



SAPIENZA
UNIVERSITÀ DI ROMA

MASTER IN CALCOLO SCIENTIFICO
A.A. 2016/2017

IDENTIFICAZIONE E IMPLEMENTAZIONE
DI UN METODO DI CLUSTERING PER LO STUDIO
DEI CAMBIAMENTI CLIMATICI IN ITALIA.

Tutor interno
Prof C. Cammarota

Tutor esterno
Dott.ssa P. Mercogliano

Candidato
Natalino Giuseppe

Introduzione

Scopo di questo lavoro di tesi è presentare un metodo di cluster analysis finalizzato a individuare delle aree omogenee in termini di proiezioni climatiche. In particolare, si usa l'approccio ibrido del *hk-means clustering* (dall'inglese *hierarchical k-means clustering*) per individuare aree del territorio nazionale omogenee in termini di variazioni climatiche attese per il periodo 2021-2050, rispetto al periodo di controllo 1981-2010, relativamente ad alcuni specifici indicatori climatici. Un'elevata risoluzione spaziale è ottenuta utilizzando il modello regionale RCM COSMO-CLM (Rockel and Geyer 2008) nella configurazione ottimizzata dal CMCC sull'Italia (Bucchignani et al. 2016; Zollo et al. 2016). Tale configurazione ha una risoluzione orizzontale di 8 km ed è forzata dal modello globale CMCC-CM con una risoluzione spaziale di 80 km. E' possibile, inoltre, valutare l'incertezza delle proiezioni climatiche confrontando i risultati di due scenari RCP (Representative Concentration Pethway): RCP4.5 e RCP8.5.

Applicando la tecnica di cluster analysis vengono individuate cinque aree omogenee, come numero ottimale di cluster. La conoscenza di queste aree potrebbe, ad esempio, aiutare a ottimizzare e a supportare successivi studi di impatti settoriali e analisi climatiche locali più mirate.

Nel primo capitolo vengono descritti i metodi applicati. Il secondo capitolo è dedicato al caso studio. Nel terzo capitolo si riportano l'idea base e i concetti matematici del *hk-means clustering* e la descrizione degli scenari.

Infine, per applicare il metodo sopra citato, è stato utilizzato il software statistico R.

Indice

Introduzione	2
1 Metodi applicati	6
1.1 Indicatori climatici selezionati	6
1.2 Tecnica della cluster analysis	15
2 Caso studio	18
2.1 Individuazione delle aree omogenee	18
2.2 Conclusioni	26
3 Appendice	27
3.1 K-Means Clustering	27
3.2 Clustering di tipo Ward	29
3.3 Gli scenari RCP4.5 e RCP8.5	29
Bibliografia	43

Elenco delle figure

1.1	Distribuzioni spaziali delle anomalie (COSMO RCP4.5)	10
1.2	Distribuzioni spaziali delle anomalie (COSMO RCP8.5)	11
1.3	Distribuzioni di frequenza delle anomalie (COSMO RCP4.5)	12
1.4	Distribuzioni di frequenza delle anomalie (COSMO RCP8.5)	12
1.5	Distribuzioni di frequenza delle anomalie dopo la pre-elaborazione(COSMO RCP4.5)	14
1.6	Distribuzioni di frequenza delle anomalie dopo la pre-elaborazione (COSMO RCP8.5)	15
2.1	Indici numero di cluster	19
2.2	Mappa dei cluster individuati (COSMO RCP4.5)	20
2.3	Distribuzioni di frequenza delle anomalie all'interno dei cluster individuati (COSMO RCP4.5)	21
2.4	Mappa dei cluster individuati (COSMO RCP8.5)	23
2.5	Distribuzioni di frequenza delle anomalie all'interno dei cluster individuati (COSMO RCP8.5)	24

Elenco delle tabelle

1.1	Titoli, abbreviazioni e unità di misura degli indicatori considerati	7
1.2	Matrice di correlazione delle anomalie (COSMO RCP4.5)	13
1.3	Matrice di correlazione delle anomalie (COSMO RCP8.5)	13
2.1	Valori medi dei cluster individuati (COSMO RCP4.5)	19
2.2	Valori medi dei cluster individuati (COSMO RCP8.5)	25

Capitolo 1

Metodi applicati

1.1 Indicatori climatici selezionati

Gli *indicatori climatici* selezionati come input della cluster analysis sono i seguenti (si veda anche la Tabella 1.1):

- *temperatura media annuale*: media annuale della temperatura media giornaliera;
- *giorni di precipitazione intensa*: media annuale del numero di giorni con precipitazione giornaliera superiore ai 20 mm;
- *frost days*: media annuale del numero di giorni con temperatura minima al di sotto dei $0^{\circ}C$;
- *summer days*: media annuale del numero di giorni con temperatura massima maggiore di $29.2^{\circ}C$ (valore medio del 95° percentile della distribuzione delle temperature massime osservate tramite E-OBS per il periodo 1981-2010¹);
- *cumulata delle precipitazioni invernali*: cumulata delle precipitazioni nei mesi invernali (Dicembre, Gennaio, Febbraio);
- *cumulata delle precipitazioni estive*: cumulata delle precipitazioni nei mesi estivi (Giugno, Luglio, Agosto);

¹Il dataset E-OBS viene comunemente utilizzato in studi scientifici.

- *copertura nevosa*: media annuale del numero di giorni per cui l'ammontare di neve superficiale è maggiore di 1 cm;
- *evaporazione*: evaporazione cumulata annuale;
- *95° percentile della precipitazione*: 95° percentile della precipitazione .

Questi indicatori sono stati individuati nell'ESPON CLIMATE project (Schmidt-Thomé and Greiving, 2013) e rappresentano i principali impatti meteo-indotti, a scala europea, su ambiente naturale, costruito, patrimonio culturale, sfera sociale ed economica. In questo lavoro sono state apportate delle modifiche per tener conto delle caratteristiche climatiche del territorio nazionale (PNACC, 2017).

Tabella 1.1: Titoli, abbreviazioni e unità di misura degli indicatori considerati

Indicatore	Abbreviazione	Unità di misura
Temperatura media annuale	Tmean	(°C)
Giorni di precipitazione intense	R20	(giorni/anno)
Frost days	FD	(giorni/anno)
Summer days	SU95p	(giorni/anno)
Cumulata delle precipitazioni invernali	WP	(mm)
Cumulata delle precipitazioni estive	SP	(mm)
Copertura nevosa	SC	(giorni/anno)
Evaporazione	Evap	(mm/anno)
95 percentile della precipitazione	R95p	(mm)

Gli *indicatori del cambiamento climatico* (o *anomalie climatiche*) sono calcolati come differenza tra i valori degli indicatori selezionati di due periodi, uno futuro (2021-2050) e uno di riferimento (1981-2010), entrambi della durata di 30 anni². In particolare, i valori degli indicatori climatici sono ottenuti: per il periodo di riferimento, usando i dati simulati dal modello climatico regionale COSMO-CLM; per il periodo futuro, usando i dati simulati dal modello climati-

²L'intervallo temporale di 30 anni è ritenuto adeguato per la caratterizzazione sia dei valori medi che degli estremi delle variabili atmosferiche di interesse (IPCC 2013)

co regionale COSMO-CLM³ considerando due scenari IPCC: RCP4.5 e RCP8.5⁴. E', quindi, possibile anche una valutazione dell'incertezza delle proiezioni climatiche rispetto ai due scenari considerati. Inoltre, le anomalie climatiche relative agli indicatori *temperatura media annuale, giorni di precipitazione intensa, frost days, summer days, copertura nevosa*, sono misurate in termini assoluti; le anomalie climatiche relative agli indicatori *cumulata delle precipitazioni invernali, cumulata delle precipitazioni estive, evaporazione cumulata annuale, 95 percentili della precipitazione* sono, invece, misurate in termini relativi.

Le distribuzioni spaziali e le distribuzioni di frequenza delle anomalie impiegate per l'individuazione delle aree omogenee con scenario RCP4.5 sono illustrate rispettivamente nella Figura 1.1 e nella Figura 1.3, mentre con scenario RCP8.5 nella Figura 1.2 e nella Figura 1.4.

Esse evidenziano come la variabilità dell'anomalia di temperatura media sul territorio nazionale sia limitata rispetto alla variabilità che si registra per gli altri indicatori, indipendentemente dallo scenario considerato. La stessa limitata variazione in frequenza si ritrova per l'indicatore R20 per entrambi gli scenari. Le variazioni nei valori di precipitazione (sia media che estrema) mostrano una maggiore variabilità per segno (aumento o diminuzione) ed intensità.

Per quanto riguarda i *frost days*, entrambi gli scenari registrano variazioni unicamente negative; più marcate per lo scenario RCP8.5. L'indicatore R95p, invece, risulta avere per entrambi gli scenari una distribuzione della variazione simile ma con valori di frequenza più elevati per lo scenario RCP8.5.

In entrambi gli scenari, poi, gli indicatori associati alla temperatura indicano un generalizzato incremento dei valori: aumento nei valori medi e dei *summer days* (SU95p).

