.00	-	527.85	120
39	Confitade, milk, bread and margarine	Fast	104	13.00	18.00	-	609.06	120
41	Barley with tomato and cheese	Fast	68	8.00	12.00	-	394.10	120
53	Wheat-biscuit, low-fat milk, whole-meal bread, fruit, poultry, vegetable	Large	123	29.00	17.00	-	673.78	180
54	Wheat-biscuit, low-fat milk, whole-meal bread, fruit, poultry, vegetable	Large	126	28.00	20.00	-	685.73	180

% CHO	% Grasa	% Proteína	% Fibra	g	mean	sd	min	max	range	CV	
10	67	5	12	15	37.1						
38	72	7	21	NA	121						
39	77	10	13	NA	135						
41	77	9	14	NA	88						
53	73	17	10	NA	169						
54	70	16	14	NA	179						
					mean	sd	min	max	range	CV	
					% CHO	72.7	3.9	67	77	10	0.1
					% Grasa	10.7	4.8	5	17	12	0.5
					% Proteína	14	3.7	10	21	11	0.3
					% Fibra	15	NA	15	15	0	NA
					g	121.5	62.9	37.1	179	141.9	0.4



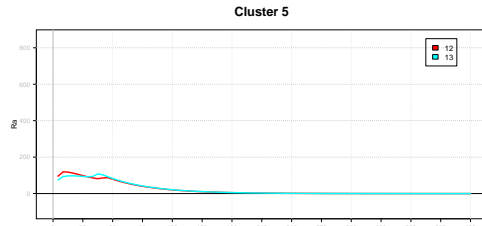
Comida	Nombre	Clasificación	CHO	Grasa	Proteína	Fibra	Energía_kCal	Tiempo_min
3	Milk, white bread, low-fat cheese, butter, oil	Medium	52.5	10.50	14.50	2.60	362.33	300
9	Pasta + psyllium low fat	Medium	52.0	-	-	1.70	240.00	240

% CHO	% Grasa	% Proteína	% Fibra	g	mean	sd	min	max	range	CV	
3	66	13	18	3	80.1						
9	66	NA	66	66	0	NA					
					mean	sd	min	max	range	CV	
					% CHO	13	NA	13	13	0	NA
					% Grasa	18	NA	18	18	0	NA
					% Proteína	14	3.7	10	21	11	0.3
					% Fibra	3	NA	3	3	0	NA
					g	80.1	NA	80.1	80.1	0	NA



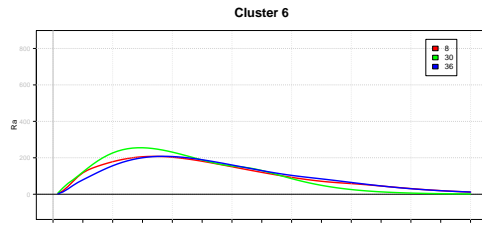
Comida	Nombre	Clasificación	CHO	Grasa	Proteína	Fibra	Energía_kCal	Tiempo_min
5	Pasta + low content of sunflower oil	Large	75	3.20	12.80	-	531.00	420

% CHO	% Grasa	% Proteína	% Fibra	g	mean	sd	min	max	range	CV	
5	82	4	14	NA	91						
					mean	sd	min	max	range	CV	
					% CHO	15.5	2.1	14	17	3	0.1
					% Grasa	13	7.1	8	18	10	0.5
					% Proteína	66.5	2.1	65	68	3	0
					% Fibra	5	2.8	3	7	4	0.8
					g	64.2	7.4	59	69.5	10.5	0.1



Comida	Nombre	Clasificación	CHO	Grasa	Proteína	Fibra	Energía_kCal	Tiempo_min
12	Chocolate Raspberry Bar	Small	10	12.50	45.00	2.00	526.66	120
13	Chocolate Daydream shake ...sucralose	Small	10	5.00	40.00	4.00	260.82	120

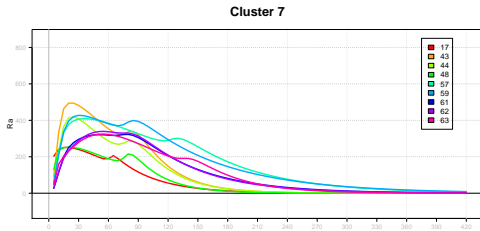
% CHO	% Grasa	% Proteína	% Fibra	g	mean	sd	min	max	range	CV	
12	14	18	65	3	69.9						
13	17	8	68	7	59						
					mean	sd	min	max	range	CV	
					% CHO	15.5	2.1	14	17	3	0.1
					% Grasa	13	7.1	8	18	10	0.5
					% Proteína	66.5	2.1	65	68	3	0
					% Fibra	5	2.8	3	7	4	0.8
					g	64.2	7.4	59	69.5	10.5	0.1



Comida	Nombre	Clasificación	CHO	Grasa	Proteína	Fibra	Energía_kCal	Tiempo_min
8	Kidney beans, wholemeal bread, satami, cheese	Medium	50	-	-	-	750.00	300
30	Pasta and tomato sauce + oil + psyllium	Medium	50	-	-	-	509.84	240
36	Standard breakfast	Medium	50	80.00	14.40	-	955.86	300

% CHO	% Grasa	% Proteína	% Fibra	g	mean	sd	min	max	range	CV	
36	35	55	10	NA	144.4						
					mean	sd	min	max	range	CV	
					% CHO	35	NA	35	35	0	NA
					% Grasa	55	NA	55	55	0	NA
					% Proteína	10	NA	10	10	0	NA
					% Fibra	NaN	NA	inf	inf	inf	NA
					g	144.4	NA	144.4	144.4	0	NA

Figura 23: Clústering aconseguit amb l'aproximació relativa (part 1).

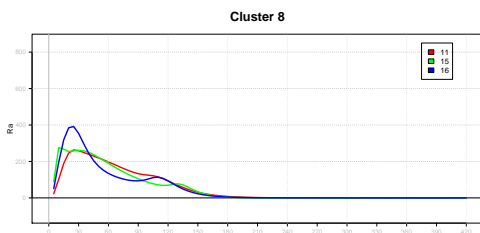


Comida	Nombre	Clasificación	CHO	Grasa	Proteína	Fibra	Energia_kCal	Tiempo_min
17	Soy Spaghetti	Small	25.0	1.10	10.60	0.80	151.91	120
43	Fiber and fruit cereal, milk, rockmelon, pineapple	Fast	50.0	2.60	10.20	-	265.12	120
44	High fiber cereal, milk, strawberries, quinoa/ut	Fast	42.0	2.50	11.00	-	234.07	120
48	Muffin, peanut butter	Small	27.3	15.40	4.50	-	265.12	120
57	White Rice	Medium	75.0	0.90	4.60	0.50	342.74	240
59	50% rolled barley	Medium	75.0	1.30	5.90	4.90	352.78	240
61	White Whole Bread	Medium	50.0	3.80	8.40	3.10	298.79	180
62	Whole Kernel Bread	Medium	50.0	2.60	7.40	12.80	258.91	180
63	B-glucan Rye Bread	Medium	50.0	2.40	10.50	17.10	270.85	180

% CHO	% Grasa	% Proteína	% Fibra	g	mean	sd	min	max	range	CV
17	67	3	28	2	37.5					
43	80	4	16	NA	62.8					
44	75	5	20	NA	55.5					
48	58	33	10	NA	47.2					
57	58	1	6	1	81					
59	85	1	7	6	87.1					
61	78	5	13	5	64.5					
62	69	4	10	18	72.8					
63	62	3	13	21	80					

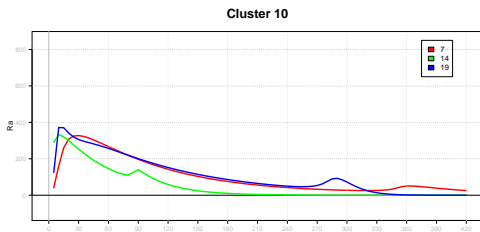
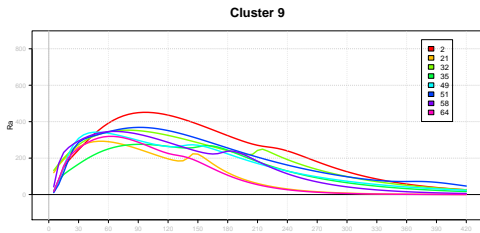
Comida	Nombre	Clasificación	CHO	Grasa	Proteína	Fibra	Energia_kCal	Tiempo_min
11	Oat Tempe	Small	25	1.80	4.60	13.70	181.76	120
15	Chocolate Daydream shake _fluctose	Small	25	1.80	14.70	1.50	177.22	120
16	Lightly Salted Soy Protein Chips	Small	25	3.80	13.50	0.00	150.89	120

% CHO	% Grasa	% Proteína	% Fibra	g	mean	sd	min	max	range	CV
11	59	4	10	30	45.1					
15	58	4	34	3	43					
16	59	9	32	0	42.3					



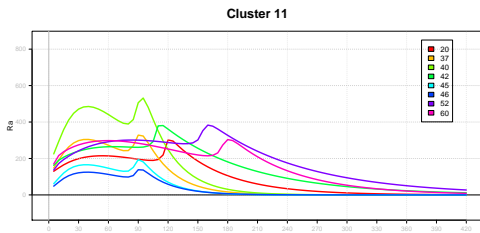
Comida	Nombre	Clasificación	CHO	Grasa	Proteína	Fibra	Energia_kCal	Tiempo_min
2	Milk, white rice, pear, bran-cookies, oil	Large	111.0	22.60	31.00	6.40	776.01	300
21	Boiled rice, corn, turkey breast	Medium	50.0	18.00	15.00	-	180	
32	High fat meal	Medium	27.0	60.00	135.00	18.00	2842.27	360
35	Standard breakfast	Medium	70.3	70.30	15.60	-	952.76	300
49	Candy bar, milk, ice-cream, heavy cream	Large	60.0	80.00	18.00	-	1111.11	360
51	Biscuit, no-fat yogurt	Large	94.0	17.00	29.00	-	673.78	360
58	100% rolled barley	Medium	75.0	1.70	7.20	9.20	351.85	240
64	Whole Meal Pasta	Medium	50.0	4.70	12.10	5.60	298.56	180

% CHO	% Grasa	% Proteína	% Fibra	g	mean	sd	min	max	range	CV
2	65	13	18	4	171					
21	60	22	18	NA	83					
32	11	25	56	8	240					
35	45	45	10	NA	156.2					
49	45	45	10	NA	178					
51	67	12	21	NA	140					
58	81	2	8	10	93.1					
64	69	6	17	8	72.4					



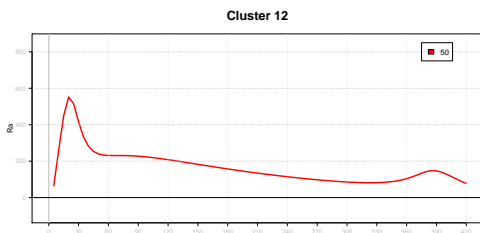
Comida	Nombre	Clasificación	CHO	Grasa	Proteína	Fibra	Energia_kCal	Tiempo_min
7	Rice pudding, sugar and cinnamon	Medium	50	-	-	-	270.00	300
14	Peanut Butter Chocolate Bar	Small	25	4.80	0.80	194.18	120	
19	High GI (285ml of water): Mashed potato	Medium	50	4.50	5.50	6.60	244.34	240

% CHO	% Grasa	% Proteína	% Fibra	g	mean	sd	min	max	range	CV
14	57	11	30	2	43.5					
19	75	7	8	10	66.6					



Comida	Nombre	Clasificación	CHO	Grasa	Proteína	Fibra	Energia_kCal	Tiempo_min
20	Baked potato, potato, turkey breast	Medium	45.0	18.00	15.00	-	180	
37	Cheese omelet, bread and margarine	Small	38.0	24.00	24.00	-	466.94	120
40	Oatmeal, milk, bread and margarine	Fast	62.0	17.00	12.00	-	452.81	120
42	Oat cereal, milk, strawberries (and orange juice) breakfast	Medium	69.0	12.80	13.10	-	441.96	300
45	Egg omelet, bread, spinach, red capsicum, grilled tomato	Small	20.2	9	17.50	-	236.94	120
46	Egg omelet, honey, bread, spinach, red capsicum	Small	15.5	10.20	17.10	-	223.08	120
52	Wheat flakes, fresh cheese, milk	Large	91.0	25.00	28.00	-	685.73	360
60	30% rolled barley	Medium	75.0	1.10	5.40	3.20	348.72	240

% CHO	% Grasa	% Proteína	% Fibra	g	mean	sd	min	max	range	CV
20	58	23	19	NA	78					
37	44	28	28	NA	86					
40	68	19	13	NA	91					
42	73	13	14	NA	94.9					
45	35	24	40	NA	42.9					
52	63	17	19	NA	144					
60	89	1	6	4	84.7					



Comida	Nombre	Clasificación	CHO	Grasa	Proteína	Fibra	Energia_kCal	Tiempo_min
50	Syrup, Skimmed milk, fat-free ice-cream	Large	80	0.80	18.00	-	399.35	360

% CHO	% Grasa	% Proteína	% Fibra	g	mean	sd	min	max	range	CV
50	81	1	18	NA	98.8					

Figura 24: Clústering aconseguit amb l'aproximació relativa (part 2).

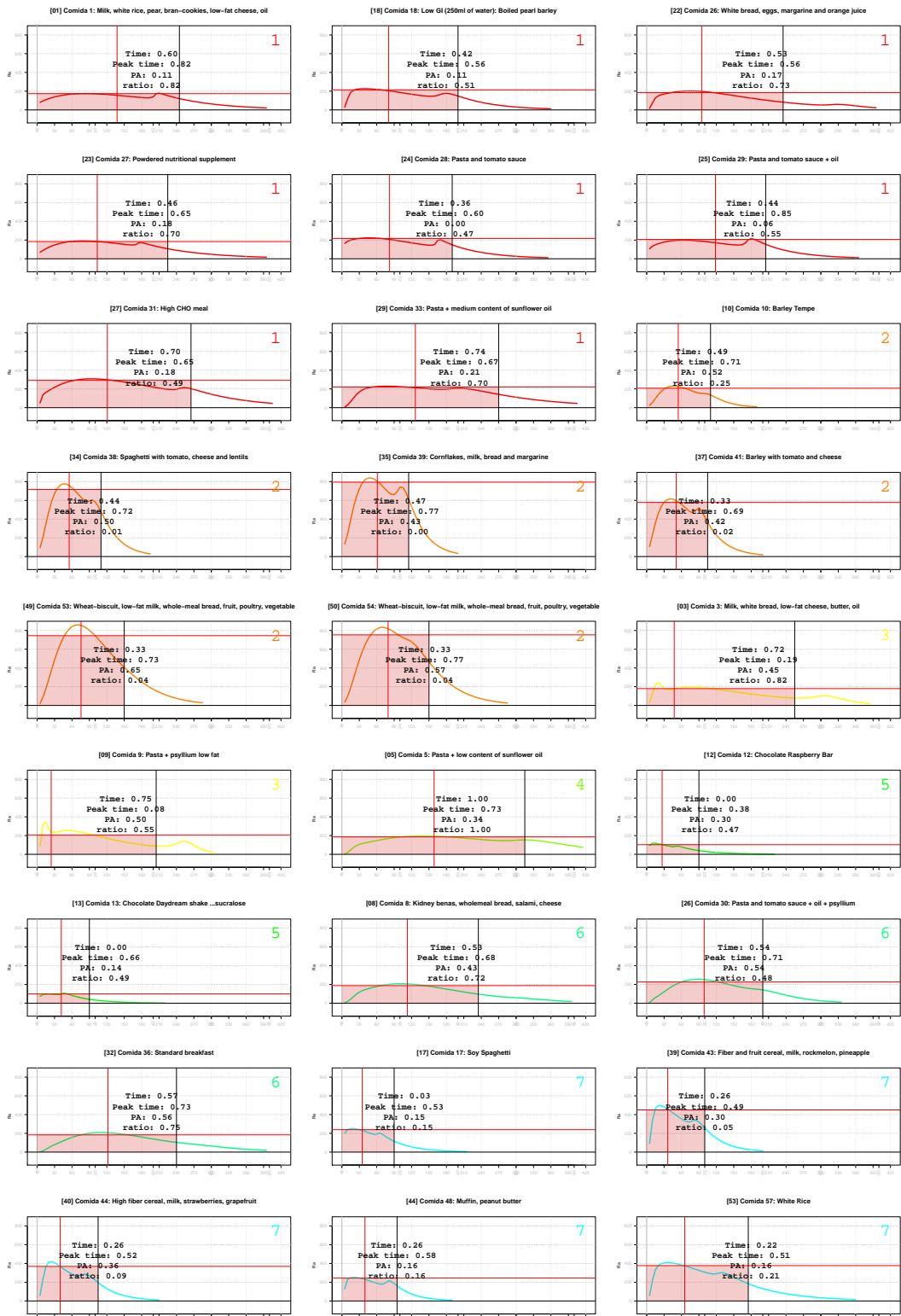


Figura 25: Descripció «relativa» de cada corba del clústering (part 1).

