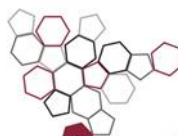




ISPRA

Istituto Superiore per la Protezione
e la Ricerca Ambientale



**Sistema Nazionale
per la Protezione
dell'Ambiente**

Variazioni della temperatura in Italia: estensione della base dati e aggiornamento della metodologia di calcolo



81/2018

STATO DELL'AMBIENTE



ISPRA

Istituto Superiore per la Protezione
e la Ricerca Ambientale



Sistema Nazionale
per la Protezione
dell'Ambiente

Variazioni della temperatura in Italia: estensione della base dati e aggiornamento della metodologia di calcolo

Informazioni legali

L'istituto Superiore per la Protezione e la Ricerca Ambientale (ISPRA), insieme alle 21 Agenzie Regionali (ARPA) e Provinciali (APPA) per la protezione dell'ambiente, a partire dal 14 gennaio 2017 fa parte del Sistema Nazionale a rete per la Protezione dell'Ambiente (SNPA), istituito con la Legge 28 giugno 2016, n.132.

Le persone che agiscono per conto dell'Istituto non sono responsabili per l'uso che può essere fatto delle informazioni contenute in questa pubblicazione.

ISPRA - Istituto Superiore per la Protezione e la Ricerca Ambientale

Via Vitaliano Brancati, 48 – 00144 Roma

www.isprambiente.gov.it

ISPRA, Stato dell'Ambiente 81/2018

ISBN 978-88-448-0917-1

Riproduzione autorizzata citando la fonte

Elaborazione grafica

ISPRA

Grafica di copertina: Franco Iozzoli

Foto di copertina: Franco Desiato

Coordinamento pubblicazione on line:

Daria Mazzella

ISPRA – Area Comunicazione

Ottobre 2018

Autori

Guido Fioravanti, Emanuela Piervitali (ISPRA)

INDICE

INTRODUZIONE	6
1. DATI	7
2. METODOLOGIA	13
2.1 Metodo di omogeneizzazione	13
2.1.1 <i>Cluster analysis</i>	13
2.1.2 <i>Omogeneizzazione delle serie di temperatura</i>	20
3. ESEMPI DI INDICATORI ED INDICI DI TEMPERATURA	26
4. CONCLUSIONI	31
BIBLIOGRAFIA	32
APPENDICE	35

INTRODUZIONE

Il Sistema nazionale per la raccolta, l'elaborazione e la diffusione di dati Climatologici di Interesse Ambientale (SCIA, www.scia.isprambiente.it) assicura da diversi anni la disponibilità di dati, statistiche e indici climatici di qualità controllata, provenienti dalle principali reti osservative nazionali e regionali, con passo temporale decadale, mensile e annuale (Desiato, 2007; Desiato et al., 2006; 2007; 2011).

Recentemente il database del sistema SCIA è stato arricchito con un dataset nazionale di serie temporali di temperatura e precipitazione con passo giornaliero, realizzato integrando serie provenienti da fonti diverse (v. Cap. 2 – Dati).

La realizzazione di un dataset nazionale, con una risoluzione spaziale e temporale più elevata rispetto ai dataset compilati per stimare le tendenze di lungo periodo su vasta scala (come quelli utilizzati nei report dell'Intergovernmental Panel on Climate Change), risponde alla crescente esigenza di studiare i cambiamenti climatici a scala regionale e locale, con particolare attenzione all'evoluzione degli estremi climatici (Thorne et al., 2011). Per quanto riguarda la risoluzione temporale, mentre molti aspetti del clima possono essere studiati sulla base dell'analisi dei dati mensili, è la disponibilità di serie giornaliere che consente l'analisi degli eventi estremi, ai quali è stata rivolta una grande attenzione nel corso degli ultimi anni per i potenziali impatti sull'ambiente e sulla salute (Heim, 2015; Thorne et al., 2011, Easterling et al., 2016). Inoltre, le serie giornaliere costituiscono un'importante risorsa per la validazione dei modelli climatologici (Alexander, 2016; Gettelman and Rood, 2016).

L'obiettivo di questo rapporto è quello di descrivere la nuova base dati, utile alla stima delle variazioni di temperatura in Italia (sia dei valori medi che degli estremi) e il processo di omogeneizzazione delle serie giornaliere di temperatura, che è stato completamente rivisto ed aggiornato rispetto a quello adottato nei precedenti lavori (Desiato et al., 2012).

Una serie omogenea è una serie la cui variabilità temporale dipende solo da fattori climatici (Peterson et al., 1998). Sebbene l'omogeneità delle serie climatiche sia un requisito fondamentale per la stima corretta delle tendenze in atto, nella realtà le serie di dati osservati di lungo periodo risultano più o meno affette da discontinuità artificiali dovute a cambiamenti delle condizioni di misura, quali lo spostamento della stazione osservativa, il cambiamento della strumentazione e nuove procedure di elaborazione dei dati (Klein Tank et al., 2009; Aguilar et al., 2003).

Il problema dell'identificazione di eventuali disomogeneità artificiali e della loro correzione è ampiamente documentato in letteratura (v., fra gli altri, Peterson et al., 1998; Aguilar et al., 2003; Reeves et al., 2007; Beaulieu et al., 2008; Venema et al.; 2012). Qui preme invece sottolineare che nel corso degli ultimi venti anni sono stati messi a disposizione diversi software per l'omogeneizzazione automatica o semi-automatica delle serie climatiche e in particolare delle serie di temperatura. Questi software semplificano notevolmente il problema dell'omogeneizzazione, consentendo all'utilizzatore di concentrarsi più sull'aspetto tecnico (scelta delle serie da omogeneizzare, controlli di qualità, valutazione della completezza e continuità delle serie, analisi dei risultati) che sull'aspetto teorico (implementazione dell'algoritmo di omogeneizzazione). Tra i software più diffusi possono essere segnalati ACMANT (Domonkos et al., 2011), Climatol (Guijarro, 2014), HOMER (Mestre et al., 2013), MASH (Szentimrey, 1999), Rhtests (Wang et al., 2010) e USHCN (Menne and Williams, 2009). Una lista esaustiva dei software disponibili è riportata alla pagina web: www.climatol.eu/tt-hom/index.html.

In questo lavoro, l'omogeneizzazione delle serie giornaliere di temperatura è stata effettuata utilizzando il software ACMANT (Adapted Caussinus-Mestre Algorithm for Networks of Temperature series), una metodologia inizialmente sviluppata nell'ambito dell'iniziativa europea COST Action ES0601 (HOME, *Advances in Homogenisation Methods of Climate Series: An Integrated Approach*) e sottoposta successivamente ad ulteriori miglioramenti (Domonkos e Coll, 2017).

Il capitolo 2 di questo rapporto descrive il nuovo database con risoluzione giornaliera del sistema SCIA e i criteri di selezione del sottoinsieme di dati utile alla stima delle variazioni di temperatura in Italia, successivamente sottoposto ad omogeneizzazione. Nel capitolo 3 viene illustrata la metodologia di elaborazione delle serie di temperatura, che include la procedura di omogeneizzazione. Nel capitolo 4 sono presentati alcuni esempi di serie di indicatori climatici aggregati a livello nazionale, calcolati sulla base del nuovo dataset, insieme al confronto con le serie nazionali riportate nei precedenti lavori (Desiato et al., 2016; 2017), basate sugli indicatori decadali, mensili e annuali.

1. DATI

Le serie giornaliere di temperatura massima e minima utilizzate per questo lavoro sono state estratte dal dataset di dati giornalieri del sistema SCIA. Tale dataset è stato realizzato tramite l'acquisizione, l'integrazione e il controllo di serie di temperatura e precipitazione provenienti da fonti diverse, principalmente: la rete sinottica di stazioni del Servizio Meteorologico dell'Aeronautica Militare e dell'ENAV e la rete dell'ex Ufficio Idrografico e Mareografico Nazionale (SIMN), poi passata in gestione e sviluppata dalle Regioni. Le serie della rete sinottica, le cui osservazioni rispondono agli standard dell'Organizzazione Meteorologica Mondiale (WMO), provengono dal database pubblico Climate Data Online (CDO) del NOAA (National Oceanic and Atmospheric Administration, www.ncdc.noaa.gov/data-access). Le serie giornaliere dell'ex SIMN, insieme a quelle provenienti da varie altre fonti (Centri Funzionali di Protezione Civile, Servizi Idrografici o Idrologici regionali) sono state invece acquisite attraverso il sistema nazionale per la condivisione dei dati idrologici HisCentral (Hydrological Information System, www.hiscentral.isprambiente.gov.it). L'archivio HisCentral copre un intervallo temporale che va approssimativamente dai primi anni del '900 fino ad oggi, ma con una forte variabilità in termini di numero, lunghezza e continuità delle serie, che variano da regione a regione. Queste serie rappresentano una parte consistente del dataset di dati giornalieri SCIA. Esse riguardano stazioni originariamente dotate di strumentazione meccanica e che a partire dagli anni '90 sono state in parte chiuse, in parte integrate o sostituite dagli strumenti automatici delle nuove reti regionali. La completa sostituzione delle stazioni meccaniche si è avuta con il trasferimento delle competenze del SIMN alle Regioni nel 2002.

Al fine di garantire una maggiore copertura temporale e spaziale, la base di dati giornalieri SCIA è stata ampliata con dati provenienti da ulteriori fonti di dati, costituite principalmente dai siti web di Centri Funzionali o Servizi Meteorologici regionali liberamente accessibili, dai quali i dati sono stati acquisiti sotto forma di tabelle HTML, file pdf o Excel (mediante tecniche di *web-scraping*).

Sebbene il database sia popolato da un numero rilevante di serie climatiche, va comunque evidenziato che solo un sottoinsieme di queste appartiene a stazioni ancora operative e che coprono un periodo sufficientemente lungo (almeno trent'anni di dati) per una stima statisticamente robusta dei trend climatici.

Una parte consistente del lavoro di realizzazione del dataset dei dati giornalieri ha riguardato: 1) l'omogeneizzazione del formato dati originario in un unico formato utile al caricamento dei dati sul database; 2) l'unificazione (*merge*) delle serie di dati che, pur derivando da fonti diverse, sono state identificate come appartenenti alla stessa stazione di misura. A questo proposito sono stati esaminati congiuntamente diversi elementi come la prossimità spaziale delle stazioni, la totale o parziale corrispondenza dei loro nomi (utilizzando algoritmi di confronto approssimativo delle stringhe come la distanza di *Levenshtein*) e soprattutto la corrispondenza dei dati negli eventuali periodi di sovrapposizione.

L'applicazione di controlli di qualità avanzati sulle serie di dati è di fondamentale importanza in quanto dati errati possono influenzare diversi aspetti dell'analisi climatologica, riducendo la correlazione tra le serie, aumentando la differenza fra le stime dei trend nelle singole serie e introducendo *bias* significativi nella stima dei trend regionali (Hunziker et al., 2018). Una volta consolidate, le serie giornaliere sono state quindi sottoposte ad una procedura automatica di controllo, che è stata sviluppata (Fioravanti et al., 2016) sulla base dei controlli di qualità applicati al dataset globale di osservazioni meteorologiche giornaliere (Global Historical Climatological Network – GHCN) del National Climatic Data Center (NCDC) della NOAA (Durre et al., 2010). La procedura comprende una sequenza di diversi test, quali i test di integrità di base, test per l'identificazione dei valori anomali, test di consistenza interna e temporale. Test di consistenza spaziale sono stati implementati per la sola temperatura in quanto per la precipitazione i risultati dei controlli spaziali richiedono comunque un'ulteriore verifica manuale da parte dell'operatore. I dati riconosciuti come errati sono esclusi dalle successive elaborazioni per il calcolo degli indicatori climatici.

Poiché l'obiettivo di questo lavoro è quello di rivedere le procedure per l'analisi statistica delle serie di anomalie nazionali di temperatura (valori medi ed estremi), sono state estratte dal database SCIA tutte le serie di temperatura massima (Tmax) e minima (Tmin) per il periodo 1961-2015, periodo in cui l'insieme delle stazioni disponibili è molto ricco e rappresentativo in termini di distribuzione spaziale dei dati.

In una prima fase sono state selezionate serie di osservazioni lunghe almeno 20 anni (con completezza del 90%), alle quali è stata applicata la procedura di omogeneizzazione descritta nel capitolo 2.1. Le serie brevi, pur non contribuendo alla stima dei trend di lungo periodo, forniscono comunque un

importante contributo in fase di omogeneizzazione delle serie più lunghe, considerando che i metodi di omogeneizzazione qui utilizzati non solo analizzano l'omogeneità di ciascuna serie su tutto il periodo 1961-2015, ma anche su sottoperiodi. Questa prima selezione di dati conta 576 serie di temperatura massima e 563 serie di temperatura minima, provenienti da stazioni con una quota che varia tra il livello del mare e i 2837 metri.

Superata la fase di omogeneizzazione, è stato selezionato un set di stazioni utili al calcolo delle anomalie medie nazionali e la stima delle tendenze. A tal fine sono state prese in esame solo le serie di temperatura massima e minima ancora operative al 2015 e che rispondono a specifici criteri di completezza e continuità, in modo da minimizzare il numero di dati mancanti e nel contempo disporre di una sufficiente densità di stazioni sul territorio italiano.

Per quanto riguarda la completezza della serie, è stata adottata la soglia di disponibilità di almeno l'80% di anni validi (45 anni) tra il 1961 e il 2015, criterio analogo a quello adottato in altri studi sulle serie climatiche (Alexander et al., 2006; Hunziker et al., 2018; Klein Tank and Können, 2003; Toreti and Desiato, 2008). Un anno è stato considerato valido se contiene almeno 9 mesi di dati e nessuna stagione totalmente priva di mesi validi. Un mese, invece, è stato considerato valido se contiene al massimo 5 giorni mancanti e non più di 3 giorni consecutivi mancanti. Relativamente alla continuità, si è deciso di selezionare le serie che presentano al massimo quattro anni consecutivi mancanti. L'adozione di tali criteri ha permesso di individuare nel set di stazioni precedentemente identificato 144 serie di temperatura massima e 139 serie di temperatura minima, la cui anagrafica è riportata in Tabella 1.1 e la cui distribuzione sul territorio nazionale è illustrata in Figura 1.1. Le serie con almeno 50 anni di dati sono 127 per Tmax e 123 per Tmin, il 30% circa delle quali proviene dalla rete sinottica, le cui stazioni sono distribuite in modo piuttosto uniforme su tutto il territorio nazionale.

Per il calcolo degli indici estremi di temperatura si è proceduto a un'ulteriore selezione delle serie della Tabella 1.1, adottando requisiti di completezza a livello mensile e annuale ancora più stringenti, poiché gli estremi sono molto sensibili ai dati mancanti. Sono stati applicati criteri adottati a livello internazionale (Alexander et al., 2006; Donat et al., 2013) che definiscono un mese valido se contiene al massimo 3 giorni mancanti e un anno valido se ha tutti i mesi validi e contiene al massimo 15 giorni mancanti.

Una schematizzazione del processo di selezione delle serie è riportato in Figura 1.2.