Con lo scenario RCP4.5 si osserva una riduzione generale delle precipitazioni estive (SP) ad esclusione del basso versante adriatico, una riduzione delle precipitazioni invernali sulle Alpi, sugli Appennini e in Calabria e nell'area centro-

³La configurazione usata del modello climatico regionale COSMO-CLM è quella ottimizzata dal CMCC (Bucchignani et al. 2016). Tale configurazione ha una risoluzione orizzontale di 8 km ed è stata forzata dal modello globale CMCC-CM con risoluzione orizzontale di 80 km (Scocci-marro et al 2011). Informazioni aggiuntive sulle caratteristiche del modello sono disponibili sul sito del consorzio europeo CLM Assembly che ne cura lo sviluppo

⁴Per ulteriori informazioni sui scenari e modelli si veda l'appendice

orientale della Sicilia, ed una riduzione complessiva dell'evaporazione su tutto il territorio, specie in parte della Puglia e in Basilicata, escludendo le Alpi (probabilmente associato all'incremento di temperatura e variazione della copertura nevosa). Per quanto riguarda l'indicatore R20 si registrano variazioni contenute nell'intero territorio nazionale ad eccezione delle Alpi, con dei picchi su quelle occidentali. Per quanto riguarda la copertura nevosa e i *frost days* si registra una loro diminuzione generale, specie sulle aree montane prevalentemente interessate da tali fenomeni.

Invece, lo scenario RCP8.5 proietta un aumento significativo delle precipitazioni estive sul basso versante adriatico, e un aumento complessivo nel centro-nord delle precipitazioni invernali e dell'evaporazione, a differenza del sud Italia dove si registra un comportamento opposto; per quanto concerne gli eventi estremi, è stimato un aumento generalizzato nella magnitudo (R95p) mentre le variazioni dell'indicatore R20 ricalcano quelle della precipitazione invernale con un aumento nell'Italia Centro-Settentrionale ed una riduzione nelle aree meridionali.

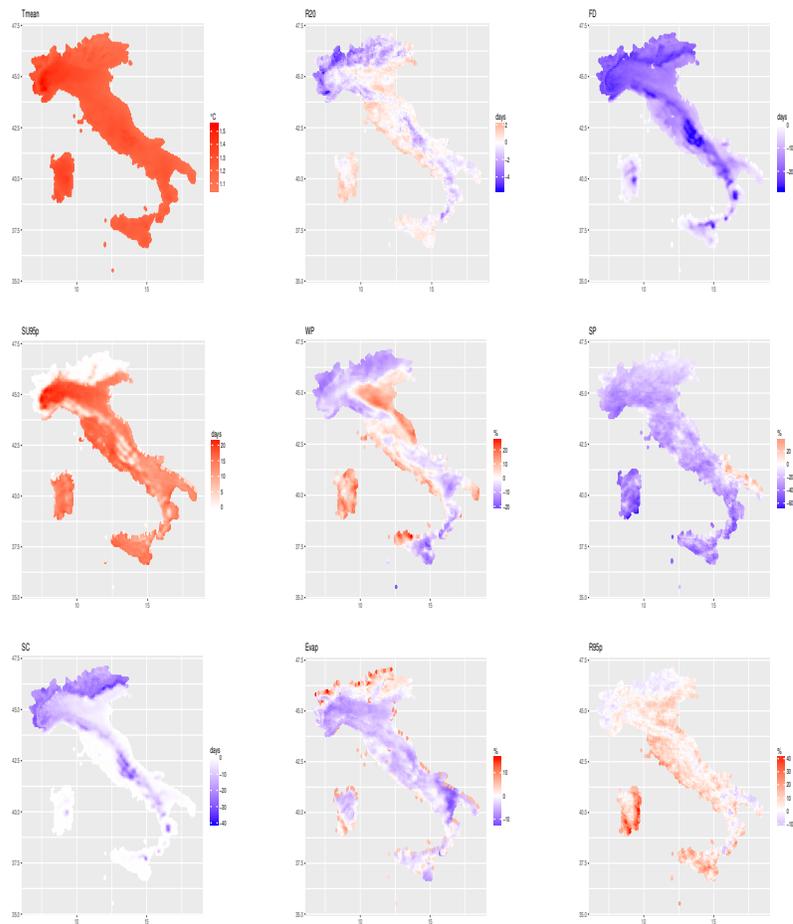


Figura 1.1: Distribuzioni spaziali delle anomalie (COSMO RCP4.5)

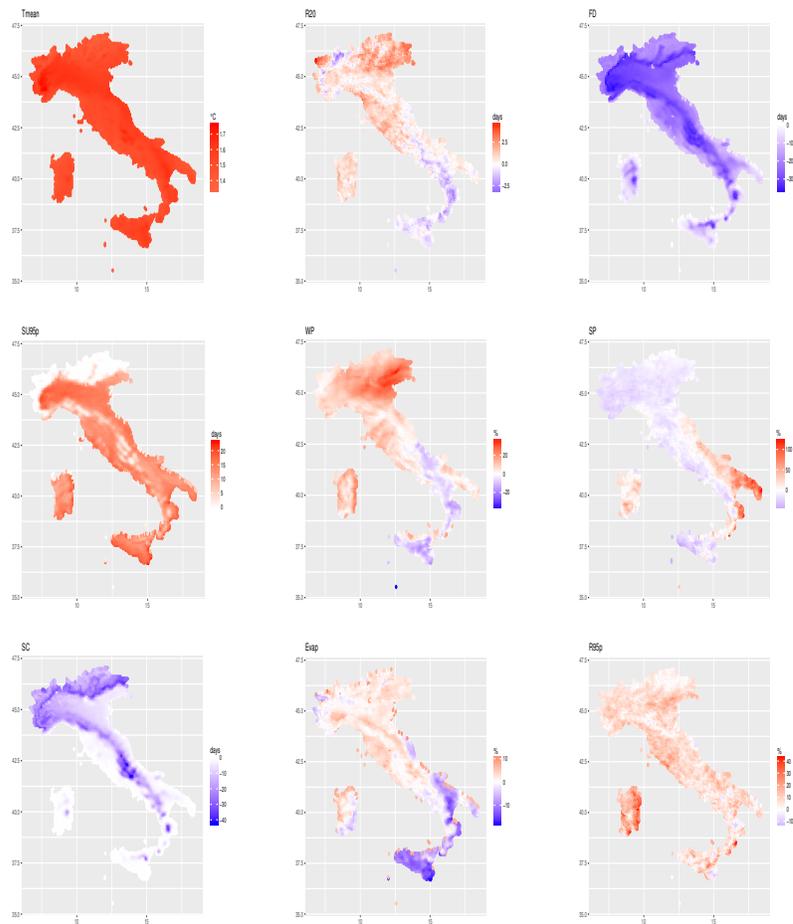


Figura 1.2: Distribuzioni spaziali delle anomalie (COSMO RCP8.5)

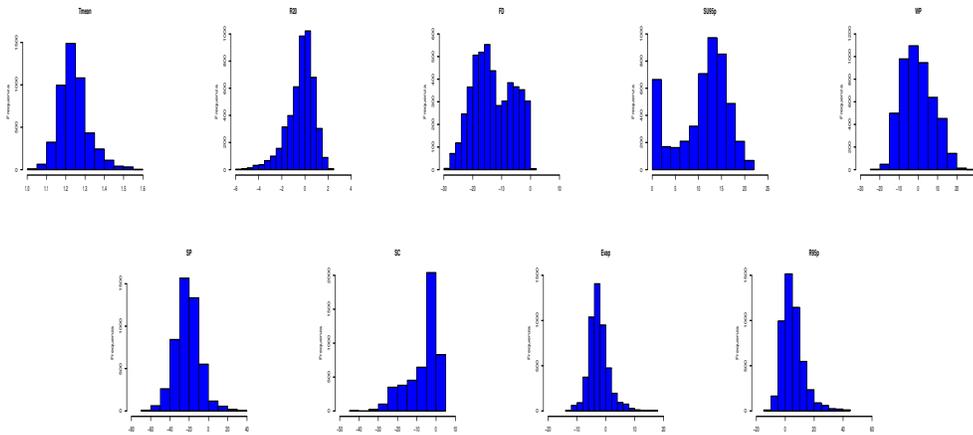


Figura 1.3: Distribuzioni di frequenza delle anomalie (COSMO RCP4.5)

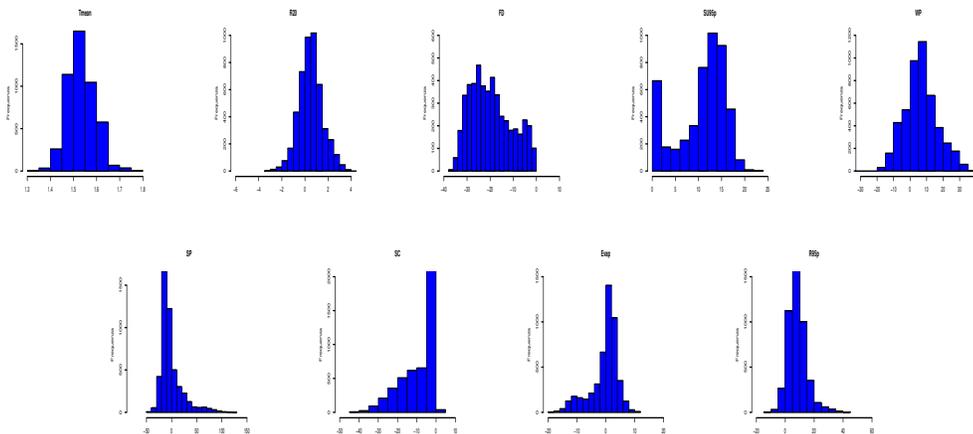


Figura 1.4: Distribuzioni di frequenza delle anomalie (COSMO RCP8.5)

Inoltre, vengono riportate le matrici di correlazioni delle anomalie: per lo scenario RCP4.5 nella Tabella 1.2; per lo scenario RCP8.5 nella Tabelle 1.3. In entrambi i casi, non si individua una correlazione significativa tra le anomalie considerate.