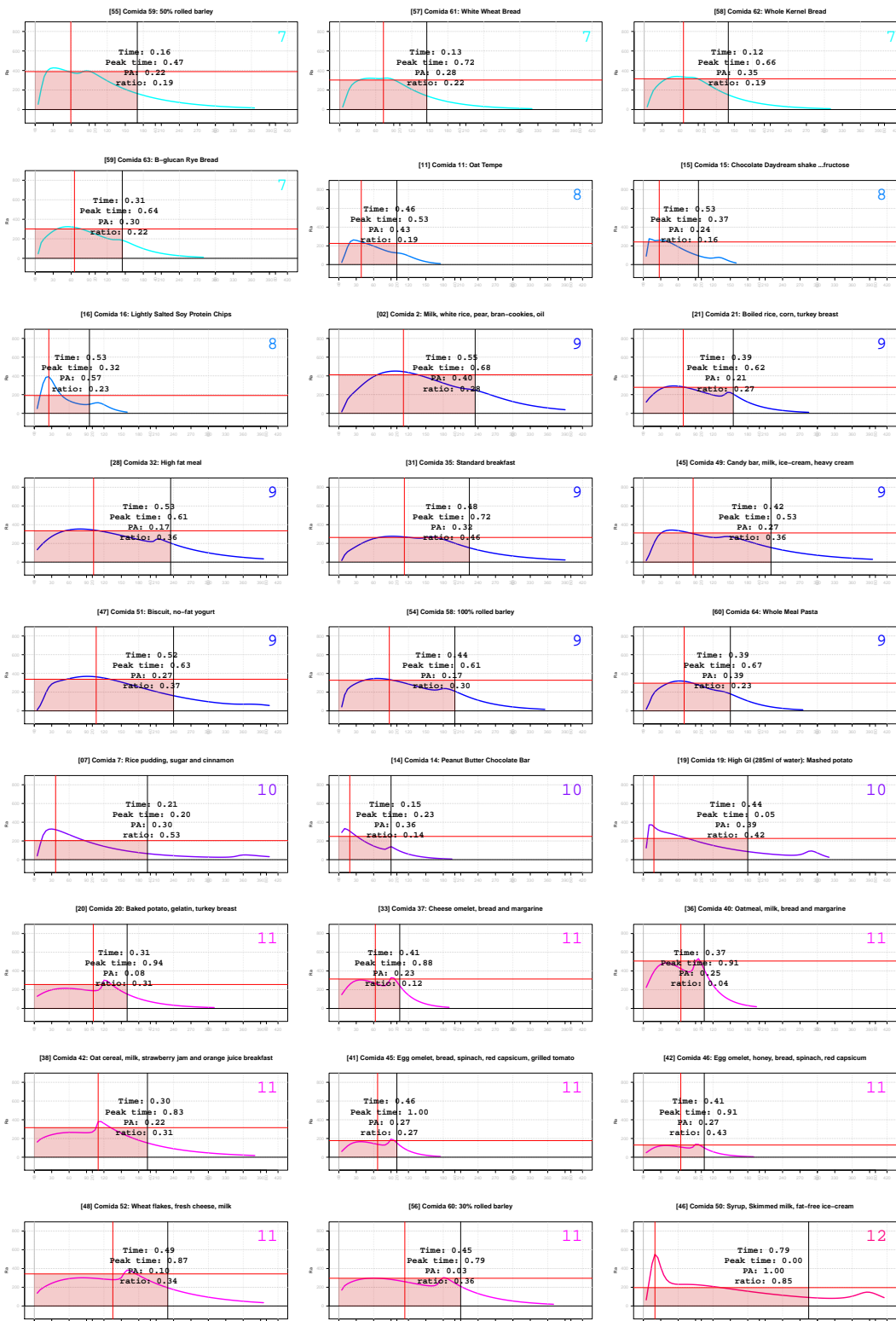
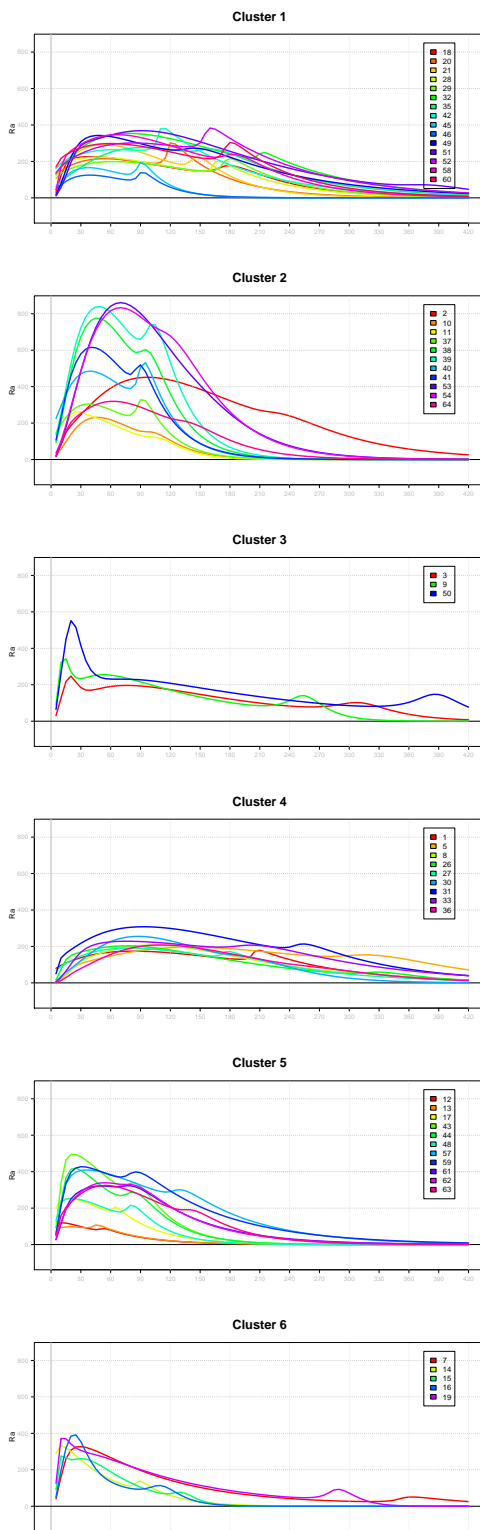


Figura 26: Descripció «relativa» de cada corba del clústering (part 2).



Item	Classification	CHO	Grasa	Proteïna	Fibra	Energia_kCal	Tempo_min		
18	Low GI (250ml of water): Baked pearl barley	Medium	50.0	1.30	9.20	5.10	243.15	240	
20	Baked potato, oil, turkey breast	Medium	45.0	18.00	15.00	-	-	180	
21	Bolled rice, corn, turkey breast	Medium	50.0	18.00	15.00	-	-	180	
28	Pasta and tomato sauce	Medium	50.0	-	-	-	-	240	
29	Pasta and tomato sauce + oil	Medium	50.0	-	-	-	-	240	
32	High fat meal	Medium	27.0	60.00	135.00	15.00	609.94	300	
42	Standard breakfast	Medium	70.3	70.30	15.80	-	-	240	
45	Oat cereal, milk, strawberry jam and orange juice breakfast	Medium	69.0	12.80	13.10	-	-	360	
46	Egg omelet, bread, spinach, red capsicum, grilled tomato	Small	20.2	0	17.50	-	-	238.94	120
46	Egg omelet, honey, bread, spinach, red capsicum	Small	18.5	10.30	17.10	-	-	223.98	120
49	Caroly bar, milk, ice-cream, heavy cream	Large	80.0	80.00	15.00	-	-	1111.11	360
51	Biscuit, no-fat yogurt	Large	94.0	17.00	29.00	-	-	673.78	360
62	Wheat flakes, both cheese, milk	Large	91.0	25.00	28.00	-	-	685.73	360
58	100% rolled barley	Medium	75.0	1.70	7.20	9.20	361.85	240	
60	30% rolled barley	Medium	75.0	1.10	5.40	3.20	348.72	240	

Comida	Nombre	Classificació	CHO	Grasa	Proteïna	Fibra	Energia_kCal	Tempo_min
2	Milk, white rice, pear, bran-cookies, oil	Large	111	22.60	31.00	6.40	776.01	300
10	Barley Tempé	Small	25	1.80	4.80	5.70	150.47	120
11	Oat Tempé	Small	25	1.80	4.60	13.70	181.76	120
37	Cheese omelet, bread and margarine	Small	38	24.00	24.00	-	466.94	120
38	Spaghetti with tomato, cheese and lentils	Fast	67	9.8	44	77	33	0.1
39	Corntflakes, milk, bread and margarine	Fast	104	13.00	18.00	-	609.06	120
40	Oatmeal, milk, bread and margarine	Fast	62	17.00	12.00	-	452.61	120
41	Barley with tomato and cheese	Fast	68	8.00	12.00	-	394.10	120
41	Wheat-biscuit, low-fat milk, whole-meal bread, fruit, poultry, vegetable	Large	123	20.00	17.00	-	673.78	360
54	Wheat-biscuit, low-fat milk, whole-meal bread, fruit, poultry, vegetable	Large	126	28.00	25.00	-	685.73	180
64	Whole Meal Pasta	Medium	50	4.70	12.10	6.60	298.56	180

% CHO	% Grasa	% Proteïna	% Fibra	g	
2	65	13	18	4	171
10	67	5	12	15	37.1
11	65	4	10	30	65.1
37	44	28	28	NA	86
38	72	7	11	NA	121
39	77	10	13	NA	135
40	68	19	13	NA	91
41	77	9	14	NA	88
53	73	17	10	NA	169
54	70	16	14	NA	179
64	69	6	17	8	72.4

Comida	Nombre	Classificació	CHO	Grasa	Proteïna	Fibra	Energia_kCal	Tempo_min
3	Milk, white bread, low-fat cheese, butter, oil	Medium	52.5	10.50	14.50	2.90	362.09	300
9	Pasta + psyllium low fat	Medium	52.0	-	-	1.70	240.00	240
50	Syrup, Skimmed milk, fat-free ice-cream	Large	80.0	0.80	18.00	-	399.35	360

% CHO	% Grasa	% Proteïna	% Fibra	g	
1	65	13	18	4	79.9
5	62	4	14	NA	51.5
26	64	17	17	3	78
27	66	18	0	76	
31	32	10	50	8	291
33	56	34	11	NA	135
36	35	55	10	NA	144.4

Comida	Nombre	Classificació	CHO	Grasa	Proteïna	Fibra	Energia_kCal	Tempo_min
12	Chocolate Raspberry Bar	Small	10.0	12.50	45.00	2.00	526.66	120
13	Chocolate Daydream shake - sucralose	Small	10.0	5.00	40.00	4.00	260.82	120
17	Soy Spaghetti	Small	25.0	1.10	10.60	0.80	151.91	120
43	Fiber and fruit cereal, milk, rockmelon, pineapple	Fast	50.0	2.80	10.20	-	265.12	120
44	High fiber cereal, milk, strawberries, grapefruit	Fast	42.0	2.80	11.00	-	234.07	120
48	Muffin, peanut butter	Small	27.3	15.40	4.50	-	285.12	120
57	White Rice	Medium	75.0	0.80	4.60	0.50	342.74	240
59	50% rolled barley	Medium	75.0	1.30	5.90	4.90	352.78	240
61	Whole Wheat Bread	Medium	50.0	3.00	8.40	3.10	266.79	180
62	Whole Kernel Bread	Medium	60.0	2.60	7.40	12.80	258.91	180
63	B-glucan Rye Bread	Medium	50.0	2.40	10.50	17.10	270.85	180

Figura 27: Clústering aconseguit amb l'aproximació relativa (k = 6).



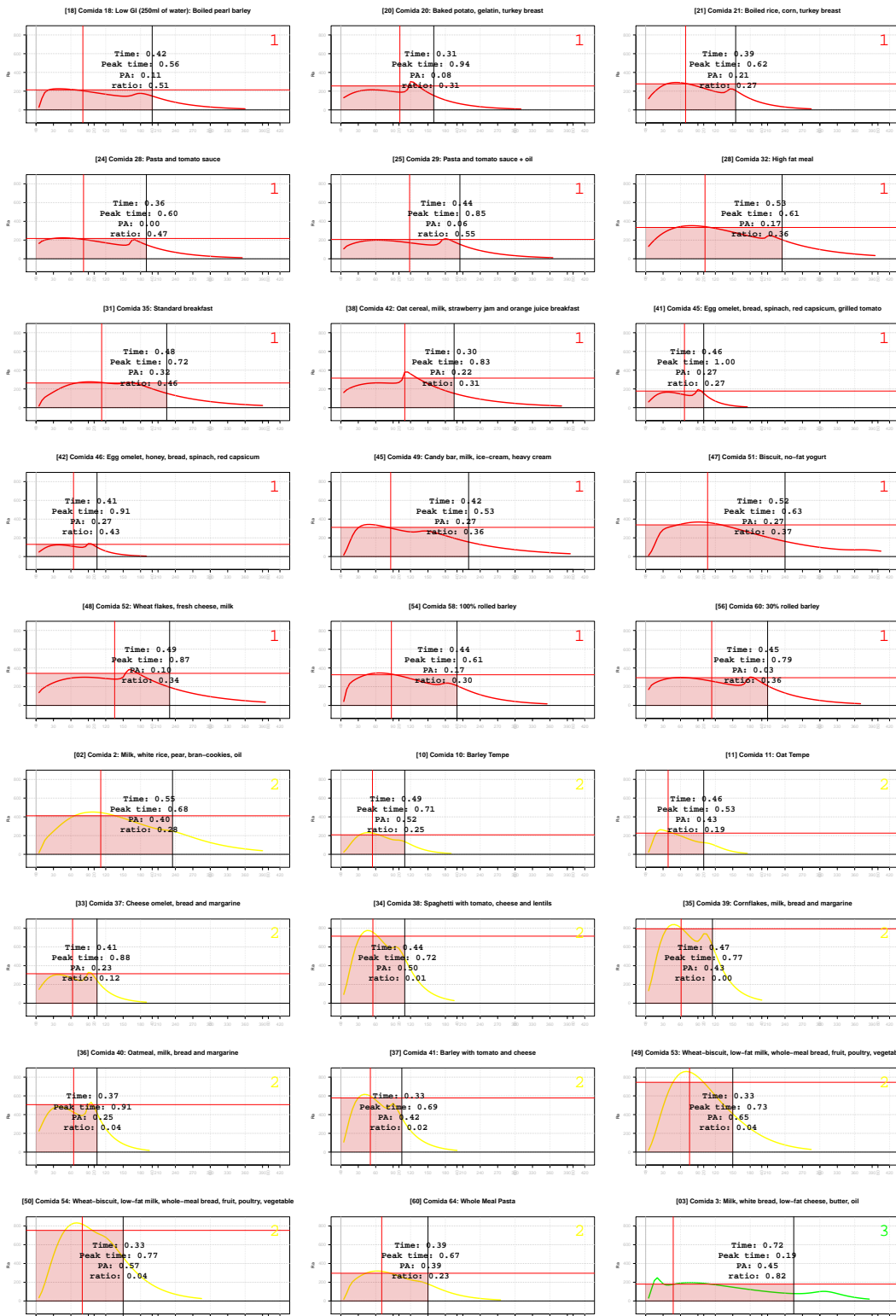


Figura 28: Descripció «relativa» de cada corba del clústering  $k = 6$  (part 1).

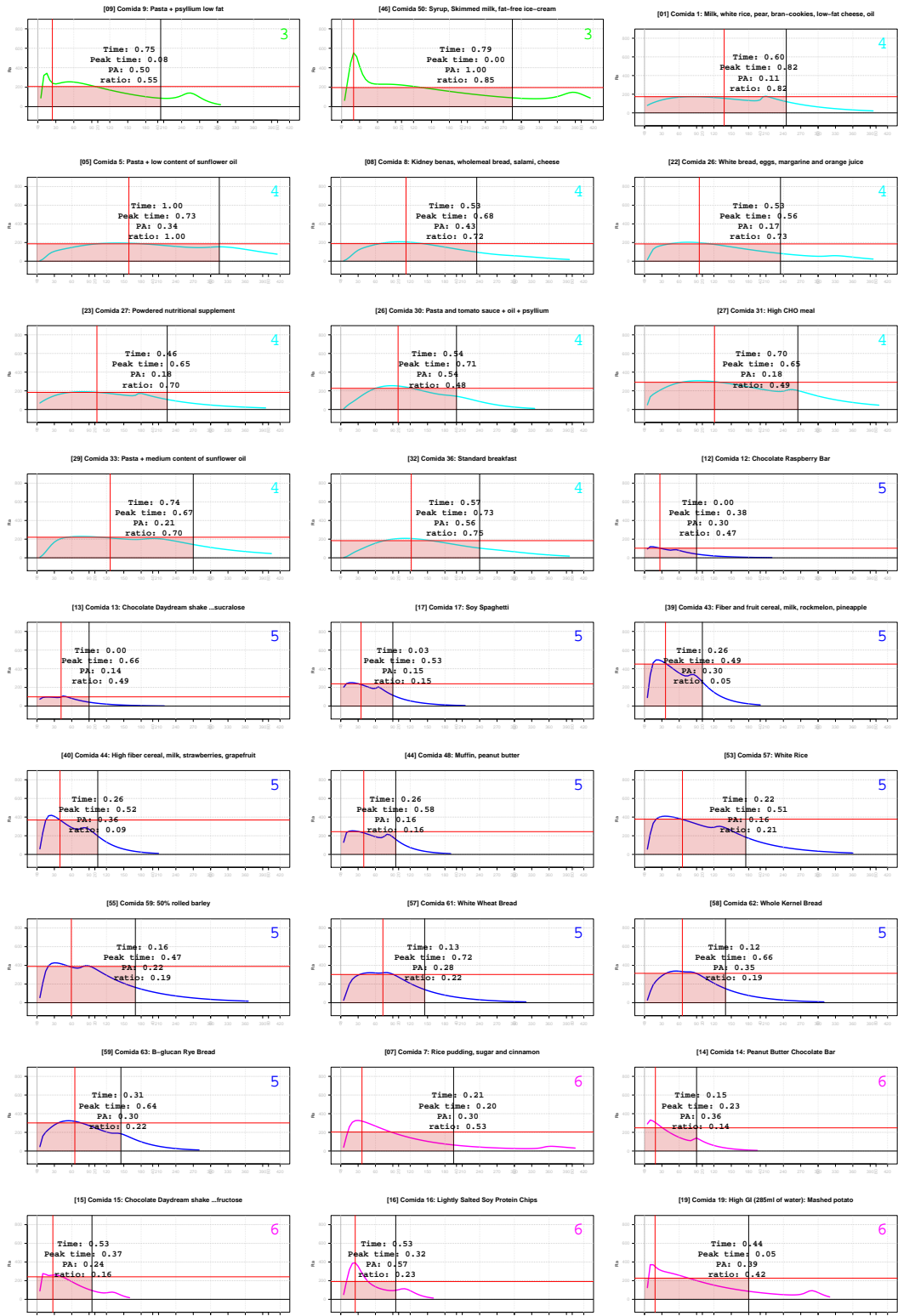


Figura 29: Descripció «relativa» de cada corba del clústering  $k = 6$  (part 2).

# B

## FIGURES COMPLETES DELS RESULTATS COMENTATS

---

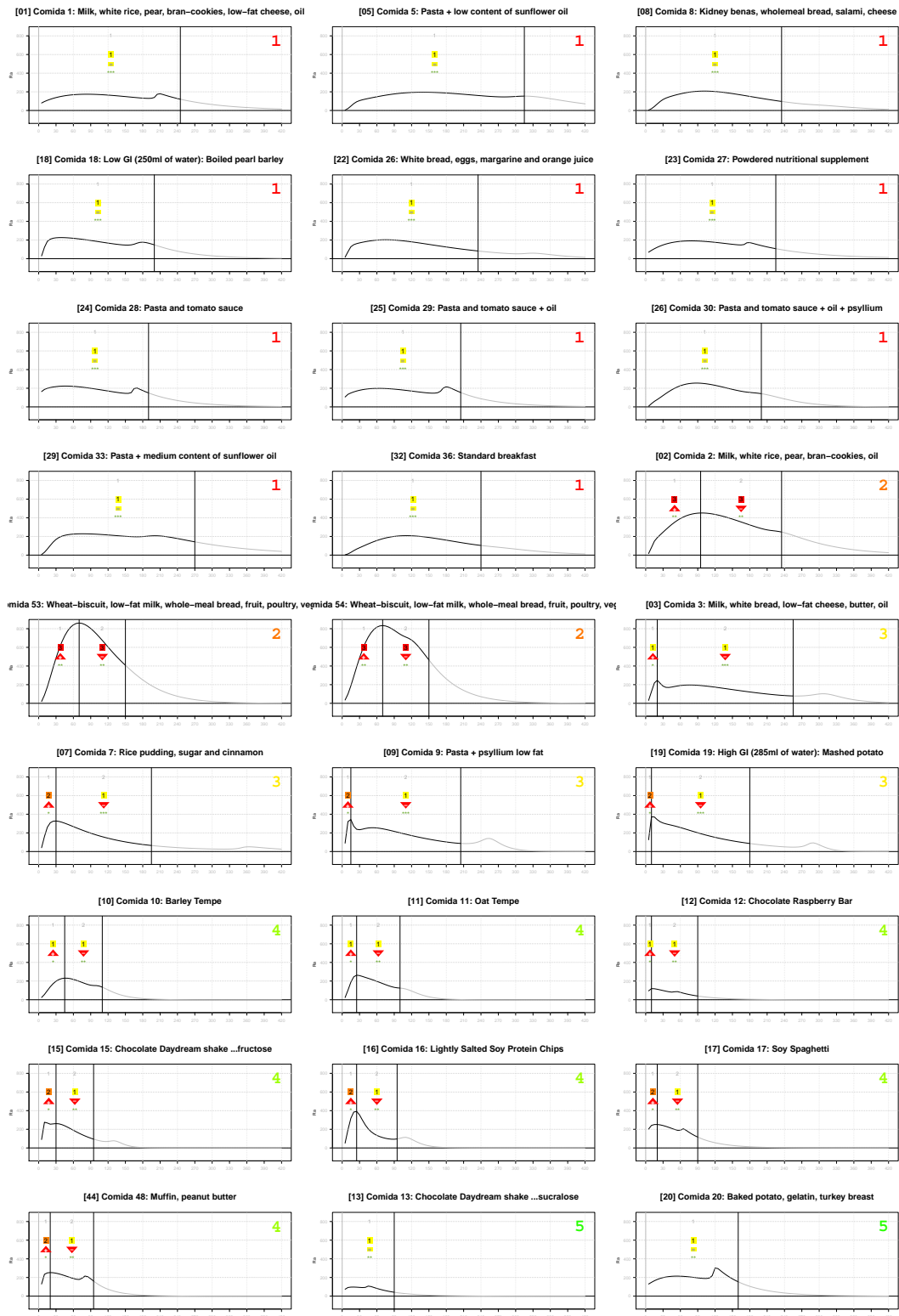


Figura 30: Descripció qualitativa del millor resultat segons DI (1a file de la Taula 7 a la pàgina 37), part 1.