Tabella 1.1 - Stazioni di temperatura massima e minima selezionate per il calcolo delle serie di anomalia media nazionale

Nome stazione	lon	lat	Quota (m)	Tmax	Tmin
RIDANNA	11.31	46.91	1337	x	x
VIPIGENO	11.43	46.90	946	x	x
ANTERSELVA DI MEZZO OSS.	12.10	46.86	1274	x	x
S.MADDALENA IN CASIES	12.24	46.84	1456	x	x
PLATA	11.18	46.82	1118	x	x
S. VALENTINO ALLA MUTA	10.53	46.75	1461	x	x
VERNAGO	10.85	46.74	1823	x	x
DOBBIACO	12.22	46.73	1226	x	x
MONTE MARIA	10.52	46.71	1256	x	x
SESTO	12.35	46.70	1490	x	x
FIÈ ALLO SCILIAR	11.51	46.51	804	x	x
TARVISIO	13.58	46.50	778	x	x
BOLZANO	11.33	46.47	241	x	x
CARESER (DIGA)	10.70	46.42	2600	-	x
PASSO ROLLE	11.78	46.30	2006	x	x
PAGANELLA	11.03	46.15	2129	x	x
AVIANO	12.60	46.03	128	x	x
LAVARONE	11.25	45.94	1155	x	x
GRESSONEY-L.T. DEJOLA	7.81	45.86	1823	x	x
TREVISO/ISTRANA	12.10	45.68	46	x	x
TREVISO/S. ANGELO	12.18	45.65	23	x	x
TRIESTE	13.75	45.65	20	x	x
VALGRISENCHE DIGA	7.61	45.62	1832	x	x
MILANO/MALPENSA	8.73	45.62	211	x	x
VENEZIA/TESSERA	12.33	45.50	6	x	x
MILANO/LINATE	9.28	45.43	103	x	x
BRESCIA/GHEDI	10.28	45.42	97	x	x
VERONA/VILLAFRANCA	10.87	45.38	68	x	x
TORINO/CASELLE	7.65	45.22	287	x	x
TORINO/BRIC DELLA CROCE	7.73	45.03	710	x	x
PIACENZA	9.73	44.92	138	x	x
BOBBIO	9.38	44.76	364	x	x
ANZOLA DELL'EMILIA	11.18	44.58	34	x	x
BEDONIA	9.63	44.51	527	x	x
ALFONSINE	12.04	44.49	7	x	x
MARINA DI RAVENNA	12.28	44.49	1	x	x
PAVULLO	10.83	44.32	687	x	x
LIGONCHIO	10.34	44.32	803	x	x
FAENZA	11.89	44.29	34	x	x
PIANDELGOTTI	10.52	44.24	1228	x	x
SESTOLA	10.77	44.23	1005	x	x
CERVIA	12.30	44.22	10	x	-
MONTE CIMONE	10.70	44.20	2173	x	x
MODIGLIANA	11.80	44.16	240	x	x
PORRETTA TERME	10.98	44.15	395	x	x
SAN CASSIANO LAMONE	11.69	44.15	437	x	x
SUVIANA	11.04	44.13	479	x	x
COTTEDEMEC	11.17	44.11	838	x	x
RIMINI	12.62	44.03	13	x	x
CAPO MELE	8.17	43.95	221	x	x
NOVAFELTRIA MARECCHIA	12.29	43.89	263	x	-
CAMPIGNA	11.75	43.87	1034	-	x
LUCCA (ORTO BOTANICO)	10.51	43.84	17	x	x
FIRENZE/PERETOLA	11.20	43.80	38	x	x
CARPEGNA	12.34	43.78	715	x	x
URBINO	12.64	43.72	471	x	x
FOSSOMBRONE	12.80	43.69	96	x	x
PISA/S. GIUSTO	10.38	43.68	6	x	x
SANT'ANGELO IN VADO	12.41	43.67	352	x	x
ANCONA TORRETTE	13.45	43.61	5	x	x
PERGOLA	12.84	43.57	242	x	x
JESI	13.23	43.53	100	x	x
FRONTONE	12.73	43.52	574	x	x
ARCEVIA	12.94	43.50	535	x	x
AREZZO	11.89	43.47	291	x	x
FONTE AVELLANA	12.73	43.47	690	x	x
AREZZO	11.85	43.47	249	x	x
FABRIANO	12.90	43.33	354	x	x
LORNANO	13.42	43.29	294	x	x

PERUGIA	12.50	43.08	205	x	x
MONTEMONACO	13.33	42.90	995	x	x
ASCOLI PICENO	13.65	42.85	78	x	x
GROSSETO	11.07	42.75	7	x	x
AMATRICE	13.30	42.63	954	x	x
LEONESSA	12.95	42.58	929	x	x
MONTE TERMINILLO	12.98	42.47	1875	x	x
PESCARA	14.20	42.43	11	x	x
MONTE ARGENTARIO	11.17	42.38	631	x	x
POSTICCIOLA	12.94	42.23	626	x	x
ALLUMIERE	11.91	42.16	568	x	-
BRACCIANO	12.18	42.10	214	x	-
CIVITAVECCHIA	11.79	42.10	7	x	x
VIGNA DI VALLE	12.22	42.08	266	x	x
TERMOLI	15.00	42.00	44	x	x
ROMA MACAO	12.51	41.90	56	x	x
COLLEGIO ROMANO	12.48	41.90	19	x	x
ROMA FIUMICINO	12.23	41.80	3	x	x
ROMA/CIAMPINO	12.58	41.78	105	x	x
MONTE S. ANGELO	15.95	41.70	844	x	x
PRATICA DI MARE	12.45	41.65	21	x	x
FROSINONE	13.30	41.63	185	x	x
CAMPOBASSO	14.65	41.57	807	x	x
LATINA	12.90	41.55	26	x	x
LATINA	12.91	41.47	17	x	x
BARI/PALESE MACCHIE	16.78	41.13	49	x	x
GRAZZANISE	14.07	41.05	10	x	x
TREVICO	15.23	41.05	1093	x	x
PONZA	12.95	40.92	185	x	x
GIOIA DEL COLLE	16.93	40.77	350	x	x
BRINDISI	17.95	40.65	10	x	x
ALGHERO	8.28	40.63	40	x	x
POTENZA	15.81	40.63	669	x	x
MARINA DI GINOSA	16.88	40.43	12	x	x
LECCE	18.15	40.23	61	x	x
LATRONICO	16.02	40.08	896	x	x
CAPO PALINURO	15.28	40.02	185	x	x
CAPO BELLAVISTA	9.72	39.93	150	x	x
S. MARIA DI LEUCA	18.35	39.82	112	x	x
VILLAPIANA SCALO	16.50	39.80	206	x	x
CASTROVILLARI	16.25	39.77	199	x	x
TORANO SCALO	16.21	39.49	93	x	x
DECIMOMANNU	8.97	39.35	28	x	x
MONTE SCURO	16.40	39.33	1720	x	x
COSENZA	16.27	39.29	269	x	x
CAGLIARI/ELMAS	9.07	39.25	5	x	x
CROTONE	17.13	39.09	1	x	x
CROTONE	17.07	39.00	161	x	-
CATANZARO	16.59	38.91	315	x	x
USTICA	13.18	38.70	251	x	x
CITTANOVA	16.08	38.35	407	x	x
MESSINA	15.55	38.20	51	x	x
PALERMO/PUNTA RAISI	13.10	38.18	21	x	x
PALERMO OSS. ASTRONOMIC	13.35	38.11	27	x	x
SAN FRATELLO	14.60	38.02	654	x	x
SAN GIUSEPPE JATO	13.19	37.97	583	-	x
TRAPANI/BIRGI	12.50	37.92	14	x	x
CIMINNA	13.56	37.89	467	x	x
LINGUAGLOSSA	15.14	37.84	512	x	x
MARSALA	12.46	37.81	5	x	x
PRIZZI	13.43	37.72	1035	x	x
PIANO DEL LEONE	13.47	37.67	876	x	-
BIVONA	13.44	37.62	554	x	x
ENNA	14.27	37.57	804	x	x
SCIACCA	13.13	37.51	110	x	x
CATANIA/FONTANAROSSA	15.05	37.47	17	x	x
CATANIA/SIGONELLA	14.92	37.40	22	x	x
PIAZZA ARMERINA	14.37	37.39	708	x	x
AGRIGENTO	13.59	37.30	212	x	-
LENTINI CITTA	15.00	37.29	40	x	-
MINEO	14.69	37.26	459	x	x
CALTAGIRONE	14.52	37.24	428	x	x
LICATA	13.93	37.10	48	x	x
GELA	14.22	37.08	65	x	x

RAGUSA	14.72	36.92	532	x	x
PANTELLERIA	11.97	36.82	170	x	x
COZZO SPADARO	15.13	36.68	51	x	x
LAMPEDUSA	12.60	35.50	20	x	x

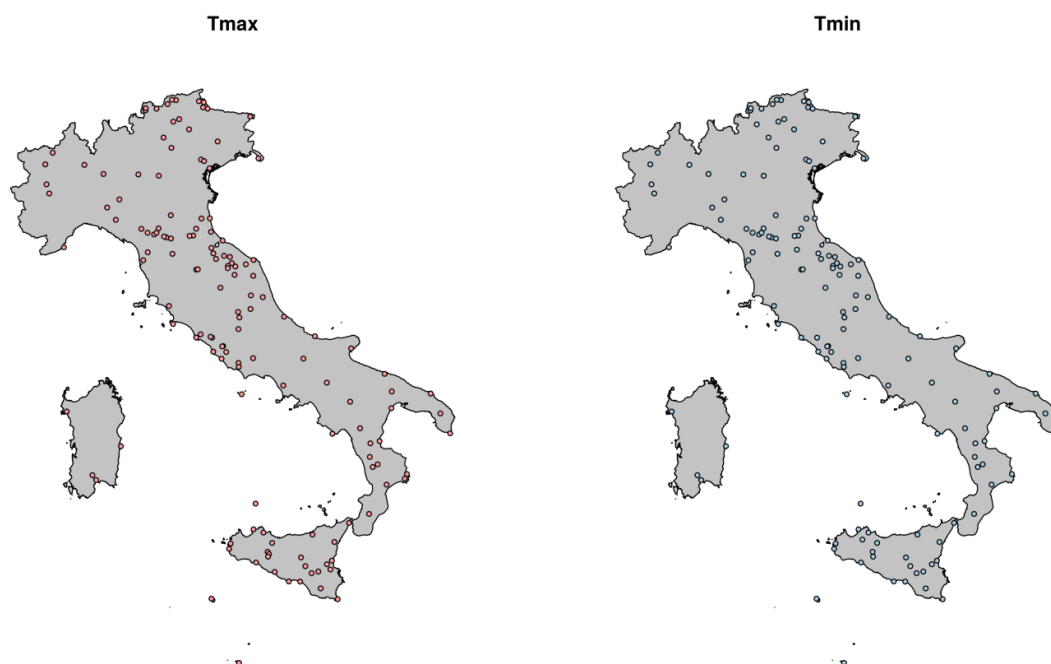


Figura 1.1 - Distribuzione spaziale delle stazioni utilizzate per il calcolo dell'anomalia media nazionale di temperatura massima (a sinistra) e minima (a destra).

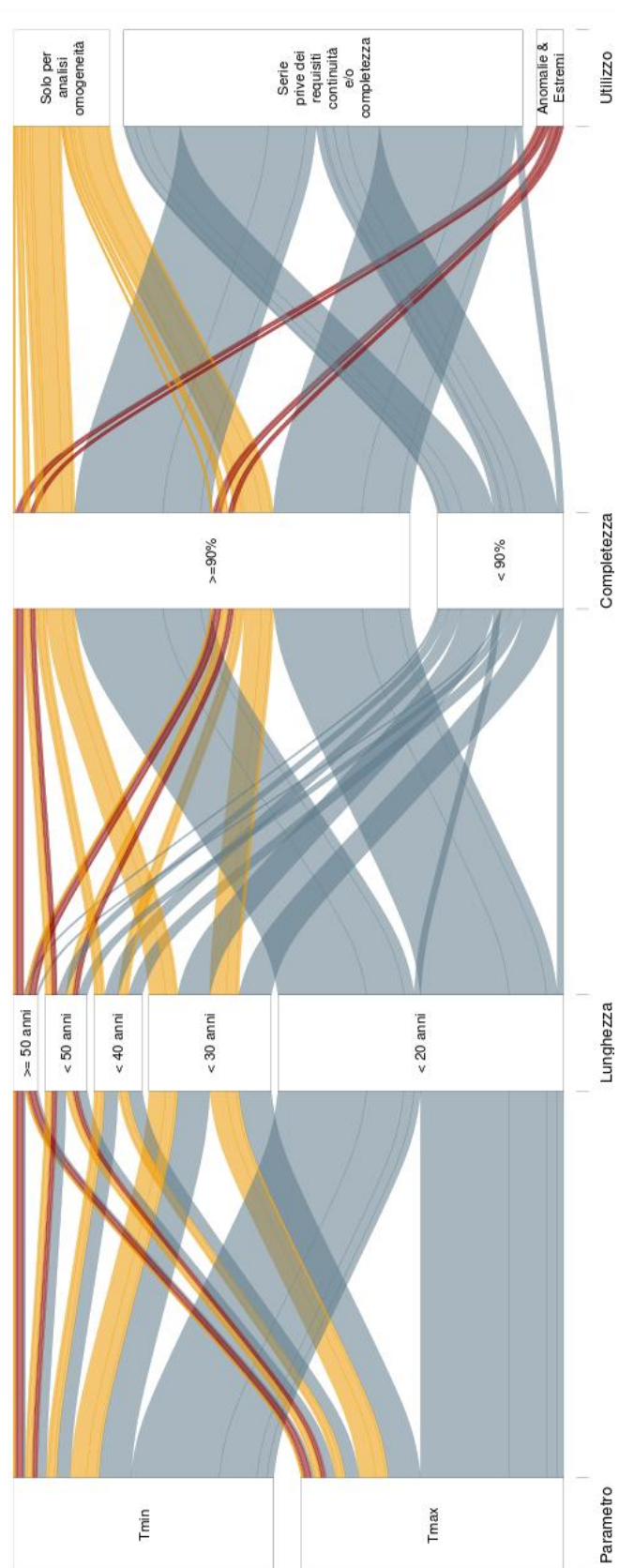


Figura 1.2 - Lo schema (alluvial diagram) illustra il processo di selezione delle serie. Le serie lunghe almeno 20 anni (e complete al 90% a livello annuale) sono state omogeneizzate. Successivamente le serie più lunghe e complete, ancora in funzione nel 2015, sono state utilizzate per il calcolo delle anomalie e degli estremi climatici.

2. METODOLOGIA

Il precedente set di stazioni (serie versione 1) utilizzato per la stima delle anomalie medie nazionali e dei relativi trend (Desiato et al., 2012; Fioravanti et al., 2015) era costituito da 68 serie di temperatura massima e 61 serie di temperatura minima per lo più appartenenti alla rete sinottica nazionale. Utilizzando il Penalized Maximum t test implementato nel software RHTest (Wang et al., 2010), le serie di questo insieme erano state riconosciute come “*potenzialmente omogenee*” ovvero prive di evidenti discontinuità artificiali (*breakpoints*) o con discontinuità *statisticamente* dubbie (che richiederebbero la validazione mediante metadati) a livello annuale, e quindi utili per la stima dei trend delle anomalie medie nazionali. Per il calcolo degli indici estremi venivano utilizzate le corrispondenti serie giornaliere (se disponibili), assumendo che l’assenza di *breakpoints* a livello annuale potesse rassicurare sull’assenza di importanti disomogeneità a livello giornaliero (Desiato et al., 2012; 2013). Il nuovo set di dati di temperatura minima e massima giornaliera (capitolo 1) migliora nettamente la situazione precedente, sia in termini di numerosità di stazioni (più del doppio), sia per la possibilità di analizzare l’omogeneità delle serie e correggerle direttamente a livello giornaliero.

2.1 Metodo di omogeneizzazione

2.1.1 Cluster analysis

L’Italia è caratterizzata da una notevole estensione in latitudine e da una complessa orografia che ne rendono molto varie le caratteristiche climatiche. L’elevata variabilità spaziale richiede che il processo di omogeneizzazione delle serie sia condotto all’interno di aree climaticamente omogenee, in modo che per ciascuna serie da omogeneizzare (serie candidata) si disponga di più serie di riferimento il cui segnale climatico sia rappresentativo dell’area di appartenenza.

La “regionalizzazione climatica” del dataset di serie di temperatura può essere condotta in diversi modi, ad esempio ricorrendo a sistemi classici di classificazione del clima (come la Köppen Climate Classification, Köppen, 1936), ampiamente ed efficacemente utilizzati in diversi campi di ricerca (Rubel et al., 2016). Il limite di questi sistemi è quello di essere rappresentativi delle condizioni climatiche medie di una regione, piuttosto che della variabilità temporale del segnale climatico, che è invece l’obiettivo principale di questo lavoro. Inoltre, in questi sistemi di classificazione è presente un certo grado di soggettività nella definizione delle classi climatiche (Geng et al., 2014).

Un approccio alternativo è l’uso di metodi statistici multivariati quali la *cluster analysis*, ovvero metodi di classificazione oggettivi (*data driven*), in cui le unità statistiche tendono a raggrupparsi in classi (*cluster*) non definite a priori, bensì ricavate dall’algoritmo di classificazione sulla base di caratteristiche dei dati stessi. L’obiettivo della *cluster analysis* è quindi il partizionamento delle unità statistiche (in questo caso le serie climatiche) in gruppi tali che unità appartenenti allo stesso gruppo siano fra loro molto simili rispetto a determinate caratteristiche dei dati, mentre unità appartenenti a gruppi distinti risultino fra loro molto dissimili. Sistemi di classificazione dei dati di questo tipo, che non usano alcuno schema predefinito, sono generalmente noti come metodi di apprendimento non supervisionato (*unsupervised learning*).

