



PROGETTO WISHES

Studio di fattibilità

a cura di

Marcella Corduas

(marcella.corduas@unina.it)

Giancarlo Ragozini

(giancarlo.ragozini@unina.it)

Unità di Ricerca:

prof. Marcella Corduas, prof. Giancarlo Ragozini, dott. Alfonso Piscitelli, dott.ssa Danila Azzolina

3 SETTEMBRE 2014

UNIVERSITA' DI NAPOLI FEDERICO II
Dipartimento di Scienze Politiche

INDICE

1.1	Introduzione	3
1.2	Piano generale del lavoro	3
1.3	Osservazioni preliminari sui dati e le limitazioni	4
Capitolo 2. Analisi statica		9
2.1	Il concetto di vulnerabilità	9
2.2	I dati	10
2.3	Le analisi preliminari: il trattamento dei dati mancanti	12
2.4	L'analisi fattoriale multipla e la rappresentazione dei Paesi	20
2.4.1	Analisi parziali	21
2.4.2	Analisi globale (o analisi media)	23
2.4.3	Analisi interstrutturale	23
2.5	I principali risultati	23
2.6	Osservazioni finali	47
Capitolo 3. I disastri naturali		49
3.1	Introduzione	49
3.2	Definizioni e dati	52
3.2	Il modello statistico	56
3.2.1	La probabilità di emergenza allorché è osservato un disastro	56
3.2.2	La probabilità che si verifichi una calamità	59
3.3	La classifica ed una visualizzazione dei Paesi a rischio	59
Capitolo 4. Le crisi umanitarie indotte dall'instabilità politica		65
4.1	Introduzione	65
4.2	Un sistema di monitoraggio fondato su GDELT	66
4.3	Le carte di controllo	72
4.3.1	Le carte di controllo CUSUM e EWMA	72
4.3.2	Il monitoraggio delle situazioni di crisi	74
4.4	Una rappresentazione dei Paesi rispetto ai precursori di instabilità	85
Conclusioni e raccomandazioni finali		91
Bibliografia		95
Appendice - Descrizione dei dati		99
A.1	Lista delle variabili (in ordine alfabetico)	99
A.2	Descrizione dell'archivio EM-DAT del CRED	122
A.3	Statistiche descrittive e dati mancanti	125
A.4	Grafico albero classificazione	131
A.5	Timeline crisi siriana	135

1.1 Introduzione

Il Progetto WISHES è stato promosso da Save the Children (STC) come primo passo per la pianificazione di un nuovo sistema informativo che operi da supporto al processo decisionale interno. L'attenzione è rivolta alla ricerca di un sistema che possa fornire dati, informazioni e un insieme di analisi statistiche finalizzate alla assegnazione ai Paesi di livelli di allarme contro il rischio di insorgenza di eventuali crisi umanitarie. Il sistema dovrebbe includere fonti informative provenienti da varie origini (dati interni, dati secondari e opinioni di esperti) ed un complesso di strumenti, anche grafici, che ne consentano l'immediata elaborazione. Inoltre, esso deve rispettare una serie di vincoli esistenti: l'ammontare delle risorse finanziarie che condiziona la scala di grandezza, il tempo di completamento del progetto, la possibilità di stabilire rapporti di cooperazione tra i componenti di STC e partner esterni al fine di migliorare il processo di raccolta ed analisi dei dati.

L'obiettivo indicato è molto ampio a causa della molteplicità di calamità che possono generare una crisi umanitaria e del livello di aggregazione considerato, ovvero i Paesi del mondo. La realizzazione di tale obiettivo dipende fortemente dalla progettazione iniziale del sistema che comporta la specificazione dei dati da raccogliere, l'organizzazione dei flussi informativi, l'identificazione delle tecniche statistiche per l'elaborazione dei dati, la costruzione di una interfaccia di comunicazione, flessibile e orientata all'utente, che faciliti l'interpretazione dei risultati anche per i non-esperti di statistica.

Questa attività è piuttosto complessa e richiede la collaborazione di esperti dei vari settori. Per questo motivo, STC ha promosso uno studio preliminare ed esplorativo al fine di ottenere una panoramica del problema, delle indicazioni per l'ulteriore fase di sviluppo ed una valutazione dei problemi e dei limiti delle fonti informative e delle metodologie disponibili.

I quesiti che devono essere affrontati sono i seguenti:

- E' possibile identificare i principali fattori che determinano la vulnerabilità dei Paesi alle calamità?
- Esiste una relazione misurabile e significativa tra questi fattori e la comparsa delle crisi umanitarie?
- E' possibile sfruttare tali relazioni per costruire un sistema che produca un livello di attenzione per i Paesi che potenzialmente possono sperimentare una crisi umanitaria in futuro?

1.2 Piano generale del lavoro

In primo luogo, per poter affrontare lo studio, il problema è stato segmentato e sono stati individuati gli aspetti che ragionevolmente potevano essere esaminati tenuto conto del tempo a disposizione e dell'ampio spettro dell'obiettivo assegnato.

Il verificarsi di un disastro può determinare l'interruzione del funzionamento di una comunità o di una società con perdite generalizzate (in termini di vite umane, peggioramento delle condizioni igienico-sanitarie, perdita dei mezzi di sussistenza, danni alle cose ed alle attività economiche etc.) con la conseguente incapacità della comunità a reagire e recuperare. Poiché queste condizioni possono essere generate da vari tipi di calamità, in questa fase iniziale, abbiamo concentrato la nostra attenzione su due categorie:

- le calamità naturali con particolare riferimento alle tipologie più frequenti (inondazioni, terremoti, temperature estreme e tempeste (cicloni, tifoni, etc.);
- le catastrofi generate dal comportamento umano, con particolare attenzione a quelle originate dal progressivo deterioramento delle condizioni sociali e politiche alla base dei conflitti.

In secondo luogo, abbiamo distinto due livelli di analisi:

- l'analisi statica, finalizzata all'identificazione dei fattori che agevolano e precorrono la crisi umanitaria ed alla rappresentazione dei Paesi in termini delle loro caratteristiche strutturali che ne determinano la vulnerabilità;
- il monitoraggio nel tempo in cui viene esplorata una possibile strategia per determinare il livello di warning da attribuire ai vari Paesi per i due tipi di catastrofi (naturali e human driven).

Entrambe le analisi hanno richiesto una fase iniziale per la selezione delle fonti secondarie e delle variabili utili per gli obiettivi dichiarati. A tal fine è stato costruito un database che comprende le principali statistiche disponibili sulle caratteristiche macro-economiche, sociali e politiche di ciascun Paese. Per un sottoinsieme selezionato di tali indicatori si è proceduto anche alla ricostruzione delle serie storiche per l'ultimo ventennio. Nella successiva fase di esplorazione dei dati, sono state realizzate le analisi preliminari delle variabili considerate ed è stata implementata una procedura per il trattamento dei valori mancanti che, come vedremo, molto frequentemente sono presenti nelle statistiche ufficiali a livello internazionale.

1.3 Osservazioni preliminari sui dati e le limitazioni

Prima di procedere alla illustrazione dei risultati è opportuno evidenziare alcuni aspetti che caratterizzano lo studio intrapreso.

Il progetto è iniziato a fine settembre del 2013, quindi, le informazioni ed i dati che verranno esaminati nel corso di questo rapporto si riferiscono, se non specificato altrimenti, al 2010. Questa scelta è essenzialmente giustificata dal fatto che il 2010 costituiva, per la gran parte degli indicatori, l'anno più recente disponibile nelle statistiche ufficiali e tale da garantire una sufficiente qualità dei dati in termini di accuratezza, completezza ed accessibilità.

In particolare, abbiamo fatto riferimento alle fonti secondarie ufficiali (quali Nazioni Unite, WHO, World Bank, etc..) che mettono a disposizione i dati statistici attraverso dei repository via web. Inoltre, sono stati considerati in relazione a tematiche molto circoscritte anche i dati forniti da centri di ricerca, organizzazioni no-profit o altre istituzioni internazionali, a condizione che essi

fossero affiancati da meta-dati che fornissero una chiara esposizione delle metodologie e degli approcci seguiti nel processo di produzione del dato statistico.

Nella costruzione del nostro archivio sono stati rilevati alcuni problemi generali relativi alla qualità dei dati internazionali, ben noti nel campo della ricerca, ma di cui in genere gli utenti non sono consapevoli.

Come ogni dato statistico anche i dati internazionali, provenienti per gran parte dalle statistiche nazionali, sono ottenuti mediante indagini censuarie, rilevazioni e stime campionarie o come stime da un modello statistico.

Per chiarire questo punto, basta riflettere sull'origine di un dato utilizzato molto comunemente: la popolazione di un Paese. La Population Division delle Nazioni Unite fornisce le stime da modelli demografici che proiettano la popolazione sotto opportune ipotesi di tendenza nei tassi di mortalità, fertilità e migrazione. Le stime sfruttano, ove disponibili, i dati dei censimenti nazionali ma anche dati provenienti da surveys (quali le Demographic and Health Surveys o le Multiple-Indicator Cluster Surveys). La Population Division stima la popolazione di 233 Paesi ed aree e, da quanto da loro stesso indicato, circa la metà di questi Paesi o aree non dispongono di statistiche demografiche ufficiali con il dettaglio necessario all'elaborazione delle proiezioni. Ciò significa che all'interno della Division viene svolto un esteso lavoro di analisi di coerenza, integrazione di dati, di modellistica ad hoc.

La diversa origine dei dati (censimenti, surveys o modelli) implica un diverso livello di precisione, per tale motivo la scelta delle fonti e la loro qualificazione diviene fondamentale. Le grandi organizzazioni internazionali come ONU, WHO, World Bank hanno un proprio staff statistico che elabora i dati al fine di migliorare la loro qualità. I dati ufficiali sono accompagnati da meta-dati che sono fondamentali per comprendere il tipo di elaborazioni effettuate.

Va segnalato, comunque, che, anche avvalendosi di fonti qualificate, i dati non sono privi di problematiche che richiedono particolare cautela nel loro utilizzo. Ad esempio, i dati relativi ad un medesimo fenomeno prodotti da istituzioni diverse possono divergere anche in misura consistente poiché sono i risultati di modelli o stime campionarie diverse. Allo stesso modo, i dati diffusi da una medesima istituzione su un certo fenomeno possono non assicurare la comparabilità tra Paesi.

Un grave problema è, inoltre, costituito dal grande ritardo con cui di solito sono pubblicati i dati. Per questo motivo, i dati a livello internazionale sono presentati nelle pubblicazioni ufficiali con una dichiarazione implicita, che talora non è neppure più segnalata agli utenti: "ogni dato si riferisce all'anno più recente". Ciò implica che all'interno di una tabella nonostante l'intestazione riporti una particolare collocazione temporale, in realtà i dati contenuti si riferiscono a periodi molto eterogenei tra loro.

La seguente tabella fornisce un esempio di come questa convenzione può trarre in inganno gli utenti che non abbiano familiarità con le statistiche ufficiali. Nel volume "*The State of The World's Children 2013*", pubblicato dall'UNICEF, è inclusa la percentuale di bambini sottopeso (si tratta di una variabile stimata

congiuntamente dal WHO, UNICEF e World Bank). La tabella seguente riporta alcune righe selezionate da tale pubblicazione.

Tabella 1.1 Malnutrizione

	Underweight (%) WHO 2007-2011*			
	moderate & severe		Severe	
Afghanistan	33	x	12	x
Albania	5		2	
Algeria	3	x	1	x
Angola	16	y	7	y
Argentina	2	x	0	x
Armenia	5		1	
Azerbaiján	8	x	2	x
Baharein				
Ecuador	6	x	-	
x Data refer to years or periods other than those specified in the column heading.				
y Data differ from the standard definition or refer to only part of a country				

Fonte: UNICEF, dati WHO

La prima riga indica il periodo di riferimento 2007-2011. Si osservi però la presenza delle note che indicano le possibili fonti di disomogeneità: la definizione (vedi nota y) ed il periodo temporale (vedi nota x). La presentazione è certamente ingannevole per l'utente non specialista.

Tabella 1.2 Malnutrizione

Country	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
Afghanistan					32,9							
Albania	17					6,6				6,3		
Algeria	5,4		11,1			3,7						
Angola		27,5						15,6				
Argentina						2,3						
Armenia	2,6					4,2					5,3	
Azerbaijan	14	5,9					8,4					
Bahrain												
Ecuador					6,2							

Fonte: World Bank, dati WHO

Difatti, il confronto di questa tabella con la serie temporale annua fornita World Bank e che riporta con ampiezza di dettagli le medesime statistiche, pone chiaramente in evidenza come per alcuni Paesi il periodo di riferimento sia ben lontano da quello dichiarato. Ad esempio, per l'Afghanistan il dato si riferisce al 2004 e nessuno dei Paesi citati ha una rilevazione nel 2011.

La stessa tabella tratta dal repository del WHO in aggiunta riporta il tipo di indagine che ha originato i dati.

Tutto ciò evidenzia come particolare cautela deve essere esercitata da parte dell'utente nell'uso delle statistiche internazionali e come la implementazione di un sistema informativo interno richieda l'acquisizione dei dati dalle fonti originali in una forma standardizzata.

L'ultimo aspetto da considerare riguarda la gamma di informazioni che sono effettivamente disponibili ed immediatamente utilizzabili per ulteriori analisi. Numerose variabili concernenti fenomeni socio-economici e politici non vengono rilevate, e quando lo sono, i dati sono così frammentati ed eterogenei che il loro uso richiede una modellazione ad hoc (questo accade ad esempio per la gran parte delle statistiche riguardanti le forze lavoro ed in particolare quelle concernenti la forza lavoro minorile così come prodotte dall'International Labour Organization - ILO).

I problemi di cui sopra sono ben noti e la produzione di statistiche ufficiali affidabili e complete è ancora uno degli obiettivi non raggiunti delle Nazioni Unite, come emerge dalle parole del comunicato del high level Panel of eminent persons, Bali, 28 Marzo 2013:

"Troppo spesso, gli sforzi di sviluppo sono stati ostacolati dalla mancanza di dati più basilari circa le condizioni economiche e sociali in cui le persone vivono ... il maggiore monitoraggio e l'avvio della valutazione a tutti i livelli e in tutti i processi di sviluppo (dalla progettazione alla realizzazione) aiuterà a guidare il processo decisionale, ad aggiornare le priorità e ad assicurare un processo di responsabilizzazione".

Il rapporto del High Level Panel delle Nazioni Unite con riferimento all'"agenda di sviluppo post-2015 redatto nel 2013 ha esplicitamente rivolto l'appello per "una rivoluzione dei dati per lo sviluppo sostenibile, con una nuova iniziativa internazionale per migliorare la qualità delle statistiche e le informazioni a disposizione dei cittadini".

La nostra esperienza ha confermato le difficoltà descritte nonché una grande inefficienza nella disseminazione dei dati. Le stesse statistiche ufficiali delle Nazioni Unite sono distribuite da una molteplicità di siti, gestiti dai vari dipartimenti, che coesistono e forniscono dati (talvolta con riferimento allo stesso insieme di variabili) caratterizzati da un grado molto diverso di qualità in termini di omogeneità ed aggiornamento.

La stessa denominazione dei Paesi ed il numero dei Paesi osservati non è uniforme tra le varie fonti.

Queste circostanze devono essere seriamente prese in considerazione da STC ogni volta che necessita di dati statistici internazionali per scopi interni.

Inoltre, nella costruzione di un database interno sarà necessario prevedere dei canali di acquisizione dei dati che forniscano le informazioni in una forma organizzata e standardizzata, di modo che siano direttamente utilizzabili.

In questo quadro, il progetto ha prestato grande attenzione alla costruzione di un database omogeneo che potesse contribuire efficacemente nelle varie fasi dello studio.

La banca dati si compone delle seguenti parti:

- a) un insieme di 131 variabili macro-economiche, sociali e politiche a livello nazionale (dati annuali, 2010). Di
- b) le serie storiche ricostruite per un piccolo gruppo di tali variabili e relative agli ultimi due decenni;
- c) l'elenco degli eventi (calamità naturali) che hanno dato origine alla crisi umanitarie nel periodo 1990-2011;
- d) la lista di eventi che costituiscono segnali di grave instabilità politico-sociale relativi ad un caso di studio su cui sperimentare le tecniche di monitoraggio.

In Appendice (A.1) sono riportate le liste delle variabili che costituiscono la base dati con la descrizione delle fonti e delle note esplicative concernente il fenomeno misurato ed eventualmente i metodi di costruzione del dato. Sono, inoltre, illustrate le caratteristiche dei due archivi EM-DAT e GDELT che hanno originato i dati descritti al punto b) e c) e di cui si dirà ampiamente nel seguito.

Capitolo 2. Analisi statica

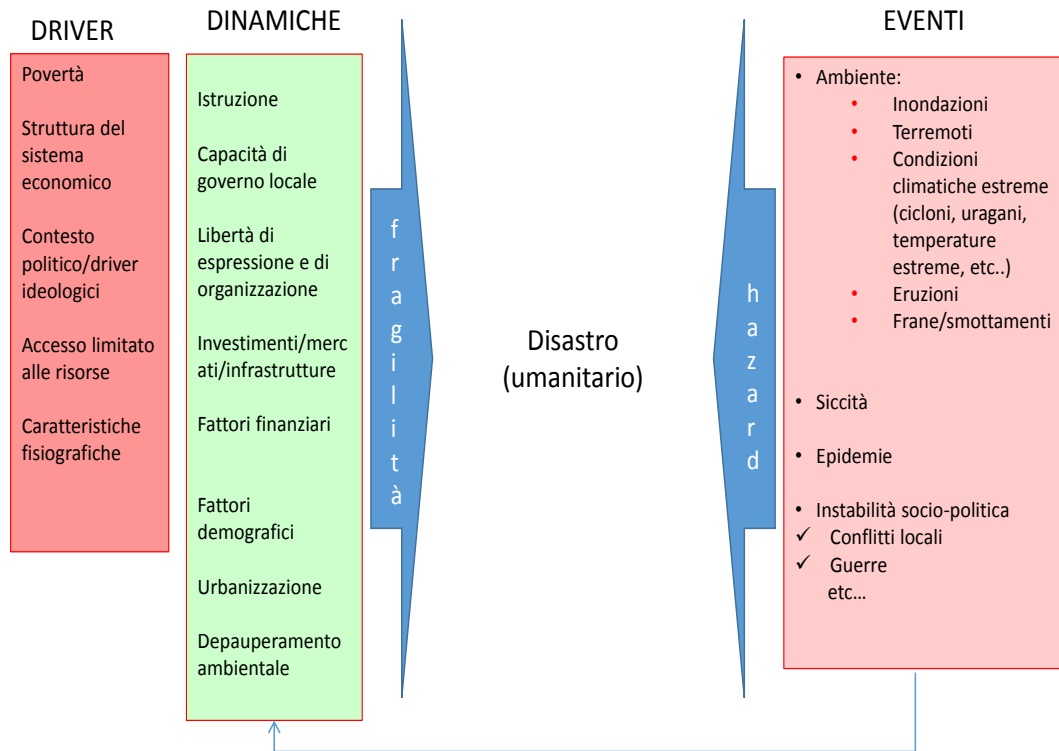
2.1 Il concetto di vulnerabilità

Le Nazioni Unite (International Strategy for Disaster Reduction) definiscono la vulnerabilità di un Paese come “the conditions determined by physical, social, economic and environmental factors or processes which increase the susceptibility of a community to the impact of hazards” (UN/ISDR, 2004). La vulnerabilità descrive quindi un concetto connesso alle condizioni della società e della collettività in tutte le sue espressioni (politiche, sociali, ambientali). L’impatto della calamità è misurato in termini di perdite, ma queste sono qualificabili in vario modo: morti, persone che perdono la casa, i beni, il lavoro, i danni economici, i danni privati, i danni alle infrastrutture e così via.

Una classificazione delle recenti definizioni, fornite in letteratura, del concetto di vulnerabilità riporta 29 descrizioni diverse (Thywissen, 2004) confermando che si tratta di un fenomeno multidimensionale la cui definizione è stata nel tempo ampliata comprendendo aspetti connessi alla attitudine, alla esposizione, alla capacità di adattamento, di reazione, e coprendo aree tematiche differenti e riferite all’ambiente, alle istituzioni, al contesto sociale ed economico (Cutter, 1996; Birkmann, 2006; Hufschmidt, 2011).

Un modello comportamentale molto diffuso per descrivere la vulnerabilità e che riteniamo fornisca una formulazione adeguata per il nostro studio, tenuto conto del livello di aggregazione a cui viene portata l’analisi, è denominato PAR. Pressure and Release (si veda, ad esempio, Wisner et al. 2004). In pratica, la vulnerabilità origina da alcune cause primarie (come la struttura del sistema economico, il sistema politico, la struttura demografica) su cui agiscono delle forze dinamiche, nel tempo e nello spazio che generano uno stato di fragilità (tasso elevato di analfabetismo, vasta disoccupazione, una elevata pressione demografica, disparità di genere, crescente urbanizzazione e depauperamento ambientale etc..) e che vanno a costituire delle condizioni di insicurezza (fragilità delle istituzioni, presenza di discriminazioni, mancata prevenzione e protezione, assenza di controlli ambientali, etc.). Il modello identifica la dipendenza delle crisi umanitarie con fattori agenti nel tempo e nello spazio e talora radicati in un passato (neppure troppo recente) di un Paese. L’intervento (release) deve quindi essere rivolto alla attenuazione di tali cause primarie e dei fattori agenti nella dinamica complessiva.

Questo modello comportamentale si applica sia ai disastri naturali (Wisner et al. 2004) che a quelli derivanti da conflitti ed instabilità socio-politica (Hughes et al., 2011).



L'analisi statica, che verr  sviluppata nelle pagine seguenti,   finalizzata a produrre un confronto tra i Paesi rispetto ai fattori che determinano la loro vulnerabilit  e che, quindi, possono precedere le crisi umanitarie.

2.2 I dati

Come abbiamo evidenziato, la vulnerabilit  di un Paese si concretizza in una diminuita capacit  di prevedere, prevenire, affrontare e assorbire l'impatto di un evento (disastro prodotto da cause naturali o da attivit  umane). Essa dipende certamente dalla struttura e dalle caratteristiche socio-economiche, politiche e geofisiche di un Paese ed   strettamente connessa (in negativo) alla capacit  di recupero di fronte ad eventi estremi. Purtroppo, quest'ultimo aspetto non pu  essere studiato direttamente in quanto non vi sono rilevazioni e indicatori specificamente indirizzati a rilevare le capacit  di prevenzione e di reazione dei Paesi che sperimentano crisi umanitarie. Pertanto, si ricorre a misure indirette.

Un'ulteriore limitazione   costituita dalla presenza nelle statistiche ufficiali di un elevato numero di valori mancanti o di rilevazioni disomogenee, di cui si   gi  detto in precedenza, e che nella situazione specifica fa s  che numerosi indicatori, che in linea teorica potrebbero descrivere la vulnerabilit , non possono essere inseriti nel database per poi essere utilizzati nelle analisi. Ad esempio, questa considerazione vale per gli indicatori relativi al lavoro minorile, la disoccupazione per classe di et , i dati sulla presenza / assenza di infrastrutture, di cui non sono diffuse statistiche di qualit  adeguata.

Nel nostro studio abbiamo preso in esame i 176 Paesi del mondo con una popolazione superiore a 300.000 abitanti al 2010. Questa scelta è giustificata dalle fatto che per molti dei Paesi esclusi (per lo più arcipelaghi e isole oceaniche o territori oltremare) non esistono statistiche affidabili con uno spettro così ampio come quello considerato nel presente studio. Inoltre, le statistiche relative al Kosovo e al Sud Sudan sono quasi inesistenti poiché queste nazioni hanno dichiarato la loro indipendenza in tempi molto recenti ed il primo censimento è stato condotto solo nel 2011 e 2008 rispettivamente. Per tale motivo, il Kosovo è stato eliminato dallo studio mentre il Sud Sudan è stato preso in esame nelle sole analisi statiche. Pertanto, le nostre analisi saranno condotte su 175 Paesi.

Il database include 131 variabili suddivise in 5 categorie che, per grandi linee, possono essere sintetizzate come segue:

- Popolazione e territorio (densità, urbanizzazione, fertilità, peso relativo della popolazione in particolari classi di età, aspettativa di vita, tassi di dipendenza, territorio)
- Economia (reddito nazionale, apertura verso l'estero, occupazione/disoccupazione, indici di parità di genere nel lavoro, indici di povertà, indebitamento e aiuti dall'estero, produzioni agricole e ambiente, infrastrutture, presenza di specifiche attività estrattive)
- Società e Politica (libertà politiche, diritti civili, regime politico, acquisizione del diritto di voto, frazionamento religioso/etnico/linguistico, forze armate e spese militari, corruzione, rifugiati, rifugiati e popolazione sfollata, criminalità)
- Istruzione (tasso di alfabetizzazione, rapporto docenti/studenti, accesso e successi per i vari gradi di istruzione, indici di parità di genere rispetto all'istruzione)
- Salute (ospedali provinciali, mortalità, mortalità per cause specifiche, malnutrizione e sottopeso, diffusione di alcune patologie, condizioni igienico-sanitarie)

Sono state, inoltre, incluse alcune variabili indicatrici: la classificazione dei Paesi della World Bank rispetto al reddito (high/upper middle/lower middle/ low income) e quelle delle Nazioni Unite rispetto al livello di sviluppo ed alle aree geografiche (developed/developing countries; macro-aree).

Per quanto riguarda le variabili concernenti l'istruzione, ed in particolare i livelli di istruzione conseguiti, abbiamo preso in esame i dati stimati da Barro e Lee (2012): Si tratta di stime fondate sui dati censuari per 146 Paesi dal 1950 al 2010 suddivise per sesso e per classi di età. I dati relativi al frazionamento etnico, linguistico e religioso sono stati tratti dallo studio di Alesina et al. (2003). Infine, per colmare i dati mancanti del reddito nazionale si è ricorso alle stime di Spenser et al. (2012) che hanno armonizzato e ricostruito i dati del GDP pro capite per 210 Paesi dal 1950 al 2015.

2.3 Le analisi preliminare: il trattamento dei dati mancanti

La mancanza di osservazioni su alcune variabili per individui specifici in un dataset riduce la nostra capacità di comprendere le relazioni e le dinamiche dei fenomeni. In passato, la tendenza più diffusa era quella di ignorare gli individui ove le osservazioni erano mancanti ovvero di comportarsi come se il dataset fosse completo. In entrambe le situazioni è evidente che la mancanza di informazioni complete può influenzare le conclusioni di uno studio e soprattutto la generalizzazione dei risultati.

L'approccio più corretto alle analisi statistiche richiede, invece, il trattamento preliminare dei dati che porti alla stima dei dati mancanti sfruttando le informazioni disponibili. In particolare, è possibile tener conto delle relazioni che legano:

- la variabile, caratterizzata da dati mancanti, con altre variabili per le quali sono presenti tutte le osservazioni;
- l'osservazione mancante di una variabile per un determinato individuo con le osservazioni rilevate in istanti temporali anteriori.

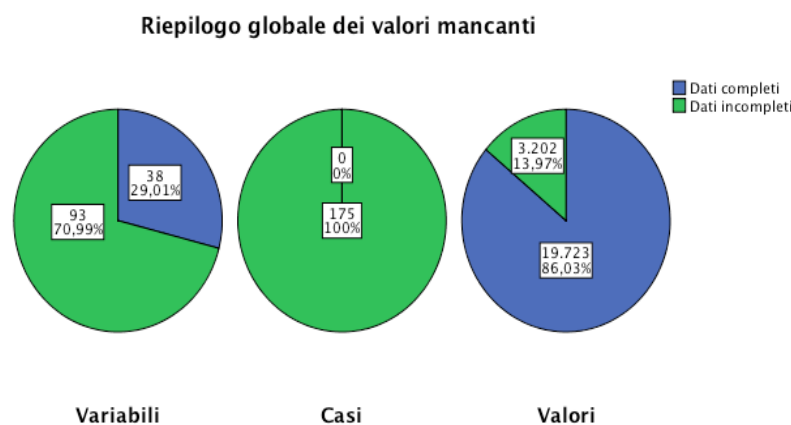


Figura 2.1 Dati mancanti

Il grafico in Figura 2.1 fornisce la situazione iniziale del dataset da noi costruito: circa il 71% delle variabili presenta dei dati mancanti, nessun Paese (=i casi) presenta tutte le informazioni in maniera completa. Considerando l'intera tabella di dati, composta di $131 \times 175 = 22925$ elementi, verificiamo che il 14% delle osservazioni è mancante. Pertanto, nonostante le carenze descritte, si può tuttavia affermare che è disponibile una grande quantità di informazioni su cui fondare i processi di imputazione statistica.

Il grafico in Figura 2.2 illustra i box-plot del numero di valori mancanti per i vari Paesi classificati in funzione del livello di reddito (1=High income, Upper-middle=2, Lower-middle=3, 4=Low). I Paesi che si stagliano al di fuori dei box-plot rappresentano le situazioni più estreme in cui la carenza di informazioni è più consistente. In genere si tratta di Paesi ove la struttura amministrativa, per motivi diversi, non opera in maniera adeguata, ovvero in cui sono intervenuti cambiamenti a livello politico istituzionale di rilievo. Molti di questi Paesi sono

privi di veri e propri istituti di statistica nazionali o di struttura burocratiche diffuse (si pensi ad esempio alla situazione del Sud Sudan che solo di recente a seguito dell'indipendenza ha realizzato il suo primo censimento coadiuvato dai programmi di sostegno delle Nazioni Unite). Caso particolare è costituito da Macao e Hong Kong, le cui rilevazioni sono spesso inglobate nelle statistiche riguardanti la Cina da cui è poi difficile trarre delle stime scorporate.

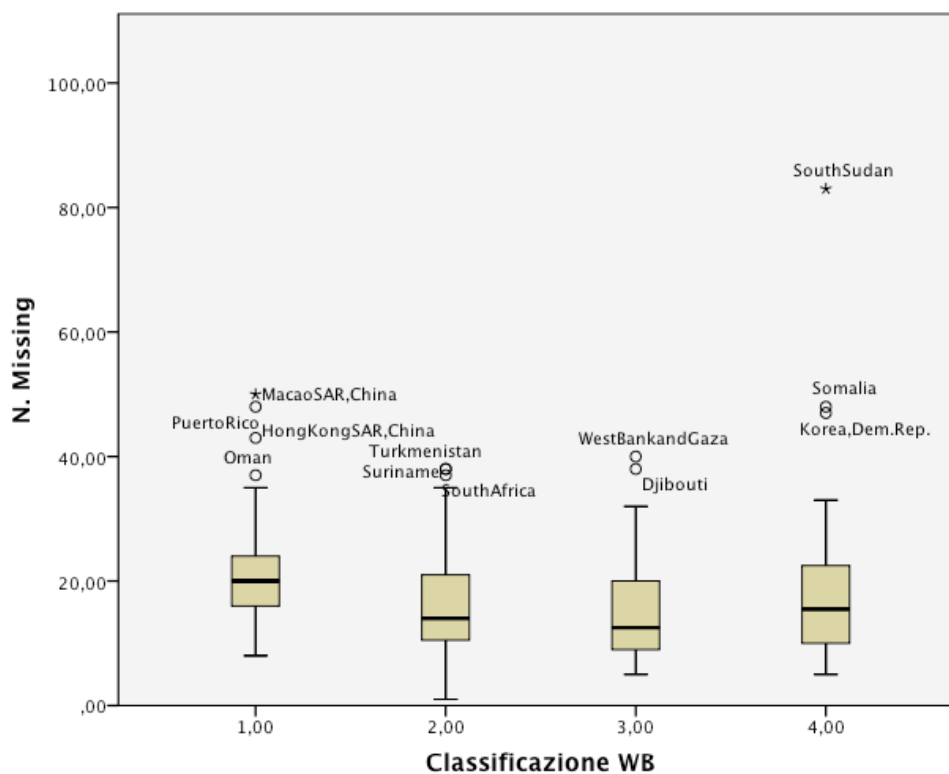


Figura 2.2 Numero di dati mancanti per Paese: box-plot

Per analizzare graficamente la situazione delle osservazioni mancanti è possibile rappresentare la matrice dei dati come in Figura 2.3. In tale rappresentazione ogni casella rappresenta un singolo dato mentre il colore rosso indica l'osservazione mancante. Dalle righe si riconosce il singolo Paese, invece, dalle colonne si identifica la singola variabile. Le righe e le colonne della matrice così rappresentate vengono permutate in modo da avere osservazioni e variabili con un numero di dati mancanti via via crescente. Questa visualizzazione consente di verificare l'addensamento dei dati mancanti e la gravità del problema per certi tipi di informazioni. Si osservi come l'ultimo caso (il Sud Sudan), così come l'ultima variabile (la malnutrizione) mostrino prevalentemente caselle rosse (ovvero dati mancanti). La natura di tale assenza è ovviamente diversa: per la malnutrizione, ad esempio, essa si riferisce sia a Paesi ove le osservazioni effettivamente non sono state rilevate ma il fenomeno è presente, sia a Paesi ove il fenomeno è praticamente assente. Dall'analisi del grafico si nota, inoltre, che circa la metà delle variabili e dei casi sono quasi completi e che la presenza di dati mancanti è piuttosto

diffusa. Risulta quindi necessario effettuare un trattamento statistico preliminare per colmare t

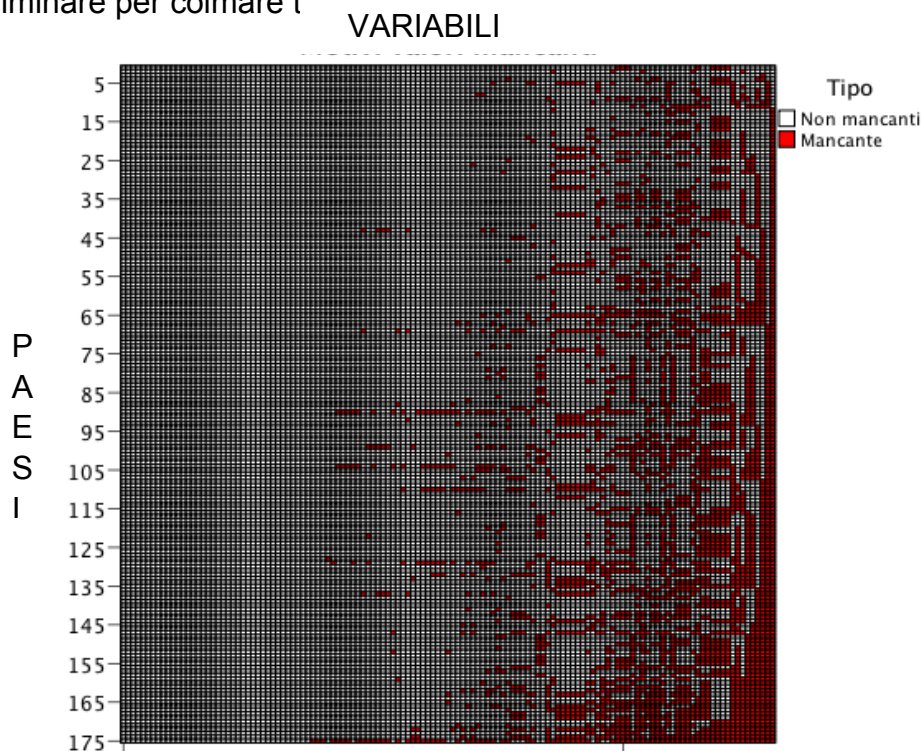


Figura 2.3 Matrice dei dati (caselle rosse=dato mancante; righe=Paesi, colonne=variabili)

Per poter definire la strategia di trattamento più idonea è necessario procedere ad analisi più dettagliate circa il meccanismo di generazione dei dati mancanti. Fra i vari criteri di classificazione dei dati mancanti, il più utilizzato è quello introdotto da Rubin (1976). Vengono distinti tre tipologie: dati mancanti completamente a caso (MCAR), mancanti in modo casuale (MAR), e mancanti non casualmente (MNAR). Questi termini sono definiti da relazioni teoriche tra valori mancanti e le variabili osservate o non osservate.

- 1) MCAR: i dati sono mancanti all'interno della tabella in modo completamente casuale, ovvero non vi è alcun meccanismo che regola l'assenza del dato in determinate posizioni della tabella;
- 2) MAR: il valore del dato mancante è indipendente dagli altri valori osservati;
- 3) MNAR: I dati mancanti dipendono dagli altri dati osservati.

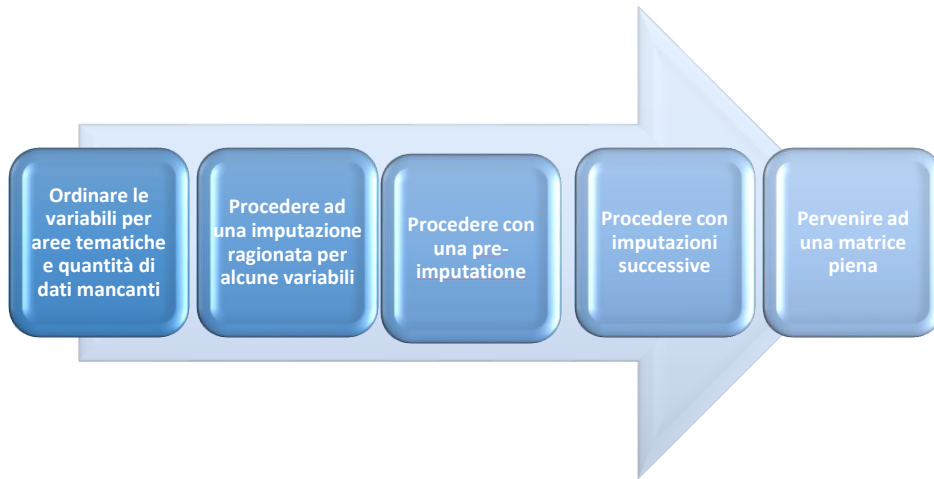
La conoscenza del tipo di processo che ha generato il dato mancante fornisce informazioni sulla tecnica da utilizzare per colmare la carenza e sul livello di distorsione che essi possono esercitare sulle susseguenti analisi statistiche. La categoria MCAR, essendo caratterizzata da uno schema del tutto casuale, provoca un impatto potenzialmente trascurabile ed assimilabile all'errore di misurazione, mentre il tipo MNAR, che vede una struttura interna di dipendenza, esercita potenzialmente il maggiore impatto.