Tabella 1.2: Matrice di correlazione delle anomalie (COSMO RCP4.5)

	Tmean	R20	FD	SU95p	WP	SP	SC	Evap	R95p
Tmean	1.00	-0.17	-0.39	0.36	-0.06	-0.37	-0.21	-0.46	-0.01
R20	-0.17	1.00	0.52	0.53	0.60	-0.13	0.65	0.01	0.63
FD	-0.39	0.52	1.00	0.37	0.37	-0.08	0.78	0.11	0.46
SU95p	0.36	0.53	0.37	1.00	0.44	-0.28	0.66	-0.50	0.37
WP	-0.06	0.60	0.37	0.44	1.00	-0.18	0.50	0.01	0.53
SP	-0.37	-0.13	-0.08	-0.28	-0.18	1.00	-0.17	0.07	-0.32
SC	-0.21	0.65	0.78	0.66	0.50	-0.17	1.00	-0.19	0.47
Evap	-0.46	0.01	0.11	-0.50	0.01	0.07	-0.19	1.00	0.14
R95p	-0.01	0.63	0.46	0.37	0.53	-0.32	0.47	0.14	1.00

Tabella 1.3: Matrice di correlazione delle anomalie (COSMO RCP8.5)

	Tmean	R20	FD	SU95p	WP	SP	SC	Evap	R95p
Tmean	1.00	-0.22	-0.43	0.43	-0.05	-0.11	-0.23	-0.18	-0.07
R20	-0.22	1.00	-0.17	-0.22	0.64	-0.12	-0.08	0.43	0.32
FD	-0.43	-0.17	1.00	0.25	-0.34	0.43	0.71	-0.48	0.17
SU95p	0.43	-0.22	0.25	1.00	-0.16	0.13	0.62	-0.36	0.20
WP	-0.05	0.64	-0.34	-0.16	1.00	-0.25	-0.19	0.54	0.26
SP	-0.11	-0.12	0.43	0.13	-0.25	1.00	0.31	-0.21	0.14
SC	-0.23	-0.08	0.71	0.62	-0.19	0.31	1.00	-0.37	0.27
Evap	-0.18	0.43	-0.48	-0.36	0.54	-0.21	-0.37	1.00	0.10
R95p	-0.07	0.32	0.17	0.20	0.26	0.14	0.27	0.10	1.00

Infine, prima di applicare una cluster analysis ai nostri dati, è necessario effettuare una pre-elaborazione degli indicatori del cambiamento climatico che consiste nei seguenti passi:

1. *Winsorization*: gli outliers che si discostano notevolmente dai valori centrali

della distribuzione delle variabili vengono rimossi ⁵. Tutti i valori minore del percentile 0.1% sono sostituiti con tale percentile; tutti i valori superiori al 99,9% percentile sono sostituiti da questa soglia.

2. *Standardizzazione*: dopo la winsorization, il set di dati è standardizzato nell'intervallo $[0, 1]$, in modo da uniformare i valori degli indicatori del cambiamento climatico con differenti unità di misura (Milligan and Cooper 1988)

La Figura 1.5 e la Figura 1.6 mostrano, rispettivamente, per lo scenario RCP4.5 e lo scenario RCP8.5, le distribuzioni di frequenza dei valori delle anomalie dopo la pre-elaborazione.

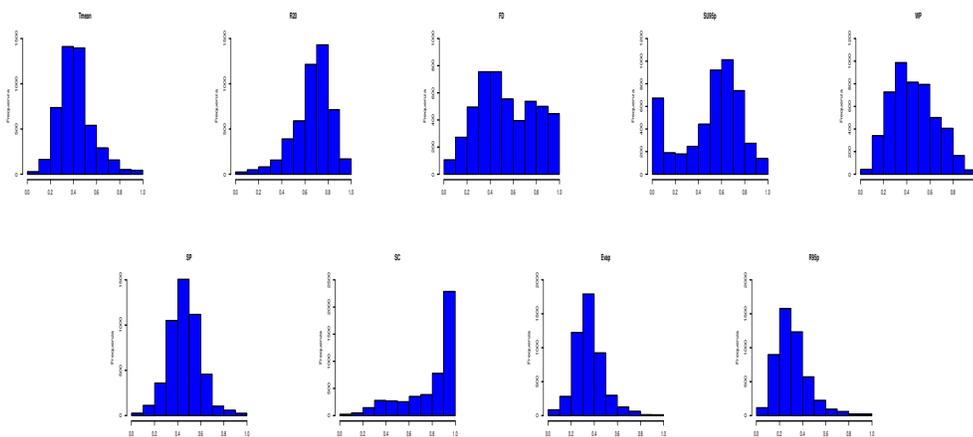


Figura 1.5: Distribuzioni di frequenza delle anomalie dopo la pre-elaborazione(COSMO RCP4.5)

⁵I valori anomali possono influenzare pesantemente la distribuzione di un indicatore.

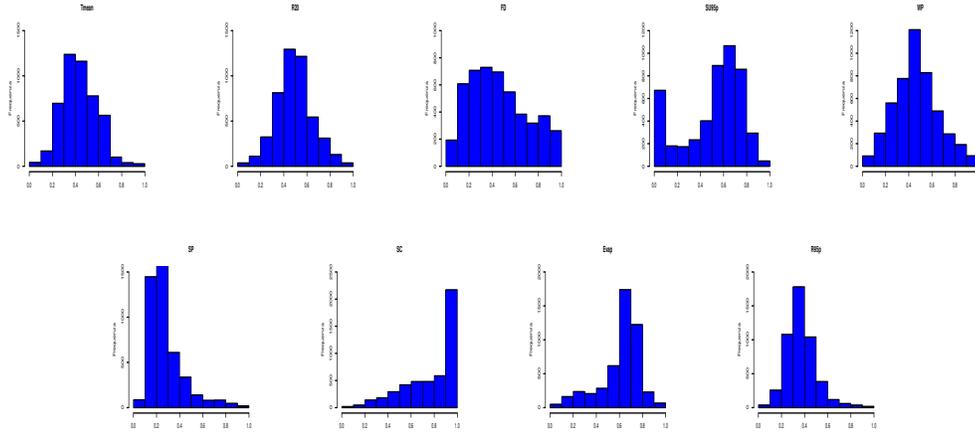


Figura 1.6: Distribuzioni di frequenza delle anomalie dopo la pre-elaborazione (COSMO RCP8.5)

1.2 Tecnica della cluster analysis

La cluster analysis classifica i dati in cluster (o gruppi) in modo tale che gli elementi appartenenti a ciascun gruppo siano piú simili tra loro rispetto agli elementi degli altri gruppi ⁶. I metodi di clustering possono essere distinti in clustering gerarchico e in clustering partizionale (Mooi and Sarstedt, 2011). La nostra analisi si basa principalmente sul secondo metodo. Il clustering gerarchico è usato solo per inizializzare il metodo partizionale. In particolare, viene applicato il k-means clustering⁷ (Macqueen, 1967; Hartigan and Wong, 1979). Il metodo k-means è un algoritmo molto semplice e veloce. Risulta vantaggioso in termini di onere computazionale qualora si debbano caratterizzare aree di vaste dimensioni. Tuttavia presenta alcuni punti deboli:

⁶In generale, la classificazione dei dati in gruppi richiede qualche metodo per calcolare la distanza o la (dis)similarità tra ciascuna coppia di dati. Ci sono diversi metodi per calcolare questa informazione sulla misura. Quello usato nel presente lavoro è la distanza euclidea:

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

dove x e y sono due vettori di lunghezza n

⁷L'idea base e i concetti chiave di matematica dei metodi sono riportati nell'Appendice.

1. Richiede all'analista di scegliere a priori il numero ottimale k di cluster.
2. Il risultato finale è sensibile alla selezione aleatoria iniziale dei centroidi.

Possibili soluzioni a questi punti deboli posso essere:

- a) Soluzione al problema 1: Non c'è una risposta definitiva alla questione. Per identificare il numero ottimale di cluster, si possono applicare diversi metodi di interpretazione e validazione della consistenza dei dati all'interno dei cluster con l'obiettivo di fornire una misura della coesione interna tra gli elementi di uno stesso cluster (compattezza) ed una misura della separazione tra un cluster e gli altri (isolamento).

Nel presente lavoro, sono utilizzati i seguenti metodi:

- *Elbow method*: considera la varianza totale intra-cluster come una funzione del numero di cluster; il numero ottimale di cluster è determinato dal punto di massimo locale di tale funzione.
- *Average Silhouette method*: valuta sia la distanza media di ciascun punto dagli altri punti del cluster di appartenenza che la distanza media dai punti del cluster più vicino; maggiore è il valore medio di tale indice più alta è la consistenza dell'analisi per il numero k di cluster corrispondenti (Kaufman and Rousseeuw, 1990);
- *Calinski-Harabasz method*: stima il rapporto tra la devianza tra i cluster e quella interna agli stessi; il numero ottimale di cluster è determinato dai massimi di tale indice (Calinski and Harabasz, 1974);

- b) Soluzione al problema 2: L'algoritmo è inizializzato da un clustering gerarchico agglomerativo definito come metodo di Ward (Ward,1963).