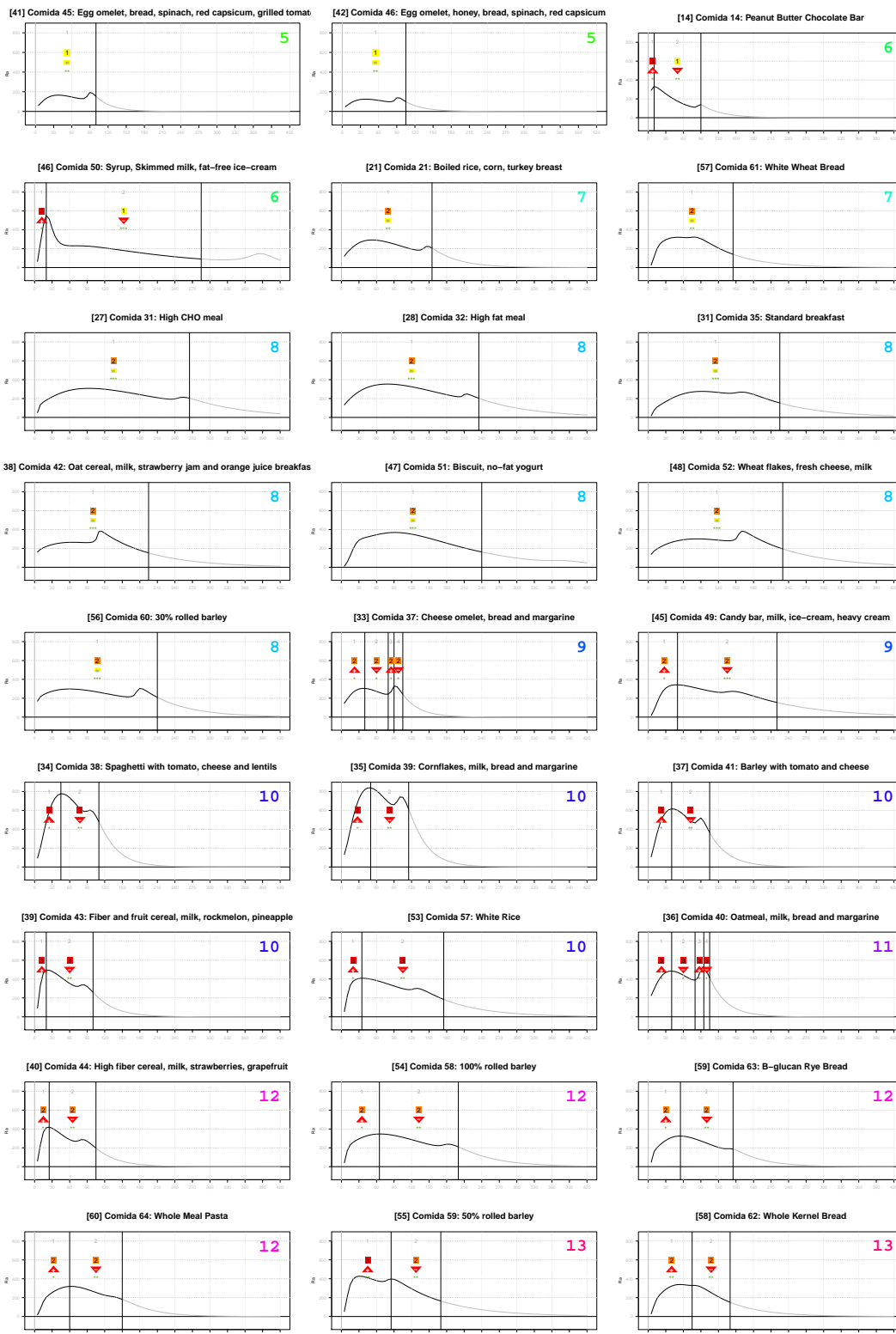


Figura 31: Descripció qualitativa del millor resultat segons DI (1a fila de la Taula 7 a la pàgina 37), part 2.

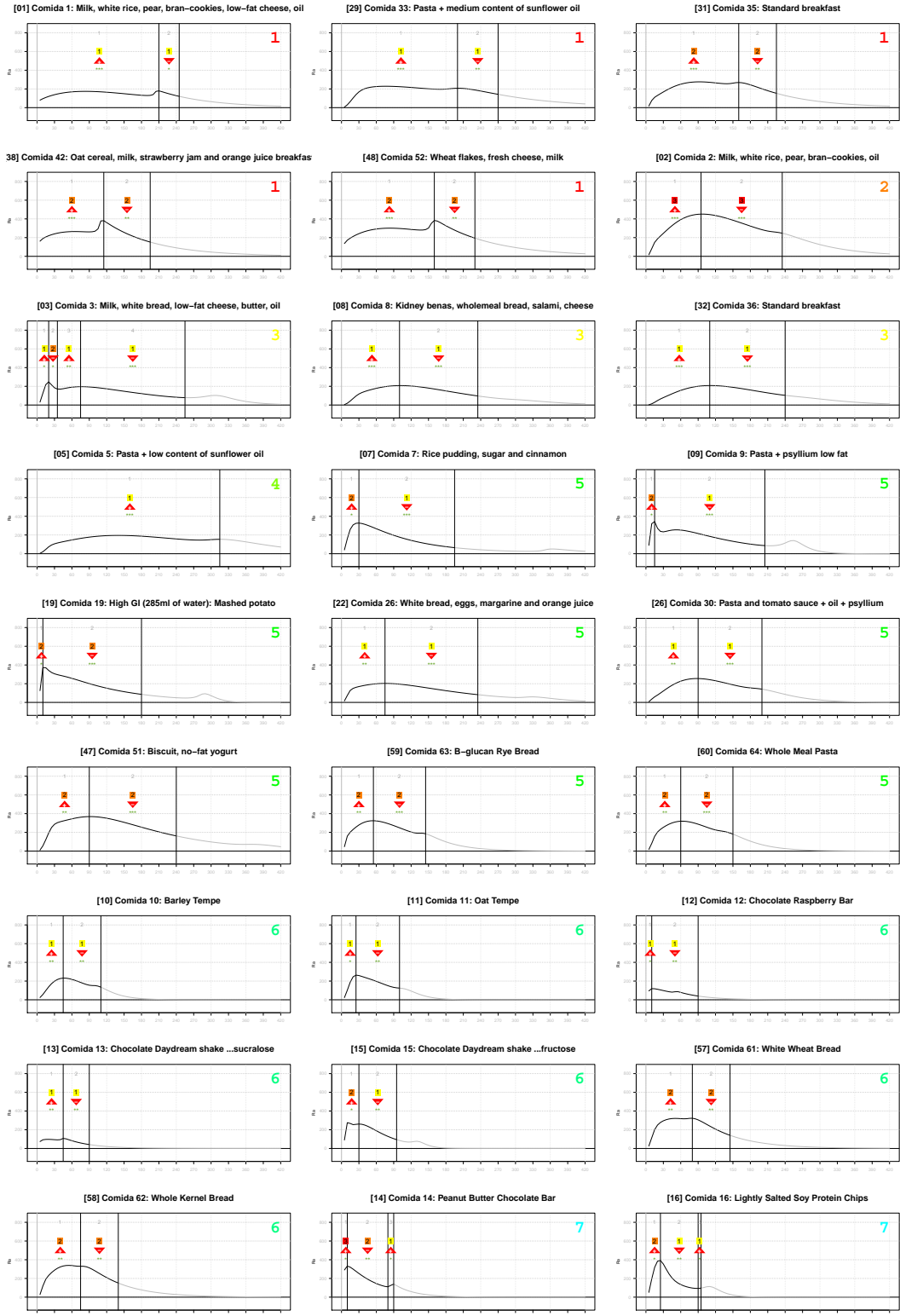


Figura 32: Descripció qualitativa del millor resultat segons la puntuació agregada, excloent CUM && C<=3 (1a fila de la Taula 8 a la pàgina 38), part 1.

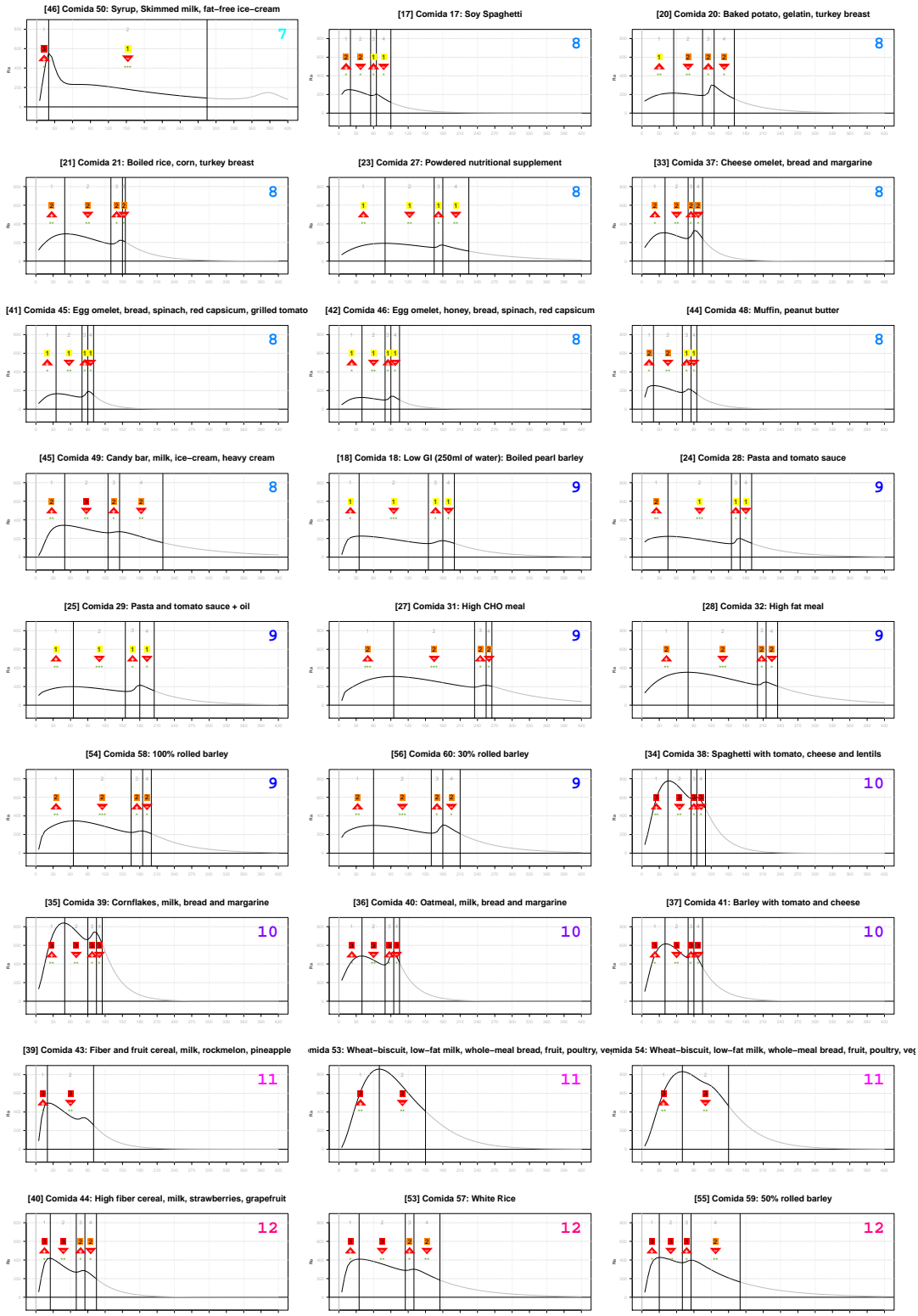


Figura 33: Descripció qualitativa del millor resultat segons la puntuació agregada, excoient CUM && C<= 3 (1a fila de la Taula 8 a la pàgina 38), part 2.

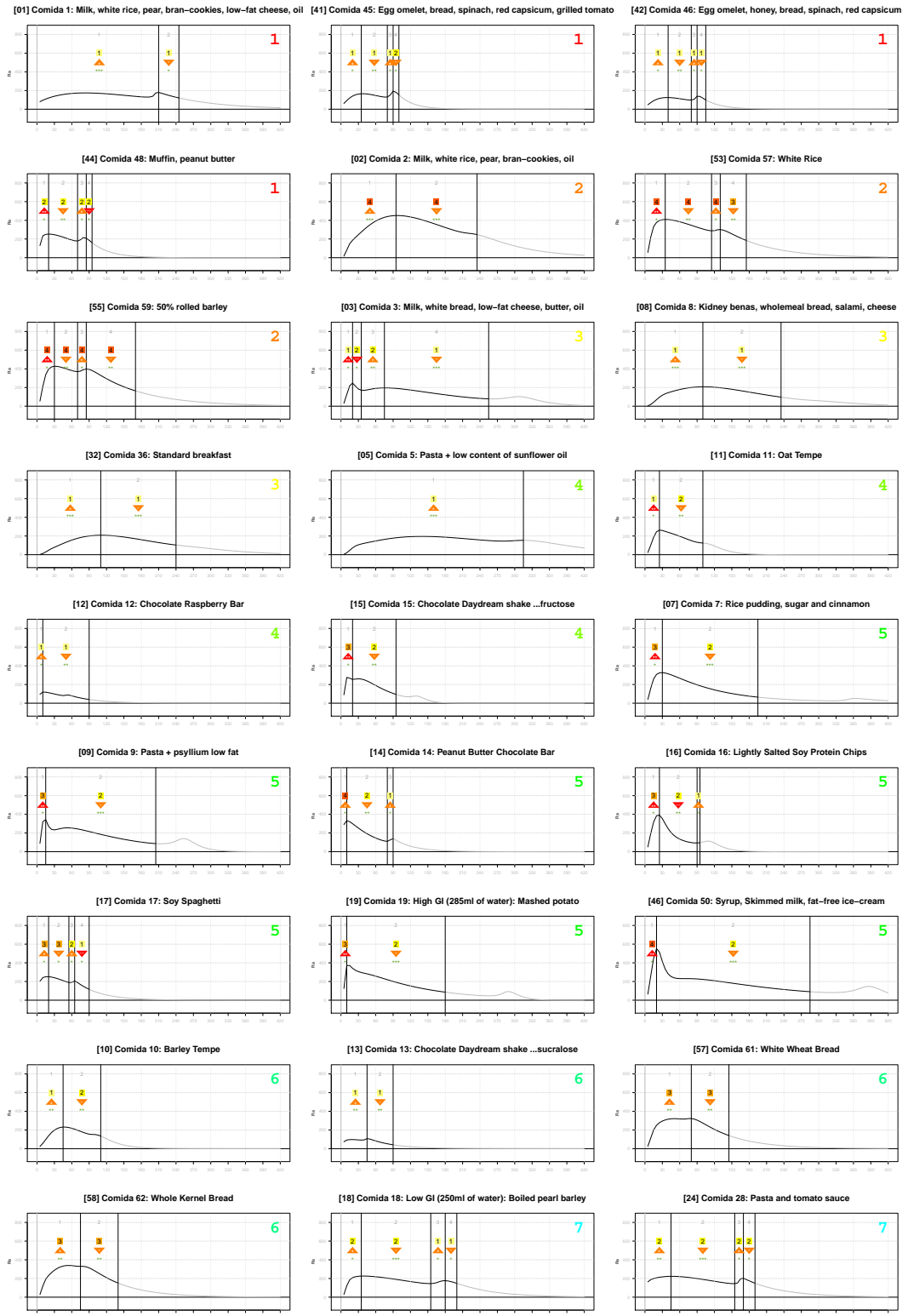


Figura 34: Descripció qualitativa del millor resultat segons la puntuació agregada, excloent  $C \leq 3$  (1a fila de la Taula 9 a la pàgina 39), part 1.