In genere, i dati non mancanti a caso non sono una situazione rara, e sono di difficile trattamento.

Nel nostro caso i dati mancanti sono numerosi e i meccanismi di generazione appaiono piuttosto complessi. Nelle situazioni in cui sono noti (ad esempio: assenza di indagini censuarie recenti) è evidente la non casualità (tutte le osservazioni sono assenti). Pertanto, è stata implementata la seguente procedura ad hoc costituita da più passi operanti in sequenza.

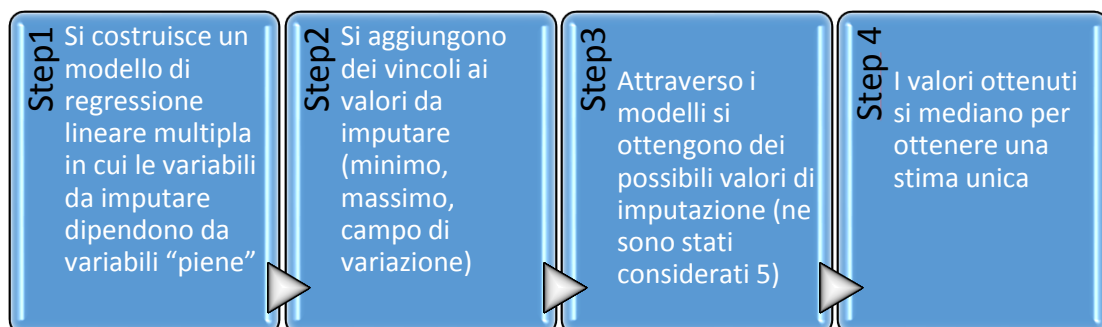
- Le variabili, raggruppate nelle aree tematiche precedentemente identificate, sono state ordinate in rapporto alla percentuale di dati mancanti in esse presenti;
- L'imputazione di un valore mancante, quando possibile, è stata effettuata recuperando l'informazione attraverso ricerche specifiche su siti specializzati e sui siti dei singoli Paesi (ciò ha riguardato, ad esempio, gli aspetti territoriali, le forme di governo, la presenza di diritti civili e politici). Inoltre, sono stati risolti con una imputazione ragionata tutte quelle situazioni ove la definizione stessa della variabile e le caratteristiche del Paese in termini di sviluppo forniva forti indicazioni sul valore della variabile (ad esempio, la malnutrizione in Italia è posta pari a 0).
- Per le variabili con più del 40% di valori mancanti e nei casi in cui era disponibile una serie storica (ovvero dati per gli ultimi due decenni) si è proceduto ad una parziale imputazione attraverso regressioni per serie storiche (33 variabili sono state pre-processate giungendo ad una parziale pre-imputazione);
- Si è iniziato, poi, il processo di stima applicando l'approccio della imputazione multipla vincolata (illustrato di seguito) ai gruppi di variabili appartenenti alla stessa area tematica ed aventi basse percentuali di valori mancanti;
- I risultati ottenuti nei passi precedenti sono stati impiegati per le imputazioni successive. Quando disponibili, sono state anche utilizzate le informazioni provenienti dalle variabili pre imputate sulla base delle serie storiche (passo 3).
- Si è pervenuti alla fine delle iterazioni ad una matrice di dati "piena".

La procedura può essere rappresentata schematicamente come segue.



Come abbiamo visto, l'imputazione multipla vincolata costituisce un passo essenziale nella procedura adottata. Tale metodo non produce una sola imputazione, ma fornisce più valori possibili per lo stesso dato mancante, mimando così la naturale variabilità dei dati. Le varie imputazioni poi possono essere combinate in una sola informazione (ad esempio, considerandone la media) che va a colmare l'osservazione mancante. Il metodo utilizzato per l'assegnazione multipla dei valori mancanti è un modello di regressione multipla vincolato.

Tralasciando gli aspetti puramente metodologici e le relative formulazioni analitiche, possiamo sintetizzare il processo di imputazione mediante uno schema che evidenzia la sequenza di passi: stima del modello di regressione, analisi di coerenza in termini di statistiche descrittive (minimo, massimo, campo di variazione), generazione di 5 possibili valori che possono costituire un sostituto per il dato mancante, calcolo della media dei 5 valori e sostituzione del dato mancante con tale media.



Al fine di mostrare qualche dettaglio operativo, possiamo ripercorrere un caso specifico in dettaglio, illustrando le operazioni svolte per le variabili che disegnano la dimensione associata al territorio coltivato ed alle produzioni agricole. Nella Tabella 2.1 sono riportate dei casi selezionati di alcune variabili

che mostrano una situazione tipica: le ultime tre colonne sono complete mentre i dati mancanti sono distribuiti piuttosto casualmente tra le rimanenti colonne della matrice.

Tabella 2.1 Alcune variabili connesse alla dimensione agricola

Country Name	Cerealyield kg per hectare	Depth of the food deficit kilocalories per person per day	Forest area	Agricultural land	Arable land	Arable land hectares per person
Afghanistan	2.011,141	#NULO!	2.070	58.124	11.947	0,274
Albania	4.781,839	65,000	28.332	43.843	22.847	0,199
Algeria	1.401,126	24,000	#NULO!	17.371	3.150	0,202
Angola	683,097	192,000	46.908	46.836	3.289	0,210
Argentina	5.041,657	25,000	10.743	53.611	13.599	0,922
Armenia	2.088,236	23,000	9.199	60.920	15.748	0,151
Australia	1.723,825	5,000	19.434	51.883	5.541	1,928
Austria	6.194,328	1,000	47.167	38.406	16.600	0,183
Azerbaijan	2.020,626	14,000	11.324	57.670	22.794	0,208
Bahamas The	7.533,333	50,000	51.449	1.469	0.859	0,025
Bahrain	#NULO!	#NULO!	0.697	11,000	1.789	0,001
Bangladesh	4.288,385	118,000	11.078	70.992	59.468	0,051
Belarus	2.813,623	5,000	42.533	43.854	27.279	0,583
Belgium	9.343,835	1,000	22.384	44.848	27.609	0,077
Belize	2.798,584	45,000	61.070	6.883	3.288	0,243
Benin	1.200,581	56,000	40.449	30.064	22.526	0,267
Bhutan	2.469,096	#NULO!	84.623	13.544	2.493	0,133
Bolivia	2.056,751	175,000	52.798	34.123	3.458	0,369
BosniaandHerzegovina	3.852,951	20,000	42.843	42.039	19.686	0,261
Botswana	373,646	227,000	20.029	45.632	0.457	0,132
Brazil	4.040,836	60,000	61.413	32.321	8.313	0,360
BruneiDarussalam	580,087	4,000	72.106	2.163	0.569	0,007
Bulgaria	4.019,885	54,000	36.174	46.536	29.320	0,422
BurkinaFaso	1.062,694	169,000	20.647	44.097	21.930	0,398
Burundi	1.321,468	653,000	#NULO!	88.396	35.626	0,100
Cambodia	2.902,828	138,000	57.183	32.036	22.660	0,278
Cameroon	1.669,295	96,000	42.132	20.308	13.116	0,301
Canada	3.490,573	4,000	34.105	6.958	4.772	1,272

Nel primo passo, quindi, viene stimato un modello di regressione lineare del tipo:

$$Y_j = \beta_0 + \beta_1 X_{1j} + \beta_2 X_{2j} + \dots + \beta_k X_{kj} + u_j, \quad j = 1, \dots, n$$

dove la variabile Y viene spiegata attraverso una relazione lineare che la lega ad una serie di variabili cosiddette esplicative (X_1, \dots, X_k). In particolare, le variabili che nella nostra procedura agiscono come variabili esplicative sono le seguenti:

- Agricultural land (%of land area)
- Arable land (%of land area)
- Arable land (hectares per person)
- GDP per capita constant 2005 US \$
- Macroarea
- Rural population (%of total population)
- Developed/Developing countries
- GNI per capita constant 2005 US \$

Si tratta di variabili per le quali non sono presenti valori mancanti.

Le variabili invece da spiegare (nella formulazione precedente la variabile Y) sono quelle che presentano valori mancanti. Di seguito ne riportiamo la lista ed i vincoli imposti in relazione ai valori minimi e massimi di modo che il valore imputato sia compreso nel campo dei valori ammissibili.

- Depth of the food deficit, kilocalories per person per day (MIN =0.0)
- Food production index (MIN =0.0)

- Cereal yield kg per hectare (MIN=0.0)
- Crop production index (MIN=0.0)
- Forest area, % of land area (MIN=0.0 MAX=100.0)
- Permanent crop land, % of land area (MIN=0.0 MAX=100.0)
- Livestock production index (MIN=0.0 MAX=100.0)

Dai vari modelli di regressione applicati attraverso metodi di imputazione consolidati implementati all'interno del software SPSS 20, si sono ottenute 5 possibili imputazioni che sono state poi combinate, per ottenere una unica stime del dato mancante.

Tabella 2.2 Valori imputati

Country Name	Cereal yield kg per hectare	Depth of the food deficit kilocalories per person per day	Forest area	Agricultural land	Arable land	Arable land hectares per person
Afghanistan	2.011.141	131.360	2.070	58.124	11.947	0.274
Albania	4.761.839	65.000	28.332	43.843	22.847	0.199
Algeria	1.401.126	24.000	0.626	17.371	3.150	0.202
Angola	663.097	192.000	46.908	46.836	3.289	0.210
Argentina	5.041.657	25.000	10.743	53.611	13.599	0.922
Armenia	2.068.236	23.000	9.199	60.920	15.748	0.151
Australia	1.723.825	5.000	19.434	51.883	5.541	1.929
Austria	6.194.328	1.000	47.167	38.406	16.600	0.163
Azerbaijan	2.020.626	14.000	11.324	57.670	22.794	0.208
Bahamas, The	7.533.333	50.000	51.449	1.499	0.899	0.025
Bahrain	8.077.582	60.744	0.697	11.000	1.789	0.001
Bangladesh	4.288.385	118.000	11.078	70.992	59.468	0.051
Belarus	2.813.623	5.000	42.533	43.854	27.279	0.583
Belgium	9.343.835	1.000	22.384	44.848	27.609	0.077
Belize	2.798.584	45.000	61.070	6.883	3.288	0.243
Benin	1.200.581	56.000	40.449	30.064	22.526	0.267
Bhutan	2.469.086	122.107	84.623	13.544	2.493	0.133
Bolivia	2.056.751	175.000	52.798	34.123	3.458	0.369
BosniaandHerzegovina	3.852.951	20.000	42.843	42.039	19.686	0.261
Botswana	373.646	227.000	20.029	45.632	0.457	0.132
Brazil	4.040.836	60.000	61.413	32.321	8.313	0.360
BruneiDarussalam	580.087	4.000	72.106	2.163	0.569	0.007
Bulgaria	4.019.885	54.000	36.174	46.536	29.320	0.422
BurkinaFaso	1.062.694	169.000	20.647	44.097	21.930	0.386
Burundi	1.321.468	663.000	6.697	88.396	35.826	0.100
Cambodia	2.902.828	138.000	57.183	32.036	22.660	0.278
Cameroon	1.669.295	96.000	42.132	20.308	13.116	0.301

Una procedura analoga è stata applicata in maniera sequenziale a tutte le variabili presenti nel nostro dataset. Le variabili riempite nei passi precedenti sono state rese disponibili come possibili variabili esplicative nei passi successivi.

Dopo la prima imputazione la situazione è migliorata come si evince dalla Figura 2.4. La percentuale di dati mancanti si è ridotta al 65,5% e i dati incompleti sono il 7,3%.

Riepilogo globale dei valori mancanti

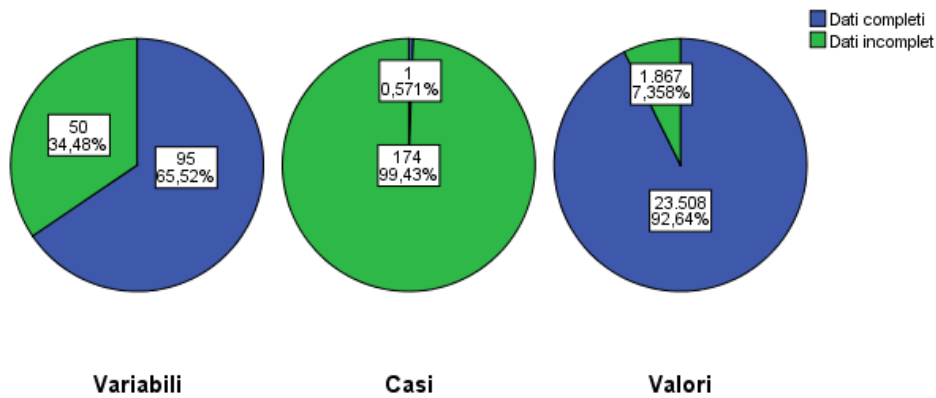


Figura 2.4 Dati mancanti dopo la prima imputazione

Il processo è stato reiterato per tutte le aree tematiche fino a giungere alla completa imputazione del dataset.

Si osservi che 3 variabili non sono state considerate nel processo di imputazione poiché per la loro natura non ha senso procedere con una imputazione statistica. In particolare, si tratta del:

- Growth rate refugee by destination country;
- Growth rate refugee by origin country;
- Refugee by origin country.

Analogamente a quanto visto all'inizio di questo paragrafo, la situazione finale può essere rappresentata con dei grafici a torta per i tre aspetti di interesse: variabili, individui ed intera tabella di dati.

Riepilogo globale dei valori mancanti

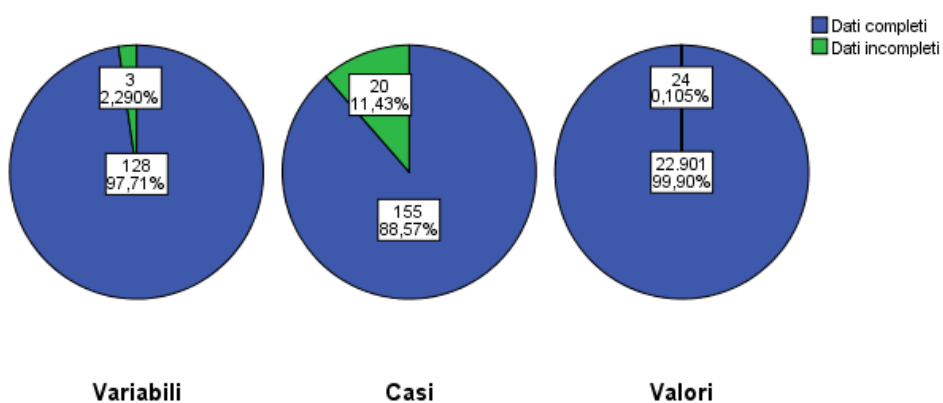


Figura 2.5 Dati mancanti a fine processo

2.4 L'analisi fattoriale multipla e la rappresentazione dei Paesi

Dopo essere pervenuti alla definizione di un database in cui i dati mancanti sono stati imputati secondo le procedure appena descritte, occorre elaborare una visione sintetica di questa ingente quantità di informazioni.

Di solito, per tale scopo, si ricorre ai metodi fattoriali che consentono di sintetizzare sia variabili quantitative che variabili qualitative riducendo la dimensione dello spazio in cui esse sono rappresentabili. L'analisi fattoriale giunge ad una rappresentazione attraverso l'identificazione di fattori o variabili latenti. I metodi fattoriali si dividono in esplorativi e confermativi (Abdi et al. 2013; Escoufier e Pages, 1988, 1994).

Nell'ambito dei metodi fattoriali esplorativi le dimensioni latenti non sono note e definite a priori. In questo caso, quindi, si esplora la struttura di relazione multivariata alla ricerca di dimensioni che siano in grado di catturare le caratteristiche e gli aspetti essenziali del fenomeno. Nelle analisi fattoriali confermativa, invece, si parte da alcune ipotesi circa la presenza di fattori latenti associati ad alcune variabili osservate e attraverso il metodo si verifica la presenza e la capacità di sintesi di tali fattori.

Nel nostro caso, la grande quantità di variabili e la presenza di poche ipotesi circa le variabili osservate e le relative aree tematiche non consente di specificare dimensioni latenti. E' stato perciò utilizzato un approccio ibrido, sostanzialmente esplorativo ma che tiene conto delle aree tematiche introdotte. In particolare, è stata applicata l'Analisi Fattoriale Multipla (AFM).

L'AFM è un metodo fattoriale molto flessibile che si presta al trattamento di numerose strutture di dati. L'AFM, infatti, consente di studiare adeguatamente matrici di dati 3-mode o 3-way, ovvero matrici in cui in cui sugli stessi soggetti sono osservate gruppi di variabili diverse (anche nella natura: quantitative, qualitative, o frequenze) o variabili osservate in tempi, occasioni, o contesti diversi. Per questo motivo, il metodo può essere utilizzato per volumi di dati ove sono presenti più dimensioni ($I \times J_k \times K$) come le matrici a più vie (*multiway*) e le matrici di dati strutturate in gruppi o temi (uno per gruppo, quale il nostro caso).

Indichiamo con \mathbf{X} la matrice *multiway* e con \mathbf{X}_k le sub-matrici o tabelle relative a gruppi di variabili. Poiché le variabili da noi prese in esame, sono tutte di natura quantitativa, l'AFM consiste in una analisi in componenti principali ponderata della matrice \mathbf{X} , in cui ogni gruppo (o blocco o tabella) è pesato per l'inverso della radice quadrata del primo autovalore di \mathbf{X}_k .

La struttura delle variabili all'interno del gruppo non viene mutata, ma allo stesso tempo il sistema di pesi fa sì che nessun tema (nessuna tabella) prevalga nell'analisi, determinando e orientando da sola l'intera analisi media. Questa caratteristica è particolarmente utile e rilevante nel nostro studio dal momento che le variabili che attengono alla dimensione economica sono certamente dominanti e avrebbero, in assenza di tale sistema di pesi, sicuramente orientato l'intera analisi, oscurando gli altri temi e dimensioni.

Inoltre, questo approccio fornisce come ulteriore vantaggio la possibilità di interpretare i risultati secondo i criteri classici dell'analisi in componenti principali e su cui torneremo in seguito.

L'AFM è una analisi in più passi che produce una molteplicità di informazioni di sintesi sui fenomeni in esame.

Volendo semplificare l'AFM si compone di:

- 1) Analisi parziali (o analisi fine)
- 2) Analisi globale (o analisi media)
- 3) Analisi dell'inter-struttura

2.4.1 Analisi parziali

In questa prima fase ogni tabella (ovvero ogni tema) viene analizzata singolarmente con il metodo fattoriale idoneo. Nel nostro caso, poiché le variabili in esame sono quantitative, si è applicata l'analisi delle componenti principali alle sottotabelle, \mathbf{X}_k , di dimensione $(I \times J_k)$. Date le diverse unità di misura e scale di grandezza delle nostre variabili, si è inoltre scelto di procedere alla loro standardizzazione. Si osservi che l'esclusione di alcune variabili, come la popolazione totale, è diretta ad evitare il pesante condizionamento dei risultati finali che, altrimenti, potrebbero ridursi al semplice ordinamento dei Paesi in termini di taglia, dal più grande al più piccolo. Dall'analisi delle sottotabelle si ottengono le rappresentazioni delle variabili nello spazio degli individui (il cosiddetto cerchio delle correlazioni) e le rappresentazioni dei Paesi nello spazio delle variabili.

Il cerchio delle correlazioni

In questa rappresentazione ogni variabile è descritta da un vettore (cioè un segmento orientato) che parte dall'origine degli assi. Le coordinate rispetto a ciascun fattore indicano la correlazione della variabile col fattore.

Nello spazio originario ciascun vettore ha una lunghezza pari ad 1 (valore della varianza di una variabile standardizzata), mentre nello spazio ridotto del cerchio delle correlazioni ciascun vettore ha una lunghezza proporzionale al suo peso nella determinazione degli assi fattoriali.

Quanto più la coordinata rispetto ad un fattore in valore assoluto si approssima all'unità (ovvero il vettore è vicino al bordo del cerchio delle correlazioni) tanto più la variabile è correlata con quel fattore e quindi contribuisce alla sua spiegazione.

Nel cerchio delle correlazioni è possibile visualizzare anche le correlazioni fra le variabili, guardando agli angoli che i vettori corrispondenti formano. Infatti la correlazione fra due variabili è pari al coseno dell'angolo da loro formato. In particolare, due vettori che presentano:

- un angolo inferiore a 90° , indicano due variabili con una correlazione positiva tanto più forte quanto più piccolo è l'angolo;
- un angolo di circa 90° , indicano due variabili fra loro linearmente indipendenti;

- un angolo superiore a 90° , indicano due variabili che sono negativamente correlate, con una correlazione tanto più forte quanto più l'angolo è grande (tendente a 180°).

Attraverso la lettura del cerchio delle correlazioni (e dei contributi assoluti) è possibile interpretare le dimensioni latenti emerse.

I fattori latenti sono il risultato di una combinazione lineare delle variabili osservate. Essi sono determinati in misura maggiore dalle variabili aventi le corrispondenti coordinate più elevate. In particolare, nella rappresentazione grafica del cerchio delle correlazioni, per ciascun asse si cercano, nelle due direzioni (verso positivo e verso negativo), le variabili che sono ad esso più vicine e con una lunghezza del vettore maggiore. Queste sono quelle più rilevanti e quindi si tenta di comprendere quale fenomeno complesso esse descrivano.

La mappa delle osservazioni

Dopo avere interpretato le dimensioni latenti emerse è possibile rappresentare le osservazioni (i Paesi, nel nostro caso) nello spazio delle variabili. In questo spazio, ogni osservazione è raffigurata come un punto sul piano (nel caso di sole due dimensioni latenti) o nello spazio (nel caso in cui due sole dimensioni non fossero sufficienti a spiegare la struttura dei dati). Due o più punti vicini in questo spazio, indicano osservazioni che hanno un profilo simile rispetto alle variabili che determinano i fattori latenti. La distanza tra due punti è di tipo euclideo e quindi essa è di facile lettura sui grafici ottenuti.

Si ricordi, infine, che le analisi svolte prescindono dalla stretta dimensione geografica. La vicinanza sulla mappa implica similitudine delle caratteristiche dei paesi. Così può accadere che Paesi geograficamente limitrofi abbiano caratteristiche simili e quindi siano vicini anche sulle mappe fattoriali, ma può accadere anche che Paesi molto lontani geograficamente si ritrovino vicini sulla mappa fattoriale

E' bene sottolineare, inoltre, che tali rappresentazioni sono proiezioni di punti definiti originariamente in spazi multivariati (sono, infatti, le misurazioni su numerose variabili) rispetto a spazi di dimensioni inferiori. Per comprendere come opera la tecnica, possiamo immaginare un esempio elementare: schiacciamo (proiettiamo) un punto definito in uno spazio tridimensionale su un piano (ovvero uno spazio bidimensionale). In tale operazione di proiezione una certa quantità di informazione va inesorabilmente persa: il metodo dell'ACP assicura che tale perdita sia la più piccola possibile.

Ovviamente, le rappresentazioni ottenute sono solo delle approssimazioni della situazione iniziale. La relativa qualità può essere giudicata sia in termini globali, considerando la quota di varianza spiegata dalla soluzione fattoriale, sia singolarmente in relazione alle variabili e ai fattori prendendo in esame i contributi relativi.

2.4.2 Analisi globale (o analisi media)

Dalle analisi parziali precedenti si ricavano i pesi da attribuire a ciascuna tabella associata ad una delle aree tematiche. Alla tabella pesata si applica l'usuale Analisi in Componenti Principali. In questo caso, vengono estratti dei fattori comuni ai diversi gruppi di variabili. Pertanto nel cerchio delle correlazioni è possibile esaminare l'intera struttura derivante dall'associazione che c'è tra i gruppi e all'interno dei gruppi. Le coordinate delle osservazioni sono una media ponderata delle coordinate delle analisi parziali (in tal senso, l'analisi si può definire "media").

Sia nel cerchio delle correlazioni che nella mappa fattoriale valgono le stesse regole di lettura e interpretazione precedentemente espresse.

2.4.3 Analisi interstrutturale

Obiettivo di questa fase dell'analisi è quello di individuare le tabelle che presentano strutture simili rispetto alle osservazioni. Ogni tabella viene trattata come un tutt'uno (e viene rappresentata come un punto nello spazio) e si giudica la similarità fra le tabelle.

2.5 I principali risultati

Dal dataset iniziale sono state selezionate 115 variabili disponibili per 175 Paesi. Queste variabili sono state divise in 5 ambiti:

- Popolazione (P)
- Istruzione (I)
- Economia (E)
- Salute (S)
- Società e Politica (SP)

L'elenco completo delle variabili diviso nelle diverse aree tematiche è riportato nell'Appendice A1. La similarità della struttura di correlazione delle aree tematiche può essere valutata attraverso l'analisi inter-strutturale sulla base dei coefficienti di similarità RV (Tabella 2.3) e, da un punto di vista grafico, attraverso la mappa riportata in Figura 2.6. I coefficienti di similarità sono compresi tra 0 (max dissimilarità) e 1 (=max similarità ovvero coincidenza dei gruppi).

Tabella 2.3. Coefficienti RV di similarità tra i gruppi

Gruppi	1	2	3	4	5	AFM
1	1,000					
2	0,696	1,000				
3	0,615	0,528	1,000			
4	0,730	0,629	0,520	1,000		
5	0,501	0,433	0,472	0,425	1,000	
AFM	0,872	0,810	0,793	0,728	0,805	1,000

Si nota che le varie aree tematiche presentano un buon grado di coerenza complessiva: i coefficienti di similarità sono tutti superiori a 0,5 e i punti nel grafico sono abbastanza vicini. Ciò significa che le aree tematiche, sebbene coerenti fra loro, non si riferiscono tutte alla stessa dimensione latente. Unica eccezione è costituita dall'area tematica relativa agli aspetti politici che si discosta un po' di più dalle restanti aree (sebbene il coefficiente RV non scenda al di sotto di 0,425). Interessante anche notare come i coefficienti RV dei gruppi con la soluzione dell'AFM siano molto elevati per tutte le aree tematiche, indicando quindi la bontà della soluzione AFM.

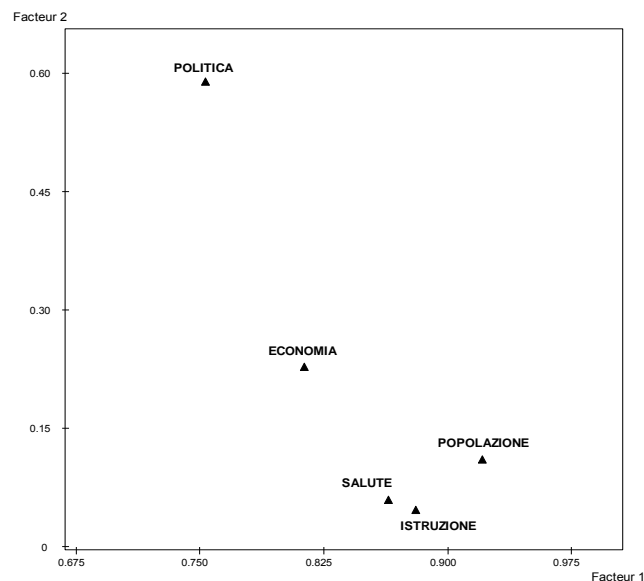


Figura 2.6 Rappresentazione delle aree tematiche

Prima di presentare i risultati dell'analisi globale si illustrano i risultati delle **analisi parziali** che poi combinati insieme forniscono il quadro complessivo.

1) Popolazione

Esaminando il cerchio delle correlazioni definito dai primi due assi fattoriali (che congiuntamente spiegano il circa il 68,3% della variabilità totale) si possono desumere il ruolo delle variabili che maggiormente li determinano (Figura 2.7). Il primo asse è fortemente orientato dalla speranza di vita media opposta agli alti tassi di fertilità e di dipendenza intergenerazionale (sia totale che specifica dei bambini). Dallo stesso lato troviamo i più elevati tassi di crescita della popolazione con la percentuale di popolazione giovanile tra i 15 e 35 anni. Quest'ultima variabile ha un ruolo significativo anche nella spiegazione del secondo asse. Ad essa si contrappone, nella direzione positiva del secondo asse, il rapporto di femminilità della popolazione. Nella mappa fattoriale, quindi, i Paesi nella parte sinistra del grafico saranno caratterizzati da una popolazione più longeva, quelli nella parte in alto a

sinistra presenteranno con una quota di donne superiore alla media, mentre i Paesi che si posizioneranno sul lato destro avranno una popolazione con alta fertilità e con fasce di giovani ampie (ovvero con una piramide della popolazione tipica dei Paesi in via di sviluppo o in transizione demografica). Tale idea è rafforzata anche dalla presenza, sul lato destro del cerchio delle correlazioni, della quota di popolazione rurale. Sul lato sinistro del grafico si trovano poi la densità della popolazione e la popolazione nelle mega-cities, mentre la concentrazione della popolazione nelle grandi città, invece, è vicina all'origine degli assi. Tali variabili non sono ben raffigurate dai due assi fattoriali e quindi non sono rilevanti nel determinare la rappresentazione finale. Invece, esse hanno un ruolo significativo nel caratterizzare il quarto asse fattoriale, che qui non viene preso in considerazione.

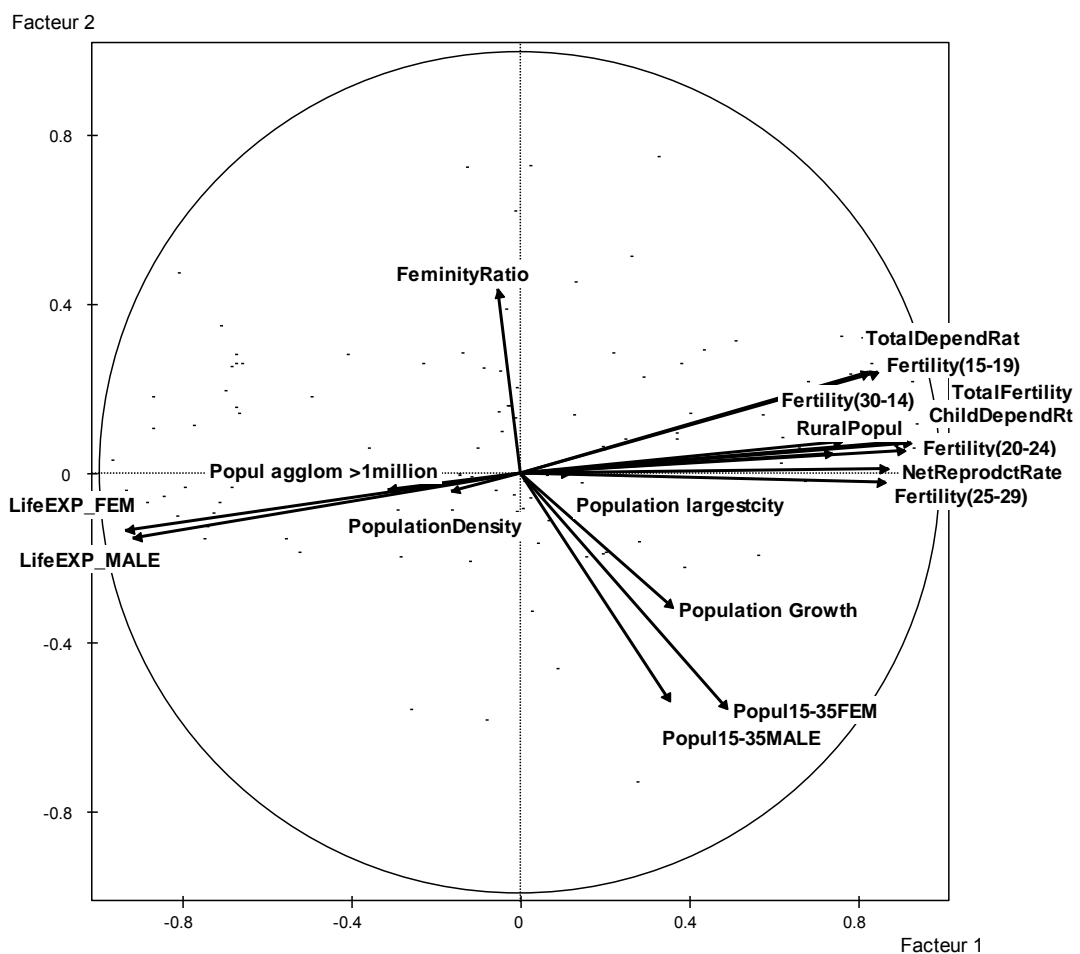


Figura 2.7 Cerchio delle correlazioni "area tematica: Popolazione"

I Paesi vengono rappresentati nella mappa fattoriale costruita sulla base dei primi due fattori (Figura 2.8). In alto a destra, si collocano per lo più i Paesi africani che possiedono una struttura della popolazione con un consistente numero di bambini ed elevati tassi di fertilità. Dal lato opposto, in alto a sinistra, si posizionano i Paesi sviluppati e industrializzati, con peso relativo più elevato della popolazione più anziana (primi fra tutti il Giappone e l'Italia).

Si osservi, in basso al centro, come la collocazione di un gruppo di Paesi arabi (Emirati Arabi, Kuwait, Oman, Qatar e Bahrain) si stacchi notevolmente da tutti gli altri Paesi. La loro posizione è dovuta principalmente ai bassi rapporti di femminilità causati dai forti squilibri demografici prodotti dall'immigrazione maschile legata all'importazione di manodopera per l'industria petrolifera.

2) Istruzione

Per quel che riguarda questa area tematica, bisogna innanzitutto precisare che i primi due assi spiegano poco più della metà dell'inerzia totale, con un forte contributo dato dal primo asse che da solo raccoglie circa il 40% della variabilità totale. Il primo asse oppone da sinistra a destra elevati tassi alfabetizzazione, adulta e giovanile, una discreta parità di genere nei vari gradi dell'istruzione, a alti rapporti di docenti e alunni specialmente nei gradi inferiori dell'istruzione (ovvero Paesi dove ci sono pochi insegnanti per molti studenti). È interessante notare come la parità di genere in fase di iscrizione nella scuola primaria sia opposta a quella nella scuola secondaria e terziaria. Ciò caratterizza fortemente i Paesi poveri e in via di sviluppo rispetto a quelli più sviluppati in cui la presenza femminile nelle scuole superiori è più elevata. Al secondo asse contribuisce, anche se aggiungendo solo 11,5% alla spiegazione complessiva, la quota di GDP destinato alla spesa pubblica per l'istruzione.

Esaminando la mappa fattoriale dei Paesi è bene ricordare che la maggior parte della variabilità è spiegata dal primo asse, che oppone a sinistra i Paesi con elevati tassi di scolarità ed istruzione (specialmente di grado superiore) ai Paesi con bassi tassi di scolarità e di istruzione.

Si noti la posizione di Timor-Leste e Lesotho, che si staccano dagli altri Paesi, e la posizione di Cuba, che invece si colloca vicina alla Danimarca. Tale effetto è causato dalla più elevata quota di reddito nazionale investita nell'istruzione pubblica che rende questi Paesi più simili ai Paesi più avanzati e quindi produce il loro allontanamento dai Paesi a loro simili per altri indicatori.