È stato dimostrato (Peterson, Ghosh and Maitra, 2010) che tale procedura è un valido metodo di inizializzazione e permette, quindi, di avere al passo iniziale una suddivisione delle classi non causale e/o arbitraria.

Pertanto, il metodo di clustering, implementato nel presente lavoro, può essere riassunto nei seguenti passi:

- Determinazione del numero ottimale k di cluster

- Calcolo del clustering gerarchico di tipo Ward e taglio del dendrogramma in k cluster
- Calcolo del centroide di ciascuno dei k cluster
- Calcolo del K-means clustering usando i centroidi, calcolati nel passo precedenti, come centri iniziali

Capitolo 2

Caso studio

2.1 Individuazione delle aree omogenee

Per individuare le aree del territorio nazionale omogenee in termini di anomalie climatiche abbiamo applicato il metodo di clustering implementato ai due insiemi di dati descritti nella sezione 1.1.

Come mostra la Figura 2.1, per entrambi gli scenari, sono stati individuati cinque cluster. La distribuzione geografica della tipologia delle aree dei cambiamenti climatici è mostrata nella Figura 2.2 per lo scenario RCP4.5 e nella Figura 2.4 per lo scenario RCP8.5. Inoltre, la Figura 2.3 e la Figura 2.5 mostrano la distribuzione delle anomalie all'interno delle singole classi, rispettivamente per lo scenario RCP4.5 e RCP8.5. Infine, La Tabella 2.1 restituisce i valori medi, in termini di anomalie, per i singoli cluster per lo scenario RCP4.5 e la Tabella 2.2 per lo scenario RCP8.5.

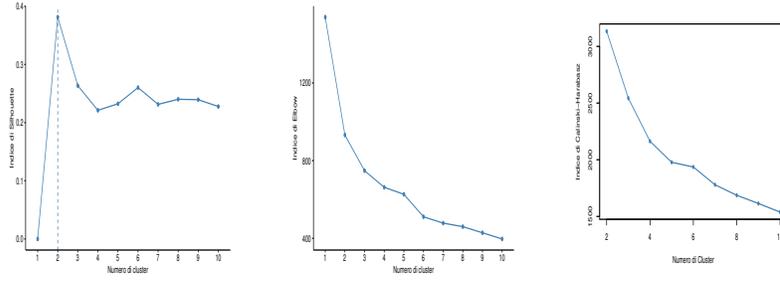


Figura 2.1: Indici numero di cluster

Tabella 2.1: Valori medi dei cluster individuati (COSMO RCP4.5)

	Tmean	R20	FD	SU95p	WP	SP	SC	Evap	R95p
Cluster 1	1.20	0.00	-6.00	12.00	-5.00	-18.00	-1.00	-3.00	4.00
Cluster 2	1.20	1.00	-9.00	14.00	8.00	-25.00	-1.00	-2.00	11.00
Cluster 3	1.30	-1.00	-19.00	9.00	-2.00	-24.00	-8.00	-3.00	3.00
Cluster 4	1.40	-1.00	-20.00	18.00	-4.00	-27.00	-12.00	-6.00	1.00
Cluster 5	1.20	-2.00	-20.00	1.00	-8.00	-15.00	-21.00	1.00	-1.00

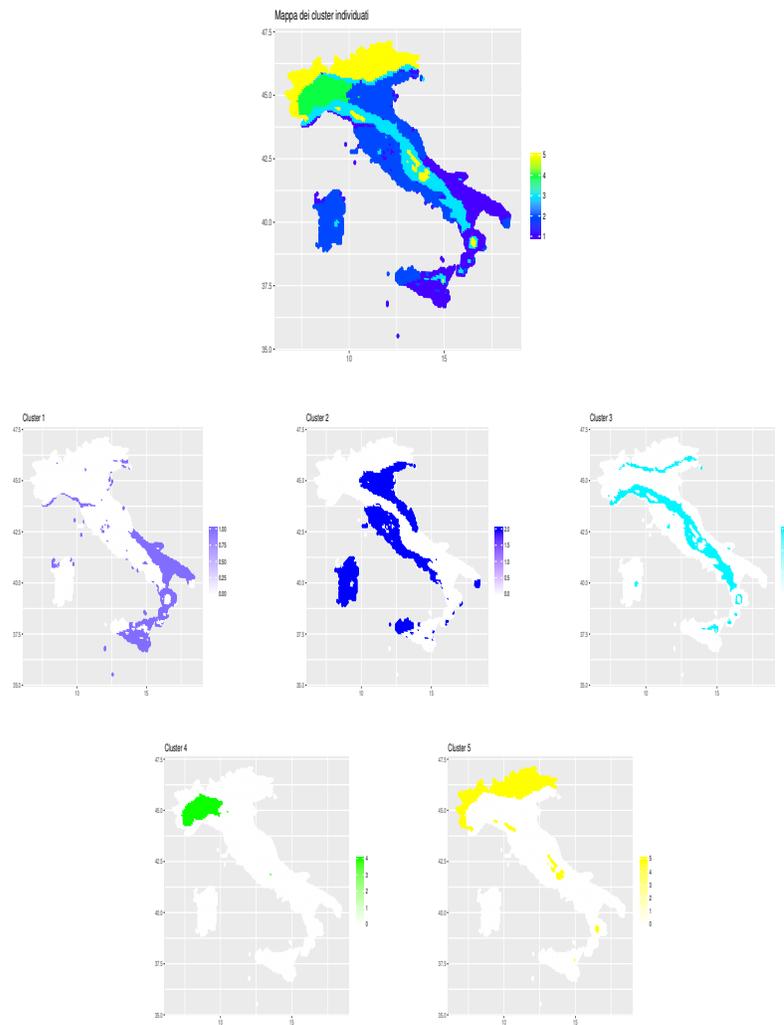


Figura 2.2: Mappa dei cluster individuati (COSMO RCP4.5)

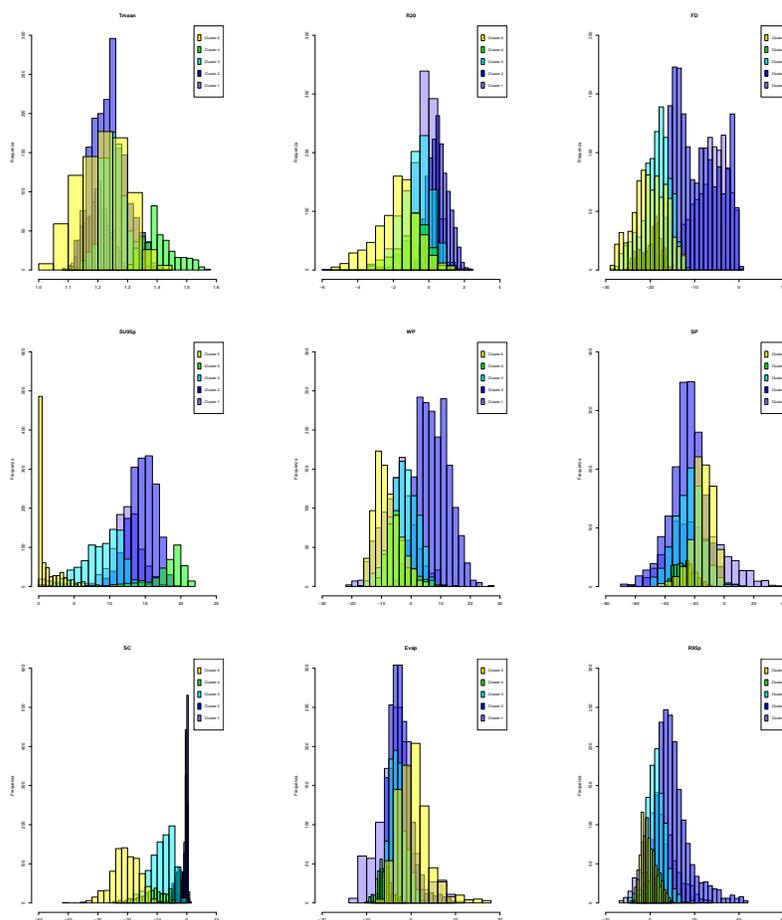


Figura 2.3: Distribuzioni di frequenza delle anomalie all'interno dei cluster individuati (COSMO RCP4.5)

Per lo scenario RCP4.5, l'analisi evidenzia le seguenti caratteristiche:

Cluster 1 (secco): questo cluster è caratterizzato da una riduzione delle precipitazioni estive e di quelle invernali; da un aumento moderato dei *summer days* (di 12 giorni/anno).

Cluster 2 (piovoso invernale-secco estivo): questo cluster è interessato da un aumento delle precipitazioni invernali (valore medio dell'aumento pari all'8%) e da una riduzione notevole di quelle estive (valore medio della riduzione pari al 25%). In generale si ha un aumento significativo sia dei fenomeni di precipitazione estremi (R95p) che dei *summer days* (di 14 giorni/anno).