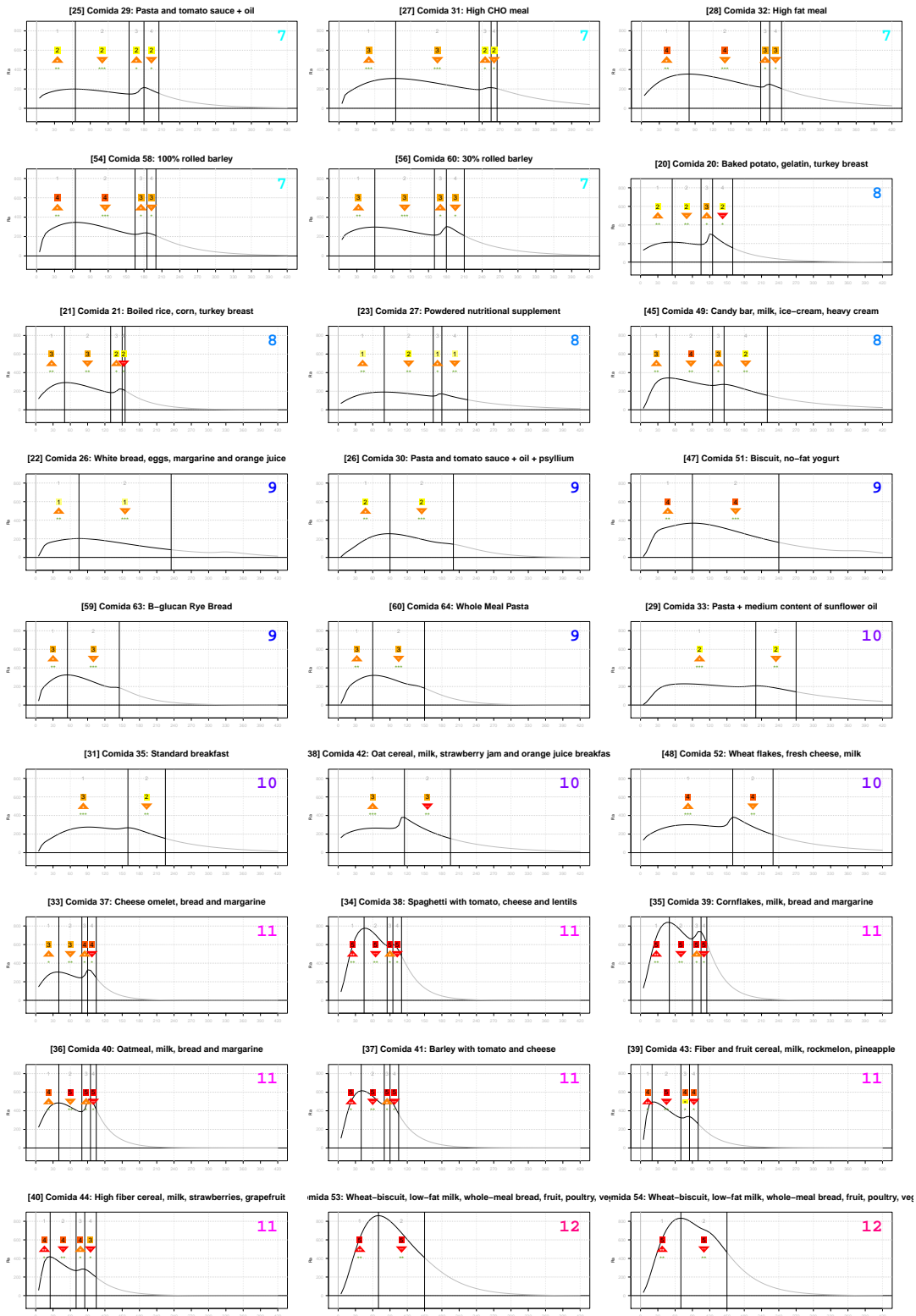


Figura 35: Descripció qualitativa del millor resultat segons la puntuació agregada, excloent  $C \leq 3$  (1a fila de la Taula 9 a la pàgina 39), part 2.

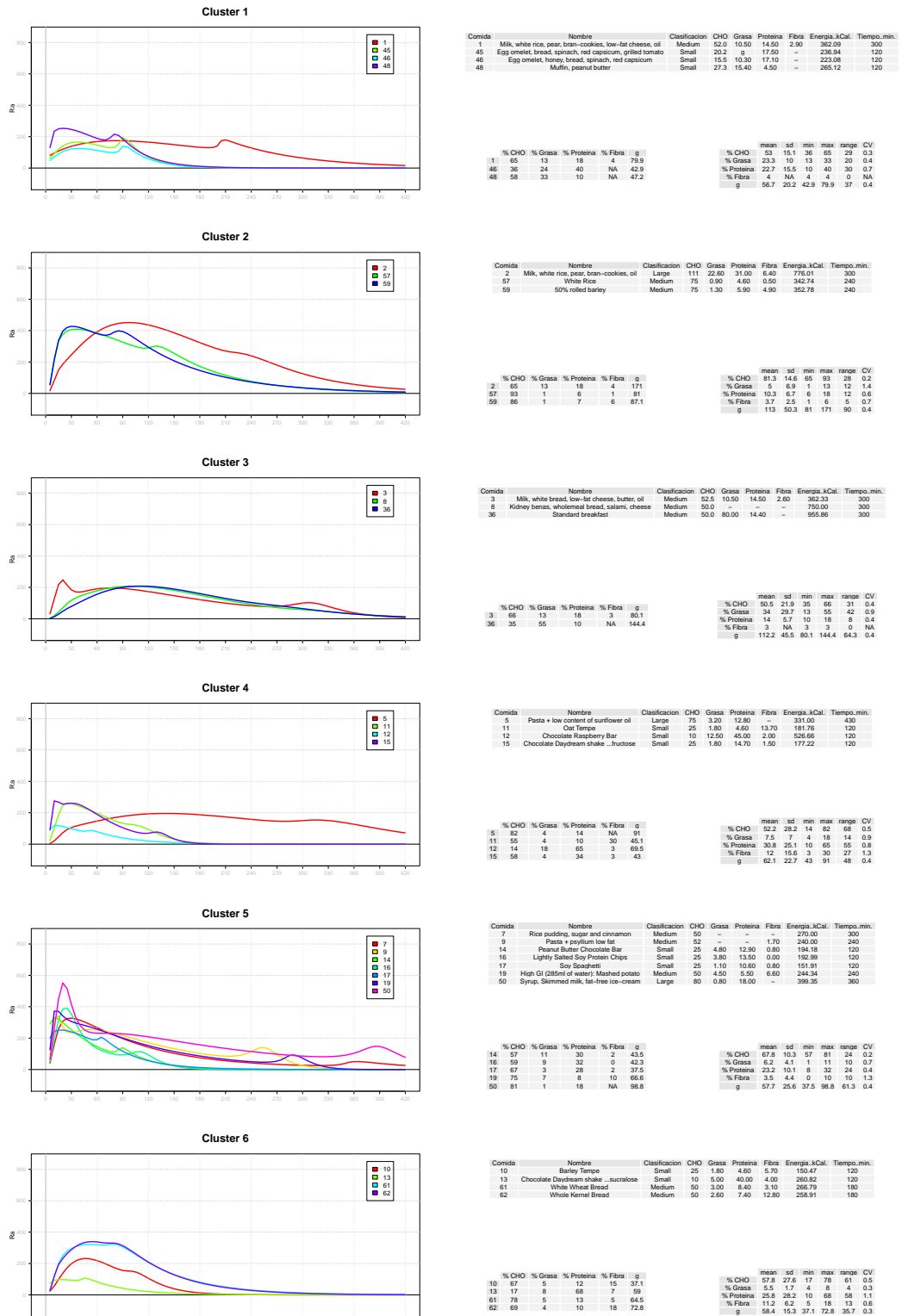
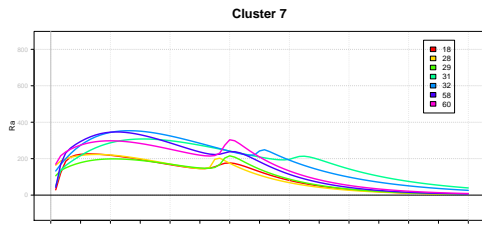
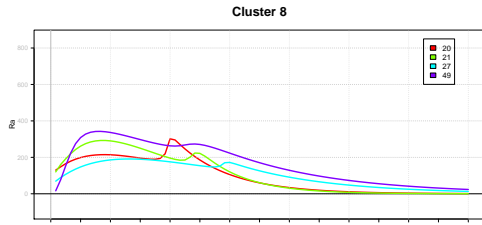


Figura 36: Clustering generat pel millor resultat segons DI i Silueta (1a fila de la Taula 12 a la pàgina 42), part 1.



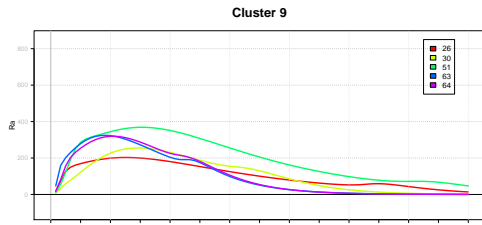
Comida	Nombre	Clasificación	CHO	Grasa	Proteína	Fibra	Energía_kCal.	Tiempo_min.
18	Low GI (250ml of water): Boiled pearl barley	Medium	50	1.30	9.20	5.10	243.15	240
28	Pasta and tomato sauce	Medium	50	-	-	-	240.04	240
29	Pasta and tomato sauce + oil	Medium	50	-	-	-	509.94	240
31	High CHO meal	Large	93	30.00	146.00	22.00	2913.92	360
32	High fat meal	Medium	27	60.00	135.00	18.00	2842.27	360
58	100% rolled barley	Medium	75	1.70	7.20	9.20	361.85	240
60	20% rolled barley	Medium	75	1.10	5.40	3.20	344.72	240

% CHO	% Grasa	% Proteína	% Fibra	g	mean	sd	min	max	range	CV	
18	75	2	14	8	65.6						
31	32	10	50	8	291						
32	11	25	56	8	240						
58	81	2	8	10	93.1						
60	89	1	6	4	84.7						
					mean	sd	min	max	range	CV	
					% CHO	57.8	34.3	11	89	78	0.6
					% Grasa	8	10.2	1	25	24	1.3
					% Proteína	26.8	24.2	6	56	50	0.9
					% Fibra	7.6	2.2	4	10	6	0.3
					g	154.9	103.1	65.6	291	225.4	0.7



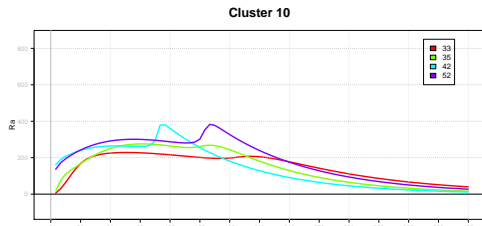
Comida	Nombre	Clasificación	CHO	Grasa	Proteína	Fibra	Energía_kCal.	Tiempo_min.
20	Baked potato, potato, turkey breast	Medium	45	18.00	15.00	-	-	180
21	Boiled rice, corn, turkey breast	Medium	50	18.00	15.00	-	-	180
27	Powdered nutritional supplement	Medium	50	12.00	14.00	0.00	366.87	300
49	Candy bar, milk, ice-cream, heavy cream	Large	80	80.00	18.00	-	1111.11	360

% CHO	% Grasa	% Proteína	% Fibra	g	mean	sd	min	max	range	CV	
20	58	23	19	NA	79						
21	60	22	18	NA	83						
27	66	16	18	0	75						
49	45	45	10	NA	178						
					mean	sd	min	max	range	CV	
					% CHO	57.2	8.8	45	66	21	0.2
					% Grasa	26.5	12.7	16	45	29	0.5
					% Proteína	16.2	4.2	10	19	9	0.3
					% Fibra	0	NA	0	0	NA	
					g	103.8	49.6	76	178	102	0.5



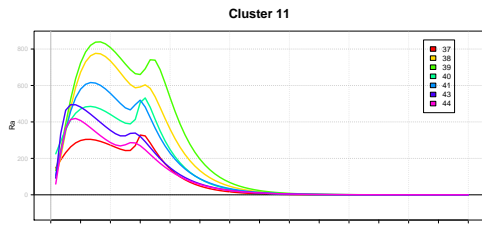
Comida	Nombre	Clasificación	CHO	Grasa	Proteína	Fibra	Energía_kCal.	Tiempo_min.
26	White bread, eggs, margarine and orange juice	Medium	50	13.00	13.00	2.00	360.90	300
30	Pasta and tomato sauce + oil + psyllium	Medium	50	-	-	-	509.94	240
51	Biscuit, no-fat yogurt	Large	94	17.00	29.00	-	673.78	360
63	B-glucan Rice Bread	Medium	50	2.40	10.50	17.10	270.85	180
64	Whole Meal Pasta	Medium	50	4.70	12.10	6.60	286.96	180

% CHO	% Grasa	% Proteína	% Fibra	g	mean	sd	min	max	range	CV	
26	64	17	17	3	79						
51	67	12	21	NA	140						
63	62	3	13	21	80						
64	69	6	17	8	72.4						
					mean	sd	min	max	range <td>CV</td>	CV	
					% CHO	65.5	3.1	62	69	7	0
					% Grasa	9.5	6.2	3	17	14	0.7
					% Proteína	17	3.3	13	21	8	0.2
					% Fibra	10.7	9.3	3	21	18	0.9
					g	92.6	31.8	72.4	140	67.6	0.3



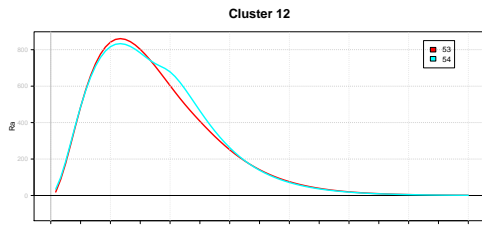
Comida	Nombre	Clasificación	CHO	Grasa	Proteína	Fibra	Energía_kCal.	Tiempo_min.
33	Pasta + medium content of sunflower oil	Large	75.0	45.50	14.50	-	952.76	420
35	Standard breakfast	Medium	70.3	70.30	15.60	-	952.76	300
42	Oat cereal, milk, strawberry jam and orange juice breakfast	Medium	69.0	12.80	13.10	-	441.86	300
52	Wheat flakes, french cheese, milk	Large	91.0	26.00	28.00	-	685.73	180

% CHO	% Grasa	% Proteína	% Fibra	g	mean	sd	min	max	range	CV	
33	56	34	11	NA	135						
35	45	45	10	NA	156.2						
42	73	13	14	NA	94.9						
52	63	17	19	NA	144						
					mean	sd	min	max	range <td>CV</td>	CV	
					% CHO	59.2	11.8	45	73	28	0.2
					% Grasa	27.2	14.9	13	45	32	0.5
					% Proteína	13.5	4	10	19	9	0.3
					% Fibra	NA	NA	inf	inf	inf	NA
					g	132.5	26.5	94.9	156.2	61.3	0.2



Comida	Nombre	Clasificación	CHO	Grasa	Proteína	Fibra	Energía_kCal.	Tiempo_min.
37	Cheese omelet, bread and margarine	Small	38	24.00	24.00	-	466.94	120
38	Spaghetti with tomato, cheese and lentils	Fast	87	9.00	25.00	-	527.85	120
39	Cornflakes, milk, bread and margarine	Fast	104	13.00	18.00	-	609.06	120
40	Oatmeal, milk, bread and margarine	Fast	62	17.00	12.00	-	452.61	120
41	Barley with tomato and cheese	Fast	68	8.00	12.00	-	394.10	120
43	Fiber and fruit cereal, milk, rockmelon, pineapple	Fast	50	2.60	10.20	-	285.12	120
44	High fiber cereal, milk, strawberries, grapefruit	Fast	42	2.50	11.00	-	234.07	120

% CHO	% Grasa	% Proteína	% Fibra	g	mean	sd	min	max	range	CV	
37	44	28	28	NA	86						
38	72	7	21	NA	121						
39	77	10	13	NA	135						
40	69	19	13	NA	91						
41	77	5	14	NA	89						
43	80	4	16	NA	62.8						
44	76	6	20	NA	55.5						
					mean	sd	min	max	range <td>CV</td>	CV	
					% CHO	71.5	2.1	70	73	3	0
					% Grasa	16.5	0.7	16	17	1	0
					% Proteína	17.8	5.5	13	26	15	0.3
					% Fibra	NA	NA	inf	inf	inf	NA
					g	81.3	28.6	55.5	135	79.5	0.3



Comida	Nombre	Clasificación	CHO	Grasa	Proteína	Fibra	Energía_kCal.	Tiempo_min.
53	Wheat-biscuit, low-fat milk, whole-meal bread, fruit, poultry, vegetable	Large	123	29.00	17.00	-	673.78	180
54	Wheat-biscuit, low-fat milk, whole-meal bread, fruit, poultry, vegetable	Large	126	29.00	20.00	-	685.73	180

% CHO	% Grasa	% Proteína	% Fibra	g	mean	sd	min	max	range	CV	
53	73	17	10	NA	169						
54	70	16	14	NA	179						
					mean	sd	min	max	range <td>CV</td>	CV	
					% CHO	71.5	2.1	70	73	3	0
					% Grasa	16.5	0.7	16	17	1	0
					% Proteína	17.8	5.5	13	26	15	0.3
					% Fibra	NA	NA	inf	inf	inf	NA
					g	174	7.1	169	179	10	0

Figura 37: Clústering generat pel millor resultat segons DI i Silueta (1a fila de la Taula 12 a la pàgina 42), part 2.

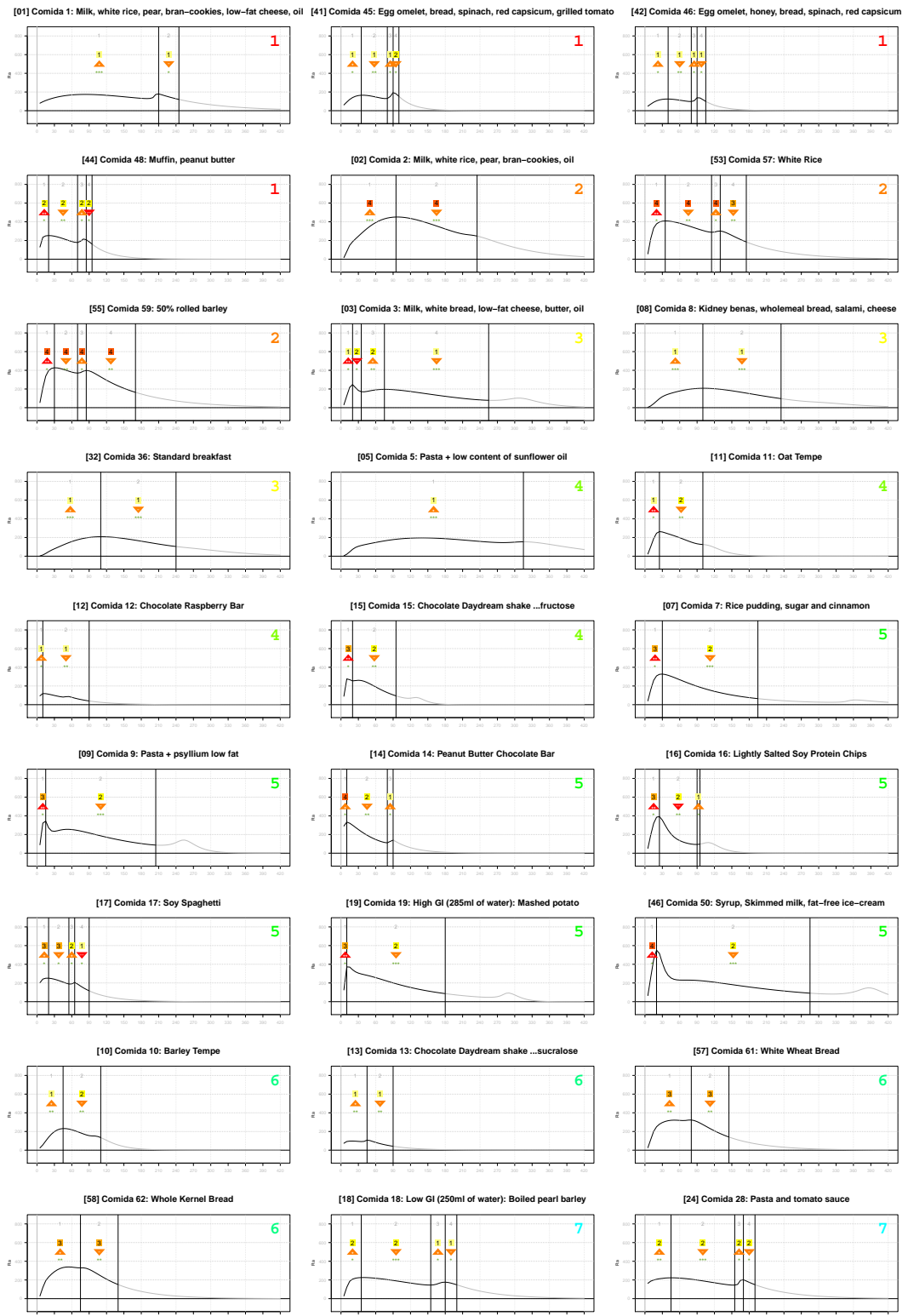


Figura 38: Descripció qualitativa del millor resultat segons DI i Silueta (1a fila de la Taula 12 a la pàgina 42), part 1.