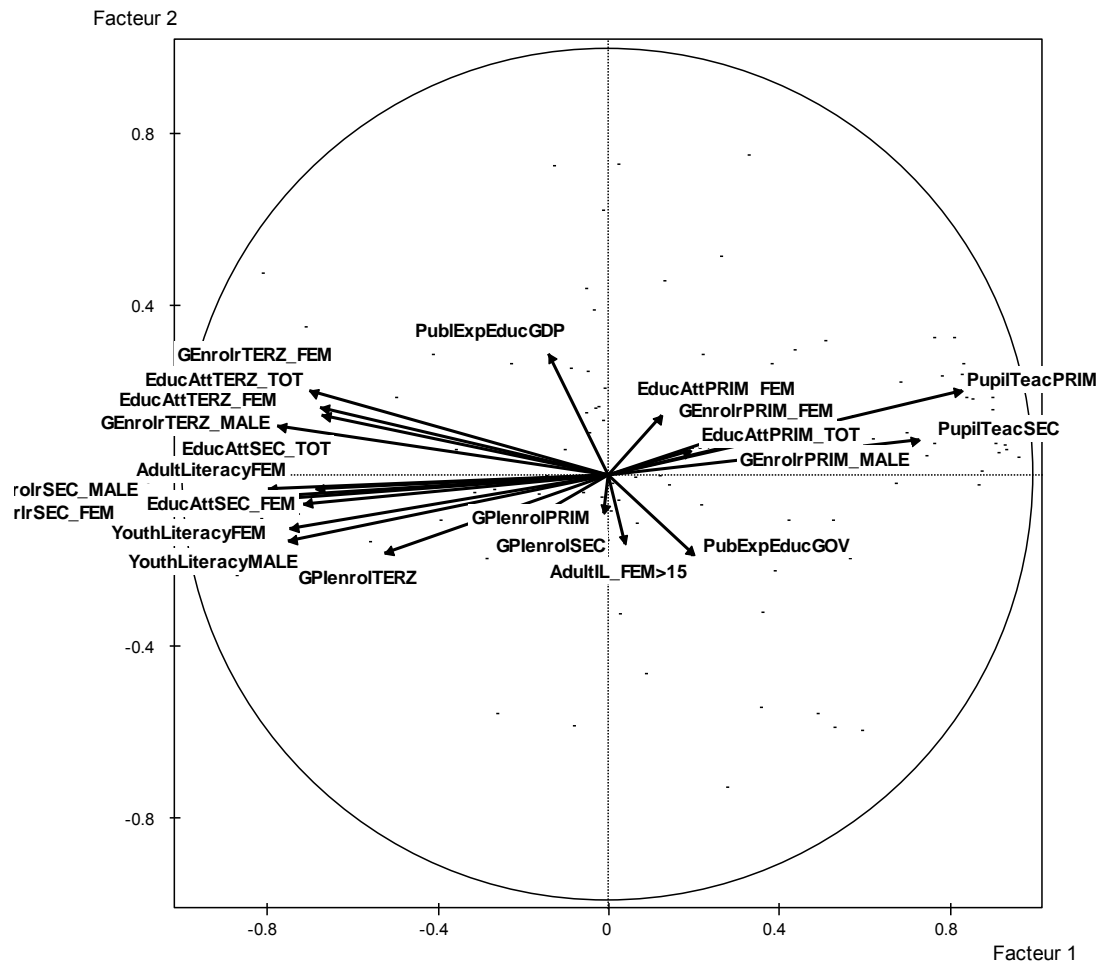


Figura 2.9 Cerchio delle correlazioni “area tematica: Istruzione”

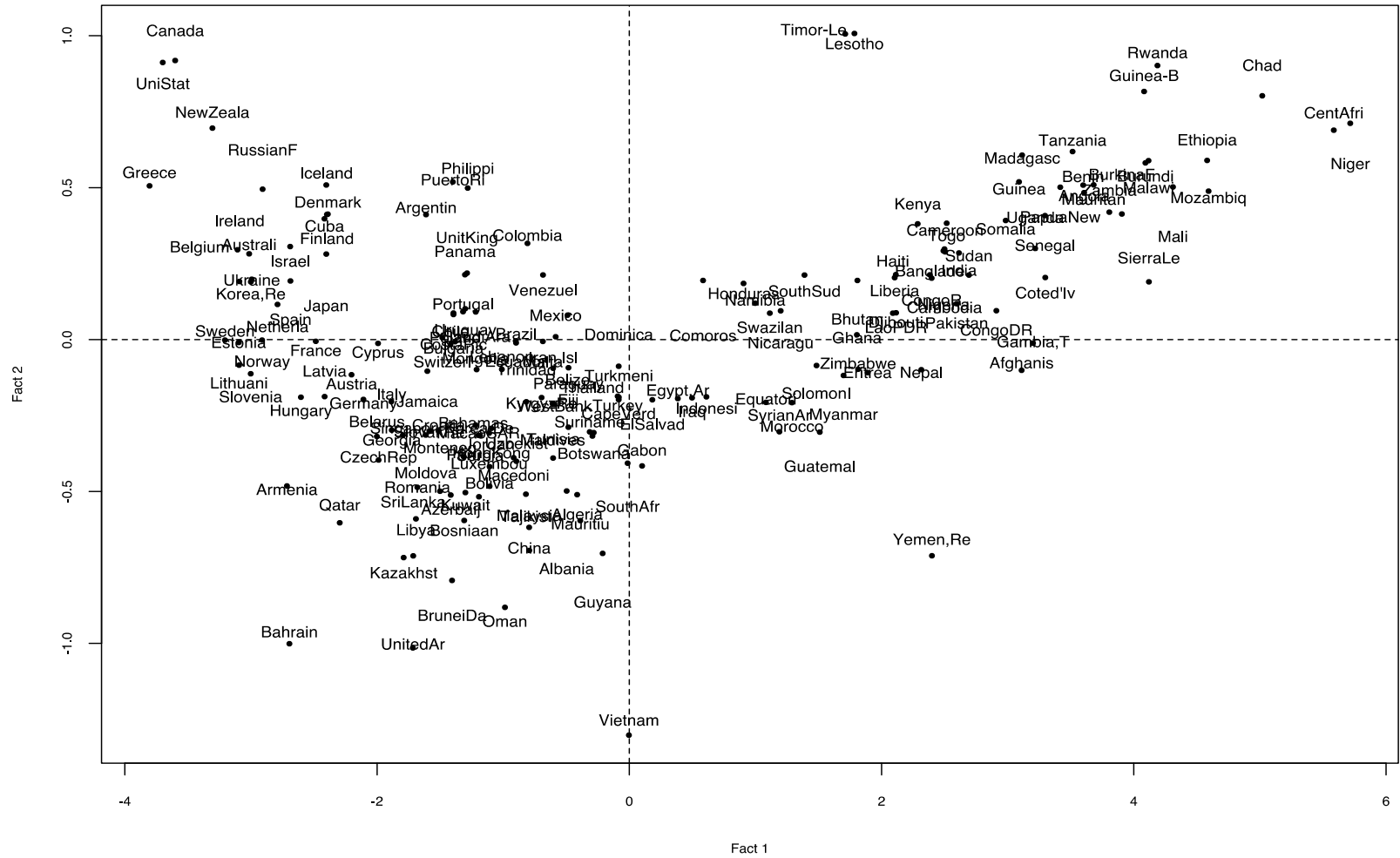


Figura 2.10 Mappa fattoriale “area tematica: Istruzione”

3) *Economia*

Come abbiamo già discusso in precedenza la dimensione economica è generalmente riconosciuta come il driver che guida la vulnerabilità di un Paese. La debolezza economica e le crisi umanitarie sono strettamente legate in un processo che si autoalimenta. Paesi con strutture economiche deboli non riescono a far fronte alle azioni di prevenzione e di rafforzamento delle capacità dirette ad attenuare l'impatto di disastri naturali e quindi ne subiscono gli effetti in misura estrema. D'altra parte il susseguirsi di disastri e quindi di crisi umanitarie fa sì che la struttura economica resti debole non essendo in grado di recuperare gli effetti negativi. La debolezza economica assieme alle spinte demografiche ed al tipo di governo è poi un elemento che gli analisti politici indicano come fattore favorente le instabilità politiche ed i conflitti veri e propri.

Questa è l'area più ricca di indicatori (38 variabili) ed eterogenea. Per tale motivo i primi due assi spiegano solo il 32,9% circa della variabilità totale. Per pervenire ad una migliore rappresentazione, è quindi necessario ricorrere anche al terzo fattore (per tale motivo si proporrà più innanzi una sintesi tridimensionale dei dati).

Il primo fattore descrive la dimensione latente associata al benessere economico, ed oppone chiaramente, a sinistra, i Paesi ricchi e industrializzati a quelli poveri. I primi sono caratterizzati oltre che da un elevato reddito pro-capite anche da un largo accesso a internet, diffuso accesso all'elettricità e uso dei cellulari, e quindi ad elevato uso di tecnologia sia per la vita quotidiana che per la produzione. I secondi, invece, presentano elevate disegualianze, un deficit di cibo, elevati tassi di malnutrizione. Sempre sul primo fattore, ma parzialmente anche sul secondo, troviamo opposti i Paesi che riescono a utilizzare il mercato dei capitali per collocare loro debito e che in parte risultano anche prestatori, opposti a quelli che ricorrono ai prestiti del Fondo Monetario internazionale.

Il secondo asse è, invece, caratterizzato dal livello di partecipazione femminile al mercato del lavoro. Tale variabile, secondo alcune teorie dello sviluppo, è un elemento trainante ed un acceleratore dello sviluppo economico, essendo connessa ad un miglioramento della struttura economica (da agricola/familiare a industrializzata e moderna), dell'educazione (più estesa), della struttura demografica (più bassa fertilità), e delle condizioni di salute (più bassa mortalità).

Infine, il terzo fattore, per il quale non si riporta la rappresentazione grafica, contribuisce per un ulteriore 8.5% alla spiegazione complessiva ed è fortemente caratterizzato dalle variabili direttamente legate alla misura della ricchezza: il prodotto interno lordo, il reddito nazionale così come dalla produzione di petrolio e dal consumo di energia elettrica. Questo asse è importante per la raffigurazione finale dei Paesi, in quanto, come vedremo, consente di staccare i Paesi produttori di petrolio ed in genere quei Paesi che sebbene dotati di ricchezze presentano enormi squilibri interni.

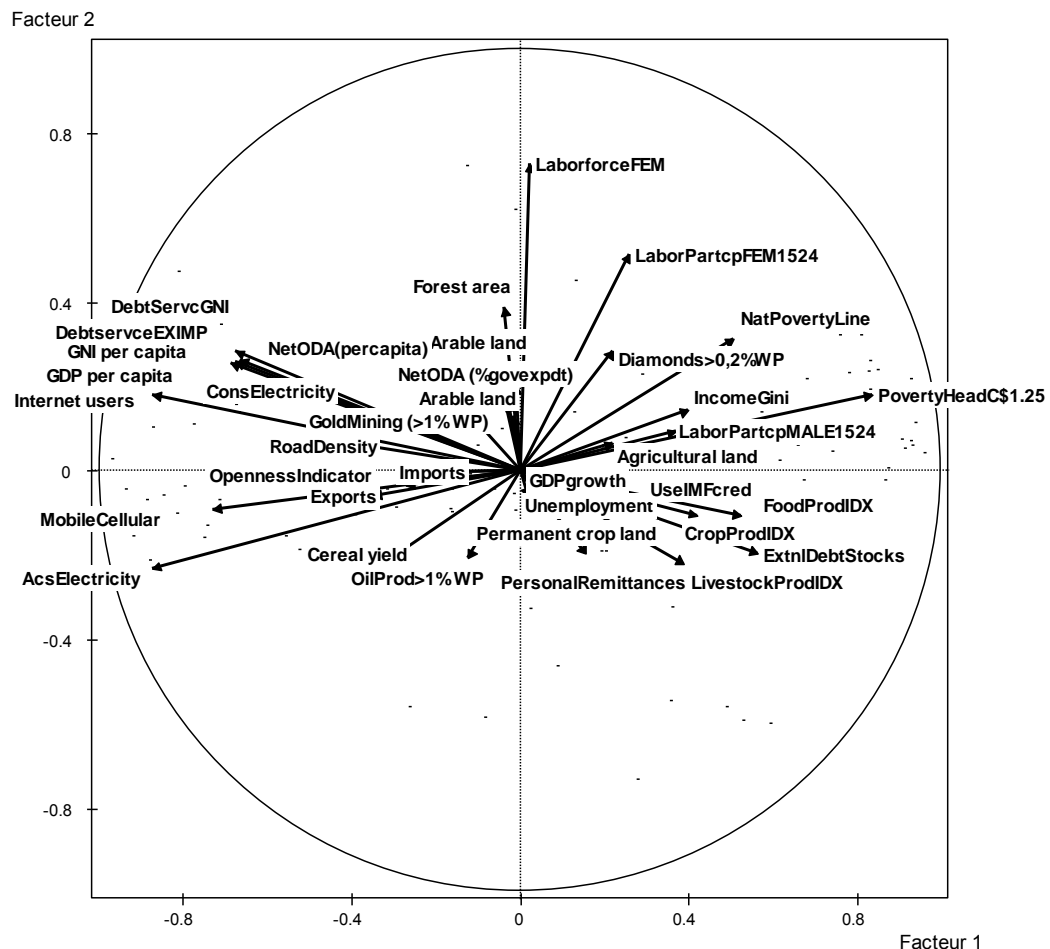
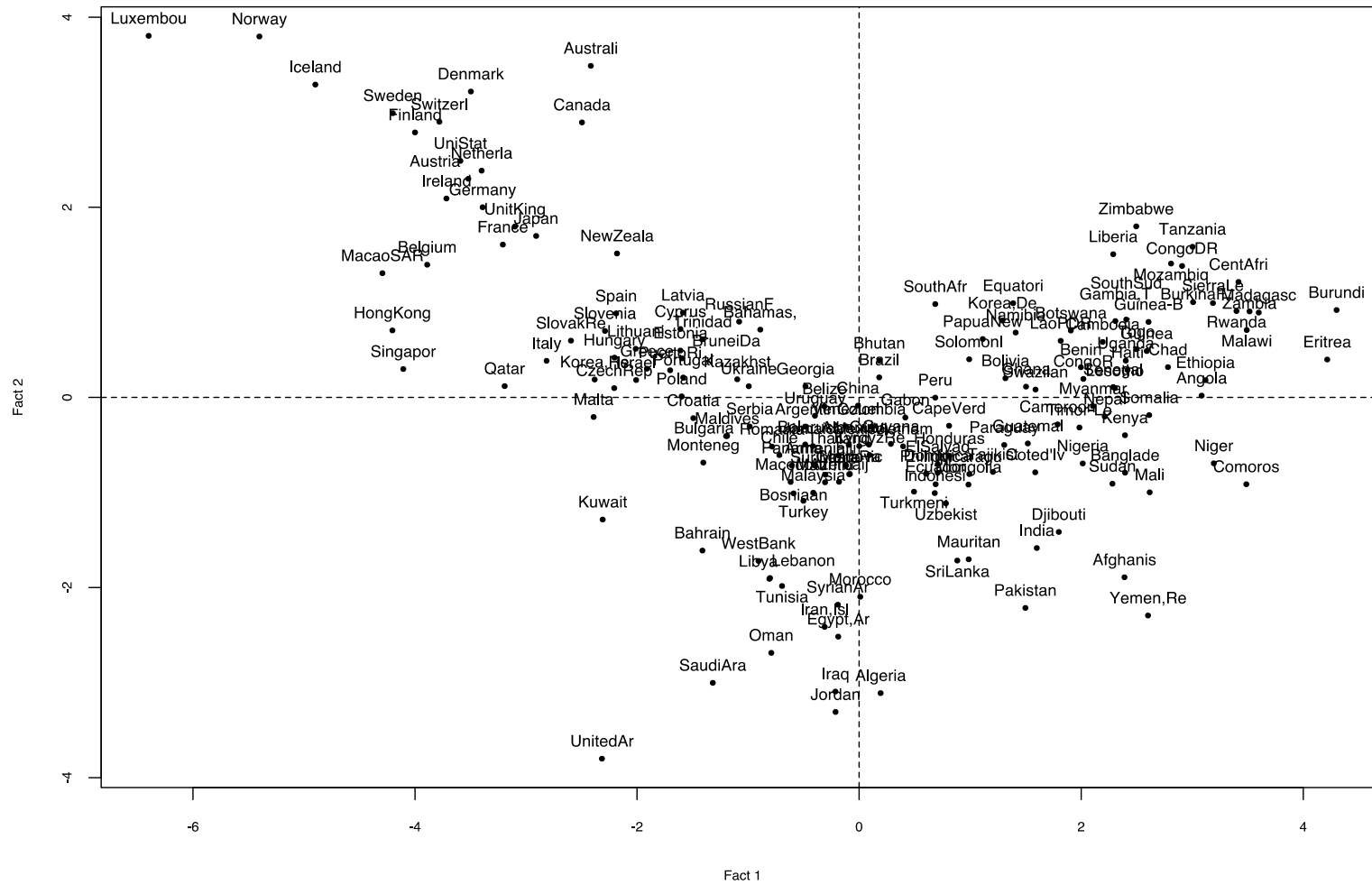


Figura 2.11 Cerchio delle correlazioni “area tematica: Economia”

La mappa fattoriale dei Paesi mostra in alto a sinistra il gruppo dei Paesi del Nord Europa con il Canada e gli Stati Uniti, l’Australia e la Nuova Zelanda ed infine il Giappone. Più in basso invece si collocano i Paesi economicamente sviluppati ma con livelli di benessere economico e sociale inferiore rispetto a quello delle economie più avanzate. In basso sinistra si posizionano alcuni Paesi arabi, molto ricchi ma con bassissimi livelli di partecipazione femminile al mercato del lavoro. Nel primo quadrante in alto a destra, troviamo i Paesi africani più poveri ove sono presenti elevati livelli di diseguglianze sociali ed economiche ed elevate quote di popolazione sotto la soglia di povertà. All’estremo nella parte bassa del grafico, infine, si collocano numerosi Paesi musulmani, alcuni dei quali con grandi diseguglianze e che sono coinvolti in conflitti interni ed esterni (questi Paesi come vedremo successivamente, verranno poi isolati rispetto al terzo asse).



4) Salute

Il cerchio delle correlazioni relativo all'area tematica della salute mostra essenzialmente la presenza di una unica dimensione latente catturata dal primo asse fattoriale, a cui corrisponde il 67.6% della variabilità totale. Il secondo fattore, invece, spiega circa il 12% della variabilità totale.

Il primo asse oppone le migliori condizioni igienico-sanitarie (acqua, vaccinazioni, igiene) e le più elevate spese destinate alla sanità, agli elevati tassi di mortalità infantile e non, alla malnutrizione e sotto-nutrizione. Il secondo asse fattoriale poi è caratterizzato maggiormente dall'elevata presenza dell'HIV fra la popolazione.

La mappa dei Paesi evidenzia la presenza di questa unica dimensione latente (che attraversa diagonalmente il primo e il terzo quadrante), che mette in ordine i Paesi ponendo all'estremo i Paesi africani ove più elevata è la prevalenza dell'HIV, della mortalità infantile e della malnutrizione.

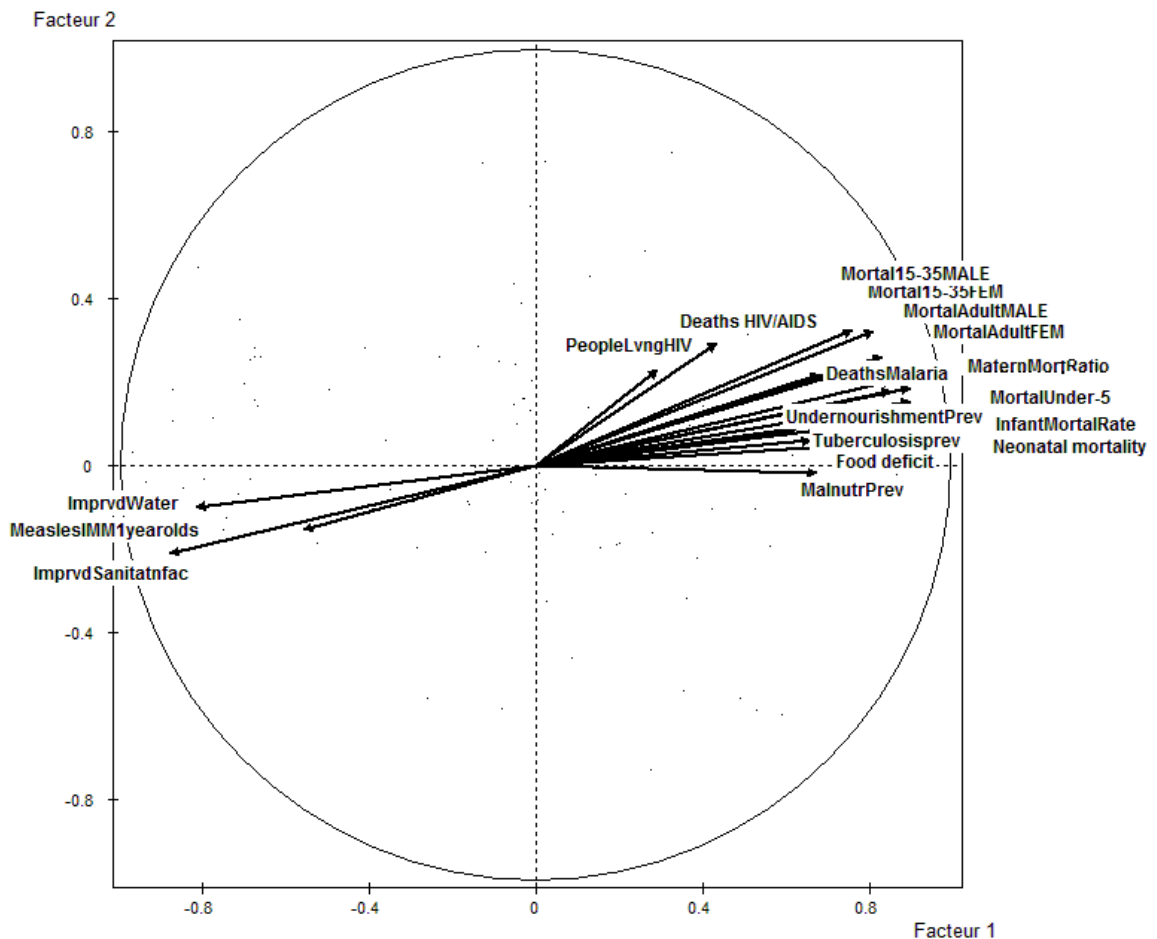


Figura 2.13 Cerchio delle correlazioni "area tematica: Salute"

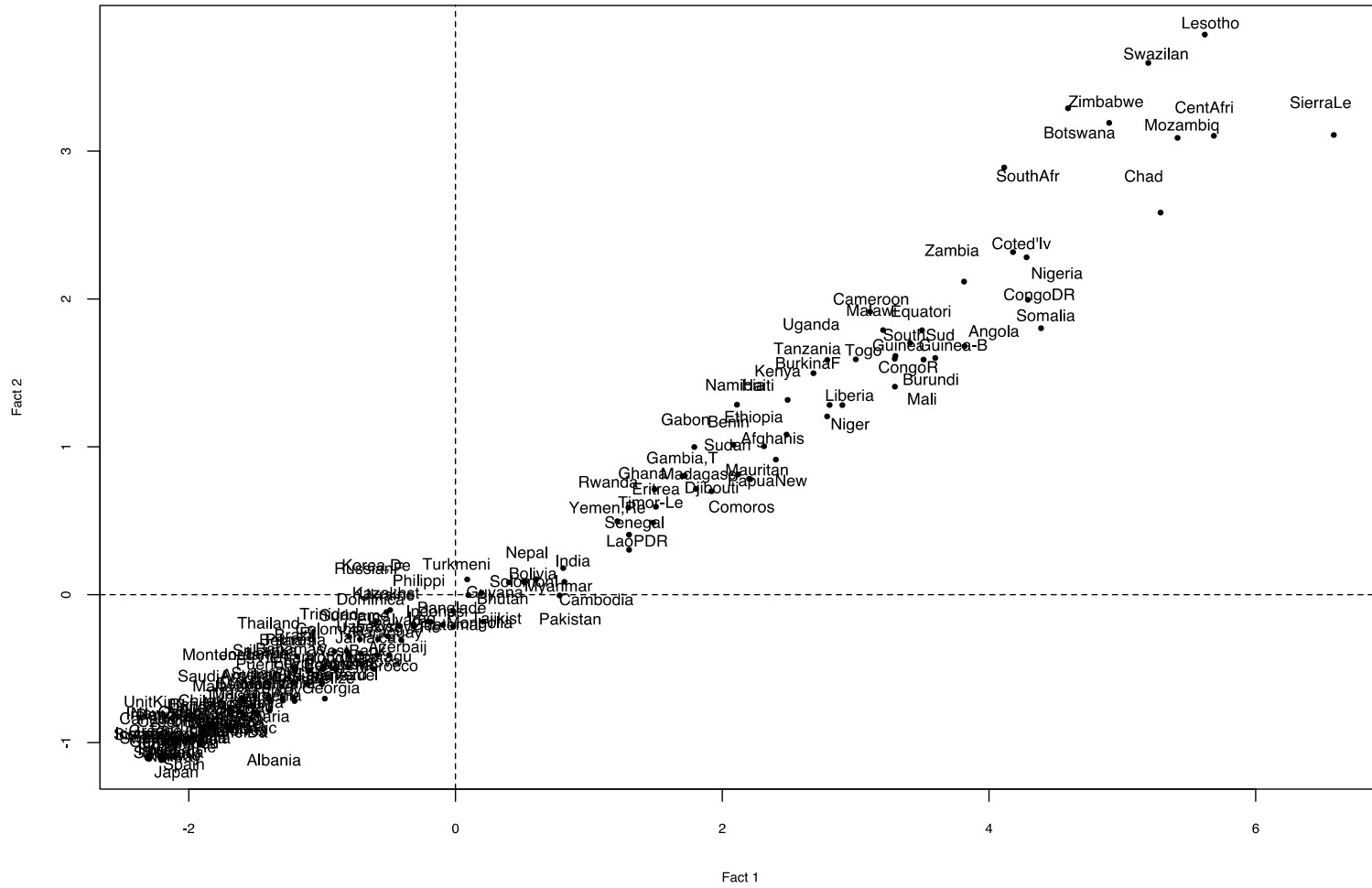
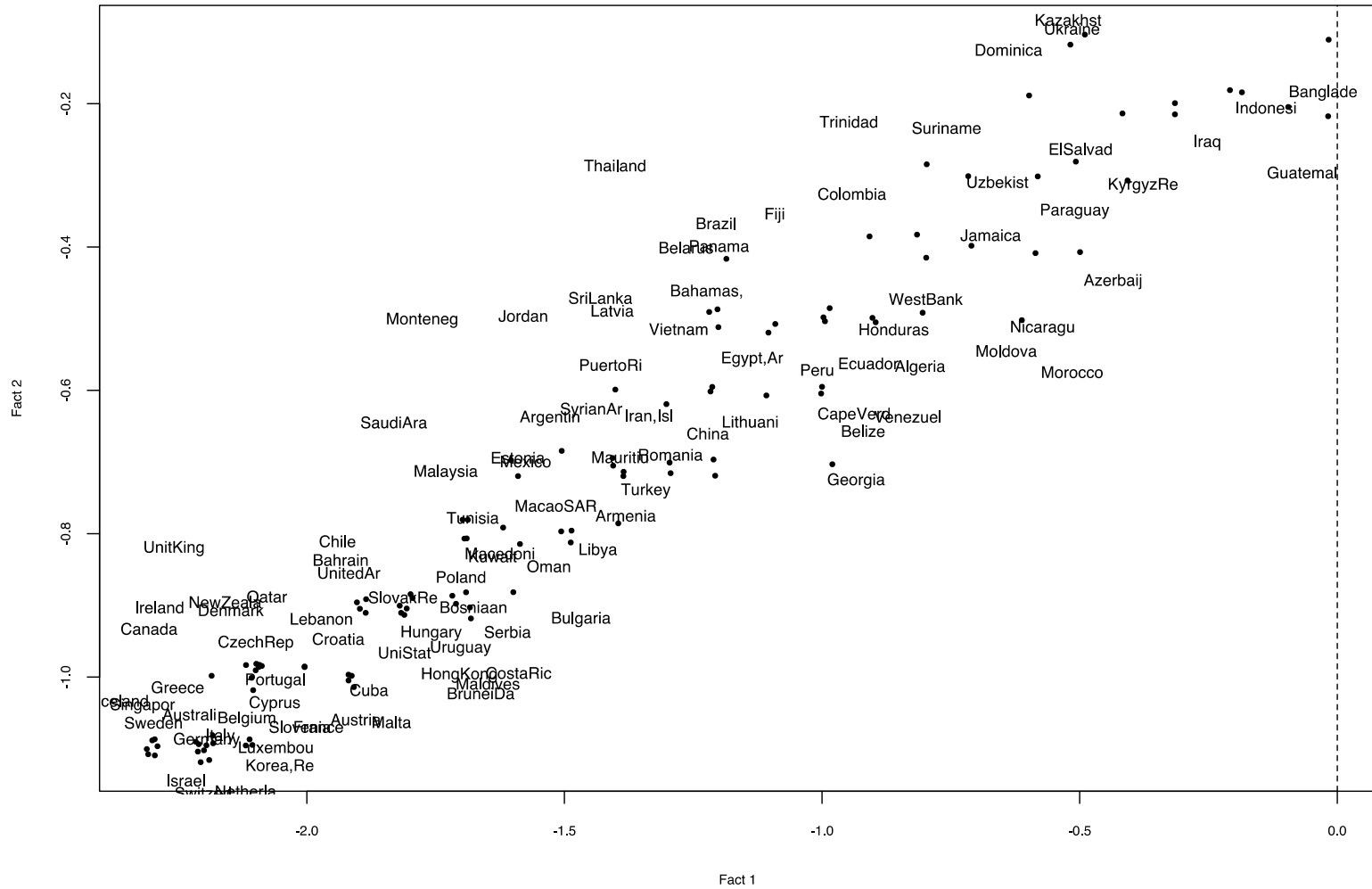


Figura 2.14.a Mappa fattoriale “area tematica: Salute



5) Società e Politica

Le variabili relative agli aspetti politici presentano una maggiore dispersione ed una minore coerenza. In realtà, in questo gruppo tematico sono stati considerati indicatori di varia natura ed origine: punteggi (come le valutazioni sui diritti politici e libertà civili), descrittori di caratteristiche della popolazione, indicatori sulla presenza di una struttura militare. Si osservi che la direzione delle scale utilizzate da alcuni indici è opposta a quella attesa: a) i punteggi per la variabile Libertà Civili e Diritti Politici assume valore da 1 a 7, con 1 =più libero e 7=meno libero; b) l'indice di corruzione rappresenta la percezione del fenomeno ed assume valore tra 0 e 100: 0=elevata percezione della corruzione, 100=bassa percezione del fenomeno.

Il primo e secondo asse spiegano complessivamente circa il 47% della variabilità totale.

Il primo asse oppone un elevato sviluppo umano, bassi livelli di corruzione e la significativa presenza delle donne in politica agli elevati tassi di frazionamento etnico e linguistico.

Il secondo asse oppone la presenza di una popolazione cristiana associata da una struttura sociale caratterizzata da un consistente frazionamento religioso alla presenza di una consistente parte della popolazione che segue la religione musulmana. Sempre sul secondo asse si associano le elevate spese in armamenti e in pubblica sicurezza con bassi livelli di diritti politici e civili (delineando regime poco democratici ed inclini ai conflitti).

La mappa dei Paesi, pur evidenziando una variabilità più elevata rispetto alle altre aree tematiche, mette in luce una direzione prevalente che attraversa in diagonale dall'angolo in alto a sinistra all'angolo in basso a destra. La direzione va dai Paesi più aperti ed in cui la tutela dei diritti è più elevata (Norvegia, Danimarca, Svezia, etc.), la presenza delle donne in politica è consistente, ai Paesi in cui il livello dei diritti politici e delle libertà civili è generalmente basso e dove la presenza militare è più consistente. I Paesi a prevalente religione islamica si posizionano nella parte inferiore del grafico e lungo il secondo asse fattoriale. Numerosi Paesi africani, poi, vengono posizionati all'opposto nella parte alta del grafico a causa del loro elevato frazionamento etnico e linguistico. I Paesi più instabili ed in guerra si collocano tra del grafico e sui margini più esterni della raffigurazione. Su tale aspetto ritorneremo in maniera più estesa nel paragrafo 4.4

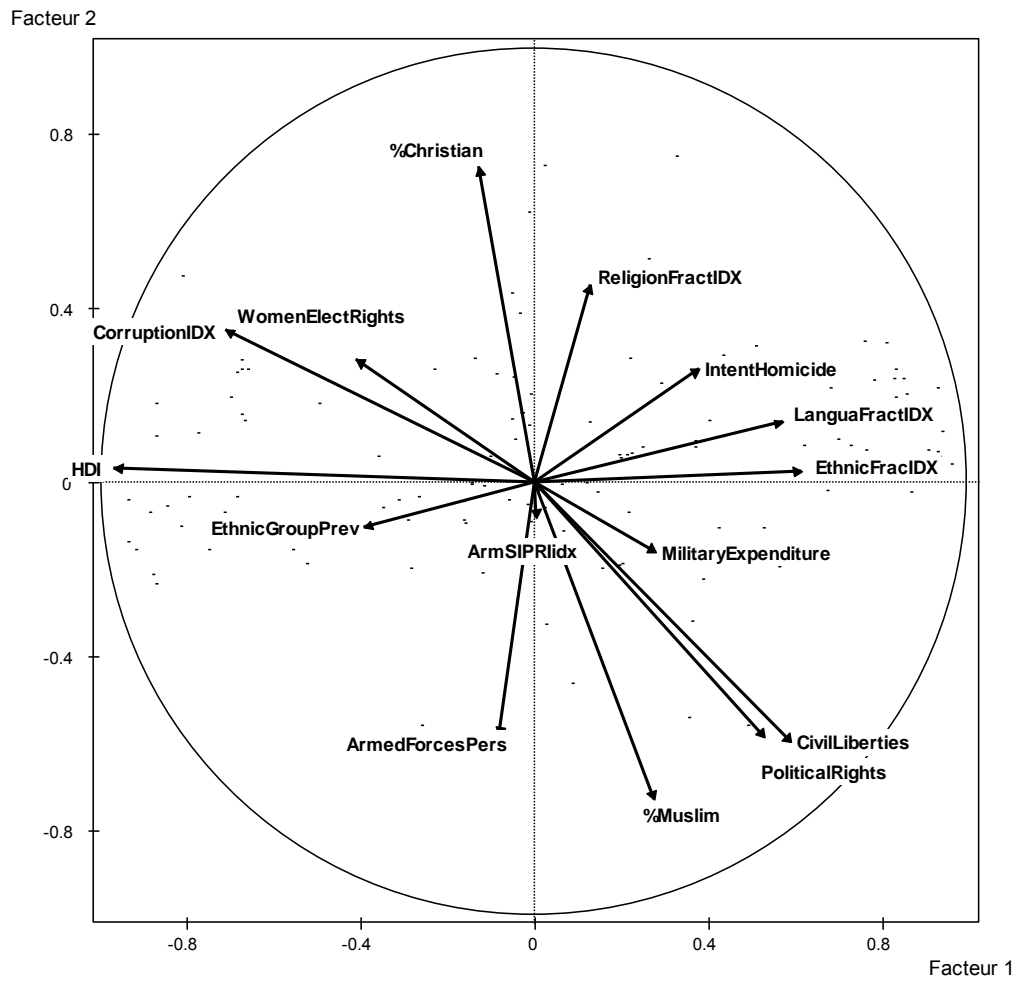


Figura 2.15 Cerchio delle correlazioni "area tematica: Società e Politica"

Utilizzando le prime 5 coordinate e costruendo la distanza euclidea di ogni Paese dalla Norvegia (ovvero costruendo un ordinamento radiale dall'esterno verso l'interno), che appare il Paese più estremo nella direzione dell'assenza di conflitti, è possibile ottenere una sorta di classifica. Ovviamente, l'interesse è per i Paesi che mostrano una maggiore distanza rispetto alla Norvegia. In particolare, si identificano (1=più lontano) i seguenti Paesi.

1. Iraq
2. Afghanistan
3. Lybia
4. Somalia
5. Sudan
6. Djibuti
7. Korea Dem
8. Chad
9. Syria
10. Cote d'Ivoire

Ricordiamo che sebbene questa lista sia prodotta considerando dati del 2010, in realtà essa ha una valenza che va oltre l'istante temporale specifico dal moneto che, come è ben noto, le situazioni di conflitto non sono frutto di eventi istantanei, ma trovano le loro radici in condizioni che si sono determinate nel tempo.

L'analisi globale

A questo punto, le analisi parziali possono essere messe assieme per generare la soluzione globale che consente di evidenziare i legami tra le variabili appartenenti ad aree tematiche differenti (ricordiamo che gli assi fattoriali sono comuni a tutte le soluzioni). Il primo fattore spiega il 31% circa della variabilità complessiva e mostra nella direzione dell'asse verso sinistra una associazione fra il benessere, lo sviluppo umano e elevati livelli di parità di genere nell'istruzione e in politica (vedi Figura 2.17). Sempre dallo stesso lato, ma un po' più in basso, si trovano le variabili che definiscono elevati livelli di scolarizzazione e istruzione, elevata speranza di vita alla nascita e migliori livelli igienico-sanitari. Opposto a ciò si trovano gli elevati tassi di mortalità, l'alta fertilità, i bassi livelli di scolarità, gli elevati livelli di malnutrizione, povertà e disuguaglianza sociale. Sul secondo fattore (che spiega un ulteriore 7,7% della variabilità totale) si oppongono, in alto, la prevalenza di cristiani nella popolazione associata alla partecipazione femminile al mercato del lavoro ed alla presenza di un significativo frazionamento religioso (quest'ultimo sembra essere un elemento positivo per lo sviluppo). Al contrario, il frazionamento etnico e linguistico si associa alla povertà, agli investimenti in sicurezza e armamenti. Questi ultimi risultano ulteriormente legati alla prevalenza della religione musulmana e a bassi livelli di diritti civili e politici goduti dalla popolazione.

Il terzo asse, infine, contribuisce per il 5,29% alla spiegazione complessiva ed è caratterizzato (nella direzione positiva) dall'elevato livello di reddito a cui non si aggiungono alcuna delle altre variabili che indicano benessere o sviluppo umano. In tal senso, si tratta di un fattore che discrimina esclusivamente la ricchezza economica (Figura 2.18).

Nella mappa riportata in Figura 2.19, ogni Paese ha coordinate pari alla media ponderata delle coordinate delle analisi parziali. I punti che rappresentano i Paesi si dispongono come in una nuvola a forma a ferro di cavallo. Tale forma è tipica delle situazioni in cui è presente una prima e forte dimensione latente, a cui però si associa una seconda dimensione, non altrettanto forte, ma significativa. In alto a destra troviamo i Paesi occidentali con elevato benessere, regimi democratici e ampie libertà politiche e civili. In alto a sinistra troviamo i Paesi molto poveri e con elevata mortalità ma che possiedono un qualche grado di democrazia e mostrano una partecipazione femminile alla vita della società. Più a destra e un po' più in basso, troviamo i Paesi molto frazionati dal punto di vista etnico e linguistico e con bassi livelli di scolarità.