Cluster 3 (caldo invernale- secco estivo): il cluster 3 è caratterizzato da una riduzione sia delle precipitazioni estive (valore medio della riduzione pari al 24%) che dei *frost days* (19 giorni/anno) e da una riduzione moderata della copertura nevosa (8 giorni/anno).

Cluster 4 (caldo - secco estivo): per questo cluster si osserva un aumento significativo dei *summer days* (18 giorni/anno) e una riduzione delle precipitazioni invernali e, soprattutto di quelle estive (valore medio della riduzione pari al 27%). Inoltre, presenta una riduzione rilevante anche dei *frost days*, della copertura nevosa e dell'evaporazione.

Cluster 5 (secco - caldo invernale): questo cluster è caratterizzato da una riduzione generale dei fenomeni di precipitazione. Inoltre, si osserva una riduzione significativa dei *frost days* (20 giorni/anno) e della copertura nevosa (21 giorni/anno)

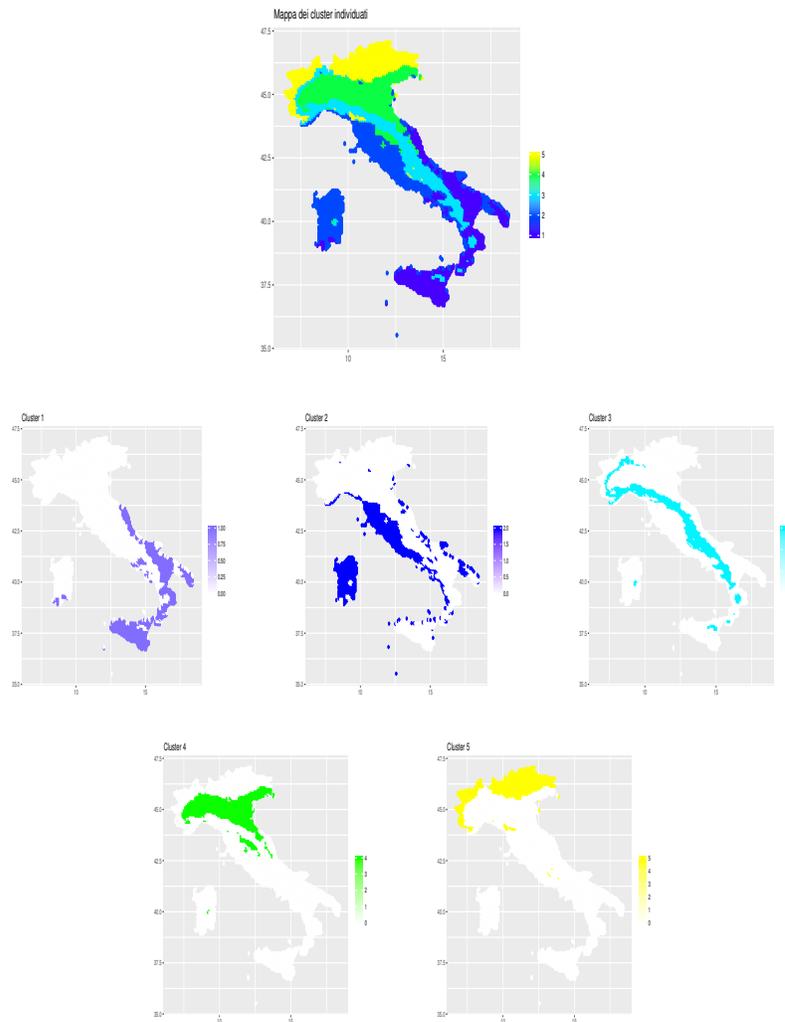


Figura 2.4: Mappa dei cluster individuati (COSMO RCP8.5)

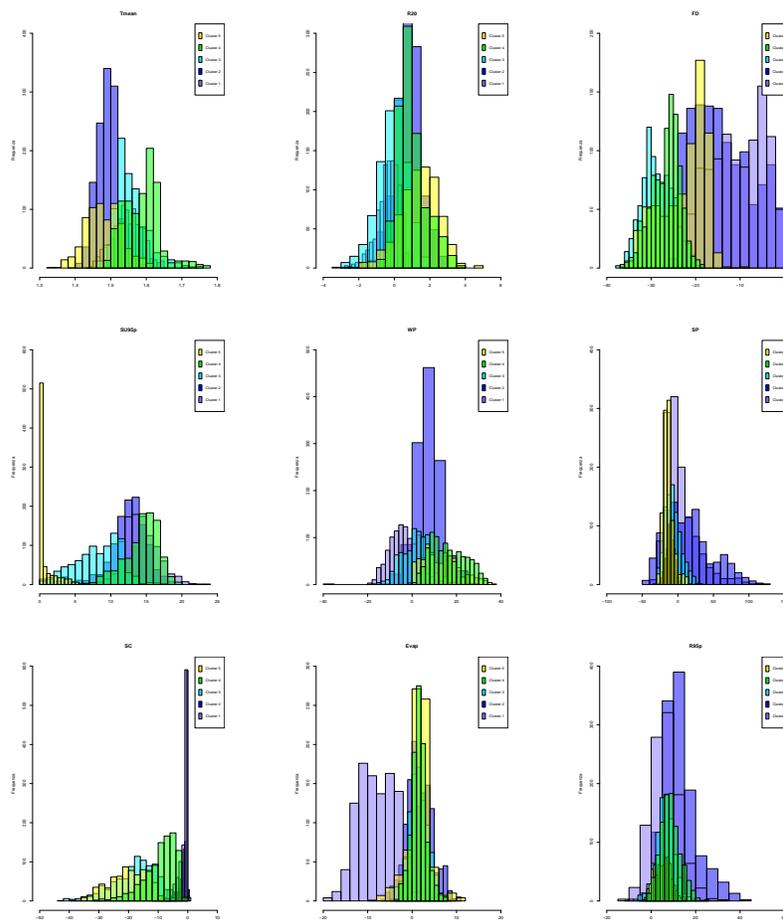


Figura 2.5: Distribuzioni di frequenza delle anomalie all'interno dei cluster individuati (COSMO RCP8.5)

Tabella 2.2: Valori medi dei cluster individuati (COSMO RCP8.5)

	Tmean	R20	FD	SU95p	WP	SP	SC	Evap	R95p
Cluster 1	1.50	0.00	-10.00	14.00	-4.00	14.00	-1.00	-8.00	6.00
Cluster 2	1.50	1.00	-14.00	12.00	7.00	3.00	-1.00	2.00	13.00
Cluster 3	1.60	0.00	-28.00	8.00	2.00	-7.00	-18.00	1.00	6.00
Cluster 4	1.50	1.00	-27.00	14.00	16.00	-14.00	-9.00	2.00	9.00
Cluster 5	1.50	1.00	-23.00	1.00	13.00	-11.00	-20.00	2.00	5.00

Per lo scenario RCP8.5 ,l'analisi evidenzia le seguenti caratteristiche:

Cluster 1 (secco invernale - caldo estivo): per questo cluster si osserva una complessiva riduzione delle precipitazioni invernali e un aumento rilevante di quelle estive; un aumento notevole dei *summer days* (14 giorni/anno); una riduzione complessiva dell'evaporazione (valore medio della riduzione pari al 8%).

Cluster 2 (piovoso-caldo estivo): questo cluster è caratterizzato da un aumento sia delle precipitazioni invernali che di quelle estive e da un aumento significativo dei fenomeni di precipitazione estremi (valore medio dell'aumento pari al 13%). Inoltre, si osserva un aumento rilevante dei *summer days* (12 giorni/anno).

Cluster 3 (caldo invernale): questo cluster è interessato da una riduzione significativa sia dei *frost days* (28 giorni/anno) che della copertura nevosa (18 giorni/anno). Inoltre, si osserva una riduzione moderata delle precipitazioni estive (valore medio della riduzione pari al 7%).

Cluster 4 (caldo-piovoso invernale-secco estivo): questo cluster è caratterizzato da un aumento significativo sia dei *summer days* (14 giorni/anno) che dei fenomeni di precipitazione estremi (valore medio dell'aumento pari al 9%). Inoltre, si osserva una rilevante riduzione delle precipitazioni estive (valore medio della riduzione pari al 14%) e un aumento significativo delle precipitazioni invernali (valore medio dell'aumento pari al 16%). Infine, il cluster 4 presenta anche una notevole riduzione dei *frost days* (27 giorni/anno).

Cluster 5 (piovoso invernale-secco estivo): il cluster 5 è caratterizzato da un aumento delle precipitazioni invernali (valore medio dell'aumento pari al 13%) e da una riduzione di quelle estive (valore medio della riduzione pari al 11%). Inoltre, si osserva una riduzione significativa sia dei *frost days* (23 giorni/anno) che

della copertura nevosa (20 giorni/anno).

2.2 Conclusioni

Lo studio fin qui condotto ha permesso di individuare, per entrambi gli scenari, cinque differenti aree omogenee in termini di proiezioni climatiche. Le due tipologie mostrano una somiglianza nell'estensione spaziale dei cluster 1,3 e 5; una differenza in quella dei cluster 2 e 4. Tra i due scenari, tuttavia, si evidenziano alcune differenze in termini di eventi estremi. Per lo scenario RCP8.5 si osserva un aumento significativo del 95° percentile della precipitazione (R95p) rispetto allo scenario RCP4.5. Inoltre, mentre le anomalie WP e SP per lo scenario RCP4.5 mostrano una riduzione (WP aumenta solo per il Cluster 2), nel caso dello scenario RCP8.5 si individuano zone in cui si osserva un aumento delle precipitazioni estive (Cluster 1 e 2) e di quelle invernali (Cluster 2,3,4,5).