In questo lavoro la *cluster analysis* è stata effettuata ricorrendo al metodo di Ward (1963), frequentemente utilizzato in ambito climatico (Pérez-Zanón et al., 2015; Calmanti et al., 2015; Muñoz-Díaz and Rodrigo, 2004; Unal et al. 2003). Il metodo di Ward viene definito come metodo gerarchico agglomerativo (*agglomerative hierarchical clustering*) in quanto tende a raggruppare le unità statistiche in *cluster* via via sempre più ampi.

Il metodo di Ward si può così schematizzare:

- inizialmente vengono definiti tanti *cluster* quante sono le singole unità statistiche (le serie di temperatura massima o minima, nel nostro caso);
- viene calcolata la distanza tra ciascuna coppia di *cluster* utilizzando una metrica opportuna (ad esempio, la metrica euclidea o la correlazione);
- le coppie di *cluster* con distanza minima l’una dall’altra vengono fuse in un nuovo *cluster*;
- il costo associato alla creazione di questo nuovo *cluster* viene calcolato in termini di *Within cluster Sum of Squares* (WSS), così definito:

$$\begin{aligned}
WSS(A, B) &= \sum_{x \in A \cup B} \|x_i - m_{A \cup B}\|^2 - \sum_{x \in A \cup B} \|x_i - m_A\|^2 - \sum_{x \in B \cup B} \|x_i - m_B\|^2 \\
&= \frac{n_A \cdot n_B}{n_A + n_B} \|m_A - m_B\|^2
\end{aligned}$$

dove m_A (m_B) rappresenta il centroide del *cluster* A (B) e n_A (n_B) il numero di unità che ad esso appartengono.

In termini molto generici, la quantità WSS ci dice di quanto aumenta la devianza interna (*within*) quando il *cluster* A viene fuso con il *cluster* B. Ad ogni iterazione l'algoritmo di classificazione tende ad aggregare le coppie di *cluster* A e B che minimizzano la dissimilarità (devianza) interna del nuovo *cluster*. L'ultimo termine a destra della formula ci dice che, a parità di distanza tra coppie di *cluster*, il metodo di Ward tende a privilegiare (unire) i *cluster* con il minor numero di unità.

- il processo di fusione termina quando tutte le unità si ritrovano in un unico grande *cluster* (il dataset nel suo insieme).

Ovviamente l'obiettivo di un metodo gerarchico agglomerativo non è quello di arrivare ad un unico grande *cluster*, bensì quello di individuare un certo numero di gruppi tale da garantire un soddisfacente livello di separazione tra le unità appartenenti a *cluster* distinti (*devianza between*) e un elevato livello di coesione interna ai vari gruppi (*devianza within*).

La successione delle aggregazioni innestate prodotte dal metodo di Ward viene efficacemente rappresentata mediante un "dendrogramma" ovvero un diagramma ad albero in cui sull'asse delle ascisse compaiono le varie unità statistiche oggetto del processo di classificazione e sull'asse delle ordinate la distanza che separa i vari *cluster*. L'analisi del dendrogramma può aiutare il ricercatore nella scelta del numero di *cluster* in cui suddividere il set di dati di partenza. L'idea è quella di scegliere un numero di *cluster* k per cui il passaggio ad un numero di $k-1$ *cluster* corrisponda nel dendrogramma ad un salto rilevante (distanza sull'asse y) nella distanza tra i gruppi (Gareth et al., 2013).

In questo lavoro la *cluster analysis* è stata applicata alle serie dei valori medi mensili di temperatura (massima e minima separatamente), utilizzando il metodo agglomerativo di Ward e la metrica euclidea.

Le serie medie mensili sono state costruite a partire dal dataset delle serie giornaliere descritto nel capitolo 1. Poiché la *cluster analysis* non ammette la presenza di dati mancanti, per ogni singola stazione i mesi non sufficientemente popolati a livello giornaliero sono stati sostituiti con il valore medio mensile calcolato su tutti gli anni disponibili. Poiché la *cluster analysis* non è l'obiettivo finale di questo lavoro, metodi di imputazione dei dati mancanti alternativi (come la ricostruzione dei dati mediante l'uso di serie limitrofe) non sono stati presi in considerazione in quanto più onerosi in termini computazionali e di programmazione.

L'applicazione della *cluster analysis* al set di dati in esame ha permesso di ottenere, sia per la temperatura massima che per la temperatura minima, una prima suddivisione del dataset in 6 gruppi di stazioni (Figura 2.1 per Tmax e Figura 2.2 per Tmin). La scelta del numero di *cluster* si è basata sia sull'analisi visiva del dendrogramma (a titolo di esempio in Figura A.1 è riportato il dendrogramma per le serie di Tmax) sia tenendo conto del numero di stazioni presenti in ogni gruppo. Gruppi scarsamente popolati potrebbero invalidare il processo di omogeneizzazione; d'altra parte gruppi molto numerosi possono essere ulteriormente suddivisi per ottenere sottogruppi di stazioni con caratteristiche climatiche più omogenee.

In Figura 2.3 viene riportata la distribuzione della quota delle stazioni in ogni *cluster*. Da queste figure è chiaramente riconoscibile un *cluster* costituito unicamente da stazioni di alta quota (*cluster* 5), con un'altitudine che va da 1900 a oltre 2700 m (quota media di circa 2200 m). Questo *cluster* conta appena 20 stazioni che appartengono quasi tutte al nord Italia, ad eccezione della stazione di Monte Terminillo nel Lazio (quota 1875). Sebbene questa stazione sia spazialmente distante dalle restanti 19 stazioni non è stata "riassegnata" agli altri *cluster*. Questo criterio di "non riassegnazione" è stato generalmente adottato per tutti i *cluster*, sulla base del fatto che, in fase di omogeneizzazione, il contributo di ciascuna serie dipende dal livello effettivo di correlazione con le restanti serie.

Il *cluster* 4 è anch'esso un *cluster* di stazioni in quota, con altitudine che varia da circa 1000 a oltre 1800 m (quota media di 1438 metri) localizzate principalmente nel nord Italia.

I *cluster* 1 e 6 sono costituiti da stazioni del centro e sud Italia. Il *cluster* 1 contiene stazioni con una distanza media dalla linea di costa di 18 km e una quota media di circa 200 m (intervallo da 0 a 800 m), mentre il *cluster* 6 contiene stazioni con una distanza media dalla linea di costa di 10 km e una quota media di 100 m (intervallo da 0 a 500 m).

I *cluster* 2 e 3 sono costituiti principalmente da stazioni del centro e nord Italia, generalmente localizzate in zone più interne e caratterizzate da una quota media rispettivamente di circa 950 m e di 350 m. La distribuzione delle stazioni all'interno del *cluster* 2 fa supporre la presenza di almeno tre sottocluster al suo interno ma si è preferito non procedere ad una sua ulteriore disaggregazione per non incorrere nel rischio di non avere sufficienti serie di riferimento in fase di omogeneizzazione.

La distribuzione spaziale delle stazioni di temperatura minima è riportata in Figura 2.2. Il box-plot in basso della Figura 2.3 mostra la distribuzione della quota di queste stazioni all'interno di ciascun *cluster*.

Per la temperatura minima il *cluster* 5 contiene un set di stazioni di montagna del nord Italia, con una quota che varia da 1000 metri circa a 2800 m e una media di 1700 metri. Questo *cluster* sembra corrispondere ai *cluster* 4 e 5 della temperatura massima, le cui stazioni hanno complessivamente una quota compresa tra circa 1000 e 2800 m.

I *cluster* 2 e 6 sono costituiti principalmente da stazioni costiere a bassa quota del centro e sud Italia (ad eccezione di una stazione in Liguria nel *cluster* 6), con una distanza media dalla linea di costa rispettivamente di 1 e 10 km. Questi due *cluster* sembrano avere le caratteristiche del *cluster* 6 della temperatura massima.

Il *cluster* 3 della temperatura minima, al quale appartengono stazioni del centro e nord Italia, sembra corrispondere al *cluster* 2 della temperatura massima, sia rispetto alla distribuzione delle stazioni al suo interno che alla quota, che varia da 350 a 1800 m (media di 1000 m). I *cluster* 1 e 4 contengono principalmente stazioni del centro e nord Italia in pianura o collina.

Il ciclo stagionale delle serie di temperatura in ogni *cluster*, riportato nella Figura 2.4 (Tmax in alto e Tmin in basso) mostra chiaramente la differenza tra i gruppi di stazioni. I *cluster* delle stazioni di montagna, ai quali si riferiscono le temperature più basse, mostrano valori medi di Tmax che vanno da circa 0°C nei mesi invernali a 15°C in quelli estivi (*cluster* 5) e valori di Tmin che variano da temperature inferiori a -5°C a circa 8°C per il *cluster* 5 e da 0 a 12°C per il *cluster* 3. Il ciclo stagionale con valori medi più elevati, che si riferisce alle stazioni costiere, indica per la Tmax (*cluster* 6) valori da 15 a 30°C e per la Tmin da 5 a 20°C e da 10 a 23°C, rispettivamente per i *cluster* 2 e 6.

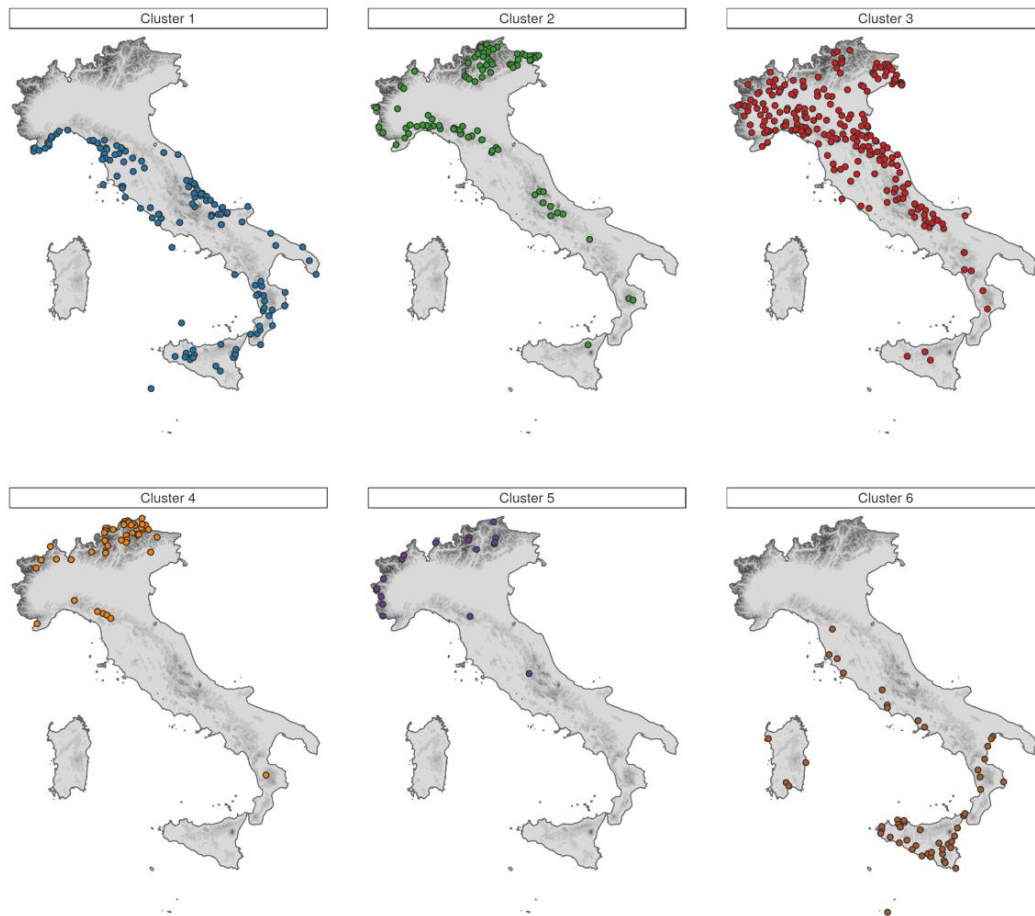


Figura 2.1 - Distribuzione spaziale delle stazioni di temperatura massima in ogni cluster. Le mappe rappresentano una prima classificazione delle stazioni ottenuta a partire dalle serie medie mensili. I cluster troppo ampi sono stati ulteriormente sottoposti a cluster analysis (i risultati non vengono qui illustrati).

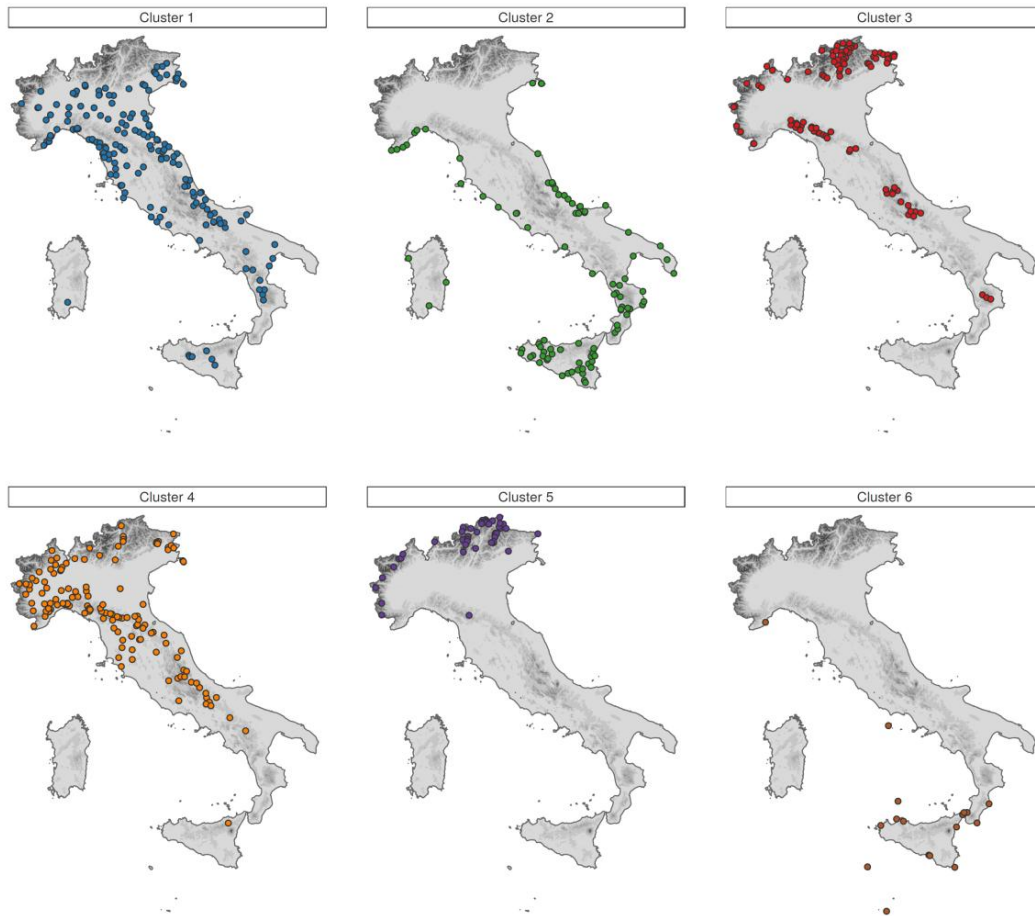


Figura 2.2 - *Distribuzione spaziale delle stazioni di temperatura minima in ogni cluster.*