Nella zona centrale ci sono i Paesi in via di sviluppo o caratterizzati da economie emergenti, in cui però dal punto di vista delle condizioni sanitarie, del sistema educativo e politico vi sono molte criticità. In basso troviamo in prevalenza i Paesi caratterizzati fortemente dalla religione islamica, molti dei quali in stato di conflitto, e che presentano elevate spese militari.

E' interessante evidenziare come in questa area del grafico si verifichi la vicinanza di Paesi come gli Emirati Arabi, il Bahrein e l'Arabia Saudita, a Paesi quali la Siria e l'Algeria. Tale vicinanza è giustificata dal fatto che la proiezione è ottenuta tenendo conto solo di due dimensioni (la proiezione è infatti nel piano) che sono molto condizionate dalla prevalenza degli islamici nella popolazione (sul primo asse) e i bassi tassi di femminilità (sul secondo asse). Se si prende in considerazione anche il terzo asse (che tiene conto esclusivamente della ricchezza e della produzione) questo gruppo di Paesi si differenzia ulteriormente; Kuwait, Qatar, Oman e Brunei si staccano dai restanti Paesi islamici e si collocano vicino ai Paesi occidentali più ricchi e sviluppati.

Da tali mappe è possibile identificare i gruppi di Paesi via via più fragili partendo dall'angolo in alto a sinistra del grafico e seguendo una serie di curve di livello che si avvicinano progressivamente all'angolo destro. I Paesi più fragili si collocano infatti nella periferia di tale raffigurazione.

Le rappresentazioni descritte possono essere utilizzate congiuntamente con altre informazioni e strumenti di analisi, di tipo qualitativo e quantitativo, per selezionare su scala globale, macroaree o Paesi da monitorare con più attenzione (si vedano l'uso proposto nei paragrafi 3.3 e 4.4).

Nella fase di sperimentazione e attuazione del sistema, l'analisi da noi condotta potrebbe, poi, essere effettuata per più istanti temporali (a distanza di 3 o 5 anni) di modo che siano prodotte delle mappe attraverso le quali identificare i mutamenti nelle posizioni relative dei Paesi.

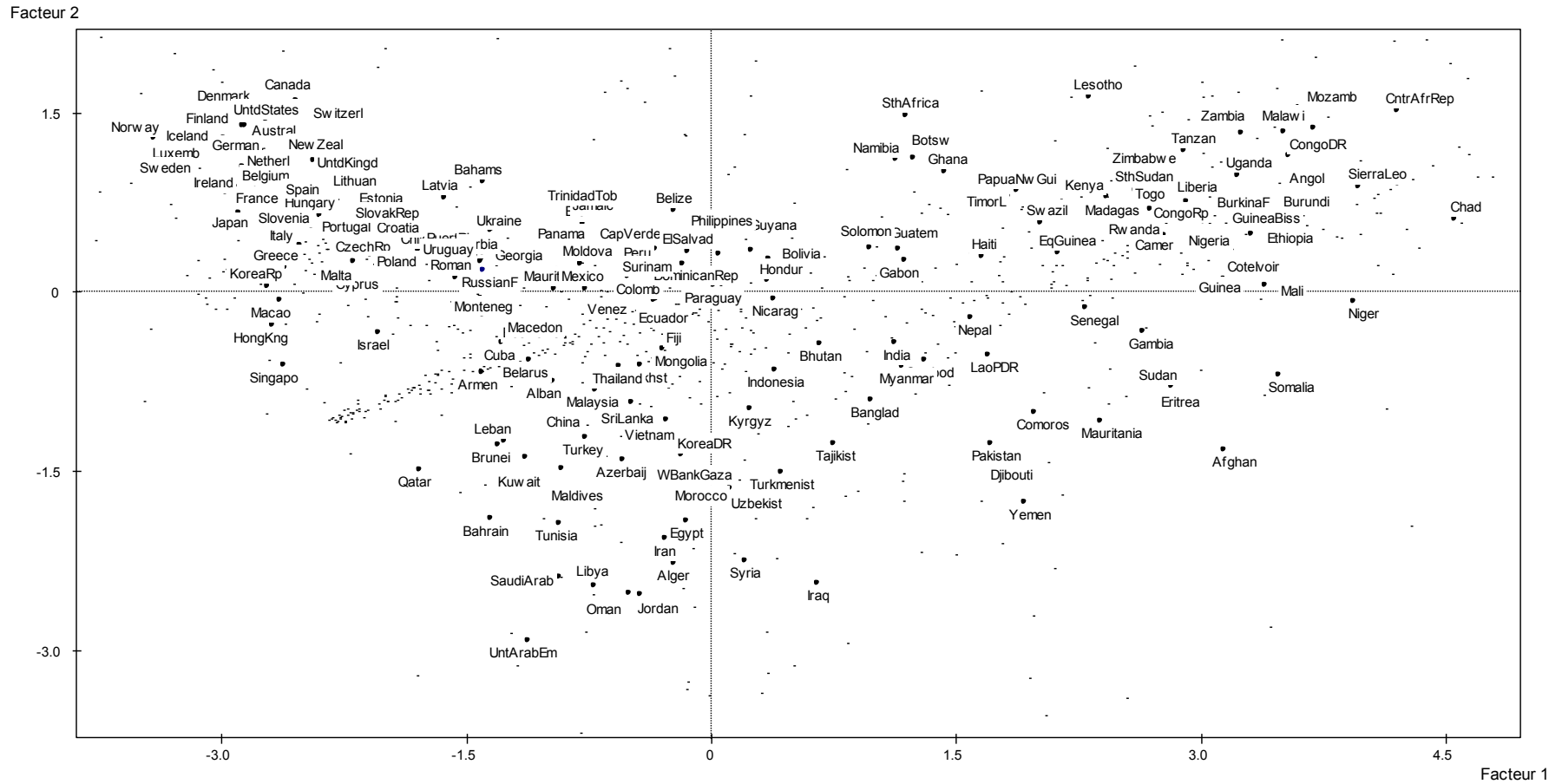


Figura 2.19 Mappa fattoriale globale (fattore 1/ fattore 2)

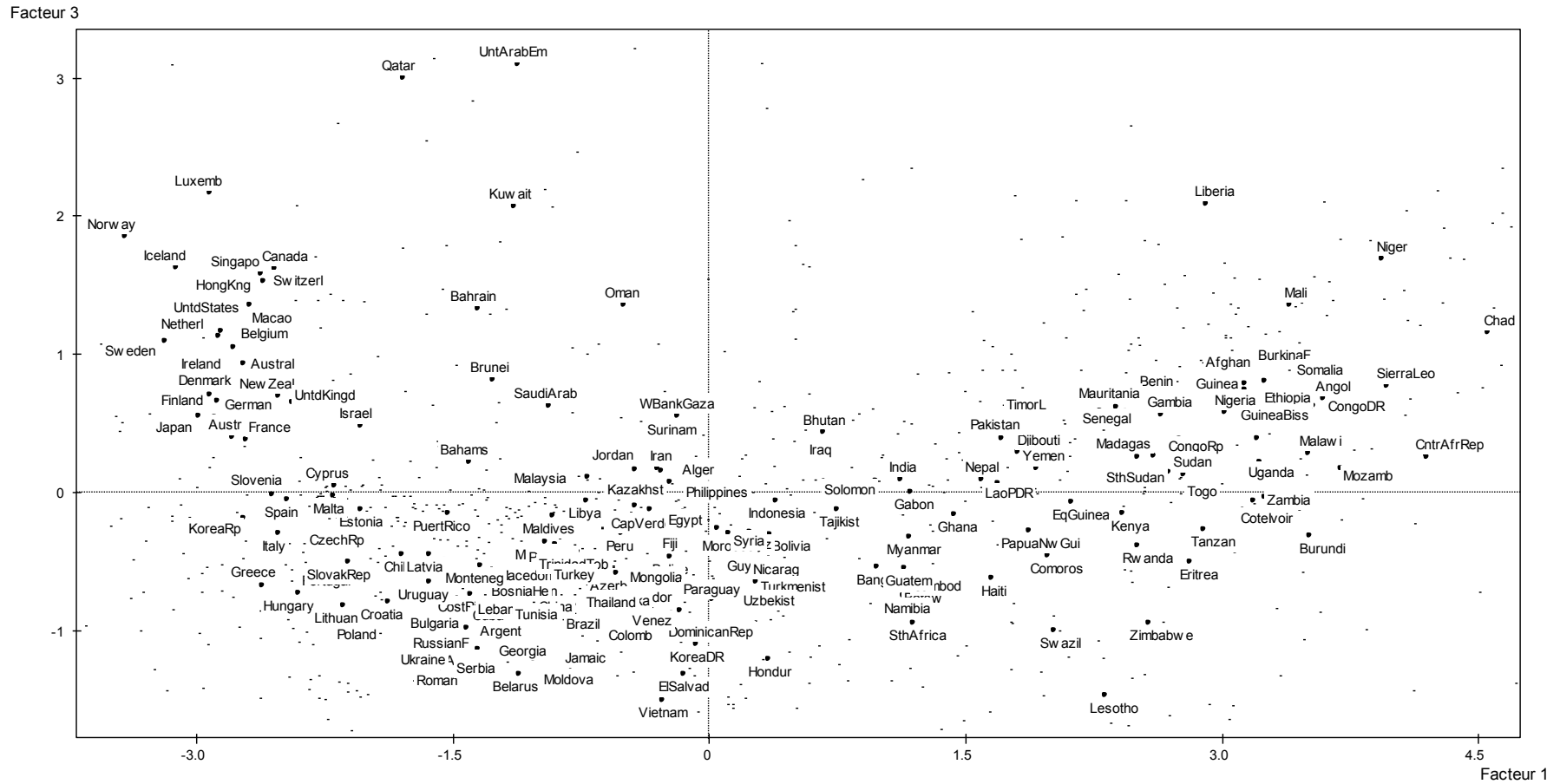


Figura 2.20 Mappa fattoriale globale (fattore 1/ fattore 3)

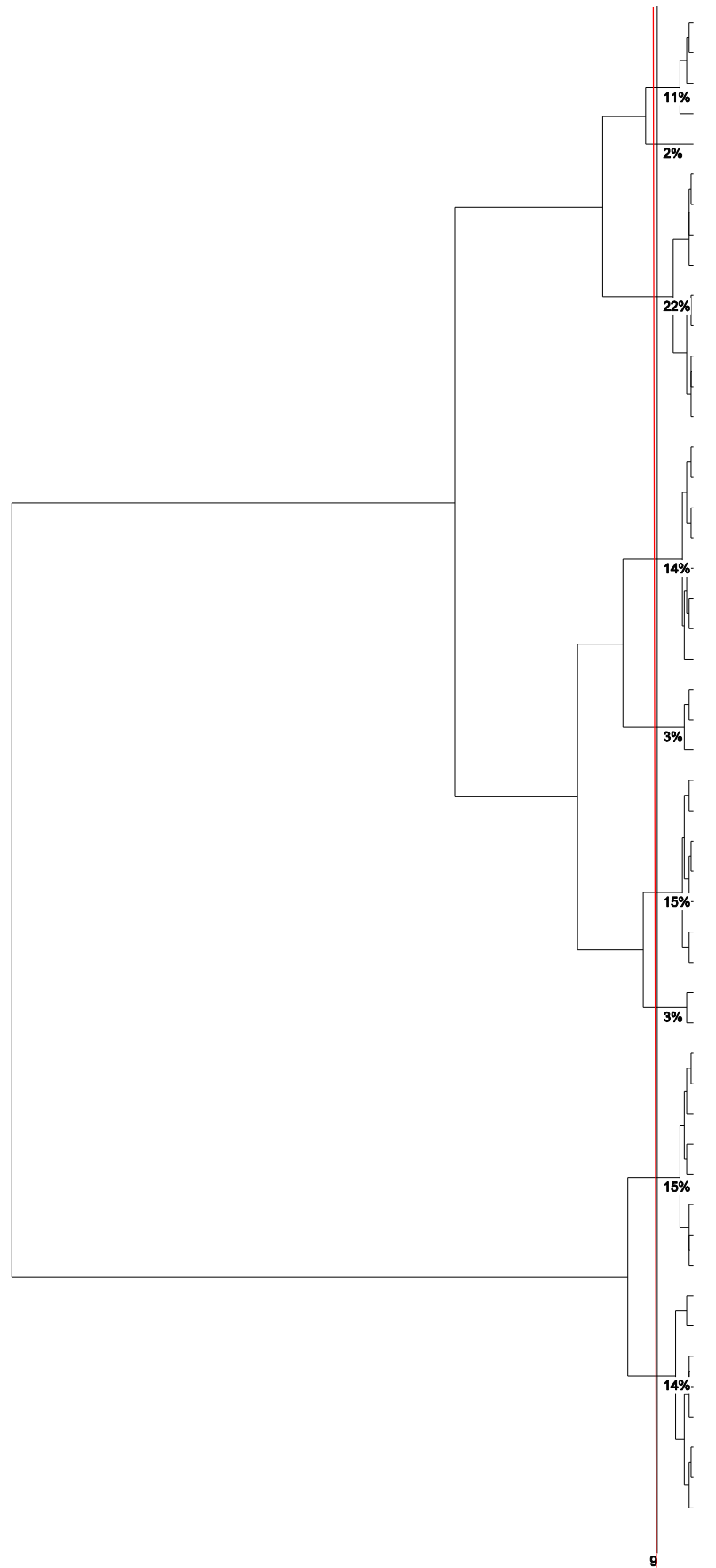


Figura 2.21 Albero di classificazione

Ricostruendo la storia recente si potrebbero ad esempio evidenziare Paesi che si spostano da aree di maggiore stabilità in aree della mappa caratterizzate da maggiori livelli di fragilità. Tale analisi potrebbe anche essere utile in un'ottica predittiva.

Per rendere ancor più evidenti questi gruppi è stata eseguita una cluster analysis gerarchica con il metodo di Ward sui primi 10 fattori dell'AFM (si veda Figura 2.21 in cui è riportato l'albero prodotto dalla classificazione per fornire una prima idea sull'ampiezza dei gruppi). A causa dell'elevato numero di Paesi non è possibile evidenziare le singole denominazioni direttamente nel grafico; esse sono riportate nell'Appendice A.4 da cui si evince la successione con cui i Paesi si aggregano nei gruppi.

Dall'analisi del grafico è tenuto conto di come la variabilità tra i gruppi (che dovrebbe essere grande) e quella all'interno nei gruppi (che dovrebbe essere piccola) varia a seconda del numero di gruppi selezionato, si è optato per una classificazione con 9 clusters. Si noti che l'analisi proposta è basata su una molteplicità di informazioni e, pertanto, essa produce necessariamente una semplificazione di una realtà molto complessa che, come già sottolineato, non tiene conto della prossimità geografica. Inoltre, da un esame approfondito dell'albero di classificazione si nota la presenza all'interno dei gruppi di sotto gruppi che potrebbero essere ulteriormente analizzati, così come si nota la presenza di alcuni Paesi che si trovano al confine tra una classe ed un'altra. Nonostante tali limitazioni, a nostro avviso, la cluster analysis fornisce risultati molto interessanti ed utili per analisi politiche successive. In uno sviluppo successivo, si potrebbero utilizzare metodi di cluster per dati geo-referenziati o vincolati alla presenza di aree di influenza e di network politici.

Di seguito, forniamo la lista dei Paesi appartenenti ai vari gruppi con una descrizione associata che può riassumere in breve

L'Africa sub-sahariana

Angola, Benin, Burkina Faso, Burundi, Cameroon, Central African Republic, Chad, Congo DR, Congo Rep, Cote d' Ivoire, Ethiopia, Guinea, Guinea Bissau, Kenya, Liberia, Madagascar, Malawi, Mali, Mozambico, Niger, Nigeria, Sierra Leone, South Sudan, Tanzania, Togo, Uganda, Zambia

Basso sviluppo umano

Afghanistan, Bangladesh, Bhutan, Cambodia, Comoros, Djibouti, Equatorial Guinea, Eritrea, Gambia, Haiti, India, LaoPDR, Mauritania, Myanmar, Nepal, Pakistan, PapuaNwGuinea, Rwanda, Senegal, Solomon, Somalia, Sudan, Timor Leste, Yemen

Il Sud dell'Africa (HIV)

Botswana, Lesotho, Namibia, South Africa, Swaziland, Zimbabwe

Le economie sudamericane e asiatiche in progresso

Belize, Bolivia, Brazil, Capo Verde, China, Colombia, Dominican Rep, Ecuador, El Salvador, Fiji, Gabon, Ghana, Guatemala, Guyana, Honduras,

Indonesia, Jamaica, Kazakhstan, Mexico, Mongolia, Nicaragua, Panama, Paraguay, Peru, Philippines, Suriname, TrinidadTobago, Venezuela.

I petrolieri

Bahrain, Kuwait, Oman, Qatar, United Arabian Emirates

Le libertà fragili

Albania, Algeria, Azerbaijan, Brunei, Egypt, Iran, Iraq, Jordan, KoreaDR, Kyrgyzan, Lebanon, Libya, Malaysia, Maldives, Morocco, Saudi Arabia, SriLanka, Syria, Tajikistan, Thailand, Tunisia, Turkey, Turkmenistan, Uzbekistan, Vietnam, West Bank Gaza

Le economie tradizionali in crisi e le economie in ripresa

Argentina, Armenia, Bahamas, Belarus, Bosnia Herzegovina, Bulgaria Chile, Costa Rica, Croatia, Cuba, Cyprus, CzechRep, Estonia, Georgia, Greece, Hungary, Israel, Italy, KoreaRep, Latvia, Lithuania, Macedonia, Malta, Mauritius, Moldova, Montenegro, Poland, Portugal, Puerto Rico, Romania, Russian Fed, Serbia, SlovakRep, Slovenia, Spain, Ukraine, Uruguay

La Cina non cinese

HongKong, Macao, Singapore

L'occidente più avanzato

Australia, Austria, Belgium, Canada, Denmark, Finland, France, German, Iceland, Ireland, Japan, Luxemburg, Netherland, New Zealand, Norway, Sweden, Switzerland, United Kingdom, United States

2.6 Osservazioni finali

Le tecniche illustrate forniscono una prima visualizzazione del posizionamento dei Paesi rispetto alle variabili che ne determinano la fragilità.

Il sistema informativo dovrebbe essere implementato da un punto di vista informatico di modo che sia possibile:

- a) selezionare sottoinsiemi di variabili per riprodurre o variare le analisi qui svolte;
- b) disporre di un'interfaccia grafica che includa strumenti di grafica dinamica, consenta la produzione dei grafici legate alle analisi qui svolte e la sovrapposizione ad essi di informazioni provenienti da variabili supplementari (ad esempio: graduatorie stabilite dall'utente, classificazioni tradizionali geografiche o economiche);
- c) evidenziare i dati originari relativi al singolo Paese.

Capitolo 3. I disastri naturali

3.1 Introduzione

Negli anni recenti, i disastri naturali hanno costituito oggetto di grande attenzione per l'apparente crescita nelle perdite in termini di vite umane, nell'ammontare dei danni economici e nell'ampiezza dei territori coinvolti. Al riguardo, vari progetti di considerevole dimensione sono stati promossi per pervenire alla costruzione di indicatori. In special modo, ricordiamo il progetto che ha originato il Disaster Risk Index (UNDP, 2004; Peduzzi et al. 2005, 2009, 2012; Peduzzi, 2012), il progetto "Hotspots" promosso dalla Columbia University in collaborazione con la World Bank (Dilley et al., 2005) e lo studio di un sistema di indicatori regionali per le Americhe sviluppato dall'Università di Colombia nell'ambito delle attività della Inter-American Development Bank. (Cardona, 2006).

Si tratta di ampi studi che partono da ipotesi iniziali diverse e che implementano definizioni e metodologie alternative (Birkmann, 2007; Mosquera-Machado e Dilley, 2009).

Il Disaster Risk Index (DRI) è stato successivamente alla sua implementazione aggiornato nella definizione ed ha assunto una più corretta denominazione: Mortality Risk Index (MRI). Esso prende il numero dei morti come riferimento nella valutazione delle perdite causate da un disastro naturale. Viene calcolato il valore atteso delle perdite come prodotto della popolazione esposta (ovvero la popolazione di una area per la frequenza con cui un certo tipo di disastro accade) per una misura della vulnerabilità specifica. Quest'ultima è stimata attraverso un modello di regressione che tiene conto di alcune caratteristiche dei Paesi. Dalla valutazione del rischio si passa poi ad una categorizzazione in 10 fasce ove vengono collocati i Paesi. L'analisi in rapporto al calcolo del rischio ricostruisce dati a livello territoriale molto circoscritti in relazione agli eventi calamitosi puntualmente identificati in base alle coordinate geo-spaziali. Il progetto Hotspots costituisce un interessante tentativo diretto alle analisi sub-regionali. In particolare, lo studio si fonda su poche informazioni (la popolazione, il GDP, il territorio, l'estensione della rete stradale) che vengono stimate mediante appositi modelli (spaziali) in relazione a delle griglie territoriali. E' possibile così elaborare degli indicatori di rischio di mortalità e di perdite economiche attese. Il progetto non prende in esame alcuna misura esplicita di vulnerabilità, non essendo disponibili informazioni macro-economiche a tale livello di dettaglio. La nozione è perciò implicita negli indicatori di rischio elaborati.

Il terzo progetto invece è multiscopo, essendo focalizzato sulla costruzione di una molteplicità di indicatori per misurare: il rischio macroeconomico e finanziario (Disaster Deficit Index), il rischio sociale ed ambientale (Local Disaster Index), le condizioni di fragilità locale (Prevalent Vulnerability Index) e le capacità manageriali e di reazione locali (Risk Management Index).

Tenuto conto del livello di dettaglio a cui si fa riferimento, tale tipo di analisi può solo essere condotta a partire dalle banche dati degli uffici statistici nazionali e dai risultati di indagini che raccolgono opinioni di esperti. Per tale motivo, lo studio può solo riferirsi ad aree geografiche limitate; nel caso specifico il progetto regionale prende in esame 12 Paesi del centro e sud America.

Numerosi altri contributi hanno poi discusso l'impatto macro-economico dei disastri naturali sottolineando anche con analisi quantitative le differenze che esistono tra i vari Paesi (si veda van Bergeijk e Lazzaroni, 2013, per una rassegna, Kahan, 2005; Strömberg, 2007):

- I Paesi ricchi e sviluppati sono meglio equipaggiati e sono in grado di resistere meglio grazie a reddito più elevato, maggiore expertise, maggiori investimenti per proteggersi dai cambiamenti climatici, un sistema assicurativo avanzato, maggiori informazioni, governi e classi politiche più impegnate nella prevenzione e management dei disastri;
- I Paesi poveri e poco sviluppati, invece, risentono maggiormente dei disastri naturali dal momento che l'economia si fonda essenzialmente sui settori ambiente-dipendenti (agricoltura, pastorizia, etc..) e sensibili alle avverse condizioni climatiche. Inoltre, si tratta di Paesi che non hanno sistemi di prevenzione e che sperimentano notevoli difficoltà ad adattarsi ai cambiamenti di qualsiasi tipo a causa delle scarse risorse, di governi fragili, dell'elevata corruzione, di scarse capacità tecnologiche. A tutto ciò si aggiunge una generale carenza di informazioni per la prevenzione tra le popolazioni, talora agevolata dall'esteso analfabetismo. La sostanziale fragilità determina come fatale conseguenza il fatto che questi Paesi sperimentano disastri molteplici e ravvicinati nel tempo
- le piccole isole e gli arcipelaghi sono più vulnerabili (Lal et al. 2009, per un caso studio).

L'esperienza dei progetti prima citati e dei contributi scientifici, così come riportata in letteratura, ci ha aiutato a definire i contorni del problema ed ad individuare le metodologie a nostro avviso più adatte.

Nel prosieguo del nostro studio nella valutazione, mediante opportune tecniche statistiche, del grado di vulnerabilità abbiamo preso in esame la "probabilità che una collettività sia esposta ad un'emergenza umanitaria, causata dall'interazione della pericolosità di un luogo con il profilo sociale della comunità" (Cutter, 1993).

Come anticipato in precedenza, abbiamo limitato l'attenzione alle calamità naturali che appaiono più frequenti e più significative rispetto al numero di persone colpite. Le tabelle seguenti riportano il numero medio annuo di disastri naturali (Tabella 3.1) e di vittime (Tabella 3.2) nel periodo 2002-2011 ripartito per continente. E' molto evidente la concentrazione territoriale: l'Asia è il

continente più colpito. Analogamente, i fenomeni idrologici e meteorologici presentano un maggior peso rispetto agli altri fenomeni considerati.

Si osservi che la classificazione delle calamità naturali che viene utilizzata (si veda Appendice A.2) include alcuni fenomeni che in questo rapporto, per motivi diversi, non verranno considerati. Difatti, sono presenti gli effetti dovuti alle attività vulcaniche (nelle attività sismiche), alle frane (nei fenomeni idrologici e geofisici), alla siccità ed agli incendi boschivi (nei fenomeni climatici).

Tabella 3.1 Numero medio di disastri naturali 2002-2011

N. di disastri naturali	Africa	Americhe	Asia	Europa	Oceania	Globale
Climatologici	7	11	8	16	1	43
Climatologici-siccità	7	3	4	1	0	16
Geo-fisici	3	7	22	2	2	36
Idrologici	46	41	82	23	5	197
Meteorologici	9	34	39	14	7	102
Totale	72	95	156	56	16	394

Fonte: Annual Disaster Statistical Review 2012, CRED

Tabella 3.2 Numero medio di vittime (morti e persone colpite) dei disastri naturali 2002-2011

N. di vittime (milioni)	Africa	Americhe	Asia	Europa	Oceania	Globale
Climatologici	10,03	0,67	19,48	0,14	0,00	30,60
Climatologici-siccità	13,83	0,69	57,32	0,13	0,00	71,97
Geo-fisici	0,08	0,83	7,13	0,01	0,07	8,12
Idrologici	2,08	4,26	111,05	0,28	0,06	117,71
Meteorologici	0,37	2,19	37,05	0,11	0,04	39,75
Totale	26,38	8,64	232,03	0,66	0,17	267,88

Fonte: Annual Disaster Statistical Review 2012, CRED

La siccità è certamente un fenomeno di rilievo, soprattutto in Asia ed Africa, ma il suo studio dovrà essere affrontato, al di fuori di questo progetto di fattibilità, e su basi specifiche. Si tratta infatti di un fenomeno completamente diverso da tutti gli altri, in quanto, la siccità colpisce prevalentemente particolari aree del pianeta e si realizza con un processo di accumulo generato da una serie di eventi concomitanti e persistenti (meteorologici e non) che determinano nel tempo il progressivo impoverimento delle risorse idriche locali. A ciò si associano le crescenti difficoltà (o impossibilità) a coltivare la terra, a svolgere le tradizionali attività per la produzione del fabbisogno alimentare locale nonché quelle azioni connesse al mantenimento della qualità del suolo coltivabile. Migrazioni forzate, conflitti locali, etc. costituiscono poi le aggravanti più frequenti che preludono ad eventi complessi quale la fame. Per tale motivo abbiamo ritenuto che occorra effettuare uno studio ad hoc per tale fenomeno in una fase successiva.

Sono state, inoltre, escluse le attività vulcaniche e le frane per il modesto contributo che esplicitano in termini di vittime (con una media annua di circa 0,5 milioni) e per la rarità degli eventi mentre, invece, gli incendi non sono stati considerati, oltre che per la modesta incidenza anche perché questa tipologia di disastro non ci sembra strettamente inerente le problematiche del presente progetto.

3.2 Definizioni e dati

Varie tipologie di 'perdite o danni' possono essere utilizzate per categorizzare l'entità di un disastro naturale: il numero di morti, di persone colpite, il valore dei danni alle strutture o più in generale dei danni economici. Tali valutazioni possono essere effettuate in termini assoluti, ad esempio, prendendo in esame "*il numero di persone colpite*", oppure in termini relativi laddove si consideri "*la percentuale di popolazione colpita nella regione*". Proseguendo in questo esempio, è evidente che le analisi basate su queste due misurazioni producono risultati diversi e quindi orientano in maniera diversa il processo decisionale. E' banale osservare che un disastro che coinvolge 500.000 persone alle isole Fiji colpisce il 58% della popolazione totale, d'altra parte un disastro della medesima entità in Cina coinvolge appena lo 0,037 % della popolazione totale, ma ciò ci fa comprendere che la scelta di una delle scale di misurazione non potrà che generare risultati contrapposti, L'ordinamento dei Paesi rispetto all'esposizione alle calamità naturali vedranno nei primi posti i Paesi più popolati, per quanto riguarda le analisi effettuate utilizzando i valori assoluti, ed invece i Paesi più piccoli e gli arcipelaghi nel caso si utilizzino le valutazioni relative.

Sulla base delle indicazioni fornite da STC questo studio ha preso come riferimento *il numero di bambini colpiti da un disastro*.

Al riguardo è bene chiarire che la definizione di 'bambino' qui adottata è quella relativa alle statistiche internazionali (che considerano la classe 0-14) e che in genere suddividono la popolazione in classi di età quinquennali: 0-4, 5-9, 10-14, 15-19 etc....

Pertanto, in questo studio preliminare ove è sostanzialmente richiesto di esplorare l'esistenza di tecniche e procedure adatte agli obiettivi posti, si è preferito utilizzare come riferimento la popolazione nella fascia di età 0-14, lasciando alla successiva fase di implementazione del progetto il raffinamento di questo indicatore per produrre una stima affidabile della popolazione nella classe di età 0-18, che costituisce il riferimento di STC. Una verifica effettuata sulla categorizzazione dei disastri naturali in fasce di emergenza, di cui si dirà a breve, utilizzando una stima molto rozza della popolazione nella classe di età 0-18 ci ha mostrato che gli effetti della limitazione si esplicita soprattutto per i disastri di entità medio-bassa.

L'attenzione dedicata negli ultimi decennio per la creazione di sistemi di early warning sempre più ampi e sofisticati, ha immediatamente posto in evidenza l'esigenza di archivi affidabili che registrino i disastri e il loro impatto. Le

Nazioni Unite attraverso UNDP (Development Programme) ed il CRED (Centre Centre for Research on the Epidemiology of Disasters) hanno notevolmente contribuito al dibattito ed alla ricerca di soluzioni che potessero essere utili per fornire informazioni affidabili alle istituzioni internazionali ed alle organizzazioni umanitarie. In particolare, nel 2006 in previsione di un primo incontro internazionale sul tema, è stato condotto uno studio per confrontare i principali dataset esistenti sia a livello mondiale (EM-DAT, NatCal, etc..) che a livello nazionale e di macro-regione (Tschoegl et al., 2006)

Il presente studio utilizza il database EM_DAT (Emergency Disasters Data Base), gestito dal CRED all'Università Cattolica di Louvain in Belgio. Questo archivio costituisce un catalogo pubblico che contiene informazioni sui disastri naturali, tecnologici e sui cosiddetti disastri complessi (fame, etc.), a partire dal 1900 ad oggi. In totale, a luglio 2014, presenta 20981 disastri di cui 13142 naturali (questo numero si riferisce ai soli records completi di tutte le informazioni). Il catalogo è aggiornato ogni giorno e rende disponibili dati concernenti l'evento (la tipologia, il luogo, la data) e l'entità del disastro (numero di morti, feriti, numero di persone colpite, l'ammontare stimato dei danni economici). I disastri naturali sono classificati in 5 sotto-categorie che includono 12 tipi di disastri e più di 30 sotto-tipologie. I disastri catalogati devono soddisfare almeno uno dei seguenti requisiti: a) aver prodotto 10 o più morti; b) aver colpito 100 o più persone; c) essere tale da far dichiarare lo stato di emergenza; d) essere tale che le autorità richiedano aiuto a livello internazionale. Le fonti con cui opera il CRED sono le principali agenzie delle Nazioni Unite (UNEP, OCHA, WFP, FAO), le Organizzazioni Non Governative, gli enti di ricerca, le compagnie assicurative internazionali (Lloyds) e le agenzie di stampa.

Il database predisposto dal CRED, per la sua vasta esperienza e ampio riconoscimento della comunità scientifica, è attualmente la fonte più utilizzata per gli studi in tema di rischi naturali. L'archivio è accessibile via web (www.emdat.be) e quasi tutte le informazioni possono essere scaricate direttamente dall'utente. Di recente, però, il CRED ha ristretto l'accesso all'archivio dettagliato e, di conseguenza, il dataset che in questo lavoro viene utilizzato può essere riprodotto solo contattando direttamente l'ente gestore. Al riguardo, si vedano le raccomandazioni finali del presente rapporto.

L'affidabilità del database muta in funzione del periodo temporale e della natura delle variabili considerate. Il numero di disastri, infatti, registrati aumenta nel tempo fino al 1995. Ciò a detta dello stesso CRED è certamente frutto della accresciuta capacità espressa nella raccolta ed elaborazione dei dati ma è anche in parte effetto di una accresciuta vulnerabilità ad eventi climatici estremi che determinano alcuni tipi di disastri (come cicloni e alluvioni). Nel nostro lavoro è stato perciò considerato il periodo 1990-2011. Ovviamente, tale finestra temporale è troppo piccola per validare una qualsiasi ipotesi di mutamento climatico. Inoltre, ad essa sfuggono inevitabilmente gli eventi naturali con periodo di ritorno molto elevati.

In relazione alla qualità dell'archivio considerato, è opportuno notare che i dati non provengono da rilevazioni ma piuttosto dall'analisi incrociata di numerose fonti informative quali: i comunicati delle autorità locali, le agenzie delle Nazioni Unite, compagnie assicurative, organizzazioni non governative etc.. Quindi, si tratta di vere e proprie stime elaborate dal CRED a cui si può attribuire un diverso grado di affidabilità: più elevato nel caso delle statistiche concernenti i morti, meno elevato nel caso del numero di persone colpite (che resta comunque una stima più grossolana e poco verificabile). A conforto di ciò, bisogna anche dire che questo tipo di informazione approssimata ha certamente una qualità analoga o superiore a quella delle informazioni sulla base delle quali usualmente vengono prese decisioni in presenza di una calamità. Da un punto di vista statistico, il consistente errore di misurazione che influenza le osservazioni fa sì che le relazioni con i fattori esplicativi che faranno parte del modello si rivelino meno forti.

Si osservi, infine, che in presenza di eventi che investono più Paesi, il database EM-DAT ripete la registrazione per ciascun Paese coinvolto con la medesima data e codice di evento. Per tale motivo, il numero di records è più ampio del numero di eventi distinti registrati.

Partendo dall'archivio EM-DAT si è proceduto alla costruzione del database per il presente studio di fattibilità. In particolare, come già anticipato l'attenzione è stata focalizzata sui seguenti disastri: terremoti, alluvioni, tempeste di varia natura, temperature estreme, che costituiscono nell'insieme complessivo gli eventi più frequenti.

Per quanto concerne gli Tsunami, nel periodo 1990-2011, nell'archivio EM-DAT sono registrati 6 eventi che hanno prodotto circa 250.000 morti e 2.900.000 vittime. Lo tsunami del 2004 è certamente l'evento più intenso di tale categoria con il 90,4% dei morti e l'82,8% dei colpiti. Sebbene, il fenomeno costituisca un evento troppo raro per essere considerato in maniera specifica in qualsiasi tipo di modello, tenuto conto delle esigenze esposte da STC, si è, analogamente a quanto fatto dal CRED, catalogato lo tsunami all'interno degli eventi sismici, di modo che fosse incluso nelle successive analisi. E' bene però sottolineare che le conseguenze dello tsunami del 2004 furono eclatanti per la contemporaneità dell'evento che coinvolse numerosi Paesi e per le vaste distruzioni alle infrastrutture, ma l'evento ebbe una dimensione assoluta (numero di persone affette) di gran lunga inferiore ad altri singoli eventi presenti nel database.

In generale, l'archivio di EM-DAT relativo ai disastri da noi selezionati e per l'arco temporale considerato contiene un numero consistente di records incompleti (in una delle variabili che qualificano l'entità del disastro) come descritto nella tabella seguente.

Tabella 3.3. Archivio EM-DAT 1990-2011

Numero records	Numero morti %records mancanti	Popolazione colpita %records mancanti
6314	27,6	22,7

Sono stati quindi considerati tutti i records ove il numero di persone colpite era presente e si è proceduto attraverso i dati sulla popolazione per fasce di età rese disponibili dalle Nazioni Unite (Population Division) alla stima per ciascun evento del numero di bambini coinvolti. Sulla base di tale stima si è poi proceduto alla categorizzazione degli eventi in fasce di emergenza secondo la classificazione di STC (Save the Children Intl, 2012).

Tabella 3.4 Classificazione STC delle categorie di emergenza

Extraordinary (Cat 1)	Large (Cat 2)	Medium (Cat 3)	Small (Cat 4)
<i>Numbers affected:</i> Significant population size affected and/or >1 million children affected	Significant population size affected and/or >100,000 children affected	>10,000 children affected	>1,000 children affected

Si osservi che per motivi operativi, per avere un ordine crescente coerente tra dizione e valore numerico, le categorie di emergenza sono state ricodificate come segue: None =0 (un disastro che colpisce meno di 1000 bambini), Low =1; Medium =2; Large=3, Extremely Large=4.

La classificazione dei disastri in categorie ha prodotto per il periodo di riferimento e per i soli records validi la seguente distribuzione.

