

POLITECNICO DI TORINO

Collegio di Ingegneria Gestionale
Corso di Laurea Magistrale in Ingegneria Gestionale

Tesi di Laurea Magistrale

Analisi dei modelli di diffusione delle certificazioni ISO 9000 in Europa



Relatori

Prof. Fiorenzo Franceschini
Prof. Domenico Augusto Francesco Maisano
Prof. Luca Mastrogiacomo

Candidato

Antonio Carrozza

Aprile 2018

DEDICA

A Flavio, Teresa e Anna,
per il loro continuo supporto.

SOMMARIO

| | |
|--|-----------|
| ABSTRACT | 6 |
| INTRODUZIONE..... | 7 |
| Capitolo 1. DIFFUSIONE DELLE CERTIFICAZIONI ISO 9000 IN EUROPA..... | 8 |
| 1.1 ISO 9000: una famiglia di norme | 8 |
| 1.2 Cenni Storici..... | 9 |
| 1.3 Impatto delle certificazioni ISO 9000 sulle prestazioni aziendali | 11 |
| 1.4 Diffusione delle certificazioni ISO 9000 in Europa | 12 |
| 1.4.1 Diversi schemi di diffusione in Europa | 15 |
| Capitolo 2. ANALISI E PREVISIONE DELLA DIFFUSIONE DELLE CERTIFICAZIONI ISO 9000 | 22 |
| 2.1 Rassegna della letteratura..... | 22 |
| 2.2 Logistic model | 24 |
| 2.3 Cluster-analysis..... | 28 |
| 2.3.1 Clustering e modello di previsione proposto da Sampaio <i>et al</i> (2011)..... | 29 |
| 2.3.2 Clustering dei Paesi europei proposto da Franceschini <i>et al.</i> (2010) | 31 |
| Capitolo 3. SCHEMI DI DIFFUSIONE DELLE CERTIFICAZIONI ISO 9000 IN EUROPA..... | 34 |
| 3.1 Clustering gerarchico | 34 |
| 3.2 Clustering in MATLAB..... | 35 |
| 3.2.1 Normalizzazione dei valori e misure di distanza | 35 |
| 3.2.2 Collegamenti tra gli oggetti | 37 |
| 3.2.3 Dendrogramma | 38 |
| 3.2.4 Verifiche dei risultati | 39 |
| 3.2.5 Creazione dei cluster..... | 44 |
| 3.3 Selezione del campione di Paesi europei e normalizzazione dei valori | 46 |
| 3.4 Misure di distanza tra i Paesi europei..... | 48 |
| 3.4.1 Distanza Euclidea..... | 48 |
| 3.4.2 Dynamic Time Warping – DTW | 49 |
| 3.4.3 Adaptive Dissimilarity Index..... | 53 |
| 3.4.4 Calcolo delle distanze DTW in Matlab | 55 |
| 3.5 Dendrogramma dei Paesi europei | 59 |

| | |
|---|-----------|
| 3.6 Cluster degli schemi di diffusione per i Paesi europei..... | 63 |
| CONCLUSIONI..... | 76 |
| Appendice..... | 78 |
| BIBLIOGRAFIA..... | 81 |
| SITOGRAFIA..... | 84 |

ABSTRACT

Le certificazioni ISO 9000 hanno assunto negli anni un'importanza sempre maggiore per imprese e organizzazioni, infatti sono state ampiamente adottate più di qualsiasi altra pratica di gestione della qualità, sia in termini di numeri, sia in termini di diffusione geografica. Questo documento presenta un'analisi dettagliata della diffusione delle certificazioni ISO 9000 in Europa, l'obiettivo è individuare i diversi modelli di diffusione sulla base dell'analisi delle serie storiche descriventi l'evoluzione negli anni del numero di certificazioni nei vari Paesi europei. L'analisi è stata effettuata applicando la tecnica del *Dynamic Time Warping* e implementando gli algoritmi di *clustering* gerarchico agglomerativo nell'ambiente di calcolo *MATLAB R2017b*. I risultati ottenuti mostrano sei diversi schemi di diffusione, quattro dei quali sono riconducibili ad un più generico modello costituito da quattro fasi: crescita, maturità, declino e stazionarietà.

Il presente studio si limita ai Paesi europei, studi futuri dovrebbero estendere la seguente analisi ad un'area geografica più ampia al fine di verificare i risultati ottenuti e le previsioni proposte per il caso europeo. Lo studio dei modelli di diffusione delle certificazioni ISO 9000 è importante sia per gli enti certificatori che le rilasciano sia per le imprese e le organizzazioni che le richiedono, le attuano e ne sostengono i costi.

[English version]

ISO 9000 certifications became over time a prominence for companies and organizations, they have actually been adopted more than other quality practices, in terms of numbers and in terms of geographical diffusion all over the world. This document shows a detailed analysis of how ISO 9000 standards are spread in Europe, the purpose is to find different diffusion models based on the analysis of the time series that describe the evolution during years of ISO 9000 in European countries.

The analysis is based on the *Dynamic Time Warping* technique and on the agglomerative hierarchical clustering algorithms implemented with *MATLAB R2017b*. The results show six different diffusion patterns, four of them relate to a generic model consisting in four phases: growth, maturity, decline and stationarity.

This study deals with European countries, future research will expand this analysis to a wider geographical area to verify the results and assumptions found in Europe. The study of ISO 9000 certifications diffusion models is essential for both certification bodies that release them and companies that require them, apply them and bear their costs.

INTRODUZIONE

Fin dal 1987, anno dell'introduzione del sistema di certificazioni ISO 9000 per la gestione della qualità, la richiesta di certificazioni è cresciuta di anno in anno coinvolgendo un numero sempre maggiore di imprese e organizzazioni.

Il seguente elaborato individua, descrive e analizza i diversi schemi di diffusione temporale delle certificazioni ISO 9000 in Europa. L'obiettivo è individuare i cluster di Paesi europei sulla base di trend comuni, osservabili dall'analisi delle serie storiche descrittive il fenomeno in questione. In particolare, il punto di partenza dell'intera analisi è l'ISO Survey 2016 (ISO, 2017), indagine annuale dell'International Organization for Standardization (ISO) riguardo al numero di certificazioni valide emesse dagli enti certificatori autorizzati.

Il lavoro è complessivamente strutturato in tre parti: le prime due sono di carattere prettamente teorico e introducono il tema della ricerca, nella terza parte invece viene descritto in dettaglio il lavoro di *cluster-analysis* effettuato e vengono analizzati i risultati ottenuti.

In particolare, nel primo capitolo viene descritta la famiglia di norme ISO 9000 con particolare attenzione al loro contenuto, alla loro importanza e al loro scopo. Viene inoltre tracciata l'origine storica della diffusione delle certificazioni e, sulla base dei dati forniti dall'ultimo ISO Survey, viene analizzato il loro stato di diffusione nel mondo e in Europa.

Nella seconda parte del documento, oltre a una rapida analisi della letteratura, ci si focalizza sulla descrizione dei vari modelli proposti da diversi ricercatori, i quali hanno cercato di descrivere e prevedere la diffusione delle ISO 9000 in Europa e nel mondo. Di questi modelli vengono messi in evidenza i limiti e la loro applicabilità ai dati aggiornati dell'ultimo ISO Survey.

Nell'ultimo capitolo infine, vengono spiegati nel dettaglio gli algoritmi e le tecniche di *clustering* adottati sulla base dei dati a disposizione, inoltre vengono identificati e analizzati i diversi schemi di diffusione delle certificazioni ISO 9000 che ad oggi possono essere individuati in Europa.

CAPITOLO 1.

DIFFUSIONE DELLE CERTIFICAZIONI ISO 9000 IN EUROPA

1.1 ISO 9000: una famiglia di norme

La sigla ISO 9000 si riferisce a una serie di norme relative a vari aspetti della gestione e del controllo della qualità. Questi standard forniscono linee guida e strumenti alle imprese e alle organizzazioni che vogliono garantire che i loro prodotti e servizi incontrino costantemente le esigenze dei clienti e che la qualità venga continuamente migliorata (ISO, 2017). All'interno di questa famiglia è possibile individuare diverse norme, di cui solo una effettivamente certificabile, ossia per la quale un'impresa o un'organizzazione può ottenere da parte di un ente certificatore autorizzato un attestato di conformità ai requisiti in essa contenuti. In particolare, la prima norma, che dà nome alla serie, è la ISO 9000:2015 (ISO, 2015), propriamente non certificabile, che descrive i concetti e i principi fondamentali della gestione della qualità e fornisce il vocabolario utile per le normative della stessa famiglia. La norma ISO 9001 (ISO, 2015), l'unica propriamente certificabile, definisce invece i requisiti per un sistema di gestione della qualità per un'impresa o un'organizzazione. Oltre a varie norme specifiche non certificabili e alle norme ISO 9002 e 9003, ritirate e incorporate nella ISO 9001:2000, si può distinguere all'interno di questa famiglia la norma non certificabile ISO 9004 (ISO, 2009). Quest'ultima fornisce le linee guida per supportare il raggiungimento del successo sostenibile all'interno di un'organizzazione attraverso la gestione della qualità. Essendo la norma ISO 9001 l'unica effettivamente certificabile da imprese e organizzazioni, il seguente elaborato studia la diffusione di questa famiglia di standard basandosi sui dati disponibili riguardo al numero di certificazioni ISO 9001 rilasciate, benché da questo punto in poi ci si riferisce in generale al numero di ISO 9000.

Una fotografia aggiornata riguardo la diffusione delle ISO 9000 rilasciate in tutto il mondo dagli enti certificatori di competenza è fornita dall'ultimo ISO *Survey* (ISO, 2017). Al 31 dicembre 2016 il numero di certificazioni rilasciate ammontava quasi a 1'106'000 in 201 Paesi (di cui il 40,8% di certificazioni in Europa), con un tasso di crescita prossimo al 7% rispetto ai dati del 2015. Questi numeri riflettono l'importanza che le certificazioni ISO 9000 hanno assunto negli anni per imprese e organizzazioni a livello globale.

In generale si possono distinguere diverse ragioni che spingono un'impresa a certificarsi: in alcuni casi le certificazioni rappresentano un vero e proprio vincolo per contratti nazionali o internazionali oppure possono essere i clienti stessi a richiederle, in altri casi le certificazioni possono essere utilizzate come strumento da parte dell'impresa per dimostrare un vantaggio sui competitors (Anderson *et al.*, 1999). A partire dalla fine degli anni novanta molti autori si sono concentrati sull'impatto che le certificazioni ISO 9000 hanno avuto sulla performance dell'impresa al fine di ottenere un vantaggio competitivo rispetto ai concorrenti. Empiricamente si è osservato infatti che la qualità sembra migliorare grazie a queste certificazioni e che la motivazione che spinge un'impresa a certificarsi può influenzare il grado stesso di miglioramento della qualità (Withers *et al.*, 2000).

Altrettanto interessante è però descrivere e analizzare in che modo e con che numero le certificazioni ISO 9000 si siano diffuse nel tempo e come tale diffusione possa essere simile o diversa nei vari Paesi. Infatti, le ISO 9000 sono state ampiamente adottate più di qualsiasi altra pratica di gestione della qualità, sia per quanto riguarda il loro numero sia per la loro diffusione a livello geografico. L'obiettivo di questo capitolo è proprio quello di descrivere l'evoluzione che le certificazioni ISO 9000 hanno avuto in Europa a partire dalla loro introduzione nel 1987.

1.2 Cenni Storici

Fin dai primi anni Ottanta iniziò a diffondersi l'idea di gestione dell'impresa in accordo con i principi fondamentali della qualità. Durante questi anni si diffondeva sempre più tra le imprese l'adozione di varie forme di *Total Quality Management* (TQM), che causavano però lo spostamento del controllo di accettazione dall'ingresso dei prodotti al controllo diretto dei processi di produzione dei fornitori. I fornitori subivano dunque continui controlli da parte dei loro clienti e questo comportava spesso la necessità di dover produrre diversi tipi di documentazione e diversi sistemi di misura a seconda del cliente. Le certificazioni ISO 9000 vennero quindi introdotte per ridurre questo peso sui fornitori (Castka e Corbett, 2013) definendo un unico standard verificabile da un terzo organo di controllo. Le ISO 9000 hanno dunque svolto il ruolo di catalizzatore delle tendenze esistenti, al fine di indurre imprese ed organizzazioni verso una logica di gestione strategica della qualità (Franceschini *et al.*, 2006).

Possiamo datare la prima vera e propria definizione di linee guida per l'applicazione dei principi fondamentali della qualità a fine anni Cinquanta negli Stati Uniti, inizialmente nel 1959 nel settore militare con la pubblicazione della normativa MIL-Q-9858 da parte del Dipartimento di difesa degli Stati Uniti, in seguito in quello nucleare, farmaceutico e dell'automotive. L'unico scopo era garantire che i prodotti rispettassero le specifiche e i requisiti tecnici di progetto. Queste direttive furono poi adottate dal British Standard Institution (BSI) che ampliò il loro campo di applicazione all'intero sistema aziendale attraverso la pubblicazione nel 1979 della norma BS 5750, che portò in seguito alla pubblicazione della serie di norme ISO 9000 (The British Standards Institution, 2017).

Negli anni Ottanta infatti, l'International Organization for Standardization (ISO), ente in principio interessato esclusivamente alle attività di misurazione nei vari settori industriali, si concentrò sulla definizione di una configurazione unica e coerente del sistema di gestione della qualità. L'ISO pubblicò nel 1987 la prima edizione della famiglia di norme ISO 9000, destinate a diventare da lì in poi il punto di riferimento per i sistemi di qualità di imprese ed organizzazioni in tutto il mondo (Franceschini *et al.*, 2006). L'obiettivo era facilitare il commercio internazionale e aumentare la competitività delle imprese europee e nordamericane attraverso un allineamento di termini, sistemi e metodologie, in un mercato sempre più selettivo caratterizzato dalla minaccia delle imprese asiatiche. Tale obiettivo poteva essere ottenuto solamente attraverso il raggiungimento di un vantaggio competitivo da parte dell'impresa in termini di soddisfazione del cliente e affidabilità del prodotto (Withers *et al.*, 2000).

La famiglia di norme ISO 9000 nasce quindi per specificare i requisiti di un'impresa o un'organizzazione al fine di fornire un'evidenza formale delle sue competenze nell'organizzare risorse e processi, con l'obiettivo di garantire il soddisfacimento degli *stakeholders* dell'impresa (Franceschini, 2002). Gli standard ISO 9000 non si focalizzano quindi sulla qualità intrinseca di prodotti o di servizi (output dell'impresa), ma sui relativi processi, ampliando la loro azione all'intero network di interazioni sulla cui base l'impresa opera. Questa estensione del campo di applicazione origina dalla consapevolezza che la qualità è una variabile strategica di cui tenere in considerazione nell'intero network che caratterizza la *value-chain* (Romano e Vinelli, 2001). Possiamo perciò definire la famiglia di norme ISO 9000 come una categoria speciale di

standard “orizzontali” di generale applicazione (Franceschini *et al.*, 2006), finalizzate a garantire la qualità di un prodotto attraverso una gestione consapevole e adeguata di risorse e processi.

1.3 Impatto delle certificazioni ISO 9000 sulle prestazioni aziendali

Come già accennato, ci sono varie ragioni che spingono un’impresa a certificarsi secondo le norme ISO 9000. Il fatto stesso che queste siano standard non specifici ma di generale applicazione le rende un’importante strumento di marketing per sviluppare, rafforzare e facilitare contratti e accordi commerciali sia con partner nazionali che internazionali, migliorando la performance dell’impresa (Franceschini *et al.*, 2010).

In generale le certificazioni ISO 9000 si fondano su due attributi chiave: la forte dipendenza dalla documentazione relativa a processi e performance e l’obiettivo di assicurare consistenza degli output. L’idea sottostante è che un’impresa che abbia definito le proprie *best practices* per i propri processi chiave, che le abbia documentate e che metta effettivamente in pratica queste procedure avrà buoni output (Castka e Corbett, 2013). Questo non significa che l’impresa non commetterà errori (difetti o output difettosi), ma piuttosto che tali errori saranno meno frequenti e soprattutto che l’impresa possiederà i mezzi e le procedure per affrontare tali problematiche e per prevenire la loro ricorrenza. Le certificazioni ISO 9000 quindi non dicono nulla riguardo l’output effettivo dell’impresa, ma ovviamente l’aspettativa è che quando un *auditor* (ente di controllo esterno) conferma che le procedure documentate dell’impresa sono adeguate e messe in pratica, l’output dovrebbe soddisfare le aspettative.

Molti ricercatori si sono concentrati sulla ricerca di evidenze empiriche della relazione che lega l’adozione delle certificazioni ISO 9000 alle prestazioni aziendali. Uno dei risultati più interessanti è che certificarsi ha un impatto positivo sulle prestazioni aziendali quando la cultura della qualità è ben sviluppata all’interno dell’impresa e quando la motivazione manageriale che spinge a certificarsi consiste nel migliorare le prestazioni e non nel conformarsi ad uno standard (Franceschini *et al.*, 2006). A tale proposito, Clougherty e Grajek (2008) hanno identificato tre principali caratteristiche delle certificazioni ISO 9000 che favoriscono le relazioni *business-to-business* aumentando le prestazioni delle imprese:

- *Quality commitment signal*. Le certificazioni rappresentano un segnale dell'impegno da parte dell'impresa nell'implementare e seguire un sistema documentato di gestione della qualità, riducendo in tal modo le barriere (linguistiche, culturali, governative o manageriali) tra le imprese legate da una relazione commerciale, specialmente se a livello internazionale;
- *Common language*. Le certificazioni definiscono un linguaggio d'uso comune che facilita le relazioni tra i diversi *business* abbattendo le barriere linguistiche e di comunicazione ed eliminandone i relativi costi;
- *Conflict settling*. Le certificazioni ISO 9000 minimizzano i conflitti, infatti possono essere adottate come strumento per mostrare la facilità con cui è possibile stringere relazioni commerciali con l'impresa stessa.

Questi tre fattori riducono quindi i costi di transazione e le asimmetrie informative, migliorando le prestazioni dell'impresa e facilitando le relazioni *business-to-business* nazionali e internazionali.

1.4 Diffusione delle certificazioni ISO 9000 in Europa

In questi ultimi anni la crescita di certificazioni ISO 9000 in Europa sembra aver raggiunto un livello di saturazione. Per la seguente analisi si considera l'Europa come definita dall'ISO *Survey* (vedi Tabella 1), i dati si riferiscono inoltre al numero di certificazioni ISO 9000 rilasciate e non al numero di siti certificati.

Tra il 1993 e il 2010 si osserva in Europa una continua crescita anno per anno del numero di certificazioni, trend che rispecchia più in generale l'andamento a livello mondiale (vedi Figura 1 e 2). L'unica eccezione è rappresentata dalla riduzione registrata nel 2003 (pari al 17% in Europa e all'11% a livello mondiale) per effetto del passaggio dalla norma ISO 9000:1994 alla ISO 9000:2000 (Franceschini *et al.*, 2010).

A partire dal 2011, a seguito di un calo del 13% in Europa (6% complessivamente nel mondo), il numero di certificazioni rilasciate è variato negli anni oscillando attorno ad un livello di saturazione di circa 450,000 certificazioni. Lo stesso comportamento si osserva anche nelle altre

macro-aree geografiche con una crescita annuale complessiva stabile intorno all'1% (vedi Figura 1). Da notare è l'aumento registrato nel 2016 che interessa però relativamente poco l'Europa, in cui infatti l'incremento effettivo è stato di meno del 3% rispetto al 2015.

Tabella 1 Numero di certificazioni ISO 9000 per macro-area geografica [Fonte: ISO (2017)]

| | Europa | Africa | Centro- e Sud- America | Nord- America | Asia orientale e Pacifico | Asia centrale e meridionale | Medio oriente | TOTALE |
|------|--------|--------|------------------------------|------------------|---------------------------------|-----------------------------------|------------------|---------|
| 1993 | 37779 | 1009 | 140 | 2613 | 4767 | 74 | 189 | 46571 |
| 1994 | 55400 | 1177 | 475 | 4915 | 7719 | 330 | 348 | 70364 |
| 1995 | 92611 | 1563 | 1220 | 10374 | 19766 | 1038 | 776 | 127348 |
| 1996 | 109961 | 2255 | 1713 | 16980 | 27885 | 1712 | 2194 | 162700 |
| 1997 | 143674 | 2555 | 2989 | 25144 | 42824 | 2963 | 3149 | 223298 |
| 1998 | 166255 | 3342 | 5221 | 33550 | 54671 | 3556 | 5251 | 271846 |
| 1999 | 190247 | 4928 | 8972 | 45166 | 81950 | 5508 | 6870 | 343641 |
| 2000 | 219173 | 4769 | 10805 | 48296 | 109217 | 6411 | 9003 | 407674 |
| 2001 | 269648 | 3903 | 14409 | 50894 | 155597 | 6348 | 9550 | 510349 |
| 2002 | 292878 | 4529 | 13679 | 53806 | 177767 | 9383 | 9724 | 561766 |
| 2003 | 242455 | 3769 | 9303 | 40185 | 185846 | 9162 | 7199 | 497919 |
| 2004 | 320748 | 4865 | 17016 | 49962 | 240938 | 13856 | 12747 | 660132 |
| 2005 | 377172 | 6763 | 22498 | 59663 | 266100 | 27966 | 13681 | 773843 |
| 2006 | 414208 | 7441 | 29382 | 61436 | 320320 | 44923 | 19195 | 896905 |
| 2007 | 431479 | 7446 | 39354 | 47600 | 354056 | 50379 | 21172 | 951486 |
| 2008 | 455303 | 8534 | 37458 | 47896 | 366491 | 44171 | 20469 | 980322 |
| 2009 | 500286 | 8435 | 35549 | 41947 | 408498 | 44432 | 24604 | 1063751 |
| 2010 | 530039 | 7667 | 49260 | 36632 | 396492 | 37596 | 18839 | 1076525 |
| 2011 | 459367 | 8164 | 51685 | 37530 | 402453 | 33577 | 17069 | 1009845 |
| 2012 | 469739 | 9674 | 51459 | 38586 | 396398 | 32373 | 19050 | 1017279 |
| 2013 | 458814 | 9816 | 52466 | 48579 | 387543 | 44847 | 20812 | 1022877 |
| 2014 | 453628 | 10143 | 50165 | 41459 | 414801 | 44790 | 21335 | 1036321 |
| 2015 | 439477 | 12154 | 49509 | 46938 | 422519 | 40822 | 22761 | 1034180 |
| 2016 | 451415 | 13378 | 52094 | 44252 | 480445 | 41370 | 22983 | 1105937 |

L'ipotesi del raggiungimento di una saturazione in Europa è confermata a livello mondiale dalla quota europea di certificazioni ISO 9000 (vedi Figura 3). Fino al 2007, nonostante una crescita continua in termini di numeri assoluti, la quota europea si riduceva per il solo effetto della rapida crescita del numero di certificazioni nei Paesi asiatici (Franceschini *et al.*, 2010), rendendo comunque l'Europa la macro-area con il maggior numero di certificazioni rilasciate. Negli ultimi anni invece, sembrerebbe che il fenomeno di crescita si sia arrestato e che il primato sia conteso per l'appunto con i Paesi asiatici. Infatti, a differenza dell'Europa, in Asia sembra che il mercato delle certificazioni ISO 9000 non conosca crisi, rendendo questa di fatto

nel 2016 la macro-area con il maggior numero di certificazioni rilasciate (con una quota superiore al 43%).

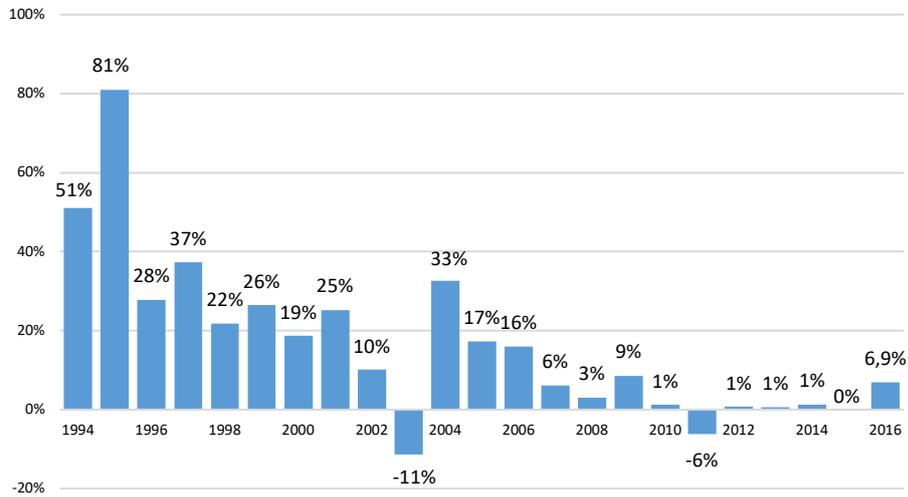


Figura 1
Tasso di crescita annuale mondiale delle certificazioni ISO 9000

Fonte: ISO (2017)

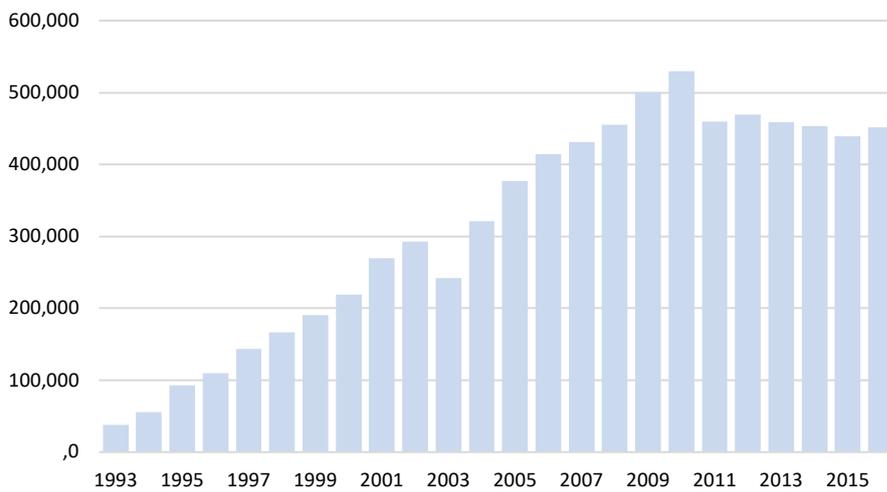


Figura 2
Diffusione delle certificazioni ISO 9000 in Europa

Fonte: ISO (2017)

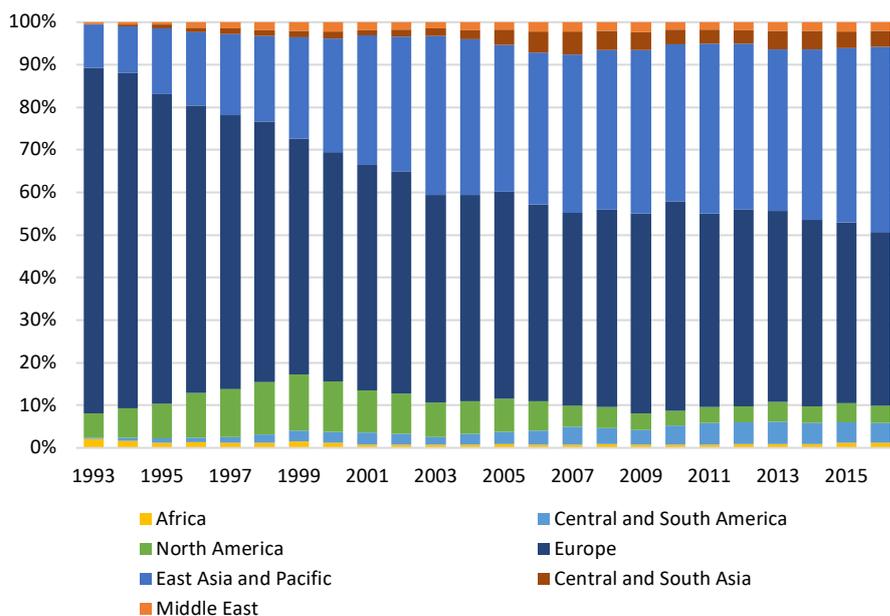


Figura 3
Quota regionale del numero di certificazioni ISO 9000

Fonte: ISO (2017)

Interessante è il trend in Centro- e Sud America in cui il numero di certificazioni ISO 9000 è cresciuto fino al 2007, ma negli anni successivi la loro diffusione sembra aver raggiunto un livello di saturazione. Come già sottolineato da Franceschini *et al.* (2010), da notare è l'andamento registrato in Nord America in cui l'adozione delle certificazioni ISO 9000 è notevolmente inferiore rispetto ad altri Paesi sviluppati. Questo comportamento è dovuto al naturale dinamismo del mercato nordamericano, che non forza le imprese a ricorrere alle certificazioni come elemento distintivo di concorrenza.