In questo caso, l'utilizzo di un approccio multi-model permetterebbe di valutare l'incertezza nella previsione di cambiamento e la robustezza del cambiamento previsto dai diversi modelli.

Allo stato attuale, però, la procedura di cluster analysis richiede la selezione di un unico modello. Pertanto, sarebbero auspicabili sviluppi futuri nei criteri implementativi della cluster analysis al fine di tener conto della valutazione dell'incertezza associata alle anomalie climatiche.

Capitolo 3

Appendice

3.1 K-Means Clustering

Il K-means clustering (MacQueen,1967) è un metodo di clustering partizionale. I dati sono divisi in un insieme di k gruppi, dove k è il numero di cluster prestabilito dall'analista. Ogni cluster è rappresentato dal suo centro (o centroide) che corrisponde alla media degli elementi appartenenti al cluster.

L'idea alla base del K-means clustering consiste nel definire i cluster in modo tale da minimizzare la varianza totale intra-cluster.

Sono disponibili diversi K-means algoritmi. Quello usato nel presente lavoro è l' Hartigan-Wong algorithm (1979), che definisce la varianza totale intra-cluster come la somma dei quadrati delle distanze (distanza euclidea) tra gli elementi di un cluster e il relativo centroide. Posto, cioè,

$$W(C_j) = \sum_{l=1}^p \sum_{i \in C_j} (x_{il} - \mu_l^j)^2$$

in cui

- p è il numero degli indicatori
- x_{il} è il valore dell'indicatore l dell'oggetto i
- μ_l^j è il valore medio dell'indicatore l nel cluster C_j

la varianza totale intra-cluster è definita dalla seguente formula:

$$W = \sum_{j=1}^k W(C_j) = \sum_{j=1}^k \sum_{l=1}^p \sum_{i \in C_j} (x_{il} - \mu_l^j)^2$$

dove

- k rappresenta il numero di cluster.

Il suo valore deve essere il più piccolo possibile.

La prima fase, quando si usa un k-means clustering, è quella di indicare il numero k dei cluster che saranno generati nella soluzione finale.

L'algoritmo inizia selezionando, in modo aleatorio, k oggetti dall'insieme di dati da utilizzare come centroidi dei cluster iniziali.

Poi, ciascuno dei rimanenti oggetti è assegnato al centroide più vicino, secondo la metrica euclidea (fase di assegnazione del cluster). Dopo questa fase, l'algoritmo calcola i nuovi centroidi di ciascun cluster (fase di aggiornamento dei centroidi). Una volta che tutti i centri sono stati ricalcolati, ogni osservazione viene ricontrollata per vedere se potrebbe essere più vicina a un differente centroide. Tutti gli oggetti, allora, sono nuovamente assegnati usando i centroidi ricalcolati.

Le due fasi di assegnazione del cluster e di aggiornamento dei centroidi sono ripetute iterativamente fino a quando o è raggiunta la convergenza (i cluster ottenuti nella corrente iterazione sono gli stessi di quelli ottenuti nella precedente iterazione) oppure è raggiunto il numero massimo di iterazioni scelto.

L' Hartigan-Wong algorithm può essere, dunque, riassunto come segue:

1. Specificare il numero k dei cluster da generare
2. Selezionare in modo aleatorio k oggetti, dall'insieme dei dati, come centroidi dei cluster iniziali
3. Assegnare ciascuna osservazione al centroide più vicino, secondo la distanza euclidea.
4. Per ciascuno dei k cluster aggiornare i centroidi calcolando i nuovi valori medi di tutti gli elementi appartenente al cluster. Il centroide di un cluster è un vettore di lunghezza p contenente le medie di tutte le variabili per gli elementi appartenenti al cluster; p è il numero di variabili.

5. Iterativamente minimizzare la varianza totale intra-cluster. Cioè iterare le fasi 3 e 4 fino a quando è raggiunta la convergenza o è raggiunto il numero massimo di iterazioni scelto.

3.2 Clustering di tipo Ward

Il clustering di tipo Ward (Ward, 1963) è un metodo di clustering gerarchico agglomerativo¹. Inizialmente, viene calcolata la distanza euclidea tra ogni coppia di osservazioni. L'insieme dei dati viene così trasformato in una matrice delle distanze a partire dalla quale, usando una specifica funzione di collegamento, vengono unite coppie di oggetti/cluster in un nuovo e più grande cluster in base alla loro somiglianza. Il metodo di Ward realizza tale classificazione gerarchica minimizzando la varianza totale interna a ciascun cluster cioè ad ogni passo vengono combinati oggetti/cluster che rappresentano la minima variazione della varianza entro i gruppi.

Al termine dell'algoritmo si ottiene un dendrogramma che è possibile tagliare ad una determinata altezza per partizionare l'insieme dei dati in cluster.

3.3 Gli scenari RCP4.5 e RCP8.5

La simulazione climatica utilizzata nel presente lavoro, si basa sui seguenti elementi:

¹Il clustering gerarchico è un approccio alternativo al clustering partizionale. Non richiede di specificare il numero di cluster e il risultato viene presentato attraverso un diagramma ad albero detto dendrogramma che permette di identificare i successivi accorpamenti o suddivisioni del campione. Può essere suddiviso in due tipi:

- *Clustering agglomerativo*: ogni osservazione è inizialmente considerata come un cluster di un singolo elemento (foglia). Successivamente, i due cluster più simili sono combinati in un nuovo e più grande cluster (nodo). Questa procedura è iterata fino a quando tutte le osservazioni diventano elementi di un unico cluster (radice).
- *Clustering divisivo*: inverso del clustering agglomerativo, inizia con la radice, in cui tutte le osservazioni sono incluse in un unico cluster. Ad ogni passo dell'iterazione, il cluster più eterogeneo è diviso in due. Il processo è iterato fino a quando ogni osservazione individua un cluster di un singolo elemento.

1. Representative Concentration Pathways (RCPs) rappresentano percorsi alternativi (traiettorie nel tempo) dei livelli di forcing radiativo (o concentrazioni di CO-equivalente) consistenti con diverse ipotesi sulla crescita economica globale, variazione della popolazione, sfruttamento delle risorse energetiche ed altri fattori socio economici. In particolare, la IPCC ne ha selezionati quattro al fine di poter disporre di scenari standard che permettano la comparazione dei risultati tra proiezioni (van Vuuren et al. 2011). Di questi quattro, nel presente lavoro, ne sono considerati due: lo scenario RCP4.5, che rappresenta uno scenario intermedio o di stabilizzazione delle emissioni e lo scenario RCP8.5 che prevede un incremento costante delle emissioni fino al 2100; il suffisso numerico rappresenta essenzialmente l'incremento del forcing radiativo stimato al 2100 rispetto all'era preindustriale (1765) in funzione delle concentrazioni e variazioni ipotizzate; al 2111, tale valore è circa pari a 2.29 W/m^2 .
2. Le stime di emissioni così ottenute sono utilizzate come forzanti per i modelli climatici globali (GCM) che permettono la rappresentazione dei principali processi atmosferici a scala globale; numerosi studi (Breugem et al. 2007; IPCC 2013a) hanno mostrato come essi siano in grado di riprodurre il clima globale e la risposta alla variazione di GHG (Green House Gases) con affidabilità maggiori per alcune variabili (temperatura) ed inferiore per altre; d'altronde, nonostante i notevoli sviluppi degli ultimi anni, le risoluzioni orizzontali all'oggi permesse a causa degli elevati oneri computazionali si attestano nell'ordine del centinaio di chilometri; a causa di tale vincolo, tali modelli sono quindi inadeguati per stime di tendenze e impatti a scala locale/regionale per le quali assumono importanza cruciale (anche rispetto alla circolazione di larga scala) le caratteristiche geografiche del territorio (distanza dal mare, orografia).
3. Per sopperire alle limitazioni degli attuali modelli GCM, sono state quindi sviluppate tecniche di regionalizzazione differenti per approccio, oneri computazionali, prerequisiti e limiti. Fondamentalmente si riconoscono metodi di regionalizzazione statistica e dinamica. I primi adottano schemi empirici sulla base di relazioni statistiche tra predittori a larga scala e variabili climatiche locali, calibrate e validate su dati osservati e successivamente

applicate alle variabili dei GCM; richiedono limitato onere computazionale e permettono analisi anche alla scala del punto stazione ma necessitano di lunghe serie di dati osservati per la definizione delle relazioni statistiche. I secondi prevedono l'uso di modelli climatici ad area limitata e a maggiore risoluzione (Regional Climate Model RCM) innestati per l'area di interesse sul modello globale dal quale traggono le condizioni al contorno (Figura 18); le risoluzioni all'oggi correntemente adottate, nell'ordine dei 10 km, permettono da un lato, una migliore risoluzione dell'orografia e dall'altro, di risolvere una frazione cospicua dei fenomeni atmosferici locali, sebbene permangano ancora degli errori (anche in tal caso maggiore affidabilità per le temperature, minore per le precipitazioni).