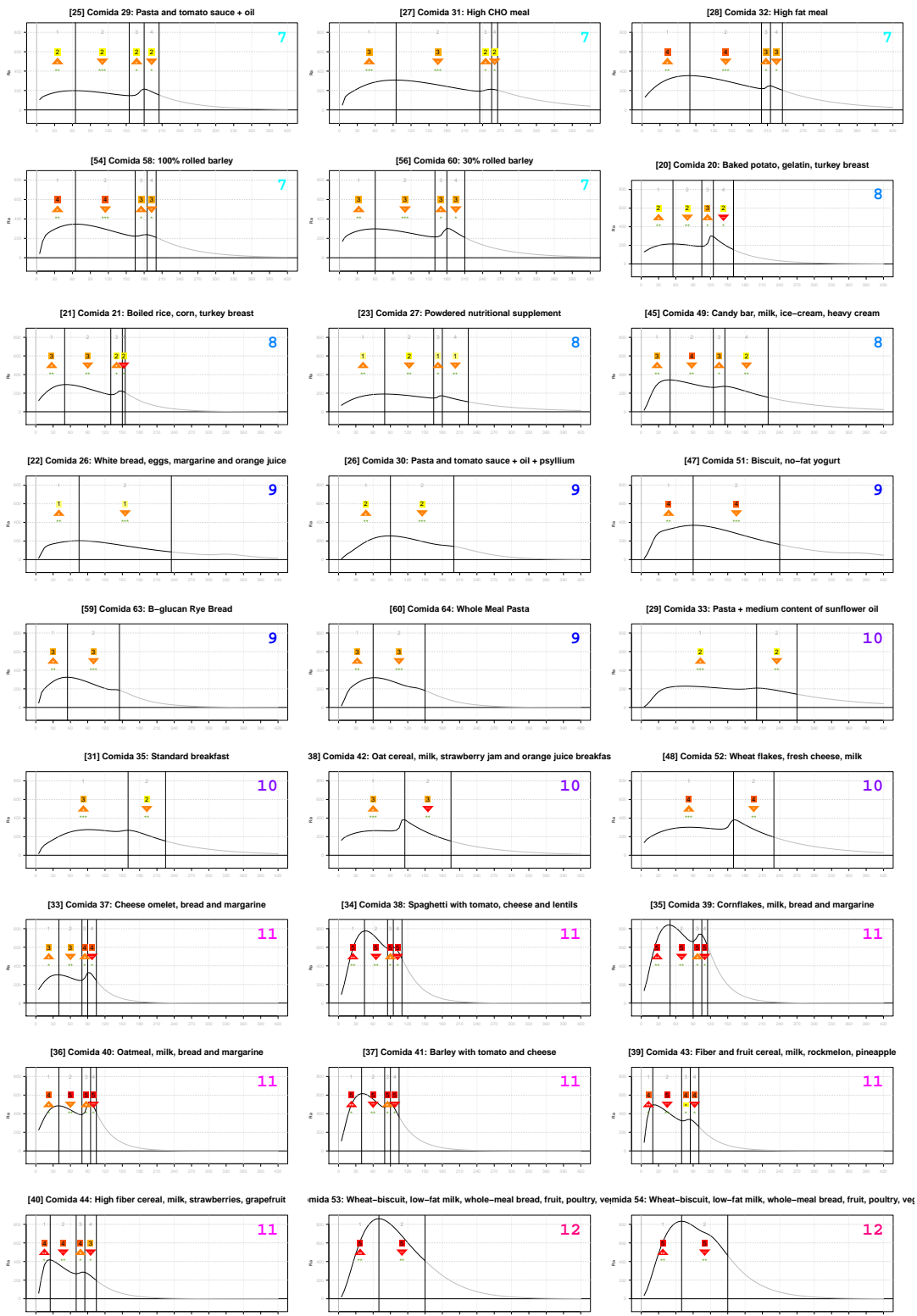
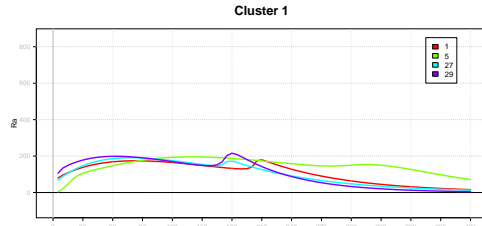
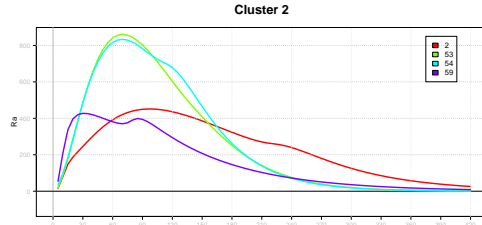


Figura 39: Descripció qualitativa del millor resultat segons DI i Silueta (1a fila de la Taula 12 a la pàgina 42), part 2.



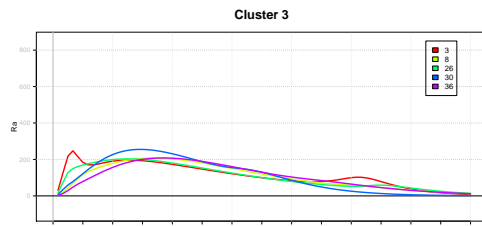
Comida	Nombre	Clasificación	CHO	Grasa	Proteïna	Fibra	Energia_kCal.	Tempo_min.
1	Milk, white rice, pear, brown-cookies, low-fat cheese, oil	Medium	52	10.50	14.50	2.90	352.00	300
5	Pasta + low content of sunflower oil	Large	75	3.20	12.80	-	331.00	430
27	Protein nutritional supplement	Medium	60	12.00	14.00	0.00	305.97	300
29	Pasta and tomato sauce + oil	Medium	50	-	-	-	509.94	240

% CHO	% Grasa	% Proteïna	% Fibra	g	mean	sd	min	max	range	CV	
1	65	13	18	4	79.9						
5	82	4	14	NA	91						
27	66	16	18	0	76						
					% CHO	71	9.5	65	82	17	0.1
					% Grasa	11	6.2	4	16	12	0.6
					% Proteïna	16.7	2.3	14	18	4	0.1
					% Fibra	2	2.8	0	4	4	1.4
					g	82.3	7.8	76	91	15	0.1



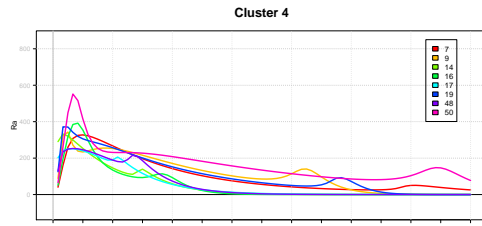
Comida	Nombre	Clasificación	CHO	Grasa	Proteïna	Fibra	Energia_kCal.	Tempo_min.
2	Milk, white rice, pear, brown-cookies, oil	Medium	52	10.50	14.50	2.90	352.00	300
53	Wheat-biscuit, low-fat milk, whole-meal bread, fruit, poultry, vegetable	Large	123	29.00	17.00	-	673.78	180
54	Wheat-biscuit, low-fat milk, whole-meal bread, fruit, poultry, vegetable	Large	126	28.00	25.00	-	685.73	180
59	50% rolled barley	Medium	75	1.90	5.90	4.90	352.78	240

% CHO	% Grasa	% Proteïna	% Fibra	g	mean	sd	min	max	range	CV	
2	65	13	18	4	171						
53	73	17	10	NA	169						
54	70	16	14	NA	179						
59	66	1	7	6	87.1						
					% CHO	73.5	9	65	86	21	0.1
					% Grasa	11.8	7.4	1	17	16	0.6
					% Proteïna	12.2	4.8	7	18	11	0.4
					% Fibra	5	1.4	4	6	2	0.3
					g	151.5	43.2	87.1	179	91.9	0.3



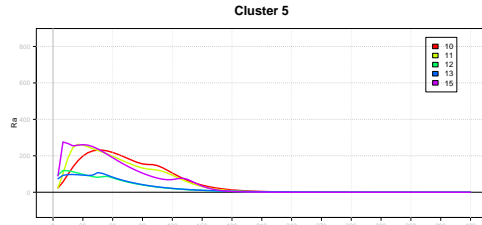
Comida	Nombre	Clasificación	CHO	Grasa	Proteïna	Fibra	Energia_kCal.	Tempo_min.
3	Milk, white bread, low-fat cheese, butter, oil	Medium	52.5	10.50	14.50	2.60	362.33	300
8	Kidney beans, wholemeal bread, salmon, cheese	Medium	50.0	-	-	-	750.00	300
26	White bread, soap, margarine and orange juice	Medium	50.0	13.00	13.00	2.00	360.90	300
30	Pasta and tomato sauce + oil + psyllium	Medium	50.0	-	-	-	609.94	240
36	Standard breakfast	Medium	60.0	80.00	14.40	-	955.86	300

% CHO	% Grasa	% Proteïna	% Fibra	g	mean	sd	min	max	range	CV	
3	66	13	18	3	80.1						
26	64	17	17	3	78						
30	35	55	10	NA	144.4						
36	66	13	18	3	80.1						
					% CHO	55	17.3	35	66	31	0.3
					% Grasa	29.3	23.2	13	55	42	0.8
					% Proteïna	15	4.4	10	18	8	0.3
					% Fibra	3	0	3	3	0	0
					g	100.8	37.7	78	144.4	66.4	0.4



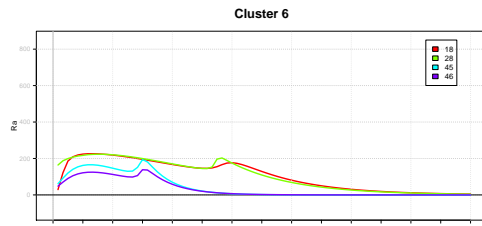
Comida	Nombre	Clasificación	CHO	Grasa	Proteïna	Fibra	Energia_kCal.	Tempo_min.
7	Rice pudding, sugar and cinnamon	Medium	50.0	-	-	-	270.00	300
9	Pasta + psyllium low fat	Medium	52.0	-	-	-	240.00	240
14	Peanut Butter Chocolate Bar	Small	29.0	4.80	12.90	0.80	194.18	120
16	Lightly Salted Soy Protein Chips	Small	25.0	3.80	13.50	0.00	192.99	120
17	Soy Soapnuts	Small	29.0	1.10	0.60	0.80	151.91	120
19	High GI (285ml of water): Mashed potato	Medium	50.0	4.50	5.50	6.60	244.34	240
48	Muffin, peanut butter	Small	27.0	15.40	4.50	-	205.12	120
50	Syrup, Skimmed milk, fat-free ice-cream	Large	80.0	0.80	18.00	-	399.35	360

% CHO	% Grasa	% Proteïna	% Fibra	g	mean	sd	min	max	range	CV	
14	57	11	30	2	43.5						
16	59	9	32	0	42.3						
17	67	3	28	2	37.5						
19	75	7	8	10	65.6						
48	58	33	10	NA	47.2						
50	81	1	18	NA	98.8						
					% CHO	66.2	10	57	81	24	0.2
					% Grasa	10.7	11.6	1	33	32	1.1
					% Proteïna	21	10.9	8	32	24	0.5
					% Fibra	3.5	4.4	0	10	10	1.3
					g	56	23.3	37.5	98.8	61.3	0.4



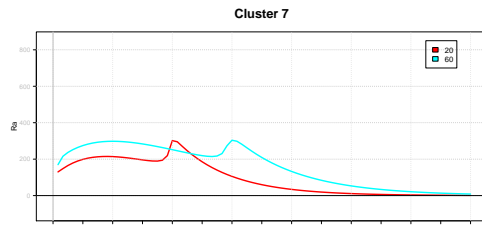
Comida	Nombre	Clasificación	CHO	Grasa	Proteïna	Fibra	Energia_kCal.	Tempo_min.
10	Barley Tempe	Small	25	1.80	4.60	5.70	150.47	120
11	Das Tempe	Small	25	1.80	4.60	13.70	181.76	120
12	Chocolate Raspberry Bar	Small	10	12.50	45.00	2.00	528.66	120
13	Chocolate Daydream shake ...sucralose	Small	10	5.00	40.00	4.00	260.82	120
15	Chocolate Daydream shake ...fructose	Small	25	1.80	14.70	1.50	177.22	120

% CHO	% Grasa	% Proteïna	% Fibra	g	mean	sd	min	max	range	CV	
10	67	5	12	10	37.1						
11	55	4	10	30	45.1						
12	14	18	65	3	69.5						
13	17	8	68	7	59						
15	58	4	34	3	43						
					% CHO	42.2	24.8	14	67	53	0.6
					% Grasa	7.8	5.9	4	18	14	0.8
					% Proteïna	37.8	27.9	10	68	58	0.7
					% Fibra	11.6	11.4	3	30	27	1
					g	50.7	13.2	37.1	69.5	32.4	0.3



Comida	Nombre	Clasificación	CHO	Grasa	Proteïna	Fibra	Energia_kCal.	Tempo_min.
18	Low GI (250ml of water): Boiled pearl barley	Medium	50.0	1.30	9.20	5.10	243.15	240
28	Pasta and tomato sauce	Medium	50.0	-	-	-	240.94	240
45	Egg omelet, bread, spinach, red capsicum, grilled tomato	Small	20.2	g	17.50	-	236.84	120
46	Egg omelet, honey, bread, spinach, red capsicum	Small	15.5	10.30	17.10	-	223.08	120

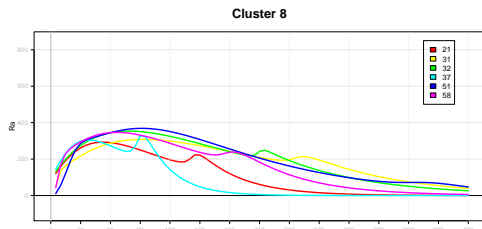
% CHO	% Grasa	% Proteïna	% Fibra	g	mean	sd	min	max	range	CV	
18	76	2	14	8	65.6						
46	36	24	40	NA	42.9						
					% CHO	56	28.3	36	76	40	0.5
					% Grasa	13	15.6	2	24	22	1.2
					% Proteïna	27	18.4	14	40	26	0.7
					% Fibra	8	NA	8	0	NA	NA
					g	54.2	16.1	42.9	65.6	22.7	0.3



Comida	Nombre	Clasificación	CHO	Grasa	Proteïna	Fibra	Energia_kCal.	Tempo_min.
20	Baked potato, gelatin, turkey breast	Medium	45	18.00	15.00	-	348.72	180
60	30% rolled barley	Medium	75	1.10	5.40	3.20	348.72	240

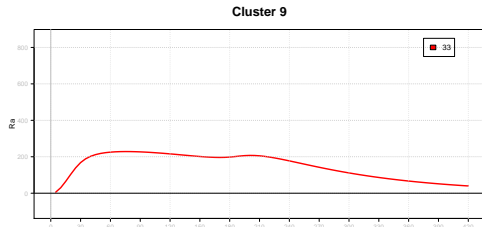
% CHO	% Grasa	% Proteïna	% Fibra	g	mean	sd	min	max	range	CV	
20	58	23	19	NA	78						
60	69	1	6	4	84.7						
					% CHO	73.5	21.9	58	89	31	0.3
					% Grasa	12	15.6	1	23	22	1.3
					% Proteïna	12.5	9.2	6	19	13	0.7
					% Fibra	4	NA	4	4	0	NA
					g	81.3	4.7	78	84.7	6.7	0.1

Figura 40: Clústering generat pel 14è resultat segons DI i Silueta (14a fila de la Taula 12 a la pàgina 42), part 1.



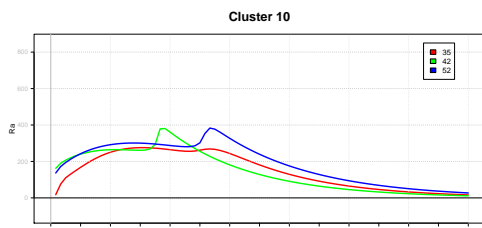
Comida	Nombre	Clasificación	CHO	Grasa	Proteína	Fibra	Energia_kCal.	Tiempo_min.
21	Boiled rice, corn, turkey breast	Medium	50	18.00	15.00	-	-	180
31	High CHO meal	Large	93	30.00	146.00	22.00	2913.92	360
32	High fat meal	Medium	27	60.00	136.00	18.00	2842.27	360
37	Cheese omelet, bread and margarine	Small	38	24.00	24.00	-	466.94	120
51	Biscuit, no-fat yogurt	Large	94	17.00	29.00	-	673.76	360
58	100% rolled barley	Medium	75	1.70	7.20	9.20	361.85	240

% CHO	% Grasa	% Proteína	% Fibra	g	mean	sd	min	max	range	CV	
21	60	22	18	NA	83						
31	32	10	50	8	291						
32	11	25	56	8	240						
37	44	28	28	NA	86						
51	67	12	21	NA	140						
58	81	2	8	10	93.1						
					% CHO	49.2	25.4	11	81	70	0.5
					% Grasa	16.5	10.1	2	28	26	0.6
					% Proteína	30.2	18.9	8	56	48	0.6
					% Fibra	8.7	1.2	8	10	2	0.1
					g	155.6	65.1	83	291	218	0.6



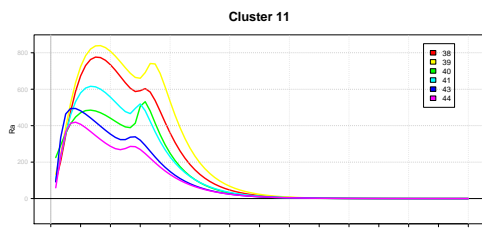
Comida	Nombre	Clasificación	CHO	Grasa	Proteína	Fibra	Energia_kCal.	Tiempo_min.
33	Pasta + medium content of sunflower oil	Large	75	45.50	14.50	-	952.76	420

% CHO	% Grasa	% Proteína	% Fibra	g	mean	sd	min	max	range	CV
33	56	34	11	NA	135					
					% CHO	56	NA	56	0	NA
					% Grasa	34	NA	34	0	NA
					% Proteína	11	NA	11	0	NA
					% Fibra	NaN	NA	inf	-inf	NA
					g	135	NA	135	0	NA



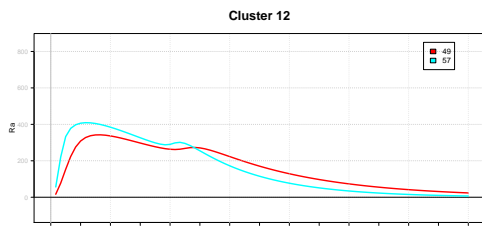
Comida	Nombre	Clasificación	CHO	Grasa	Proteína	Fibra	Energia_kCal.	Tiempo_min.
35	Standard breakfast	Medium	70.3	70.30	15.60	-	952.76	300
42	Oat cereal, milk, strawberry jam and orange juice breakfast	Medium	69.0	12.80	13.10	-	441.86	300
52	Wheat flakes, fresh cheese, milk	Large	91.0	26.00	28.00	-	660.73	360