L'Europa e l'Asia rimangono comunque le macro-aree in cui le certificazioni ISO 9000 si sono maggiormente diffuse e che hanno dunque riconosciuto la loro importanza. Ma mentre in Asia il fenomeno di diffusione non sembra fermarsi, in Europa, benché primo continente a dimostrare un forte interesse nei confronti delle certificazioni, la crescita sembra essersi arrestata. In questo contesto risulta quindi interessante indagare come all'interno dell'Europa i vari Paesi abbiano reagito in maniera diversa alla diffusione delle certificazioni ISO 9000.

1.4.1 Diversi schemi di diffusione in Europa

Se si mettono a confronto i diversi Paesi, il fenomeno di diffusione delle certificazioni ISO 9000 non risulta essere "sincrono" (Franceschini *et al.*, 2006). Infatti, in alcuni Paesi le certificazioni

hanno trovato larga diffusione fin dalla loro introduzione, nel Regno Unito ad esempio già dai primi anni novanta il numero di certificazioni rilasciate superava la soglia delle 30'000 (vedi Tabella 2). In altri Paesi invece, come ad esempio in Italia, il processo di diffusione è stato posteriore. Il punto di partenza e la dinamica di diffusione dipendono in particolare dalla politica locale e dalle condizioni economiche in ciascun Paese (Franceschini *et al.*, 2004).

Tabella 2 Numero di certificazioni ISO 9000 per i sei Paesi europei con maggior PIL [Fonte: ISO (2017)]

| | France | Germany | Italy | Russian Federation | Spain | UK |
|------|--------|---------|--------|-----------------------|-------|-------|
| 1993 | 1586 | 1534 | 864 | 5 | 320 | 28096 |
| 1994 | 3359 | 3470 | 2008 | 8 | 586 | 36825 |
| 1995 | 5536 | 10236 | 4814 | 22 | 1492 | 52595 |
| 1996 | 8079 | 12979 | 7321 | 56 | 2496 | 53099 |
| 1997 | 11920 | 20656 | 12134 | 95 | 4268 | 56696 |
| 1998 | 14194 | 24055 | 18095 | 132 | 6412 | 58963 |
| 1999 | 16028 | 30150 | 21069 | 541 | 8699 | 63700 |
| 2000 | 17170 | 32500 | 30367 | 1134 | 12576 | 63725 |
| 2001 | 20919 | 41629 | 48109 | 1517 | 17749 | 66760 |
| 2002 | 19870 | 35802 | 61212 | 1710 | 28690 | 60960 |
| 2003 | 15073 | 23598 | 64120 | 962 | 31836 | 45465 |
| 2004 | 21769 | 26654 | 84485 | 3816 | 40972 | 50884 |
| 2005 | 21700 | 39816 | 98028 | 4883 | 47445 | 45612 |
| 2006 | 21349 | 46458 | 105799 | 6398 | 57552 | 40909 |
| 2007 | 22981 | 45195 | 115359 | 11527 | 65112 | 35517 |
| 2008 | 23837 | 48324 | 118309 | 16051 | 68730 | 41150 |
| 2009 | 23065 | 47156 | 130066 | 53152 | 59576 | 41193 |
| 2010 | 29713 | 50583 | 143305 | 62265 | 59854 | 43293 |
| 2011 | 29215 | 49540 | 142853 | 13308 | 53057 | 41943 |
| 2012 | 29198 | 51701 | 136547 | 12488 | 59418 | 42304 |
| 2013 | 29598 | 56303 | 135939 | 11764 | 42644 | 42843 |
| 2014 | 29112 | 55344 | 139416 | 11213 | 35995 | 39982 |
| 2015 | 27844 | 52995 | 132870 | 9084 | 32730 | 40161 |
| 2016 | 23403 | 66233 | 150143 | 5083 | 34438 | 37901 |

Se si prendono in considerazione i trend di diffusione in Figura 4, osserviamo che se nel Regno Unito e in Belgio il numero di certificazioni è cresciuto costantemente fino agli anni Duemila e sembra aver raggiunto nell'ultimo decennio un livello di saturazione, in Romania invece il numero di certificazioni è cresciuto costantemente fino al 2015 e ha subito una decrescita solo nell'ultimo anno. Si possono interpretare quindi i vari trend più che come diversi schemi di diffusione, come istantanee dello stesso tipo di fenomeno, ma scattate in diversi momenti della sua evoluzione (Franceschini *et al.*, 2010).

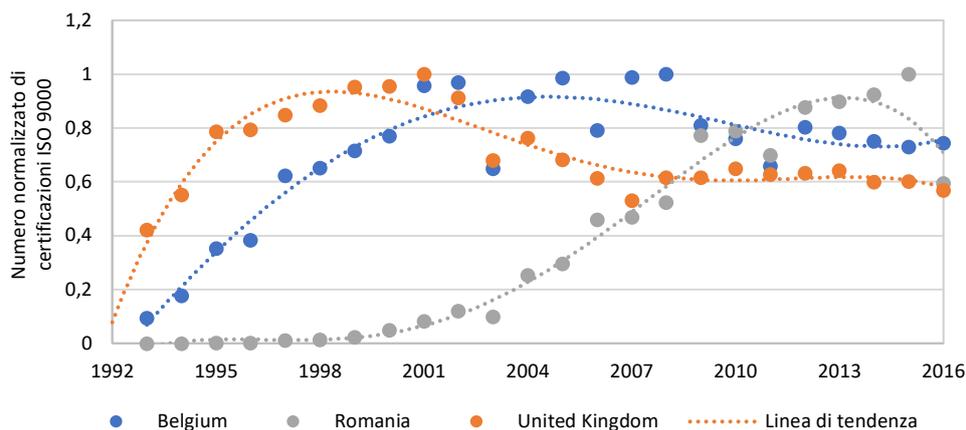


Figura 4
Diversi trend
di diffusione
in Europa¹

Nel 2010 Franceschini *et al.* (2010) hanno cercato di descrivere il fenomeno di evoluzione delle ISO 9000 identificando 3 diverse fasi (vedi Figura 5):

1. *Growth phase*. Fase di crescita in cui il numero di certificazioni cresce costantemente in quanto le imprese sono attratte dai benefici derivanti dalla loro implementazione;
2. *Maturity phase*. È propriamente la fase di saturazione in cui il numero di certificazioni si mantiene costante negli anni. Durante questa fase, diversi fattori come la riduzione del gap competitivo tra imprese certificate e non-certificate, e il numero limitato di imprese potenzialmente certificabili, rendono le ISO 9000 sempre meno appetibili. Questo effetto di saturazione dipende fortemente dalla struttura economica e produttiva di ciascun Paese (Franceschini *et al.*, 2006);
3. *Decline phase*. Fase di declino in cui il numero di certificazioni rilasciate decresce progressivamente anno per anno.

Questo modello può essere applicato ai dati aggiornati dell'ultimo ISO Survey (2017) ma con una certa attenzione. Se considera il caso dell'Irlanda e del Regno Unito (vedi Figura 6), è possibile identificare chiaramente le tre fasi di crescita, maturità e declino, si può osservare però che negli ultimi anni sembra esserci una nuova fase stazionaria a valle della fase di declino dei primi anni Duemila.

¹ Per ragioni grafiche e di confronto dei dati si fa riferimento al numero normalizzato di certificazioni ISO 9000 come variabile di analisi. Tale numero è definito come il numero di certificazioni rilasciate diviso il numero massimo di certificazioni rilasciate nel periodo 1993-2016 in un paese.

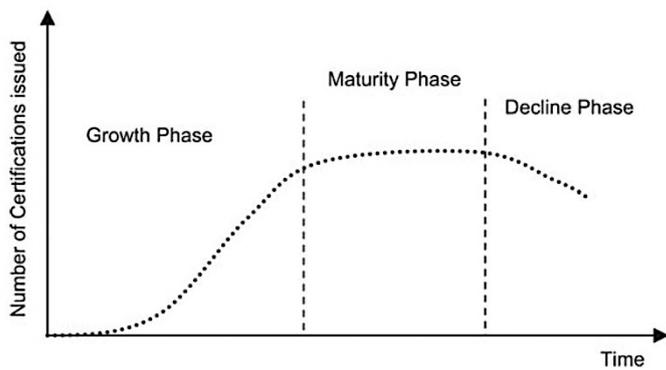


Figura 5
 Pattern di diffusione delle certificazioni

Fonte: Franceschini et al. (2010)

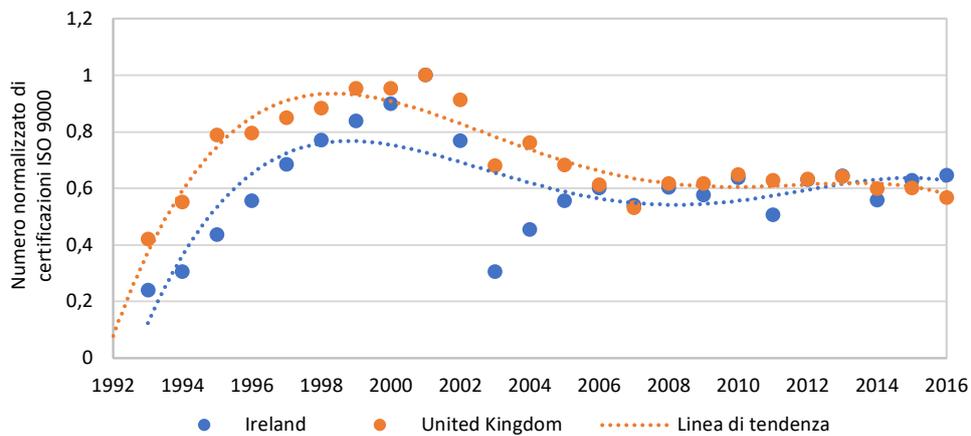


Figura 6
 Trend di diffusione in Irlanda e Regno Unito

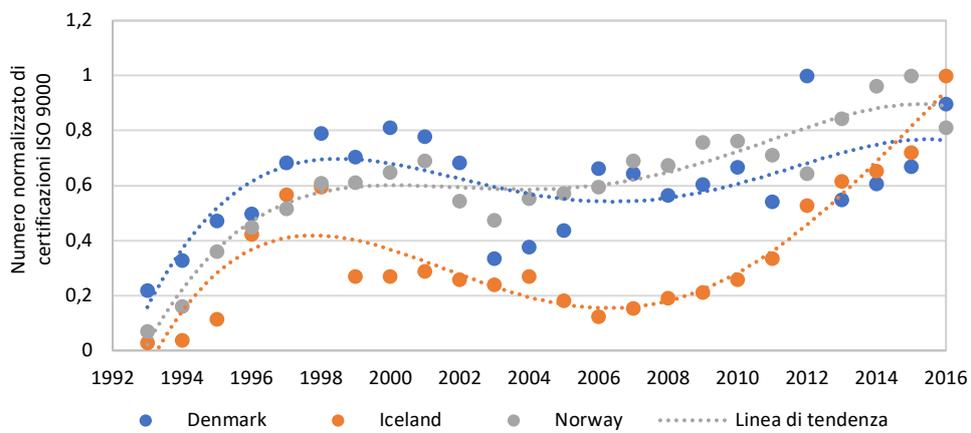


Figura 7
 Trend di diffusione in Danimarca, Islanda e Norvegia

Se si analizza invece il trend in Paesi come la Danimarca, l'Islanda e la Norvegia (vedi Figura 7), si osserva, invece, che a valle della fase di maturità il fenomeno di diffusione sembra essersi rimesso in moto con una crescita progressiva del numero di certificazioni rilasciate negli ultimi anni. Questo andamento è ben evidente in Islanda: qui si può notare infatti che a seguito di una prima fase di declino nei primi anni Duemila, a partire dal 2007 il numero di ISO 9000 rilasciate è notevolmente cresciuto. Per i dati effettivi relativi al numero di certificazioni ISO 9000 per i principali Paesi europei si rimanda alla Tabella 3 e 4.

Con i dati aggiornati dell'ultimo ISO *Survey* si può dunque affermare che il modello ipotizzato da Franceschini *et al.* non riesce a spiegare completamente il fenomeno di diffusione delle ISO 9000 nei vari Paesi europei. In particolare, non spiega correttamente ciò che accade in seguito al raggiungimento della fase di maturità o declino. Questo modello rappresenta comunque il punto di partenza per il seguente lavoro di ricerca che ha l'obiettivo di concentrarsi su ciò che accade in ciascun Paese in termini di numero di certificazioni ISO 9000 rilasciate, attraverso una comparazione e una *cluster-analysis* dei diversi schemi di diffusione.

| | 1993 | 1994 | 1995 | 1996 | 1997 | 1998 | 1999 | 2000 | 2001 | 2002 | 2003 | 2004 | 2005 |
|----------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| Austria | 200 | 434 | 1133 | 1824 | 2627 | 3245 | 3421 | 3826 | 4000 | 4094 | 2809 | 3259 | 3368 |
| Belarus | | | | | 6 | 14 | 26 | 58 | 78 | 115 | 102 | 447 | 658 |
| Belgium | 464 | 870 | 1716 | 1871 | 3042 | 3176 | 3495 | 3760 | 4670 | 4725 | 3167 | 4471 | 4810 |
| Bulgaria | | | 3 | 14 | 42 | 96 | 199 | 259 | 469 | 625 | 842 | 1685 | 2220 |
| Croatia | 0 | 2 | 22 | 38 | 96 | 121 | 336 | 302 | 415 | 590 | 580 | 966 | 1273 |
| Czech Republic | 18 | 47 | 180 | 366 | 746 | 1443 | 1500 | 3855 | 5627 | 8489 | 2565 | 10781 | 12743 |
| Denmark | 608 | 916 | 1314 | 1387 | 1902 | 2200 | 1962 | 2258 | 2163 | 1900 | 935 | 1050 | 1219 |
| Finland | 324 | 496 | 772 | 951 | 1445 | 1450 | 2105 | 1651 | 1870 | 1872 | 1861 | 1784 | 1914 |
| France | 1586 | 3359 | 5536 | 8079 | 11920 | 14194 | 16028 | 17170 | 20919 | 19870 | 15073 | 21769 | 21700 |
| Germany | 1534 | 3470 | 10236 | 12979 | 20656 | 24055 | 30150 | 32500 | 41629 | 35802 | 23598 | 26654 | 39816 |
| Greece | 46 | 90 | 248 | 348 | 682 | 764 | 1050 | 2173 | 2325 | 3180 | 1615 | 2572 | 3255 |
| Hungary | 23 | 58 | 309 | 423 | 1341 | 1660 | 3282 | 4672 | 6362 | 9254 | 7750 | 10207 | 15464 |
| Iceland | 3 | 4 | 12 | 44 | 59 | 62 | 28 | 28 | 30 | 27 | 25 | 28 | 19 |
| Ireland | 893 | 1132 | 1617 | 2056 | 2534 | 2854 | 3100 | 3330 | 3700 | 2845 | 1132 | 1683 | 2055 |
| Italy | 864 | 2008 | 4814 | 7321 | 12134 | 18095 | 21069 | 30367 | 48109 | 61212 | 64120 | 84485 | 98028 |
| Latvia | | | | 1 | 1 | 14 | 39 | 94 | 67 | 93 | 73 | 484 | 561 |
| Lithuania | | | 2 | 3 | 29 | 40 | 91 | 173 | 202 | 280 | 324 | 487 | 591 |
| Luxembourg | 10 | 21 | 48 | 46 | 89 | 106 | 113 | 136 | 108 | 148 | 110 | 108 | 147 |
| Netherlands | 1502 | 2718 | 5284 | 7986 | 10380 | 10570 | 10620 | 11036 | 12745 | 13198 | 9917 | 6402 | 9160 |
| Norway | 172 | 400 | 890 | 1109 | 1273 | 1503 | 1509 | 1600 | 1703 | 1344 | 1171 | 1368 | 1410 |
| Poland | 1 | 16 | 130 | 260 | 669 | 768 | 1012 | 2075 | 2622 | 3091 | 3216 | 5753 | 9718 |
| Portugal | 85 | 181 | 389 | 535 | 819 | 944 | 1131 | 1696 | 2474 | 3061 | 3417 | 4733 | 5820 |
| Romania | | 6 | 42 | 61 | 214 | 269 | 466 | 1032 | 1670 | 2463 | 2052 | 5183 | 6097 |
| Russia | 5 | 8 | 22 | 56 | 95 | 132 | 541 | 1134 | 1517 | 1710 | 962 | 3816 | 4883 |
| Slovakia | 5 | 11 | 59 | 135 | 404 | 575 | 560 | 522 | 827 | 1544 | 1148 | 2008 | 2050 |
| Slovenia | 16 | 43 | 99 | 152 | 467 | 502 | 521 | 843 | 1026 | 973 | 465 | 1811 | 2114 |
| Spain | 320 | 586 | 1492 | 2496 | 4268 | 6412 | 8699 | 12576 | 17749 | 28690 | 31836 | 40972 | 47445 |
| Sweden | 365 | 618 | 1095 | 1931 | 2789 | 3489 | 3786 | 4358 | 4652 | 4039 | 3107 | 4687 | 4744 |
| Switzerland | 569 | 945 | 2065 | 3701 | 4653 | 6426 | 7124 | 8660 | 8605 | 10299 | 8300 | 11549 | 12413 |
| Turkey | 65 | 106 | 434 | 606 | 1284 | 1607 | 1672 | 2278 | 2949 | 3941 | 3248 | 5009 | 10929 |
| Ukraine | 1 | 4 | 8 | 14 | 30 | 56 | 82 | 151 | 269 | 893 | 308 | 934 | 1375 |
| UK | 28096 | 36825 | 52595 | 53099 | 56696 | 58963 | 63700 | 63725 | 66760 | 60960 | 45465 | 50884 | 45612 |

Fonte: (ISO, 2017)

Tabella 3
 Numero di
 certificazioni
 ISO 9000 per i
 principali Paesi
 europei
 (1993-2005)

| | 2006 | 2007 | 2008 | 2009 | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 |
|----------------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| Austria | 3806 | 4203 | 4272 | 4277 | 5161 | 4138 | 4562 | 4637 | 4211 | 4470 | 3922 |
| Belarus | 882 | 1308 | 1749 | 2014 | 151 | 171 | 170 | 130 | 2905 | 3657 | 175 |
| Belgium | 3865 | 4822 | 4875 | 3950 | 3715 | 3207 | 3915 | 3812 | 3661 | 3562 | 3634 |
| Bulgaria | 3097 | 4663 | 5323 | 5322 | 6248 | 5001 | 6037 | 5378 | 5729 | 5441 | 5951 |
| Croatia | 1676 | 2073 | 2302 | 2567 | 2102 | 2117 | 2584 | 2636 | 2806 | 2529 | 2659 |
| Czech Republic | 12811 | 10458 | 10089 | 14031 | 16242 | 12697 | 10679 | 12679 | 13229 | 10648 | 10568 |
| Denmark | 1840 | 1794 | 1574 | 1683 | 1856 | 1505 | 2780 | 1527 | 1689 | 1865 | 2498 |
| Finland | 1986 | 1804 | 1975 | 2243 | 2147 | 2265 | 2403 | 2838 | 2648 | 2596 | 2592 |
| France | 21349 | 22981 | 23837 | 23065 | 29713 | 29215 | 29198 | 29598 | 29112 | 27844 | 23403 |
| Germany | 46458 | 45195 | 48324 | 47156 | 50583 | 49540 | 51701 | 56303 | 55344 | 52995 | 66233 |
| Greece | 4753 | 5132 | 6747 | 5034 | 4322 | 4168 | 4796 | 7293 | 5445 | 6187 | 7303 |
| Hungary | 15008 | 10473 | 10187 | 7122 | 8083 | 6825 | 7232 | 7186 | 6909 | 5789 | 6559 |
| Iceland | 13 | 16 | 20 | 22 | 27 | 35 | 55 | 64 | 68 | 75 | 104 |
| Ireland | 2225 | 1999 | 2237 | 2136 | 2359 | 1875 | 2331 | 2388 | 2067 | 2323 | 2393 |
| Italy | 105799 | 115359 | 118309 | 130066 | 143305 | 142853 | 136547 | 135939 | 139416 | 132870 | 150143 |
| Latvia | 625 | 342 | 500 | 708 | 809 | 787 | 791 | 923 | 1000 | 1115 | 866 |
| Lithuania | 697 | 809 | 815 | 1111 | 1207 | 1168 | 1165 | 1110 | 1214 | 1238 | 1150 |
| Luxembourg | 145 | 197 | 246 | 247 | 107 | 156 | 163 | 259 | 150 | 252 | 209 |
| Netherlands | 18922 | 18922 | 13597 | 12260 | 11213 | 11072 | 11417 | 11415 | 10429 | 10381 | 10326 |
| Norway | 1467 | 1703 | 1666 | 1871 | 1882 | 1756 | 1589 | 2080 | 2377 | 2467 | 2002 |
| Poland | 8115 | 9184 | 10965 | 12707 | 12195 | 10984 | 10105 | 10527 | 9574 | 10681 | 12152 |
| Portugal | 5851 | 5283 | 5128 | 5051 | 5588 | 4638 | 6650 | 7041 | 8006 | 7498 | 7160 |
| Romania | 9426 | 9633 | 10737 | 15865 | 16200 | 14345 | 18014 | 18450 | 18984 | 20524 | 12209 |
| Russia | 6398 | 11527 | 16051 | 53152 | 62265 | 13308 | 12488 | 11764 | 11213 | 9084 | 5083 |
| Slovakia | 2195 | 2840 | 3476 | 3475 | 3895 | 3787 | 4281 | 3891 | 4598 | 5683 | 5716 |
| Slovenia | 2182 | 1886 | 1945 | 1688 | 1701 | 1658 | 1595 | 1993 | 1672 | 1481 | 1848 |
| Spain | 57552 | 65112 | 68730 | 59576 | 59854 | 53057 | 59418 | 42644 | 35995 | 32730 | 34438 |
| Sweden | 4839 | 5233 | 5377 | 5346 | 5687 | 4901 | 4846 | 4613 | 4998 | 4316 | 4041 |
| Switzerland | 10984 | 11077 | 11724 | 11581 | 12110 | 10358 | 11542 | 12030 | 11205 | 12218 | 11212 |
| Turkey | 12350 | 12802 | 13217 | 13705 | 10680 | 9446 | 7608 | 7178 | 8969 | 8538 | 6889 |
| Ukraine | 1808 | 2150 | 2453 | 3252 | 2592 | 1207 | 1091 | 1275 | 1297 | 1052 | 1382 |
| UK | 40909 | 35517 | 41150 | 41193 | 43293 | 41943 | 42304 | 42843 | 39982 | 40161 | 37901 |

Fonte: (ISO, 2017)

Tabella 4
 Numero di
 certificazioni
 ISO 9000 per i
 principali Paesi
 europei
 (2006-2016)

CAPITOLO 2.

ANALISI E PREVISIONE DELLA DIFFUSIONE DELLE CERTIFICAZIONI ISO 9000

2.1 Rassegna della letteratura

Il fenomeno delle certificazioni ISO 9000 ha suscitato un forte interesse da parte dei ricercatori ed è stato ampiamente studiato e approfondito, analizzando in particolare sia i fattori che spingono le imprese e le organizzazioni a implementare le ISO 9000 sia i benefici che ne derivano. Risulta invece scarsa la letteratura focalizzata sulla diffusione di questi standard. In particolare, quest'ultima si basa sui modelli di diffusione delle innovazioni tecnologiche. Infatti, il fenomeno di adozione delle nuove tecnologie segue un andamento a curva-a-S (curva logistica): in una prima fase il numero di membri di un sistema sociale che adottano la nuova tecnologia è limitato, ma con il passare del tempo il numero di adozioni cresce fino a raggiungere un certo livello di saturazione mentre il tasso di crescita si riduce.

Uno dei primi studi di carattere quantitativo circa la diffusione delle certificazioni ISO 9000 è stato pubblicato da Saraiva e Duarte (2003), i quali hanno cercato di predire come le certificazioni ISO 9000 si sarebbero sviluppate nel mondo tra il 2002 e il 2006. A partire da un indice calcolato sulla base del numero di certificazioni ISO 9000 per abitante in ciascun Paese, questi autori hanno utilizzato un semplice modello di regressione lineare identificando in quegli anni i Paesi leader in termini di certificazioni ISO 9000 e hanno inoltre osservato come il livello di saturazione stesse per essere raggiunto.

In modo simile, Franceschini *et al.* (2004) hanno studiato la crescita delle certificazioni ISO 9000 in sei Paesi europei proponendo come strumento di analisi l'utilizzo del *logistic model* (modello della curva logistica), associando quindi la diffusione delle ISO 9000 a fenomeni di crescita biologici. Il modello suggerito propone una previsione della diffusione delle ISO 9000 e rileva come già al tempo si stesse delineando una sorta di *saturation effect* (effetto saturazione) (Franceschini *et al.*, 2006). Utilizzando infatti i dati dell'ISO Survey dal 1991 al 2001 e i dati riguardo al numero di imprese in ciascun Paese, gli autori avevano predetto i livelli di saturazione che sarebbero stati raggiunti nei successivi cinque anni nei Paesi europei analizzati.

Nel 2009 Sampaio *et al.* hanno concluso che esiste una relazione lineare tra il numero di ISO 9000 rilasciate per 1'000 abitanti e la percentuale di imprese certificate per ciascun Paese dell'Unione Europea. Gli autori hanno inoltre osservato come in alcuni Paesi i valori di saturazione predetti da Franceschini *et al.* fossero significativamente inferiori rispetto ai valori registrati con i nuovi dati resi disponibili dall'ISO.

Altri studi si sono concentrati invece sul legame tra le certificazioni ISO 9000 e le certificazioni ISO 14000². Marimon *et al.* (2006) hanno dimostrato che le due famiglie di standard seguono schemi di diffusione molto simili. Applicando infatti il modello proposto da Franceschini *et al.* (2004), sulla base dei dati dell'ISO Survey 2003, hanno osservato che entrambe le famiglie di standard avrebbero raggiunto nel 2007 un livello di saturazione prossimo al 95% (circa 800'000 certificazioni per le ISO 9000 e poco più di 70'000 per le ISO 14000). Questi autori hanno inoltre verificato che i settori pionieri delle certificazioni ISO 9000 hanno svolto lo stesso ruolo leader per quanto riguarda la diffusione delle certificazioni ISO 14000. Altri studi a riguardo sono stati condotti da ricercatori come Corbett e Kirsch (2001), i quali hanno constatato che la straordinaria crescita del numero di ISO 9000 è un fattore chiave per spiegare la crescita stessa delle certificazioni ISO 14000.

Un recente studio portato avanti da Castka e Corbett (2013) rappresenta un'importante monografia che ha l'obiettivo di fornire una panoramica completa e dettagliata circa la letteratura e la ricerca riguardo le certificazioni ISO 9000, ISO 14000 e altri standard, soffermandosi in particolare su tre questioni chiave: perché le organizzazioni adottano questi standard, quali effetti hanno e come vengono gestiti. Questo lavoro ha l'obiettivo di raccogliere tutta la letteratura disponibile riguardo le certificazioni e di fungere quindi da manuale per quanto riguarda la loro diffusione, l'impatto sulle performance e riguardo alla loro governance. L'aspetto interessante consiste nell'analisi dettagliata della letteratura circa l'adozione e la diffusione delle ISO 9000, sia a livello micro (con indagini e casi aziendali), in cui è stato chiesto direttamente alle imprese di fornire le motivazioni circa la decisione di certificarsi e i benefici tratti, sia a livello macro (tipicamente in termini di Paesi) sulla base dei database disponibili.

² Secondo la definizione fornita dall'ISO la famiglia di standard ISO 14000 fornisce strumenti pratici per imprese e organizzazioni di ogni tipo per la gestione delle loro responsabilità ambientali. International Organization for Standardization, www.iso.org/iso-14001-environmental-management (ultimo accesso 16-01-2018).

2.2 Logistic model

La prima fase del processo di diffusione delle certificazioni ISO 9000 presenta un comportamento molto simile al fenomeno di crescita delle popolazioni biologiche in un habitat di risorse limitate, oppure in generale al processo di diffusione delle nuove tecnologie (Franceschini *et al.*, 2004). Infatti, fino a un certo punto (corrispondente al raggiungimento del livello di saturazione) il processo è ben descrivibile attraverso il modello demografico logistico proposto dal matematico e statistico belga Pierre Verhulst (1838) e adottato per il fenomeno delle certificazioni ISO 9000 da Franceschini *et al.* (2004). Le ipotesi definite per l'applicazione del modello sono le seguenti:

- Il modello considera il numero totale di imprese certificate senza considerare la loro dimensione o il loro settore industriale;
- Il livello di saturazione è influenzato dalla concorrenza di mercato e dalle politiche economiche locali;
- La diffusione è influenzata da incentivi da parte dei governi locali e dal numero di enti certificatori;
- Non ci sono fattori esogeni che possano modificare l'evoluzione naturale della popolazione analizzata (numero di imprese certificate secondo le ISO 9000).