Tabella 3.5 Classificazione dei records in categorie di emergenza per il periodo 1990-2011

valori	Categorie di emergenza	Numero di records	%
0	<=1000 children	1948	41,26
1	1000<children<=10.000	1260	26,69
2	10.000children=<100.000	923	19,55
3	100.000<children=< 1 million	455	9,64
4	>1 million children	135	2,86

Si evince con chiarezza che circa il 41% dei records si riferisce ad emergenze di livello estremamente ridotto (meno di 1000 bambini coinvolti) e che non rientrano nelle fasce di emergenza per le quali si suppone un intervento. D'altra parte, solo il 13% dei records è costituito da emergenze nazionali di grande impatto.

Un ulteriore problema, che è strettamente legato alla fase di indagine sulle relazioni tra disastri ed indicatori di vulnerabilità (macro-economici, socio-politici e ambientali), concerne la ricostruzione di serie storiche complete ed affidabili, per almeno due decenni, per gli indicatori da utilizzare nella successiva fase. Il problema, già evidenziato in precedenza (vedi paragrafo 2.3) circa i dati mancanti e la qualità delle statistiche internazionali, si riflette in maniera significativa nella stessa selezione delle variabili per la modellistica.

In conclusione, per il nostro studio è stato costruito un dataset che contiene: il Paese, l'anno (t) in cui ha colpito la calamità, il livello di emergenza, le variabili esplicative al tempo ($t-3$). Per poter consentire, infatti, l'applicazione del modello a fini di proiezione, si è considerato un ritardo temporale di 3 anni, di

modo che con i dati disponibili (fino al 2011) fosse possibile estrapolare almeno i risultati all'anno 2014. Questo aspetto costituisce un elemento che andrà approfondito nella seconda fase di sperimentazione più ampia sia rispetto alla possibilità di migliorare l'archivio di partenza che rispetto al affinamento della modellistica applicata.

3.2 Il modello statistico

Il confronto tra i Paesi è effettuato in base alla probabilità che un Paese sia colpito da una calamità naturale che generi un'emergenza di livello prefissato (extremely large, large, medium and small). Indicheremo tale probabilità con:

$$Pr(\text{calamità naturale} \cap \text{emergenza di livello } j), \quad j = 0, 1, \dots, 4$$

In particolare, essa verrà calcolata come prodotto:

$$Pr(\text{calamità naturale} \cap \text{emergenza di livello } j) = Pr(\text{calamità naturale}) Pr(\text{emergenza di livello } j | \text{calamità naturale}).$$

Il primo termine, $Pr(\text{emergenza di livello } j | \text{calamità naturale})$, indica la probabilità che si abbia un'emergenza di livello j essendo stata osservata una certa calamità. Si osservi che tale termine è strettamente legato ai fattori che determinano la vulnerabilità di un Paese. Pertanto, essa verrà stimata mediante un modello statistico che pone in relazione una opportuna trasformazione di tale quantità con un insieme di variabili che ne spieghino la variazione nei diversi Paesi

Il secondo termine: $Pr(\text{calamità naturale})$, indica la probabilità con cui una calamità naturale colpisce un certo Paese. In esso si estrinsecano gli effetti delle condizioni fisio-grafiche, delle aree climatiche, delle caratteristiche geologiche e idrologiche del territorio.

3.2.1 La probabilità di emergenza allorché è osservato un disastro

Il modello utilizzato appartiene alla classe dei *proportional odds models* che consente di modellare simultaneamente la probabilità

Indicando con Y il livello di emergenza e tralasciando nella notazione il condizionamento alla realizzazione di una calamità naturale, allora: $Pr(Y \leq j)$ indica la probabilità che il livello di emergenza sia pari o inferiore a j , dove il livello j indica uno dei livelli prima codificati (da 0 a 4 a seconda della gravità del disastro).

Nella formulazione più elementare, il modello lega $\text{logit}(Pr(Y \leq j))$, $j=0,1,\dots,4$ alle variabili esplicative (X_1, X_2, \dots, X_k) mediante una relazione lineare, ovvero:

$$\text{logit}(Pr(Y \leq j)) = a_j + b_1 X_1 + \dots + b_k X_k$$

dove:

$$\text{logit}(Pr(Y \leq j)) = \log \frac{Pr(Y \leq j)}{1 - Pr(Y \leq j)}$$

indica una particolare trasformazione della probabilità che il livello di emergenza sia pari o inferiore a j , mentre (a_0, a_1, \dots, a_4) e (b_1, b_2, \dots, b_k) sono i coefficienti del modello. La trasformazione è reversibile, nel senso che allorché il modello è stato stimato, e quindi i coefficienti del modello sono stati calcolati, è possibile stimare la $\Pr(Y \leq j)$, per i vari valori di j , da cui poi ricavare la probabilità che un'emergenza sia di un livello j fissato.

Senza entrare nei dettagli metodologici per i quali si rinvia ad uno dei testi fondamentali quali Agresti (2002), di seguito illustreremo le variabili selezionate per la costruzione del modello ed i risultati ottenuti.

L'analisi statistica sulla vulnerabilità dei Paesi ha evidenziato alcune dimensioni fondamentali di cui si è tentato di tener conto nella scelta delle variabili.

In particolare, sono entrate a far parte del modello finale le seguenti variabili:

- il GDP pro-capite; la percentuale di territorio destinata alla agricoltura;
- la densità della popolazione; il tasso di crescita della popolazione urbana;
- l'altitudine media del Paese e la percentuale di territorio che ricade nella fascia tropicale
- il tasso di dipendenza per la popolazione 0-14, l'indice di parità di genere nel lavoro.

Tabella 3.6 Modello stimato

Variabile	Coefficiente	Std.Error	t-value
GDP (log)	-0.480926	0.033419	-14.391
Population density (log)	0.056389	0.026877	2.098
Elevation (sqrt)	0.022031	0.003641	6.051
Agricultural land (log)	0.190119	0.057123	3.328
urban growth (log)	0.810107	0.307272	2.636
tropical	0.979164	0.091195	10.737
labour parity index	0.004494	0.001486	-3.025
Child dependency	-2.425796	0.252032	-9.625

NOTA: Alcune variabili sono state preliminarmente trasformate. In particolare, alle variabili (opportunitamente traslate) è stata applicata la trasformazione logaritmo (GDP, densità della popolazione, il tasso di crescita della popolazione urbana, l'indice di parità di genere nel lavoro, la percentuale di territorio destinata alla agricoltura) e la radice quadrata (altitudine media).

Intercetta	Coefficiente	Std.Error	t-value
0 1	-1.6207	0.8382	-1.9336
1 2	-0.3621	0.8380	-0.4320
2 3	0.9160	0.8387	1.0922
3 4	2.5474	0.8427	3.0230

Le variabili hanno un ruolo diverso a seconda del peso e del contributo che esprimono nello spiegare la dinamica del fenomeno di interesse. In particolare, le variabili sono entrate a far parte del modello con un processo step-wise (si parte dal modello più semplice e si complica la struttura aggiungendo ulteriori

variabili esplicative fino a quando un indicatore di bontà del modello non dimostri che gli ulteriori guadagni in termini di spiegazione del fenomeno siano trascurabili). Nella tabella seguente riportiamo la strategia seguita nella costruzione del modello e l'indicatore AIC che fornisce una misura per comparare i modelli (il modello migliora quando AIC diminuisce).

Modello	GDP (log)	Populati on density (log)	Elevation (sqrt)	Agricult ural land (log) ^r	Urban growth (log)	tropic al	labour parity index	Child depen dency	AIC
A	•								11284,11
B	•	•							11249,83
C	•	•	•						11245,44
D	•	•	•	•					11181,17
E	•	•	•	•	•				11111,0
F	•	•	•	•	•	•			11106,60
G	•	•	•	•	•	•	•		11100,29
H	•	•	•	•	•	•	•	•	11067,03

Il modello A costituisce il modello più elementare ove viene utilizzato solamente il GDP (espresso in logaritmo) come variabile esplicativa. Il modello man mano viene complicato aggiungendo ulteriori variabili, fino a giungere al modello finale indicato con H, dove l'indice raggiunge il valore più basso.

E' bene osservare che le variabili che intervengono nella spiegazione del fenomeno di interesse sono fattori che caratterizzano la struttura del Paese e, pertanto, sono variabili che mutano, in genere lentamente. Nonostante, sia possibile identificare delle tendenze di crescita/decrecita, tali dinamiche si esplicitano con gradualità laddove il Paese non subisca un tracollo politico ed economico catastrofico. Quindi è lecito attendersi che il modello produca delle stime che per molti Paesi non varieranno grandemente da anno ad anno. In seguito, andrà sperimentato un periodo sufficientemente ampio per procedere all'aggiornamento del modello.

Inoltre, occorre sottolineare che nel corso dello studio sono state, in aggiunta a quelle elencate, prese in esame altre variabili che dalle analisi statiche apparivano essere delle determinanti importanti nella specificazione delle dimensioni fondamentali per la rappresentazione dei Paesi rispetto alla vulnerabilità: l'aspettativa di vita, l'indice di parità di genere rispetto all'aspettativa di vita, la mortalità infantile, la percentuale di popolazione urbana, la percentuale di territorio coperto da foreste, le emissioni di CO₂, l'indicatore di efficienza dei governi sviluppato dalla World Bank. Tali variabili però non sono risultate significative nella descrizione del modello.

Il modello individuato è parsimonioso nel senso che poche variabili contribuiscono alla spiegazione della probabilità di interesse. In realtà, come già abbiamo discusso in precedenza, la dimensione fondamentale è quella economica, per questa ragione, essendo molte variabili connesse al reddito nazionale di fatto non aggiungono alcun contributo nella spiegazione

complessiva. La parsimonia, peraltro, è una caratteristica che assicura la reale applicabilità dei modelli: modelli semplici sono più agevolmente controllabili rispetto alla coerenza e consistenza, sono più facilmente riproducibili nel tempo, esplicitano generalmente relazioni forti.

Non sono state prese in considerazione specifiche variabili concernenti i tassi di alfabetizzazione e l'educazione a causa della scarsa qualità dei dati e della mancanza di serie storiche complete per gli ultimi due decenni.

Inoltre, inizialmente sono stati esplorati altre formulazioni del modello per tener conto dell'effetto specifico dei diversi tipi di disastri considerati (terremoti, inondazioni, temperature estreme e tempeste). L'ipotesi di lavoro non è stata però a questo stadio di analisi confermata; il che ci porta a credere che, essendo la probabilità che stiamo valutando connessa al concetto di vulnerabilità (e quindi ai fattori strutturali del Paese), l'intensità dell'evento piuttosto che la tipologia della calamità sia rilevante nella sua determinazione. Al riguardo, sebbene EM-DAT fornisca l'indicazione della localizzazione degli eventi non fornisce una misura di intensità e la ricostruzione di tale dato, pur tentato da altri progetti (ad esempio, il progetto promosso dal Development programme delle Nazioni Unite per la costruzione del Disaster Risk Index), non era un obiettivo perseguibile in questa fase.

3.2.2 La probabilità che si verifichi una calamità

La definizione fornita dal UNDR0 (1979) indica come "hazard occurrence probability", la probabilità che accada una particolare calamità naturale con uno specificato livello di intensità in un particolare periodo di tempo". Come già abbiamo discusso in questo studio non sono presenti informazioni sull'intensità degli eventi naturali. L'unica stima possibile, su cui basare questo studio preliminare, è la frequenza relativa con cui in un Paese è stata osservata un certo tipo di calamità naturale.

In particolare, poiché nel nostro modello, come già evidenziato non vi è uno specifico riferimento alla tipologia delle catastrofi, considereremo semplicemente la frequenza relativa con cui nell'ultimo ventennio si sono verificate calamità appartenenti alle quattro tipologie esaminate.

3.3 La classifica ed una visualizzazione dei Paesi a rischio

Sfruttando il modello descritto nel paragrafo precedente sono state stimate le probabilità che si verifichi un'emergenza di un certo livello che sono divenute la base per procedere ad un ordinamento dei Paesi rispetto alle 4 fasce di emergenza.

Small	1000<children<=10.000
Medium	10.000children=<100.000
Large	100.000<children=< 1 million
Extremely Large	>1 million children

Di seguito riportiamo le classifiche di interesse con particolare riferimento ai Paesi che ne costituiscono il primo 15% per l'anno 2014. Come già evidenziato in apertura di questo rapporto, la scelta effettuata da STC di operare su una scala di valori assoluti (il numero di bambini colpiti) fa sì che i Paesi più popolosi (la Cina in particolare) risultino ai primi posti.

Nella predisposizione delle classifiche si è ovviamente tenuto conto del fatto che in alcuni Paesi le fasce di emergenza di livello più elevato non possono verificarsi, essendo la numerosità della popolazione presente in quei territori non compatibile con le numerosità indicate nella definizione di tali fasce (ad esempio, in Islanda non possono verificarsi emergenze "Large" o "Extremely Large" dal momento che la popolazione 0-14, al 2010, è di circa 66mila bambini). Le classifiche contengono, perciò, un numero diverso di Paesi.

Tabella 3.7 Classifica (prima fascia: top 15%)

Small	Medium	Large	Extremely Large
CHN	CHN	CHN	CHN
PHL	IND	IND	IND
USA	PHL	PHL	PHL
IND	IDN	VNM	VNM
IDN	BGD	IDN	IDN
BGD	VNM	BGD	BGD
MEX	USA	HTI	HTI
IRN	IRN	THA	THA
VNM	THA	PAK	BDI
PAK	MEX	IRN	PAK
AFG	PAK	USA	IRN
BRA	AFG	MEX	ETH
THA	HTI	AFG	USA
RUS	BRA	ETH	LKA
TUR	COL	LKA	MEX
ROM	LKA	BDI	AFG
COL	PER	BRA	COL
AUS	ETH	COL	MDG
JPN	MDG	MDG	BRA
PER	KEN	PER	KEN
HTI	RUS	KEN	TJK
FRA	TUR	TJK	PER
DZA	ROM	CRI	CRI
NGA	TJK	NPL	NPL
LKA	CRI	TZA	TZA
ARG	NGA	HND	YEM
KEN	TZA	YEM	HND

Sulla base della precedente classifica per categoria di emergenza è necessario stabilire un'unica lista. A tal fine è stata utilizzata un metodo di aggregazione robusto che tiene conto del fatto che le liste contengono un numero diverso di Paesi (Kolde et al. 2012). La procedura è robusta nel senso che essa evita che la classifica finale risenta grandemente di situazioni in cui uno stesso Paese occupi posizione estreme nelle graduatorie intermedie (ad esempio è primo in una lista ed ultimo in un'altra).

Si osservi, infine, che, a nostro avviso, la classifica può essere utilizzata per grandi fasce, cioè per fette ampie della distribuzione (ad esempio, il primo 10% dei Paesi) mentre non va commentato specificamente confrontando le posizioni di due singoli Paesi, poiché le differenze in termini di probabilità potrebbero essere non significative.

1	China
2	Philippines
3	India
4	Indonesia
5	Bangladesh
6	Vietnam
7	Iran
8	Pakistan
9	USA
10	Thailandia
11	Mexico
12	Afghanistan
13	Haiti
14	Colombia
15	Brasil
16	SriLanka
17	Peru
18	Kenia
19	Madagascar
20	Ethiopia
21	CostaRica
22	Cuba
23	Nicaragua
24	Tanzania
25	Honduras
26	Turkey
27	Tajikistan

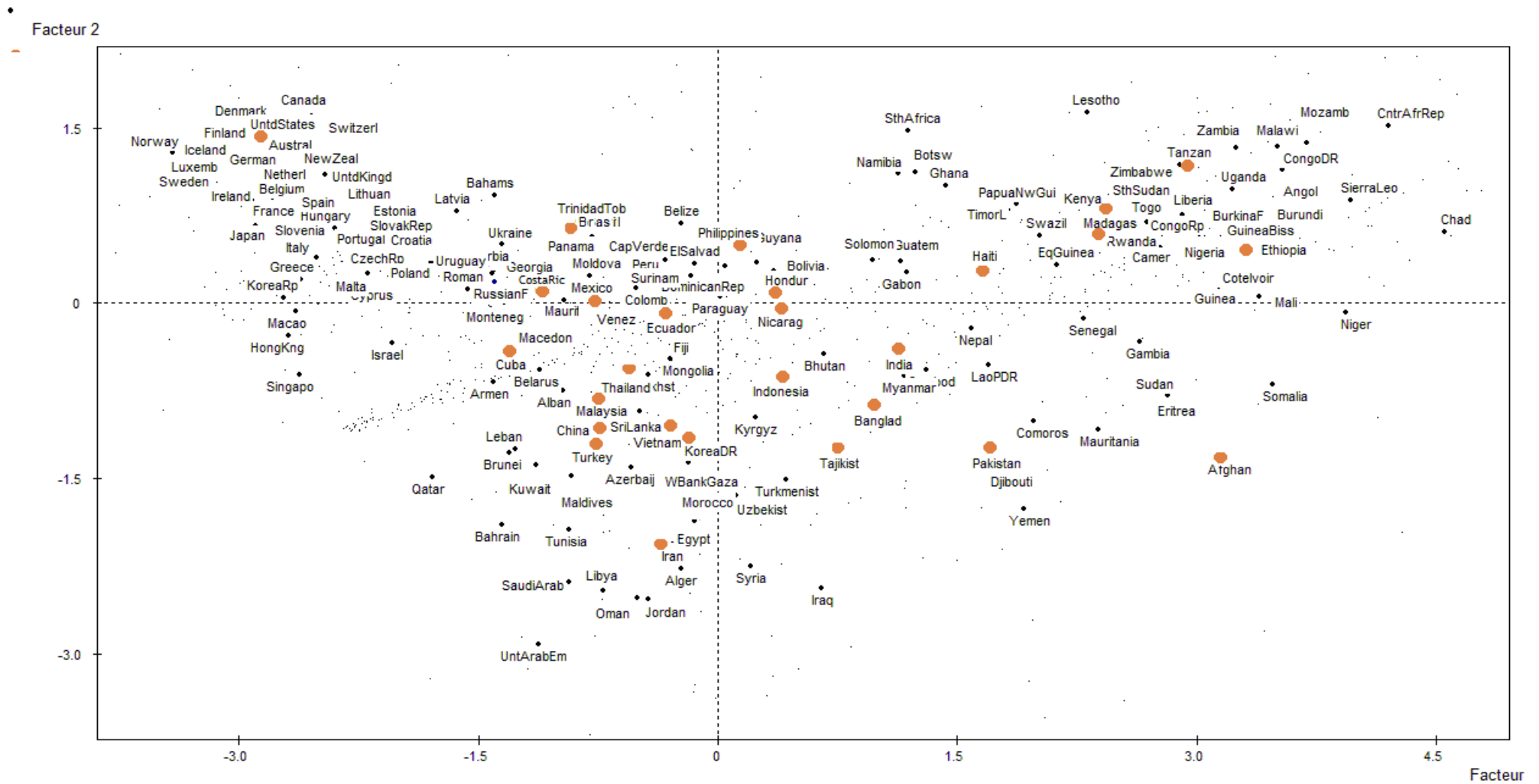
E' opportuno sottolineare che la classifica ottenuta si basa su due assunti che possono essere oggetto di consistenti miglioramenti nella successiva fase di sperimentazione e attuazione. In primo luogo, nella categorizzazione delle emergenze non si è tenuto conto del riferimento geografico specifico dell'evento, che sembra essere disponibile nell'archivio CRED ma a cui non è dato accesso. Laddove, fosse possibile restringere l'area geografica di

incidenza dell'evento, anche in maniera approssimata, sarebbe possibile, sfruttando la stima della popolazione per griglie territoriali, meglio valutare la quantità di bambini potenzialmente coinvolti. Tale approccio necessita comunque la ricerca di collaborazioni adeguate, anche, per sfruttare esperienze di progetti precedenti che abbiano già effettuato la ricostruzioni di tali dati.

In secondo luogo, poi, per quanto riguarda la stima della probabilità con cui una certa tipologia di disastro avviene in un Paese, questa è stata semplicemente riferita alla frequenza relativa osservata nell'ultimo ventennio. E' evidente che nel corso dell'anno si accumulano ulteriori informazioni che fanno sì che eventi specifici assumano una probabilità maggiore (si pensi ad esempio alle previsioni meteorologiche a livello globale per macro-aree del pianeta che vengono prodotte dagli enti di sorveglianza sovranazionale). Per taluni tipi di disastri di origine meteorologica (tempeste, inondazioni, temperature estreme) sarebbe possibile studiare un'ulteriore miglioramento del modello introducendo un meccanismo di adattamento della stima con le informazioni provenienti da tali sistemi di sorveglianza,

Infine, abbiamo poi visualizzato la posizione di questi Paesi nella mappa fattoriale. Come abbiamo discusso in precedenza, seppur riferita ad un anno particolare, la raffigurazione fornita da tale mappa consente di selezionare le aree maggiormente bisognevoli di interventi in quanto meno dotate di fattori di resistenza.

L'esame congiunto della graduatoria e della mappa fattoriale della fragilità consente una migliore disamina dei risultati. Gli Stati Uniti, ad esempio, si collocano al 9° posto nella graduatoria di rischio di calamità (ciò è causato dal grande numero di alluvioni, tornado e altri fenomeni climatici cui va soggetto il Paese). Tuttavia, essi si collocano nell'area di non attenzione della mappa fattoriale. Diversa, invece, è la situazione del Pakistan e della Thailandia (rispettivamente al 8° e 9° della graduatoria), che si posizionano in due zone di attenzione della mappa fattoriale, essendo caratterizzati da crescenti livelli di fragilità.



Capitolo 4. Le crisi umanitarie indotte dall'instabilità politica

4.1 Introduzione

Le recenti riflessioni in tema di instabilità politica e delle cause che determinano conflitti (nelle varie forme) hanno evidenziato come lo studio dei fattori che determinano la debolezza, la fragilità ed il fallimento delle forme di governo sia un elemento fondamentale per costruire degli adeguati sistemi di preallarme e di prevenzione. Tali fattori sono strettamente connessi alla vulnerabilità della dimensione sociale così come espressa dalla struttura della società e dell'economia, dalla struttura demografica, dalle diseguaglianze, dalle forme di governo, dalla capacità di interagire con gli altri governi, dal contesto territoriale (presenza di Paesi vicini instabili o in guerra).

Marshall (2008), Mata e Ziaja (2009), Hughes et al. (2011) hanno sviluppato un'interessante sintesi dei principali progetti che sono stati intrapresi nelle ultime decadi ed una meta-analisi delle pubblicazioni prodotte fornendo una presentazione organica degli indicatori e modelli che sono stati presi in esame. Le linee di ricerca seguite negli studi, che hanno implementato delle forme di modelli statistici, sono generalmente diretti all'analisi politica e finalizzati alla successiva attuazione di meccanismi di prevenzione.

In genere, nella stima della probabilità di conflitto vengono inserite nel modello variabili esplicative ritardate, di 3-5 anni, prendendo in esame periodi temporali molto ampi. Sono inoltre modellati separatamente i vari possibili eventi che sono originati da situazioni di conflitto: guerre, genocidi, guerre etniche, rivoluzioni, etc...Il un numero di variabili che possono agire da pre-allerta è enorme: SIPRI (Stockholm International Peace Research Institute) ne elenca ben 1200. Gli studi quantitativi sono poi molto spesso integrati con analisi qualitative.

Le esperienze realizzate sono talora di vasta portata sia in termini di risorse finanziarie ed umane. Il progetto più noto è PITF Political Instability Task Force, che fu promosso dalla CIA nel 1994 per conto del governo statunitense ed ha coinvolto noti studiosi provenienti da vari settori (non solo analisti politici, ma anche economisti, econometrici, statistici etc..) (Esty et al. 1995; Gurr et al. 1999; Goldstone et al., 2005; King e Zeng, 2001). PITF è giunto alla sua quinta release e propone vari modelli focalizzati su temi particolari (guerre etniche, la dinamica dei Paesi musulmani, le rivolte etniche, i genocidi, le situazioni di transizione democratica etc..) ed aree geografiche (globale, africa subsahariana, Africa, etc). In letteratura è possibile reperire, per grandi linee, la base del progetto e con qualche ritardo il dataset dei conflitti (presso il website del Center for global policy presso la Mason University).

Vi sono numerosi altri progetti sul tema, anche se in alcuni casi l'implementazione non è realizzata o resa accessibile in maniera continuativa ed organica. Tra gli altri, citiamo: POLITY IV (Center for Systemic Peace), ICEWS (o'Brien, 2010), ACTOR (o'Brien, 2002; 2003), il modello dei conflitti e

delle rivolte etniche di Gurr e Marshal (2000; 2005); i modelli per le rivoluzioni e le guerre civili (si veda Sambanis, 2001; Fearon e Laitin, 2003); il modello per le instabilità politiche in Africa (ad esempio, Marshall, 1999); i modelli per i genocidi e i politicidi (Harff e Gurr, 1998).

Oltre alla modellistica, sono stati poi proposti numerosi indicatori fondati su analisi quantitative ed opinioni di esperti: il Conflict Barometer (Heidelberg Institute for International Conflict Research), il Fragility State Index /Failed State Index (Fund for Peace), Global Peace Index e il Global Terrorism Index (Institute for Economics and Peace), CIFP Failed and Fragile State (Carment, 2003, 2008, 2013), Index of State weakness (Rice e Patrick, 2008), State Fragility Index (Center for Systemic Peace).

Alla luce di tali esperienze e tenuto conto dell'obiettivo posto da STC centrato sulla necessità di disporre informazioni a breve termine sfruttando delle metodologie omnibus che possano essere facilmente implementate, abbiamo concentrato la nostra attenzione sulla possibilità di costruire un sistema di monitoraggio piuttosto che un vero e proprio modello. In tal modo si giungerebbe ad uno strumento che possa in maniera automatica fornire informazioni sul peggioramento delle condizioni di sicurezza delle popolazioni e che sono preludio di crisi umanitarie. A nostro avviso, infatti, l'evoluzione delle situazioni di crisi ha certamente radici in una serie di fattori (povertà economica, pressione demografica, maturità democratica etc-) ma essendo i conflitti eventi tutto sommato rari rispetto ai disastri ambientali, ci sembra che il patrimonio di dati a disposizione per l'ultimo ventennio sia poco informativo. Per questo motivo, per far fronte anche alla grande fluidità con cui gli eventi che fanno precipitare un Paese instabile o soggetto a conflitti, nella crisi umanitaria, come illustreremo nel paragrafo seguente, abbiamo preso in considerazione un archivio di news su temi di interesse per le analisi politiche.

4.2 Un sistema di monitoraggio fondato su GDELT

GDELT è un archivio di recente formazione che cataloga in una forma sintetica le news quotidiane, relative a temi di interesse socio-politico, che vengono pubblicate dalle maggiori fonti internazionali, nazionali, regionali e locali. Il database è aggiornato quotidianamente, copre gli ultimi 25 anni e tutti i Paesi del mondo (Leetaru K., Schrod, 2013). Ciascuna notizia è caratterizzata attraverso un sistema di codici che consentono di identificare numerosi chiavi tra cui l'evento, gli attori, il ruolo, il tipo di evento, il numero di menzioni, il numero di articoli, il numero di fonti, il tono medio, un punteggio di gravità. Il catalogo fa riferimento a circa 300 categorie ed a fonti che si esprimono in più di 100 lingue.

I records sono geo-referenziati relativamente al luogo ove l'evento accade e che è specificato nella notizia. Inoltre, essi contengono informazioni concernenti il gruppo etnico, religioso degli attori coinvolti. Circa un quarto di bilione di records sono disponibili in una forma completamente libera e gratuita.

Nei mesi in cui questo progetto è stato svolto, l'archivio GDELT ha subito, dopo una breve interruzione, importanti modifiche. Il sito (oggi, <http://gdeltproject.org/>) è sostanzialmente mutato rispetto alla versione originaria ed il volume di informazioni è notevolmente cresciuto. Il database è sostenuto da Google Ideas e sono oggi disponibili una serie di supporti che consentono di produrre le proprie analisi nonché una vasta selezione di commenti sulle situazioni di crisi in corso. Numerosi studiosi hanno iniziato ad usare questo archivio per le loro analisi e si va costituendo una rete di conoscenze sempre più interessante per i risultati raggiunti ed il rilievo degli studiosi coinvolti. Al momento sono anche in sperimentazione il monitoraggio di altri fonti connesse ai social media.

L'analisi statistica di tale mole di informazioni richiede, in un sistema a regime, tecniche specifiche connesse al trattamento di big data mediante tecniche di data mining.

La scelta delle aggregazioni possibili da un archivio così dettagliato è un elemento cruciale (Yonamine, 2011). Per dare un'idea della raffinatezza delle keywords a disposizione riportiamo il dettaglio di alcuni codici: le proteste, gli scontri, le violenze di massa.

14: PROTEST

- 140: Engage in political dissent, not specified below
- 141: Demonstrate or rally, not specified below
 - 1411: Demonstrate for leadership change
 - 1412: Demonstrate for policy change
 - 1413: Demonstrate for rights
 - 1414: Demonstrate for change in institutions, regime
- 142: Conduct hunger strike, not specified below
 - 1421: Conduct hunger strike for leadership change
 - 1422: Conduct hunger strike for policy change
 - 1423: Conduct hunger strike for rights
 - 1424: Conduct hunger strike for change in institutions, regime
- 143: Conduct strike or boycott, not specified below
 - 1431: Conduct strike or boycott for leadership change
 - 1432: Conduct strike or boycott for policy change
 - 1433: Conduct strike or boycott for rights
 - 1434: Conduct strike or boycott for change in institutions, regime
- 144: Obstruct passage, block, not specified below
 - 1441: Obstruct passage to demand leadership change
 - 1442: Obstruct passage to demand policy change
 - 1443: Obstruct passage to demand rights
 - 1444: Obstruct passage to demand change in institutions, regime
- 145: Protest violently, riot, not specified below
 - 1451: Engage in violent protest for leadership change
 - 1452: Engage in violent protest for policy change
 - 1453: Engage in violent protest for rights
 - 1454: Engage in violent protest for change in institutions, regime

19: FIGHT

- 190: Use conventional military force, not specified below
- 191: Impose blockade, restrict movement
- 192: Occupy territory
- 193: Fight with small arms and light weapons
- 194: Fight with artillery and tanks
- 195: Employ aerial weapons, not specified below
- 1951: Employ precision-guided aerial munitions
- 1952: Employ remotely piloted aerial munitions
- 196: Violate ceasefire

20: USE UNCONVENTIONAL MASS VIOLENCE

- 200: Use unconventional mass violence, not specified below
- 201: Engage in mass expulsion
- 202: Engage in mass killings
- 203: Engage in ethnic cleansing
- 204: Use weapons of mass destruction, not specified below
 - 2041: Use chemical, biological, or radiological weapons
 - 2042: Detonate nuclear weapons

E' possibile dall'archivio selezionare un Paese (ovvero una città od un'area geografica attraverso le coordinate), un tema (ad esempio le proteste), un periodo di tempo (l'anno 2011), la frequenza (ogni giorno) e quindi aggregare le informazioni provenienti dalle news per ottenere, ad esempio, il numero di articoli che riportano a quel tema, oppure il numero di fonti che ne parlano oppure il numero di eventi (il numero di proteste in atto in un giorno).

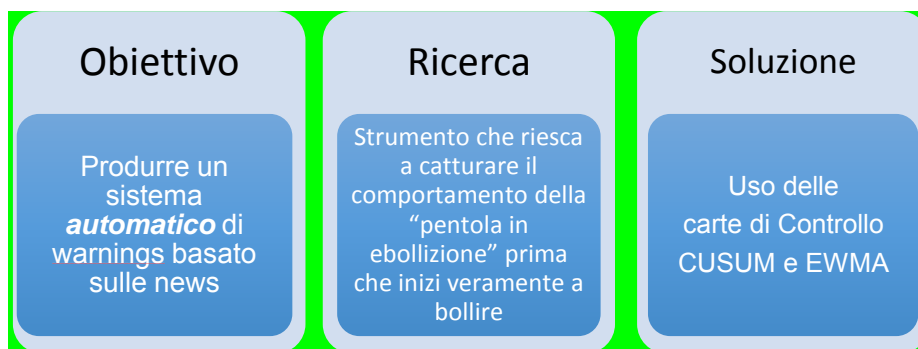
In questo lavoro abbiamo scelto un caso per individuare un approccio, di semplice implementazione, che possa fornire dati utili per un futuro sistema informativo di STC. Le condizioni che si determinano in un Paese che sperimenta una crisi che sfocia in fenomeni estremi di instabilità politica, di conflitti interni o anche di guerra, sono frutto di periodi più o meno lunghi di incubazione in cui si pongono in atto una serie di azioni violente. La situazione può essere raffigurata con una metafora come "una pentola che via via raggiunge lo stato di ebollizione". Un sistema di allarme dovrebbe essere in grado di segnalare che il momento in cui parte il processo di ebollizione associando diversi livelli di attenzione al perdurare delle condizioni di pericolo, di modo che possa essere posto in atto una procedura più raffinata di monitoraggio.

Rispetto alla identificazione di possibili luoghi ove si potrebbe generare una crisi che richieda l'intervento di organizzazioni umanitarie, l'elemento che ci sembra importante è disporre di un sistema che in maniera automatica possa fornire una prima segnalazione da inserire all'interno di una cornice più ampia derivata da altri strumenti di analisi.

E' bene sottolineare che in questa fase preliminare, non viene affrontato il problema dell'architettura del sistema informativo, essendo noi concentrati a

trovare risposte ai quesiti fondamentali concernenti la possibilità di riconoscere relazioni e strumenti utili per l'obiettivo posto da STC. Questo problema deve essere oggetto della seconda fase ove sarà necessario tra l'altro integrare varie tipi di analisi sia quantitative che qualitative.

Pertanto, l'analisi delle news, tenuto conto dell'ampiezza delle fonti che GDELT considera, ci sembra una strada interessante per gli obiettivi informativi di breve/brevissimo periodo. Sebbene si possa ipotizzare che vi siano fenomeni di sfocamento del rilievo degli eventi dovuto al noto effetto del calo dell'attenzione dei media, questo effetto è molto attenuato in un archivio come GDELT ove non sono presenti solo le fonti internazionali e le agenzie di stampa più note (quali Reuters, France Press, etc.), ma anche la stampa regionale e locale che certamente non abbandona facilmente dei temi importanti che preludono a crisi di rilievo.



Abbiamo quindi considerato il caso della Siria, in particolare, sono stati considerati tutti gli eventi riconducibili alle parole chiave: proteste e scontri.

E' bene sottolineare che GDELT sta anche perseguendo un obiettivo più raffinato di ricostruzione del dato del numero di vittime legato a ciascun evento, ma dalle analisi svolte sul breve periodo disponibile nel marzo del 2014, allorché si è consultato il sito, l'algoritmo di ricostruzione doveva essere ancora molto migliorato per enucleare la stima delle vittime dalle news e contestualizzarle all'istante temporale corretto.

Nelle Figure 4.1.a-b riportiamo il grafico del conteggio degli eventi riconducibili alle parole chiave: proteste e scontri, e riferiti alle news catalogate ogni giorno dal 1.1.2009 al 10/3/2014. Dal grafico delle corrispondenti serie storiche lisciate (attraverso un media del numero di eventi riferita ad una finestra di 13 giorni) evidenzia molto chiaramente l'andamento del singolo fenomeno che ha, fasi di espansione e contrazione. Dal confronto tra i pattern di proteste e scontri è molto evidente lo shift da una forma di lotta all'altra e l'avanzare dell'escalation delle violenze. In Appendice A.5, solo a titolo informativo abbiamo riportato una time-line della crisi.