% CHO	% Grasa	% Proteína	% Fibra	g	mean	sd	min	max	range	CV	
35	45	45	10	NA	156.2						
42	73	13	14	NA	94.9						
52	63	17	19	NA	144						
					% CHO	60.3	14.2	45	73	28	0.2
					% Grasa	25	17.4	13	45	32	0.7
					% Proteína	14.3	4.5	10	19	9	0.3
					% Fibra	NaN	NA	inf	-inf	NA	
					g	131.7	32.4	84.9	156.2	61.3	0.2



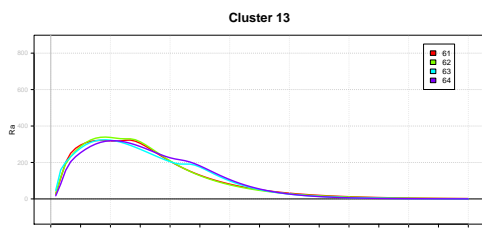
Comida	Nombre	Clasificación	CHO	Grasa	Proteína	Fibra	Energia_kCal.	Tiempo_min.
38	Spaghetti with tomato, cheese and lentils	Fast	87	9.00	25.00	-	527.86	120
39	Cornflakes, milk, bread and margarine	Fast	104	13.00	18.00	-	620.66	120
40	Oatmeal, milk, bread and margarine	Fast	62	17.00	12.00	-	452.61	120
41	Barley with tomato and cheese	Fast	68	8.00	12.00	-	384.10	120
43	Fiber and fruit cereal, milk, rockmelon, pineapple	Fast	50	2.60	10.20	-	265.12	120
44	High fiber cereal, milk, strawberries, grapefruit	Fast	42	2.50	11.00	-	234.07	120

% CHO	% Grasa	% Proteína	% Fibra	g	mean	sd	min	max	range	CV	
38	72	7	21	NA	121						
39	77	10	13	NA	135						
40	68	19	13	NA	91						
41	77	9	14	NA	86						
43	80	4	16	NA	62.8						
44	76	6	20	NA	55.5						
					% CHO	75	4.3	68	80	12	0.1
					% Grasa	9	5.4	4	19	15	0.6
					% Proteína	16.2	3.5	13	21	8	0.2
					% Fibra	NaN	NA	inf	-inf	NA	
					g	92.2	31.3	55.5	135	79.5	0.3



Comida	Nombre	Clasificación	CHO	Grasa	Proteína	Fibra	Energia_kCal.	Tiempo_min.
49	Candy bar, milk, ice-cream, heavy cream	Large	80	80.00	18.00	-	1111.11	360
57	White Rice	Medium	75	0.90	4.60	0.50	342.74	240

% CHO	% Grasa	% Proteína	% Fibra	g	mean	sd	min	max	range	CV	
49	45	45	10	NA	178						
57	93	1	6	1	81						
					% CHO	69	33.9	45	93	48	0.5
					% Grasa	23	31.1	1	45	44	1.4
					% Proteína	8	2.8	6	10	4	0.4
					% Fibra	1	NA	1	1	0	NA
					g	129.5	68.6	81	178	97	0.5



Comida	Nombre	Clasificación	CHO	Grasa	Proteína	Fibra	Energia_kCal.	Tiempo_min.
61	White Wheat Bread	Medium	50	3.00	8.40	3.10	266.79	180
62	Whole Kernel Bread	Medium	50	2.60	7.40	12.80	258.91	180
63	B-glucan Rye Bread	Medium	50	2.40	10.50	17.50	270.85	180
64	Whole Meal Pasta	Medium	50	4.70	12.10	5.60	298.56	180

% CHO	% Grasa	% Proteína	% Fibra	g	mean	sd	min	max	range	CV	
61	78	5	13	5	64.5						
62	69	4	10	18	72.8						
63	62	3	13	21	80						
64	69	6	17	8	72.4						
					% CHO	69.5	6.6	62	78	16	0.1
					% Grasa	4.5	1.3	3	6	3	0.3
					% Proteína	13.2	2.9	10	17	7	0.2
					% Fibra	13	7.7	5	21	16	0.6
					g	72.4	6.3	64.5	80	15.5	0.1

Figura 41: Clústering generat pel 14è resultat segons DI i Silueta (14a fila de la Taula 12 a la pàgina 42), part 2.

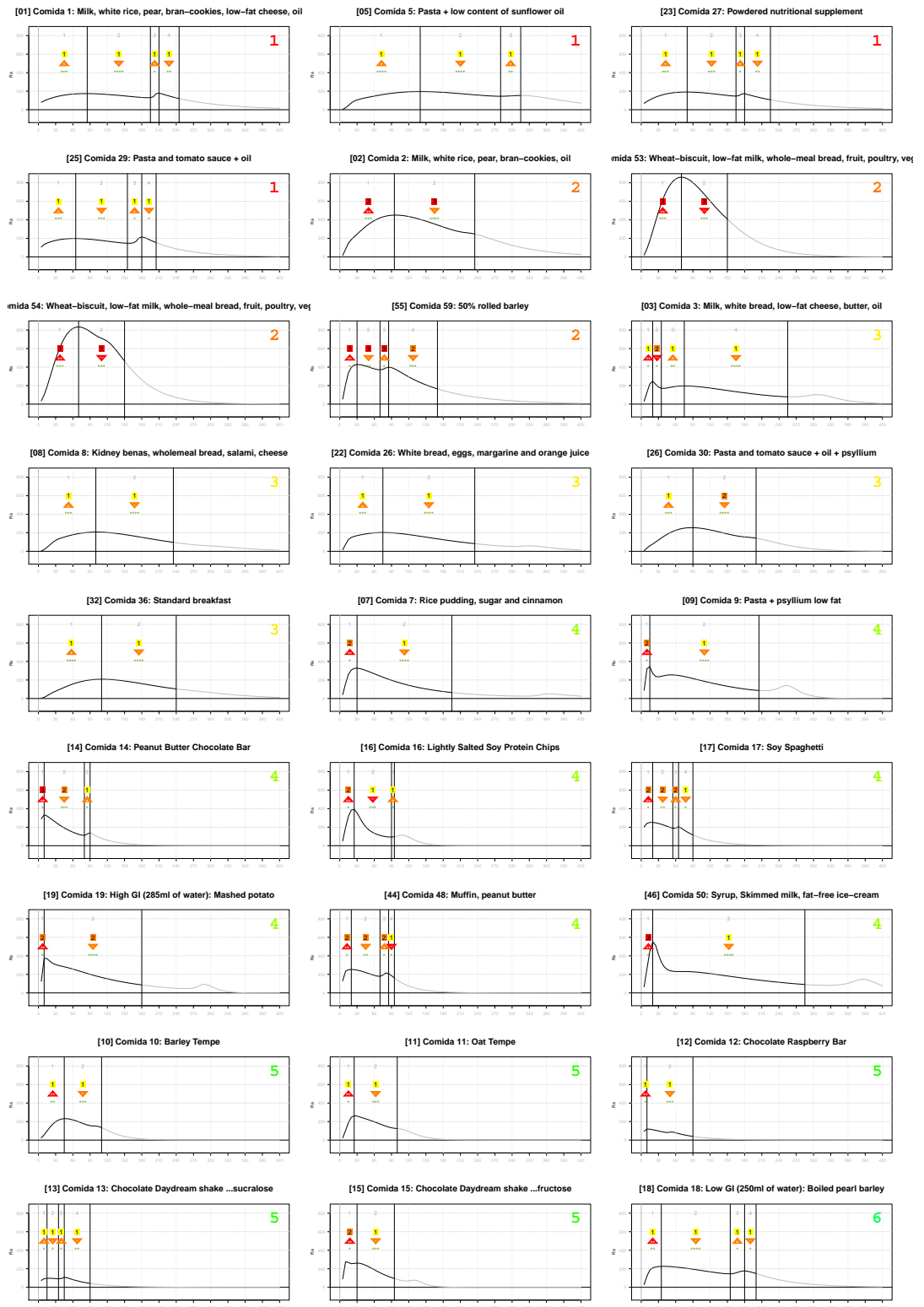


Figura 42: Descripció qualitativa del 14è resultat segons DI i Silueta (1a fila de la Taula 12 a la pàgina 42), part 1.



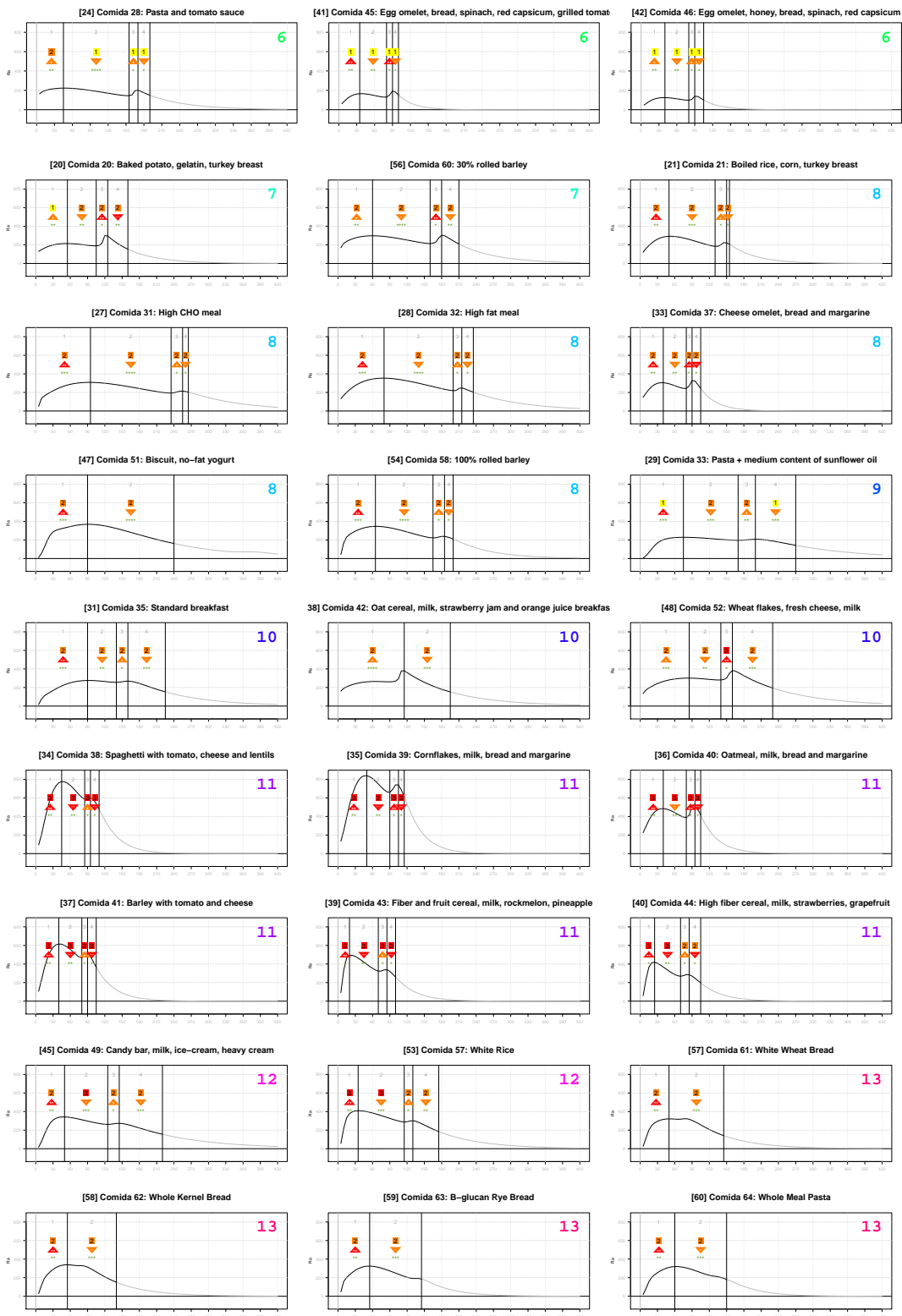


Figura 43: Descripció qualitativa del 14è resultat segons DI i Silueta (1a fila de la Taula 12 a la pàgina 42), part 2.

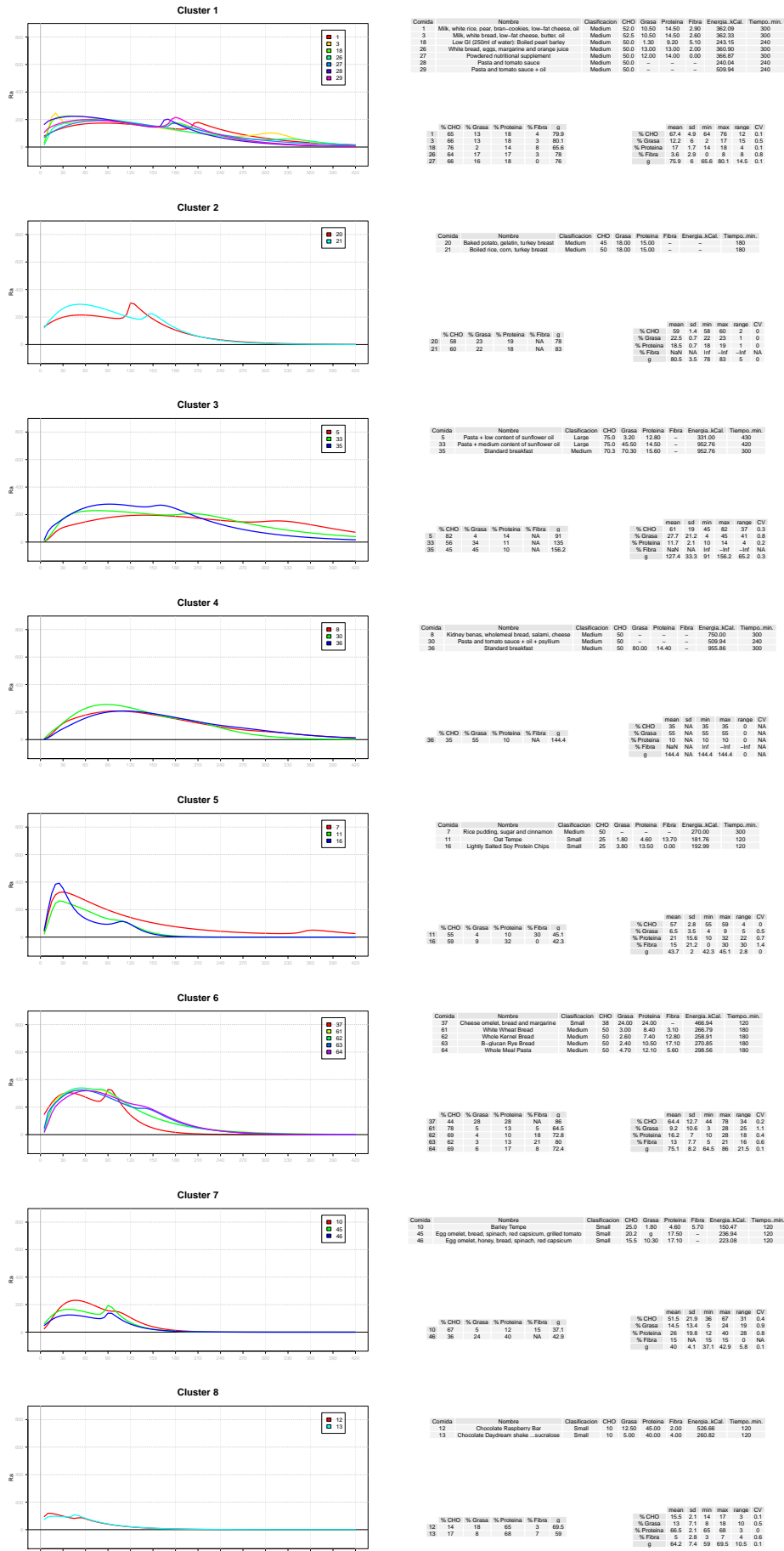
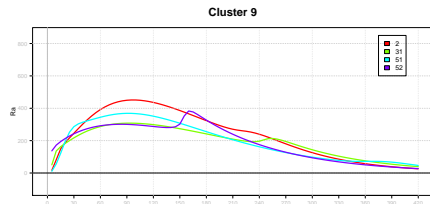


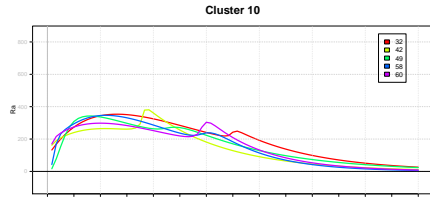
Figura 44: Clustering original del projecte ClosedLoop4Meals, part 1.