La curva-logistica-modificata per una certa popolazione ha la seguente espressione:

$$N(t) = \frac{N_0 \cdot K}{N_0 + (K - N_0)e^{-r_0 t}} - N_0$$

dove:

- $N(t)$ rappresenta il numero della popolazione: imprese certificate secondo le norme ISO 9000 nel tempo;
- r_0 è il tasso di crescita della popolazione in assenza di competizione interspecifica;
- N_0 è un fattore che consente di garantire la condizione $N(0) = 0$;
- $N(\infty) = (K - N_0)$ rappresenta il livello di saturazione della popolazione, coincide con il numero totale di imprese interessate al processo di certificazione.

Franceschini *et al.* (2004) hanno applicato questo modello al tasso percentuale di imprese certificate, ossia al rapporto tra il numero di certificazioni emesse e il numero complessivo di

imprese³, sia per il caso italiano sia più in generale per alcuni Paesi europei. Come prima considerazione viene osservato che il numero di imprese certificate è molto inferiore rispetto al numero complessivo di imprese presenti in un Paese. Questo sta a significare che il livello di saturazione, rappresentato dalla variabile $N(\infty)$, è solamente una frazione del numero di potenziali imprese e che quindi la diffusione delle certificazioni ISO 9000 risulta essere un fenomeno limitato. Gli autori, applicando il modello, hanno osservato in particolare un andamento di crescita secondo una curva-a-S propria del modello logistico, prevedendo nel caso italiano una stima del valore di $N(\infty)$ pari al 13% del numero totale di imprese (vedi Figura 8). Più ampiamente, il modello è stato applicato ad altri Paesi europei con una simile struttura economico-imprenditoriale (Francia, Germania, Paesi Bassi, Regno Unito e Spagna) definendo per ciascuno una previsione del livello di saturazione inferiore al 10% del numero totale di imprese (vedi Tabella 4).

Sampaio *et al.* (2009b), analizzando la relazione lineare tra il numero di certificazioni ISO 9000pc⁴ e la percentuale di imprese certificate con più di dieci dipendenti (vedi Figura 9), in base ai dati dell'ISO Survey per l'anno 2004, hanno osservato che proprio l'Italia, oltre ad avere un valore elevato di ISO 9000pc (pari a 1.47), era anche il Paese con il maggior numero di imprese certificate (42%), con un valore di molto superiore rispetto a quello predetto da Franceschini *et al.* (2004), lo stesso vale per Francia, Germania, Paesi Bassi, Regno Unito e Spagna per i quali si osservano valori percentuali di imprese certificate superiori rispetto a quelli predetti.

Il modello della curva-logistica-modificata proposto da Franceschini *et al.* (2004) può essere applicato dunque solamente alla fase iniziale del fenomeno di diffusione, fino al raggiungimento del livello di *plateau*, una volta raggiunto questo livello entrano in gioco altri meccanismi che influenzano il fenomeno ed è necessaria l'applicazione di un modello più appropriato (Franceschini *et al.*, 2006).

L'aspetto interessante che emerge da questa ricerca non è solamente come effettivamente il modello della curva-logistica-modificata si applichi bene al fenomeno di diffusione delle ISO

³ L'analisi si concentra esclusivamente sulla classe delle *corporation companies* in quanto categoria con il maggior numero di imprese certificate e bacino naturale di diffusione delle ISO 9000 (Franceschini *et al.*, 2004).

⁴ Numero di certificazioni ISO 9000 pro capite, ossia per 1'000 abitanti.

9000, ma l'interpretazione dell'effetto di saturazione. In accordo con gli autori infatti, a fronte di un certo periodo di crescita segue una fase di saturazione, questo suggerisce che quando il numero di imprese certificate raggiunge un certo limite, la certificazione perde la sua connotazione iniziale e diventa meno attrattiva per le imprese rimanenti non certificate (Franceschini *et al.*, 2006). Proprio in questa fase si verifica quindi una riduzione del gap competitivo tra imprese certificate e non-certificate facendo sì che il numero di imprese potenzialmente interessate a certificarsi decresca.

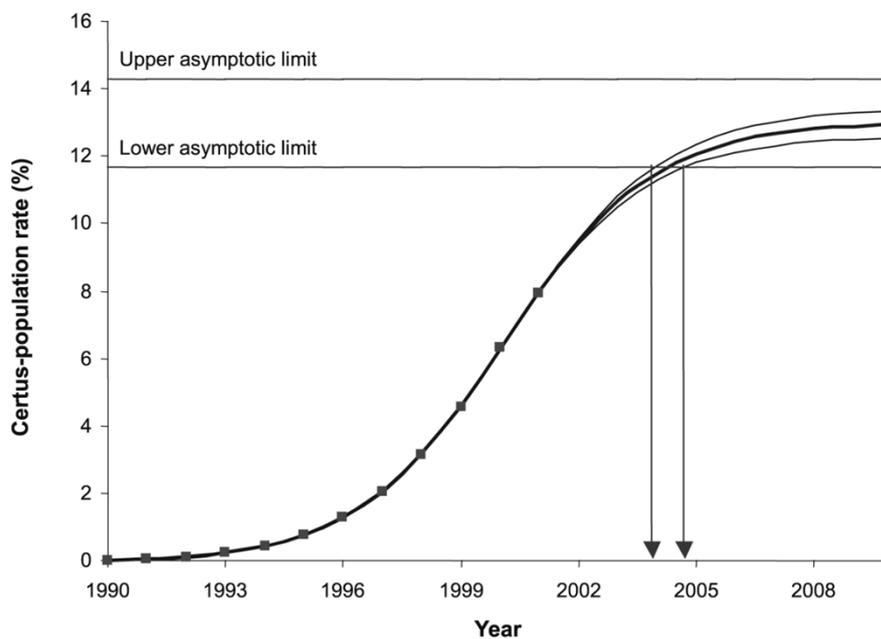


Figura 8
 Previsione della crescita delle imprese italiane certificate secondo le ISO 9000⁵

Fonte: Franceschini *et al.* (2004)

⁵ La curva mostra l'andamento del tasso percentuale di *corporation* italiane certificate nel tempo. Si possono osservare i dati empirici (punti sulla curva), la curva logistica (linea spessa) e gli intervalli di confidenza posti al 95% (curve sottili). Ponendo una tolleranza del 10% rispetto al valore asintotico del 13% si ottengono i limiti asintotici superiore e inferiore. La fase di saturazione predetta comincia ad inizio 2004 quando la curva di previsione incontra il limite asintotico inferiore (Franceschini *et al.* 2004).

| Parametri | Italia | | Regno Unito | | Paesi Bassi | | Germania | | Spagna | | Francia | |
|----------------------------------|--------------------|------------------------------|-------------------|------------------------------|-------------------|------------------------------|-------------------|------------------------------|-------------------|------------------------------|-------------------|------------------------------|
| | Valori | Intervallo di confidenza 95% | Valori | Intervallo di confidenza 95% | Valori | Intervallo di confidenza 95% | Valori | Intervallo di confidenza 95% | Valori | Intervallo di confidenza 95% | Valori | Intervallo di confidenza 95% |
| N_0 (%) | 0.063 | (0.058;0.067) | 0.080 | (0.008;0.152) | 0.016 | (0.004;0.027) | 0.023 | (0.005;0.041) | 0.004 | (0.002;0.006) | 0.012 | (0.005;0.019) |
| K (%) | 13.07 | (12.66;13.48) | 9.55 | (9.08;10.01) | 6.93 | (6.63;7.23) | 8.12 | (7.36;8.89) | 3.01 | (2.63;3.40) | 2.03 | (1.87;2.19) |
| r_0 (year ⁻¹) | 0.53 | (0.52;0.54) | 0.68 | (0.56;0.81) | 0.88 | (0.76;1.00) | 0.68 | (0.57;0.78) | 0.55 | (0.51;0.60) | 0.62 | (0.53;0.70) |
| Previsione in cinque anni (% CC) | anno 2006 12.45 | (12.12;12.77) | anno 2005 9.47 | (9.05;9.88) | anno 2005 6.91 | (6.62;7.20) | anno 2005 8.07 | (7.35;8.79) | anno 2006 2.91 | (2.59;3.23) | anno 2005 2.01 | (1.87;2.16) |

Note: La tabella mostra i valori dei parametri del modello e i rispettivi intervalli di confidenza (95%). È riportata anche la previsione in cinque anni del valore asintotico di saturazione come percentuale delle *corporation companies* (% CC).

Tabella 5
Risultati del modello logistico per
Italia, Regno Unito, Paesi Bassi,
Germania, Spagna e Francia.

Fonte: Franceschini et al. (2004)

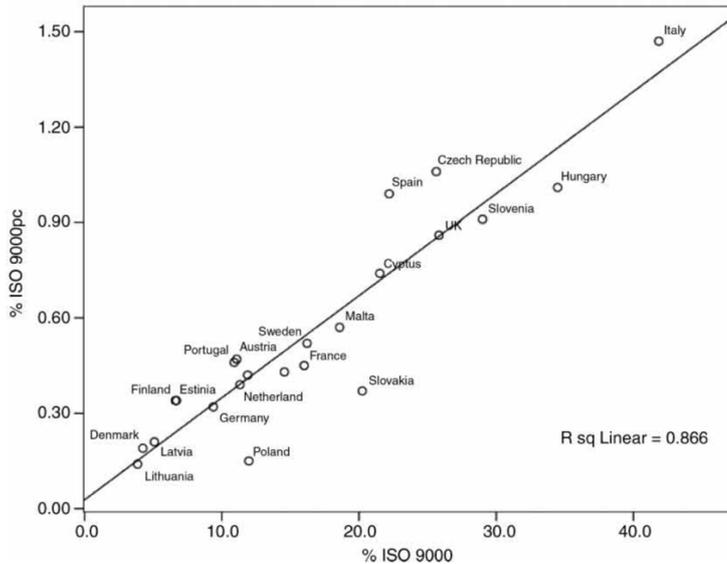


Figura 9
 Relazione tra la percentuale di ISO 9000pc e la percentuale di imprese certificate

Fonte: Sampaio et al. (2009b)

2.3 Cluster-analysis

Con il termine *clustering* ci si riferisce a una varietà di procedure che consentono di formare gruppi (cluster) di entità che siano omogenei al loro interno ma il più possibile diversi tra loro (Aldenderfer e Blashfield, 1984). Nello specifico, i metodi di *clustering* costituiscono delle procedure statistiche multivariate che partono da un data set contenente informazioni riguardo alle singole entità e le riorganizzano in gruppi il più possibile omogenei, massimizzando quindi la varianza tra i cluster e minimizzando quella al loro interno.

Per la sua natura, il *clustering* risulta essere uno strumento valido per identificare i diversi schemi di diffusioni delle certificazioni ISO 9000 nei Paesi. In quest'ottica, gli oggetti da raggruppare sono i Paesi stessi e le entità del data set rappresentanti i Paesi sono le serie storiche descrittive l'evoluzione negli anni del numero di certificazioni ISO 9000 in ciascun Paese.

Il processo di *clustering* di serie storiche si può suddividere in quattro fasi (Hastie et al., 2009):

- Rappresentazione delle serie storiche. Fase che definisce il modello con cui rappresentare le serie. Queste possono essere trattate in modo grezzo, utilizzando il data set originario (*Time-domain continuous representation*), oppure si possono trasformare in un dominio comune confrontabile (*Transformation based*), se le serie storiche sono continue possono essere discretizzate (*Discretisation based*) oppure si può cercare di definire il modello originario che le ha generate (*Generative models*).

- Definizione di una misura di similarità al fine di confrontare le serie storiche. Esistono diverse misure di distanza tra cui ad esempio la distanza *Minkowski* (un cui caso particolare è la distanza Euclidea), il *Dynamic Time Warping* e il coefficiente di correlazione di *Pearson*.
- Definizione e applicazione degli algoritmi di *clustering*. La scelta dell'algoritmo dipende dal tipo di dati che si hanno a disposizione. Si possono distinguere gli algoritmi in due famiglie:
 - *Clustering* partizionale. Questi algoritmi creano delle partizioni degli oggetti analizzati in un numero noto di gruppi, dove ogni oggetto può appartenere ad un solo gruppo. In questa famiglia rientra ad esempio l'algoritmo *k-means* (MacQueen, 1967)
 - *Clustering* gerarchico. Questi algoritmi creano una gerarchia dei gruppi producendo una struttura ad albero detta dendrogramma. A questa famiglia appartiene ad esempio l'algoritmo *single-linkage clustering* (Gower e Ross, 1969).
- Validazione. Fase in cui si verifica che gli algoritmi implementati forniscano risultati coerenti con le informazioni di partenza.

Le varie fasi sono analizzate nel dettaglio nel capitolo successivo in cui viene mostrato come la tecnica di *clustering* gerarchico è stata applicata ai dati dell'ultimo *ISO Survey* per l'anno 2016 (ISO, 2017).

Diversi autori, tra cui Sampaio *et al.* e Franceschini *et al.*, hanno applicato i metodi di *clustering* per identificare i gruppi di Paesi sulla base dell'evoluzione delle certificazioni ISO 9000 al loro interno, definendo in questo modo i vari aspetti della diffusione delle certificazioni, le variabili che la influenzano e sviluppando modelli matematici al fine di descrivere l'evoluzione delle ISO 9000 Paese per Paese.

2.3.1 Clustering e modello di previsione proposto da Sampaio *et al* (2011)

Sampaio *et al.* (2011), partendo dai risultati ottenuti da ricerche precedenti, hanno effettuato una *cluster-analysis* sulla base del data set relativo al numero di certificazioni ISO 9001 fornito dall'*ISO Survey* 2005, al fine di identificare gruppi di Paesi nel mondo aventi schemi di diffusione simili tra loro. La loro analisi non si fonda sul numero assoluto di certificazioni ma sul numero di ISO 9001 per 1'000 abitanti per Paese (ISO 9000pc) con l'obiettivo di normalizzare

i valori delle serie storiche sulla base della dimensione dei Paesi (misurata in termini di numerosità della popolazione). Questa scelta può essere criticata in quanto la popolazione non è un corretto indice di misura della dimensione di un Paese per un'analisi relativa alla numerosità delle certificazioni ISO 9000, le quali interessano piuttosto imprese e organizzazioni. Gli autori inoltre prendono in considerazione solamente i Paesi che tra il 1993 e il 2005 hanno raggiunto il valore di 0.05 ISO 9000pc, al di sotto di questo valore si considera il fenomeno di diffusione della certificazione ISO 9001 limitato. Questa scelta porta alla perdita di informazioni ma consente di definire una scala temporale che consideri l'origine ($t = 0$) come l'anno in cui un determinato Paese ha superato la soglia di 0.05 ISO 9000pc, eliminando così dall'analisi eventuali problematiche relative ai differenti anni in cui la diffusione delle certificazioni ha preso piede nei vari Paesi.

L'algoritmo di *clustering* applicato è il *k-means*, introdotto per la prima volta da MacQueen nel 1967. Questo algoritmo consiste in un metodo partizionale iterativo che prevede di distribuire inizialmente gli oggetti del data set in un numero definito a priori di cluster, definendo per ciascuno il corrispondente baricentro. A questo punto si alloca nei vari cluster ciascun Paese che abbia valore del baricentro vicino a quello del cluster, in seguito si calcolano nuovamente i baricentri dei gruppi ottenuti e si riallocano i Paesi iterando il processo fino a quando non è più necessario spostare i Paesi da un cluster ad un altro. Teoricamente sarebbe necessario prendere in considerazione tutte le possibili partizioni iniziali del data set; tuttavia, essendo questo un processo oneroso dal punto di vista computazionale, è possibile applicare procedure euristiche che consentono di ridurre il numero possibile di partizioni.

I risultati ottenuti da Sampaio *et al.* (2011) mostrano la definizione di tre diversi cluster:

- Cluster 1, composto dai Paesi leader in termini di ISO 9000pc in quanto hanno raggiunto valori compresi tra 1.00 e 1.80. All'interno di questo gruppo troviamo alcuni Paesi europei come l'Italia, la Repubblica Ceca, l'Ungheria, la Serbia, la Spagna e la Svizzera;
- Cluster 2, comprende la maggior parte dei Paesi analizzati con valori di ISO 9000pc inferiori a 1.00 e presentano schemi di crescita più attenuati con un tasso di crescita inferiore rispetto a quello dei Paesi del cluster 1;

- Cluster 3, composto da Australia, Irlanda, Lettonia, Paesi Bassi, Nuova Zelanda e Regno Unito. Gli schemi di diffusione in questi Paesi sono caratterizzati da una decrescita di certificazioni tra il 1993 e il 2005. L'aspetto interessante è che alcuni Paesi di questo cluster appartengono al Commonwealth e hanno quindi tra loro importanti relazioni commerciali ed economiche. Il fatto che presentino anche schemi di diffusione simili sta ad indicare come l'adozione di sistemi per la gestione della qualità sia influenzata dalle relazioni economico-commerciali tra i Paesi (Sampaio *et al.*, 2011).

Sulla base dei risultati precedentemente ottenuti e considerando la relazione positiva tra il numero di imprese certificate secondo l'ISO 9001 per 1'000 abitanti ($ISO\ 9000_{pc}$) e il reddito nazionale lordo pro capite (GNI_{pc}), gli autori hanno sviluppato un modello di regressione multipla per ciascun cluster al fine di descrivere e prevedere l'evoluzione delle certificazioni ISO 9000. Il modello presenta tre diverse espressioni (ognuna relativa a ciascun cluster) con cui è possibile prevedere il numero di certificazioni ISO 9001 per ciascun Paese sulla base della sua popolazione, del reddito nazionale lordo pro capite e del numero di certificazioni nell'anno precedente. Di seguito vengono riportate le espressioni delle funzioni ottenute:

$$\text{Cluster 1: } ISO9000_{pc(t)} = 0.444 + 0.634 * (ISO9000_{pc(t-1)})^2$$

$$\text{Cluster 2: } ISO9000_{pc(t)} = 0.121 + 1.220 * (ISO9000_{pc(t-1)})^2 + 2.10E^{-6} * GNI_{pc}$$

$$\text{Cluster 3: } ISO9000_{pc(t)} = -0.953 + 0.489 * (ISO9000_{pc(t-1)})^2 + 0.132 * Ln(GNI_{pc})$$

Fonte: Sampaio *et al.* (2011)

Rispetto a Franceschini *et al.* (2004), attraverso questa ricerca Sampaio *et al.* (2011) suggeriscono che non si può parlare di un unico modello di previsione delle certificazioni ISO 9000 valido per tutti i Paesi, si dovrebbero individuare piuttosto diversi modelli sulla base dei diversi schemi di evoluzione delle ISO 9000 in ciascun Paese.

2.3.2 Clustering dei Paesi europei proposto da Franceschini *et al.* (2010)

In modo simile, Franceschini *et al.* (2010) hanno realizzato una *cluster-analysis* dei Paesi europei sulla base dell'ISO Survey per il 2008, al fine di creare una tassonomia della diffusione delle

certificazioni ISO 9000 in Europa. Per questa analisi gli autori hanno definito una misura di similarità (*dissimilarity metric* – $d_{A,B}$) per poter confrontare le diverse dinamiche di diffusione:

$$d_{A,B} = 1 - \text{corr}(n_A, n_B)$$

dove n_A e n_B sono il numero di certificazioni rilasciate rispettivamente nel Paese A e nel Paese B tra il 1992 e il 2007. In particolare:

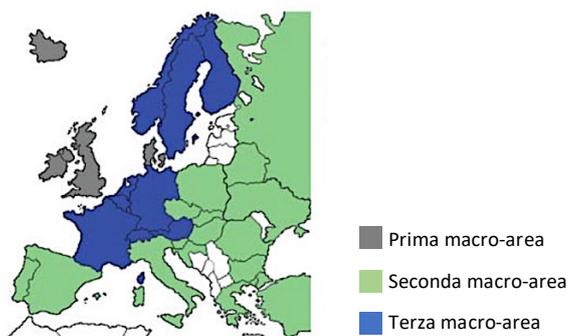
- $d_{A,B} \rightarrow 0$ indica un'elevata correlazione tra gli schemi di diffusione dei Paesi;
- $d_{A,B} \rightarrow 1$ indica una bassa correlazione tra gli schemi di diffusione dei Paesi;
- $d_{A,B} > 1$ indica una correlazione negativa.

Applicando l'algoritmo del *single-linkage clustering* (Gower e Ross, 1969) gli autori hanno individuato tre cluster escludendo dall'analisi i Paesi europei caratterizzati da una carenza di dati o con un numero molto basso di certificazioni ISO 9000 rilasciate. Le macro-aree individuate (corrispondenti ai tre cluster) sono:

- *Constant growth area*. Cluster costituito da quei Paesi in cui tra il 1993 e il 2007 le ISO 9000 sono state costantemente in crescita. A questo gruppo appartiene la maggior parte dei Paesi dell'Est-Europa insieme all'Italia e alla penisola Iberica, ossia Paesi caratterizzati da un mercato delle certificazioni relativamente giovane (Franceschini *et al.*, 2010);
- *Saturation area*. Questo cluster è costituito dai principali Paesi dell'Europa centro-occidentale insieme alla penisola scandinava, nei quali dopo una fase di crescita iniziale durata fino ai primi anni Duemila il numero di certificazioni si è stabilizzato attorno a un valore di saturazione;
- *Decline area*. Cluster costituito da Danimarca, Irlanda, Islanda e Regno Unito. In questi Paesi si osserva una fase di declino delle certificazioni. In accordo con Franceschini *et al.* (2010), il passaggio dalla versione ISO 9000:1994 alla ISO 9000:2000 è avvenuta in un momento in cui questi Paesi avevano ormai già raggiunto il livello di saturazione, i costi legati al passaggio alla nuova versione hanno rappresentato quindi un disincentivo per le imprese, riducendo la diffusione delle ISO 9000.

In Figura 10 è possibile osservare le tre macro-aree individuate. Un aspetto interessante è che sulla base dei risultati ottenuti, Franceschini *et al.* (2010) fanno notare che il fenomeno delle

ISO 9000 è “altamente contagioso”: si parla propriamente di un effetto a catena tra quei Paesi confinanti in cui le dinamiche del mercato delle certificazioni risultano essere simili. Gli autori attribuiscono a tale fenomeno tre motivi: pressione da parte dei competitors, un’impresa ricorre alle certificazioni per colmare il gap competitivo rispetto a concorrenti che operano nello stesso mercato locale; pressione da parte di imprese *customer*, le imprese clienti possono coinvolgere nel processo di certificazione i fornitori, i quali a loro volta coinvolgono i propri fornitori e così via; e infine economie e politiche simili che possono indurre le imprese a certificarsi. Inoltre, gli autori fanno notare che se le imprese che si trovano a dover definire per la prima volta un processo di gestione della qualità possono essere incentivate ad adottare le ISO 9000, per le imprese che possiedono già un buon sistema di gestione della qualità le certificazioni non fanno che aggiungere costi, documentazioni e ritardi, oltre al fatto che con il tempo e all’aumentare della loro diffusione si riduce il gap competitivo tra imprese certificate e non-certificate. Questo disincentiva le imprese a certificarsi e il numero di ISO 9000 rilasciate decresce di conseguenza.



Fonte: Franceschini et al. (2010)

Figura 10
Macro-aree ottenute applicando la cluster-analysis

CAPITOLO 3.

SCHEMI DI DIFFUSIONE DELLE CERTIFICAZIONI ISO 9000 IN EUROPA

3.1 Clustering gerarchico

Per poter identificare i diversi schemi di diffusione delle certificazioni ISO 9000 in Europa, sono state utilizzate le serie storiche relative al numero di certificazioni negli anni in ciascun Paese europeo come oggetto dell'implementazione degli algoritmi di *clustering* gerarchico agglomerativo.

Come anticipato nel Capitolo 2⁶, i metodi di *clustering* gerarchico consentono di produrre una rappresentazione gerarchica in cui, in ciascun livello vengono creati nuovi cluster fondendo insieme i cluster definiti nei livelli inferiori. Al livello più basso della gerarchia, ciascun cluster contiene un singolo oggetto. Al livello più alto, invece, un unico cluster contiene tutti gli oggetti dell'insieme di partenza (Hastie *et al.*, 2009). Le strategie di *clustering* gerarchico si basano su una misura di distanza tra gli oggetti definita a priori e si dividono in due tipologie: agglomerative (*bottom-up*) e divisive (*top-down*). Le strategie agglomerative partono dal basso e ad ogni livello della gerarchia fondono ricorsivamente una data coppia di cluster in un unico singolo cluster. Le strategie divisive, invece, partono dall'apice della gerarchia e ad ogni livello dividono ciascun cluster esistente in due nuovi cluster. Ciascun livello rappresenta quindi un particolare raggruppamento degli oggetti in cluster disgiunti e l'intera gerarchia è pertanto una sequenza ordinata di questi raggruppamenti. Graficamente l'output ottenuto è detto dendrogramma o albero gerarchico, propriamente quest'albero non rappresenta un singolo set di cluster, ma piuttosto una gerarchia multilivello dove i cluster ad un dato livello sono raggruppati in un unico cluster al livello superiore. In ultima analisi spetta quindi a chi analizza il dendrogramma definire quali siano i cluster in cui raggruppare gli oggetti di partenza.

Per il seguente studio si è scelto di implementare gli algoritmi di *clustering* gerarchico agglomerativo al fine di raggruppare i diversi Paesi europei secondo una precisa gerarchia. A tale scopo è stato utilizzato il programma *MATLAB R2017b*, sia implementando le funzioni rese

⁶ Vedi pp. 24-25 del cap. 2 di questo studio.

disponibili dall'ambiente di calcolo stesso, sia scrivendo direttamente i codici di funzioni accessorie laddove necessario.

3.2 Clustering in MATLAB

Per l'implementazione degli algoritmi di *clustering* gerarchico agglomerativo in *MATLAB* è necessario seguire i seguenti step (The MathWorks, Inc., 2017):

- Identificare un criterio di similarità (o dissimilarità) per ogni coppia di Paesi del data set.
- Raggruppare le coppie di Paesi in un albero gerarchico sulla base delle informazioni ottenute nello step precedente.
- Suddividere l'albero in cluster definendo un criterio con cui tagliare il dendrogramma, ottenendo in questo modo una partizione dei dati raggruppati in cluster.

L'ambiente di calcolo *MATLAB* mette a disposizione la funzione `clusterdata` (The MathWorks, Inc., 2017) che esegue autonomamente tutti gli step. Per la natura di questa ricerca e per la particolarità dei dati a disposizione, si è deciso di eseguire i vari passi singolarmente allo scopo di poter definire i giusti criteri e le opportune funzioni da applicare.

3.2.1 Normalizzazione dei valori e misure di distanza

Il primo passo consiste nel normalizzare i valori del data set per rendere il numero di certificazioni tra Paesi confrontabili. Questa operazione consente di trasformare gli oggetti analizzati in uno spazio comune in cui essi siano equiparabili (rappresentazione *transformation based*).

Il secondo passo consiste invece nel definire una distanza tra i Paesi attraverso un criterio di similarità (o dissimilarità). Di default *MATLAB* suggerisce di utilizzare la funzione `pdist` (The MathWorks, Inc., 2017) che calcola la distanza Euclidea per ogni coppia di Paesi. In generale, per un data set costituito da m Paesi, è possibile definire un numero di coppie pari a:

$$\frac{m(m-1)}{2}$$

Data in ingresso la matrice X contenente i Paesi del data set sulle righe e i corrispondenti valori del numero di certificazioni ISO 9000 nei diversi anni sulle colonne, la funzione `pdist` restituisce un vettore di dimensione $1 \times m(m - 1)/2$ in cui ciascun elemento corrisponde alla distanza per una data coppia di Paesi. Per implementare questa funzione si digita nella *Command Window* di *MATLAB*:

```
>> D = pdist(X);
```

A scopo esemplificativo, se si considera un data set costituito dal numero di certificazioni in cinque Paesi in un orizzonte temporale di due anni, la matrice X avrà dimensione 5×2 :

$$X = \begin{pmatrix} 1 & 2 \\ 3 & 5 \\ 2 & 2 \\ 4 & 2 \\ 4 & 3 \end{pmatrix}$$

In *MATLAB* la matrice in ingresso può essere definita in questo modo:

```
>> X = [1 2; 3 5; 2 2; 4 2; 4 3];
```

Dall'implementazione della funzione `pdist` si ottiene:

```
>> D = pdist(X)
```

```
D =
```

```
Column 1 through 7
3.6056    1.0000    3.0000    3.1623    3.1623    3.1623    2.2361
Column 8 through 10
2.0000    2.2361    1.0000
```

Si può osservare che partendo da un data set di cinque paesi il vettore D risultante ha dimensione 1×10 e contiene le informazioni di distanza tra i Paesi. È possibile formattare D in una matrice in cui sia le righe sia le colonne corrispondono ai Paesi del data set e un suo elemento generico d_{ij} corrisponde alla distanza misurata tra il Paese sulla riga i e il Paese sulla colonna j . Questa matrice è detta *dissimilarity matrix* o matrice delle distanze (`Ddist`). A tale scopo si utilizza il comando `squareform` digitando:

```
>> Ddist = squareform(D);
```

Se si implementa questo comando nell'esempio precedente dei cinque paesi si ottiene:

```
>> Ddist = squareform(D)
Ddist =
      0      3.6056      1.0000      3.0000      3.1623
    3.6056      0      3.1623      3.1623      2.2361
    1.0000      3.1623      0      2.0000      2.2361
    3.0000      3.1623      2.0000      0      1.0000
    3.1623      2.2361      2.2361      1.0000      0
```

Si può osservare quindi che il vettore `D` corrisponde esattamente alla matrice triangolare superiore (o inferiore) di `Ddist`. Questa matrice consente tuttavia di visualizzare in modo più semplice e immediato la relazione tra le informazioni di distanza generate dalla funzione `pdist` e gli oggetti del data set di partenza.