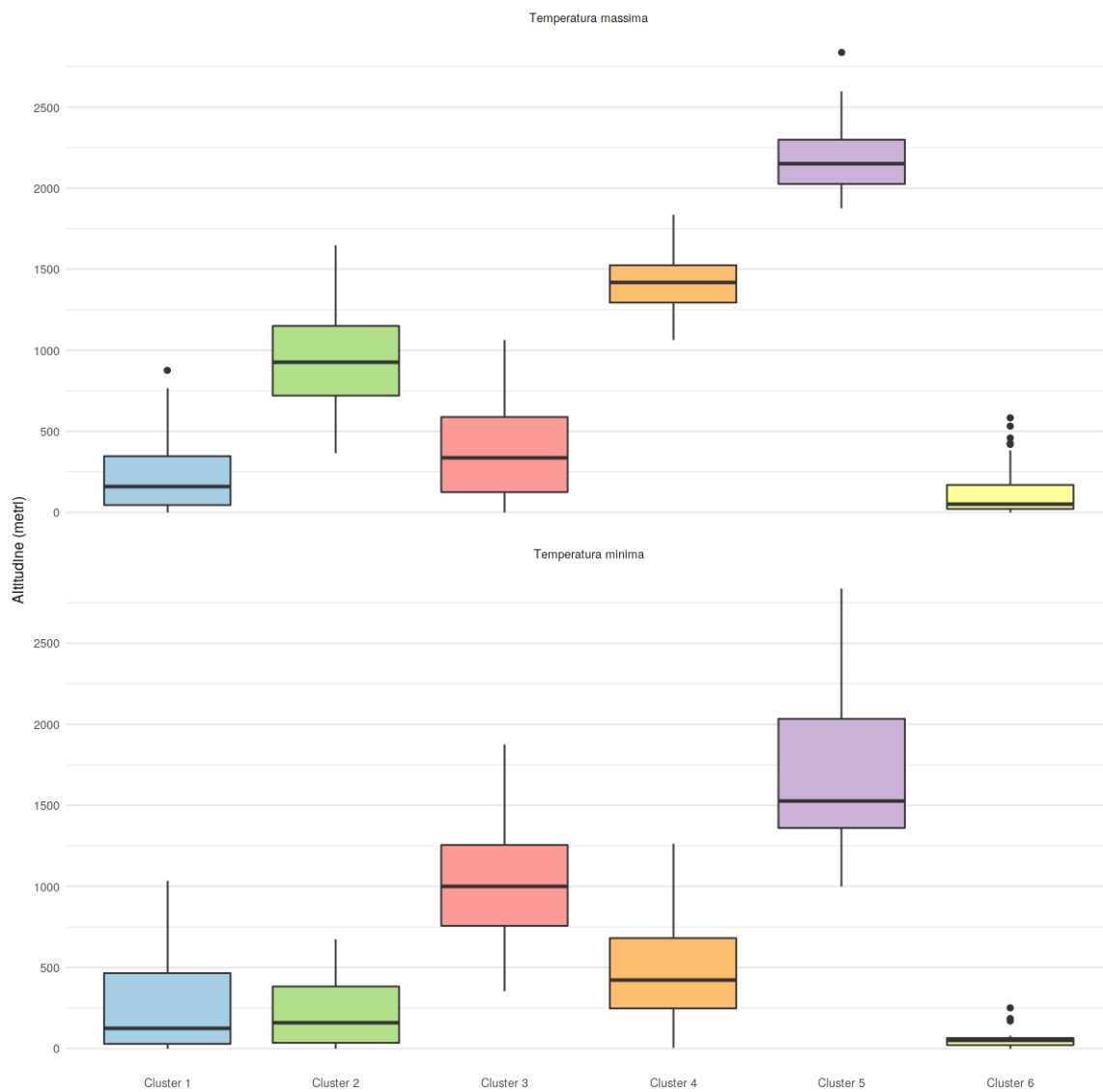


Figura 2.3 - Box-plot della quota delle stazioni nei sei cluster di Tmax (in alto) e Tmin (in basso).

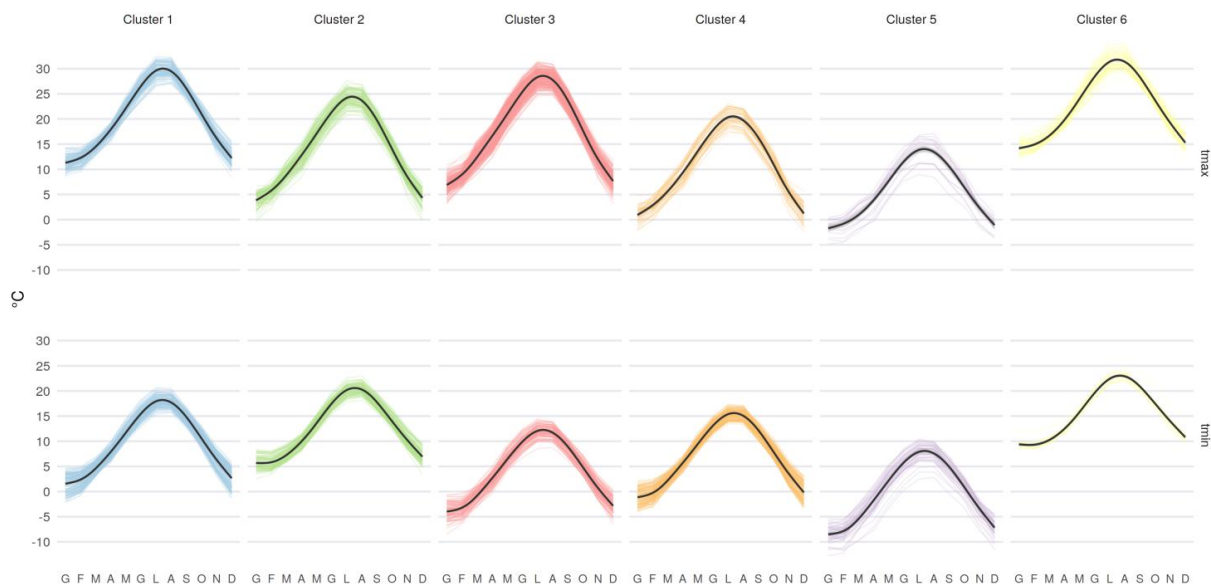


Figura 2.4 - Ciclo stagionale della temperatura massima (in alto) e minima (in basso) per ogni cluster: la linea nera indica il valore medio del cluster, mentre le linee colorate indicano il ciclo stagionale medio di ogni stazione.

2.1.2 Omogeneizzazione delle serie di temperatura

Il primo passo nel processo di omogeneizzazione consiste nell'identificare le eventuali disomogeneità (*breakpoints*) e la loro significatività statistica. I metodi comunemente utilizzati a tal proposito ricadono sotto la definizione di “metodi relativi” in quanto non vengono applicati direttamente alla serie oggetto di analisi (serie candidata) ma alla serie costruita come differenze (ad esempio per la temperatura) o rapporti (nel caso della precipitazione) fra la serie candidata e la corrispondente serie di riferimento. La serie di riferimento, quando opportunamente costruita, è una serie priva di disomogeneità e rappresentativa del segnale climatico dell'area di appartenenza. Metodi di omogeneizzazione assoluti, ovvero test statistici applicati direttamente alla serie candidata, sono generalmente sconsigliati in quanto si basano sul presupposto (non plausibile) di stazionarietà del clima (Mamara et al., 2013).

L'analisi e il confronto dei metodi di omogeneizzazione più diffusi è stato oggetto dell'iniziativa europea COST Action ES0601 (HOME, *Advances in Homogenisation Methods of Climate Series: An Integrated Approach*; www.homogenisation.org) e del più recente progetto MULTITEST (www.climatol.eu/MULTITEST). Al fine di condividere aggiornamenti, metodologie e software necessari per il controllo di qualità e l'omogeneizzazione delle serie temporali, vengono inoltre organizzati periodicamente seminari, sotto l'egida della WMO e con il supporto del Servizio Meteorologico Ungherese (WMO, 2014; 2011).

Fra le metodologie sviluppate in ambito HOME, per questo lavoro sono state individuate e utilizzate in una prima fase due diversi metodi di omogeneizzazione: Climatol (Guijarro et al., 2016) e ACMANT (Adapted Caussinus-Mestre Algorithm for Networks of Temperature series, Domonkos e Coll, 2017). Entrambi sono metodi relativi e utilizzano un approccio completamente automatico per identificare e correggere le disomogeneità multiple nelle serie, facilitando il lavoro in presenza di una grande mole di dati. Questi metodi hanno il vantaggio di non richiedere valutazioni soggettive da parte dell'operatore, valutazioni che potrebbero richiedere l'uso di metadati che spesso non sono disponibili, come nel nostro caso.

Un approccio completamente automatico è ad esempio adottato dalla NOAA, che utilizza il Pairwise Homogeneity Algorithm (Menne and Williams, 2009) per l'omogeneizzazione del dataset di temperatura media GHCN-M (Global Historical Climatology Network-Monthly; Lawrimore et al., 2011).

Sia Climatol che ACMANT provvedono alla costruzione della serie di riferimento per ciascuna serie candidata. Climatol, basato sull'implementazione del test di Alexandersson (1986), è stato efficacemente utilizzato per omogeneizzare le serie di temperatura delle stazioni del Servizio Meteorologico Nazionale Greco (Mamara et al., 2013) e le serie di precipitazione del dataset dell'Agenzia Meteorologica Spagnola (Luna et al., 2012); ACMANT, sviluppato sulla base dell'algoritmo di Cassinus-Mestre (Cassinus and Mestre, 2004), è stato applicato con risultati soddisfacenti per l'omogeneizzazione delle serie di temperatura nella regione montuosa dei Pirenei (Pérez-Zanón et al., 2015).

Di seguito viene riportata una breve descrizione dei due metodi.

Climatol è un software implementato in R (R Core Team, 2015) la cui procedura si può riassumere nei seguenti passi:

1. aggregazione delle serie giornaliere a livello mensile;
2. identificazione delle serie limitrofe per ciascuna serie candidata: le serie limitrofe vengono identificate sulla base della distanza dalla serie candidata, all'interno della regione climatica di appartenenza (definita dalla *cluster analysis*), includendo le serie che si sovrappongono anche solo parzialmente (serie brevi) alla serie candidata, al fine di sfruttare al meglio il set di dati disponibili;
3. stima della serie di riferimento: le serie di riferimento vengono costruite a partire da una media ponderata delle stazioni limitrofe tenendo conto della distanza di separazione dalla serie candidata;
4. l'analisi di omogeneità viene condotta sulla serie ottenuta sottraendo la serie di riferimento alla serie dei dati originali normalizzati, utilizzando il test di Alexandersson (1986). L'analisi si svolge in due fasi: inizialmente il test viene condotto su sotto-periodi di 20 anni per individuare i *breakpoint* che sfuggirebbero applicando il test alla serie intera; nella seconda fase il test viene applicato a tutta la serie per identificare eventuali *breakpoint* sfuggiti al controllo precedente. Per ogni disomogeneità individuata, l'algoritmo procede a suddividere la

serie in due serie indipendenti, una precedente e l'altra successiva al *breakpoint*, e a calcolare i fattori di correzione;

5. correzione delle serie giornaliere, applicando i fattori di correzione individuati a livello mensile.

Per quanto riguarda ACMANT il processo di omogeneizzazione può essere così sintetizzato (WMO, 2011; Domonkos et al., 2011):

1. creazione delle serie di riferimento: le serie di riferimento vengono costruite secondo il metodo di Peterson and Easterling (1994) che tiene conto della correlazione tra serie candidata e serie limitrofe. Le serie di riferimento sono sottoposte a un processo di pre-omogeneizzazione in modo di eliminare gli errori più consistenti;
2. identificazione delle disomogeneità: ACMANT utilizza una ricerca dei *breakpoint* in due fasi (*bivariate detection*), entrambe basate sull'algoritmo di Caussinus-Mestre. La prima fase (*main detection*) avviene a livello annuale, utilizzando segmenti di serie lunghi almeno 3 anni; la collocazione a livello mensile, all'interno della serie temporale, del *breakpoint* individuato a livello annuale viene stimata utilizzando una finestra di 48 mesi centrata sul *breakpoint* in esame. Nella seconda fase (*second detection*) vengono ricercate ulteriori disomogeneità a livello mensile, dopo aver eliminato i *breakpoint* identificati nella prima fase. Per calcolare i fattori di correzione viene utilizzata l'analisi ANOVA (analisi della varianza), che fornisce anche i valori finali per riempire i buchi nelle serie, causati dai dati mancanti e dall'eliminazione degli *outliers*;
3. correzioni finali: in questa fase vengono eliminate le discontinuità statisticamente non significative e l'analisi ANOVA viene nuovamente applicata con il set ridotto di *breakpoint*.

In fase di valutazione le due diverse metodologie, applicate al dataset delle serie giornaliere di temperatura, hanno fornito complessivamente risultati fra loro consistenti sia in termini di numero di *breakpoint* individuati, sia rispetto al Root Mean Square Error (RMSE) calcolato fra le serie originali e le corrispondenti serie omogeneizzate. Poiché in termini di correlazione le serie prodotte da ACMANT mostrano, all'interno di ciascun *cluster*, valori leggermente più alti di quelli associati alle serie di Climatol, nella parte restante di questo rapporto sono riportati i grafici e i risultati riferiti solo alle serie di ACMANT.

Le Figure 2.5 e 2.6 mostrano alcuni esempi di omogeneizzazione per quattro serie di temperatura minima e massima. Ciascun grafico riporta l'andamento sia della serie originale che della serie omogeneizzata. Si può osservare che i due tipi di serie, a partire da un certo anno, coincidono. Questo perché le serie vengono corrette nella parte che precede ciascuna discontinuità individuata, mantenendo inalterata la parte più recente, che si assume corretta. Così facendo le serie omogeneizzate possono essere estese, nella parte più recente, via via che nuovi dati si rendono disponibili e fino a che non si ritenga necessario ripetere l'analisi di omogeneità delle serie.

Il box-plot di Figura 2.7 mostra la distribuzione del *Root Mean Square Error* (RMSE) fra le serie omogeneizzate e le corrispondenti serie originali, per ogni *cluster*. Sia per Tmax che per Tmin la mediana è inferiore a 1°C per tutti i *cluster*. Le serie con $RMSE \geq 2^\circ C$, ovvero quelle che presentano consistenti disomogeneità (6 serie di Tmax e 5 di Tmin), sono state considerate poco affidabili e quindi escluse da ulteriori elaborazioni.

La correlazione fra le serie all'interno di ciascun *cluster* è mostrata mediante *heatmap* in Figura 2.8. Il grafico di esempio si riferisce alle serie di temperatura massima per il *cluster* 6. Appare evidente che il processo di omogeneizzazione comporta un generale aumento delle correlazioni fra le serie.

La maggiore coerenza delle serie all'interno di ciascun *cluster*, frutto del processo di omogeneizzazione, è evidente in Figura 2.9 dove sono riportate le anomalie annuali rispetto al trentennio 1961-1990 della temperatura minima delle stazioni del *cluster* 2. Un risultato analogo si evince dalla Figura 2.10. La mappa illustra i trend nel periodo 1981-2015 per Tmax, calcolati sulla base delle serie originali (in alto) e di quelle omogeneizzate (in basso). È evidente la maggiore coerenza spaziale di quest'ultima.