Comida	Nombre	Classificacio	CHO	Grasa	Proteina	Fibra	Energia_kCal	Temps.min
2	Milk, white rice, oat, bran-cookies, oil	Large	111	22.60	31.00	6.40	776.01	300
51	High-meat	Large	93	30.00	146.00	-	293.00	360
51	Biscuit, no-fat yogurt	Large	94	17.00	29.00	-	673.78	360
52	Wheat flakes, fresh cheese, milk	Large	91	25.00	28.00	-	665.79	360

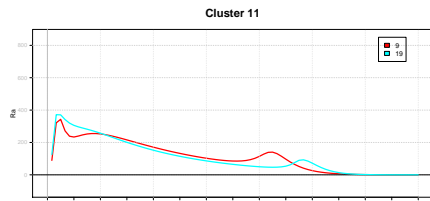
% CHO	% Grasa	% Proteina	% Fibra	g	mean	sd	min	max	range	CV
2	65	13	18	4	171	56.8	16.6	32	67	56
51	52	10	58	8	291	15	2.9	10	17	7
51	67	12	21	NA	140	% Proteina	27	15.4	18	50
52	63	17	19	NA	144	% Fibra	5	2.8	4	4
					g	186.5	71	140	291	151



Comida	Nombre	Classificacio	CHO	Grasa	Proteina	Fibra	Energia_kCal	Temps.min
32	High fat meal	Medium	27	60.00	135.00	18.00	2842.27	360
42	Oat cereal, milk, strawberry jam and orange juice breakfast	Medium	69	12.80	13.10	-	441.95	300
49	Candy bar, milk, ice-cream, heavy cream	Large	80	80.00	18.00	-	1111.11	360
58	100% rolled barley	Medium	75	1.70	7.20	9.30	361.85	240
60	30% rolled barley	Medium	75	1.10	5.40	3.20	348.72	240

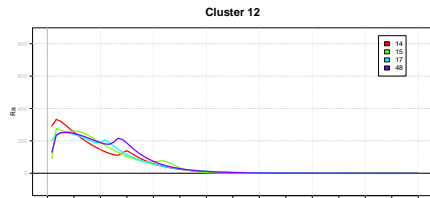
% CHO	% Grasa	% Proteina	% Fibra	g	mean	sd	min	max	range	CV
32	11	25	66	8	240	59.8	31.9	11	89	0.5
42	73	13	14	NA	94.9	% Grasa	17.2	18.3	1	45
49	45	45	10	NA	178	% Proteina	19.8	21	6	56
58	81	2	8	10	93.1	% Fibra	7.3	3.1	4	10
60	69	1	6	4	84.7	g	138.1	68.4	84.7	240



Comida	Nombre	Classificacio	CHO	Grasa	Proteina	Fibra	Energia_kCal	Temps.min
9	Pasta + coriolum low fat	Medium	52	-	-	1.70	240.00	240
19	High GI (285ml of water): Mashed potato	Medium	50	4.50	5.50	6.60	244.34	240

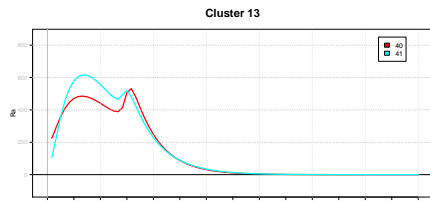
% CHO	% Grasa	% Proteina	% Fibra	g	mean	sd	min	max	range	CV
9	75	NA	75	0	NA	% CHO	75	NA	75	0
19	75	7	8	10	66.6	% Grasa	7	NA	7	0
					g	66.6	NA	66.6	0	NA



Comida	Nombre	Classificacio	CHO	Grasa	Proteina	Fibra	Energia_kCal	Temps.min
14	Peanut Butter Chocolate Bar	Small	25.0	4.80	13.00	0.80	194.18	120
15	Chocolate Daydream shake...fructose	Small	26.0	1.80	14.70	1.50	177.22	120
17	Soyl Sploshetti	Small	25.0	1.10	10.60	0.80	153.91	120
48	Muffin, peanut butter	Small	27.3	15.40	4.50	-	265.12	120

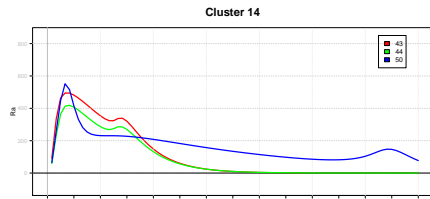
% CHO	% Grasa	% Proteina	% Fibra	g	mean	sd	min	max	range	CV
14	57	11	38	2	48.5	% CHO	69	4.7	67	9
15	58	4	34	3	43	% Grasa	12.8	14	3	33
17	67	3	28	2	37.5	% Proteina	14	7.1	9	19
48	58	33	10	NA	47.2	% Fibra	2.3	0.6	2	3
					g	42.8	4	37.6	47.2	2.7



Comida	Nombre	Classificacio	CHO	Grasa	Proteina	Fibra	Energia_kCal	Temps.min
40	Oatmeal, milk, bread and margarine	Fast	62	17.00	12.00	-	452.61	120
41	Barley with tomato and cheese	Fast	68	8.00	12.00	-	394.10	120

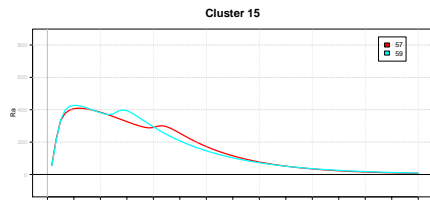
% CHO	% Grasa	% Proteina	% Fibra	g	mean	sd	min	max	range	CV
40	68	19	13	NA	91	% CHO	72.5	6.4	68	77
41	77	9	14	NA	88	% Grasa	14	7.1	9	19
					g	89.5	2.1	88	91	3



Comida	Nombre	Classificacio	CHO	Grasa	Proteina	Fibra	Energia_kCal	Temps.min
43	Fiber and fruit cereal, milk, rockmelon, pineapple	Fast	50	2.60	16.00	-	285.12	120
44	High fiber cereal, milk, strawberries, grapefruit	Fast	42	2.50	11.00	-	234.07	120
50	Syrup, Skimmed milk, fat-free ice-cream	Large	80	0.80	18.00	-	399.35	360

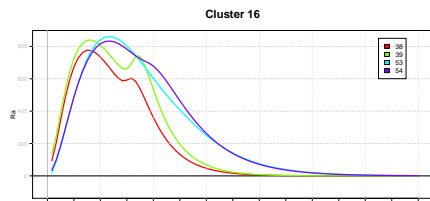
% CHO	% Grasa	% Proteina	% Fibra	g	mean	sd	min	max	range	CV
43	80	4	16	NA	62.8	% CHO	79	2.6	76	81
44	76	5	20	NA	55.5	% Grasa	3.3	2.1	1	5
50	81	1	18	NA	98.8	% Proteina	18	2	16	20
					g	72.4	23.2	55.5	98.8	43.3



Comida	Nombre	Classificacio	CHO	Grasa	Proteina	Fibra	Energia_kCal	Temps.min
57	White Rice	Medium	75	0.90	4.80	0.50	342.74	240
59	50% rolled barley	Medium	75	1.30	5.90	4.90	352.78	240

% CHO	% Grasa	% Proteina	% Fibra	g	mean	sd	min	max	range	CV
57	53	1	6	1	81	% CHO	89.5	4.9	86	93
59	66	1	7	6	87.1	% Grasa	1	0	1	0
					g	84	4.3	81	87.1	6.1

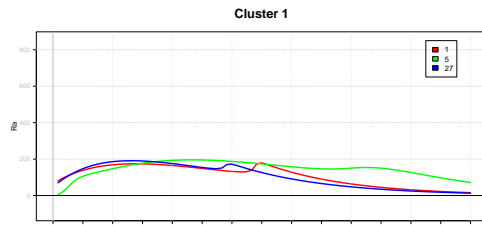


Comida	Nombre	Classificacio	CHO	Grasa	Proteina	Fibra	Energia_kCal	Temps.min
38	Spaghetti with tomato, cheese and lentils	Fast	67	9.00	25.00	-	527.85	120
39	Corntflakes, milk, bread and margarine	Fast	104	13.00	18.00	-	609.08	120
53	Wheat-biscuit, low-fat milk, whole-meal bread, fruit, poultry, vegetable	Large	123	20.00	27.00	-	679.78	180
54	Wheat-biscuit, low-fat milk, whole-meal bread, fruit, poultry, vegetable	Large	126	28.00	25.00	-	685.73	180

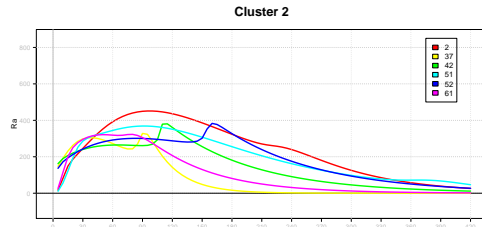
% CHO	% Grasa	% Proteina	% Fibra	g	mean	sd	min	max	range	CV
38	72	7	21	NA	121	% CHO	73	2.9	70	77
39	77	10	13	NA	135	% Grasa	12.5	4.8	7	17
53	73	17	10	NA	169	% Proteina	14.5	4.7	10	21
54	70	16	14	NA	179	% Fibra	NaN	NA	inf	-inf
					g	151	27.5	121	178	58

Figura 45: Clústering original del projecte ClosedLoop4Meals, part 2.



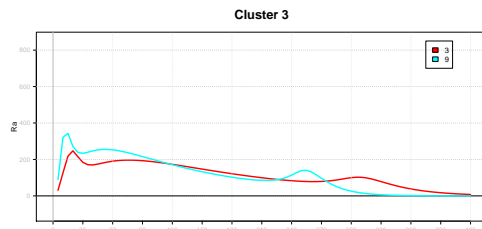
Comida	Nombre	Classificacio	CHO	Grasa	Proteina	Fibra	Energia_kCal	Tiempo_min.
1	Milk, white rice, pear, bran-cookies, low-fat cheese, oil	Medium	52	10,50	14,50	2,90	362,29	300
5	Pasta + low content of sunflower oil	Large	75	3,20	12,80	-	331,00	430
27	Powdered nutritional supplement	Medium	50	12,00	14,00	0,00	366,87	300

% CHO	% Grasa	% Proteina	% Fibra	g	mean	sd	min	max	range	CV	
1	65	13	18	4	79,9						
5	82	4	14	NA	91						
27	66	16	18	0	76						
					mean	sd	min	max	range	CV	
					% CHO	71	9,5	65	82	17	0,1
					% Grasa	11	6,2	4	16	12	0,6
					% Proteina	16,7	2,3	14	19	4	0,1
					% Fibra	2	2,8	0	4	4	1,4
					g	62,3	7,8	76	91	15	0,1



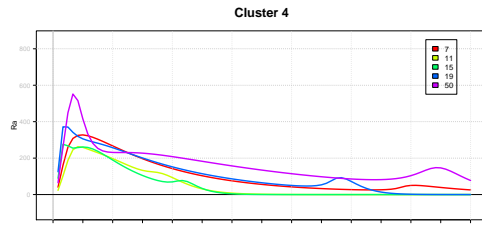
Comida	Nombre	Classificacio	CHO	Grasa	Proteina	Fibra	Energia_kCal	Tiempo_min.
2	Milk, white rice, pear, bran-cookies, oil	Large	111	22,80	31,00	6,40	776,01	300
37	Cheese omelet, bread and margarine	Small	38	24,00	24,00	-	466,94	120
42	Oat cereal, milk, strawberry jam and orange juice breakfast	Medium	69	12,80	13,10	-	441,86	300
51	Biscuit, no-fat yogurt	Large	94	17,00	29,00	-	673,78	360
52	Wheat flakes, fresh cheese, milk	Large	91	25,00	29,00	-	696,73	360
61	White Wheat Bread	Medium	50	3,00	8,40	3,10	266,79	180

% CHO	% Grasa	% Proteina	% Fibra	g	mean	sd	min	max	range	CV	
2	65	13	18	4	171						
37	44	28	28	NA	86						
42	73	13	14	NA	84,9						
51	67	12	21	NA	140						
52	63	17	19	NA	144						
61	78	5	13	5	64,5						
					mean	sd	min	max	range	CV	
					% CHO	65	11,7	44	78	34	0,2
					% Grasa	14,7	7,6	5	28	23	0,5
					% Proteina	18,8	5,4	13	28	15	0,3
					% Fibra	4,5	0,7	4	5	1	0,2
					g	116,7	40,9	64,5	171	106,5	0,4



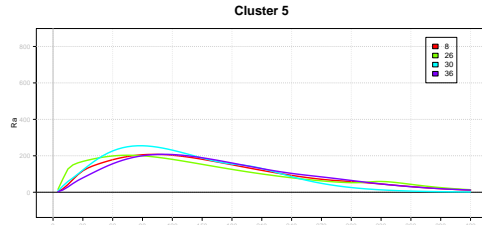
Comida	Nombre	Classificacio	CHO	Grasa	Proteina	Fibra	Energia_kCal	Tiempo_min.
3	Milk, white bread, low-fat cheese, butter, oil	Medium	52,5	10,50	14,50	2,60	362,33	300
9	Pasta + psyllium low fat	Medium	52,0		1,70		240,00	240

% CHO	% Grasa	% Proteina	% Fibra	g	mean	sd	min	max	range	CV	
3	66	13	18	3	80,1						
					mean	sd	min <td>max</td> <td>range</td> <td>CV</td>	max	range	CV	
					% CHO	66	NA	66	66	0	NA
					% Grasa	13	NA	13	13	0	NA
					% Proteina	18	NA	18	18	0	NA
					% Fibra	3	NA	3	3	0	NA
					g	80,1	NA	80,1	80,1	0	NA



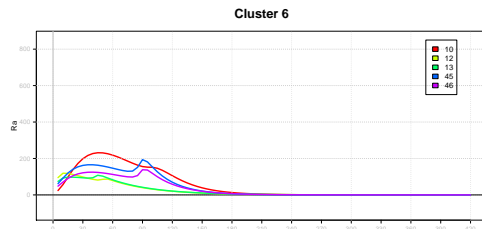
Comida	Nombre	Classificacio	CHO	Grasa	Proteina	Fibra	Energia_kCal	Tiempo_min.
7	Rice pudding, sugar and cinnamon	Medium	50				270,00	300
11	Oat Tempe	Small	25	1,80	4,60	13,70	161,76	120
15	Chocolate Daydream shake ...fructose	Small	25	1,80	14,70	1,50	177,22	120
19	High GI (26% of water): Mashed potato	Medium	50	4,50	9,50	6,80	244,54	240
50	Syrup, Steamed milk, fat-free ice-cream	Large	80	0,80	10,00	-	393,35	360

% CHO	% Grasa	% Proteina	% Fibra	g	mean	sd	min	max	range	CV	
7	55	4	10	30	45,1						
11	58	4	34	3	43						
15	58	4	34	3	43						
19	75	7	9	10	66,6						
50	81	1	18	NA	98,8						
					mean	sd	min	max	range	CV	
					% CHO	67,2	12,7	55	81	26	0,2
					% Grasa	4	2,4	1	7	6	0,6
					% Proteina	17,5	11,8	8	34	26	0,7
					% Fibra	14,3	14	3	30	27	1
					g	63,4	26,8	43	98,8	55,8	0,4



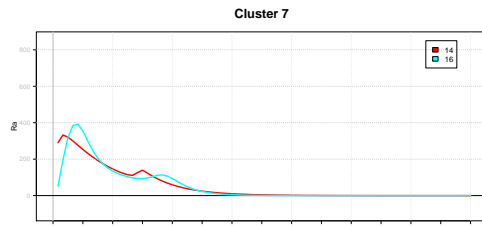
Comida	Nombre	Classificacio	CHO	Grasa	Proteina	Fibra	Energia_kCal	Tiempo_min.
8	Kidney beans, wholemeal bread, salami, cheese	Medium	50				750,00	300
26	White bread, eggs, margarine and orange juice	Medium	50	13,00	13,00	2,00	360,90	300
30	Pasta and tomato sauce + oil + psyllium	Medium	50				509,94	240
36	Standard breakfast	Medium	50	80,00	14,40	-	605,66	300

% CHO	% Grasa	% Proteina	% Fibra	g	mean	sd	min	max	range	CV	
26	64	17	17	3	78						
36	35	55	10	NA	144,4						
					mean	sd	min	max	range	CV	
					% CHO	49,5	20,9	35	64	29	0,4
					% Grasa	36	26,9	17	55	38	0,7
					% Proteina	13,5	4,9	10	17	7	0,4
					% Fibra	3	NA	3	3	0	NA
					g	111,2	47	78	144,4	66,4	0,4



Comida	Nombre	Classificacio	CHO	Grasa	Proteina	Fibra	Energia_kCal	Tiempo_min.
10	Banley Tempe	Small	25,0	1,80	4,60	5,70	150,47	120
12	Chocolate Raspberry Bar	Small	10,0	12,50	45,00	2,00	526,66	120
13	Chocolate Daydream shake ...sucrose	Small	10,0	5,00	40,00	4,00	260,82	120
45	Egg omelet, bread, spinach, red capsicum, oriled tomato	Small	20,2	g	17,50	-	236,94	120
46	Egg omelet, honey, bread, spinach, red capsicum	Small	15,5	10,30	17,10	-	223,08	120

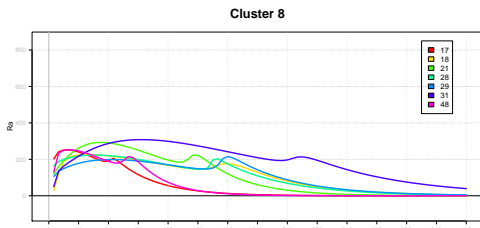
% CHO	% Grasa	% Proteina	% Fibra	g	mean	sd	min	max	range	CV	
10	67	5	12	15	37,1						
12	14	18	65	3	69,5						
13	17	8	68	7	59						
46	36	24	40	NA	42,9						
					mean	sd	min	max	range	CV	
					% CHO	33,5	24,4	14	67	53	0,7
					% Grasa	13,8	9,8	5	24	19	0,6
					% Proteina	46,2	26,1	12	68	56	0,6
					% Fibra	6,3	6,1	3	15	12	0,7
					g	52,1	14,8	37,1	69,5	32,4	0,3



Comida	Nombre	Classificacio	CHO	Grasa	Proteina	Fibra	Energia_kCal	Tiempo_min.
14	Peanut Butter Chocolate Bar	Small	25	4,80	12,90	0,80	194,18	120
16	Lightly Salted Soy Protein Chnps	Small	25	3,80	13,50	0,00	192,99	120

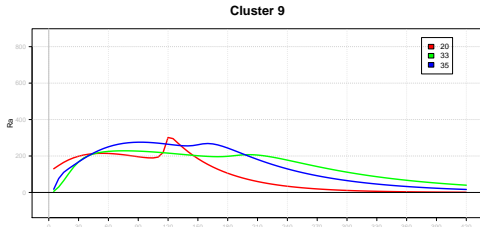
% CHO	% Grasa	% Proteina	% Fibra	g	mean	sd	min	max	range	CV	
14	57	11	30	2	43,5						
16	59	9	32	0	42,3						
					mean	sd	min	max	range	CV	
					% CHO	58	1,4	57	59	2	0
					% Grasa	10	1,4	9	11	2	0,1
					% Proteina	31	1,4	30	32	2	0
					% Fibra	1	1,4	0	2	2	1,4
					g	42,9	0,8	42,3	43,5	1,2	0

Figura 46: Clústering proper a l'original i de relativa qualitat (19a fila de la Taula 13 a la pàgina 46), part 1.