3.2.2 Collegamenti tra gli oggetti

Determinata la distanza tra i Paesi, la funzione `linkage` (The MathWorks, Inc., 2017) raggruppa questi oggetti in cluster sulla base delle informazioni contenute nella matrice delle distanze (o più propriamente nel vettore `D`). Questa funzione crea inizialmente dei cluster binari attraverso l'unione di Paesi vicini, questi cluster vengono poi fusi tra loro in cluster binari sempre più grandi fino a quando tutti i Paesi del data set sono collegati in una struttura gerarchica ad albero (dendrogramma). La funzione `linkage` si implementa digitando nella *Command Window* di *MATLAB*:

```
>> Z = linkage(D);
```

`Z` è una matrice $(m - 1) \times 3$, dove le prime due colonne identificano i Paesi o i cluster collegati, la terza colonna contiene l'informazione relativa alla distanza tra questi oggetti. La funzione `linkage` utilizza la distanza tra i Paesi come informazione di partenza, a sua volta però deve essere in grado di determinare la distanza tra i cluster che crea, a tale proposito utilizza di default la tecnica del *single-linkage*⁷.

Ad esempio, dato il vettore delle distanze `D` generato dalla funzione `pdist` nel caso del data set costituito da cinque Paesi, la funzione `linkage` fornisce in output la seguente matrice `Z`:

```
>> Z = linkage(D)
Z =
```

⁷ È possibile però specificare l'utilizzo di una delle seguenti tecniche alternative: *Complete Linkage*, *Weighted average distance* (WPGMA), *Unweighted average distance* (UPGMA), *Centroid distance* (UPGMC), *Weighted center of mass distance* (WPGMC) e *Inner squared distance* (algoritmo della minima varianza di Ward).

| | | |
|--------|--------|--------|
| 4.0000 | 5.0000 | 1.0000 |
| 1.0000 | 3.0000 | 1.0000 |
| 6.0000 | 7.0000 | 2.0000 |
| 2.0000 | 8.0000 | 2.2361 |

In quest'output ciascuna riga della matrice identifica un collegamento tra gli oggetti del data set o tra i cluster generati. Le prime due colonne identificano i Paesi collegati, la terza colonna contiene l'informazione di distanza tra questi oggetti. Nell'esempio la funzione raggruppa inizialmente i Paesi 4 e 5 che presentano un valore di distanza pari a 1.0000 e prosegue collegando i Paesi 1 e 3 che hanno a loro volta distanza pari a 1.0000. La terza riga indica che la funzione `linkage` raggruppa gli oggetti 6 e 7, tuttavia nel data set di partenza i Paesi sono cinque quindi propriamente l'oggetto 6 rappresenta il cluster binario creato contenente i Paesi 4 e 5, mentre l'oggetto 7 identifica il cluster binario costituito dai paesi 1 e 3. Ogni volta che `linkage` raggruppa due Paesi in un cluster assegna a quest'ultimo un indice partendo dal valore $m + 1$, dove m corrisponde al numero di Paesi del data set di partenza.

3.2.3 Dendrogramma

Per comprendere meglio la gerarchia creata dalla funzione `linkage` è possibile visualizzare graficamente l'albero gerarchico attraverso la funzione `dendrogram` (The MathWorks, Inc., 2017), la quale viene implementata nella *Command Window* di *MATLAB* digitando:

```
>> dendrogram(Z)
```

MATLAB restituisce un grafico in cui i numeri sull'asse orizzontale rappresentano gli indici dei Paesi che costituiscono il data set. I collegamenti tra i Paesi o tra i cluster binari del livello inferiore sono rappresentati attraverso delle U-rovesciate, la cui altezza identifica la distanza tra gli oggetti. Il valore della distanza è leggibile sull'asse verticale graduata a sinistra del grafico. In Figura 11 viene mostrato il dendrogramma risultante per il set di cinque Paesi considerato negli esempi precedenti. Si può osservare che il collegamento che rappresenta il cluster contenente i Paesi 1 e 3 ha altezza pari a 1, così come il collegamento che identifica il cluster dei Paesi 4 e 5. Invece, il collegamento rappresentante il cluster che raggruppa il Paese 2 con i Paesi 1, 3, 4 e 5 (cluster 8) ha altezza con valore pari a 2.2361, quest'ultimo rappresenta la distanza calcolata dalla funzione `linkage` tra gli oggetti 2 e 8.

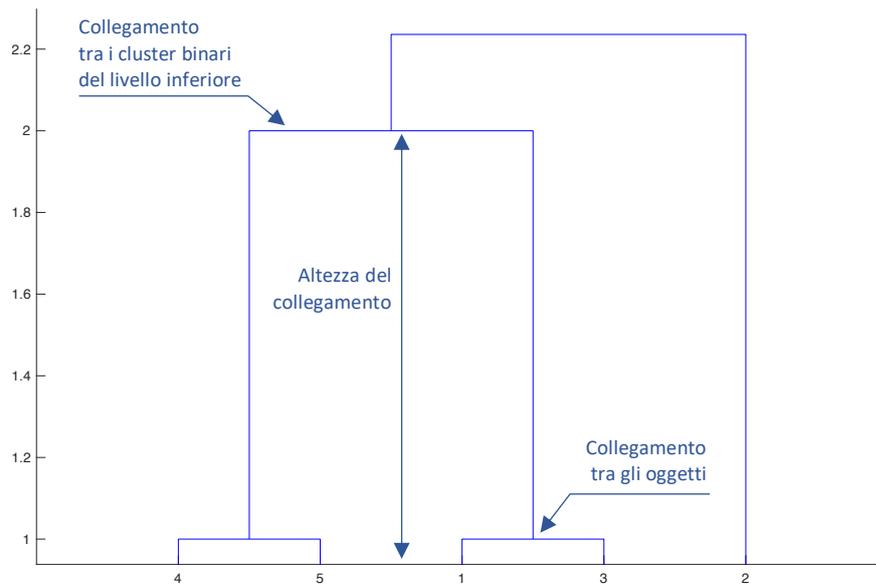


Figura 11
Esempio di
dendrogramma per il
set di cinque
Paesi

Fonte: MATLAB R2017b

3.2.4 Verifiche dei risultati

Una volta ottenuto l'albero gerarchico, è importante verificare che le distanze tra i Paesi sul dendrogramma (definite dalle altezze dei collegamenti) riflettano accuratamente le distanze originali determinate attraverso la misura di similarità (o dissimilarità), è opportuno inoltre indagare le divisioni naturali che esistono lungo i collegamenti tra gli oggetti. Questo tipo di analisi si concretizza con due verifiche distinte:

- *Verify Dissimilarity* – Verifica delle distanze
- *Verify Consistency* – Verifica della consistenza

Per quanto riguarda la prima verifica, si può osservare che nel dendrogramma due generici Paesi del data set iniziale sono collegati tra loro ad un certo livello dell'albero gerarchico. L'altezza del collegamento identifica graficamente la distanza tra due cluster che contengono i due Paesi. Questa altezza è definita *cophenetic distance*. Una modalità per verificare quanto il dendrogramma generato dalla funzione `Linkage` rifletta i dati di partenza, consiste nel confrontare le *cophenetic distance* con le informazioni contenute nella matrice delle distanze. A tal proposito, *MATLAB* consente di implementare la funzione `cophenet` (The MathWorks, Inc., 2017), la quale fornisce in output il *cophenetic correlation coefficient*. Questo coefficiente mostra la bontà di adattamento (*goodness of fit*) degli algoritmi di *clustering* gerarchico in maniera simile al coefficiente di correlazione nel caso di regressioni, più il valore di questo

coefficiente si avvicina a 1, più il dendrogramma riflette in maniera accurata i dati di partenza. Le informazioni necessarie per il calcolo del *cophenetic correlation coefficient* sono contenute nella matrice delle distanze e nella matrice *cophenet*. Quest'ultima viene determinata sulla base delle informazioni contenute nella matrice dei collegamenti Z. A titolo esemplificativo, in Tabella 6 vengono riportate la matrice triangolare inferiore delle distanze e la matrice triangolare inferiore *cophenet* per il caso dei cinque Paesi precedentemente analizzato. Per implementare la funzione *cophenet* si digita nella *Command Window* di *MATLAB*:

```
>> c = cophenet(Z,D);
```

dove *c* è il *cophenetic correlation coefficient*, Z è la matrice risultante implementando la funzione *linkage* e D è il vettore delle distanze. Nell'esempio dei cinque Paesi il valore di *c* risultante è pari a:

```
>> c = cophenet(Z,D)
c =
```

0.8618

Essendo questo valore prossimo a 1 si può affermare che i collegamenti creati dalla funzione *linkage* riflettono in maniera accurata le distanze di partenza tra i Paesi. Tipicamente si considerano accettabili valori superiori a 0.75.

Il *cophenetic correlation coefficient* è quindi un coefficiente di correlazione calcolato tra la matrice delle distanze e la matrice *cophenet*. Se si considerano il vettore delle distanze D e il vettore Y contenente le *cophenetic distance* (ossia la trasformazione in forma vettoriale della matrice *cophenet* triangolare inferiore – o superiore), si può definire formalmente il *cophenetic correlation coefficient* come:

$$c = \frac{\sum_{i<j}(D_{ij} - d)(Y_{ij} - y)}{\sqrt{\sum_{i<j}(D_{ij} - d)^2 \sum_{i<j}(Y_{ij} - y)^2}}$$

dove:

- D_{ij} è la distanza tra il Paese *i* e il Paese *j* nel vettore D;
- Y_{ij} è la *cophenetic distance* tra il Paese *i* e il Paese *j* nel vettore Y;
- *d* e *y* sono i valori medi di D e Y

Il *cophenetic correlation coefficient* può essere utilizzato inoltre per testare su uno stesso data set di partenza quale sia il miglior criterio di calcolo delle distanze tra i Paesi e quale sia l'algoritmo di *clustering* più adatto da applicare attraverso la funzione `Linkage`.

| Matrice delle Distanze | | | | | |
|-------------------------------|--------|--------|--------|--------|---|
| Paesi | 1 | 2 | 3 | 4 | 5 |
| 1 | 0 | | | | |
| 2 | 3.6056 | 0 | | | |
| 3 | 1.0000 | 3.1623 | 0 | | |
| 4 | 3.0000 | 3.1623 | 2.0000 | 0 | |
| 5 | 3.1623 | 2.2361 | 2.2361 | 1.0000 | 0 |

| Matrice Cophenet | | | | | |
|-------------------------|--------|--------|--------|--------|---|
| Paesi | 1 | 2 | 3 | 4 | 5 |
| 1 | 0 | | | | |
| 2 | 2.2361 | 0 | | | |
| 3 | 1.0000 | 2.2361 | 0 | | |
| 4 | 2.0000 | 2.2361 | 2.0000 | 0 | |
| 5 | 2.0000 | 2.2361 | 2.0000 | 1.0000 | 0 |

Tabella 6
Matrice triangolare inferiore delle distanze e Matrice triangolare inferiore Cophenet per il set di cinque Paesi

Per quanto riguarda la seconda verifica, un modo per determinare la divisione naturale dei Paesi in cluster consiste nel confrontare sul dendrogramma l'altezza di ciascun collegamento con l'altezza dei collegamenti a esso sottostanti. Infatti, un collegamento che ha relativamente la stessa altezza di quelli sottostanti, indica l'assenza di distinzioni in termini di distanza tra i Paesi collegati a quel dato livello della gerarchia. In questo caso si parla di elevato livello di consistenza. Al contrario, un collegamento la cui altezza differisce molto dalle altezze dei collegamenti a esso sottostanti, sta a indicare che i Paesi uniti a questo livello della gerarchia sono in realtà molto più lontani tra loro rispetto ai Paesi uniti ai livelli inferiori. In questo caso si parla di inconsistenza del collegamento rispetto a quelli sottostanti (in Figura 12 viene mostrato un esempio di consistenza e inconsistenza dei collegamenti). In ambito di *cluster-analysis*, l'inconsistenza può essere utilizzata come criterio per la divisione naturale dei Paesi in cluster. In particolare, in *MATLAB* la funzione `cluster` utilizza quest'informazione per determinare automaticamente dove e come suddividere gli elementi del data set.

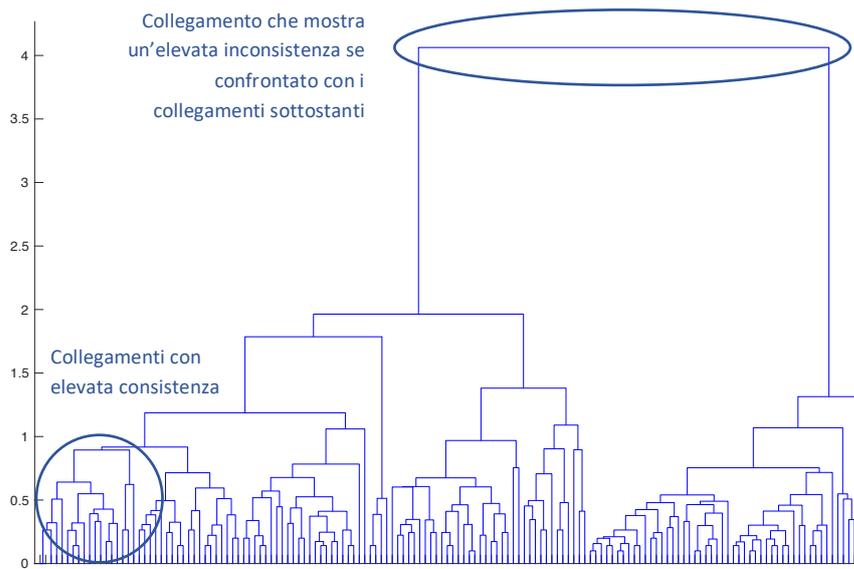


Figura 12
Esempio di consistenza e inconsistenza dei collegamenti in un dendrogramma

Fonte: MATLAB R2017b

È possibile quantificare il livello d'inconsistenza di ciascun collegamento attraverso l'*inconsistency coefficient*, che confronta l'altezza di ciascun collegamento del dendrogramma con l'altezza dei collegamenti ad esso sottostanti. In *MATLAB*, attraverso la funzione `inconsistent` (The MathWorks, Inc., 2017), è possibile generare una lista dei valori dei coefficienti d'inconsistenza relativi a ciascun collegamento. Questa funzione confronta ciascun collegamento con i collegamenti a meno di due livelli sottostanti, questo limite viene definito *depth of the comparison* (profondità del confronto) ed è fissato per limitare il lavoro computazionale dell'algoritmo⁸. Per implementare la funzione `inconsistent` si digita nella *Command Window* di *MATLAB*:

```
>> I = inconsistent(Z);
```

La funzione fornisce una matrice **I** di dimensione $(m - 1) \times 4$, le cui righe corrispondono ai collegamenti definiti nella matrice **Z** con lo stesso ordine, le colonne invece definiscono per ogni riga *k*:

- La media delle altezze di tutti i collegamenti inclusi nel calcolo (colonna 1);
- La deviazione standard di tutti i collegamenti inclusi nel calcolo (colonna 2);

⁸ È possibile definire un livello diverso di profondità. Per la seguente analisi è stata mantenuta una profondità pari a due livelli della gerarchia.

- Il numero di collegamenti inclusi nel calcolo (colonna 3);
- Il valore del coefficiente d'inconsistenza del collegamento k (colonna 4).

Nell'esempio dei cinque Paesi, la matrice I risultante dall'implementazione della funzione `inconsistent` è la seguente:

```
>> I=inconsistent(Z)
I =
```

| | | | |
|--------|--------|--------|--------|
| 1.0000 | 0 | 1.0000 | 0 |
| 1.0000 | 0 | 1.0000 | 0 |
| 1.3333 | 0.5774 | 3.0000 | 1.1547 |
| 2.1180 | 0.1669 | 2.0000 | 0.7071 |

In questo esempio, la prima riga rappresenta il collegamento tra i Paesi 4 e 5. Poiché entrambi i Paesi sono nodi *leaf*, ossia alla base della gerarchia, il valore del coefficiente d'inconsistenza per questo cluster è pari a zero. La seconda riga rappresenta il collegamento tra i Paesi 1 e 3, anch'essi nodi *leaf*. La terza riga identifica il collegamento tra i due cluster precedenti, la terza colonna in questo caso indica che i collegamenti considerati nel calcolo sono tre: il collegamento in questione e i due collegamenti sottostanti (che uniscono i Paesi 4 e 5 e i Paesi 1 e 3). La prima colonna indica la media delle altezze di questi tre collegamenti, la seconda colonna indica la deviazione standard calcolata sulle altezze dei tre collegamenti e infine la quarta colonna contiene il valore d'inconsistenza pari a 1.1547, calcolato come la differenza tra l'altezza del collegamento analizzato (2.0000) e la media delle altezze (1.3333), normalizzata secondo la deviazione standard (0.5774).

Formalmente, per ogni collegamento k l'*inconsistency coefficient* (contenuto nella quarta colonna di I) viene calcolato come:

$$I(k, 4) = \frac{Z(k, 3) - I(k, 1)}{I(k, 2)}$$

dove:

- $Z(k, 3)$ è l'altezza sul dendrogramma del collegamento k (informazione contenuta nella terza colonna della matrice Z);
- $I(k, 1)$ è la media delle altezze dei collegamenti inclusi nel calcolo, ossia k e tutti i collegamenti sottostanti (prima colonna della matrice I);

- $I(k, 2)$ è infine la deviazione standard calcolata su tutti i collegamenti inclusi (seconda colonna della matrice I).

L'*inconsistency coefficient* assume quindi il significato di una standardizzazione dei collegamenti definiti nel dendrogramma per rendere questi tra loro confrontabili. Ciascun collegamento viene pertanto considerato come una variabile aleatoria con media μ pari al valore medio delle altezze dei collegamenti considerati (ossia del collegamento analizzato più i collegamenti ad esso sottostanti) e con deviazione standard σ calcolata sulle altezze dei collegamenti considerati. Più è alto il valore del coefficiente d'inconsistenza del collegamento in questione, più i Paesi uniti a quel dato livello della gerarchia sono distanti tra loro rispetto ai Paesi collegati ai livelli inferiori.

3.2.5 Creazione dei cluster

Per la definizione dei cluster, *MATLAB* consente l'utilizzo della funzione `cluster` (The MathWorks, Inc., 2017) che permette di creare i cluster in due diversi modi:

- Specificando un numero N di cluster arbitrari;
- Trovando la suddivisione naturale dei Paesi.

La prima modalità consiste nello specificare a priori il numero di cluster con cui si intende suddividere i Paesi, attraverso la funzione `cluster` e l'opzione `maxclust` si indica nella *Command Window* di *MATLAB* il numero N di cluster desiderati:

```
>> T = cluster(Z, 'maxclust', N);
```

L'output consiste in un vettore T che ha la stessa dimensione del data set originario, ossia ciascun elemento del vettore si riferisce in ordine ad uno specifico Paese, il valore dell'elemento invece corrisponde al numero del cluster in cui il Paese viene inserito. T è quindi un vettore di valori compresi tra 1 e N .

Ad esempio, nel caso dei cinque Paesi, è possibile specificare che il numero di cluster con cui la funzione `cluster` deve partizionare il data set di partenza sia pari a due. In questo caso, la funzione crea un cluster contenente i Paesi 1, 3, 4 e 5 e un secondo cluster contenente il Paese 2:

```
>> T = cluster(Z, 'maxclust', 2)
T =
```

```
2
1
2
2
2
```

Se invece si impone un numero di cluster pari a tre la funzione `cluster` raggruppa i Paesi 1 e 3 in un cluster, i Paesi 4 e 5 in un secondo cluster e il Paese 2 in un terzo cluster:

```
>> T = cluster(Z, 'maxclust', 3)
T =
```

```
2
3
2
1
1
```

Per il seguente studio non è stato possibile conoscere a priori il numero di cluster in cui raggruppare i Paesi europei, si è quindi optato per la divisione naturale. A tal proposito, il coefficiente d'inconsistenza può essere considerato come indice per identificare i diversi cluster. Attraverso la funzione `cluster` infatti, si fissa un certo valore del coefficiente d'inconsistenza e la funzione, partendo dai nodi *leaf*, analizza ciascun collegamento mantenendo intatti quelli che presentano un valore del coefficiente inferiore al valore fissato e spezzando tutti gli altri. La funzione `cluster` si implementa come segue:

```
>> T = cluster(Z, 'cutoff', a);
```

dove `a` è il valore di taglio fissato. L'output consiste in un vettore `T` che, come nel caso precedente ha la stessa dimensione del data set originario e ciascun elemento rappresenta il numero del cluster in cui il singolo Paese viene inserito.

Nell'esempio dei cinque Paesi, se si impone un coefficiente di taglio pari a 1.2, la funzione `cluster` raggruppa tutti i Paesi in un singolo cluster in quanto nessun collegamento dell'albero gerarchico presenta un coefficiente d'inconsistenza con valore superiore a 1.2 in `I`:

```
>> T = cluster(Z, 'cutoff', 1.2)
T =
```

```
1
1
1
1
1
```

Se si abbassa il coefficiente di taglio al valore 0.8, la funzione `cluster` raggruppa i Paesi in tre cluster distinti:

```
>> T = cluster(Z, 'cutoff', 0.8)
T =
     2
     3
     2
     1
     1
```

Questo output indica che i Paesi 1 e 3 sono raggruppati in un primo cluster, i Paesi 4 e 5 in un secondo cluster e infine il Paese 2 è contenuto nel terzo cluster.

In alternativa, *MATLAB* consente di individuare i cluster con un taglio orizzontale del dendrogramma attraverso l'opzione `criterion`, in questo modo il criterio di suddivisione dei Paesi in cluster si basa sulla distanza piuttosto che sul coefficiente d'inconsistenza.

Nei paragrafi successivi viene spiegato come gli algoritmi sopra citati sono implementabili nel caso dei dati aggiornati relativi al numero di certificazioni ISO 9000 per i Paesi europei.

3.3 Selezione del campione di Paesi europei e normalizzazione dei valori

Per poter identificare in modo appropriato i diversi schemi di diffusione delle certificazioni ISO 9000 in Europa, è possibile selezionare solo un campione di Paesi europei tra tutti quelli presenti nel data set dell'ultimo *ISO Survey* (ISO, 2017). In particolare, si possono escludere i Paesi in cui il fenomeno di diffusione delle ISO 9000 risulta essere limitato, concentrato negli ultimi anni oppure in cui i dati a disposizione sono scarsi. Il criterio adottato consiste nel considerare solamente i Paesi in cui, nell'orizzonte temporale 1993-2016, sono state registrate almeno 1'000 certificazioni e in cui entro il 2006 sono state registrate almeno 500 certificazioni. In tabella 7 viene riportato l'elenco dei Paesi europei selezionati in Tabella 8 invece i Paesi esclusi dall'analisi⁹.

⁹ Per i dati annuali del numero di certificazioni ISO 9000 relativi ai paesi europei si rimanda alle tabelle 3 e 4 nelle pp. 15-16 del cap. 1 di questo studio.

In seguito, i valori relativi al numero di certificazioni ISO 9000 possono essere normalizzati secondo il criterio del valore massimo:

$$X_t^n = \frac{X_t}{X_{max}}$$

dove X_t corrisponde al numero di certificazioni ISO 9000 in un Paese nell'anno t , X_{max} è il numero massimo di certificazioni in un Paese nell'orizzonte temporale 1993-2016 e X_t^n è il valore normalizzato risultante. La scelta di questo criterio è legata al fatto che la scala relativa al numero di certificazioni può essere considerata di rapporto poiché lo zero assoluto è immediatamente identificabile come assenza di certificazioni. Pertanto, la normalizzazione rispetto al valore massimo, in quanto trasformazione lineare, è ammissibile per questo tipo di scale.

| Paese | Paese |
|-------------------|----------------|
| 1 Austria | 16 Lituania |
| 2 Bielorussia | 17 Paesi Bassi |
| 3 Belgio | 18 Norvegia |
| 4 Bulgaria | 19 Polonia |
| 5 Croazia | 20 Portogallo |
| 6 Repubblica Ceca | 21 Romania |
| 7 Danimarca | 22 Russia |
| 8 Finlandia | 23 Slovacchia |
| 9 Francia | 24 Slovenia |
| 10 Germania | 25 Spagna |
| 11 Grecia | 26 Svezia |
| 12 Ungheria | 27 Svizzera |
| 13 Irlanda | 28 Turchia |
| 14 Italia | 29 Ucraina |
| 15 Lettonia | 30 Regno Unito |

Tabella 7
Campione di 30 Paesi europei selezionati

| Paese | Paese |
|----------------------|------------------|
| Albania | Kosovo |
| Andorra | Liechtenstein |
| Armenia | Lussemburgo |
| Azerbaigian | Macedonia |
| Bosnia ed Erzegovina | Malta |
| Cipro | Moldavia |
| Estonia | Princ. Di Monaco |
| Georgia | Montenegro |
| Gibilterra (UK) | San Marino |
| Islanda | Serbia |

Tabella 8
Paesi europei esclusi dall'analisi

3.4 Misure di distanza tra i Paesi europei

Per misurare la distanza tra due serie storiche si possono distinguere almeno due diversi approcci (Douzal Chouakria e Naidu Nagabhushan, 2007):

- APPROCCIO PARAMETRICO

Consiste in una proiezione delle serie storiche sulla base di uno spazio funzionale corrispondente ad esempio a un'approssimazione polinomiale, ARIMA o all'approssimazione DTF (Trasformata Discreta di Fourier) delle serie stesse. La distanza è determinata sulla base dei coefficienti delle funzioni (Caiado *et al.*, 2006; Garcia-Escudero e Gordaliza, 2005; Kakizawa *et al.*, 1998; Maharaj, 2000; Serban e Wasserman, 2005).

- APPROCCIO NON PARAMETRICO

Consiste in una valutazione della distanza tra le serie storiche sulla base della loro descrizione temporale. Il problema principale per questo tipo di approcci consiste nell'includere l'informazione di interdipendenza temporale tra le osservazioni che definiscono le serie storiche (Heckman e Zamar, 2000; Keller e Wittfeld, 2004; Oates *et al.*, 1999).

All'interno degli approcci non parametrici le principali misure di distanza utilizzate sono la distanza Euclidea e il *Dynamic Time Warping*, di seguito spiegati nel dettaglio.

3.4.1 Distanza Euclidea

Siano $S_1 = (u_1, \dots, u_n)$ e $S_2 = (v_1, \dots, v_n)$ due serie storiche di n valori osservati negli istanti di tempo (t_1, \dots, t_n) , si definisce distanza Euclidea δ_E tra S_1 e S_2 :

$$\delta_E(S_1, S_2) = \left(\sum_{i=1}^n (u_i - v_i)^2 \right)^{1/2}$$

La distanza Euclidea è probabilmente la misura più intuitiva e diffusa, richiede però che due generiche serie storiche abbiano lo stesso numero di osservazioni n . Nel data set fornito dall'ultimo ISO Survey per l'anno 2016 (ISO, 2017), si può osservare che per alcuni Paesi le osservazioni coprono solamente un intervallo dell'intero orizzonte temporale, proprio perché non c'è sincronizzazione tra i fenomeni di diffusione delle certificazioni nei diversi Paesi (Franceschini *et al.*, 2006). Non è possibile quindi calcolare direttamente la distanza Euclidea tra le serie. Al contrario, il *Dynamic Time Warping* consente invece di espandere o comprimere

i segnali rispetto al tempo, ridefinendoli in uno spazio comune confrontabile. Douzal e Naidu (2007) fanno notare inoltre che, utilizzando la distanza Euclidea, la similarità tra due serie storiche dipende esclusivamente dalla prossimità dei valori in ogni istante temporale considerato. Questo sta a significare che le osservazioni vengono considerate indipendenti tra loro e che δ_E rimane invariato a fronte di trasformazioni che modificano l'ordine delle coppie di osservazioni, in altre parole la distanza Euclidea non tiene conto dell'interdipendenza tra i valori osservati.

3.4.2 Dynamic Time Warping – DTW

Il *Dynamic Time Warping* (DTW) è una tecnica che consente di allineare due sequenze, anche con numero di osservazioni diverse, calcolandone la distanza. L'algoritmo DTW risulta essere dunque un buon approccio per risolvere il problema del disallineamento temporale delle serie storiche relative alla diffusione delle certificazioni ISO 9000 nei Paesi Europei.