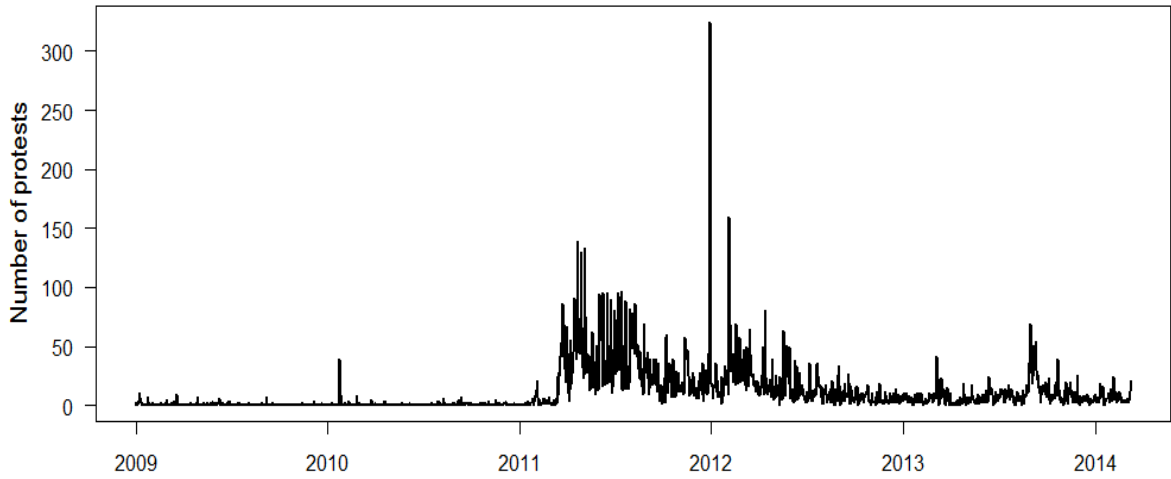


Figura 4.1.a Proteste (dati giornalieri)

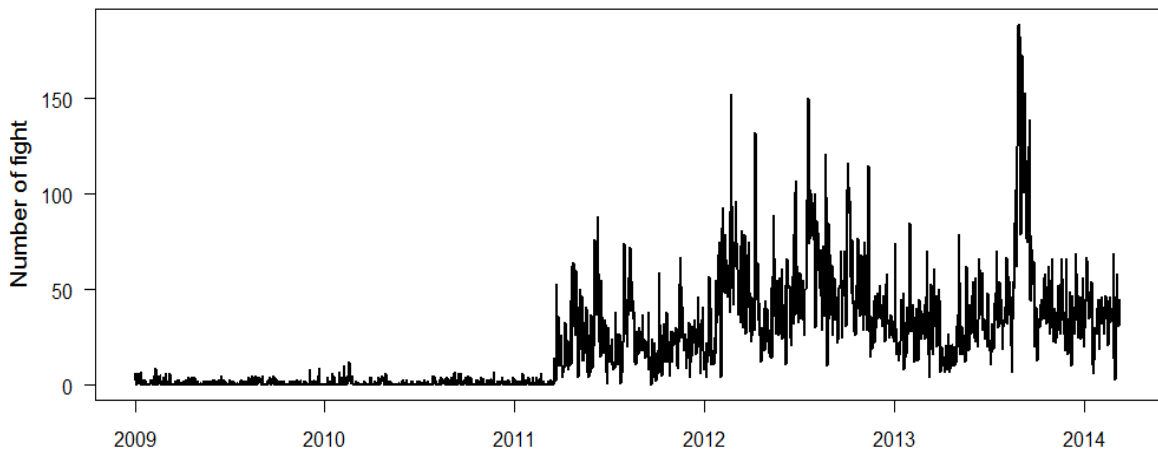


Figura 4.1.b Scontri (dati giornalieri)

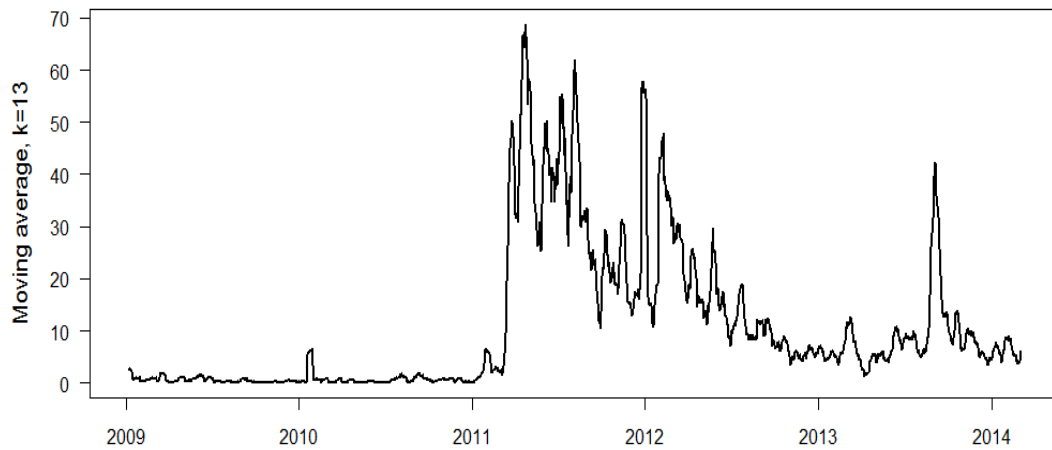


Figura 4.2.a Proteste (media mobile di 13 termini)

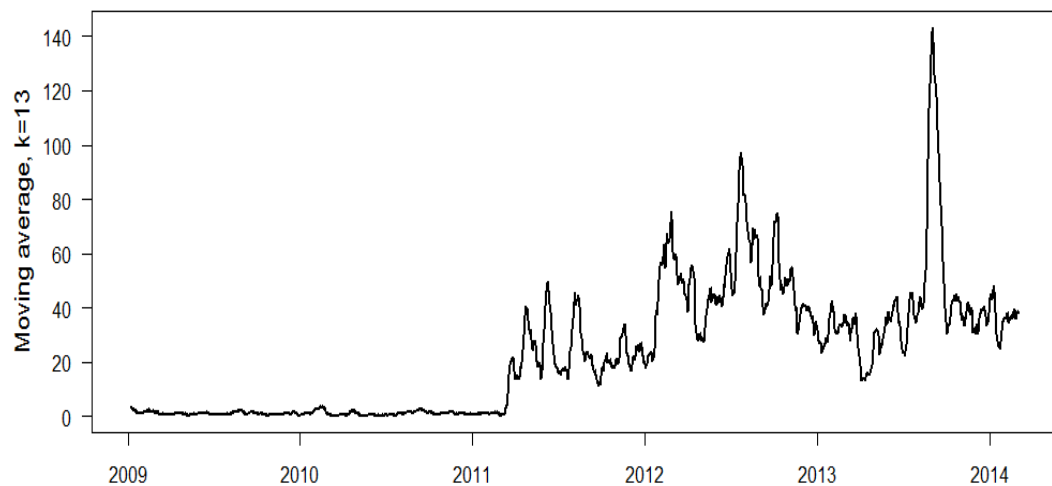


Figura 4.2.b Scontri (media mobile di 13 termini)

Nel seguito, simuleremo la situazione del monitoraggio in cui ogni istante si aggiunge un'informazione ulteriore circa il numero di scontri (proteste) che sono riportati nelle news implementando un processo fondato sulle carte di controllo. L'obiettivo è quello di catturare in maniera automatica l'istante in cui si ha un cambiamento nel pattern della serie.

Si osservi che in questa fase sono stati utilizzati i dati aggregati a livello di mesi per poter verificare l'esistenza di relazioni sfruttabili nella seconda fase di sperimentazione e attuazione del sistema informativo, ma la medesima metodologia può essere implementata considerando un diverso livello di aggregazione (settimanale o quindicinale) per consentire una maggiore flessibilità ed aderenza dei risultati. Tale scelta ovviamente dipende

strettamente dalle modalità e dalle risorse con cui si intende poi affrontare la fase di disegno e costruzione del sistema informativo.

Di seguito illustreremo in breve gli aspetti salienti delle carte di controllo che costituiscono lo strumento statistico che utilizzeremo per analizzare alcuni dei fenomeni di cui GDELT fornisce informazioni.

4.3 Le carte di controllo

Il numero di news relative ad eventi potenzialmente conflittuali o il numero di eventi critici costituiscono degli indicatori il cui andamento nel tempo può essere monitorato. In particolare, essi possono essere assimilati ad un processo che assume, nei periodi di stabilità, a seconda dell'area geografica (o Paese o regione), un certo livello medio ed una certa variabilità. Quando nel territorio di interesse si producono condizioni destabilizzanti e si pongono in atto le forze e gli acceleratori che portano al delinearsi di situazioni di instabilità grave e quindi di conflitti ed atti violenti, allora il livello medio di tali indicatori cambia drasticamente senza che vi sia più la possibilità di tornare a livelli pre-conflitto in tempi rapidi.

Per poter scoprire in modo precoce, questi segnali di un mutamento nei livelli medi di conflittualità riportata nelle news è necessario tenere sotto controllo costante l'andamento del processo. A nostro avviso, questa analisi è assimilabile a quanto generalmente viene realizzato nel controllo statistico dei processi con *le carte di controllo*.

4.3.1 Le carte di controllo CUSUM e EWMA

Le carte di controllo sono strumenti utilizzati per monitorare processi sequenziali (tipicamente di produzione) in cui i dati vengono raccolti ordinatamente nel tempo. Attraverso questo strumento, viene esaminato l'andamento passato del processo, per controllare e valutare l'andamento in corso, e per prevedere i futuri risultati.

Sebbene le carte di controllo siano utilizzate per vari scopi, l'obiettivo tradizionale è quello di monitorare un processo per identificare gli scostamenti rispetto ad un comportamento cosiddetto 'normale' e 'consistente con le attese.

Nel nostro contesto in cui stiamo invece operando, la carta di controllo può essere utile per evidenziare il momento in cui inizia un periodo di crisi ed il cui permanere prelude alla crisi umanitaria.

Esistono diversi tipi di carte di controllo che consentono di monitorare aspetti rilevanti del processo. La variabile oggetto di monitoraggio può essere numerica oppure qualitativa; il monitoraggio può riferirsi ad un solo fenomeno) o più fenomeni allo stesso tempo.

Aspetto di interesse per i nostri fini è che le carte di controllo forniscono una rappresentazione grafica dei dati relativi al processo e consentono di distinguere le cause speciali della variabilità dalle cause comuni. Le cause

comuni sono così dette poiché producono la normale e naturale variabilità del fenomeno (quella che per noi è la variabilità del numero di notizie di eventi conflittuali in un periodo di stabilità); le cause speciali sono quelle che, allorché intervengono, determinano un andamento del processo che è estraneo a quanto osservato nel periodo di normalità (ad esempio l'inizio del periodo di instabilità). Il processo è "in controllo" fino a quando su di esso agiscono solo cause comuni. Il processo va "fuori controllo" quando intervengono cause speciali.

Le carte di controllo sono dei test statistici di ipotesi sequenziali (illustrati graficamente) per determinare di volta in volta se il processo è in controllo (ipotesi nulla) o se il processo è fuori controllo (ipotesi alternativa).

In questo lavoro abbiamo optato per due tipi di carte di controllo per monitorare il livello medio del processo: CUSUM e EWMA.

Si suppone che i dati osservati provengano da un processo osservato nel tempo. Si seleziona un periodo in cui certamente il processo è in controllo (nel senso che opera secondo le attese). Si raccolgono dei campioni in questo periodo e per ciascun campione si calcola un qualche indicatore statistico. La carta di controllo consiste nel grafico rispetto al tempo di tale indicatore per i vari campioni osservati sequenzialmente. La scelta di tale statistica è critica ed è condizionata dall'obiettivo per il quale l'analisi statistica è condotta.

Nella situazione di nostro interesse, il processo è costituito dal conteggio degli eventi (ad esempio, il numero di scontri in un mese). Nella fase di successiva sperimentazione, potrebbe essere utile raffinare la tecnica per introdurre possibili miglioramenti che tengano conto della dipendenza tra i dati.

Nel periodo in cui il processo è sotto controllo, la media di questo processo è molto bassa (se si tratta di un Paese pacifico) ovvero ha un livello contenuto e paragonabile a quello della naturale conflittualità del Paese. E' necessario disporre quindi di una carta di controllo che sia in grado di individuare la presenza di un aumento, anche piccolo, ma costante del livello medio di conflittualità, che poi col tempo potrebbe innescare un conflitto più rilevante.

Le carte CUSUM (cumulative sum ovvero a somme cumulate) risultano utili a tale scopo poiché prendono in considerazione la somma degli scarti delle osservazioni dal livello medio di riferimento. Tale somma viene calcolata tutte le volte che una nuova osservazione si aggiunge all'insieme di dati disponibili. Pertanto, uno scostamento sistematico, anche piccolo, cumulandosi nel tempo ad altri scostamenti sistematici, renderà sempre più evidente l'allontanamento del processo dalla sua condizione di normalità.

La carta di controllo CUSUM è rappresentata di solito attraverso il grafico di due traiettorie¹: la prima è data dalla somma progressiva dei soli scostamenti

¹Nelle carte CUSUM si utilizzano due statistiche dette rispettivamente CUSUM unilaterale superiore e CUSUM unilaterale inferiore che sono definite come segue:

$$C_i^+ = \max[0, x_i - (\mu_0 + K) + C_{i-1}^+]$$

$$C_i^- = \max[0, (\mu_0 + K) - x_i + C_{i-1}^-]$$

dove μ_0 è il valore target, mentre K è l'ammontare di scostamento ammesso. La statistica che viene mostrata in un grafico è:

$$C_i = \sum_j^i \bar{x}_j - \mu_0,$$

positivi dal valore centrale, la seconda dalla medesima somma ma dei soli scarti negativi. Per tale motivo le carte CUSUM sono molto più sensibili nell'individuare scostamenti dal livello medio della caratteristica che si sta monitorando. Tale aspetto può però, in taluni casi, costituire uno svantaggio, dal momento che le carte CUSUM possono risentire di cambiamenti avvenuti in tempi passati e piuttosto remoti, perché la statistica test conserva tutte le informazioni dall'inizio del controllo del processo.

Tale inconveniente viene risolto dalle carte EWMA (Exponential weighted moving average, ovvero le carte di controllo a medie mobili con pesi esponenziali), ove le osservazioni hanno peso diverso. In particolare, alle osservazioni passate viene assegnato un peso che va via via degradando a zero a mano a mano che l'osservazione è più lontana nel tempo di modo che le informazioni recenti raccolgono un peso maggiore.²

In questo caso se il Paese in analisi passa un periodo di instabilità che però non sfocia in un conflitto di larga portata, mentre nella carta CUSUM questo spostamento resterebbe comunque visibile, nella carta EWMA tale effetto dopo un certo periodo si perde nella carta.

Per tali motivi noi proponiamo l'uso congiunto di entrambe le carte CUSUM e EWMA.

L'utilizzo delle carte di controllo è molto semplice poiché esse mostrano un tracciato della statistica di interesse assieme a delle bande costituite dai limiti di controllo, che dipendono dalla variabilità del processo nello stato di controllo (ovvero dalle cause comuni) e dal livello di significatività prescelto. Generalmente, i limiti vengono posizionati a ± 3 scarti quadratici medi della misura considerata (si ricordi che nella distribuzione normale l'intervallo $\mu \pm 3\sigma$ contiene il 99.73% dei valori della popolazione). Laddove vi siano degli sforamenti in alto o in basso del tracciato dell'indicatore alla base della carta, occorrerà concludere che si è verificata una anomalia rispetto alla situazione standard e quindi è lecito far scattare un livello di allarme.

Nel nostro caso, saranno le condizioni pregresse di un Paese a suggerire sia il periodo temporale che il livello di riferimento del processo in controllo. In particolare, questi devono essere identificati dagli analisti qualitativi.

4.3.2. Il monitoraggio delle situazioni di crisi

La procedura che si propone è la seguente. Dato il territorio di riferimento e considerati i dati giornalieri del numero di news, si considera la serie delle medie mensili (o settimanali per avere una carta più sensibile) in un periodo di relativa stabilità. Su questi dati del "processo in controllo" si calcolano il livello medio e lo scarto quadratico medio del processo che poi verranno utilizzati come riferimento. Poiché il monitoraggio avviene istante dopo istante e quindi generati numerosi grafici che rappresentano l'aggiornamento della carta di

² La statistica test assume la forma $z_i = \lambda x_i + (1 - \lambda)z_{i-1}$, in cui λ è il peso della nuova osservazione x_i rispetto alla storia passata z_{i-1} . Per convenzione z_0 viene posto uguale a μ_0 , ovvero il valore di riferimento.

controllo, per dare un'idea della loro successione riportiamo i grafici effettuati per gli "scontri", utilizzando come periodo di riferimento dal gennaio 2009 al dicembre 2009 e monitorando fino ai seguenti istanti:

- a) dicembre 2010
- b) marzo 2011
- c) giugno 2011
- d) dicembre

Per la sola carta EWMA riportiamo anche il grafico all'ultimo istante disponibile, febbraio 2014, per far apprezzare la differenza tra i due approcci. In realtà, va sempre ricordato che la carta di controllo serve ad identificare i momenti in cui il processo va fuori controllo e non per monitorare l'andamento della crisi.

La Figura 4.3 riporta la carta di controllo CUSUM per il periodo di riferimento e per il primo anno di monitoraggio, il 2010. Nella lettura vanno considerati i due percorsi (solo i dati positivi o solo i dati negativi) separatamente e le bande di riferimento. Noi siamo interessati, ai soli sforamenti rispetto alla banda superiore.

Si osservi come entrambe le carte di controllo individuano il marzo 2011 come inizio del periodo di crisi e come i mesi successivi vanno tutti nella stessa direzione dimostrando che il processo ha intrapreso un percorso oramai fuori controllo e per il quale non è possibile prevedere un rientro nelle bande di stabilità.

L'applicazione della carta di controllo con una frequenza più alta (dati quindicinali o settimanali), sebbene possa risentire di una maggiore volatilità, può certamente dare ulteriori informazioni sul progressivo avanzare della crisi.

Da un punto di vista della costruzione di un sistema informativo bisogna inquadrare questo tipo di tecnica all'interno di un'architettura complessiva che è diretta alla contemporanea analisi di tutti i Paesi del mondo. Per cui un sistema di questo tipo, con parole chiave ben selezionate, può consentire una esplorazione a largo raggio delle possibilità insite nell'uso di GDELT.

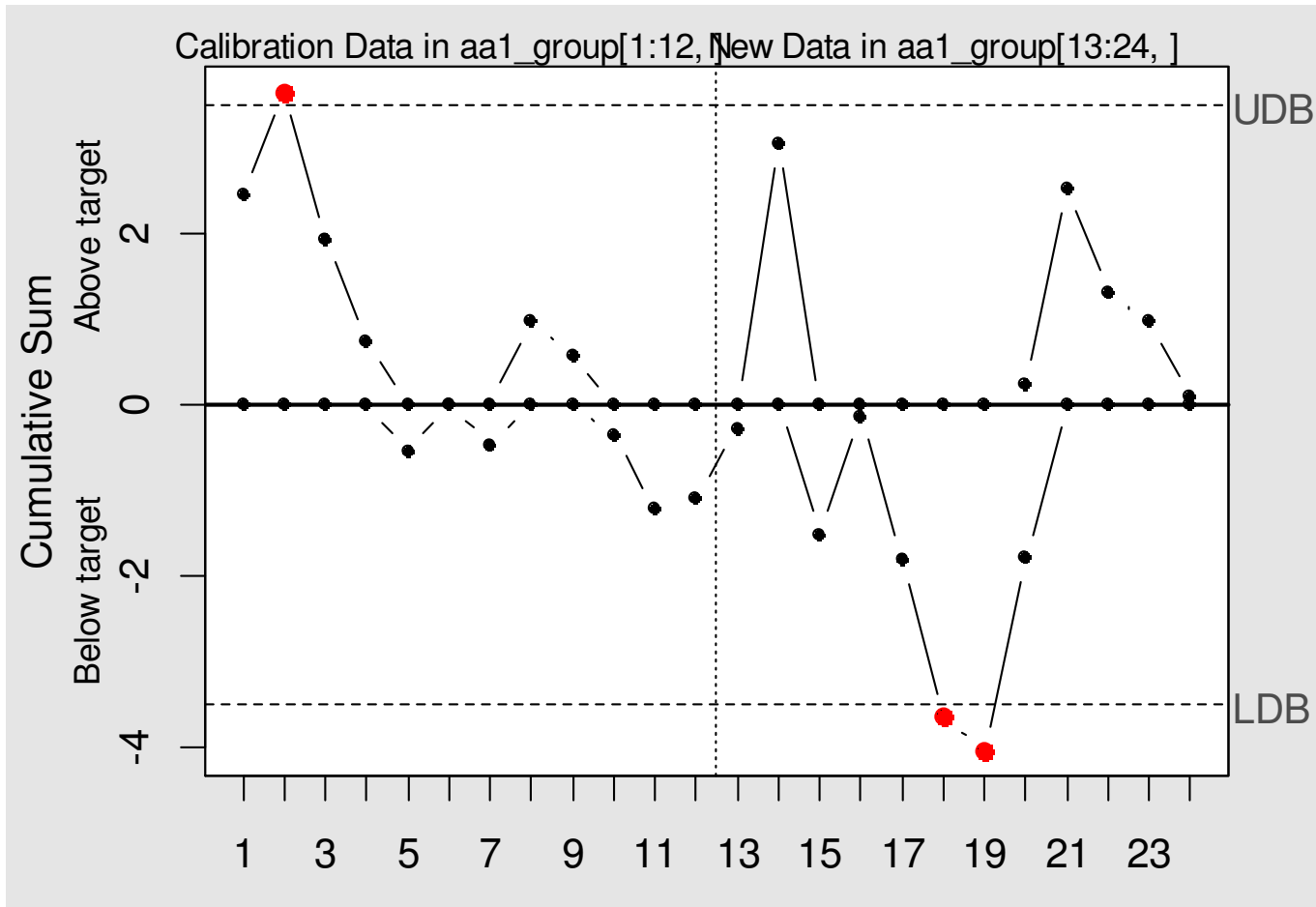


Figura 4.3 Carta di controllo CUSUM (osservazioni 01.2009-12.2010)

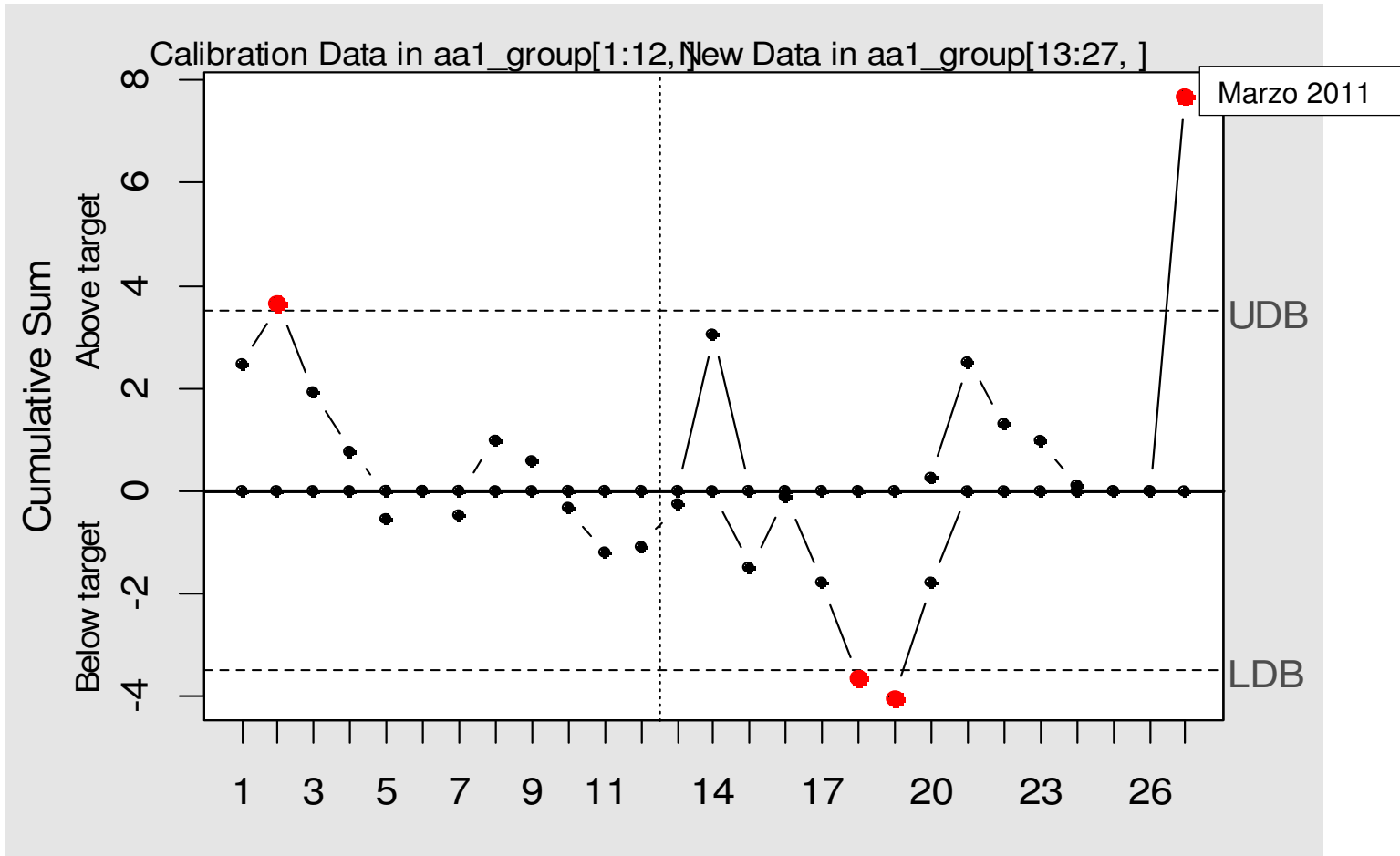


Figura 4.4 Carta di controllo CUSUM (osservazioni 01.2009-03.2011)

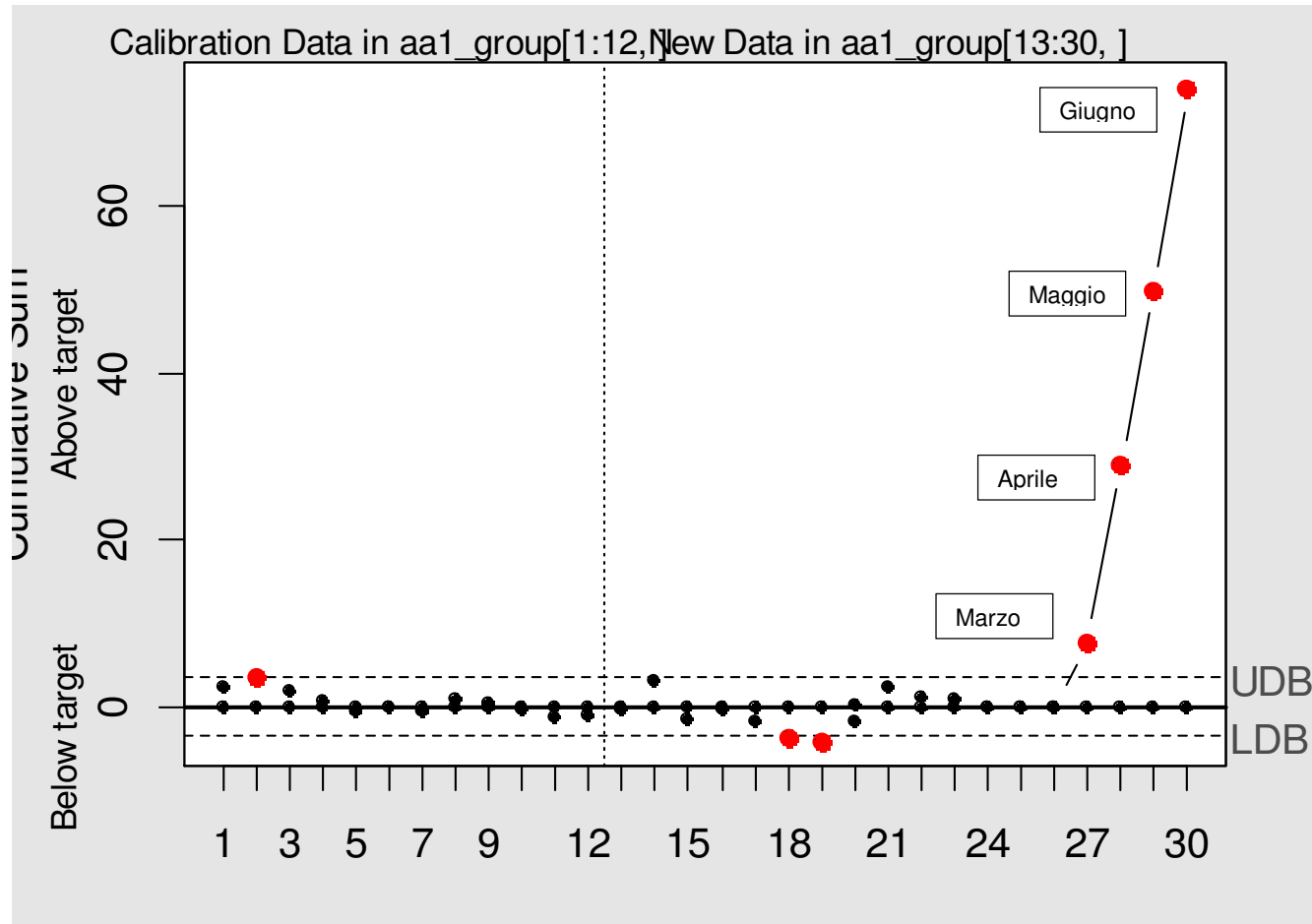


Figura 4.5 Carta di controllo CUSUM (osservazioni 01.2009-06.2011)

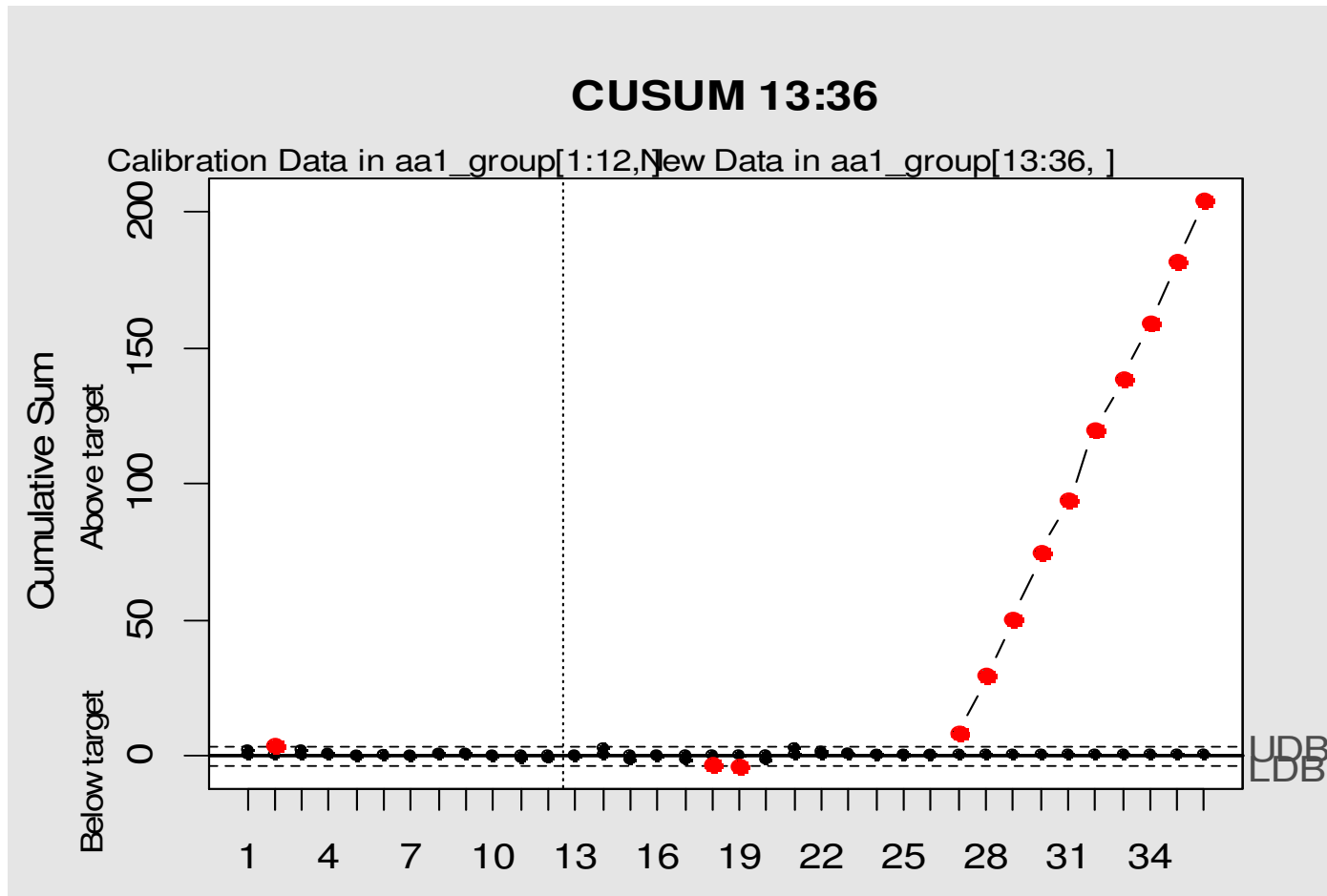


Figura 4.6 Carta di controllo CUSUM (osservazioni 01.2009-12.2011)

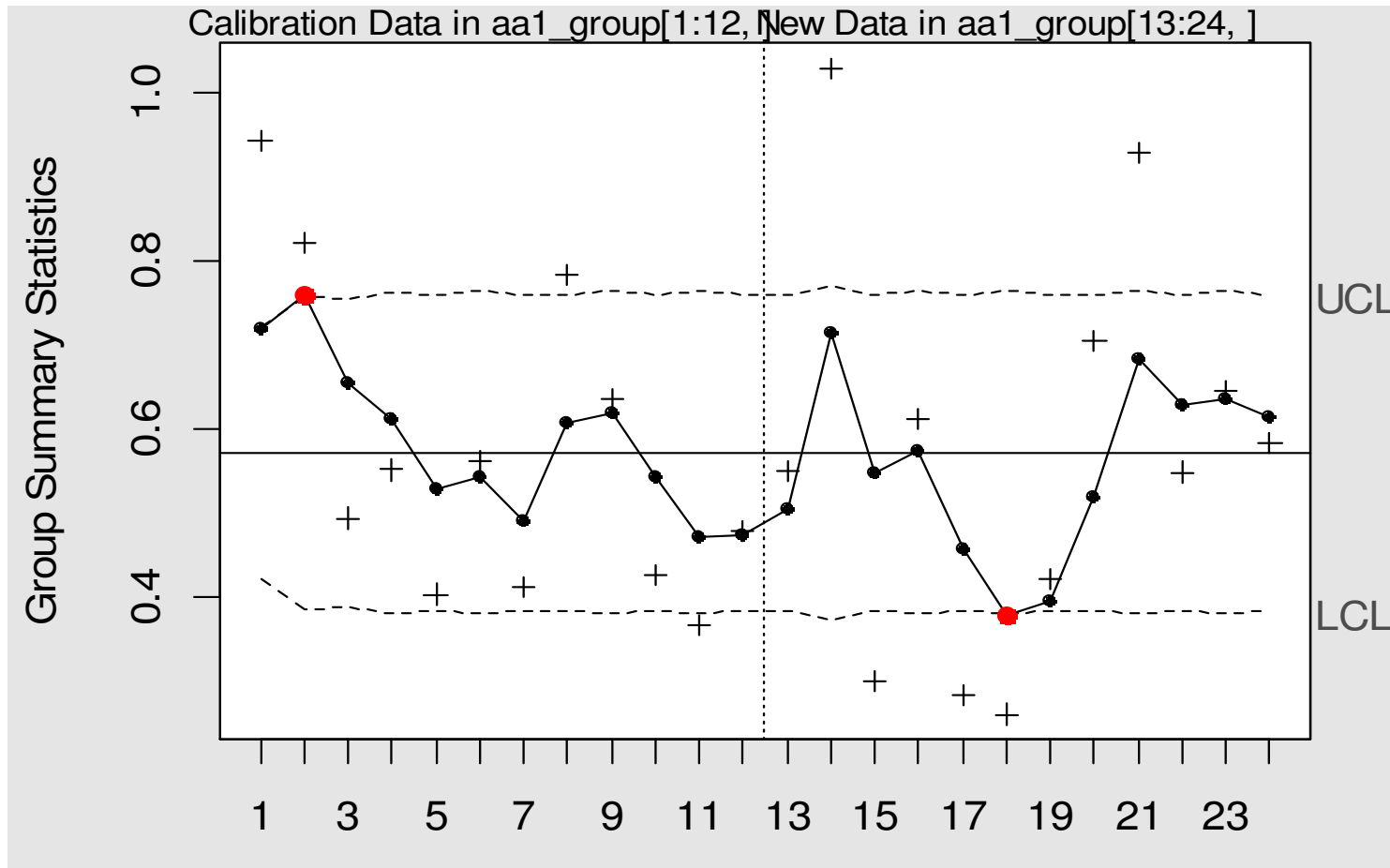


Figura 4.7 Carta di controllo EWMA

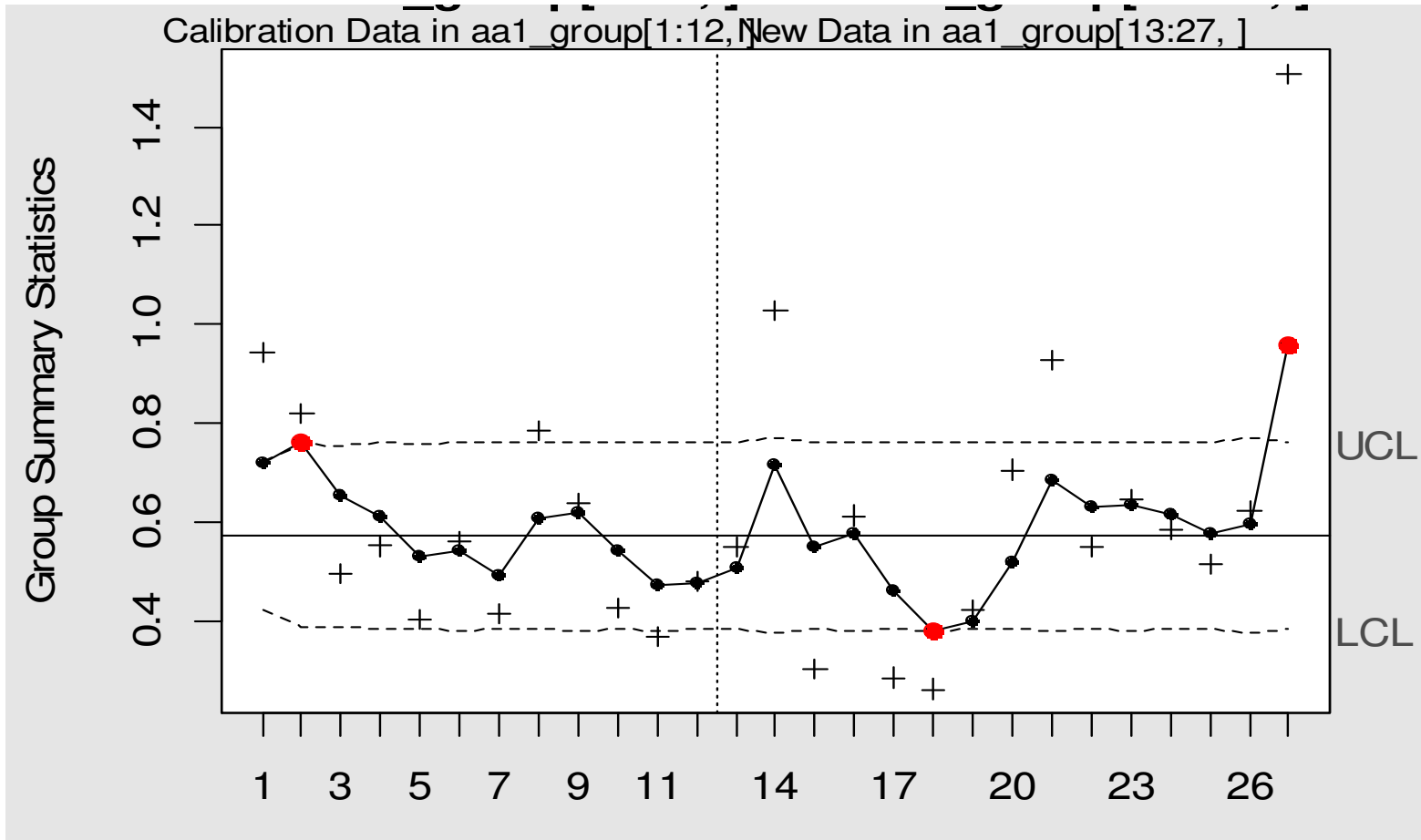


Figura 4.8 Carta di controllo EWMA (osservazioni 01.2009-03.2011)