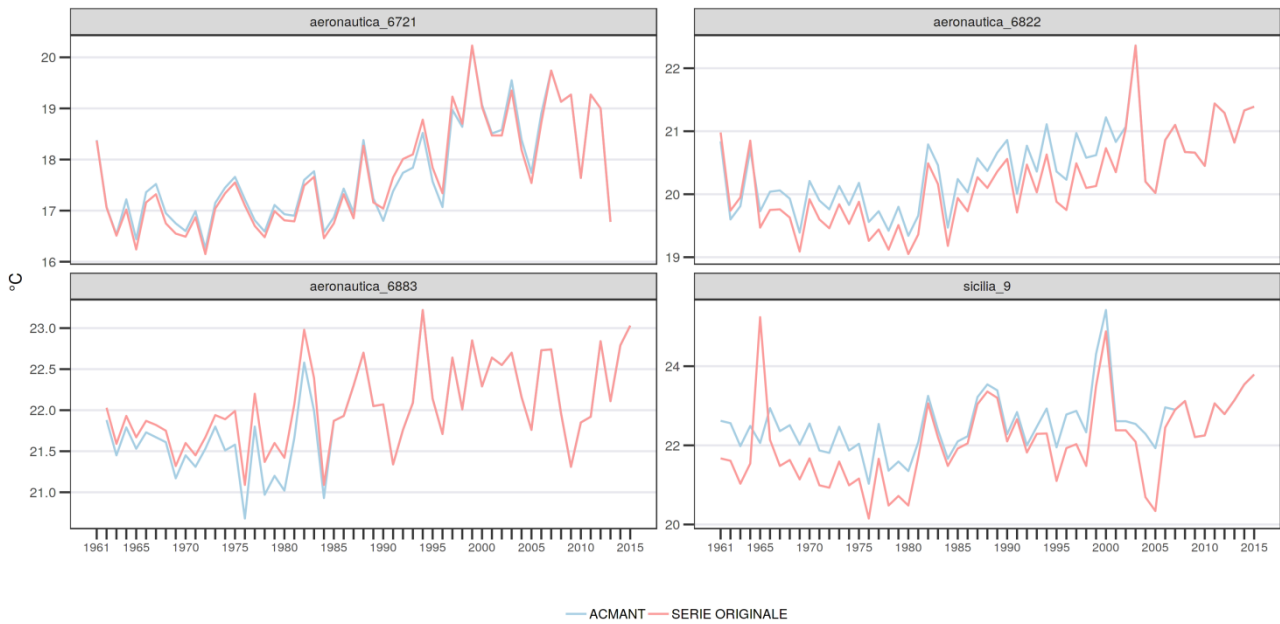


Figura 2.5 - Esempi di omogeneizzazione di serie di temperatura massima. Serie originale (in rosso) e serie omogeneizzata con ACMANT (in blu). In alto a sinistra: Brescia/Ghedi (rete sinottica, cluster 3) con 6 breakpoint (novembre 1963, dicembre 1969, novembre 1989, aprile 1998, giugno 2002 e novembre 2006); in alto a destra: Pratica di Mare (rete sinottica, cluster 1), con 3 breakpoint (gennaio 1965, ottobre 1993 e gennaio 2002); in basso a sinistra: Trapani/Birgi (rete sinottica, cluster 6) con 2 breakpoint (gennaio 1974 e maggio 1984); in basso a destra: Marsala (Regione Sicilia, Osservatorio Acque, cluster 6) con 7 breakpoint (febbraio 1966, marzo 1981, aprile 1994, luglio 2000, ottobre 2003 e aprile 2006).

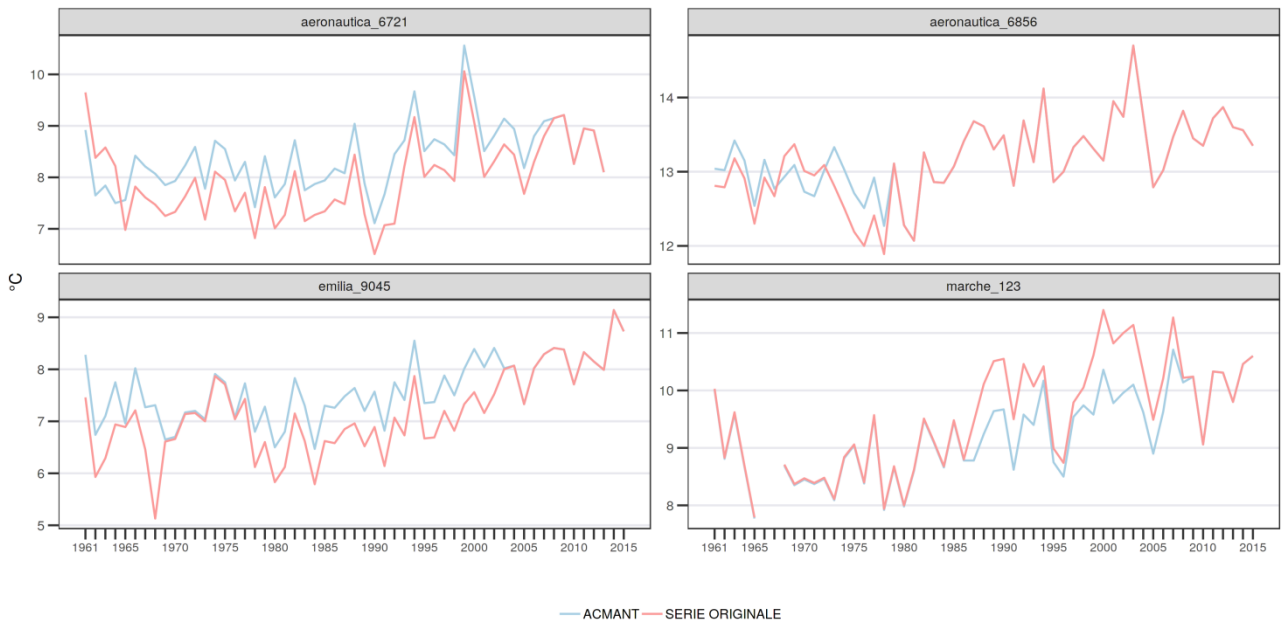


Figura 2.6 - Esempi di omogeneizzazione di serie di temperatura minima. Serie originale (in rosso) e serie omogeneizzata con ACMANT (in blu). In alto a sinistra: Brescia/Ghedi (rete sinottica, cluster 3) con 4 breakpoint (dicembre 1964, maggio 1992, novembre 1992 e aprile 2007); in alto a destra: Brindisi (rete sinottica, cluster 2), con 3 breakpoint (settembre 1967, febbraio 1972 e marzo 1978); in basso a sinistra: Bobbio (rete regionale Emilia Romagna, cluster 4) con 5 breakpoint (dicembre 1967; ottobre 1968, luglio 1977, marzo 2000 e dicembre 2002); in basso a destra: Arcevia (rete regionale Marche, cluster 3) con 5 breakpoint (marzo 1987, agosto 1993, novembre 1998, marzo 2004 e febbraio 2008).

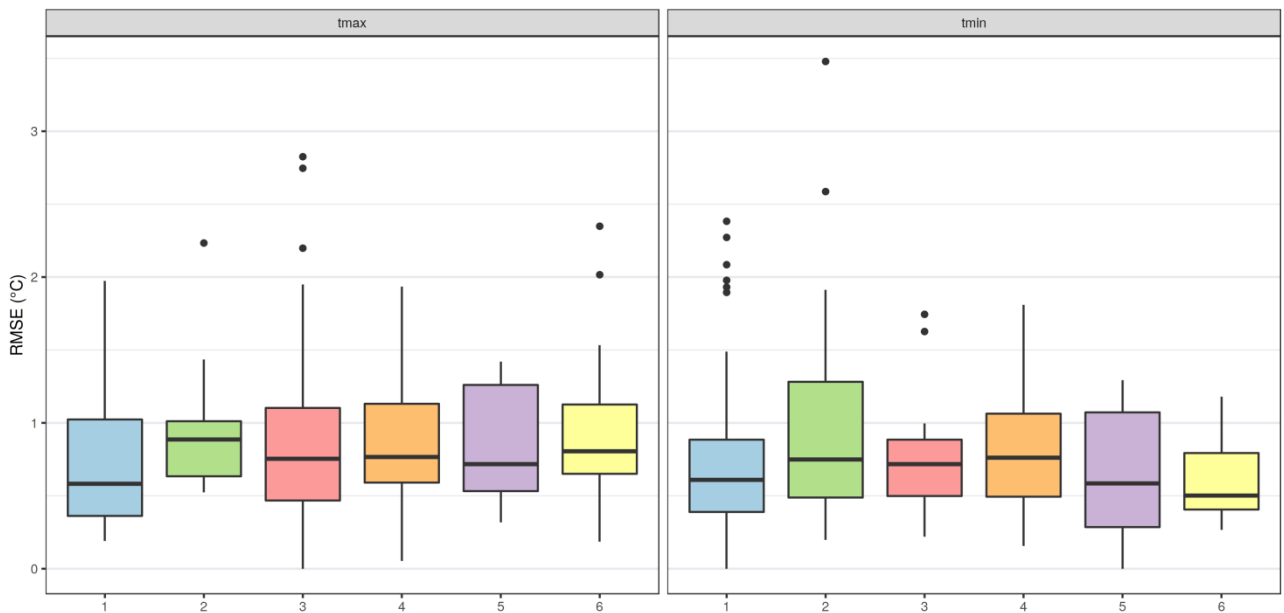


Figura 2.7 - Box-plot dell'RMSE (serie omogeneizzate rispetto alle corrispondenti serie originali) per ogni cluster. A sinistra Tmax, a destra Tmin.

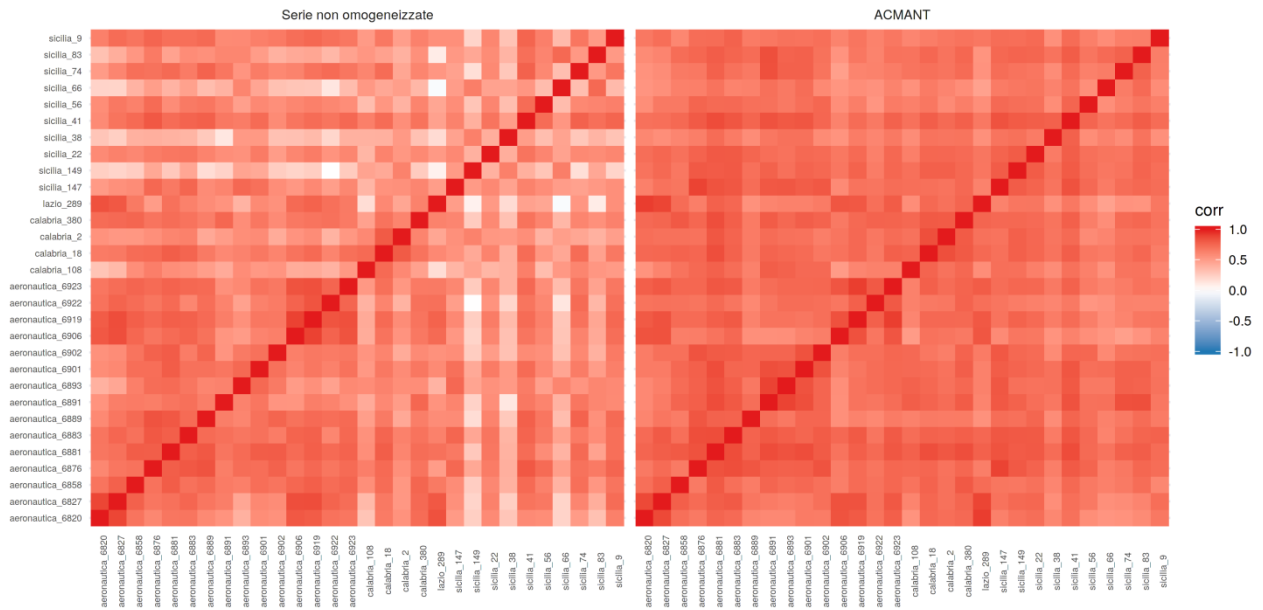


Figura 2.8 - Heatmap dei coefficienti di correlazione fra le serie di Tmax del cluster 6 prima dell'omogeneizzazione (a sinistra) e dopo l'omogeneizzazione (a destra).

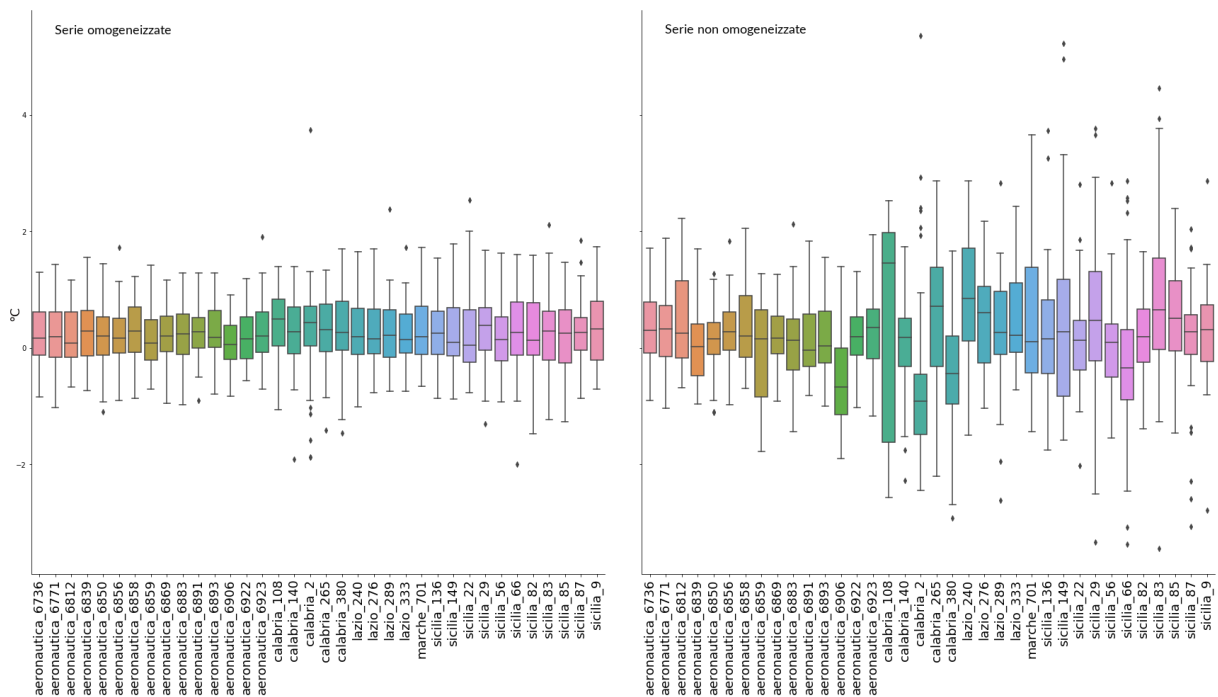


Figura 2.9 - Box-plot delle anomalie annuali di temperatura minima nel periodo 1961-2015, stazioni del cluster 2. Serie omogeneizzate (a sinistra) e serie non omogeneizzate (a destra).

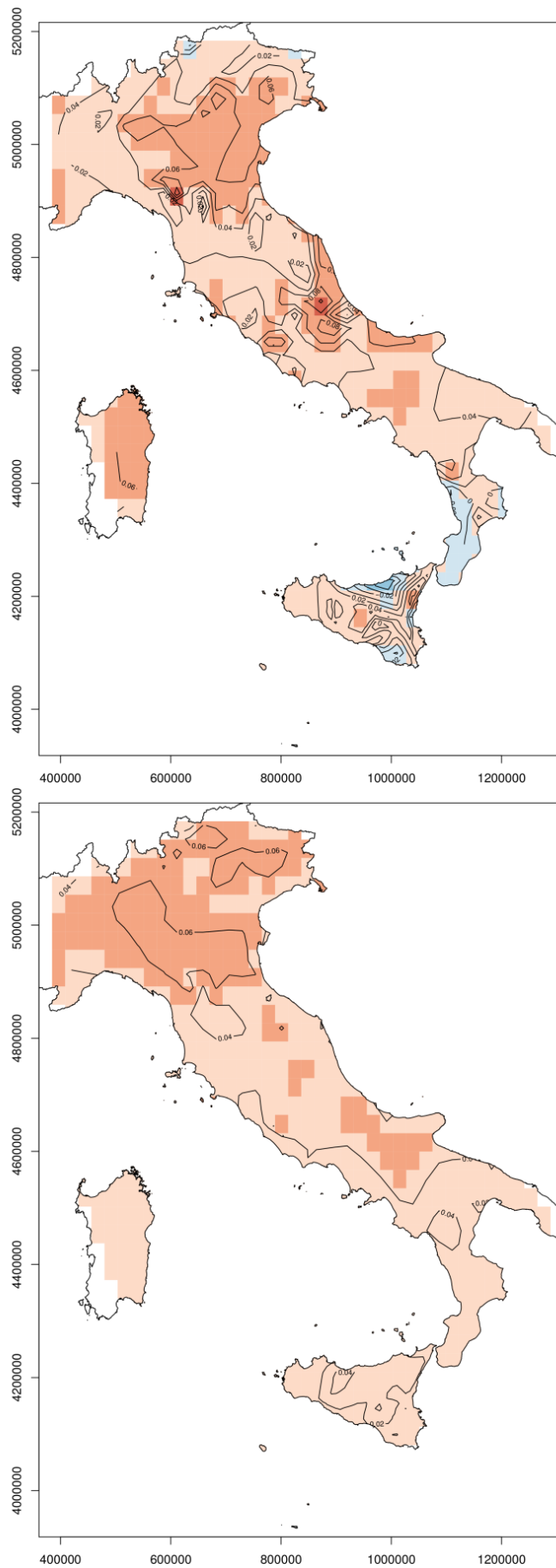


Figura 2.10 - Trend della temperatura massima ($^{\circ}\text{C}/\text{anno}$) dal 1981 al 2015. In alto: mappa dei trend delle serie non omogeneizzate; in basso: mappa dei trend delle serie prodotte da ACMANT.