Il *Dynamic Time Warping* è stato ampiamente studiato da Sankoff e Kruskal (1983). Nel 1994 Berndt e Clifford hanno analizzato l'utilizzo del DTW in ambito del riconoscimento vocale, infatti il problema dell'individuazione delle parole all'interno di un discorso presenta aspetti simili all'individuazione di pattern di serie storiche per un fenomeno continuo. In generale, il DTW consente di trovare un allineamento ottimale tra due segnali attraverso la loro distorsione (*warp*) non lineare rispetto alla variabile tempo.

In particolare, date due serie storiche $S_1 = (u_1, \dots, u_n)$ e $S_2 = (v_1, \dots, v_m)$, è possibile definire una griglia di dimensioni $m \times n$ in cui ciascun elemento (i, j) corrisponde all'allineamento tra il generico elemento u_i e il generico elemento v_j . In Figura 13 si può osservare un esempio di due generiche serie con numero di osservazioni diverse, $x[t] = (1, 1, 2, 3, 2, 0)$ e $y[t] = (0, 1, 1, 2, 3, 2, 1)$, e la corrispondente griglia: ogni riga della griglia corrisponde agli istanti di tempo relativi alla serie $y[t]$ e ogni colonna corrisponda agli istanti di tempo relativi alla serie $x[t]$. Ciascun elemento della griglia contiene la distanza $d(u_i, v_j)$ tra due generici valori u_i e v_j , con $i = 1, \dots, n$ e $j = 1, \dots, m$. Per calcolare questi valori è necessario definire una misura di distanza tra gli elementi delle serie, le misure proposte da Berndt e Clifford (1994) sono:

- $\delta(i, j) = |u_i - v_j|$ (magnitudo della differenza)¹⁰

¹⁰Per la natura dei dati della seguente ricerca la magnitudo della differenza coincide con la distanza Euclidea:

- $\delta(i, j) = (u_i - v_j)^2$ (quadrato della differenza)

Definita la griglia, il DTW determina le coppie di elementi delle due sequenze che garantiscono l'allineamento migliore e che definiscono il *warping path* (W) ottimale. Quest'ultimo è propriamente un percorso sulla griglia che allinea gli elementi di S_1 e di S_2 minimizzando la distanza tra le due serie:

$$W = w_1, w_2, \dots, w_k$$

W è dunque una mappatura tra S_1 e S_2 , dove ogni w_k corrisponde ad un punto $(i, j)_k$ della griglia, nel caso in cui non ci sia differenza temporale tra le serie, W coincide con la diagonale.

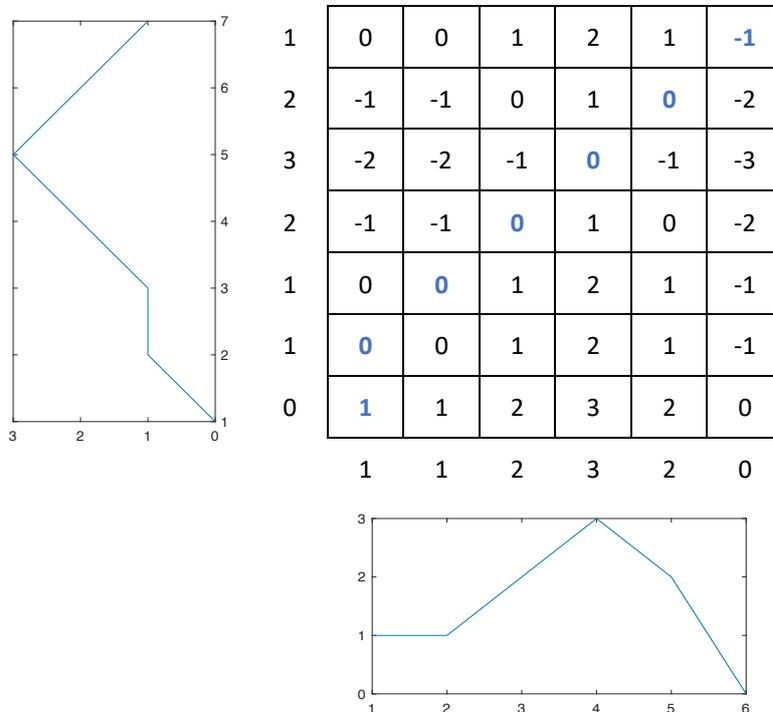


Figura 13
Esempio di griglia tra due serie storiche $x[t]$ (sull'asse orizzontale) e $y[t]$ (sull'asse verticale)

Nota: In questo esempio i valori sulla griglia sono determinati come semplice differenze tra i valori di $x[t]$ e di $y[t]$. In blu il *warping path* ottimale.

Fonte: *Dynamic Programming Example* (Coleman, 2017)

Poiché l'algoritmo confronta ogni punto di una serie con tutti i punti dell'altra, il numero di possibili W è molto alto, si osserva però una sequenza di w_k con valori minimi tipicamente

$$\delta(i, j) = |u_i - v_j| = \sqrt{(u_i - v_j)^2}$$

attorno alla diagonale. Per ridurre il campo di ricerca di tutti i possibili *warping path* vengono imposte le seguenti restrizioni:

1. Monotonia rispetto al tempo:

$$i_{k-1} \leq i_k \text{ e } j_{k-1} \leq j_k ;$$

2. Continuità per non omettere informazioni importanti:

$$i_k - i_{k-1} \leq 1 \text{ e } j_k - j_{k-1} \leq 1;$$

3. *Warping window*. Gli elementi della griglia ammissibili devono cadere all'interno di una finestra con ampiezza ω , questo garantisce che il *warping path* non si allontani eccessivamente dalla diagonale:

$$|i_k - j_k| \leq \omega, \text{ con } \omega > 0 \text{ (ampiezza della finestra);}$$

4. *Slope constraint*. Il numero di *warping path* ammissibili può essere contenuto imponendo un limite di pendenza, in questo modo si evita che una parte piccola di una serie venga allineata ad una parte lunga dell'altra;

5. Condizioni al contorno. Limitano il campo di ricerca garantendo che i *warping path* seguano una direzione diagonale. Tipicamente si impone che il primo elemento di W sia il primo elemento della diagonale in basso a sinistra e l'ultimo in alto a destra:

$$i_1 = 1, j_1 = 1, i_n = n \text{ e } j_m = m.$$

La distanza tra le serie viene computata come la minima distanza totale cumulata su tutti i potenziali W individuati:

$$DTW(S_1, S_2) = \min_w \left[\sum_{k=1}^p \delta(w_k) \right]$$

dove $\delta(w_k)$ corrisponde al k -esimo elemento di W e δ è la misura di distanza scelta. In questo problema si possono definire tre variabili:

- *Variabile stage*, corrisponde al tempo e impone un ordine monotono sugli eventi;
- *Variabile state*, sono i singoli punti sulla griglia;
- *Variabile decisionale*, meno semplici da individuare, corrispondono alle restrizioni sui percorsi ammissibili tra due punti sulla griglia.

L'algoritmo per il calcolo della distanza cumulata può essere riscritto secondo la seguente relazione iterativa:

$$\gamma(i, j) = \delta(i, j) + \min[\gamma(i - 1, j), \gamma(i - 1, j - 1), \gamma(i, j - 1)]$$

La distanza cumulata $\gamma(i, j)$ è ottenuta sommando alla distanza tra gli elementi allineati (generico w_k) la distanza minima cumulata dei punti vicini. L'algoritmo è simmetrico in quanto si considerano entrambi i punti al di fuori della diagonale della griglia, si avrebbe asimmetria se si considerasse solamente uno dei punti, $(i - 1, j)$ oppure $(i, j - 1)$ (Berndt e Clifford, 1994). Nell'esempio delle serie $x[t]$ e $y[t]$ (Figura 13) si può osservare come il DTW determini le coppie di elementi delle due sequenze che garantiscono l'allineamento migliore e che definiscono il *warping path* (W) ottimale sulla griglia. A ogni w_k corrisponde quindi una coppia di osservazioni di $x[t]$ e di $y[t]$ che consentono di ridefinire quest'ultime in un nuovo spazio temporale in cui risultano essere allineate e con lo stesso numero di osservazioni. In questo esempio le due sequenze risultanti sono $x'[t] = (1, 1, 1, 2, 3, 2, 0)$ e $y'[t] = (0, 1, 1, 2, 3, 2, 1)$, si può notare che per allineare i due segnali l'algoritmo ripete due volte la prima osservazione di $x[t]$.

In Figura 14 si può osservare graficamente come l'algoritmo allinei su un nuovo spazio temporale distorto i due segnali originariamente non sincronizzati.

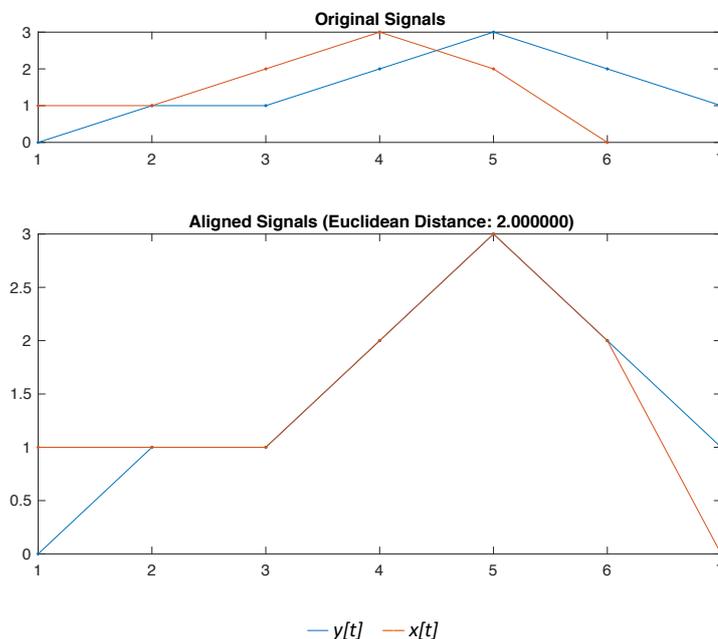


Figura 14
Segnali $x[t]$ e $y[t]$
originali e allineati
mediante l'algoritmo
DTW.

Fonte: MATLAB R2017B

Il DTW ricostruisce e allinea le due serie considerando in ordine ciascun elemento del *warping path* ottimale e i valori dei segnali ad esso associati. La distanza tra le due serie $x[t]$ e $y[t]$,

calcolata applicando la magnitudo della differenza per ogni coppia di valori dei segnali, risulta essere pari a 2, esattamente uguale alla somma dei valori del *warping path* ottimale ($W = 1,0,0,0,0,0,1$), come mostrato in Figura 14.

3.4.3 Adaptive Dissimilarity Index

Il *Dynamic Time Warping*, similmente alla distanza Euclidea, ignora la struttura temporale delle serie storiche poiché la distanza misurata si basa esclusivamente sulla differenza tra i valori ($\delta(i, j) = |u_i - v_j|$ oppure $\delta(i, j) = (u_i - v_j)^2$), indipendentemente dal comportamento di crescita o decrescita attorno a tali valori (Douzal e Naidu, 2007). Per queste ragioni Douzal e Naidu (2007) hanno cercato di definire un indice di dissimilarità (*dissimilarity index*) che contenga sia la misura convenzionale di distanza tra i valori sia la correlazione temporale tra le serie in termini di andamento. La similarità rispetto all'andamento viene definita facendo riferimento a due aspetti: la robustezza della monotonia (*strength of the monotonicity*) e la vicinanza dei tassi di crescita (*closeness of the growth rates*). Due serie storiche sono dunque simili rispetto al loro andamento se in ogni intervallo di osservazioni $[t_i, t_{i+1}]$ crescono o decrescono simultaneamente (monotonia) con lo stesso tasso di crescita (vicinanza dei tassi di crescita).

Per prima cosa, viene definita una misura di correlazione temporale in termini di andamento (*temporal correlation for behavior proximity measure*) partendo dalla definizione del classico coefficiente di correlazione e includendo in esso l'informazione di interdipendenza delle osservazioni delle serie (attraverso una relazione temporale di *neighboring* del primo ordine). Date due serie storiche $S_1 = (u_1, \dots, u_n)$ e $S_2 = (v_1, \dots, v_n)$ di n valori osservati negli istanti di tempo $[t_1, \dots, t_n]$, il coefficiente di correlazione temporale tra S_1 e S_2 viene definito in questo modo¹¹:

$$CORT(S_1, S_2) = \frac{\sum_{i=1}^{n-1} (u_{(i+1)} - u_i)(v_{(i+1)} - v_i)}{\sqrt{\sum_{i=1}^{n-1} (u_{(i+1)} - u_i)^2} \sqrt{\sum_{i=1}^{n-1} (v_{(i+1)} - v_i)^2}}$$

¹¹ Per la dimostrazione analitica della formula si rimanda a Douzal Chouakria, A. e Naidu Nagabhushan, P. (2007). Adaptive dissimilarity index for measuring time series proximity. *Advances in Data Analysis and Classification*, Vol. 1, pp. 9 e 10.

dove $CORT(S_1, S_2)$ è definito nell'intervallo $[-1, 1]$. Se $CORT(S_1, S_2) = 1$, in ogni intervallo di tempo le serie crescono o decrescono simultaneamente con lo stesso tasso di crescita (stesso andamento). Se $CORT(S_1, S_2) = -1$, in ogni intervallo di tempo se S_1 cresce allora S_2 decresce (o viceversa) con lo stesso tasso di crescita (andamenti opposti). Infine, se $CORT(S_1, S_2) = 0$ non c'è monotonia tra le serie e i loro tassi di crescita sono stocasticamente linearmente indipendenti, ossia hanno andamenti diversi (Douzal e Naidu, 2007).

L'indice di dissimilarità (D) viene dunque definito sulla base della misura convenzionale di distanza tra i valori (δ_{conv} – ad esempio distanza Euclidea o DTW), sulla base del coefficiente di correlazione temporale ($CORT$) e di una funzione adattiva modulante (f):

$$D(S_1, S_2) = f(CORT(S_1, S_2)) \cdot \delta_{conv}(S_1, S_2)$$

dove $f(x)$ è una funzione adattiva modulante esponenziale così definita:

$$f(x) = \frac{2}{1 + \exp(kx)}$$

con $k \geq 0$. Nel caso in cui le serie abbiano andamenti diversi ($CORT \rightarrow 0$), $f(x)$ tende ad 1 qualunque sia il valore attribuito a k e D è esattamente pari a δ_{conv} . Nel caso in cui le serie abbiano invece andamenti simili oppure opposti ($|CORT| \rightarrow 1$), il parametro k modula i contributi della similarità in D sia rispetto ai valori sia rispetto all'andamento. Infatti, se $k = 0$, la similarità rispetto all'andamento non ha nessun contributo su D che coincide totalmente con la distanza convenzionale δ_{conv} (contributo della similarità tra i valori). Se $k > 0$ invece, il contributo della similarità rispetto all'andamento cresce al crescere di k , mentre il contributo della similarità rispetto ai valori decresce. In Figura 15 è possibile osservare l'andamento della funzione adattiva modulante rispetto al coefficiente di correlazione temporale. Si può notare come in effetti la funzione adattiva modulante vada ad amplificare la distanza tra due generiche serie storiche se queste tendono ad avere andamenti opposti, viceversa la funzione riduce la loro distanza se queste hanno andamenti simili e infine non ha nessun effetto sulla distanza se le serie hanno andamenti indipendenti.

L'indice proposto da Douzal e Naidu risulta molto interessante e ha dimostrato dei risultati migliori in termini di *cluster-analysis* rispetto alla distanza Euclidea o al *Dynamic Time Warping*. Data però l'ipotesi di partenza per cui due generiche serie presentano lo stesso numero di osservazioni (ipotesi necessaria per il calcolo del coefficiente di correlazione temporale),

non è possibile utilizzare questo indice per misurare la distanza tra le serie descrittive il fenomeno di diffusione delle certificazioni ISO 9000 nei Paesi europei. Per quest'ultime è quindi opportuno utilizzare il *Dynamic Time Warping* con la consapevolezza dei limiti e delle approssimazioni del metodo.

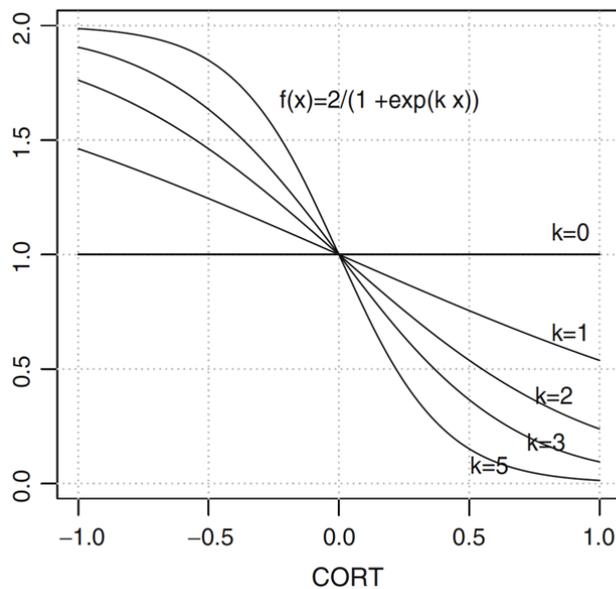


Figura 15
Effetto adattivo
modulante

Fonte: Douzal e Naidu (2007)

3.4.4 Calcolo delle distanze DTW in Matlab

In ambiente *MATLAB* è possibile calcolare la distanza DTW tra due serie storiche, definite in ingresso come vettori, attraverso la funzione `dtw` (The MathWorks, Inc., 2017). Se si considerano ad esempio le serie relative all'Austria e alla Bulgaria e si definiscono per questi Paesi i vettori *A* e *B* del numero normalizzato di ISO 9000 rispetto al valore massimo, la funzione `dtw` si implementa digitando nella *Command Window* di *MATLAB*:

```
>> [dist,a,b] = dtw(A,B);
```

In output vengono visualizzati due grafici: il primo mostra i segnali originali, il secondo i segnali allineati (vedi Figura 16). Dal primo grafico si può osservare come i segnali risultino essere disallineati dal punto di vista temporale per effetto del diverso numero di osservazioni nell'orizzonte 1993-2016 (ventiquattro osservazioni nel caso dell'Austria e ventidue nel caso della Bulgaria). Nel secondo grafico si può osservare il risultato dell'implementazione del DTW: l'algoritmo mantiene per entrambe le serie la stessa origine ($t = 1$) e le dilata su un orizzonte temporale maggiore, definendo quindi un nuovo spazio temporale. È possibile notare come

gli andamenti originari di crescita, stazionarietà oppure oscillazione vengano comunque mantenuti e allineati. La dilatazione dei segnali nel tempo è dovuta al fatto che questi vengono ridefiniti sulla base del *warping path* ottimale individuato con opportune ripetizioni delle osservazioni dei segnali originari. Per poter comprendere come vengono definite le nuove sequenze allineate, la funzione `dtw` in *MATLAB* fornisce in output due vettori, definiti arbitrariamente a e b, contenenti una sequenza monotona crescente i cui valori corrispondono agli indici relativi alle osservazioni di A e B:

```
>> [dist,a,b] = dtw(A,B)
```

```
a = [1, 1, 1, 1, 1, 1, 2, 2, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13,  
     14, 15, 16, 17, 18, 19, 20, 20, 21, 22, 23, 24]
```

```
b = [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 13, 13, 13, 13, 13,  
     13, 13, 13, 14, 15, 15, 16, 17, 18, 19, 20, 21, 21, 22]
```

Questi vettori consentono di individuare il *warping path* ottimale sulla griglia (vedi Tabella 9). Nel caso dell’Austria, ad esempio, il segnale risultante presenta la prima osservazione (numero normalizzato di ISO 9000 nel 1993) ripetuta per sei volte consecutive, la seconda osservazione (numero normalizzato di ISO 9000 nel 1994) per tre volte consecutive e la ventesima osservazione (numero normalizzato di ISO 9000 nel 2012) per due volte. Complessivamente i due segnali allineati vengono definiti su un nuovo orizzonte temporale di 32 osservazioni (vedi Tabella 9 e Figura 16).

La funzione `dtw` fornisce inoltre in output la variabile arbitrariamente definita `dist` che corrisponde al valore della distanza DTW calcolata tra i segnali, ossia la minima distanza cumulata tra tutti i *working path*:

```
dist = 1.4448
```

Di default, la funzione `dtw` utilizza la distanza Euclidea, coincidente con la magnitudo della differenza tra due valori. In alternativa, è possibile utilizzare il quadrato della differenza:

```
>> dtw(A,B, 'squared')
```

In questo caso il grafico non cambia ma la distanza è pari a:

```
dist = 0.1470.
```

| | | | | | | | | | | | | | | | | | | | | | | | | | |
|-------|---------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
| XXII | 0,95 | 0,91 | 0,87 | 0,73 | 0,60 | 0,44 | 0,32 | 0,29 | 0,21 | 0,18 | 0,16 | 0,41 | 0,32 | 0,30 | 0,22 | 0,14 | 0,12 | 0,12 | 0,05 | 0,15 | 0,07 | 0,05 | 0,14 | 0,09 | 0,19 |
| XXI | 0,87 | 0,83 | 0,79 | 0,65 | 0,52 | 0,36 | 0,24 | 0,21 | 0,13 | 0,10 | 0,08 | 0,33 | 0,24 | 0,22 | 0,13 | 0,06 | 0,04 | 0,04 | 0,13 | 0,07 | 0,01 | 0,03 | 0,05 | 0,00 | 0,11 |
| XX | 0,92 | 0,88 | 0,83 | 0,70 | 0,56 | 0,41 | 0,29 | 0,25 | 0,18 | 0,14 | 0,12 | 0,37 | 0,29 | 0,26 | 0,18 | 0,10 | 0,09 | 0,09 | 0,08 | 0,12 | 0,03 | 0,02 | 0,10 | 0,05 | 0,16 |
| XIX | 0,86 | 0,82 | 0,78 | 0,64 | 0,51 | 0,35 | 0,23 | 0,20 | 0,12 | 0,09 | 0,07 | 0,32 | 0,23 | 0,21 | 0,12 | 0,05 | 0,03 | 0,03 | 0,14 | 0,06 | 0,02 | 0,04 | 0,04 | 0,01 | 0,10 |
| XVIII | 0,97 | 0,93 | 0,88 | 0,75 | 0,61 | 0,46 | 0,34 | 0,30 | 0,22 | 0,19 | 0,17 | 0,42 | 0,33 | 0,31 | 0,23 | 0,15 | 0,14 | 0,14 | 0,03 | 0,16 | 0,08 | 0,07 | 0,15 | 0,10 | 0,21 |
| XVII | 0,80 | 0,76 | 0,72 | 0,58 | 0,45 | 0,29 | 0,17 | 0,14 | 0,06 | 0,03 | 0,01 | 0,26 | 0,17 | 0,15 | 0,06 | 0,01 | 0,03 | 0,03 | 0,20 | 0,00 | 0,08 | 0,10 | 0,02 | 0,07 | 0,04 |
| XVI | 1,00 | 0,96 | 0,92 | 0,78 | 0,65 | 0,49 | 0,37 | 0,34 | 0,26 | 0,22 | 0,21 | 0,46 | 0,37 | 0,35 | 0,26 | 0,19 | 0,17 | 0,17 | 0,00 | 0,20 | 0,12 | 0,10 | 0,18 | 0,13 | 0,24 |
| XV | 0,85 | 0,81 | 0,77 | 0,63 | 0,50 | 0,34 | 0,22 | 0,19 | 0,11 | 0,08 | 0,06 | 0,31 | 0,22 | 0,20 | 0,11 | 0,04 | 0,02 | 0,02 | 0,15 | 0,05 | 0,03 | 0,05 | 0,04 | 0,01 | 0,09 |
| XIV | 0,85 | 0,81 | 0,77 | 0,63 | 0,50 | 0,34 | 0,22 | 0,19 | 0,11 | 0,08 | 0,06 | 0,31 | 0,22 | 0,20 | 0,11 | 0,04 | 0,02 | 0,02 | 0,15 | 0,05 | 0,03 | 0,05 | 0,04 | 0,01 | 0,09 |
| XIII | 0,75 | 0,71 | 0,66 | 0,53 | 0,39 | 0,24 | 0,12 | 0,08 | 0,03 | 0,05 | 0,05 | 0,20 | 0,11 | 0,09 | 0,01 | 0,01 | 0,08 | 0,08 | 0,25 | 0,06 | 0,14 | 0,15 | 0,07 | 0,12 | 0,01 |
| XII | 0,50 | 0,46 | 0,41 | 0,28 | 0,14 | 0,01 | 0,13 | 0,17 | 0,25 | 0,28 | 0,30 | 0,05 | 0,14 | 0,16 | 0,24 | 0,32 | 0,33 | 0,33 | 0,50 | 0,31 | 0,39 | 0,40 | 0,32 | 0,37 | 0,26 |
| XI | 0,36 | 0,32 | 0,27 | 0,14 | 0,00 | 0,15 | 0,27 | 0,31 | 0,39 | 0,42 | 0,44 | 0,19 | 0,28 | 0,30 | 0,38 | 0,46 | 0,47 | 0,47 | 0,64 | 0,45 | 0,53 | 0,54 | 0,46 | 0,51 | 0,40 |
| X | 0,27 | 0,23 | 0,19 | 0,05 | 0,08 | 0,24 | 0,36 | 0,39 | 0,47 | 0,51 | 0,52 | 0,27 | 0,36 | 0,38 | 0,47 | 0,54 | 0,56 | 0,56 | 0,73 | 0,53 | 0,61 | 0,63 | 0,55 | 0,60 | 0,49 |
| IX | 0,13 | 0,10 | 0,05 | 0,08 | 0,22 | 0,37 | 0,49 | 0,53 | 0,61 | 0,64 | 0,66 | 0,41 | 0,50 | 0,52 | 0,60 | 0,68 | 0,69 | 0,69 | 0,87 | 0,67 | 0,75 | 0,76 | 0,68 | 0,73 | 0,63 |
| VIII | 0,10 | 0,06 | 0,02 | 0,12 | 0,25 | 0,41 | 0,53 | 0,56 | 0,64 | 0,67 | 0,69 | 0,44 | 0,53 | 0,55 | 0,64 | 0,71 | 0,73 | 0,73 | 0,90 | 0,70 | 0,78 | 0,80 | 0,72 | 0,77 | 0,66 |
| VII | 0,08 | 0,04 | 0,01 | 0,14 | 0,28 | 0,43 | 0,55 | 0,59 | 0,67 | 0,70 | 0,72 | 0,47 | 0,56 | 0,58 | 0,66 | 0,74 | 0,75 | 0,75 | 0,92 | 0,73 | 0,81 | 0,82 | 0,74 | 0,79 | 0,68 |
| VI | 0,04 | 0,00 | 0,04 | 0,18 | 0,31 | 0,47 | 0,59 | 0,62 | 0,70 | 0,73 | 0,75 | 0,50 | 0,59 | 0,61 | 0,70 | 0,77 | 0,79 | 0,79 | 0,96 | 0,76 | 0,84 | 0,86 | 0,77 | 0,82 | 0,72 |
| V | 0,03 | 0,01 | 0,05 | 0,19 | 0,32 | 0,48 | 0,60 | 0,63 | 0,71 | 0,74 | 0,76 | 0,51 | 0,60 | 0,62 | 0,71 | 0,78 | 0,80 | 0,80 | 0,97 | 0,77 | 0,85 | 0,87 | 0,78 | 0,83 | 0,73 |
| IV | 0,02 | 0,02 | 0,07 | 0,20 | 0,34 | 0,49 | 0,61 | 0,65 | 0,73 | 0,76 | 0,78 | 0,53 | 0,62 | 0,64 | 0,72 | 0,80 | 0,81 | 0,81 | 0,98 | 0,79 | 0,87 | 0,88 | 0,80 | 0,85 | 0,74 |
| III | 0,01 | 0,03 | 0,08 | 0,21 | 0,35 | 0,50 | 0,62 | 0,66 | 0,73 | 0,77 | 0,79 | 0,54 | 0,62 | 0,65 | 0,73 | 0,81 | 0,82 | 0,82 | 0,99 | 0,80 | 0,88 | 0,89 | 0,81 | 0,86 | 0,75 |
| II | 0,002 | 0,04 | 0,08 | 0,22 | 0,35 | 0,51 | 0,63 | 0,66 | 0,74 | 0,77 | 0,79 | 0,54 | 0,63 | 0,65 | 0,74 | 0,81 | 0,83 | 0,83 | 1,00 | 0,80 | 0,88 | 0,90 | 0,81 | 0,86 | 0,76 |
| I | 0,0005 | 0,04 | 0,08 | 0,22 | 0,35 | 0,51 | 0,63 | 0,66 | 0,74 | 0,77 | 0,79 | 0,54 | 0,63 | 0,65 | 0,74 | 0,81 | 0,83 | 0,83 | 1,00 | 0,80 | 0,88 | 0,90 | 0,82 | 0,87 | 0,76 |
| | | 0,04 | 0,08 | 0,22 | 0,35 | 0,51 | 0,63 | 0,66 | 0,74 | 0,78 | 0,79 | 0,54 | 0,63 | 0,65 | 0,74 | 0,81 | 0,83 | 0,83 | 1,00 | 0,80 | 0,88 | 0,90 | 0,82 | 0,87 | 0,76 |
| | | I | II | III | IV | V | VI | VII | VIII | IX | X | XI | XII | XIII | XIV | XV | XVI | XVII | XVIII | XIX | XX | XXI | XXII | XXIII | XXIV |
| | | Austria | | | | | | | | | | | | | | | | | | | | | | | |

Note: La tabella mostra la griglia risultante dall'implementazione del DTW tra le serie storiche A (Austria) e B (Bulgaria). Le osservazioni si riferiscono al numero normalizzato di certificazioni ISO 9000 nell'orizzonte temporale 1993-2016. Il numero di decimali dopo la virgola è stato ridotto a due per motivi di spazio. La prima osservazione per entrambe le serie si trova in basso a sinistra, l'ultima in alto a destra. La misura di distanza utilizzata per il calcolo degli elementi della griglia è la magnitudo della differenza. In blu viene evidenziato il *warping path* ottimale.