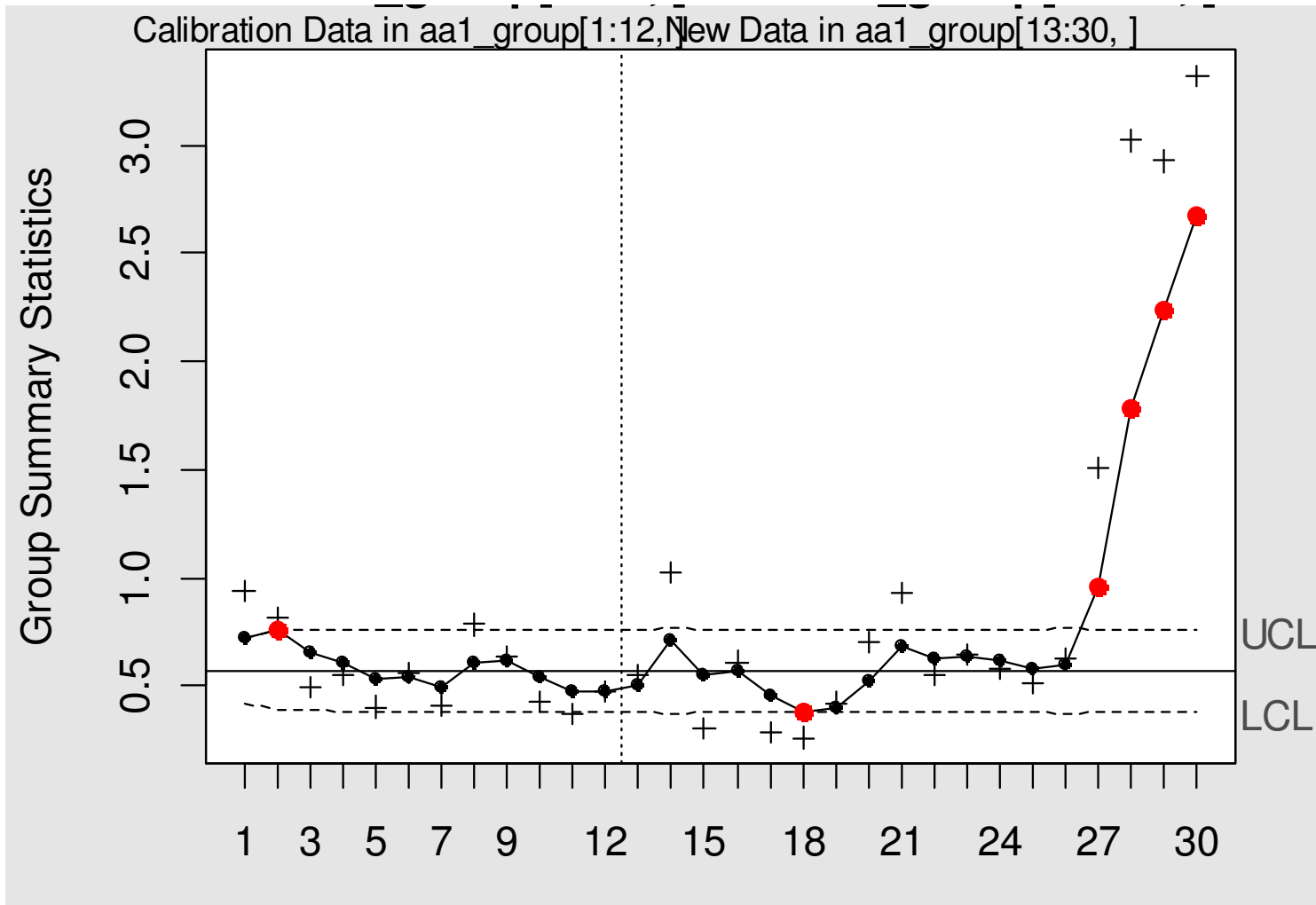


Figura 4.9 Carta di controllo EWMA (osservazioni 01.2009-06.2011)

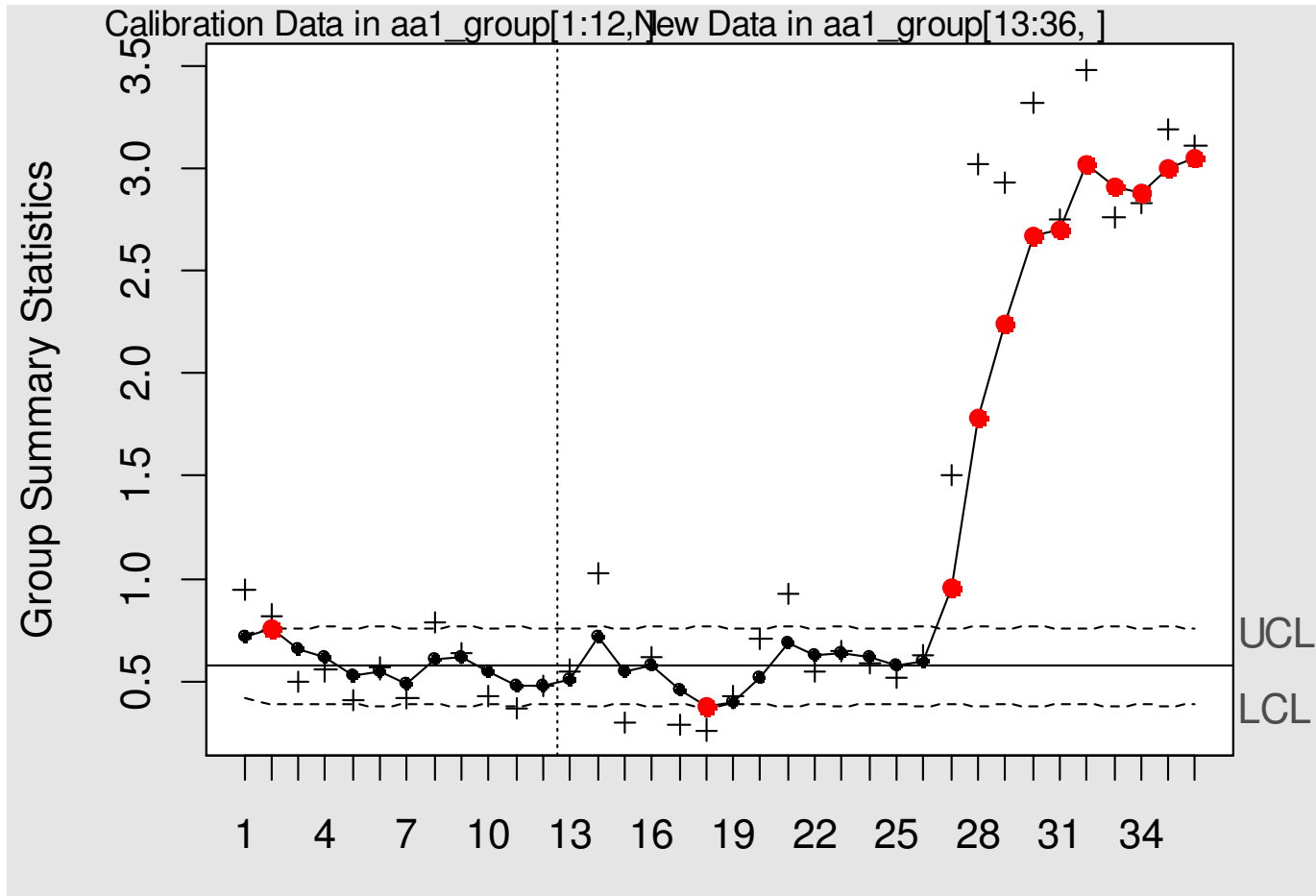


Figura 4.10 Carta di controllo EWMA (osservazioni 01.2009-12.2011)

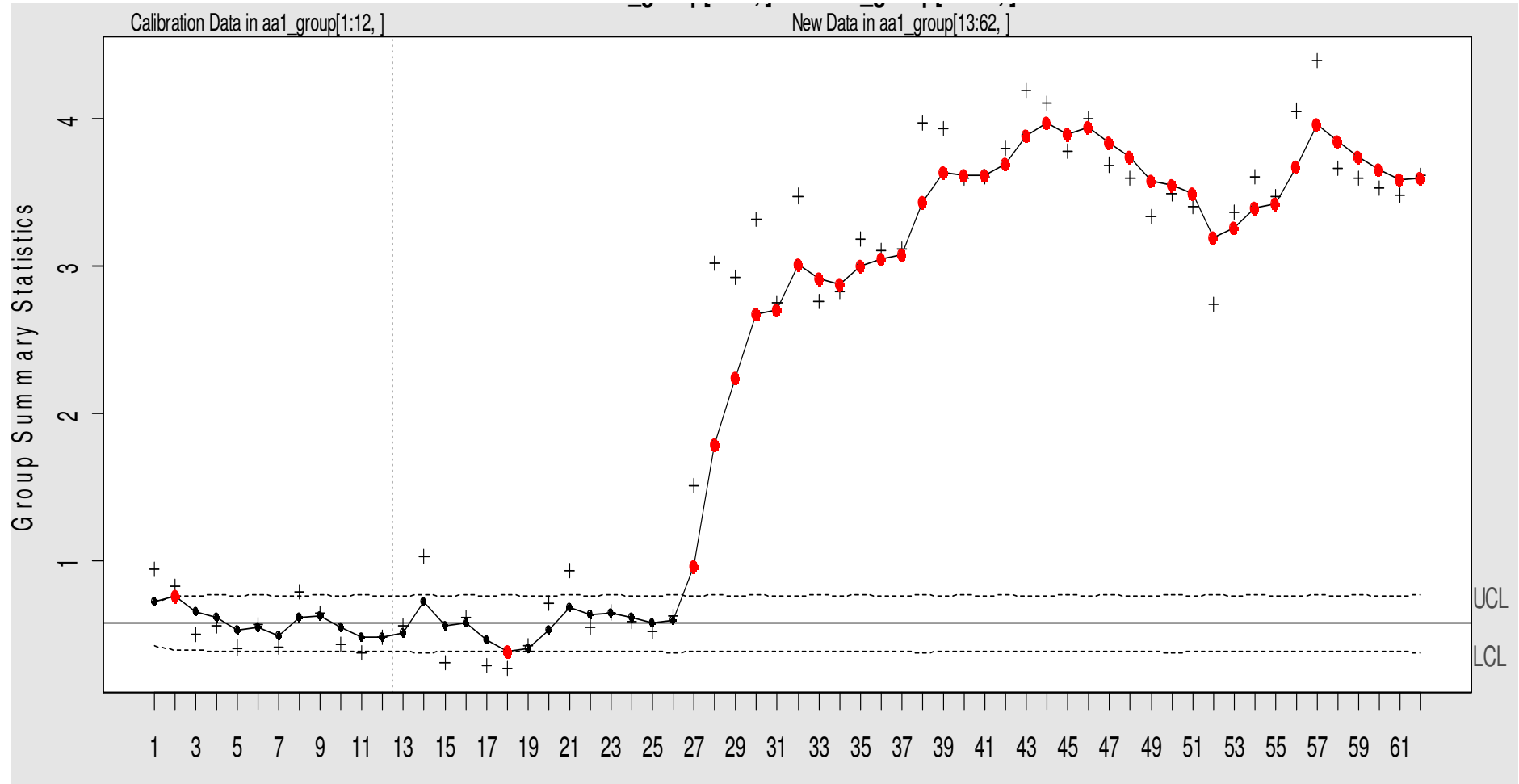


Figura 4.11 Carta di controllo EWMA (osservazioni 01.2009-02.2014)

Si osservi che per poter comprendere quando un processo in fase esplosiva rientra nei canoni di normalità, si potrebbe utilizzare la carta di controllo in forma completamente opposta. La carta di controllo verrebbe attestata e riferita al periodo di crisi, dopodiché verrebbero identificati gli sforamenti verso il basso valutandone la persistenza nel tempo. Questi istanti temporali potrebbero indicare il probabile inizio del processo di rientro nella condizione 'normale' fornendo il punto di partenza per le successive analisi qualitative più approfondita. Ovviamente, questi segnali positivi dovrebbero essere molteplici, nel senso che dovrebbero riferirsi a più indicatori, al fine di evitare che la diminuita presenza nelle news sia intesa come assenza del fenomeno. In ogni caso, questo aspetto ci sembra fuori dall'obiettivo del presente progetto che è, invece, focalizzato sull'identificazione dei momenti di crisi.

4.4 Una rappresentazione dei Paesi rispetto ai precursori di instabilità

Sebbene le carte di controllo possano essere utilizzate isolatamente come strumento di monitoraggio automatico su larga scala (se ovviamente alimentate da opportune procedure informatiche in grado di estrarre ed elaborare automaticamente le informazioni dall'archivio considerato), a nostro avviso, un sistema di early warning maggiormente efficace si gioverebbe dell'uso congiunto delle mappe fattoriali su cui "accendere" le spie di attenzione.

Abbiamo, quindi, effettuato una rappresentazione dei Paesi, nuovamente, mediante una tecnica multivariata, sulla base delle sole variabili che gli studi e le teorie politiche recenti hanno individuato come significative nell'accrescere la probabilità di conflitti o la fragilità di uno stato. In particolare, tenendo conto della descrizione fornita da Marshall (2008) abbiamo verificato che nel nostro dataset erano presenti 57 variabili di interesse:

- caratterizzazione socio-demografica (popolazione rurale, superficie agricola, tasso di urbanizzazione, densità della popolazione, pressione demografica delle classi di età giovani, analfabetismo/alfabetismo, deficit in calorie, migliori condizioni igienico-sanitarie, HIV, aspettativa di vita, mortalità materna, malnutrizione, mortalità infantile)
- contesto politico-religioso (libertà politiche, diritti civili, corruzione, diritti di voto, frazionamento etnico, religioso, linguistico, %popolazione musulmana, %pop.religione prevalente, %pop.gruppo etnico prevalente)
- sicurezza (spese militari, armi, indice SIPRI, rifugiati e sfollati, caratteristiche del territorio)
- commercio (importazioni, esportazioni, indicatore di apertura ai commerci)
- economia (GDP, tasso di crescita del GDP, Indice di Gini, produzione di petrolio, indici di povertà, disoccupazione, CO2, deforestazione, energia elettrica, aiuti dall'estero, debito nazionale)

In questo caso si è scelto di sintetizzare i dati attraverso una Analisi in Componenti Principali (descritta nel paragrafo relativo alla Analisi Fattoriale Multipla). I primi tre assi fattoriali spiegano circa il 43% della variabilità totale, con il primo che da solo raccoglie più di un quarto dell'informazione totale.

Esaminando il cerchio delle correlazioni del primo e secondo fattore e del primo e del terzo fattore, oltre che i contributi assoluti e relativi, si ritrova una struttura fattoriale molto simile, ma non sovrapponibile, a quella risultante dall'Analisi Fattoriale Multipla in relazione.

Si noti che il secondo e terzo asse sono stati ribaltati rispetto alle soluzioni fornite in automatico (il ribaltamento degli assi fattoriale non ha alcun effetto sulle analisi), per avere una raffigurazione analoga quella già discussa in precedenza (paragrafo 2.4.3).

Sulla sinistra troviamo la dimensione caratterizzata dal un ampio benessere socio economico, in particolare con alta aspettativa di vita, alto reddito e alta scolarità. Sulla destra all'opposto c'è la povertà, la malnutrizione e elevati tassi di mortalità e ampia diffusione di malattie come la malaria e l'HIV.

Il secondo asse oppone in basso destra la percentuale di musulmani a quella, in alto a sinistra di cristiani. Sempre sul secondo asse si associano livelli bassi di libertà civili e politiche con elevate spese in armamenti e allo stesso tempo una quota prevalente di popolazione giovanile. Dall'altro lato invece si trova la presenza delle donne in politica e in economia.

Il terzo asse invece è maggiormente caratterizzato dalla struttura economica di alti redditi, il ricorso al debito estero, il consumo e la produzione di energia elettrica, così come la produzione di petrolio.

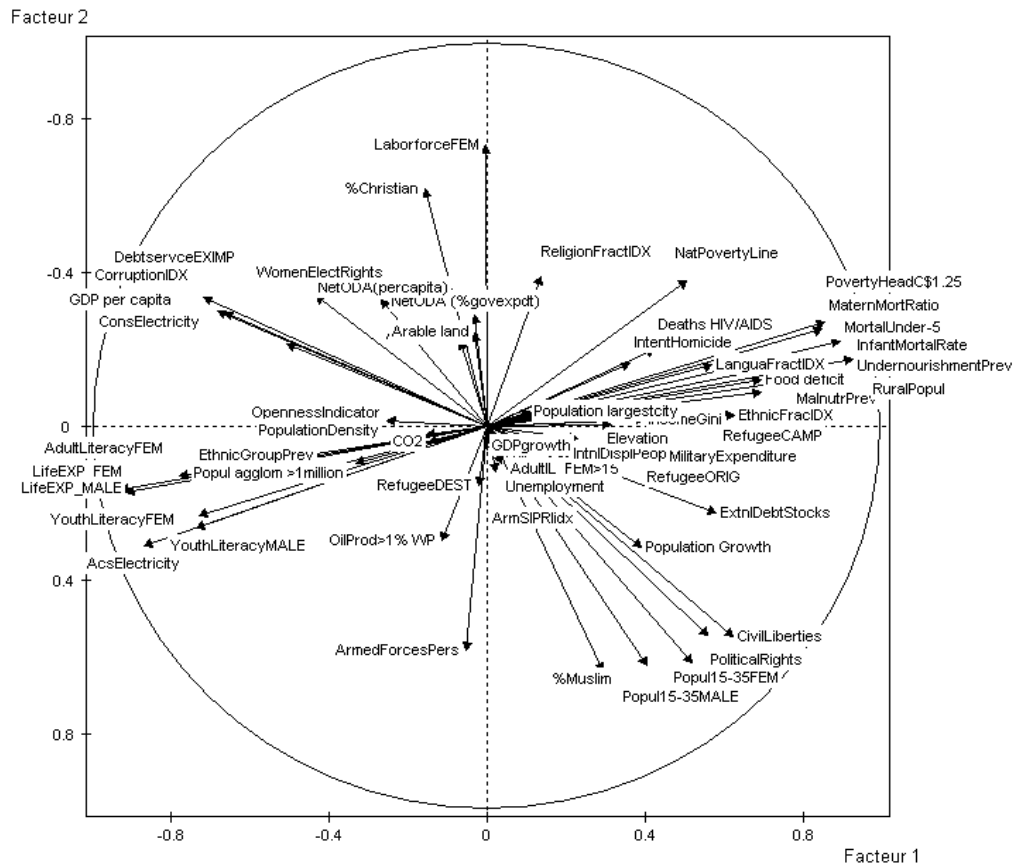


Figura 4.12 Cerchio delle correlazioni “variabili influenti conflitti/instabilità” (fattore 1/ fattore2)

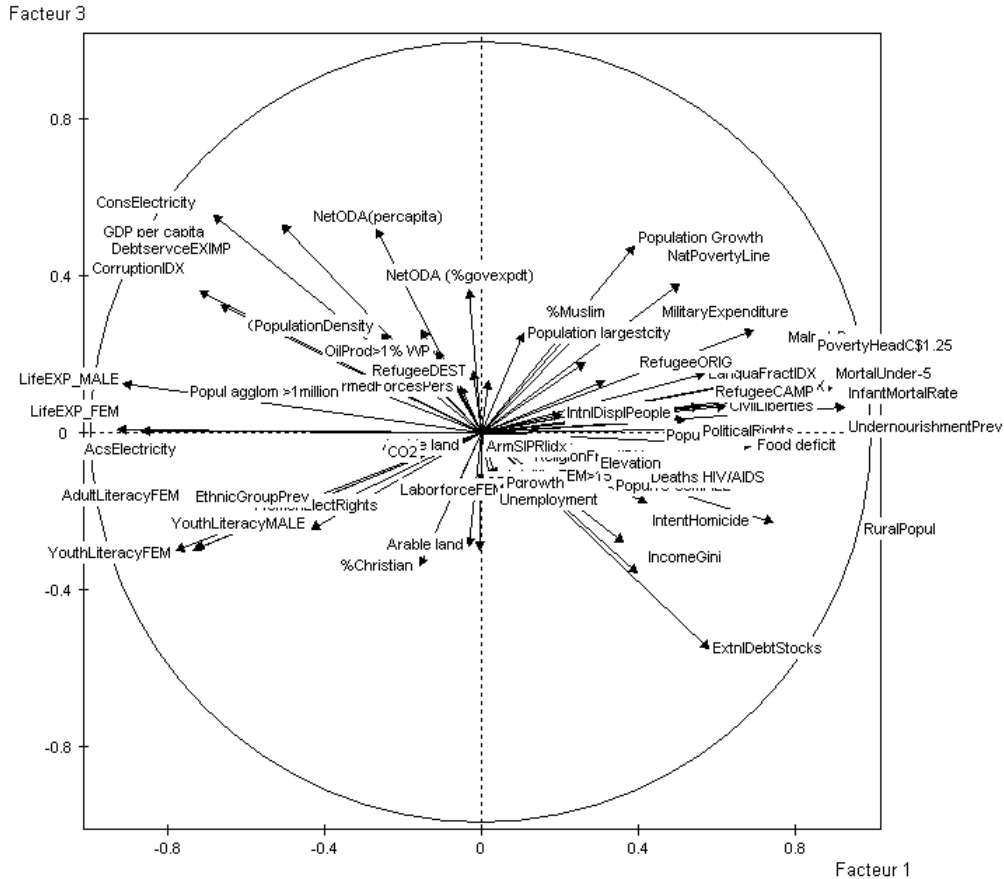


Figura 4.13 Cerchio delle correlazioni “variabili influenti conflitti/instabilità” (fattore 1/ fattore3)

A questo punto abbiamo raffigurato i Paesi sui primi tre assi fattoriali e abbiamo colorato in rosso i Paesi in cui c'è un conflitto in atto e in viola quelli in cui c'è un conflitto interno anche di tipo etnico. Si nota come la maggior parte dei Paesi in conflitto si trova sul bordo della nuvola dei punti, lungo una serie di curve di crescente instabilità potenziale. Ricordiamo anche che la mappa è costruita con delle variabili che rappresentano la foto al 2010, mentre i conflitti sono rilevati al 2013. I principali modelli attualmente in uso per il calcolo della probabilità di conflitto, essendo tarati per mettere in moto eventuali interventi, utilizzano variabili ritardate tra 3 e 5 anni. Per cui, la raffigurazione ottenuta, a nostro avviso, è in grado di descrivere piuttosto bene le aree geografiche di instabilità via via crescente.

L'uso congiunto della graduatoria e della mappa fattoriale della fragilità consente di migliorare l'utilizzo della graduatoria stessa. Si considerino ad esempio gli Stati Uniti che si collocano al 9° posto nella graduatoria di rischio di calamità (si pensi alla grande quantità di alluvioni, tornado e altri fenomeni climatici cui va soggetto il paese). Gli Stati Uniti, tuttavia, si posizionano nell'area di non attenzione della mappa fattoriale. Diversa, invece, la situazione del Pakistan e della Thailandia (rispettivamente al 8° e 9° della graduatoria) che invece di collocano in due zone di attenzione della mappa fattoriale, caratterizzate da crescenti livelli di fragilità.

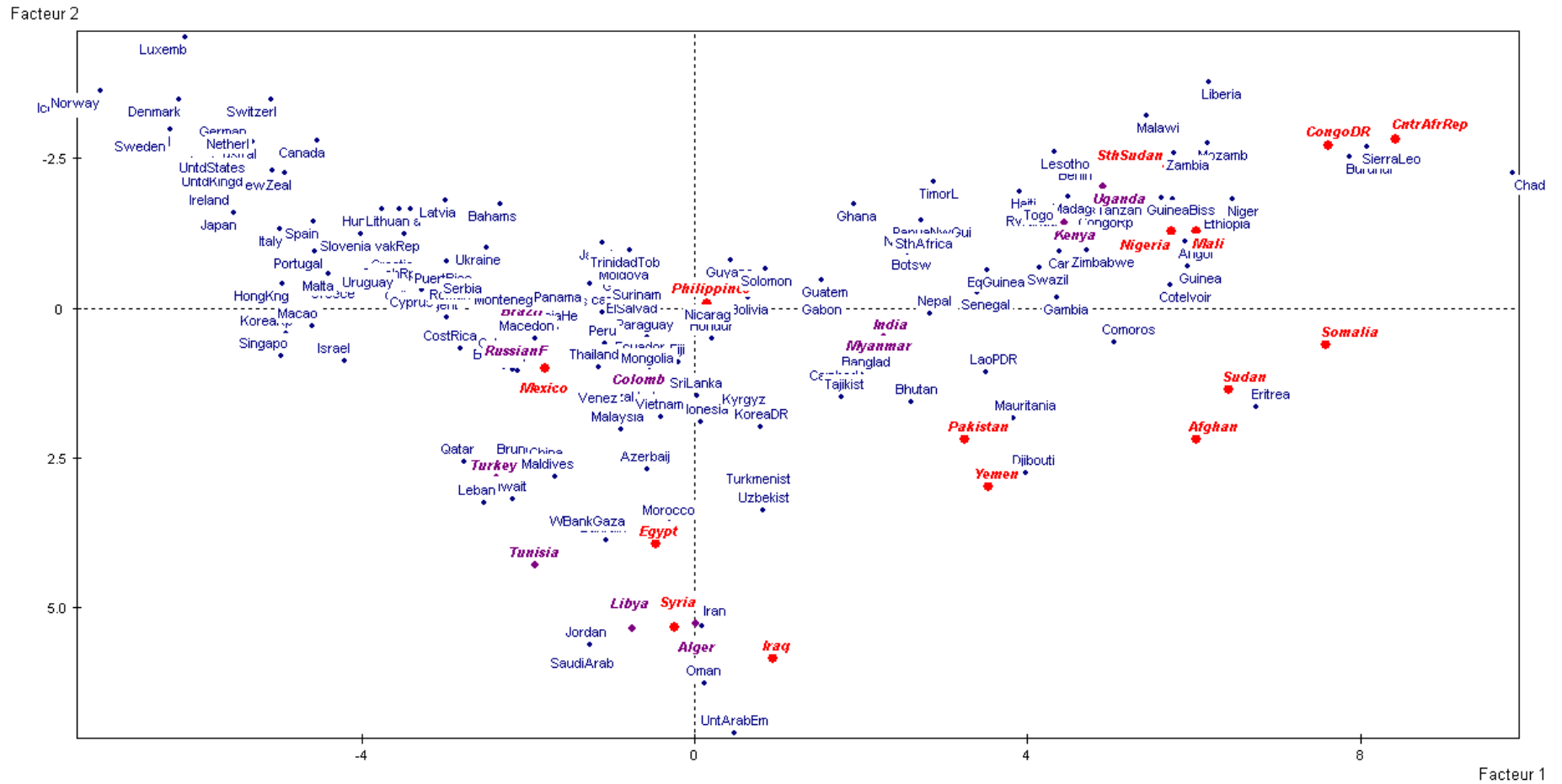


Figura 4.14 Mappa fattoriale (fattore1/fattore2)

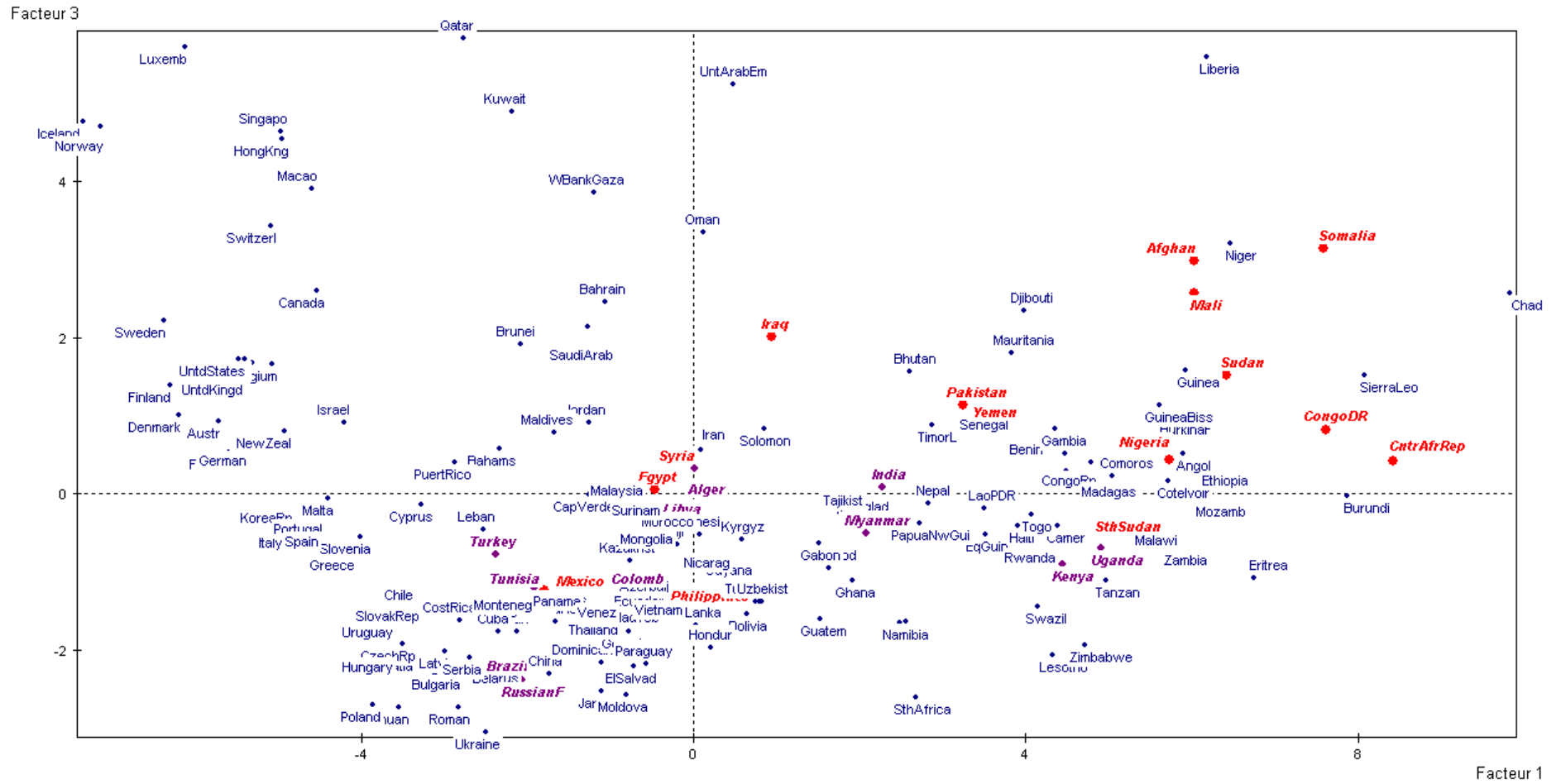


Figura 4.15 Mappa fattoriale (fattore1/fattore3)

Conclusioni e raccomandazioni finali

In questo paragrafo conclusivo illustriamo in breve i punti su cui crediamo debba essere effettuato uno sforzo nella successiva fase di implementazione e sperimentazione.

Ricordiamo che il quesito iniziale era stabilire se esistono relazioni nei dati, rilevati a livello internazionale, che possano essere utilmente sfruttate per fornire un supporto informativo al processo decisionale interno.

Nel corso dei colloqui iniziali e nell'outline del presente progetto di fattibilità approvato da STC, è stato evidenziato come il sistema informativo che si ritiene possa essere implementato necessita di più elementi:

- a) le informazioni provenienti da dati secondari (che è quanto preso in esame in questo report);
- b) la verifica e riorganizzazione delle basi dati interne per finalità di analisi. Questo è un percorso che STC dovrebbe intraprendere al suo interno, tenuto conto della propria struttura organizzativa e delle disomogeneità in termini di standard di dati raccolti e di storage che certamente saranno presenti nelle varie sedi nazionali;
- c) l'organizzazione di un sistema di panels di esperti (dove per esperti si intendo persone le quali lavorano nella propria organizzazione o in organizzazioni similari e che per posizione e tipo di attività sono più a contatto con le realtà nazionali) attraverso le proprie sedi nel mondo.

STC dovrà scegliere la successiva strada da intraprendere e quindi la scala di grandezza che vuole attribuire a questo progetto. Il prosieguo può essere, infatti, focalizzato all'interno dell'organizzazione nello spirito di migliorare i propri flussi informativi interni in rapporto alla elaborazione preventiva di warnings, oppure può costituire uno strumento di collegamento verso altre organizzazioni che, per know-how o per finalità istituzionali, condividono le medesime aspirazioni verso la costruzione di un early warning system autonomo.

Noi, qui ci limiteremo a considerare, la prima ipotesi e, pertanto, daremo le indicazioni strettamente essenziali.

Ritornando quindi all'obiettivo di questo studio preliminare, sulla base di quanto descritto nei precedenti Capitoli, si è dimostrato che utilizzando solo fonti di dati secondari è possibile individuare delle strategie di analisi che forniscono dei quadri di sintesi utili per il posizionamento strategico delle azioni di STC e per la individuazione dei Paesi che potrebbero necessitare di interventi.

Nella successiva fase di sperimentazione, analisi e implementazione, è però necessario procedere

1. all'aggiornamento ed automatizzazione della base dati;
2. all'affinamento e generalizzazione delle tecniche;

3. alla implementazione di una procedura informatica che consenta: l'acquisizione ed il trattamento automatico preliminare dei dati e la necessaria stima dei dati mancanti; l'utilizzo flessibile dei dati presenti nel database e la realizzazione di analisi (anche mediante opportuni strumenti grafici).

L'aggiornamento del database necessita, come già accennato nel corso del report, di collaborazioni specifiche con il CRED (Centre for Research on the Epidemiology of Disasters) e GDELT per verificare l'esistenza all'interno dei loro database di ulteriori informazioni (attualmente non visibili da web) e se i risultati delle sperimentazioni condotte al loro interno possano essere integrate nel presente progetto. Per quanto concerne i dati del CRED sarebbe inoltre auspicabile riuscire a migliorare la stima della popolazione colpita da un disastro naturale in funzione delle coordinate geo-spaziali.

Occorre verificare, inoltre, attraverso quali canali sia possibile automatizzare l'aggiornamento della base dati (sia per quanto concerne l'insieme delle variabili considerate in un singolo anno di riferimento che delle serie storiche utilizzate nel corso del report).

LA disponibilità, a regime, di un database così come descritto nel report, potrà fornire a STC un vantaggio ulteriore, poiché consentirà nel tempo di costruire un patrimonio dati omogeneo che potrà essere utile anche per altre tipologie di analisi.

Per quanto riguarda l'affinamento e generalizzazioni delle tecniche investigate, riteniamo che altre collaborazioni siano auspicabili in relazione a due punti specifici:

- a) per quanto concerne i disastri naturali, è necessario individuare fonti e dati concernenti le previsioni climatiche a breve termine. Tali informazioni potrebbero aiutare a migliorare la stima della probabilità che si verifichino in determinati Paesi calamità meteo-dipendenti. In tale direzione, laddove si individuasse una collaborazione utile occorrerà verificare la forma da adottare nella modellistica affinché questo meccanismo di updating possa essere incorporato.
- b) per quanto riguarda, invece, i conflitti e le instabilità, è necessario promuovere una collaborazione con analisti politici, per individuare le keywords da monitorare per identificare i punti di rottura ed i periodi da utilizzare come riferimento per le carte di controllo.

La strategia individuata nel corso del nostro lavoro necessita di una ampia fase di sperimentazione che produrrà certamente ulteriori affinamenti delle tecniche illustrate. Soprattutto in relazione alle analisi dirette alla identificazione delle condizioni che portano a conflitti ed instabilità, l'ampliamento dello studio ad altre situazioni potrebbe generare la necessità di ulteriori strumenti.