3. ESEMPI DI INDICATORI ED INDICI DI TEMPERATURA

Partendo dalle serie omogeneizzate è stata ricalcolata la serie di anomalia media italiana di temperatura (massima e minima) dal 1961 al 2015. Poiché la distribuzione delle stazioni sul territorio italiano non è propriamente uniforme, si è proceduto al calcolo delle anomalie delle singole stazioni, alla loro spazializzazione su griglia regolare e quindi al calcolo del valore di anomalia medio annuale. Questa metodologia garantisce una media rappresentativa dell'intero territorio nazionale, anche quando la disponibilità dei dati delle stazioni non presenta una distribuzione uniforme.

A titolo di esempio in Figura 3.1 è mostrata la mappa di anomalia della temperatura massima per il 2015 rispetto alla media 1961-1990. Le anomalie sono riportate su una griglia a risoluzione 50 km x 50 km ottenuta mediante *ordinary kriging* (Li and Heap, 2008).

Le serie delle anomalie medie nazionali di temperatura 1961-2015 calcolate utilizzando le serie giornaliere omogeneizzate (serie versione 2) sono mostrate in Figura 3.2 (temperatura media), 3.3 (temperatura minima) e 3.4 (temperatura massima). Tutti e tre i grafici confermano il riscaldamento nel corso del periodo analizzato già evidenziato nelle serie di anomalie calcolate in precedenza, sulla base delle serie valutate omogenee a livello annuale secondo l'RHtest (serie versione 1), con valori di anomalia quasi sempre positivi a partire dagli anni '80. Lievi differenze nelle stime di anomalia e nella classifica degli anni più caldi fra le serie della versione 2 e le corrispondenti della versione 1 possono essere attribuite al differente set di stazioni utilizzato e alle procedure di elaborazione dei dati.

La Tabella 3.1 riporta i trend stimati per le nuove serie di temperatura e per quelle della versione 1, sull'intero periodo 1961-2015. Le nuove stime mettono in evidenza trend positivi e statisticamente significativi per tutte le variabili, come già ottenuto dalle precedenti elaborazioni. Il trend della nuova serie risultano leggermente più bassi rispetto alle stime precedenti, con differenze di -0.04 °C/10anni per T_{min} e -0.02 °C per T_{max}.

Per quanto riguarda gli estremi di temperatura, utilizzando il nuovo dataset di serie giornaliere omogeneizzate, sono state ricalcolate le serie nazionali relative ad un set di indici estremi definiti dall'ETCCDI (Expert Team on Climate Change Detection and Indices) della Commissione per la Climatologia della WMO, già selezionati e descritti in Desiato et al. (2013). In generale i risultati mostrano chiaramente la tendenza all'aumento degli indici relativi agli estremi di caldo e alla riduzione di quelli relativi agli estremi di freddo negli ultimi 50 anni, in coerenza con le precedenti stime. Nella Figura 3.5 sono riportate le serie delle anomalie rispetto alla media 1961-1990 dei giorni e delle notti calde (percentuale di giorni in un anno rispettivamente con T_{max} > 90° percentile e T_{min} > 90° percentile della corrispondente distribuzione nel periodo climatologico); nella Figura 3.6 sono mostrate le serie delle anomalie dei giorni e delle notti fredde (percentuale di giorni in un anno rispettivamente con T_{max} < 10° percentile e T_{min} < 10° percentile della corrispondente distribuzione sul periodo climatologico). Gli anni in cui si osservano le maggiori anomalie positive di giorni caldi (2003, 2007, 2012, 2015) e delle notti calde (2003, 2012, 2014, 2015) corrispondono a quelli evidenziati dalle precedenti serie, sebbene con alcune differenze nei singoli valori di anomalia. Lo stesso avviene per i giorni e le notti fredde.

Tabella 3.1 - Trend stimati (e relative deviazioni standard) della temperatura in Italia dal 1961 al 2015; serie nazionali calcolate con la nuova metodologia (versione 2) e con la precedente (versione 1). Tutti i trend sono statisticamente significativi al livello del 5%.

	T _{max} (°C/decade)	T _{min} (°C/decade)	T _{media} (°C/decade)
versione 2	0.32±0.04	0.24±0.03	0.28±0.03
versione 1	0.33±0.04	0.28±0.03	0.30±0.03

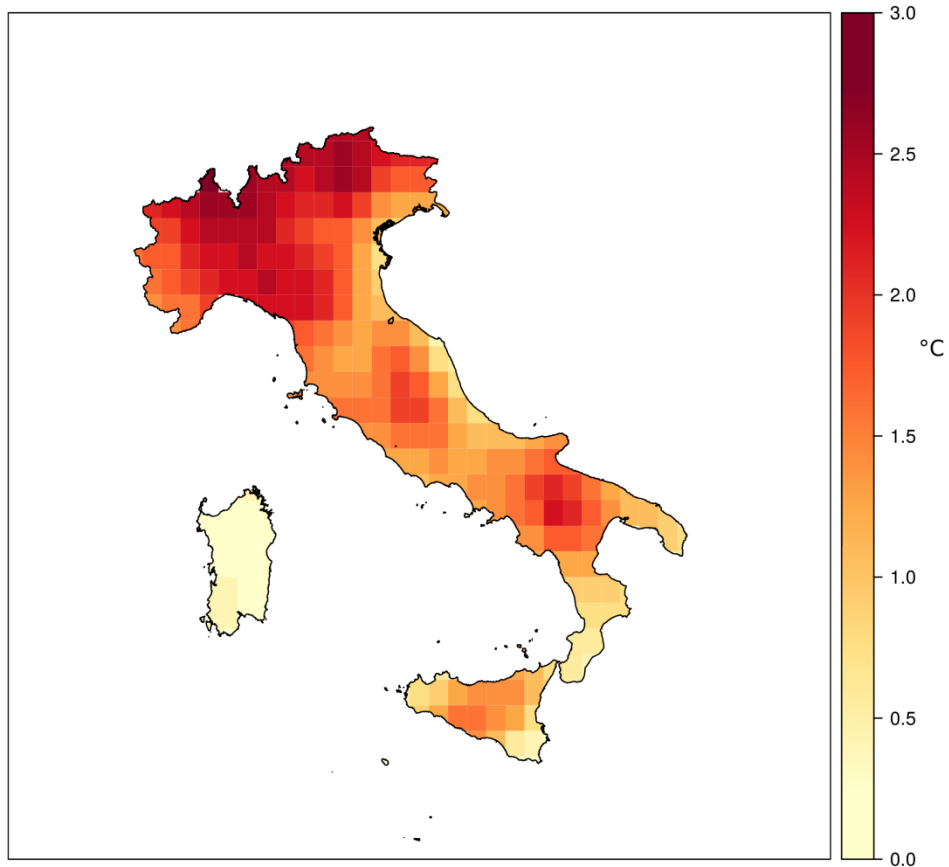


Figura 3.1 - Mappa di anomalia della temperatura massima per il 2015, rispetto al valore normale 1961-1990.

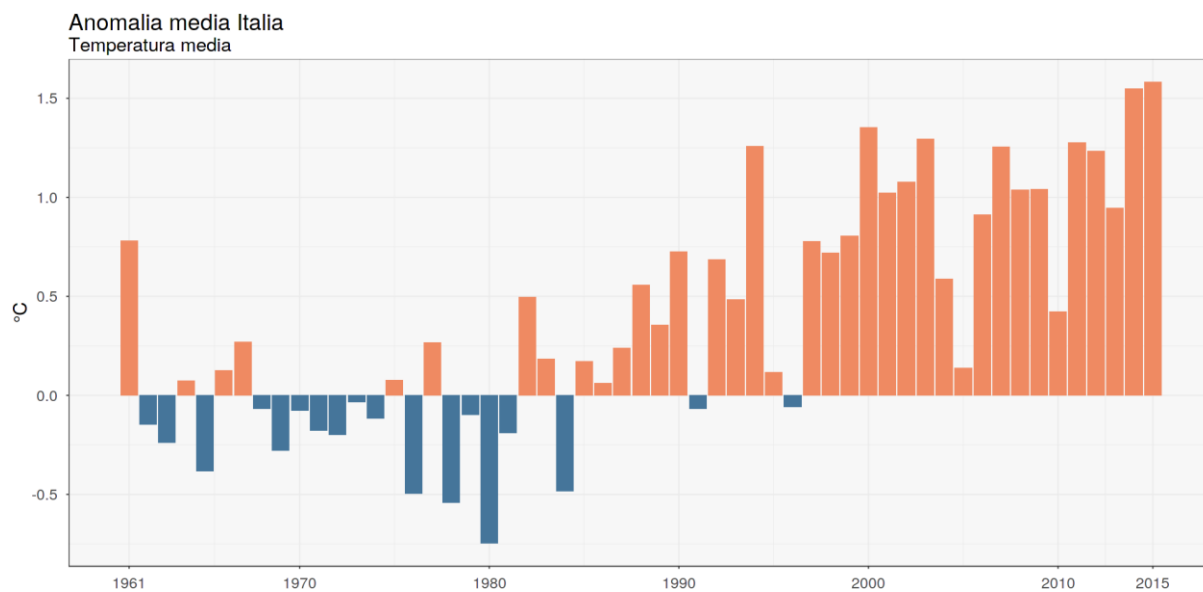


Figura 3.2 - Serie delle anomalie medie in Italia della temperatura media rispetto al valore normale 1961-1990.

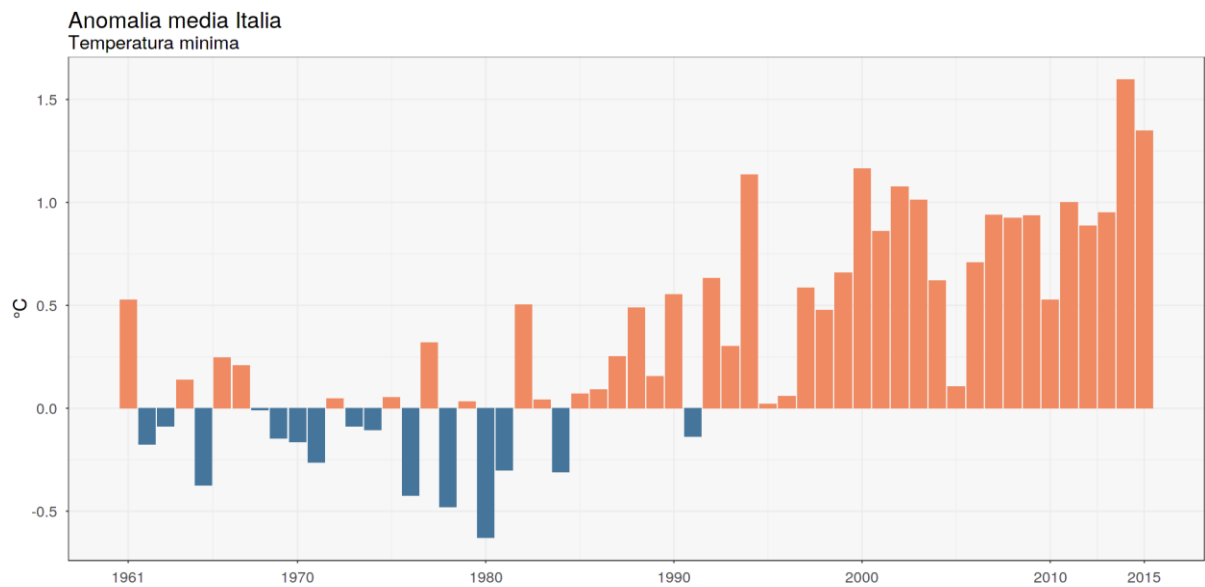


Figura 3.3 - Serie delle anomalie medie in Italia della temperatura minima rispetto al valore normale 1961-1990.

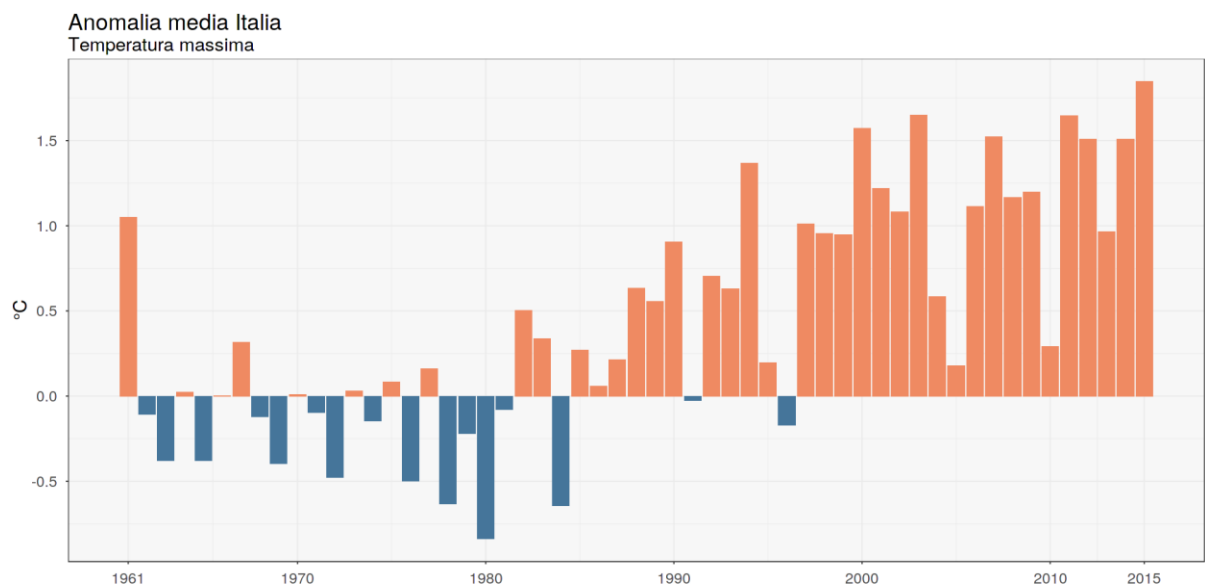


Figura 3.4 - Serie delle anomalie medie in Italia della temperatura massima rispetto al valore normale 1961-1990.

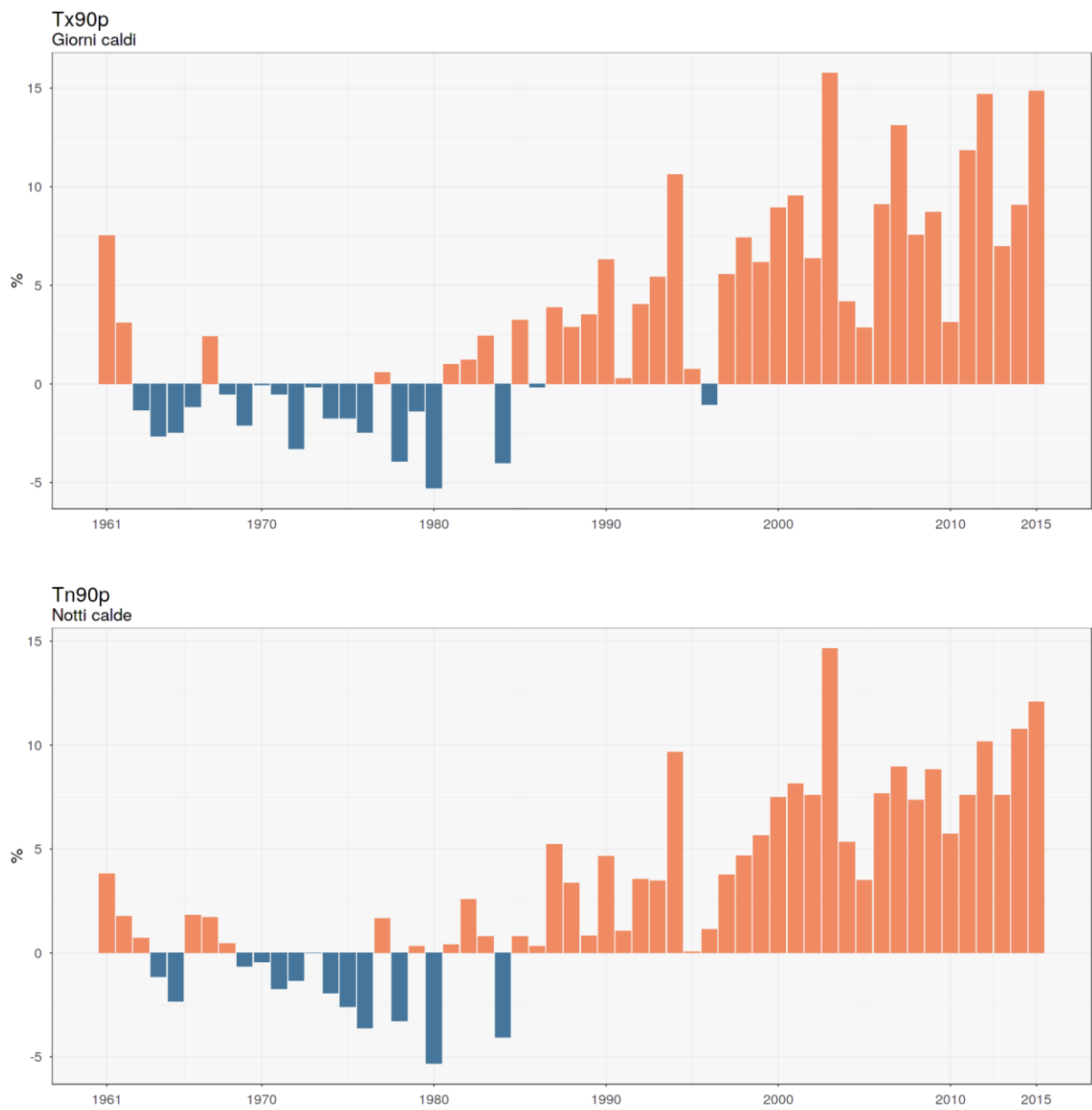


Figura 3.5 - Serie delle anomalie medie annuali del numero di giorni caldi (in alto) e di notti calde (in basso), espresso in percentuale di giorni/anno, rispetto al valore normale 1961-1990.