Tabella 9
Griglia risultante tra
Austria e Bulgaria con
warping path ottimale.

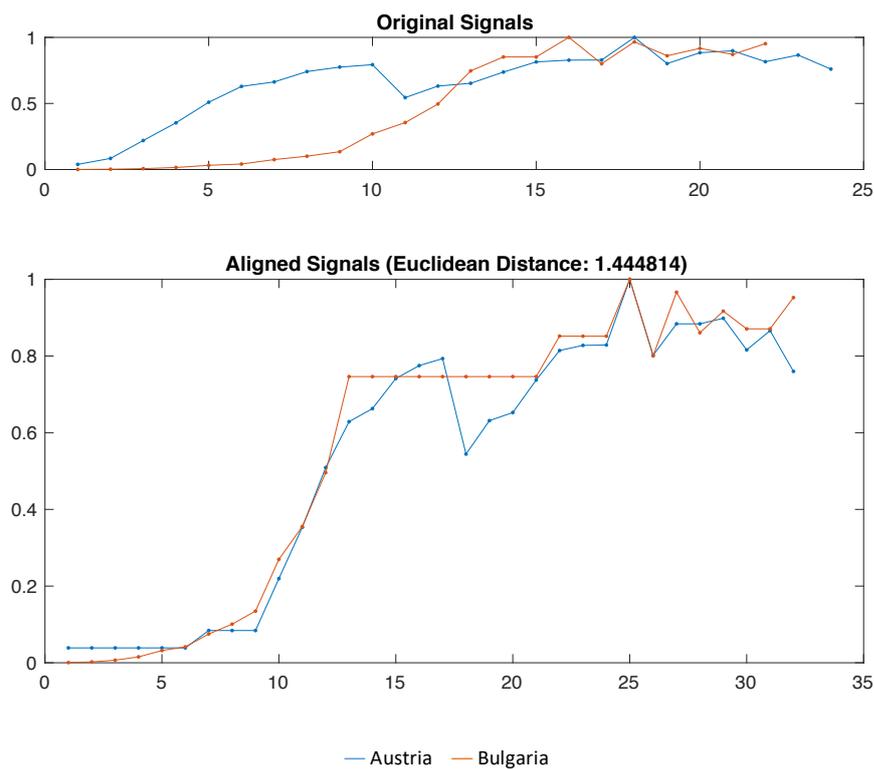


Figura 16
 Segnali originali e
 allineati relativi ad
 Austria e Bulgaria

Fonte: MATLAB R2017B

Per poter calcolare la distanza tra tutti i Paesi europei del campione di analisi in coppie (30 Paesi con 435 possibili combinazioni in coppie), viene definita in ingresso la matrice X in cui ciascuna riga si riferisce ad un Paese, mentre le colonne corrispondono agli anni dell'intero orizzonte temporale analizzato (1993-2016) e contengono per ciascuna riga il corrispondente numero normalizzato di certificazioni ISO 9000 per quell'anno. Come già osservato in precedenza, il numero di osservazioni nell'orizzonte temporale varia da Paese a Paese in quanto per alcuni il fenomeno di diffusione delle ISO 9000 è posteriore rispetto ad altri. Questo comporta che per alcune righe della matrice X le prime colonne contigue risultino essere prive di dati, rendendo di fatto impossibile implementare la funzione dtw tra le righe, ossia per ciascuna coppia di Paesi. Per ovviare a questo problema è stata definita in *MATLAB* la funzione `shiftvector` (vedi Appendice), in grado di ridimensionare le righe della matrice in vettori di dimensione inferiore, laddove queste abbiano un numero minore di osservazioni (come nel caso della Bulgaria attraverso il vettore B nell'esempio precedentemente analizzato).

Allo stesso modo, per poter implementare la funzione dtw per tutte le coppie di Paesi, è stata definita la funzione `dtwmatrix` (vedi Appendice) che calcola direttamente la distanza DTW

tra le righe di X (opportunamente ridimensionate attraverso la funzione `shiftvector`) fornendo in output la matrice delle distanze `Ddist`, in cui sia le righe sia le colonne corrispondono ai Paesi europei del campione e l'elemento generico d_{ij} corrisponde alla distanza DTW tra il Paese sulla riga i e il Paese sulla colonna j .

Una volta ottenuta la matrice delle distanze questa deve essere trasformata nel vettore riga D richiesto in ingresso dalla funzione `Linkage`, questo vettore contiene le distanze relative a ciascuna coppia di Paesi (D ha dimensione 1×435 nel caso di 30 Paesi). A tale scopo è stata definita la funzione `transformdist` (vedi Appendice) che trasforma la matrice triangolare inferiore di `Ddist` nel vettore delle distanze D .

Definite queste funzioni accessorie, è possibile implementare gli algoritmi di *clustering* gerarchico agglomerativo come spiegati nel paragrafo 3.2 di questo capitolo¹².

3.5 Dendrogramma dei Paesi europei

La scelta legata alla metrica più adatta da adottare attraverso la funzione `dtwmatrix` per il calcolo delle distanze DTW tra i Paesi (magnitudo o quadrato della differenza), può essere eseguita considerando tutte le possibili combinazioni con le diverse tecniche di *clustering* implementabili attraverso la funzione `Linkage`, calcolando quindi per ciascuna combinazione il *cophenetic correlation coefficient*. La combinazione per la quale il coefficiente si avvicina maggiormente a 1 definisce la metrica di calcolo delle distanze e il metodo di *clustering* gerarchico più adatti per i dati a disposizione. Come già anticipato, la funzione `Linkage` utilizza la distanza tra i Paesi come informazione di partenza con cui costruire i cluster, a sua volta però deve determinare la distanza tra i cluster che crea, a tale scopo esistono diverse tecniche¹³:

- *Single-Linkage*, nota anche come tecnica *nearest-neighbor*. La distanza tra due gruppi viene definita come la distanza tra i due elementi più vicini appartenenti ai due gruppi.

¹² Vedi pp. 32 e ss. del cap. 3 di questo studio.

¹³ Oltre a quelle elencate si possono aggiungere le tecniche *Centroid* (UPGMC), *Median* (WPGMC) e *Ward* (*minimum variance algorithm*) appropriate solo nel caso della distanza Euclidea come misura di distanza tra gli elementi del data set di partenza, quindi non applicabili nel caso del *Dynamic Time Warping*.

- *Complete-linkage*, nota anche come tecnica *furthest-neighbor*. La distanza tra due gruppi viene definita come la distanza tra i due elementi più lontani appartenenti ai due gruppi.
- *Simple-average*, nota anche come *weighted pair-group method* (WPGMA). Questo algoritmo definisce la distanza tra due gruppi come la distanza media pesata tra i loro membri.
- *Group-average*, nota anche come *unweighted pair-group method* (UPGMA). Attraverso questa tecnica la distanza tra due gruppi viene definita come la distanza media tra ciascuno dei loro membri.

In Tabella 10 sono riportati i valori del *cophenetic correlation coefficient* per ciascuna combinazione. Dall'implementazione in *MATLAB*, il valore del coefficiente più elevato si ottiene nel caso della magnitudo della differenza e della tecnica *Group-average* (UPGMA), *c* è pari a 0.8044. Per ottenere questo valore si digita in *MATLAB*:

```
>> Ddist = dtwmatrix(X);
>> D = transformdist(D);
>> Z = linkage(D, 'average');
>> c = cophenet(Z,D);
```

L'applicazione della magnitudo della differenza è contenuta nella funzione `dtwmatrix` (vedi Appendice) con cui si ottiene la matrice delle distanze `Ddist`. La tecnica *Group-average* invece si implementa specificandola nell'argomento della funzione `linkage`. La verifica delle distanze è quindi automaticamente eseguita poiché si può considerare il valore di *c* sufficientemente vicino a 1 per poter affermare che i collegamenti creati da `linkage` attraverso la tecnica *Group-average* riflettono bene le distanze tra i Paesi ottenute con il *Dynamic Time Warping*. La Tabella 11 presenta i valori delle distanze DTW calcolate tra i principali Paesi europei in termini di PIL e consente alcune osservazioni preliminari: Paesi come Francia, Germania e Italia sembrano avere schemi di diffusione delle certificazioni ISO 9000 simili, hanno infatti una distanza tra loro inferiore a 1.4, la Russia e il Regno Unito sembrano invece essere molto distanti rispetto agli altri, inoltre è notevole la distanza proprio tra questi due Paesi (con un valore superiore a 11), sembra quindi che essi abbiano schemi di diffusione ben diversi.

| Metrica (DTW) | Tecnica (linkage) | Cophenetic correlation coefficient (c) |
|----------------------------|-------------------|--|
| Magnitudo della differenza | Single-linkage | 0.7474 |
| | Complete-linkage | 0.7535 |
| | Simple-average | 0.7510 |
| | Group-average | 0.8044 |
| Quadrato della differenza | Single-linkage | 0.7337 |
| | Complete-linkage | 0.7533 |
| | Simple-average | 0.7706 |
| | Group-average | 0.7844 |

Tabella 10
 Combinazioni di
 Calcolo delle distanze
 e dei collegamenti
 tra i Paesi

| | Francia | Germania | Italia | Russia | Spagna | UK |
|----------|---------|----------|--------|---------|--------|----|
| Francia | 0 | | | | | |
| Germania | 1.3119 | 0 | | | | |
| Italia | 1.2302 | 1.3662 | 0 | | | |
| Russia | 6.5725 | 6.8320 | 5.9362 | 0 | | |
| Spagna | 2.1651 | 2.6221 | 2.2652 | 3.4507 | 0 | |
| UK | 4.1867 | 5.4015 | 6.8543 | 11.1284 | 4.0560 | 0 |

Tabella 11
 Matrice delle
 distanze per i sei
 Paesi europei con
 maggior PIL

Una volta trasformata D_{dist} in D , la matrice Z è ottenuta implementando in *MATLAB* la funzione `linkage` e specificando la tecnica *Group-average*:

```
>> Z = linkage(D, 'average')
```

In Tabella 12 viene riportato l'output della funzione. Z ha dimensione 29×3 e ciascuna riga definisce un collegamento tra Paesi o cluster, le prime due colonne identificano gli oggetti collegati mentre la terza colonna contiene l'informazione relativa alla distanza tra questi oggetti. Per il campione di Paesi analizzato la funzione `linkage` inizia raggruppando l'Italia (14) e la Lituania (16) che presentano la minima distanza nella matrice D_{dist} (valore pari a 0.7233). La funzione continua raggruppando la Bulgaria (4) e la Croazia (5) con una distanza di 0.8009. Nella terza riga `linkage` crea un collegamento tra il Portogallo (20) e l'oggetto 31, in particolare quest'ultimo corrisponde al cluster binario formato raggruppando l'Italia e la Lituania (prima riga). Pertanto, quando vengono raggruppati due oggetti in un cluster binario la funzione assegna a quest'ultimo un indice partendo dal valore $m + 1$, dove m è il numero di Paesi del data set. Nel caso in analisi l'algoritmo parte quindi dal valore 31 e complessivamente crea 29 collegamenti ($m - 1$). Per agevolare la lettura della matrice Z , in Tabella 12 è stata inserita una colonna contenente l'indice di ciascun collegamento creato.

Attraverso la funzione dendrogram è possibile visualizzare graficamente i collegamenti definiti nella matrice Z attraverso un dendrogramma (vedi Figura 17). I numeri lungo l'asse orizzontale rappresentano gli indici dei Paesi, i loro collegamenti (numerati secondo gli indici definiti in Z) sono rappresentati invece attraverso delle U-rovesciate, l'altezza di ciascun collegamento identifica la distanza tra gli oggetti collegati e il valore della distanza è leggibile sull'asse verticale di sinistra.

| N° Collegamento | Colonna 1 | Colonna 2 | Colonna 3 |
|-----------------|-----------|-----------|-----------|
| 31 | 14 | 16 | 0.7233 |
| 32 | 4 | 5 | 0.8009 |
| 33 | 20 | 31 | 0.8435 |
| 34 | 1 | 26 | 0.9783 |
| 35 | 23 | 33 | 0.9788 |
| 36 | 19 | 32 | 0.9901 |
| 37 | 8 | 9 | 1.0408 |
| 38 | 35 | 36 | 1.1084 |
| 39 | 10 | 18 | 1.1325 |
| 40 | 21 | 38 | 1.2404 |
| 41 | 27 | 40 | 1.2811 |
| 42 | 25 | 28 | 1.2874 |
| 43 | 12 | 29 | 1.3076 |
| 44 | 37 | 39 | 1.3819 |
| 45 | 15 | 41 | 1.4991 |
| 46 | 34 | 44 | 1.5014 |
| 47 | 45 | 46 | 1.6012 |
| 48 | 13 | 30 | 1.6501 |
| 49 | 6 | 24 | 1.7271 |
| 50 | 11 | 47 | 1.7325 |
| 51 | 42 | 43 | 1.7620 |
| 52 | 3 | 49 | 2.1184 |
| 53 | 50 | 52 | 2.1972 |
| 54 | 7 | 48 | 2.3810 |
| 55 | 2 | 22 | 2.4049 |
| 56 | 17 | 54 | 2.6059 |
| 57 | 51 | 53 | 3.2034 |
| 58 | 56 | 57 | 4.1236 |
| 59 | 55 | 58 | 5.1761 |

Tabella 12
Matrice Z

| Numero Paese | | Numero Paese | | Numero Paese | |
|--------------|-------------|--------------|-------------|--------------|-------------|
| 1 | Austria | 11 | Grecia | 21 | Romania |
| 2 | Bielorussia | 12 | Ungheria | 22 | Russia |
| 3 | Belgio | 13 | Irlanda | 23 | Slovacchia |
| 4 | Bulgaria | 14 | Italia | 24 | Slovenia |
| 5 | Croazia | 15 | Lettonia | 25 | Spagna |
| 6 | Rep. Ceca | 16 | Lituania | 26 | Svezia |
| 7 | Danimarca | 17 | Paesi Bassi | 27 | Svizzera |
| 8 | Finlandia | 18 | Norvegia | 28 | Turchia |
| 9 | Francia | 19 | Polonia | 29 | Ucraina |
| 10 | Germania | 20 | Portogallo | 30 | Regno Unito |

Legenda
Numero Paese

In Figura 17 si può osservare un'elevata concentrazione di Paesi collegati in cluster nella parte in basso a sinistra del dendrogramma, è possibile ipotizzare che questi siano caratterizzati da un simile fenomeno di diffusione delle ISO 9000. A vista risulta però difficile identificare cluster chiari e disgiunti, non è quindi possibile specificare a priori il numero di cluster con cui suddividere i Paesi attraverso la funzione `cluster` e l'opzione `maxc lust`, per determinarli è necessario ricorrere al coefficiente d'inconsistenza come parametro di divisione naturale.

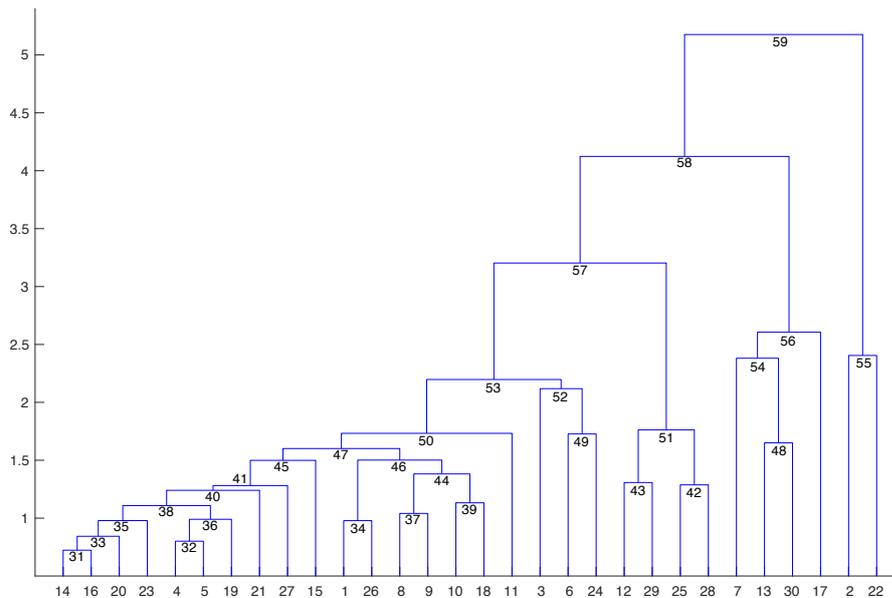


Figura 17
Dendrogramma dei Paesi europei con indici dei collegamenti

| Numero Paese | Numero Paese | Numero Paese |
|---------------|----------------|----------------|
| 1 Austria | 11 Grecia | 21 Romania |
| 2 Bielorussia | 12 Ungheria | 22 Russia |
| 3 Belgio | 13 Irlanda | 23 Slovacchia |
| 4 Bulgaria | 14 Italia | 24 Slovenia |
| 5 Croazia | 15 Lettonia | 25 Spagna |
| 6 Rep. Ceca | 16 Lituania | 26 Svezia |
| 7 Danimarca | 17 Paesi Bassi | 27 Svizzera |
| 8 Finlandia | 18 Norvegia | 28 Turchia |
| 9 Francia | 19 Polonia | 29 Ucraina |
| 10 Germania | 20 Portogallo | 30 Regno Unito |

Legenda
Numero Paese

3.6 Cluster degli schemi di diffusione per i Paesi europei

Come già spiegato nei paragrafi 3.2.4 e 3.2.5 di questo capitolo¹⁴, attraverso la funzione `inconsistent` è possibile quantificare il livello d'inconsistenza dei collegamenti creati attraverso l'*inconsistency coefficient*, che confronta l'altezza di ciascun collegamento con l'altezza

¹⁴ Vedi pp. 34 e ss. del cap. 3 di questo studio.

dei collegamenti a meno di due livelli sottostanti nel dendrogramma. Quest'informazione viene poi utilizzata dalla funzione `cluster` per suddividere i Paesi in cluster.

Se si implementa la funzione `inconsistent` si ottiene in output la matrice `I` di dimensione 29×4 , in cui le righe corrispondono ai collegamenti definiti nella matrice `Z` (vedi Tabella 13). Per agevolare la lettura di `I` è stata inserita in Tabella una colonna aggiuntiva contenente l'indice di ciascun collegamento. Si può osservare che la prima riga identifica il collegamento tra l'Italia (14) e la Lituania (16), a cui la funzione `linkage` assegna l'indice 31. Poiché entrambi gli oggetti sono nodi *leaf* alla base della gerarchia, il coefficiente d'inconsistenza per questo cluster è pari a zero. La seconda riga identifica il collegamento tra la Bulgaria (4) e la Croazia (5), anche in questo caso entrambi i Paesi sono nodi *leaf* quindi il coefficiente d'inconsistenza è pari a 0. La terza riga si riferisce invece al collegamento 33 che raggruppa il cluster binario contenente l'Italia e la Lituania con il Portogallo (20). In questo caso la terza colonna indica che i collegamenti inclusi nel calcolo sono 2: il collegamento 33 in questione e il collegamento sottostante nella gerarchia, ossia il 31. La prima colonna contiene la media delle altezze di questi due collegamenti (media delle distanze), la seconda colonna la deviazione standard tra i due collegamenti e infine la quarta colonna contiene il valore del coefficiente d'inconsistenza, pari a 0.7071, calcolato come differenza tra l'altezza del collegamento 33 e la media delle altezze dei collegamenti 31 e 33, normalizzata secondo la deviazione standard. Procedendo in questo modo vengono quindi analizzati tutti i collegamenti del dendrogramma definiti dalla funzione `linkage`.

A questo punto, per suddividere il dendrogramma in cluster si implementa la funzione `cluster`. Fissato un certo valore di taglio del coefficiente d'inconsistenza, la funzione analizza ciascun collegamento nel dendrogramma e mantiene intatti i collegamenti che presentano un valore del coefficiente inferiore al valore fissato, spezza invece tutti gli altri. In prima analisi si può osservare che il valore minimo assunto dai coefficienti d'inconsistenza per i vari collegamenti è pari a 0.7071, si può pensare pertanto di imporre un coefficiente di taglio pari a 0.71:

```
>> T = cluster(Z, 'cutoff', 0.71);
```

In output la funzione fornisce il vettore `T` di dimensione 1×30 , ciascun elemento rappresenta il numero di cluster a cui il corrispondente Paese viene assegnato. Complessivamente si ottengono 14 cluster.

| Numero Collegamento | Media altezze collegamenti inclusi | Deviazione standard collegamenti inclusi | Numero Collegamenti inclusi | Coefficiente d'inconsistenza |
|---------------------|------------------------------------|--|-----------------------------|------------------------------|
| 31 | 0,7233 | 0 | 1 | 0 |
| 32 | 0,8009 | 0 | 1 | 0 |
| 33 | 0,7834 | 0,0850 | 2 | 0,7071 |
| 34 | 0,9783 | 0 | 1 | 0 |
| 35 | 0,9112 | 0,0957 | 2 | 0,7071 |
| 36 | 0,8955 | 0,1338 | 2 | 0,7071 |
| 37 | 1,0408 | 0 | 1 | 0 |
| 38 | 1,0258 | 0,0718 | 3 | 1,1511 |
| 39 | 1,1325 | 0 | 1 | 0 |
| 40 | 1,1744 | 0,0933 | 2 | 0,7071 |
| 41 | 1,2608 | 0,0288 | 2 | 0,7071 |
| 42 | 1,2874 | 0 | 1 | 0 |
| 43 | 1,3076 | 0 | 1 | 0 |
| 44 | 1,1851 | 0,1765 | 3 | 1,1150 |
| 45 | 1,3901 | 0,1542 | 2 | 0,7071 |
| 46 | 1,2872 | 0,2741 | 3 | 0,7816 |
| 47 | 1,5339 | 0,0583 | 3 | 1,1545 |
| 48 | 1,6501 | 0 | 1 | 0 |
| 49 | 1,7271 | 0 | 1 | 0 |
| 50 | 1,6668 | 0,0928 | 2 | 0,7071 |
| 51 | 1,4524 | 0,2684 | 3 | 1,1539 |
| 52 | 1,9227 | 0,2767 | 2 | 0,7071 |
| 53 | 2,0160 | 0,2487 | 3 | 0,7286 |
| 54 | 2,0155 | 0,5168 | 2 | 0,7071 |
| 55 | 2,4049 | 0 | 1 | 0 |
| 56 | 2,4934 | 0,1590 | 2 | 0,7071 |
| 57 | 2,3876 | 0,7393 | 3 | 1,1035 |
| 58 | 3,3110 | 0,7646 | 3 | 1,0629 |
| 59 | 3,9016 | 1,3989 | 3 | 0,9111 |

Tabella 13
Matrice I

In Figura 18 i cluster sono messi in evidenza sul dendrogramma e in legenda viene indicato per ciascun Paese il cluster di appartenenza secondo il vettore T. In Figura 19, 20 e 21 sono riportati i diversi schemi di diffusione delle certificazioni ISO 9000 per ciascun cluster, l'ordine segue la disposizione dei Paesi sul dendrogramma. Da una prima analisi grafica, si può osservare che effettivamente la suddivisione in 14 cluster risulta essere ridondante, alcuni gruppi infatti presentano schemi di diffusione molto simili e sono quindi potenzialmente raggruppabili in un unico cluster. Se si considerano ad esempio i cluster I (Italia, Lituania, Portogallo, Slovacchia), II (Bulgaria, Croazia, Polonia), III (Romania), IV (Svizzera) e V (Lettonia), questi presentano schemi di diffusione analoghi e nel dendrogramma sono uniti complessivamente dal collegamento 45, è pertanto possibile ipotizzare di raggrupparli in un unico cluster. Allo stesso

modo, se si mantenesse intatto il collegamento 46 e i collegamenti sottostanti, si potrebbero unire i cluster VI (Austria, Svezia), VII (Finlandia, Francia) e VIII (Germania, Norvegia), così come per i cluster XI (Ungheria, Ucraina) e XII (Spagna, Turchia), in questo caso mantenendo intatto il collegamento 51 e i collegamenti ad esso sottostanti.