Il modello COSMO CLM può essere utilizzato per effettuare simulazioni climatiche con risoluzioni spaziali comprese nell'intervallo tra 1 e 50 km, sebbene la formulazione non idrostatica delle equazioni che regolano la dinamica lo renda maggiormente adatto per effettuare simulazioni che utilizzano griglie con risoluzioni orizzontali inferiori a 20 km (Bohm et al. 2006). La formulazione matematica del modello COSMO LM è basata sulle equazioni di Navier-Stokes per un fluido compressibile. L'atmosfera è trattata come un fluido multifase (composto da aria secca, vapore acqueo e acqua allo stato liquido e solido) per cui vale l'equazione dei gas perfetti, soggetto alla forza di gravità e a quella di Coriolis. Il modello numerico, inoltre, comprende diverse parametrizzazioni, che consentono di tenere conto, almeno in maniera statistica, dell'effetto di diversi fenomeni che avvengono su scale spaziali maggiori rispetto a quelle che il modello riesce a risolvere esplicitamente, ma che hanno effetti significativi sulle scale d'interesse meteorologiche (un esempio di tali fenomeni è l'interazione dei flussi atmosferici con l'orografia). In particolare le principali parametrizzazioni incluse nel modello sono:

- Turbolenza su scala sub griglia
- Strato superficiale
- Nubi e precipitazione su scala di griglia
- Nubi su scala sub griglia

- Convezione umida
- Convezione superficiale
- Radiazione
- Modello di suolo
- Dati di terreno e superficie

Informazioni aggiuntive sulle caratteristiche del modello sono disponibili sul sito del consorzio europeo CLM Assembly che ne cura lo sviluppo.

Listato del programma

```
#pacchetti necessari
#library(Hmisc)
#library(sp)
library(RColorBrewer)#palette
library(ggplot2)#grafico mappe
library(factoextra)#cluster + slohouette e elbow indici
library(NbClust)#30 indici
library(vegan)#Calinski Harabasz indice
library(cstab)#stability based method
library(xtable)

##Input matrice dati completa
X5=read.csv(file.choose(),header=F)
colnames(X5)=c("Tmean", "R20", "FD", "SU95p", "WP", "SP", "SC", "Evap", "R95p")
##Input coordinate
X=read.csv(file.choose(),header=F)
Y=X[,2]
X=X[,1]
#
# distribuzioni spaziali: grafico anomalie di ogni singola variabile
#
#anomalia prima variabile
Fac1=data.frame(X,Y,X5[,1])
sp1=ggplot(Fac1, aes(x=X, y=Y, color=X5[,1]))+geom_point()+
xlab("")+ylab("")+ggtitle("Tmean")+labs(color="C")
sp1+scale_color_gradient2(low="blue", mid="white",high="red", space="Lab")
#sp+scale_color_gradientn(colours=Y10rRd(3))
#anomalia seconda variabile
Fac2=data.frame(X,Y,X5[,2])
sp2=ggplot(Fac2, aes(x=X, y=Y, color=X5[,2]))+geom_point()+
xlab("")+ylab("")+ggtitle("R20")+labs(color="days")
sp2+scale_color_gradient2(low="blue", mid="white",high="red", space="Lab")
```

```

#anomalie terza variabile
Fac3=data.frame(X,Y,X5[,3])
sp3=ggplot(Fac3, aes(x=X, y=Y, color=X5[,3]))+geom_point()+
xlab("")+ylab("")+ggtitle("FD")+labs(color="days")
sp3+scale_color_gradient2(low="blue", mid="white",high="red", space="Lab")
#anomalie quarta variabile
Fac4=data.frame(X,Y,X5[,4])
sp4=ggplot(Fac4, aes(x=X, y=Y, color=X5[,4]))+geom_point()+
xlab("")+ylab("")+ggtitle("SU95p")+labs(color="days")
sp4+scale_color_gradient2(low="blue", mid="white",high="red", space="Lab")
#anomalie quinta variabile
Fac5=data.frame(X,Y,X5[,5])
sp5=ggplot(Fac5, aes(x=X, y=Y, color=X5[,5]))+geom_point()+
xlab("")+ylab("")+ggtitle("WP")+labs(color="%")
sp5+scale_color_gradient2(low="blue", mid="white",high="red", space="Lab")
#anomalie sesta variabile
Fac6=data.frame(X,Y,X5[,6])
sp6=ggplot(Fac6, aes(x=X, y=Y, color=X5[,6]))+geom_point()+
xlab("")+ylab("")+ggtitle("SP")+labs(color="%")
sp6+scale_color_gradient2(low="blue", mid="white",high="red", space="Lab")
#anomalie settima variabile
Fac7=data.frame(X,Y,X5[,7])
sp7=ggplot(Fac7, aes(x=X, y=Y, color=X5[,7]))+geom_point()+
xlab("")+ylab("")+ggtitle("SC")+labs(color="days")
sp7+scale_color_gradient2(low="blue", mid="white",high="red", space="Lab")
#anomalie ottava variabile
Fac8=data.frame(X,Y,X5[,8])
sp8=ggplot(Fac8, aes(x=X, y=Y, color=X5[,8]))+geom_point()+
xlab("")+ylab("")+ggtitle("Evap")+labs(color="%")
sp8+scale_color_gradient2(low="blue", mid="white",high="red", space="Lab")
#anomalie nona variabile
Fac9=data.frame(X,Y,X5[,9])
sp9=ggplot(Fac9, aes(x=X, y=Y, color=X5[,9]))+geom_point()+
xlab("")+ylab("")+ggtitle("R95p")+labs(color="%")

```

```

sp9+scale_color_gradient2(low="blue", mid="white",high="red", space="Lab")
#
#distribuzioni frequenze
#
hist(X5[,1], main="Tmean",xlab="", ylab="Frequenza",col="blue",ylim=c(0,1600))
hist(X5[,2], main="R20",xlab="", ylab="Frequenza",col="blue",
xlim=c(-6,4),ylim=c(0,1000))
hist(X5[,3], main="FD",xlab="", ylab="Frequenza",col="blue",
xlim=c(-40,10),ylim=c(0,600))
hist(X5[,4], main="SU95p",xlab="", ylab="Frequenza",col="blue",
xlim=c(0,25),ylim=c(0,1000))
hist(X5[,5], main="WP",xlab="", ylab="Frequenza",col="blue",
xlim=c(-30,40),ylim=c(0,1200))
hist(X5[,6], main="SP",xlab="", ylab="Frequenza",col="blue",
xlim=c(-80,150),ylim=c(0,1600))
hist(X5[,7], main="SC",xlab="", ylab="Frequenza",col="blue",
xlim=c(-50,10),ylim=c(0,2000))
hist(X5[,8], main="Evap",xlab="", ylab="Frequenza",col="blue",
xlim=c(-20,20),ylim=c(0,1500))
hist(X5[,9], main="R95p",xlab="", ylab="Frequenza",col="blue",
xlim=c(-20,60),ylim=c(0,1500))
#
#matrice di correlazione
#
C=cor(X5)
#
#pre-elaborazione dei dati
#
#winsorizzazione
X5_winsor=X5
for(i in 1:9){
p_min=quantile(X5_winsor[,i],.001, #na.rm=T,
type=5)
p_max=quantile(X5_winsor[,i],.999, #na.rm=T,

```

```

type=5)
for(j in 1:4816){
  if(X5_winsor[j,i]<p_min){
    X5_winsor[j,i]=p_min
  } else if (X5_winsor[j,i]>p_max){
    X5_winsor[j,i]=p_max
  }
}
}

#standardizzazione
X5_standard=X5_winsor
for(i in 1:9){
  m=min(X5_standard[,i])
  M=max(X5_standard[,i])
  for(j in 1:4816){
    #if (is.na(X5_standard[j,i])){
    X5_standard[j,i]=(X5_standard[j,i]-m)/(M-m)
  }
}

#X5_standard=read.csv(file.choose(),header=F)
###distribuzione delle frequenze dopo la pre-elaborazione###
hist(X5_standard[,1],main="Tmean",xlab="", ylab="Frequenza",col="blue",
ylim=c(0,1500))
hist(X5_standard[,2],main="R20",xlab="", ylab="Frequenza",col="blue",
ylim=c(0,1500))
hist(X5_standard[,3],main="FD",xlab="", ylab="Frequenza",col="blue",
ylim=c(0,1000))
hist(X5_standard[,4],main="SU95p",xlab="", ylab="Frequenza",col="blue",
ylim=c(0,1200))
hist(X5_standard[,5],main="WP",xlab="", ylab="Frequenza",col="blue",
ylim=c(0,1200))
hist(X5_standard[,6],main="SP",xlab="", ylab="Frequenza",col="blue",
ylim=c(0,1500))
hist(X5_standard[,7],main="SC",xlab="", ylab="Frequenza",col="blue",