Figura 3.6 - Serie delle anomalie medie annuali del numero di giorni freddi (in alto) e di notti fredde (in basso), espresso in percentuale di giorni/anno, rispetto al valore normale 1961-1990.

4. CONCLUSIONI

In questo rapporto è stata descritta la nuova metodologia di selezione e omogeneizzazione delle serie utili al calcolo delle anomalie medie nazionali di temperatura, sia dei valori medi che degli estremi. Il nuovo approccio è stato applicato al dataset nazionale di temperatura con passo giornaliero, recentemente realizzato integrando serie temporali provenienti da diverse fonti.

Rispetto alle precedenti elaborazioni si segnalano i seguenti miglioramenti: 1) il nuovo dataset di serie con passo giornaliero è più ricco di stazioni per il periodo in esame rispetto al precedente e quindi più rappresentativo dell'intero territorio nazionale; 2) è stata rivista ed aggiornata la procedura di omogeneizzazione delle serie di temperatura, applicando una nuova metodologia che include la regionalizzazione climatica (*cluster analysis*) e la correzione a livello giornaliero delle serie disomogenee.

Il confronto fra le nuove serie nazionali di temperatura e quelle riportate nei precedenti lavori mostra alcune piccole differenze nei valori di anomalia, dovute al diverso dataset utilizzato e alla diversa metodologia di calcolo. I nuovi risultati confermano sostanzialmente le stime precedenti, indicando un chiaro segnale di riscaldamento negli ultimi 50 anni sia per i valori medi di temperatura che per gli estremi. I nuovi trend di temperatura (massima, minima e media) sono tutti positivi e statisticamente significativi, in accordo con le precedenti stime; le nuove stime indicano valori leggermente più bassi (+0.24°C/10 anni per T_{min} e +0.28°C/10 anni per T_{max}) rispetto a quelli calcolati sulle serie precedenti.

BIBLIOGRAFIA

- Aguilar E., Auer I., Brunet M., Peterson T.C., Wieringa J., 2003. Guidelines on climate metadata and homogenization. WCDMP-No. 53, WMO-TD No. 1186. World Meteorological Organization, Geneva.
- Alexander L.V., 2016. Global observed long-term changes in temperature and precipitation extremes: a review of progress and limitations in IPCC assessments and beyond. *Weather and Climate Extremes*, 11: 4–16. doi: 10.1016/j.wace.2015.10.007.
- Alexander L. V., Zhang X., Peterson T. C., Caesar J., Gleason B., Klein Tank A.M.G., Haylock M., Collins D., Trewin B., Rahimzadeh F., Tagipour A., Rupa Kumar K., Revadekar J., Griffiths G., Vincent L., Stephenson D.B., Burn J., Aguilar E., Brunet M., Taylor M., New M., Zhai P., Rusticucci M., Vazquez-Aguirre J.L., 2006. Global observed changes in daily climate extremes of temperature and precipitation. *J. Geophys. Res.*, 111, D05109, doi:10.1029/2005JD006290.
- Alexandersson H., 1986. A homogeneity test applied to precipitation data. *J. Climatol.*, 6: 661–675. doi: 10.1002/joc.3370060607.
- Beaulieu C., Seidou O., Ouarda T.B.M.J., Zhang X., Boulet G., Yagouti A., 2008. Intercomparison of homogenization techniques for precipitation data. *Water Resour. Res.*, 44, doi:10.1029/2006WR005615.
- Calmanti S., Dell'Aquila A., Maimone F., Pelino V., 2015. Evaluation of climate patterns in a regional climate model over Italy using long-term records from SYNOP weather stations and cluster analysis. *Clim. Res.* 62:173-188. doi:10.3354/cr01256.
- Cassinus H. and Mestre O., 2004. Detection and Correction of Artificial Shifts in Climate Series, *J. R. Stat. Soc.*, series C, 53: 405-425. doi:10.1111/j.1467-9876.2004.05155.x.
- Desiato F., Lena F. e Toreti A., 2006. Un sistema per tutti – climatologia: i dati italiani. *Sapere*, Anno 72, n. 2: 62-69.
- Desiato F., Lena F. e Toreti A., 2007. SCIA: a system for a better knowledge of the Italian climate. *Bollettino di Geofisica Teorica ed Applicata*, 48, n. 3: 351-358.
- Desiato F., 2007. SCIA. un sistema nazionale unico per i dati climatologici. *ARPA Rivista*, N. 1, gennaio-febbraio 2007.
- Desiato F., Fioravanti G., Frascchetti P., Perconti W. e Toreti A., 2011. Climate indicators for Italy: calculation and dissemination. *Adv. Sci. Res.*, 6: 147-150. doi:10.5194/asr-6-147-2011.
- Desiato F., Fioravanti G., Frascchetti P., Perconti W., Piervitali E., 2012. Elaborazione delle serie temporali per la stima delle tendenze climatiche. *Rapporto ISPRA / Stato dell'Ambiente 32/2012*.
- Desiato F., Fioravanti G., Frascchetti P., Perconti W., Piervitali E., 2013. Variazioni e tendenze degli estremi di temperatura in Italia, *Rapporto ISPRA / Stato dell'Ambiente 37/2013*.
- Desiato F., Fioravanti G., Frascchetti P., Perconti W., Piervitali E., Pavan V., 2016. Gli indicatori del clima in Italia nel 2015. *Rapporto ISPRA / Stato dell'Ambiente 65/2016*.
- Desiato F., Fioravanti G., Frascchetti P., Perconti W., Piervitali E., Pavan V., 2017. Gli indicatori del clima in Italia nel 2016. *Rapporto ISPRA / Stato dell'Ambiente 72/2017*.
- Donat M.G., Alexander L.V., Yang H., Durre I., Vose R., Dunn R., Willett K., Aguilar E., Brunet M., Caesar J., Hewitson B., Jack C., Klein Tank A.M.G., Kruger A.C., Marengo J., Peterson T.C., Renom M., Oria Rojas C., Rusticucci M., Salinger J., Sekele S.S., Srivastava A.K., Trewin B., Villarroya C., Vincent L.A., Zhai P., Zhang X., Kitching S., 2013. Updated analyses of temperature and precipitation extreme indices since the beginning of the twentieth century: The HadEX2 dataset. *J. Geophys. Res.* 118: 2098-2118. doi: 10.1002/jgrd.50150.
- Domonkos P., 2011. Adapted Caussinus-Mestre algorithm for networks of temperature series (OMO). *Int. J. Geosci.*, 2, 293–309, doi: 10.4236/ijg.2011.23032.
- Domonkos P., Poza, R. and Efthymiadis D., 2011. Newest developments of OMO, *Adv. Sci. Res.*, 6, 7-11, <https://doi.org/10.5194/asr-6-7-2011>.
- Domonkos P. and Coll J., 2017. Homogenisation of temperature and precipitation time series with OMO3: method description and efficiency tests. *Int. J. Climatol.*, 37: 1910–1921. doi:10.1002/joc.4822.
- Durre I., Menne M.J., Gleason B.E., Houston T.G. and Vose R.S., 2010. Comprehensive automated quality assurance of daily surface observations, *J. Appl. Meteor. Climatol.*, 8, 1615–1633. doi: 10.1175/2010JAMC2375.1.
- Easterling D.R., Kunkel K.E., Wehner M.F., Sun L., 2016. Detection and attribution of climate extremes in the observed record. *Weather and Climate Extremes*, 11: 17-27. doi: 10.1016/j.wace.2016.01.001.

- Easterling D.R., Kunkel K.E., Wehner M.F., Sun L., 2016. Detection and attribution of climate extremes in the observed record. *Weather Clim. Extrem.* 11, 17–27. doi: 10.1016/j.wace.2016.01.001.
- Fioravanti G., Fioravanti G., Piervitali E., Desiato F., 2015. Recent changes of temperature extremes over Italy: an index-based analysis, *Theor. Appl. Climatol.*, 123: 473–486. doi:10.1007/s00704-014-1362-1.
- Fioravanti G., Frascchetti P., Perconti W., Piervitali E., Desiato F., 2016. Controlli di qualità delle serie di temperatura e precipitazione. Rapporto ISPRA / Stato dell’Ambiente 66/2016.
- Geng Q., Wu P., Zhao X., Wang Y., 2014. Comparison of classification methods for the divisions of wet/dry climate regions in Northwest China. *Int. J. Climatol.*, 34: 2163–2174. doi: 10.1002/joc.3827.
- Gottelman A. and Rood R. B., 2016. *Demystifying Climate Models. A Users Guide to Earth System Models.* Springer. doi: 10.1007/978-3-662-48959-8_1.
- Guijarro J. A., 2014. User’s guide to climatol: an R contributed package for homogenization of climatological series. Balearic Islands Office, State Meteorological Agency, available at <http://www.climatol.eu/>
- Guijarro J.A., 2016. Package ‘climatol’. <https://cran.r-project.org/web/packages/climatol/climatol.pdf>.
- Heim R.R., 2015. An overview of weather and climate extremes – Products and trends. *Weath. and Clim. Extremes*, 10: 1–9. doi: 10.1016/j.wace.2015.11.001.
- Hunziker S., Brönnimann S., Calle J., Moreno I., Andrade M., Ticona L., Huerta A., Lavado-Casimiro W., 2018. Effects of undetected data quality issues on climatological analyses. *Clim. Past*, 14, 1–20. 10.5194/cp-14-1-2018.
- Gareth J., Witten D., Hastie T., Tibshirani R., 2013. *An Introduction to Statistical Learning with applications in R.* Springer. doi: 10.1007/978-1-4614-7138-7.
- Klein Tank A.M.G. and Können G.P., 2003. Trends in Indices of Daily Temperature and Precipitation Extremes in Europe, 1946–99. *J. Climate.*, 16, 3665–3680. doi: 10.1175/1520-0442(2003)016<3665:TIODT>2.0.CO;2.
- Klein Tank A. M.G., Zwiers F. W., Zhang X., 2009. Guidelines on Analysis of extremes in a changing climate in support of informed decisions for adaptation. *Climate Data and Monitoring WCDMP*, 72, WMO-TD N. 1500, 56pp.
- Köppen W., 1936. *Das geographische System der Klimate (The geographic system of climates).* – In: Köppen, W. R. Geiger (Hrsg.): *Handbuch der Klimatologie*, Bd. 1, Teil C. – Born-traeger, Berlin, 44 pp.
- Li J., Heap A.D., 2008. A review of spatial interpolation methods for environmental scientists. *Geoscience Australia, Record 2008/23*, 137 pp.
- Lawrimore J.H., Menne M.J., Gleason B.E., Williams C.N., Wuertz D.B., Vose R. S. and Rennie J., 2011. An overview of the Global Historical Climatology Network monthly mean temperature data set, version 3. *J. Geophys. Res.*, 116, D19121, doi:10.1029/2011JD016187.
- Luna M. Y., Guijarro J.A. and López J.A., 2012. A monthly precipitation database for Spain (1851–2008): reconstruction, homogeneity and trends. *Adv. Sci. Res.*, 8, 1–4. doi: 10.5194/asr-8-1-2012.
- Mamara A., Argiriou A.A. and Anadranistakis M., 2013. Homogenization of mean monthly temperature time series of Greece. *Int. J. Climatol*, 33: 2649–2666. doi:10.1002/joc.3614.
- Menne M. J. and Williams C.N., Jr., 2009. Homogenization of temperature series via pairwise comparisons. *J. Climate*, 22: 1700–1717. doi:10.1175/2008JCLI2263.1.
- Mestre O., Domonkos P., Picard F., Auer I., Robin S., Lebarbier E., Böhm R., Aguilar E., Guijarro J., Vertachnik G., Klancar M., Dubuisson B. and Stepanek P., 2013. HOMER: a homogenization software – methods and applications, *Idojaras*, 117, 47–67.
- Muñoz-Díaz D. and Rodrigo F., 2004. Spatio-temporal patterns of seasonal rainfall in Spain (1912–2000) using cluster and principal component analysis: comparison. *Ann. Geophys.*, 22: 1435–1448. doi:10.5194/angeo-22-1435-2004.
- Pérez-Zanón N., Sigró J., Domonkos P., Ashcroft L., 2015. Comparison of HOMER and OMO homogenization methods using a central Pyrenees temperature dataset. *Adv. Sci. Res.*, 12, 111–119. doi: 10.5194/asr-12-111-2015.
- Peterson T.C. and Easterling D.R., 1994. Creation of homogeneous composite climatological reference series. *Int. J. Climatol.* 14, 671–679. doi: 10.1002/joc.3370140606.
- Peterson T.C., Easterling D.R., Karl T.R., Groisman P., Nicholls N., Plummer N., Torok S., Auer I., Böhm R., Gullett D., Vincent L., Heino R., Tuomenvirta H., Mestre O., Szentimrey T., Salinger J., Førland E., Hanssen-Bauer I., Alexandersson H., Jones P., Parker D., 1998. Homogeneity

-
- Adjustments of 'In Situ' Atmospheric Climate Data: A Review. *Int. J. Climatol.*, 18: 1493-1518. doi:10.1002/(SICI)1097-0088(19981115)18:13<1493::AID-JOC329>3.0.CO; 2-T.
- R Core Team, 2015. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.
- Reeves, J., Chen, J., Wang, X.L., Lund R., and Lu Q., 2007. A review and comparison of changepoint detection techniques for climate data. *J. Appl. Meteorol. Climatol.* 46, 900–915. doi: 10.1175/JAM2493.1.
- Rubel F., Brugger K., Haslinger K., Auer I., 2016. The climate of the European Alps: Shift of very high resolution Köppen-Geiger climate zones 1800–2100. *Meteorologische Zeitschrift*, 26: 115–125. doi: 10.1127/metz/2016/0816.
- Szentimrey T. 1999. Multiple Analysis of Series for Homogenization (MASH). Second Seminar for Homogenization of Surface Climatological Data, WMO-WCDMP-41, 27–46.
- Thorne P.W., Willett K.M., Allan R.J., Bojinski S., Christy J.R., Fox N., Gilbert S., Jolliffe I., Kennedy J.J., Kent E.C., Klein Tank A., Lawrimore J., Parker D.E., Rayner N., Simmons A., Song L., Stott P.A., Trewin B., 2011. Guiding the creation of a comprehensive surface temperature resource for twenty-first-century climate science. *Bull. Amer. Meteorol. Soc.*, 92, ES40–ES47. doi:10.1175/2011BAMS3124.1.
- Toreti A. and Desiato F., 2008. Temperature trend over Italy from 1961 to 2004. *Theor. Appl. Climatol.*, 91, 51-58. doi:10.1007/s00704-006-0289-6.
- Unal Y., Kindap T., Karaca M., 2003. Redefining the climate zones of Turkey using cluster analysis. *Int. J. Climatol.* 23: 1045–1055. doi:10.1002/joc.910.
- Venema V., Mestre O., Aguilar E., Auer I., Guijarro J.A., Domonkos P., Vertacnik G., Szentimrey T., Stepanek P., Zahradnicek P., Viarre J., Müller-Westermeier G., Lakatos M., Williams C.N., Menne M., Lindau R., Rasol D., Rustemeier E., Kolokythas K., Marinova T., Andresen L., Acquaforte F., Fratianni S., Cheval S., Klancar M., Brunetti M., Gruber C., Duran M.P., Likso T., Esteban P., and Brandsma T., 2012. Benchmarking monthly homogenization algorithms. *Clim. Past* 8, 89–115. doi:10.5194/cp-8-89-2012.
- Wang, X.L., Chen H., Wu Y., Feng Y. and Pu Q., 2010. New techniques for detection and adjustment of shifts in daily precipitation time series. *J. Appl. Met. Climatol.*, no. 49, pp. 2416–2436. doi:10.1175/2010JAMC2376.1.
- Ward J. H., 1963. Hierarchical grouping to optimise an objective function. *J. Amer. Stat. Assoc.*, 58, 236–244. doi: 10.2307/2282967.
- WMO, 2011. Seventh Seminar for Homogenization and Quality Control in Climatological Databases jointly organized with the Meeting of COST ES0601 (HOME) Action MC Meeting (Budapest, Hungary, 24-27 October 2011). *Climate Data and Monitoring, WCDMP-No. 78*.
- WMO, 2014. Eighth Seminar for Homogenization and Quality Control in Climatological Databases and Third Conference on Spatial Interpolation Techniques in Climatology and Meteorology. *Climate Data and Monitoring, WCDMP-No. 84*.