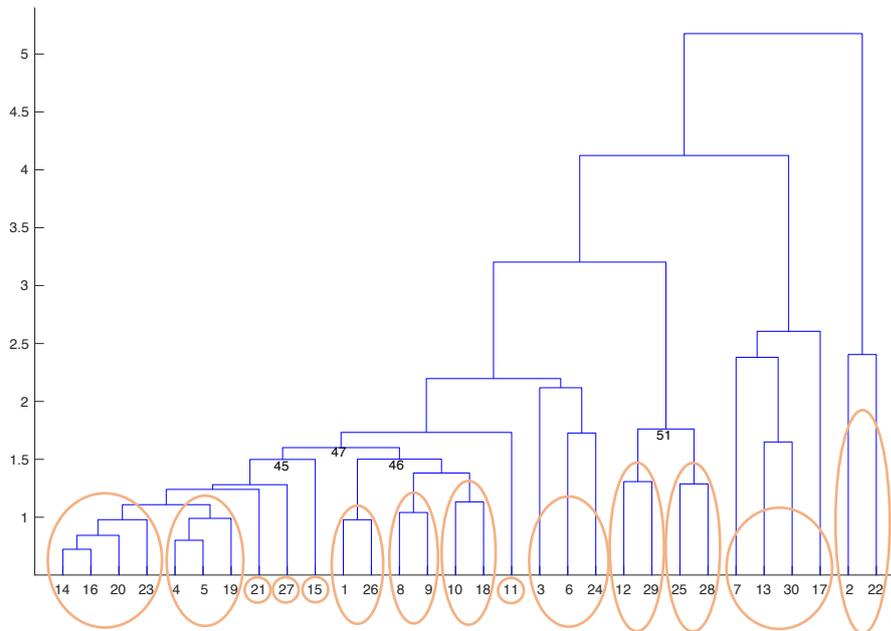


Figura 18
Dendrogramma con
suddivisione dei
Paesi
europei in 14 cluster

| Numero Paese | Cluster | Numero Paese | Cluster | Numero Paese | Cluster |
|---------------|---------|----------------|---------|---------------|---------|
| 1 Austria | 6 | 11 Grecia | 9 | 21 Romania | 3 |
| 2 Bielorussia | 14 | 12 Ungheria | 11 | 22 Russia | 14 |
| 3 Belgio | 10 | 13 Irlanda | 13 | 23 Slovacchia | 1 |
| 4 Bulgaria | 2 | 14 Italia | 1 | 24 Slovenia | 10 |
| 5 Croazia | 2 | 15 Lettonia | 5 | 25 Spagna | 12 |
| 6 Rep. Ceca | 10 | 16 Lituania | 1 | 26 Svezia | 6 |
| 7 Danimarca | 13 | 17 Paesi Bassi | 13 | 27 Svizzera | 4 |
| 8 Finlandia | 7 | 18 Norvegia | 8 | 28 Turchia | 12 |
| 9 Francia | 7 | 19 Polonia | 2 | 29 Ucraina | 11 |
| 10 Germania | 8 | 20 Portogallo | 1 | 30 UK | 13 |

Legenda
Numero Paese e
cluster di
appartenenza
secondo il vettore T

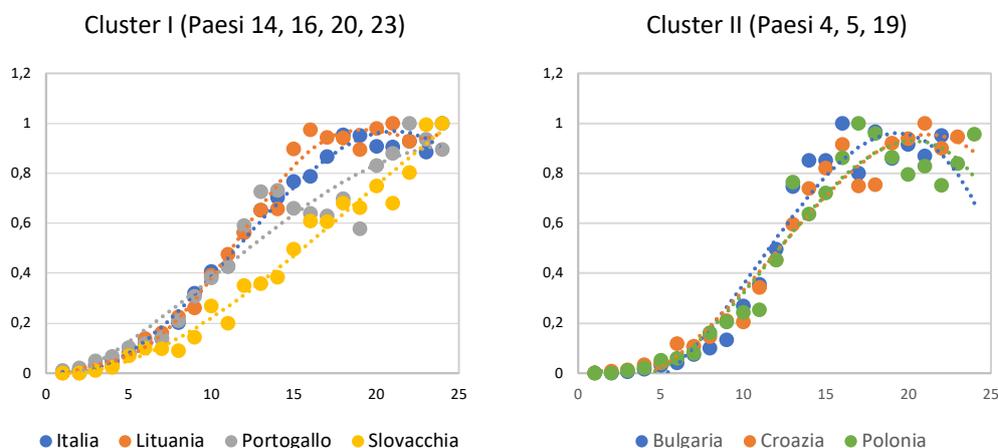


Figura 19
Cluster I e II

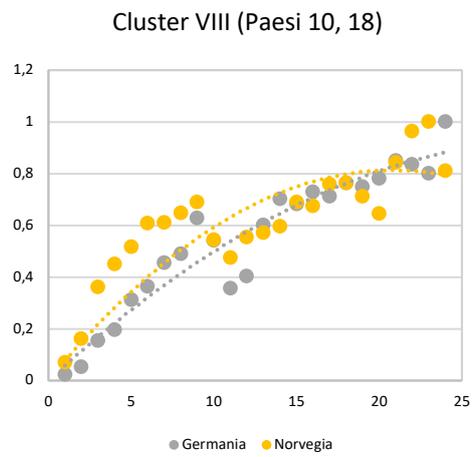
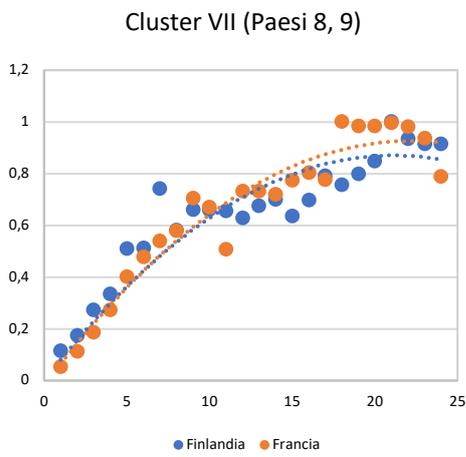
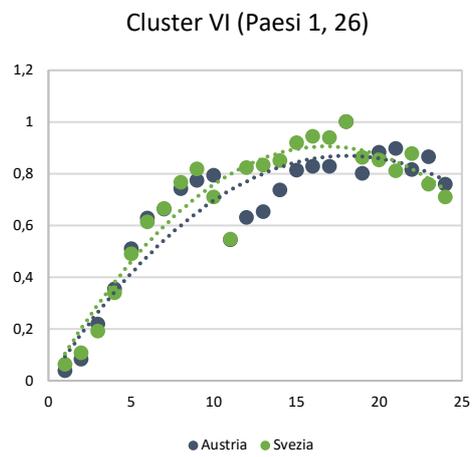
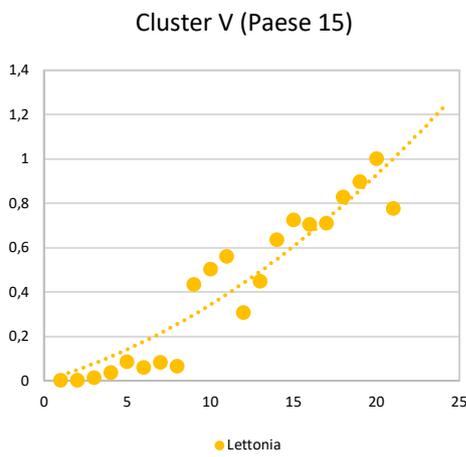
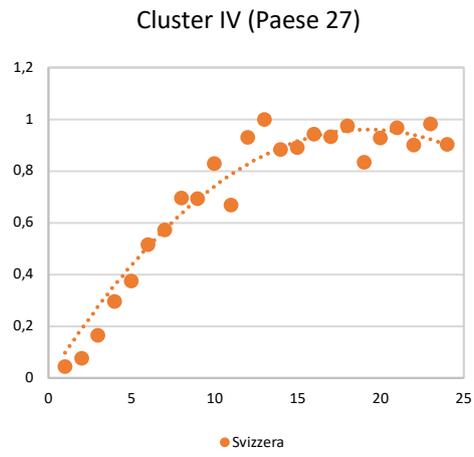
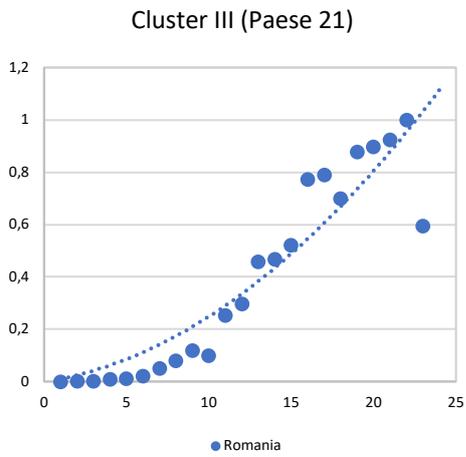


Figura 20
Cluster III,
IV, V, VI,
VII e VIII

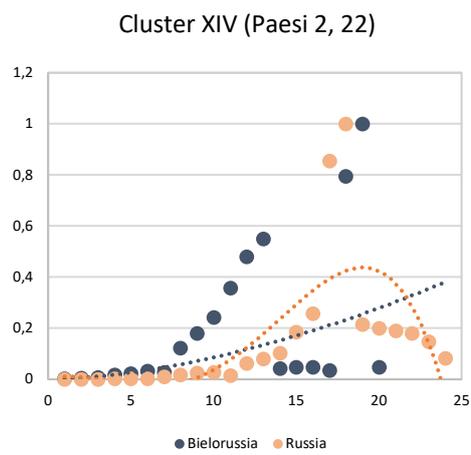
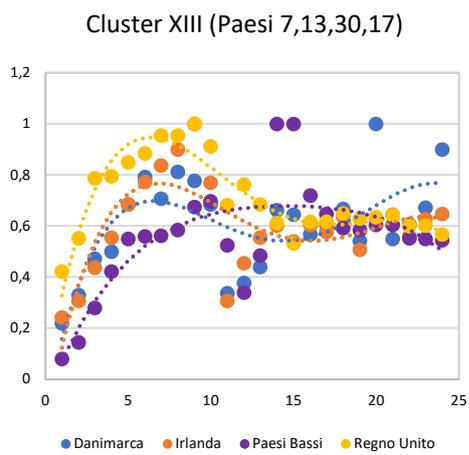
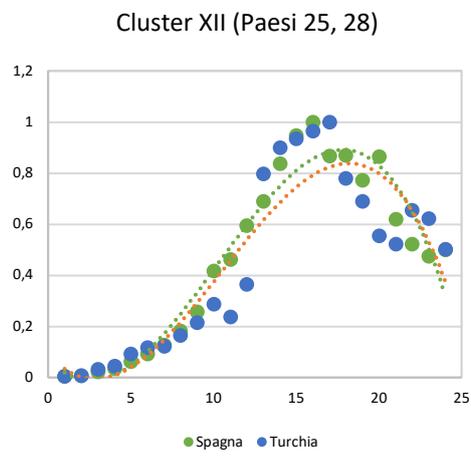
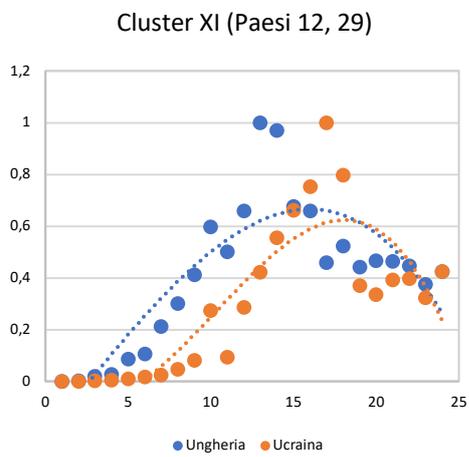
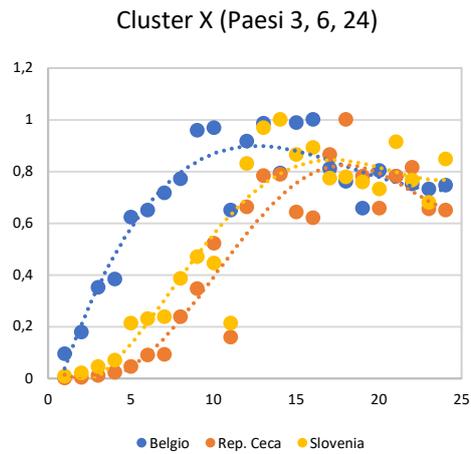
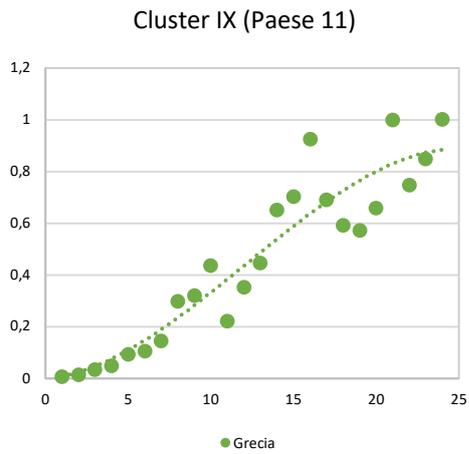


Figura 21
Cluster IX,
X, XI, XII,
XIII e XIV

Al contrario, i cluster X (Belgio, Repubblica Ceca, Slovenia), XIII (Danimarca, Irlanda, Regno Unito, Paesi Bassi) e XIV (Bielorussia, Russia) non sembrano avere andamenti simili agli altri, in prima analisi è quindi ipotizzabile che questi possano essere considerati come cluster ben definiti. Da notare infine il cluster IX, costituito dalla Grecia, che presenta uno schema di diffusione intermedio tra i cluster a sinistra nel dendrogramma. Anche la Grecia in prima analisi può essere considerata pertanto come un cluster indipendente.

Sulla base delle osservazioni fatte in precedenza, si può pensare di modificare il valore di taglio del coefficiente d'inconsistenza per definire nuovi cluster, mantenendo però i vincoli gerarchici all'interno del dendrogramma. In particolare, per non spezzare i collegamenti 45 (coeff. d'inconsistenza pari a 0.7071), 46 (coeff. d'inconsistenza pari a 0.7816) e 51 (coeff. d'inconsistenza pari a 0.1539) è necessario imporre un valore di taglio superiore a 0.1539 ma contemporaneamente inferiore a 0.1545, per spezzare il collegamento 47 (vedi Tabella 13 e Figura 18). Si può fissare quindi un valore di taglio pari a 0.1541:

```
>> T = cluster(Z, 'cutoff', 0.1541);
```

Complessivamente si ottengono 7 cluster. In Figura 22, i cluster sono messi in evidenza sul dendrogramma e in legenda viene indicato per ciascun Paese il cluster di appartenenza secondo il vettore T. Di seguito vengono analizzati i vari schemi di diffusione secondo l'ordine di disposizione dei Paesi sul dendrogramma.

Il cluster I (vedi Figura 23) nasce dall'unione dei precedenti cluster I, II, III, IV e V ed è il più numeroso in termini di Paesi che lo compongono. Analizzando lo schema di diffusione delle certificazioni ISO 9000 per questo raggruppamento, si può osservare un andamento crescente per l'intero intervallo di tempo considerato (1993-2016), solamente negli ultimi cinque anni circa si nota una saturazione del numero di certificazioni. In accordo con il modello proposto da Franceschini *et al.* (2010)¹⁵, per questo cluster sono quindi ben identificabili le fasi di crescita (*growth phase*) e maturità (*maturity phase*) del fenomeno di diffusione delle ISO 9000. Tuttavia, si può notare che la Svizzera sembra discostarsi dagli altri Paesi, nel dendrogramma

¹⁵ Vedi pag. 13 del cap. 1 di questo studio.

questo Paese si trova tra i cluster I e II e rientra in questo raggruppamento dall'implementazione della funzione `cluster`, se non si tenesse conto dei vincoli gerarchici si potrebbe pensare di spostare questo Paese nel prossimo cluster (II).

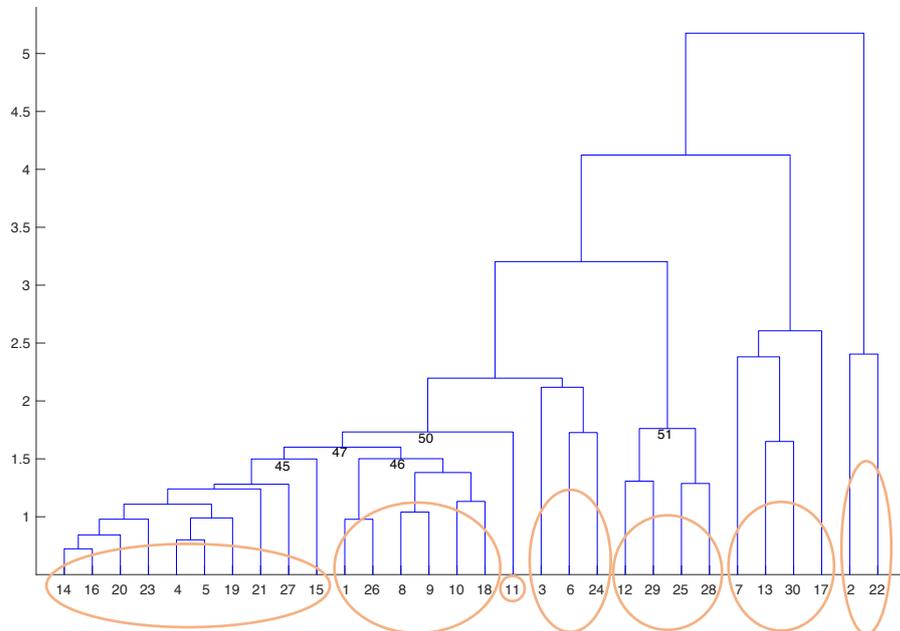


Figura 22
Dendrogramma con suddivisione dei Paesi europei in 7 cluster

| Numero Paese | Cluster | Numero Paese | Cluster | Numero Paese | Cluster |
|---------------|---------|----------------|---------|---------------|---------|
| 1 Austria | 2 | 11 Grecia | 3 | 21 Romania | 1 |
| 2 Bielorussia | 7 | 12 Ungheria | 5 | 22 Russia | 7 |
| 3 Belgio | 4 | 13 Irlanda | 6 | 23 Slovacchia | 1 |
| 4 Bulgaria | 1 | 14 Italia | 1 | 24 Slovenia | 4 |
| 5 Croazia | 1 | 15 Lettonia | 1 | 25 Spagna | 5 |
| 6 Rep. Ceca | 4 | 16 Lituania | 1 | 26 Svezia | 2 |
| 7 Danimarca | 6 | 17 Paesi Bassi | 6 | 27 Svizzera | 1 |
| 8 Finlandia | 2 | 18 Norvegia | 2 | 28 Turchia | 5 |
| 9 Francia | 2 | 19 Polonia | 1 | 29 Ucraina | 5 |
| 10 Germania | 2 | 20 Portogallo | 1 | 30 UK | 6 |

Legenda
Numero Paese e cluster di appartenenza secondo il vettore T

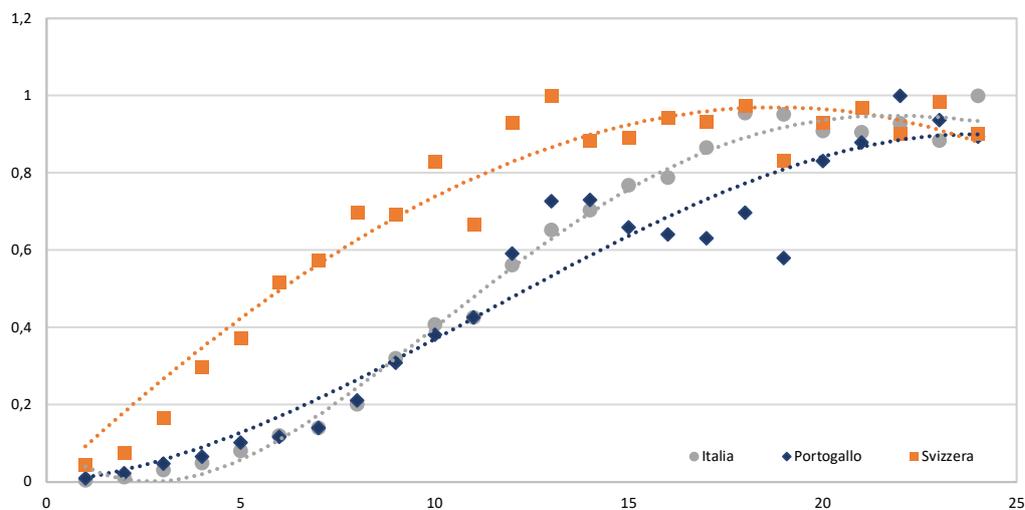


Figura 23
Cluster I

Note: Bulgaria, Croazia, Italia, Lettonia, Lituania, Polonia, Portogallo, Romania, Slovacchia, Svizzera appartengono a questo cluster.

Il cluster II (vedi Figura 24) è il secondo grande raggruppamento in termini di numero di Paesi. Nasce dall'unione dei precedenti cluster VI, VII e VIII ed è costituito dai principali Paesi dell'Europa centrale (Francia, Germania, Austria) e della penisola scandinava (Norvegia, Svezia, Finlandia). Per questi Paesi si può osservare come la crescita del numero di certificazioni ISO 9000 si sia attenuata a partire dai primi anni Duemila, da notare nel 2003 (anno 11 nel grafico) come il passaggio alla normativa aggiornata abbia causato un calo del numero di certificazioni, da allora la crescita si è inizialmente rallentata (crescita con tasso inferiore), fermandosi in seguito con il raggiungimento del livello di saturazione.

Il cluster III (vedi Figura 25) è costituito solo dalla Grecia la quale, come già osservato, ha un andamento intermedio tra i cluster I e II, nel dendrogramma infatti questo Paese unisce i due gruppi attraverso il collegamento 50. Questo cluster definisce uno schema di diffusione unico: tendenzialmente si osserva una curva-a-S con una fase di crescita e di saturazione come nei cluster precedenti, si può notare però un'oscillazione del numero di certificazioni attorno al trend con una periodicità di circa cinque anni.

Il cluster IV (vedi Figura 26) definisce uno schema di diffusione molto simile al cluster II. In particolare, il Belgio sembra avere un andamento discostante rispetto a Rep. Ceca e Slovenia, se non si tenesse conto della gerarchia, potrebbe essere inserito nel cluster II o in alternativa si potrebbe pensare di unire i cluster II e IV in un unico raggruppamento.

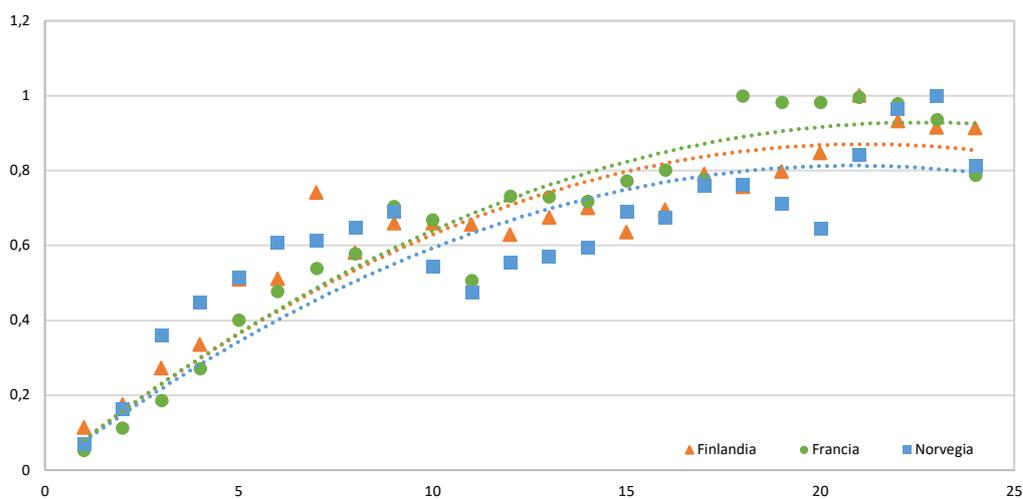


Figura 24
Cluster II

Note: Austria, Finlandia, Francia, Germania, Norvegia e Svezia appartengono a questo cluster.

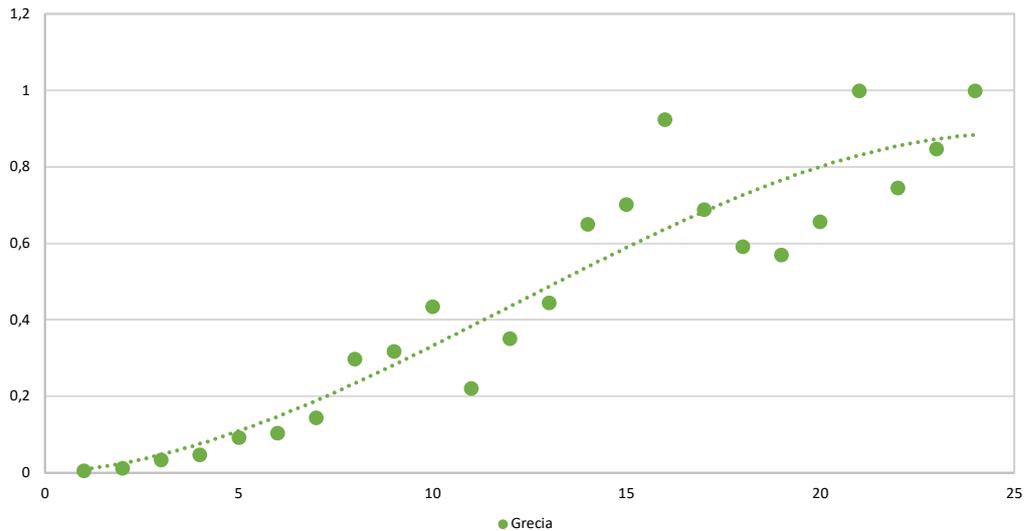


Figura 25
Cluster III

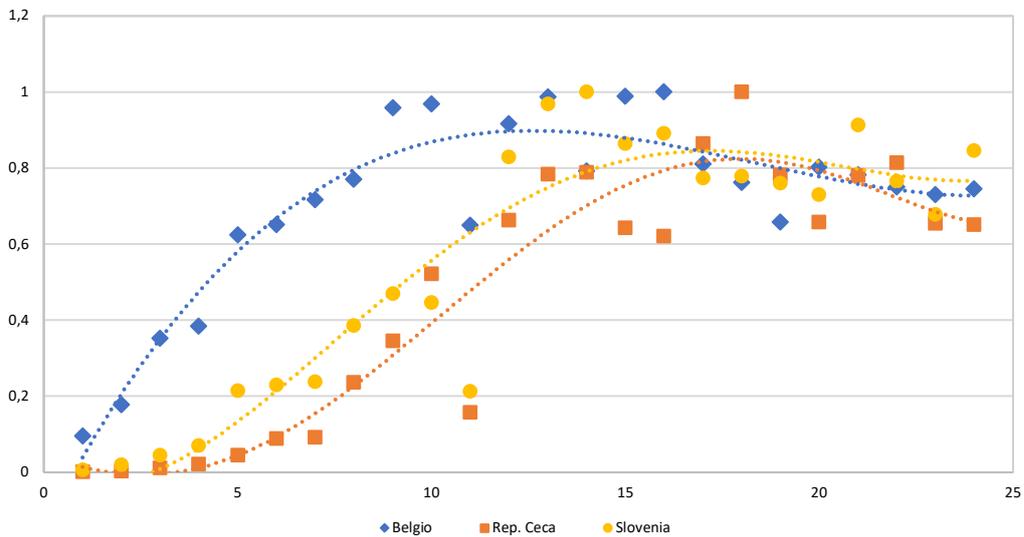


Figura 26
Cluster IV

Note: Belgio, Repubblica Ceca e Slovenia appartengono a questo cluster.

Il cluster V (vedi Figura 27), ottenuto unendo i precedenti cluster XI e XII, definisce uno schema di diffusione delle certificazioni caratterizzato dalle fasi di crescita (*growth phase*), saturazione (*maturity phase*) e declino (*decline phase*) concentrate nei primi vent'anni dell'orizzonte temporale. Si osserva quindi un chiaro andamento a U-rovesciata e, a valle della fase di declino, si può notare una fase di stazionarietà del numero di ISO 9000 negli ultimi anni. Si potrebbe considerare questo cluster come una fotografia dello stesso fenomeno la quale anticipa quello che potrebbe succedere in futuro per i Paesi dei cluster precedenti e presenta un andamento già verificatosi per i Paesi del prossimo cluster.

Il cluster VI (vedi Figura 28), costituito da Danimarca, Irlanda, Regno unito e Paesi Bassi, può essere considerato come il più evoluto in termini di diffusione delle certificazioni ISO 9000. In questo raggruppamento sono presenti infatti tutti i trend dei cluster precedenti secondo un andamento a U-rovesciata (crescita, saturazione e declino) nel primo decennio dell'orizzonte temporale considerato, in seguito si osserva una fase di stazionarietà del numero di certificazioni attorno ad un valore stabile per gli ultimi 10 anni. I cluster precedentemente analizzati potrebbero essere quindi considerati come istantanee dello stesso fenomeno, scattate in diversi momenti della sua evoluzione. Con i dati ad oggi disponibili si potrebbe ipotizzare pertanto che l'andamento delle certificazioni preveda complessivamente una prima fase di crescita, poi una fase di saturazione, una successiva fase di declino e infine una fase di stazionarietà.

L'ultimo cluster (vedi Figura 29) contiene al suo interno la Russia e la Bielorussia. Per questi Paesi la diffusione delle certificazioni sembra avere un andamento molto particolare: il fenomeno risulta infatti limitato nell'intero orizzonte temporale analizzato e presenta dei picchi concentrati in pochi anni.

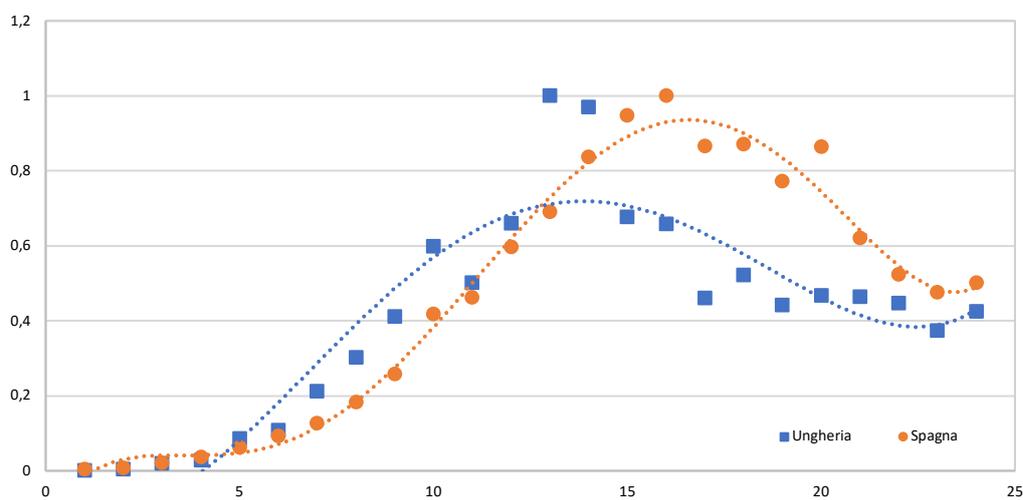


Figura 27
Cluster V

Note: Spagna, Turchia, Ucraina e Ungheria appartengono a questo cluster.

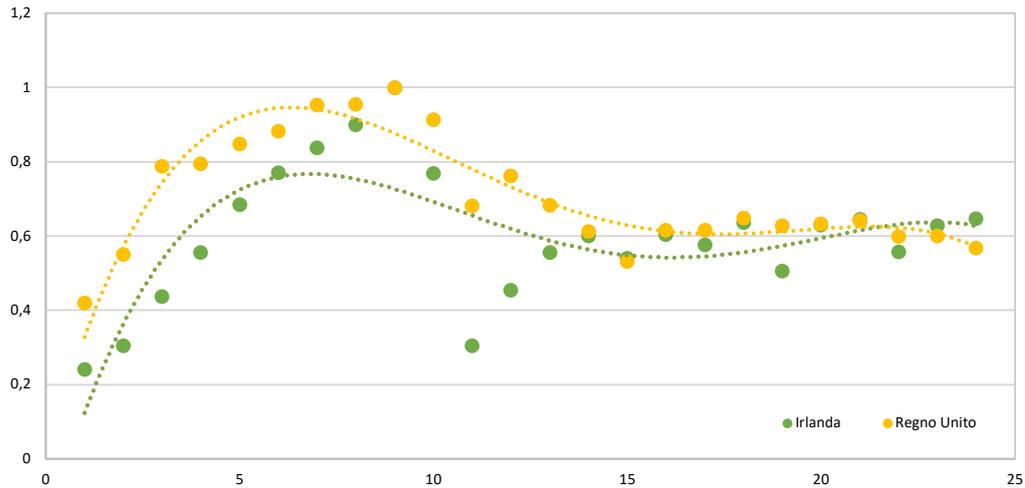


Figura 28
Cluster VI

Note: Danimarca, Irlanda, Paesi Bassi e Regno Unito appartengono a questo cluster.