```

```

ylim=c(0,2500))
hist(X5_standard[,8],main="Evap",xlab="", ylab="Frequenza",col="blue",
ylim=c(0,2000))
hist(X5_standard[,9],main="R95p",xlab="", ylab="Frequenza",col="blue",
ylim=c(0,2000))
#
#cluster analysis
#
#
#numero ottimale di cluster
#
#metodo average silhouette###
fviz_nbclust(X5_standard,kmeans,method="silhouette")+
labs(title="", subtitle = "")+xlab("Numero di cluster")+
ylab("Indice di Silhouette")
#+ geom_vline(xintercept = 4, linetype = 2)
#Calinski_Harabasz indice#
ch=cascadeKM(X5_standard, 2, 10, iter=100, criterion="calinski")
plot(2:10, ch$results[2,1:9],"o", pch=16, col="steelblue", main="",
xlab="Numero di Cluster",
ylab="Indice di Calinski-Harabasz")
#Elbow method
fviz_nbclust(X5_standard, kmeans, method = "wss") +
# geom_vline(xintercept = 5, linetype = 2)+
labs(title="", subtitle = "")+xlab("Numero di cluster")+ylab("Indice di Elbow")
#
#hk-cluster
#
hc.ris=eclust(X5_standard,"hclust",k=5,hc_metric = "euclidean",
hc_method = "ward.D", graf=F)
grp=hc.ris$cluster
c_centri=aggregate(X5_standard, list(grp), mean)
c_centri=c_centri[,-1]
km.ris=eclust(X5_standard, "kmeans", k=c_centri, graph = F)

```

```

#
#Mappe cluster
#
#grafico i cluster insieme
Z=km.ris$cluster
Graf=data.frame(X,Y,Z)
sp=ggplot(Graf, aes(x=X, y=Y, color=Z))+geom_point()+
xlab("")+ylab("")+ggtitle("Mappa dei cluster individuati")+labs(color="")
sp+scale_color_gradientn(colours=topo.colors(5))
#sp+scale_color_gradient2(low="antiquewhite2",high="antiquewhite4", space="Lab")
#grafico ciascun cluster
#cluster 1
Z1=Z
for(i in 1:4816){
if(Z1[i]>1){
Z1[i]=0
}
}
Graf1=data.frame(X,Y,Z1)
sp=ggplot(Graf1, aes(x=X, y=Y, color=Z1))+geom_point()+
xlab("")+ylab("")+ggtitle("Cluster 1")+labs(color="")
#sp+scale_color_gradientn(colours=heat.colors(3))
sp+scale_color_gradient2(low="blue", mid="white",high="lightslateBlue", space="Lab")
#cluster 2
Z2=Z
for(i in 1:4816){
if(Z2[i]!=2)
Z2[i]=0
}
Graf2=data.frame(X,Y,Z2)
sp=ggplot(Graf2, aes(x=X, y=Y, color=Z2))+geom_point()+
xlab("")+ylab("")+ggtitle("Cluster 2")+labs(color="")
#sp+scale_color_gradientn(colours=heat.colors(3))
sp+scale_color_gradient2(low="blue", mid="white",high="blue", space="Lab")

```

```

#cluster 3
Z3=Z
for(i in 1:4816){
  if(Z3[i]!=3)
  Z3[i]=0
}
Graf3=data.frame(X,Y,Z3)
sp=ggplot(Graf3, aes(x=X, y=Y, color=Z3))+geom_point()+
  xlab("")+ylab("")+ggtitle("Cluster 3")+labs(color="")
#sp+scale_color_gradientn(colours=heat.colors(3))
sp+scale_color_gradient2(low="blue", mid="white",high="turquoise1", space="Lab")
#cluster 4
Z4=Z
for(i in 1:4816){
  if(Z4[i]!=4)
  Z4[i]=0
}
Graf4=data.frame(X,Y,Z4)
sp=ggplot(Graf4, aes(x=X, y=Y, color=Z4))+geom_point()+
  xlab("")+ylab("")+ggtitle("Cluster 4")+labs(color="")
#sp+scale_color_gradientn(colours=heat.colors(3))
sp+scale_color_gradient2(low="blue", mid="white",high="green", space="Lab")
#cluster 5
Z5=Z
for(i in 1:4816){
  if(Z5[i]!=5)
  Z5[i]=0
}
Graf5=data.frame(X,Y,Z5)
sp=ggplot(Graf, aes(x=X, y=Y, color=Z5))+geom_point()+
  xlab("")+ylab("")+ggtitle("Cluster 5")+labs(color="")
#sp+scale_color_gradientn(colours=heat.colors(3))
sp+scale_color_gradient2(low="blue", mid="white",high="yellow", space="Lab")

```

```

#
#####Distribuzioni frequenze anomalie cluster
#
XX_5=data.frame(X5,Z)
#variabile 1 cluster 1:5
Z11=matrix(0,nrow =1230, ncol=1)
Z12=matrix(0,nrow=1028,ncol=1)
Z13=matrix(0,nrow=881,ncol=1)
Z14=matrix(0,nrow=716,ncol=1)
Z15=matrix(0,nrow=961,ncol=1)
s1=0
s2=0
s3=0
s4=0
s5=0
for (i in 1:4816){
if (XX_5[i,10]==1){
s1=s1+1
Z11[s1]=XX_5[i,9]
} else if (XX_5[i,10]==2){
s2=s2+1
Z12[s2]=XX_5[i,9]
} else if (XX_5[i,10]==3){
s3=s3+1
Z13[s3]=XX_5[i,9]
} else if (XX_5[i,10]==4){
s4=s4+1
Z14[s4]=XX_5[i,9]
} else if (XX_5[i,10]==5){
s5=s5+1
Z15[s5]=XX_5[i,9]
}
}
}

```

```

hist(Z11,xlim=c(-20,60), ylim=c(0,400), breaks=15,
col=rgb(0.52,0.44,1,0.5), main="R95p", xlab="",
ylab="Frequenza")
hist(Z12, breaks=15,
col=rgb(0,0,1,0.5) , main="", xlab="",
ylab="", add=T)
hist(Z13, breaks=15,
col=rgb(0,0.96,1,0.5), main="", xlab="",
ylab="",add=T)
hist(Z14, breaks =15,
col=rgb(0,1,0,0.5), main="", xlab="",
ylab="",add=T)
hist(Z15, breaks=15,
col=rgb(1,1,0,0.5), main="", xlab="",
ylab="",add=T)
legend("topright", c("Cluster 5","Cluster 4","Cluster 3","Cluster 2","Cluster 1"),
fill=c("yellow","green","turquoise1",
"blue","lightslateblue")
,cex=0.8)

```

```

#####tabella valori medi
a=c(1.2,0,-6,12,-5,-18,-1,-3,4)
b=c(1.2,1,-9,14,8,-25,-1,-2,11)
c=c(1.3,-1,-19,9,-2,-24,-8,-3,3)
d=c(1.4,-1,-20,18,-4,-27,-12,-6,1)
e=c(1.2,-2,-20,1,-8,-15,-21,1,-1)
D=matrix(0,nrow=5,ncol=9)
D[1,]=a
D[2,]=b
D[3,]=c
D[4,]=d
D[5,]=e
f=data.frame(D)
colnames(f)=c("Tmean (C)","R20 (giorni/anno)","FD (giorni/anno)",

```

```
"SU95p (giorni/anno)", "WP (%)", "SP (%)", "SC (giorni/anno)", "Evap (%)",  
"R95p (%)")  
row.names(f)=c("Cluster 1", "Cluster 2", "Cluster 3", "Cluster 4", "Cluster 5")
```

Bibliografia

- [1] Schmidt-Thom P, Greiving S (2013) European climate vulnerabilities and adaptation: a spatial planning perspective. Wiley-Blackwell
- [2] Bucchignani E, Montesarchio M, Zollo AL, Mercogliano P (2016) High-resolution climate simulations with COSMO-CLM over Italy: performance evaluation and climate projections for the 21st century. *Int J Climatol* 36:735756. doi: 10.1002/joc.4379
- [3] Scoccimarro E, Gualdi S, Bellucci A, et al (2011) Effects of Tropical Cyclones on Ocean Heat Transport in a High-Resolution Coupled General Circulation Model. *J Clim* 24:43684384. doi: 10.1175/2011jcli4104.1
- [4] Milligan, G.W. and Cooper, M.C. (1988) A study of standardization of variables in cluster analysis. *Journal of Classification*, 5 (2), 181204.
- [5] Macqueen, J. (1967) Some methods of classification and analysis of multivariate observations, 281297. Retrieved from <http://www.citeulike.org/user/filippone/article/903715> (accessed 1 June 2011).
- [6] Hartigan, J.A. and Wong, M.A. (1979) A K-means clustering algorithm. *Applied Statistics*, 28 (1), 100108. doi: 10.1016/S0167-8655(03)00146-6.
- [7] Kaufman, L. and Rousseeuw, P.J. (1990) Finding Groups in Data. (eds L. Kaufman and P.J. Rousseeuw), John Wiley Sons, Inc., New York (p. 368). doi: 10.1002/9780470316801.

- [8] Calinski, T., and Harabasz, J. (1974) A dendrite method for cluster analysis. *Communications in Statistics Theory and Methods*, 3 (1), 127. doi: 10.1080/03610927408827101.
- [9] Ward, J.H. (1963) Hierarchical grouping to optimize an objective function. *Journal of American Statistical Association*, 58, 236-244.
- [10] Peterson, A.D., Ghosh, A.P. and Maitra, R. (2010) A systematic evaluation of different methods for initializing the K-means clustering algorithm. *Knowledge Creation Diffusion Utilization*, 111.
- [11] ESPON (2011) *ESPON Climate Change and Territorial Effects on Regions and Local Economies*. Draft Final Report.
- [12] PNACC (2017), *Analisi della condizione climatica attuale e futura*