APPENDICE

Per definire i *cluster* nei quali suddividere le stazioni di temperatura (massima e minima) è stato esaminato il dendrogramma, cercando di identificare *cluster* con un buon numero di stazioni, ma non eccessivamente numerosi. In questo lavoro si è preso come soglia per suddividere un *cluster* in sottocluster il superamento di 99 unità (serie), il massimo numero di serie accettato dal software di omogeneizzazione ACMANT.

Nel caso della temperatura massima (Figura A.1), per esempio, la sola analisi visiva del dendrogramma potrebbe suggerire una divisione delle stazioni in 7 gruppi, tuttavia con questa suddivisione sono presenti 3 gruppi con più di 99 unità. Si è preferito quindi partire da un'aggregazione in 6 gruppi e sottoporre nuovamente a *cluster analysis* i gruppi 1 e 3, che risultano molto numerosi. Poiché l'analisi ha portato alla suddivisione del gruppo 1 in due sottocluster (1.1 e 1.2) e del gruppo 3 in 5 sottocluster (3.1, 3.2, 3.3, 3.4 e 3.5), la suddivisione finale del dataset di temperatura massima consiste in 11 *cluster*. Un approccio analogo è stato adottato per la temperatura minima: dopo la prima suddivisione delle stazioni in 6 *cluster*, i gruppi 1, 3 e 4 sono stati ulteriormente suddivisi rispettivamente in 3, 2 e 2 sottocluster, per un totale di 10 *cluster*.

Questa suddivisione garantisce una buona densità di stazioni all'interno di ogni *cluster*, presupposto fondamentale per il successivo processo di omogeneizzazione, basato sul confronto fra serie candidata e serie limitrofe. Mediamente il numero di stazioni limitrofe per serie candidata è di 12 stazioni nel raggio di 100 km e di 22 stazioni nel raggio di 200 km sia per Tmax che per Tmin (media calcolata su tutti i *cluster*).

Inoltre, sia per Tmax che per Tmin, la correlazione fra coppie di serie all'interno dei vari *cluster* si mantiene al di sopra di 0.5 fino a distanze di circa 400 km; questo permette di costruire valide serie di riferimento per ogni serie candidata. Nelle Figure A.2a e A.2b sono mostrati a titolo di esempio i correlogrammi delle differenze prime delle serie mensili per la Tmax, che riportano il coefficiente di correlazione fra coppie di stazioni in funzione della loro reciproca distanza, per gli 11 *cluster*. I coefficienti di correlazione più elevati si riferiscono ai *cluster* 4 e 5, che rappresentano le stazioni di montagna.

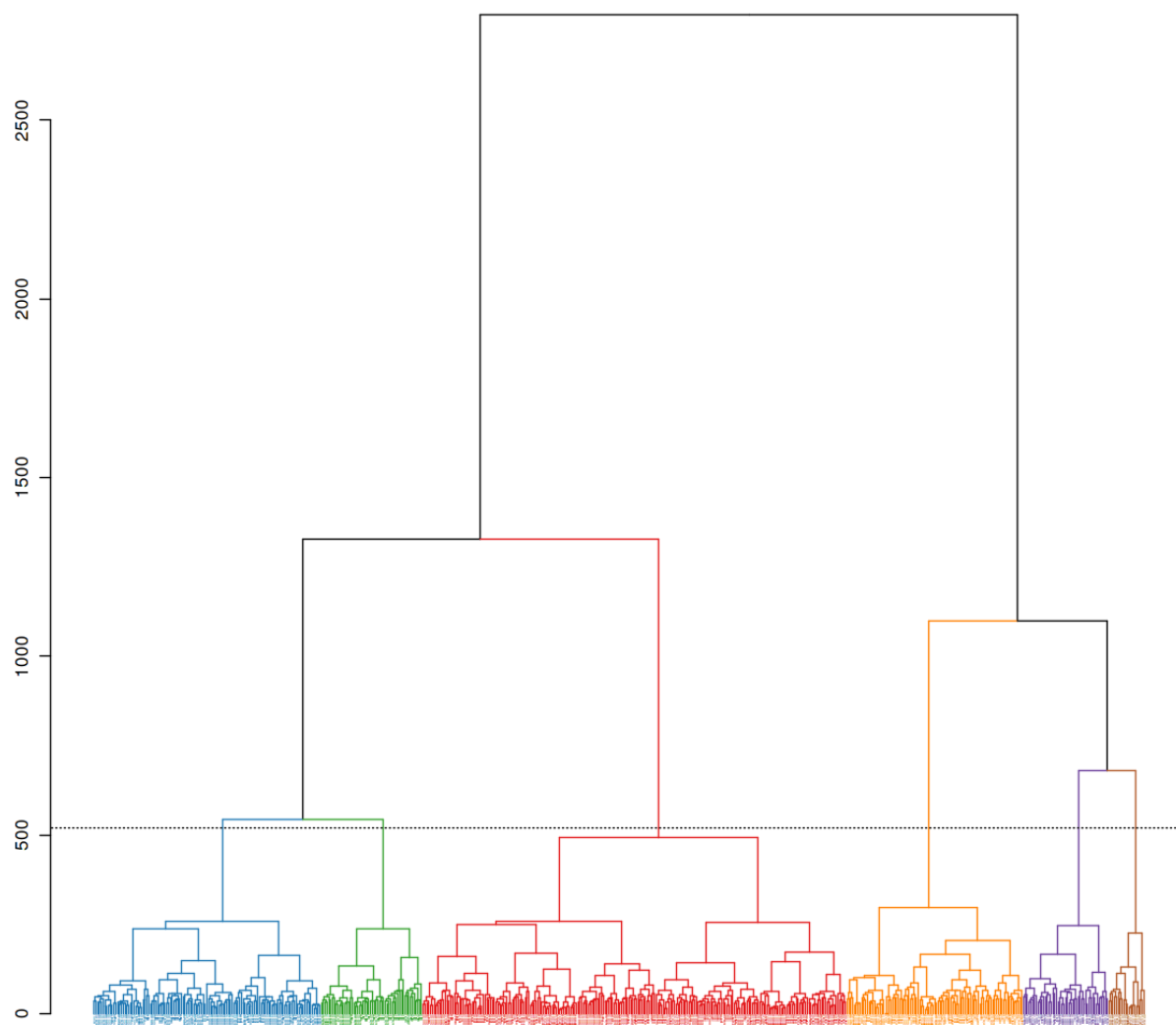


Figura A.1 - Dendrogramma relativo alla cluster analysis delle serie mensili di temperatura massima. La linea tratteggiata divide il dataset in sei cluster.

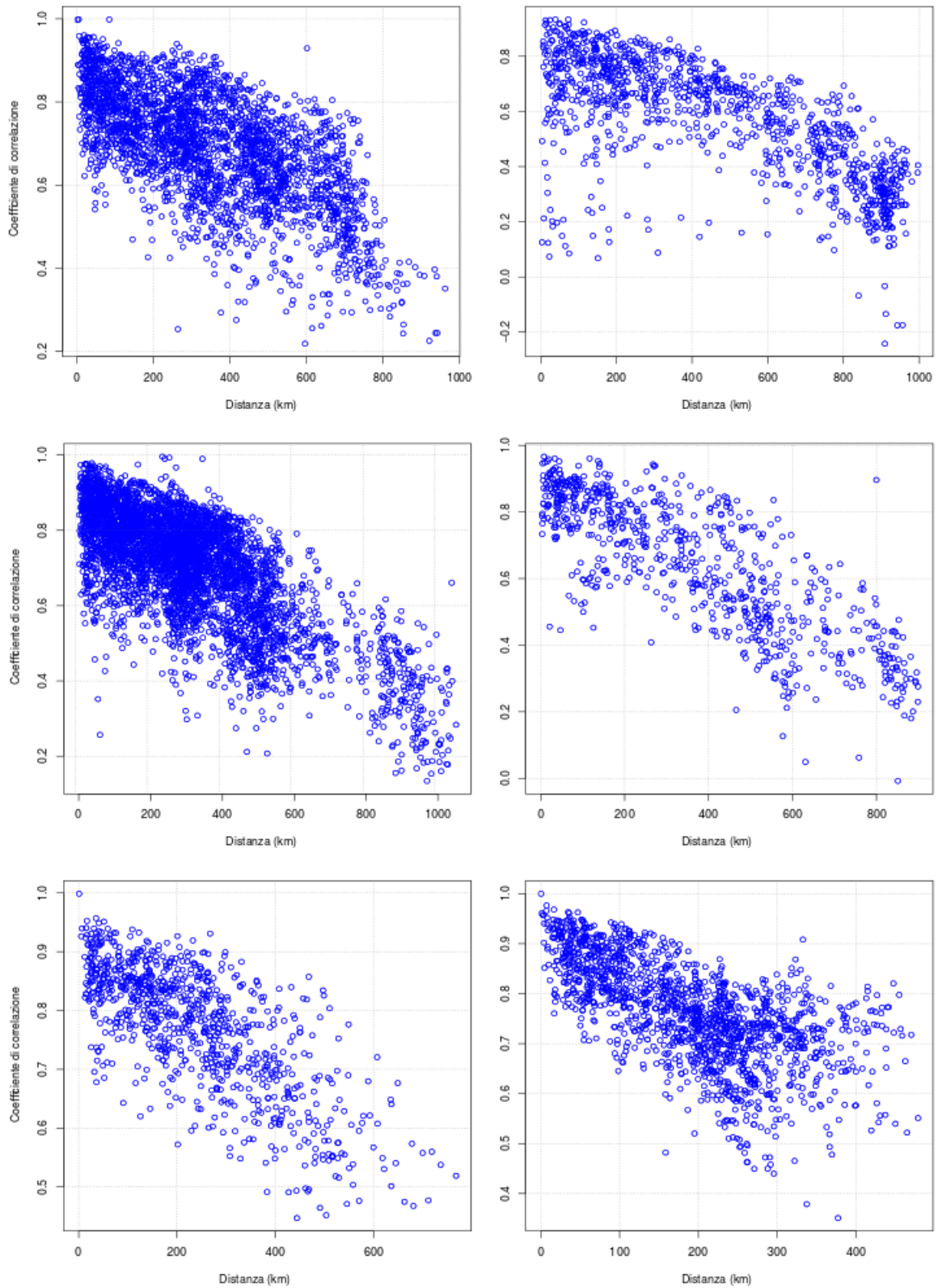


Figura A.2a - Correlogramma delle differenze prime delle serie mensili di coppie di stazioni all'interno di ogni cluster. Da sinistra a destra: cluster 1.1 e 1.2 in alto, cluster 2 e 3.1 al centro, cluster 3.2 e 3.3 in basso.

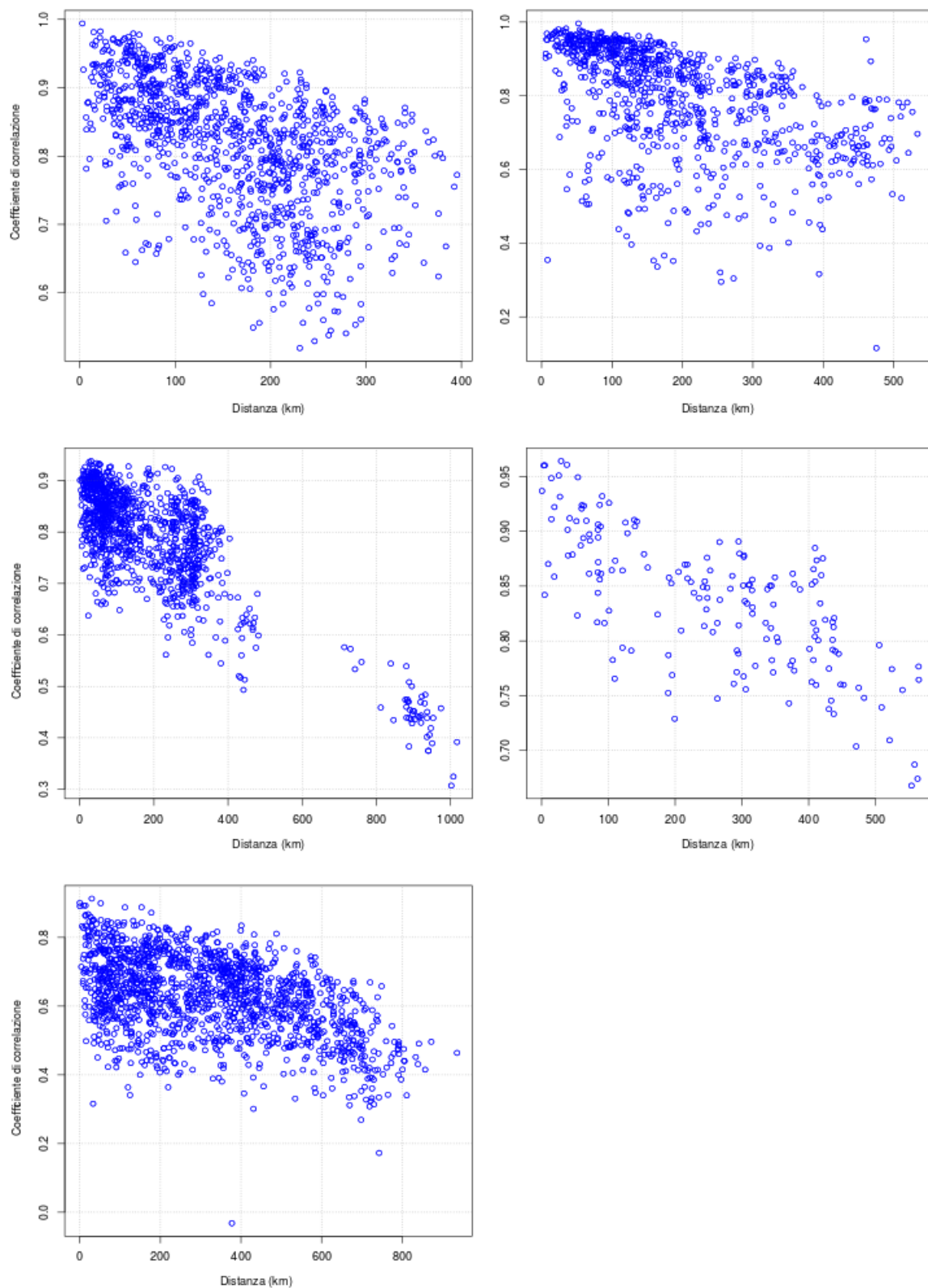


Figura A.2b - Come in figura A.2a, per i cluster 3.4 e 3.5 in alto, 4 e 5 al centro, 6 in basso