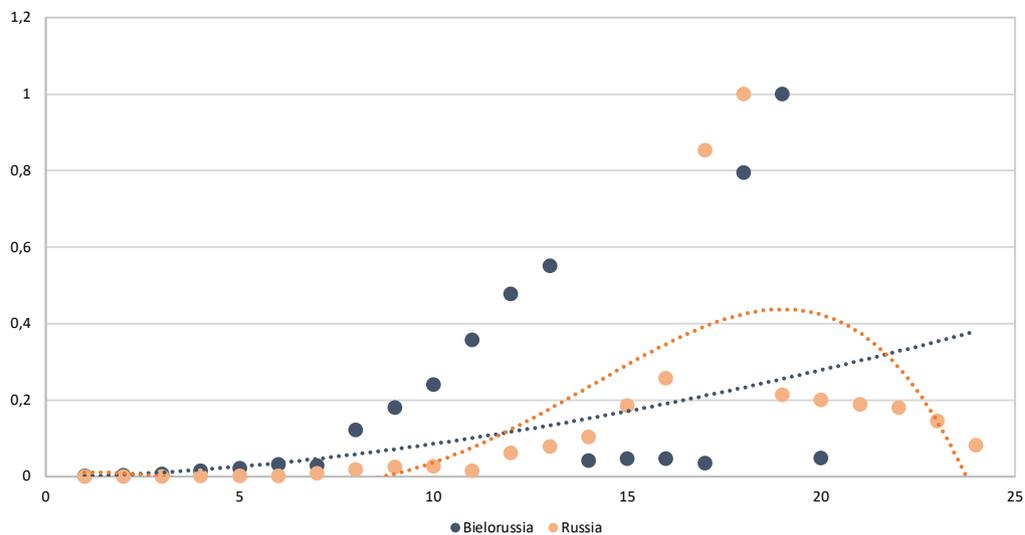


Figura 29
Cluster VII

Note: Bielorussia e Russia appartengono a questo cluster.

Se non si tenesse conto della gerarchia definita dal dendrogramma si potrebbero apportare alcune modifiche ai risultati ottenuti sulla base delle osservazioni precedenti. In particolare, inserendo la Svizzera nel cluster II, si osserva un andamento chiaramente più simile ai Paesi appartenenti a questo cluster rispetto a quelli del cluster I (vedi Figura 30). Allo stesso modo, spostando il Belgio nel cluster II (vedi Figura 30), si nota un'evidente somiglianza in termini di andamento con gli altri componenti di questo cluster. Questa osservazione suggerisce inoltre di unire i cluster II e IV (vedi Figura 31).

Dall'unione dei cluster II e IV si ottiene un chiaro andamento comune per i Paesi che ne fanno parte, caratterizzato da una prima fase di crescita delcertificazioni concentrata nei primi 10

anni e una successiva fase di saturazione che potrebbe anticipare le posteriori fasi di declino e di stazionarietà osservabili nei cluster V e VI. La Repubblica Ceca e la Slovenia sembrano discostarsi leggermente nella prima fase di crescita ma hanno lo stesso andamento degli altri Paesi a partire dall'anno 11 nel grafico. Unire i cluster II e IV può essere considerato quindi come una buona soluzione nonostante i vincoli gerarchici risultanti dall'implementazione degli algoritmi e definiti nel dendrogramma. In quest'ottica però l'albero gerarchico rappresenta una linea guida per l'interpretazione e l'analisi del fenomeno nei vari Paesi.

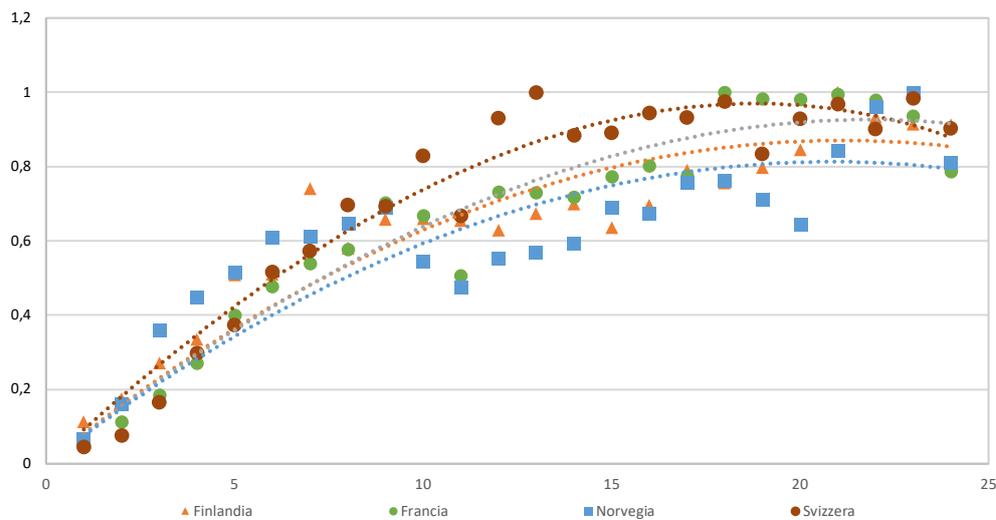


Figura 30
Cluster II con
Belgio e
Svizzera

Note: Austria, Finlandia, Francia, Germania, Norvegia, Svezia e Svizzera appartengono a questo cluster

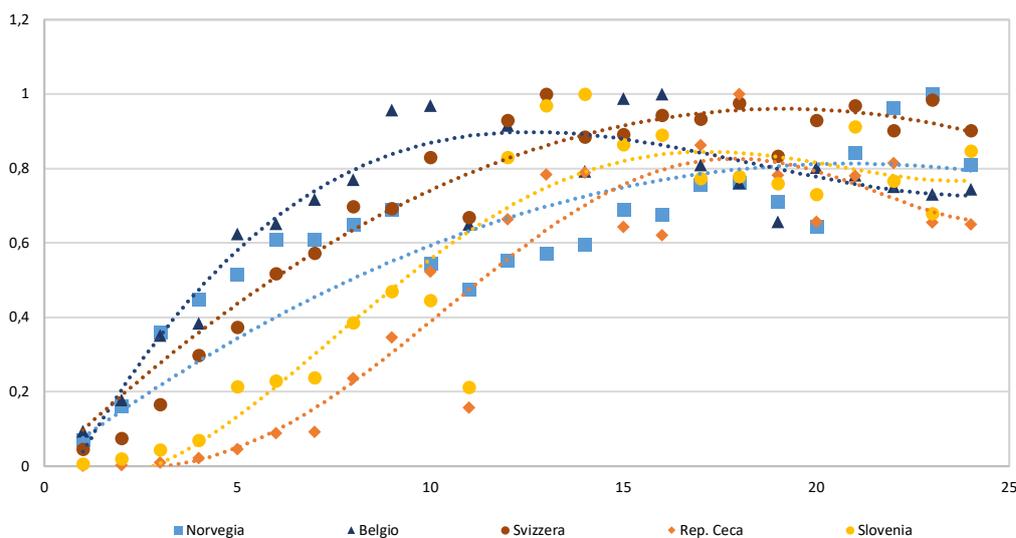


Figura 31
Unione dei
cluster
II e IV

Note: Austria, Belgio, Finlandia, Francia, Germania, Norvegia, Repubblica Ceca, Slovenia, Svezia e Svizzera appartengono a questo cluster

CONCLUSIONI

In accordo con l'ultimo *ISO Survey*, il numero di certificazioni ISO 9000 rilasciate al 31 dicembre 2016 ammontava a 1'106'000 in 201 Paesi con una crescita del 7% rispetto al 2015. Nonostante questi numeri riflettano l'importanza internazionale che le ISO 9000 hanno assunto, dall'analisi del caso europeo si può osservare che alcuni Paesi hanno raggiunto una certa saturazione o una certa stabilità circa il numero di certificazioni ISO 9000 rilasciate.

Dall'analisi presentata in questo documento, in Europa si possono distinguere sei diversi modelli di diffusione delle certificazioni, quattro di questi possono essere considerati come istantanee dello stesso fenomeno scattate in diversi momenti della sua evoluzione, gli altri due invece possono essere considerati come schemi indipendenti. In generale, la diffusione delle certificazioni ISO 9000 in Europa sembra assumere un andamento che prevede quattro fasi consecutive:

- *Fase di crescita*: il numero di certificazioni cresce negli anni in maniera costante;
- *Fase di maturità*: il numero di certificazioni raggiunge un valore massimo di saturazione che si mantiene costante per un certo numero di anni;
- *Fase di declino*: il numero di certificazioni decresce gradualmente negli anni;
- *Fase di stazionarietà*: il numero di certificazioni si stabilizza negli anni attorno a un valore indicativamente intermedio tra il numero minimo e il numero massimo di certificazioni ISO 9000 registrate nel Paese.

Le prime due fasi sono ben osservabili nel cluster I, in cui i Paesi sembrano aver raggiunto una saturazione delle certificazioni solo negli ultimi anni, e nel cluster risultante dall'unione del II e del IV. Per i Paesi appartenenti a quest'ultimo gruppo il fenomeno di diffusione delle ISO 9000 risulta essere più maturo, poiché la fase di crescita sembra essersi conclusa ormai da più di dieci anni con la successiva saturazione del numero di ISO 9000 da cui, ad oggi, i Paesi non sembrano ancora essere usciti. Le ultime due fasi sono ben osservabili invece nei cluster V e VI. La principale differenza tra questi due gruppi consiste nel fatto che i Paesi appartenenti al cluster V sembrano essere usciti da pochi anni dalla fase di declino e per ciascun Paese sembra che il numero di ISO 9000 si stia stabilizzando intorno a un livello stazionario. I Paesi appartenenti al cluster VI invece, apparentemente pionieri della diffusione di queste certificazioni,

sono caratterizzati ormai da più di dieci anni da una stazionarietà del numero di ISO 9000 rilasciate. Complessivamente per questi quattro cluster, come osservabile, il fenomeno sembra essere lo stesso, cambiano le dinamiche e le tempistiche con cui si evolve.

Diverso è il discorso per gli altri due schemi di diffusione rappresentati dai cluster III e VII. La Grecia, che da sola costituisce il cluster III, è caratterizzata da un andamento complessivamente di crescita delle certificazioni ma con continue fluttuazioni nell'ultimo decennio. Questo fenomeno può essere ricondotto alla particolare situazione politico-economica che il Paese sta vivendo e che si ripercuote direttamente sulle imprese e sulle organizzazioni, ossia gli enti interessati alle certificazioni ISO 9000. Infine, altrettanto particolare è lo schema di diffusione per la Bielorussia e la Russia, limitato per l'intero orizzonte temporale a poche certificazioni e caratterizzato da picchi di numero di certificazioni concentrati in alcuni anni, in questo caso l'analogia tra i Paesi può essere ricondotta alla loro vicinanza geo-politica.

I precedenti risultati sono stati ottenuti mediante l'implementazione degli algoritmi di *clustering* gerarchico agglomerativo attraverso l'ambiente di calcolo *MATLAB R2017b* sia sfruttando le funzioni rese disponibili dal software di calcolo stesso, sia introducendo nuovi codici che rendessero più agevole il lavoro computazionale. A tale scopo è stato necessario ridurre il numero di Paesi europei analizzati ad un campione, questa scelta ha portato alla perdita di informazioni importanti ma ha consentito di effettuare l'analisi in maniera più congrua. Infine, essenziale è stata l'interpretazione dei risultati circa la formazione dei cluster di Paesi, al di là degli output numerici prodotti dagli algoritmi.

Lo studio dei modelli di diffusione delle certificazioni ISO 9000 è importante sia per gli enti certificatori al fine di comprendere gli esiti del loro business sia per le imprese e le organizzazioni che le richiedono, le attuano e ne sostengono i costi. A tal proposito, studi futuri dovrebbero concentrarsi sulla verifica dei risultati ottenuti e le previsioni proposte in questo documento, focalizzandosi in particolare sulla fase stazionaria individuata circa il numero di certificazioni ISO 9000 emesse, individuandone i driver e le variabili che la influenzano. Un altro ambito di ricerca interessante consiste nell'estendere la seguente analisi ad un'area geografica più ampia per poter osservare se i risultati riscontrati in Europa si presentano anche negli altri continenti interessati alle certificazioni ISO 9000.

Shiftvector

La funzione `shiftvector` prevede in ingresso un vettore riga `X` e fornisce in uscita il vettore riga ridimensionato `XN`. Attraverso un ciclo-for definito sul numero di elementi del vettore in ingresso, la funzione parte dal primo elemento positivo¹⁶ e lo copia nel primo elemento del vettore in output `XN`. Il risultato è quindi un vettore che ha gli stessi elementi di partenza ma con dimensione inferiore in quanto vengono eliminati gli elementi iniziali che non presentano valori numerici. Poiché per il campione di Paesi analizzato l'assenza di osservazioni riguarda solamente i primi anni contigui dell'orizzonte temporale 1993-2016, non si hanno problemi di eventuali buchi o salti di osservazioni negli anni, in questo caso infatti la soluzione migliore sarebbe quella di risalire al numero di certificazioni ISO 9000 per quell'anno, o per gli anni mancanti, sulla base dell'andamento degli anni precedenti e successivi. Di seguito viene riportato il codice della funzione `shiftvector`:

```
function [XN]=shiftvector(X) % function to create a new vector NX which contains
                             % all the X-vector's elements shifted to the left
                             % where there are vacancies

m = size(X,2);                % #columns vector X
k = 1;                        % initialization variable k
XN = [];                       % initialization vector XN

    for j=1:m                  % move along the columns i.e. along the elements, if
        if X(j)>=0             % the row's element is positive then it is shifted
            XN(k)=X(j);        % to the left starting from the first element of the
            k=k+1;             % same row in the new matrix XN
        end
    end
end
```

¹⁶ Essendo il numero di certificazioni ISO 9000 non negativo è sufficiente imporre come condizione la positività dell'elemento del vettore.

Dtwmatrix

La funzione `dtwmatrix` consente di calcolare la distanza DTW implementando ricorsivamente la funzione `dtw` tra i vettori riga della matrice `X`. L'output è la matrice delle distanze `Ddist` in cui sia le righe sia le colonne corrispondono ai Paesi e l'elemento generico d_{ij} corrisponde alla distanza DTW tra il Paese sulla riga i e il Paese sulla colonna j . La funzione prevede in particolare un doppio ciclo-for in cui fissato ciascun vettore riga della matrice `X` (opportunamente ridimensionato attraverso la funzione `shiftvector`), viene calcolata la distanza DTW rispetto a tutti gli altri vettori righe di `X` (anch'essi di volta in volta opportunamente ridimensionati attraverso la funzione `shiftvector`). Il valore della distanza calcolato viene quindi iterativamente inserito nella matrice `Ddist` in output. Di seguito viene riportato il codice della funzione `dtwmatrix`:

```
function [Ddist] = dtwmatrix(X) % function to compute Dynamic Time Warping
                                % between the rows of matrix X

n = size(X,1); % #rows matrix X
Ddist = []; % initialization matrix Ddist
Y = []; % initialization vector Y
Z = []; % initialization vector Z

    for i=1:n % fix i-vector i.e. i-row
        Y=shiftvector(X(i,:)); % tidy i-vector up
        for j=1:n % fix j-vector i.e. j-row
            Z=shiftvector(X(j,:)); % tidy j-vector up
            D = dtw(Y,Z); % compute DTW (Euclidean) between i and j
            Ddist(i,j)=D; % enter DTW value into Ddist
        end
    end
end
```

Transformdist

La funzione `transformdist` trasforma la matrice triangolare inferiore di `Ddist` nel vettore riga `D`. Attraverso un doppio ciclo-for la funzione inserisce in `D` ciascun elemento della matrice triangolare inferiore partendo dalla prima colonna e procedendo in ordine di riga, quindi, in maniera iterativa per ciascuna colonna e per ciascuna sua riga i valori della matrice vanno a definire il vettore riga `D`. Similmente, la stessa funzione può essere definita per la matrice triangolare superiore, partendo in questo caso dalla prima riga e procedendo in ordine di co-

lonne. La successione con cui i valori vengono inseriti in D è molto importante, in quanto innanzitutto questa riflette l'ordine con cui i Paesi sono disposti sulle righe e sulle colonne della matrice delle distanze e in secondo luogo perché sulla base delle distanze in D la funzione linkage crea i collegamenti e quindi i cluster tra i Paesi. Di seguito viene riportato il codice della funzione transformdist:

```
function [D]=transformdist(Ddist) % function to transform the lower triangular
                                % part of Ddist into the vector D

n = size(Ddist,1);                % #rows matrix Ddist (equal to #columns)
D = [];                            % initialization vector D
k = 1;                            % initialization variable k

    for j=1:n                      % fix j-column starting from the 1st
        for i=j+1:n              % fix each i-row starting from the j+1
            D(k) = Ddist(i,j);  % enter the value Ddist(i,j) in the vector D
            k=k+1;              % update variable k
        end
    j=j+1;                        % update variable j
    end
end
```

BIBLIOGRAFIA

- Aldenderfer, M. e Blashfield, R. (1984). *“Cluster analysis”*. Beverly Hills: Sage Publications.
- Anderson, S., Daly, J. e Johnson, M. (1999). “Why firms seek ISO 9000 certification: regulatory compliance or competitive advantage?”. *Production and Operations Management*, Vol. 8(1), pp. 28-43.
- Berndt, D. e Clifford, J. (1994). “Using dynamic time warping to find patterns in time series”. *Proceeding AAAIWS'94 Proceedings of the 3rd International Conference on Knowledge Discovery and Data Mining*, pp. 359-70.
- Caiado, J., Crato, N. e Pena, D. (2006). “A periodogram-based metric for time series classification”. *Computational Statistics & Data Analysis*, Vol. 50(10), pp. 2668-84.
- Castka, P. e Corbett, C. (2013). “Management systems standards: diffusion, impact and governance of ISO 9000, ISO 14000, and other management standards”. *Foundations and Trends® in Technology, Information and Operations Management*, Vol. 7(3-4), pp. 161-379.
- Clougherty, J. e Grajek, M. (2008). “The impact of ISO 9000 diffusion on trade and FDI: a new institutional analysis”. *Journal of International Business Studies*, Vol. 39(4), pp. 613-33.
- Coleman, J. (2017, 31 Gennaio). *Dynamic programming exemple (dynamic time warping)*. Tratto da Phonetics Laboratory - University of Oxford:
http://www.phon.ox.ac.uk/jcoleman/old_SLP/Lecture_5/DTW_explanation.html
- Corbett, C. e Kirsch, D. (2001). “International diffusion of ISO 14000 certification”. *Production and Operations Management*, Vol. 10(3), pp. 327-42.
- Douzal Chouakria, A. e Naidu Nagabhushan, P. (2007). “Adaptive dissimilarity index for measuring time series proximity”. *Advances in Data Analysis and Classification*, Vol. 1, pp. 5-21.
- Franceschini, F. (2002). *“Advanced Quality Function Deployment”*. Boca Roton, FL., US: CRC Press.
- Franceschini, F., Galetto, M. e Gianni, G. (2004). “A new forecasting model for the diffusion of ISO 9000 standard certifications in European countries”. *International Journal of Quality & Reliability Management*, Vol. 21(1), pp. 35-50.
- Franceschini, F., Galetto, M. e Cecconi, C. (2006). “A worldwide analysis of ISO 9000 standard diffusion”. *Benchmarking: An International Journal*, Vol. 13(4), pp. 523-41.
- Franceschini, F., Galetto, M., Maisano, D. e Mastrogiacomo, L. (2010). “Clustering of European countries based on ISO 9000 certification diffusion”. *International Journal of Quality & Reliability Management*, Vol. 27(5), pp. 558-75.
- Garcia-Escudero, L. e Gordaliza, A. (2005). “A proposal for robust curve clustering”. *Journal of Classification*, Vol. 22(2), pp. 185-201.
- Gower, J. e Ross, G. (1969). “Minimum spanning trees and single linkage cluster analysis”. *Applied Statistics*, Vol. 18(1), pp. 54-64.
- Hastie, T., Tibshirani, R. e Friedman, J. (2009). *“The Elements of Statistical Learning”*. New York: Springer.
- Heckman, N. e Zamar, R. (2000). “Comparing the shapes of regression functions”. *Biometrika*, Vol. 87(1), pp. 135-44.
- ISO. (1994). ISO 9002:1994. *“Quality systems – Model for quality assurance in production, installation and servicing”*, 2. ed. Ginevra.

- ISO. (1994). ISO 9003:1994. "Quality systems – Model for quality assurance in final inspection and test", 2. ed. Ginevra.
- ISO. (2000). ISO 9001:2000. "Quality management systems – Requirements", 3. ed. Ginevra.
- ISO. (2009, 11). ISO 9004: 2009. "Managing for the sustained success of an organization – A quality management approach", 3. ed. N° 46. Ginevra.
- ISO. (2015, 09). ISO 9000:2015. "Quality management systems – Fundamentals and vocabulary", 4. ed. N° 51. Ginevra.
- ISO. (2015). ISO 9001:2015. "Quality management systems – Requirements", 5. ed. N° 29. Ginevra.
- ISO. (2017, 20 Dicembre). "ISO 9000 - Quality management". Tratto da International Organization for Standardization: <https://www.iso.org/iso-9001-quality-management.html>
- ISO. (2017, 09). "The ISO Survey of Management System Standard Certifications 2016". Ginevra.
- Kakizawa, Y., Shumway, R. e Taniguchi, M. (1998). "Discrimination and clustering for multivariate time series". *Journal of the American Statistical Association*, Vol. 93(441), pp. 328-40.
- Keller, K. e Wittfeld, K. (2004). "Distances of time series components by means of symbolic dynamics". *International Journal of Bifurcation and Chaos*, Vol. 14(2), pp. 693-704.
- MacQueen, J. (1967). "Some Methods for Classification and Analysis of MultiVariate Observations". *Proc. of the fifth Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 1, pp. 281-97.
- Maharaj, E. (2000). "Cluster of time series". *Journal of Classification*, Vol. 17(2), pp. 297-314.
- Marimon Viadiu, F., Casadesús Fa, M. e Heras Saizarbitoria, I. (2006). "ISO 9000 and ISO 14000 standards: an international diffusion model". *International Journal of Operations & Production Management*, Vol. 26(2), pp. 141-65.
- Oates, T., Firoiou, L. e Cohen, P. (1999). "Clustering time series with Hidden Markov Models and Dynamic Time Warping". In *Proceedings of the IJCAI-99 Workshop on Neural, Symbolic and Reinforcement Learning Methods for Sequence Learning*, pp. 17-21.
- Romano, P. e Vinelli, A. (2001). "Quality management in a supply chain perspective. Strategic and operative choices in a textile-apparel network". *International Journal of Operations & Production Management*, Vol. 21, pp. 446-60.
- Sampaio, P., Saraiva, P. e Guimarães Rodrigues, A. (2009a). "A statistical analysis of ISO 9000 related data for ultra-peripheral and Portugues regions". *Quality Management Journal*, Vol. 16(2), pp. 44-58.
- Sampaio, P., Saraiva, P. e Guimarães Rodrigues, A. (2009b). "An analysis of ISO 9000 data in the wolrd and the European Union". *Total Quality Manangement & Business Excellence*, Vol. 20(12), pp. 1303-20.
- Sampaio, P., Saraiva, P. e Guimarães Rodrigues, A. (2011). "ISO 9001 certification forecasting models". *International Journal of Quality & Reliability Managment*, Vol. 28(1), pp. 5-26.
- Sankoff, D. e Kruskal, J. (1983). "Time warp, string edits, and macromolecules: the theory and practice of sequence comparizon". Reading, Massachusetts: Addison-Wesley.
- Saraiva, P. e Duarte, B. (2003). "ISO 9000: some statistical results for a worldwide phenomenon". *Total Quality Management & Business Excellence*, Vol. 14(10), pp. 1169-78.

- Serban, N. e Wasserman, L. (2005). "CATS: Clustering after transformation and smoothing". *Journal of the American Statistical Association*, Vol. 100(471), pp. 990-99.
- The British Standards Institution. (2017, 20 Dicembre). "Our History". Tratto il giorno 20 Dicembre 2017 da Website BSI: <https://www.bsigroup.com/en-GB/about-bsi/our-history/>
- The MathWorks, Inc. (2017, 13 Settembre). "Cluster". Tratto a Febbraio 2018 da MATLAB R2017b Documentation: <https://it.mathworks.com/help/stats/cluster.html>
- The MathWorks, Inc. (2017, 13 Settembre). "Clusterdata". Tratto a Febbraio 2018 da MATLAB R2017b Documentation: <https://it.mathworks.com/help/stats/clusterdata.html>
- The MathWorks, Inc. (2017, 13 Settembre). "Cophenet". Tratto a Febbraio 2018 da MATLAB R2017b Documentation: <https://it.mathworks.com/help/stats/cophenet.html>
- The MathWorks, Inc. (2017, 13 Settembre). "Dendrogram". Tratto a Febbraio 2018 da MATLAB R2017b Documentation: <https://it.mathworks.com/help/stats/dendrogram.html>
- The MathWorks, Inc. (2017, 13 Settembre). "Dtw". Tratto a Febbraio 2018 da MATLAB R2017b Documentation: https://de.mathworks.com/help/signal/ref/dtw.html?s_tid=doc_ta
- The MathWorks, Inc. (2017, 13 Settembre). "Hierarchical Clustering". Tratto a Febbraio 2018 da MATLAB R2017b Documentation: https://it.mathworks.com/help/stats/hierarchical-clustering.html?s_tid=srchtitle
- The MathWorks, Inc. (2017, 13 Settembre). "Inconsistent". Tratto a Febbraio 2018 da MATLAB R2017b Documentation: <https://it.mathworks.com/help/stats/inconsistent.html>
- The MathWorks, Inc. (2017, 13 Settembre). "Linkage". Tratto a Febbraio 2018 da MATLAB R2017b Documentation: <https://it.mathworks.com/help/stats/linkage.html>
- The MathWorks, Inc. (2017, 13 Settembre). "Pdist". Tratto a Febbraio 2018 da MATLAB R2017b Documentation: <https://it.mathworks.com/help/stats/pdist.html>
- Withers, B., Ebrahimpour, M. e Hikmet, N. (2000). "Does ISO 9000 certification affect the dimensions of quality used for competitive advantage?". *European Management Journal*, Vol. 18(4),pp. 431-43.

SITOGRAFIA

British Standard Institution: www.bsigroup.com (ultimo accesso 20-12-2017).

International Organization for Standardization: www.iso.org (ultimo accesso 16-01-2018).

The MathWorks, Inc.. *Cluster*. MATLAB R2017b Documentation:
<https://it.mathworks.com/help/stats/cluster.html> (ultimo accesso 10-2-2018).

The MathWorks, Inc.. *Clusterdata*. Tratto MATLAB R2017b Documentation:
<https://it.mathworks.com/help/stats/clusterdata.html> (ultimo accesso 10-2-2018).

The MathWorks, Inc.. *Cophenet*. MATLAB R2017b Documentation:
<https://it.mathworks.com/help/stats/cophenet.html> (ultimo accesso 10-2-2018).

The MathWorks, Inc.. *Dendrogram*. MATLAB R2017b Documentation:
<https://it.mathworks.com/help/stats/dendrogram.html> (ultimo accesso 10-2-2018).

The MathWorks, Inc.. *Dtw*. MATLAB R2017b Documentation:
https://it.mathworks.com/help/signal/ref/dtw.html?s_tid=doc_ta
(ultimo accesso 15-2-2018).

The MathWorks, Inc.. *Hierarchical Clustering*. MATLAB R2017b Documentation:
https://it.mathworks.com/help/stats/hierarchical-clustering.html?s_tid=srchtitle (ultimo accesso 10-2-2018).

The MathWorks, Inc.. *Inconsistent*. MATLAB R2017b Documentation:
<https://it.mathworks.com/help/stats/inconsistent.html> (ultimo accesso 10-2-2018).

The MathWorks, Inc.. *Linkage*. MATLAB R2017b Documentation:
<https://it.mathworks.com/help/stats/linkage.html> (ultimo accesso 10-2-2018).

The MathWorks, Inc.. *Pdist*. MATLAB R2017b Documentation:
<https://it.mathworks.com/help/stats/pdist.html> (ultimo accesso 10-2-2018).