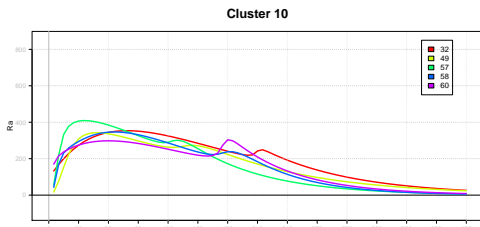
Comida	Nombre	Classificacio	CHO	Grasa	Proteina	Fibra	Energia_kCal	Tempo_min.
17	Soy Spaghetti	Small	25.0	1.10	10.00	0.80	151.91	120
18	Low GI (250ml of water): Boiled pearl barley	Medium	50.0	1.30	9.20	5.10	243.15	240
21	Boiled rice, corn, turkey breast	Medium	50.0	18.00	15.00	-	-	180
28	Pasta and tomato sauce	Medium	50.0	-	-	-	240.04	240
29	Pasta and tomato sauce + oil	Medium	50.0	-	-	-	509.94	240
31	High CHO meal	Large	93.0	30.00	146.00	22.00	2913.92	360
48	Muffin, peanut butter	Small	27.3	15.40	4.50	-	265.12	120

% CHO	% Grasa	% Proteina	% Fibra	g	mean	sd	min	max	range	CV	
17	67	3	28	2	37.5						
18	76	2	14	8	65.6						
21	60	22	18	NA	61						
28	32	10	50	8	291						
48	58	33	10	NA	47.2						
					mean	sd	min	max	range	CV	
					% CHO	68.6	16.5	32	76	44	0.3
					% Grasa	14	13.3	2	33	31	0.9
					% Proteina	24	16	10	60	40	0.7
					% Fibra	6	3.5	2	8	6	0.6
					g	104.9	105.5	37.5	291	253.5	1



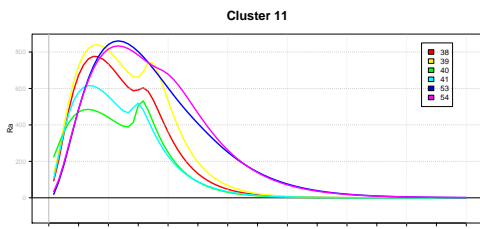
Comida	Nombre	Classificacio	CHO	Grasa	Proteina	Fibra	Energia_kCal	Tempo_min.
20	Baked potato, potato, turkey breast	Medium	45.0	18.00	15.00	-	-	180
33	Pasta + medium content of sunflower oil	Large	75.0	45.50	14.50	-	952.76	420
35	Standard breakfast	Medium	70.3	70.30	15.60	-	952.76	300

% CHO	% Grasa	% Proteina	% Fibra	g	mean	sd	min	max	range	CV	
20	58	23	19	NA	78						
33	56	34	11	NA	135						
35	45	45	10	NA	156.2						
					mean	sd	min	max	range	CV	
					% CHO	53	7	45	68	13	0.1
					% Grasa	34	11	23	45	22	0.3
					% Proteina	13.3	4.9	10	19	9	0.4
					% Fibra	NaN	NA	Inf	Inf	NA	
					g	123.1	40.4	78	156.2	78.2	0.3



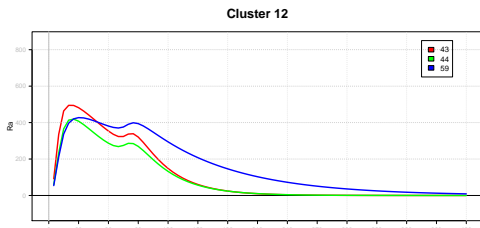
Comida	Nombre	Classificacio	CHO	Grasa	Proteina	Fibra	Energia_kCal	Tempo_min.
32	High fat meal	Medium	27	60.00	135.00	18.00	2642.27	360
49	Candy bar, milk, ice-cream, heavy cream	Large	80	80.00	18.00	-	1111.11	360
57	White Rice	Medium	75	0.90	4.60	0.50	342.74	240
58	100% rolled barley	Medium	75	1.70	7.20	9.20	351.85	240
60	30% rolled barley	Medium	75	1.10	5.40	3.20	348.72	240

% CHO	% Grasa	% Proteina	% Fibra	g	mean	sd	min	max	range	CV	
32	11	25	56	8	240						
49	81	2	8	10	93.1						
57	93	1	6	1	81						
58	89	1	6	4	84.7						
					mean	sd	min	max	range	CV	
					% CHO	63.8	35.1	11	93	82	0.5
					% Grasa	14.8	19.8	1	45	44	1.3
					% Proteina	17.2	21.8	6	56	50	1.3
					% Fibra	5.8	4	1	10	9	0.7
					g	135.4	70.8	81	240	159	0.5



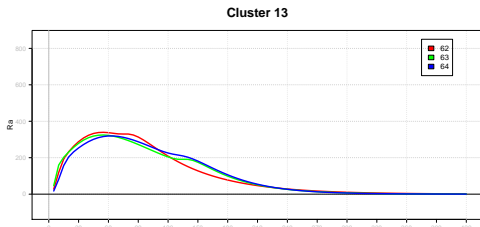
Comida	Nombre	Classificacio	CHO	Grasa	Proteina	Fibra	Energia_kCal	Tempo_min.
38	Spaghetti with tomato, cheese and lentils	Fast	87	9.00	25.00	-	527.85	120
39	Cornticles, milk, bread and margarine	Fast	104	13.00	16.00	-	602.06	120
40	Oatmeal, milk, bread and margarine	Fast	82	17.00	12.00	-	452.61	120
41	Barley with tomato and cheese	Fast	68	6.00	12.00	-	324.10	120
53	Wheat-biscuit, low-fat milk, whole-meal bread, fruit, poultry, vegetable	Large	123	29.00	17.00	-	673.78	180
54	Wheat-biscuit, low-fat milk, whole-meal bread, fruit, poultry, vegetable	Large	126	28.00	25.00	-	685.73	180

% CHO	% Grasa	% Proteina	% Fibra	g	mean	sd	min	max	range	CV	
38	72	7	21	NA	121						
39	77	10	13	NA	135						
40	68	19	13	NA	91						
41	77	9	14	NA	86						
53	73	17	10	NA	169						
54	70	16	14	NA	179						
					mean	sd	min	max	range	CV	
					% CHO	72.8	3.7	68	77	9	0.1
					% Grasa	13	4.9	7	19	12	0.4
					% Proteina	14.2	3.7	10	21	11	0.3
					% Fibra	NaN	NA	Inf	Inf	NA	
					g	130.5	38.2	88	179	91	0.3



Comida	Nombre	Classificacio	CHO	Grasa	Proteina	Fibra	Energia_kCal	Tempo_min.
43	Fiber and fruit cereal, milk, rockmelon, pineapple	Fast	50	2.60	10.20	-	265.12	120
44	High fiber cereal, milk, strawberries, grapefruit	Fast	42	2.50	11.00	-	234.07	120
59	30% rolled barley	Medium	75	1.30	5.90	4.90	352.78	240

% CHO	% Grasa	% Proteina	% Fibra	g	mean	sd	min	max	range	CV	
43	80	4	16	NA	62.8						
44	76	6	20	NA	55.5						
59	86	1	7	6	87.1						
					mean	sd	min	max	range	CV	
					% CHO	82.7	5	76	86	10	0.1
					% Grasa	3.3	2.1	1	5	4	0.6
					% Proteina	14.3	6.7	7	20	13	0.5
					% Fibra	6	NA	6	0	NA	
					g	68.5	16.5	55.5	87.1	31.6	0.2



Comida	Nombre	Classificacio	CHO	Grasa	Proteina	Fibra	Energia_kCal	Tempo_min.
62	Whole Kernel Bread	Medium	50	2.60	7.40	12.80	258.91	180
63	B-glucose Rye Bread	Medium	50	2.40	10.50	17.10	270.85	180
64	Whole Meal Pasta	Medium	50	4.70	12.10	5.60	298.56	180

% CHO	% Grasa	% Proteina	% Fibra	g	mean	sd	min	max	range	CV	
62	69	4	10	18	72.8						
63	62	3	13	21	80						
64	69	6	17	8	72.4						
					mean	sd	min	max	range	CV	
					% CHO	66.7	4	62	69	7	0.1
					% Grasa	4.3	1.5	3	6	3	0.4
					% Proteina	13.3	3.5	10	17	7	0.3
					% Fibra	15.7	6.8	8	21	13	0.4
					g	75.1	4.3	72.4	80	7.6	0.1

Figura 47: Clústering proper a l'original i de relativa qualitat (19a fila de la Taula 13 a la pàgina 46), part 2.

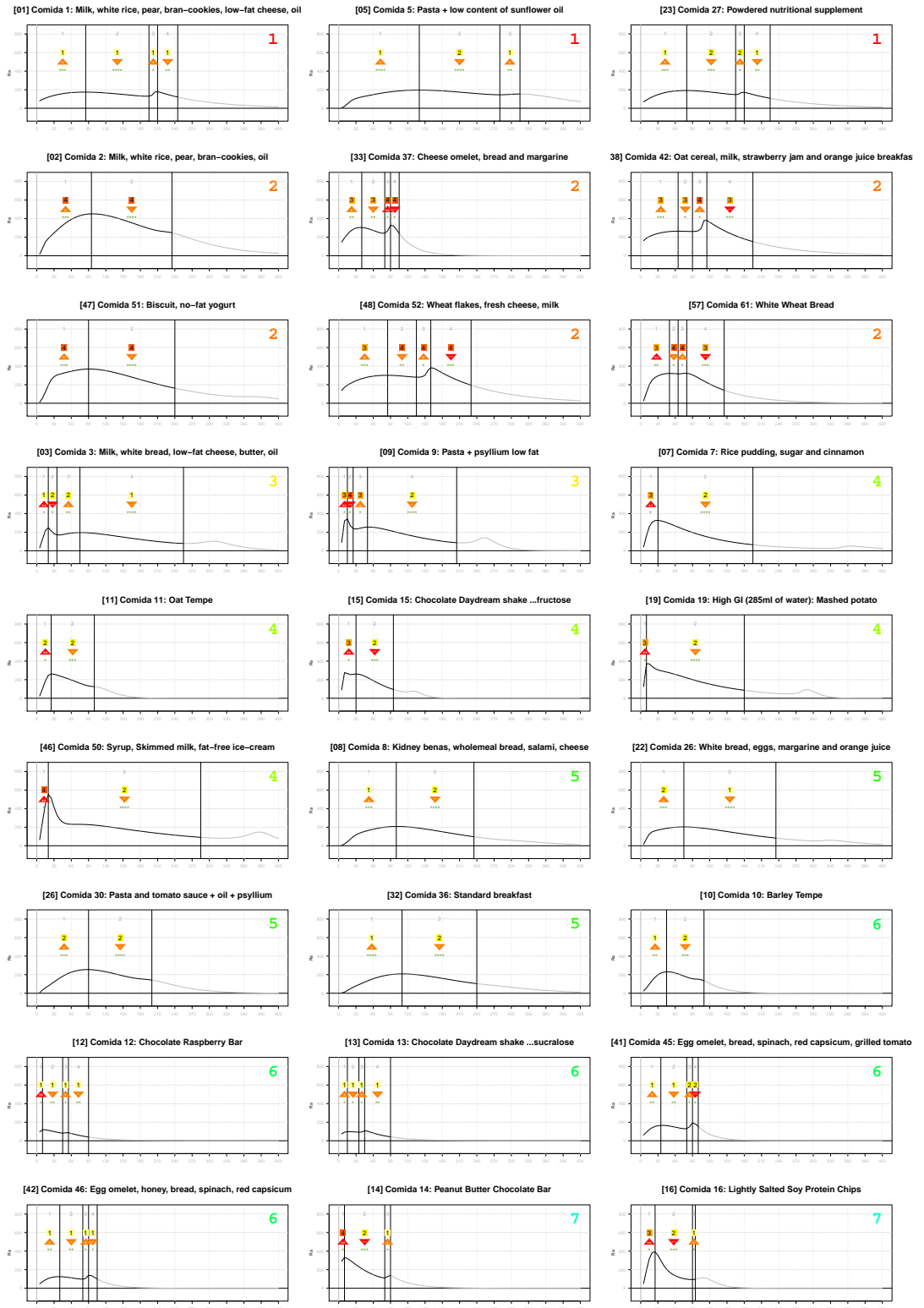


Figura 48: Descripció qualitativa del clústering elegit (19a fila de la Taula 13 a la pàgina 46), part 1.

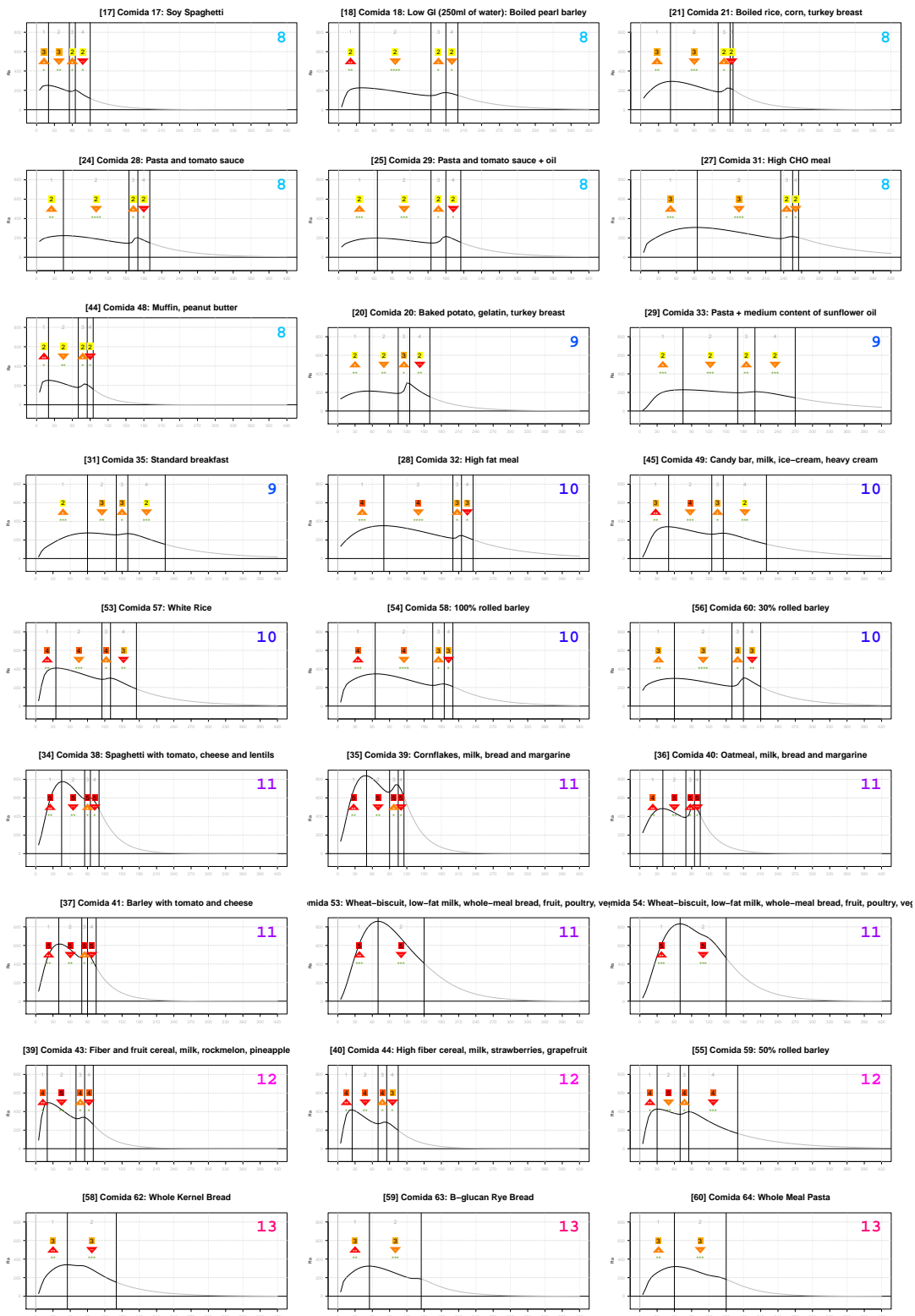


Figura 49: Descripció qualitativa del clústering elegit (19a fila de la Taula 13 a la pàgina 46), part 2.





## BIBLIOGRAFIA

---

- [1] *IDF Diabetes Atlas*. International Diabetes Federation, Brussels, Belgium, 6th edition, 2013. ISBN 2-930229-85-3. <http://www.idf.org/diabetesatlas>. (Cited on pages 3, 10, and 11.)
- [2] LG Abril and JMG Ruíz. Una metodología para la construcción de histogramas. Aplicación a los ingresos de los hogares andaluces. *asepelt.org*, 2000. URL <http://www.asepelt.org/ficheros/File/Anales/2000-Oviedo/Trabajos/PDF/207.pdf>. (Cited on page 27.)
- [3] J Bondía, J Vehí, C C Palerm, and P Herrero. El páncreas artificial: control automático de infusión de insulina en diabetes mellitus tipo 1. *Revista Iberoamericana de Automática e Informática Industrial RIAI*, 7(2):5–20, 2010. (Cited on page 12.)
- [4] Guy Brock, Vasyl Pihur, Susmita Datta, and Somnath Datta. clvalid: An r package for cluster validation. *Journal of Statistical Software*, 25(4):1–22, 3 2008. ISSN 1548-7660. URL <http://www.jstatsoft.org/v25/i04>. (Cited on pages 18 and 34.)
- [5] Phyllis A Crapo, Gerald Reaven, and Jerrold Olefsky. Plasma glucose and insulin responses to orally administered simple and complex carbohydrates. *Diabetes*, 25(9):741–747, 1976. (Cited on page 60.)
- [6] J. C. Dunn. A fuzzy relative of the isodata process and its use in detecting compact well-separated clusters. *Journal of Cybernetics*, 3(3):32–57, 1973. doi: 10.1080/01969727308546046. (Cited on page 17.)
- [7] Martin Ester, Hans peter Kriegel, Jörg S, and Xiaowei Xu. A density-based algorithm for discovering clusters in large spatial databases with noise. pages 226–231. AAAI Press, 1996. (Cited on page 16.)
- [8] Zoe Falomir, Lledó Museros, Luis Gonzalez-Abril, and Francisco Velasco. Measures of similarity between qualitative descriptions of shape, colour and size applied to mosaic assembling. *Journal of Visual Communication and Image Representation*, 24(3): 388 – 396, 2013. ISSN 1047-3203. doi: <http://dx.doi.org/10.1016/j.jvcir.2013.01.013>. URL <http://www.sciencedirect.com/science/article/pii/S1047320313000205>. (Cited on page 31.)
- [9] Kenneth D. Forbus. *Handbook of knowledge representation*, chapter Qualitative Modeling, pages 361–394. Elsevier, 2008. (Cited on page 14.)

- [10] Julia Handl, Joshua Knowles, and Douglas B Kell. Computational cluster validation in post-genomic data analysis. *Bioinformatics (Oxford, England)*, 21(15):3201–12, August 2005. ISSN 1367-4803. doi: 10.1093/bioinformatics/bti517. URL <http://www.ncbi.nlm.nih.gov/pubmed/15914541>. (Cited on page 17.)
- [11] Pau Herrero, Jorge Bondia, Cesar C Palerm, Josep Vehí, Pantelis Georgiou, Nick Oliver, and Christofer Toumazou. A simple robust method for estimating the glucose rate of appearance from mixed meals. *Journal of diabetes science and technology*, 6(1):153–162, 2012. (Cited on pages 21 and 22.)
- [12] Srinivas Laxminarayan, Jaques Reifman, and Garry M Steil. Use of a food and drug administration-approved type 1 diabetes mellitus simulator to evaluate and optimize a proportional-integral-derivative controller. *Journal of diabetes science and technology*, 6(6):1401–1412, 2012. (Cited on page 10.)
- [13] Y. Leal. *Enfoques metodológicos para mejorar la precisión y fiabilidad de los sistemas de monitorización continua de glucosa en pacientes con diabetes tipo 1 y en pacientes críticos mediante técnicas lineales y no lineales*. Universitat de Girona, Catalunya, 2013. <http://www.tdx.cat/handle/10803/120377>. (Cited on pages 3, 10, and 13.)
- [14] J. MacQueen. Some methods for classification and analysis of multivariate observations, 1967. URL <http://projecteuclid.org/euclid.bsm/1200512992>. (Cited on page 15.)
- [15] Fionn Murtagh. Complexities of hierarchic clustering algorithms: the state of the art. *Computational Statistics Quarterly*, 1(2):101–113, 1984. (Cited on pages 15 and 17.)
- [16] Yang Qu, Kyungmi Lee, and Ickjai Lee. Qualitative Cluster Reasoning Framework for Decision Making. *2009 Fifth International Joint Conference on INC, IMS and IDC*, pages 1321–1326, 2009. doi: 10.1109/NCM.2009.79. URL <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5331444>. (Cited on page 15.)
- [17] Peter J. Rousseeuw. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20(0):53 – 65, 1987. ISSN 0377-0427. doi: [http://dx.doi.org/10.1016/0377-0427\(87\)90125-7](http://dx.doi.org/10.1016/0377-0427(87)90125-7). URL <http://www.sciencedirect.com/science/article/pii/0377042787901257>. (Cited on page 17.)
- [18] Wikipedia. Cluster analysis — Wikipedia, the free encyclopedia, 2014. URL [http://en.wikipedia.org/w/index.php?title=Cluster\\_analysis#Algorithms](http://en.wikipedia.org/w/index.php?title=Cluster_analysis#Algorithms). [Online; accessed 15-November-2014]. (Cited on page 15.)

- [19] G. Williams and J.C. Pickup. *Handbook of Diabetes*. Blackwell Publishing, 2004. (Cited on page 3.)