Per quanto concerne, infine, l'implementazione del software specifico, è bene evidenziare che tutte le analisi statistiche illustrate sono riproducibili in R (che è un ambiente di programmazione statistico free-share) ma anche in molti packages specialistici. Pertanto, è stata assicurata una sufficiente libertà per

le successive implementazioni. Riteniamo, però, importante che nella fase di disegno del software sia prevista, nel team a ciò dedicato, la presenza di uno statistico al fine di dirigere correttamente le fasi di calcolo nonché l'implementazione di strategie flessibili di trattamento dei dati e di produzione delle rappresentazioni grafiche (anche dinamiche).

In conclusione, poi, va segnalato che un sistema di analisi così complesso non può essere totalmente gestito da un sistema automatizzato, ma richiede la presenza di un team qualificato che possa intervenire nelle varie fasi. Il processo alla base delle analisi può essere descritto mediante un protocollo di azioni che potrà ricalcare quanto effettuato nel presente report e quanto emergerà dalla fase di sperimentazione.

L'investimento che STC dovrebbe affrontare in termini di risorse umane, affinché il sistema informativo possa essere efficacemente sfruttato da STC per le proprie finalità interne, richiede la costituzione di una unità di analisi, ove sia presente almeno uno statistico metodologo, con particolari conoscenze nel campo delle analisi multivariate e della modellistica per dati ordinali, di un esperto capace di implementare software ottimizzato in R (se STC opta per questa soluzione), di un esperto nella creazione/gestione dell'interfaccia per l'utente. A tale gruppo andranno poi affiancati alcuni analisti qualitativi e di scenario.

Bibliografia

- Abdi, H., Williams, L.J., Valentin, D. (2013). Multiple factor analysis: principal component analysis for multitable and multiblock data sets. *WIREs Computational Statistics*, doi: 10.1002/wics.1246.
- Agresti A. (2002). *Categorical data analysis* J. Wiley, Hoboken N.J.
- Alesina A., Devleeschauwer A., Easterly W., Kurlat S., Wacziarg R. (2003). Fractionalization. *Journal of Economic growth*, 8(2), 155-194.
- Barro R. J., Lee J. W. (2012). A new dataset of educational attainment in the world, 1950–2010. *Journal of Development Economics*, 184-198.
- Basher R. (2006). Global early warning systems for natural hazards: systematic and people-centred. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 364(1845), 2167-2182.
- Birkmann J. (2007). Risk and vulnerability indicators at different scales: Applicability, usefulness and policy implications. *Environmental Hazards*, 7, 20–31.
- Bui T., Cho S., Sankaran S., Sovereign M. (2000). A framework for designing a global information network for multinational humanitarian assistance/disaster relief. *Information Systems Frontiers*, 1, 427-442.
- Carment D. (2003). Assessing state failure: implications for theory and policy. *Third World Quarterly*, 24(3), 407-427.
- Carment D., Prest S., & Samy Y. (2008). Determinants of state fragility and implications for aid allocation. *An Assessment Based on the Country Indicators for Foreign Policy Project*. *UNU World Institute for Development Economics Research*.
- Carment D., Samy Y., & Landry J. (2013). Transitioning Fragile States: A Sequencing Approach. *Fletcher Forum of World Affairs*, 125 151.
- Cardona, O.D. (2006). A System of Indicators for Disaster Risk Management in the Americas. In *Measuring Vulnerability to Hazards of Natural Origin: Towards Disaster Resilient Societies*, (ed. J. Birkmann), United Nations University Press, Tokyo.
- Cutter S. L. (1996). Vulnerability to environmental hazards. *Progress in Human Geography*, 20, 529-539.
- Dilley M. (2005). Natural disaster hotspots: a global risk analysis (Vol. 5). World Bank Publications.
- Escofier, B, Pages, J. (1988) *Analyses Factorielles Simples et Multiples: Objectifs, Methodes, Interpretation*, Dunod, Paris.
- Escofier, B, Pages, J. (1994). Multiple factor analysis (AFMULT package). *Computational Statistics and Data Analysis*, 18, 121--140.
- Esty D. C., Goldstone J. A., Gurr T. R., Surko P. T., Unger A. N. (1995). Working papers: State failure task force report. *McLean, VA: Science Applications International Corporation*, 30.
- Fearon J.D., Laitin D. (2003). Ethnicity, Insurgency, and Civil War. *American Political Science Review*, 97, 75-90.
- Goldstone J.A., Bates R.H., Gurr T.R, Lustik M., Marshall M.G., Ulfelder J., Woodward M. (2005). A Global Forecasting Model of Political Instability. *Annual Meeting of the American Political Science Association*, Washington, DC.
- Guha-Sapir D., Vos F., Below R., Ponserre S. (2011). Annual disaster statistical review 2010. *Centre for Research on the Epidemiology of Disasters*.
- Gur T. R., Harff B., Levy M., Dabelko G. D., Surko P. T., Unger A. N. (1999). State Failure Task Force Report: Phase II Findings. *Environmental Change Security Project Report*, 5, 50.

- Gurr and Marshall M.G. (2000). Assessing Risks of Future Ethnic Wars. in Gurr, ed., *Peoples versus States: Minorities at Risk in the New Century*. United States Institute of Peace Press, Washington, DC.
- Gurr T. R., Woodward M. R., Marshall M. G. (2005). *Forecasting Instability: Are Ethnic Wars and Muslim Countries Different?*. Political Instability Task Force.
- Hughes B.B., Moyer J.D., Sisk T.D.(2011). *Vulnerability to intrastate conflict*. US Institute of Peace, publication n.72.
- Legros G., Havet I., Bruce N., Bonjour S.(2009). *The energy access situation in developing countries*. UNDP and World Health Organization.
- Hales D., Miller J. (2010). *Triggers of State Failure*. (No. DRDC-Toronto-CR-2008-054). Cae Professional Services, Ottawa (CA).
- Harff B., Gurr T. R. (1998). Systematic early warning of humanitarian emergencies. *Journal of Peace Research*, 35(5), 551-579.
- Heideberg Institute for International Conflict Research (2012). *Conflict barometer N. 21*, Department of political science, University of Heidelberg.
- Hufschmidt G. (2011). A comparative analysis of several vulnerability concepts. *Nat Hazards*, 58, 621–643.
- Hughes B.B., Moyer J.D., Sisk T.D. (2011). *Vulnerability to intrastate conflict: Evaluating Quantitative Measures*. United States Institute of Peace
- Human Development Report.(2011). *Technical notes. Calculating the Human Development Index*.
- ILO (2012). *Estimates of the Economically Active Population: 1990-2010 (sixth edition)*.
- Kahn M. E. (2005). The death toll from natural disasters: the role of income, geography, and institutions. *Review of Economics and Statistics*, 87(2), 271-284.
- Kaiser R., Spiegel P. B., Henderson A. K., Gerber M. L. (2003). The application of geographic information systems and global positioning systems in humanitarian emergencies: lessons learned, programme implications and future research. *Disasters*, 27, 127-140.
- King G., Zeng L. (2001). Improving forecasts of state failure. *World Politics*, 53, 623-658.
- Kolde, R. Laur, S., Adler, P. Vilo J. (2012). Robust rank aggregation for gene list integration and meta-analysis, *Bioinformatics*, 28, 573-580.
- Lal P.N., Singh R., Holland P. (2009). *Relationship between natural disasters and poverty: a Fiji case study*. ISDR.
- Leetaru K., Schrodt P. A. (2013, April). GDELT: Global data on events, location, and tone, 1979–2012. In *ISA Annual Convention*, vol.2, 4.
- Marshall M.G. (1999). *Third World War. System, Process, and Conflict Dynamics*. Rowman & Littlefield. Publishers, Inc.
- Marshall M. G. (2008). Fragility, instability, and the failure of states. *Council on Foreign Relations*.
- Mata J.F., Ziaja S. (2009). *Users' Guide on Measuring Fragility*. German Development Institute / Deutsches Institut für Entwicklungspolitik (DIE) and the United Nations Development Programme (UNDP).
- Mosquera-Machado S., Dillely M. (2009). A comparison of selected global disaster risk assessment results. *Natural hazards*, 48, 439-456.
- O'Brien S.P. (2002). Anticipating the Good, the Bad, and the Ugly An Early Warning Approach to Conflict and Instability Analysis. *Journal of Conflict Resolution*, 46, 791-811.
- O'Brien S.P. (2003). Analyzing Complex Threats for Operations and Readiness (ACTOR). Paper presented at the *Annual meeting of the Military Operations Research Society Symposium*.

- O'Brien S.P. (2010). Crisis early warning and decision support: Contemporary approaches and thoughts on future research. *International Studies Review*, 12, 87-104.
- Peduzzi P., Herold H. D. C. (2005). Mapping disastrous natural hazards using global datasets. *Natural Hazards*, 35, 265-289.
- Peduzzi P., Dao H., Herold, C., Mouton F. (2009). Assessing global exposure and vulnerability towards natural hazards: the Disaster Risk Index. *Natural Hazards and Earth System Science*, 9, 1149-1159.
- Peduzzi P. (2012). *Risk and Global Change: Global Perspective on Disaster Risk Reduction*. ICG Symposium–Oslo, UNEP/GRID-Geneva.
- Peduzzi P., Chatenoux B., Dao H., De Bono A., Herold, C., Kossin J., Nordbeck O. (2012). Global trends in tropical cyclone risk. *Nature Climate Change*, 2, 289-294.
- Pew Research Center (2012). The Global Religious Landscape. A Report on the Size and the Distribution of the World's Major Religious Groups as of 2010. *Global religious Futures Project*.
- Rice S.E., Patrick S. (2008). *Index of State Weakness in the Developing World*. <http://www.brookings.edu/research/reports/2008/02/weak-states-index>.
- Rubin, D. (1976). Inference and missing data. *Biometrika*, 63, 581–592.
- Save the Children International (2012). *Categorisation Framework for Humanitarian Crises and Response*. Document issued on 18 July 2012.
- Sambanis N. (2001). Do Ethnic and Non-ethnic Civil Wars Have the Same Causes?: A Theoretical and Empirical Inquiry (Part 1). *The Journal of Conflict Resolution*, 45, 259-282.
- Schmeidl S., Jenkins J. C. (1998). The early warning of humanitarian disasters: Problems in building an early warning system. *International Migration Review*, 32, 471-486.
- Spencer L J., Gubbins P., Murray C.J.L., Gakidou E. (2012). Developing a comprehensive time series of GDP per capita for 210 countries from 1950 to 2015. *Population Health Metrics*, 10-12.
- Southern Sudan Centre for, Statistics and Evaluation Census (2010). *Statistical Yearbook for Southern Sudan 2010*.
- Strömberg D. (2007). Natural disasters, economic development, and humanitarian aid. *The Journal of Economic Perspectives*, 21, 199-222.
- Taras R., Ganguly R. (2010). *Understanding ethnic conflict*. Longman (4th edition).
- Tellis A. J., Szayna T. S., Winnefeld J. A. (1997). *Anticipating Ethnic Conflict* (No. RAND-MR-853-A). Rand Arroyo Center Santa Monica (US).
- Tschoegl L., Below R., Guha-Sapir R. (2006). *An Analytical Review of Selected Datasets on Natural Disasters and Impacts*. UNDP/CRED Workshop on Improving Compilation of Reliable Data on Disaster Occurrence and Impact, Bangkok.
- Thywissen K. (2004). Core Terminology of Disaster Reduction. (United Nations University – Institute for Environment and Human Security (UNU-EHS).
- United Nations (2013). *A new global partnership: eradicate poverty and transform economies through sustainable development*. Report of the high-level panel of eminent persons on the post-2015 development agenda. United Nations Publications.
- UNDRO (1979). Natural Disasters and Vulnerability Analysis. in Report of Expert Group Meeting. Risk evaluation (9-12 July, 1979).
- UNICEF (2012). *The State of The World Children 2012*. United Nations.
- UNISDR (2009). *Terminology on disaster risk reduction*. United Nations.
- UNISDR (2013). *Global Assessment Report on Disaster Risk Reduction* (Annex I).
- UNISDR (2013). *Réduction du risque de catastrophe: Bilan mondial*. United Nations.

- van Bergeijk P. A.G., Lazzaroni S. (2013). Macroeconomics of Natural Disasters: Meta-Analysis and Policy Options. Working Paper, Harvard Center for Risk Analysis
- Ward M., Beger A., Cutler J., Matthew D., et al. (2013) *Comparing GDELT and ICEWS Event Data*. Duke University, <http://mdwardlab.com/biblio/comparing-gdelt-and-icews-event-data>.
- Wisner B., Blaikie P, Cannon T., Davis I. (2004). *At Risk: Natural Hazards, People's Vulnerability and Disasters*. Routledge, London.
- Yonamine J. (2011). Working with Event Data: A Guide to Aggregation Choices. *Technical Report of the Department of Political Science Pennsylvania State University*.

Appendice - Descrizione dei dati

A.1 Lista delle variabili (in ordine alfabetico)

	VARIABLE NAME	SOURCE	DESCRIPTION	FURTHER SPECIFICATION
E	Access to electricity (% of population)	International Energy Agency, World Energy Outlook 2010.	Access to electricity is the percentage of population with access to electricity. Electrification data are collected from industry, national surveys and international sources. Data on access to electricity are collected by the IEA from industry, national surveys, and international sources. To estimate the number of people without access, population data comes from OECD statistics in conjunction with the United Nations Population Division reports World Urbanization Prospects: the 2009 Revision Population Database, and World Population Prospects: the 2010 Revision. Electricity access data is adjusted to be consistent with demographic patterns of urban and rural population.	Due to differences in definitions and methodology from different sources, data quality may vary from country to country. Where country data appeared contradictory, outdated or unreliable, the IEA Secretariat made estimates based on cross-country comparisons and earlier surveys.
I	Adult (15+) literacy rate (%). Female	UNESCO	Adult literacy rate is the percentage of female population ages 15 and above who can, with understanding, read and write a short, simple statement on their everyday life.	
I	Adult illiterate population. % female >15	UNESCO	Adult literacy rate is the percentage of female population ages 15 and above who can not, with understanding, read and write a short, simple statement on their everyday life.	
P	Age-specific fertility rates 15-19 (births per 1,000 women)	UN Population Division	Number of births to women in a particular age group, divided by the number of women in that age group. The age groups used are: 15-19, 20-24,45-49. The data refer to five-year periods running from 1 July to 30 June of the initial and final years.	
P	Age-specific fertility rates 20-24 (births per 1,000 women)	UN Population Division	Number of births to women in a particular age group, divided by the number of women in that age group. The age groups used are: 15-19, 20-24,45-49. The data refer to five-year periods running from 1 July to 30 June of the initial and final years.	
P	Age-specific fertility rates 25-29 (births per 1,000 women)	UN Population Division	Number of births to women in a particular age group, divided by the number of women in that age group. The age groups used are: 15-19, 20-24,45-49. The data refer to five-year periods running from 1 July to 30 June of the initial and final years.	
P	Age-specific fertility rates 30-34 (births per 1,000 women)	UN Population Division	Number of births to women in a particular age group, divided by the number of women in that age group. The age groups used are: 15-19, 20-24,45-49. The data refer to five-year periods running from 1 July to 30 June of the initial and final years.	
E	Agricultural land (% of land area)	Food and Agriculture Organization	Agricultural land constitutes only a part of any country's total area, which can include areas not suitable for agriculture, such as forests, mountains, and inland water bodies. Three components of the agricultural land are: a) arable land - land defined by the FAO as land under temporary crops (double-cropped areas are counted once), temporary meadows for mowing or for pasture, land under market or kitchen gardens, and land temporarily fallow, b) permanent pasture - land used for five or more years for forage, including natural and cultivated crops, and c) and under permanent crops - land cultivated with crops that occupy the land for long periods and need not be replanted after each harvest, such as cocoa, coffee, and rubber; land under flowering shrubs, fruit trees, nut trees, and vines is included, but land under trees grown for wood or timber is not. Agricultural land is also sometimes classified as irrigated and non-irrigated land. In arid and semi-arid countries agriculture is often confined to irrigated land, with very little farming possible in non-irrigated areas.	

E	Arable land (% of land area)	Food and Agriculture Organization	<p>Arable land (in hectares) includes land defined by the FAO as land under temporary crops (double-cropped areas are counted once), temporary meadows for mowing or for pasture, land under market or kitchen gardens, and land temporarily fallow</p> <p>Temporary fallow land refers to land left fallow for less than five years. The abandoned land resulting from shifting cultivation is not included in this category.</p> <p>Data for "Arable land" are not meant to indicate the amount of land that is potentially cultivable.</p>	
E	Arable land (hectares per person)	Food and Agriculture Organization	<p>Arable land (in hectares) includes land defined by the FAO as land under temporary crops (double-cropped areas are counted once), temporary meadows for mowing or for pasture, land under market or kitchen gardens, and land temporarily fallow</p> <p>Temporary fallow land refers to land left fallow for less than five years. The abandoned land resulting from shifting cultivation is not included in this category.</p> <p>Data for "Arable land" are not meant to indicate the amount of land that is potentially cultivable.</p>	
S	Armed forces personnel (% of total labor force)	International Institute for Strategic Studies, The Military Balance.	<p>Military data on manpower represent quantitative assessment of the personnel strengths of the world's armed forces. The numbers are based on the most accurate data available to, or, on the best estimate that can be made by the International Institute for Strategic Studies (IISS) at the time of its annual publication. The IISS collects the data from national governments.</p> <p>Armed forces personnel comprise all servicemen and women on full-time duty (including conscripts and long-term assignments from the Reserves). Reserve describes formations and units not fully manned or operational in peacetime, but which can be mobilized by recalling reservists in an emergency. IISS estimates of effective reservist strengths on the numbers available within five years of completing full-time service, unless there is good evidence that obligations are enforced for longer. Although paramilitary forces whose training, organization, equipment and control suggest they may be used to support or replace regular military forces, they are not included in the armed forces personnel. Home Guard units are counted as paramilitary.</p> <p>Data are shown as percentage of total labor force. According to International Labour Organization (ILO) armed forces occupations include all jobs held by members of the armed forces. Members of the armed forces are those personnel who are currently serving in the armed forces, including auxiliary services, whether on a voluntary or compulsory basis, and who are not free to accept civilian employment and are subject to military discipline. Included are regular members of the army, navy, air force and other military services, as well as conscripts enrolled for military training or other service for a specified period. Excluded are persons in civilian employment of government establishments concerned with defense issues; police (other than military police); customs inspectors and members of border or other armed civilian services; persons who have been temporarily withdrawn from civilian life for a short period of military training or retraining, according to national requirements, and members of military reserves not currently on active service.</p>	
S	Arms imports (SIPRI trend indicator\$)	Stockholm International Peace Research Institute (SIPRI).	<p>The TIV is based on the known unit production costs of a core set of weapons and is intended to represent the transfer of military resources rather than the financial value of the transfer.</p> <p>SIPRI calculates the volume of transfers to, from and between all parties using the TIV and the number of weapon systems or subsystems delivered in a given year. This data is intended to provide a common unit to allow the measurement of trends in the flow of arms to particular countries and regions over time. Therefore, the main priority is to ensure that the TIV system remains consistent over time, and that any changes introduced are backdated.</p> <p>SIPRI TIV figures do not represent sales prices for arms transfers. They should therefore not be directly compared with gross domestic product (GDP), military expenditure, sales values or the financial value of export licences in an attempt to measure the economic burden of arms imports or the economic benefits of exports. They are best used as the raw data for calculating trends in international arms transfers over periods of time, global percentages for suppliers and recipients, and percentages for the volume of transfers to or from particular states.</p>	<p>Stockholm International Peace Research Institute (SIPRI)'s Arms Transfers Program collects data on arms transfers from open sources. Since publicly available information is inadequate for tracking all weapons and other military equipment, SIPRI covers only what it terms major conventional weapons.</p>

P	Average annual rate of population change (percentage)	UN Population Division	Population increment over a period, that is, the difference between the population at the end of the period and that at the beginning of the period. Refers to five-year periods running from 1 July to 30 June of the initial and final years. Data are presented in percentage	
E	Cereal yield (kg per hectare)	Food and Agriculture Organization	A cereal is a grass cultivated for the edible components of their grain, composed of the endosperm, germ, and bran. Cereal yield is measured as kilograms per hectare of harvested land. Cereal grains are grown in greater quantities and provide more food energy worldwide than any other type of crop; cereal crops therefore can also be called staple crops Cereals production includes wheat, rice, maize, barley, oats, rye, millet, sorghum, buckwheat, and mixed grains. Production data on cereals relate to crops harvested for dry grain only. Cereal crops harvested for hay or harvested green for food, feed, or silage and those used for grazing are excluded.	The FAO allocates production data to the calendar year in which the bulk of the harvest took place. Most of a crop harvested near the end of a year will be used in the following year.
P	Child dependency ratio (ratio of population aged 0-14 per 100 population 15-64)	UN Population Division	Child Dependency Ratio (Age 0-14 / Age 15-64) De facto population as of 1 July of the year indicated.	
S	Civil Liberties	Freedom House	The Freedom in the World survey provides an annual evaluation of the progress and decline of freedom in 195 countries and 14 related and disputed territories. The survey, which includes both analytical reports and numerical ratings, measures freedom according to two broad categories: political rights and civil liberties. Political rights ratings are based on an evaluation of three subcategories: electoral process, political pluralism and participation, and functioning of government. Civil liberties ratings are based on an evaluation of four subcategories: freedom of expression and belief, associational and organizational rights, rule of law, and personal autonomy and individual rights. Each country is assigned a numerical rating from 1 to 7 for both political rights and civil liberties, with 1 representing the most free and 7 the least free. The ratings are determined by the total number of points (up to 100) each country receives on 10 political rights questions and 15 civil liberties questions; countries receive 0 to 4 points on each question, with 0 representing the smallest degree and 4 the greatest degree of freedom. The average of the political rights and civil liberties ratings, known as the freedom rating, determines the overall status: Free (1.0 to 2.5), Partly Free (3.0 to 5.0), or Not Free (5.5 to 7.0). Freedom House also assigns upward or downward trend arrows to countries which saw general positive or negative trends during the year that were not significant enough to result in a ratings change.	The survey assigns the designation of electoral democracy to countries that have met certain minimum standards. The numerical benchmark for a country to be listed as an electoral democracy is a total of 7 points or more (out of a possible 12) for the 3 political rights subcategory questions on electoral process, as well as a total of 20 points or more (out of a possible 40) for all 10 political rights questions. Freedom House does not maintain a culture-bound view of freedom. The methodology of the survey is grounded in basic standards of political rights and civil liberties, derived in large measure from relevant portions of the Universal Declaration of Human Rights. These standards apply to all countries and territories, irrespective of geographical location, ethnic or religious composition, or level of economic development.
E	CO2 emissions (kt)	World Bank	Carbon dioxide emissions are those stemming from the burning of fossil fuels and the manufacture of cement. They include carbon dioxide produced during consumption of solid, liquid, and gas fuels and gas flaring. Source: Carbon Dioxide Information Analysis Center, Environmental Sciences Division, Oak Ridge National Laboratory, Tennessee, United State	The 1950 to present CO ₂ emission estimates are derived primarily from energy statistics published by the United Nations (2013), using the methods of Marland and Rotty (1984).
	Colonial Origin		This variable defines the colonial origin of the countries as: None=0; Dutch=1; Spain=2; South Africa=3; Denmark=4; UK=5; France=6; Portugal=7; Italy=8; Australia=9; USA=11; Belgium=12; Japan=14; exURSS=15	
S	Corruption Index (100=no corruption)	Trasparenzy International	Each source is standardised to be compatible with other available sources, for aggregation to the CPI scale. The standardisation converts all the data sources to a scale of 0-100 where 0 = highest level of perceived corruption, and 100 = lowest level of perceived corruption.	(Trasparenzy International (2012) <i>Corruption Perceptions Index 2012</i> . Technical Methodology Note.)

E	Crop production index (2004-2006 = 100)	Food and Agriculture Organization	<p>FAO indices of agricultural production show the relative level of the aggregate volume of agricultural production for each year in comparison with the base period 2004-2006. They are based on the sum of price-weighted quantities of different agricultural commodities produced after deductions of quantities used as seed and feed weighted in a similar manner. The resulting aggregate represents, therefore, disposable production for any use except as seed and feed.</p> <p>All the indices at the country, regional and world levels are calculated by the Laspeyres formula.</p> <p>Production quantities of each commodity are weighted by 2004-2006 average international commodity prices and summed for each year. To obtain the index, the aggregate for a given year is divided by the average aggregate for the base period 2004-2006.</p> $CPI = \frac{\text{CropProduced}(\text{Year}) * \text{meanPriceCrop}(2004-2006)}{\text{meancrop_produced}(2004-2006) * \text{meanPriceCrop}(2004-2006)}$	Crop production index shows agricultural production for each year relative to the base period 2004-2006. It includes all crops except fodder crops. Regional and income group aggregates for the FAO's production indexes are calculated from the underlying values in international dollars, normalized to the base period 2004-2006.
H	Deaths due to HIV/AIDS (per 100000 population)	World Health Organization UNAIDS	<p>Empirical data from different HIV surveillance sources are consolidated to obtain estimates of the level and trend on HIV infection and of mortality in adults and children. Standard methods and tools for HIV estimates that are appropriate to the pattern of the HIV epidemic are used. However, to obtain the best possible estimates, judgement needs to be used as to the quality of the data and how representative it is of the population. UNAIDS and WHO produce country-specific estimates of mortality due to HIV/AIDS every two years</p> <p>To calculate mortality rates, the total population are derived from the World Population Prospects: The 2006 Revision (UN Population Division, 2007).</p>	The HIV/AIDS mortality rates of adults and of children aged less than 15 years are leading indicators of the level of impact of the HIV/AIDS epidemic and of the impact of interventions, particularly the scaling-up of treatment and prevention of mother-to-child transmission in countries with generalized HIV epidemics.
H	Deaths due to malaria (per100000 population)	World Health Organization	<p>The malaria death rate is expressed as the number of deaths due to malaria per 100 000 population per year with the population of a country derived from projections made by the UN Population Division.</p> <p>The number of malaria deaths is derived by one of two methods:</p> <p>(i) by multiplying the estimated number of P. falciparum malaria cases in a country by a fixed case-fatality rate. This method is used for all countries outside the African Region and for countries in the African Region where estimates of case incidence are derived from routine reporting systems and where malaria comprises less than 5% of all deaths in children under 5 as described in the Global Burden of Disease Incremental Revision for 2004 (GBD2004).</p> <p>(ii) A case fatality rate of 0.45% is applied to the estimated number of P.falciparum cases for countries in the African Region and a case fatality rate of 0.3% for P. falciparum cases in other regions. (In situations where the fraction of all deaths due to malaria deaths is small, the use of a case fatality rate in conjunction with estimates of case incidence is considered to provide a better guide to the levels of malaria mortality than attempts to estimate the fraction of deaths due to malaria.)</p> <p>(iii) For countries in the African Region where malaria comprises 5% or more of all deaths in children under 5, the number of deaths are derived from an estimate of the number of people living at high, low or no risk of malaria. Malaria deaths rates for these populations are inferred from longitudinal studies of malaria deaths as recorded in the published literature.</p>	
H	Depth of the food deficit (kilocalories per person per day)	Food and Agriculture Organization	<p>The depth of the food deficit indicates how many calories would be needed to lift the undernourished from their status, everything else being constant. The average intensity of food deprivation of the undernourished, estimated as the difference between the average dietary energy requirement and the average dietary energy consumption of the undernourished population (food-deprived), is multiplied by the number of undernourished to provide an estimate of the total food deficit in the country, which is then normalized by the total population.</p> $FD = (\text{Kcal}(\text{require}) - \text{kcal}(\text{consumption})) * \text{pop_undernourished} / \text{population_total}$ <p>(The State of Food Insecurity in the World)</p>	

E	Diamonds (gems+ind) >0,2% world production	British Geological survey, National Environment Reserach Council	Total diamond produced for gems and industries (carates) divided by total carates produced in the world (%), considering only a percentage greather than 0,2% Dummy variable=1 producer country with aproduction >0.2% world production; 0 otherwise	BritMinSurv
I	Educ attain % completed Secondary-FEMALE	Barro/WB	Barro, R. J., & Lee, J. W. (2012). A new dataset of educational attainment in the world, 1950–2010. <i>Journal of Development Economics</i> . The Barro-Lee Dataset provides educational attainment data for 146 countries in 5-year intervals from 1950 to 2010. It also provides information about the distribution of educational attainment of the adult population over age 15 and over age 25 by sex at seven levels of schooling— no formal education, incomplete primary, complete primary, lower secondary, upper secondary, incomplete tertiary, and complete tertiary. Average years of schooling at all levels—primary, secondary, and tertiary—are also measured for each country and for regions in the world. To reduce measurement error, the new estimates are constructed using recently available census/survey observations from consistent census data, disaggregated by age group, and new estimates of mortality rate and completion rate by age and by education.	The dataset is provided by the World Bank.
I	Educ attain % completed Secondary-TOTAL	Barro/WB	Barro, R. J., & Lee, J. W. (2012). A new dataset of educational attainment in the world, 1950–2010. <i>Journal of Development Economics</i> .	The dataset is provided by the World Bank.
I	Educ attain % completed Tertiary- TOTAL	Barro/WB	Barro, R. J., & Lee, J. W. (2012). A new dataset of educational attainment in the world, 1950–2010. <i>Journal of Development Economics</i> .	The dataset is provided by the World Bank.
I	Educ attain % completed Tertiary-FEMALE	Barro/WB	Barro, R. J., & Lee, J. W. (2012). A new dataset of educational attainment in the world, 1950–2010. <i>Journal of Development Economics</i> .	The dataset is provided by the World Bank.
I	Educ attain % completed Primary - TOTAL	Barro/WB	Barro, R. J., & Lee, J. W. (2012). A new dataset of educational attainment in the world, 1950–2010. <i>Journal of Development Economics</i> .	The dataset is provided by the World Bank.
I	Educ attain % completed Primary -FEMALE	Barro/WB	Barro, R. J., & Lee, J. W. (2012). A new dataset of educational attainment in the world, 1950–2010. <i>Journal of Development Economics</i> .	The dataset is provided by the World Bank.
S	Election Rights – Women admission	Inter-Parliamentary Union	Computed as the difference between the 2012 and the year of obtaining universal suffrage.	
E	Electric power consumption (kWh per capita)	International Energy Agency (IEA Statistics OECD/IE)	Electric power consumption per capita (kWh) is the production of power plants and combined heat and power plants less transmission, distribution, and transformation losses and own use by heat and power plants, divided by midyear population . Energy data are compiled by the International Energy Agency (IEA). IEA data for economies that are not members of the Organisation for Economic Co-operation and Development (OECD) are based on national energy data adjusted to conform to annual questionnaires completed by OECD member governments. Electricity consumption is equivalent to production less power plants' own use and transmission, distribution, and transformation losses less exports plus imports. It includes consumption by auxiliary stations, losses in transformers that are considered integral parts of those stations, and electricity produced by pumping installations. Where data are available, it covers electricity generated by primary sources of energy - coal, oil, gas, nuclear, hydro, geothermal, wind, tide and wave, and combustible renewables. Neither production nor consumption data capture the reliability of supplies, including breakdowns, load factors, and frequency of outages.	
S	Ethnic Fractionalization Index		See: Alesina, A., Devleeschauwer, A., Easterly, W., Kurlat, S., & Wacziarg, R. (2003). Fractionalization. <i>Journal of Economic growth</i> , 8(2), 155-194.	

E	Exports of goods and services (% of GDP)	World Bank	Exports of goods and services represent the value of all goods and other market services provided to the rest of the world. They include the value of merchandise, freight, insurance, transport, travel, royalties, license fees, and other services, such as communication, construction, financial, information, business, personal, and government services. They exclude compensation of employees and investment income (formerly called factor services) and transfer payments. (National accounts data, and OECD National Accounts data files).	
E	External debt stocks (% of exports of goods, services and primary income)	World Bank.	Total external debt is debt owed to nonresidents repayable in foreign currency, goods, or services. Total external debt is the sum of public, publicly guaranteed, and private nonguaranteed long-term debt, use of IMF credit, and short-term debt. Short-term debt includes all debt having an original maturity of one year or less and interest in arrears on long-term debt. Exports of goods, services and primary income is the sum of goods exports, service exports and primary income receipts. Data are in current U.S. dollars. Source: International Debt Statistics	
P	Feminity ratio	UN Population Division	Number of Females per 100 Males	
E	Food production index (2004-2006 = 100)	Food and Agriculture Organization	Food production index covers food crops that are considered edible and that contain nutrients. Coffee and tea are excluded because, although edible, they have no nutritive value. The FAO indices of agricultural production show the relative level of the aggregate volume of agricultural production for each year in comparison with the base period 2004-2006. They are based on the sum of price-weighted quantities of different agricultural commodities produced after deductions of quantities used as seed and feed weighted in a similar manner. The resulting aggregate represents, therefore, disposable production for any use except as seed and feed. All the indices at the country, regional and world levels are calculated by the Laspeyres formula. Production quantities of each commodity are weighted by 2004-2006 average international commodity prices and summed for each year. To obtain the index, the aggregate for a given year is divided by the average aggregate for the base period 2004-2006. The commodities covered in the computation of indices of agricultural production are all crops and livestock products originating in each country. Practically all products are covered, with the main exception of fodder crops. The "international commodity prices" are used in order to avoid the use of exchange rates for obtaining continental and world aggregates, and also to improve and facilitate international comparative analysis of productivity at the national level. These "international prices," expressed in so-called "international dollars," are derived using a Geary-Khamis formula for the agricultural sector. This method assigns a single "price" to each commodity. For example, one metric ton of wheat has the same price regardless of the country where it was produced. The currency unit in which the prices are expressed has no influence on the indices published	$FPI = \frac{\text{FoodProduced}(\text{Year } t) * \text{meanPricefood}(2004-2006)}{\text{meanPricefood}(2004-2006)}$
E	Forest area (% of land area)	Food and Agriculture Organization	Forest area is land under natural or planted stands of trees of at least 5 meters in situ, whether productive or not, and excludes tree stands in agricultural production systems (for example, in fruit plantations and agroforestry systems) and trees in urban parks and gardens. The Food and Agriculture Organization (FAO) provides detail information on forest cover, and adjusted estimates of forest cover. The survey uses a uniform definition of forest. Although FAO provides a breakdown of forest cover between natural forest and plantation for developing countries, forest data used to derive this indicator data does not reflect that breakdown. Total land area does not include inland water bodies such as major rivers and lakes. Variations from year to year may be due to updated or revised data rather than to change in area. The indicator is derived by dividing total area under forest of a country by country's total land area, and multiplying by 100.	

E	GDP per capita (constant 2005 US\$)	World Bank	GDP at purchaser's prices is the sum of gross value added by all resident producers in the economy plus any product taxes and minus any subsidies not included in the value of the products. It is calculated without making deductions for depreciation of fabricated assets or for depletion and degradation of natural resources. Data are in constant 2005 U.S. dollars. Dollar figures for GDP are converted from domestic currencies using 2000 official exchange rates. For a few countries where the official exchange rate does not reflect the rate effectively applied to actual foreign exchange transactions, an alternative conversion factor is used. (National accounts data, and OECD National Accounts data files.	Data are in constant 2005 U.S. dollars.
E	GDP per capita growth (annual %)	World Bank	Annual percentage growth rate of GDP per capita based on constant local currency. Aggregates are based on constant 2005 U.S. dollars. GDP per capita is gross domestic product divided by midyear population. GDP at purchaser's prices is the sum of gross value added by all resident producers in the economy plus any product taxes and minus any subsidies not included in the value of the products. It is calculated without making deductions for depreciation of fabricated assets or for depletion and degradation of natural resources.	
I	Gender parity index for gross intake ratio. Primary	UNESCO	The Gender Parity Index (GPI) is a socioeconomic index usually designed to measure the relative access to education of males and females. In its simplest form, it is calculated as the quotient of the number of females by the number of males enrolled in a given stage of education (primary, secondary, etc.).	
I	Gender parity index for gross enrolment ratio. Secondary. All programmes	UNESCO	The indicator is defined as the ratio of female to male values of Gross enrolment Ratio. The Gross enrolment ratio is defined as the number of pupils or students enrolled in the secondary level of education, regardless of age, expressed as a percentage of the official school-age population corresponding to the same level of education..	
I	Gender parity index for gross enrolment ratio. Tertiary.	UNESCO	The indicator is defined as the ratio of female to male values of Gross enrolment Ratio. The Gross enrolment ratio is defined as the number of pupils or students enrolled in the tertiary level of education, regardless of age, expressed as a percentage of the official school-age population corresponding to the same level of education..	
E	GNI per capita (constant 2005 US \$)	World Bank	GNI per capita is gross national income divided by midyear population. GNI (formerly GNP) is the sum of value added by all resident producers plus any product taxes (less subsidies) not included in the valuation of output plus net receipts of primary income (compensation of employees and property income) from abroad. Data are in constant 2005 U.S. dollars.	
E	Gold Mining (>1% world production)	British Geological survey, National Environment Research Council	Total gold produced for gems and industries (Kilograms) divided by total carates produced in the world (%), considering only a percentage grather than 1% Dummy variable=1 producer country with a production >1 % world production; 0 otherwise	BritMinSurv
I	Gross intake ratio primary female	UNESCO	Number of new entrants in Grade 1 of primary education, regardless of age, expressed as a percentage of the population at the official primary school entrance age.	
I	Gross intake ratio primary male	UNESCO	Number of new entrants in Grade 1 of primary education, regardless of age, expressed as a percentage of the population at the official primary school entrance age.	
I	Gross enrollment ratio. Secondary all programmes female	UNESCO	Number of pupils (female) enrolled in secondary, regardless of age, expressed as a percentage of the population (female) in the theoretical age group for secondary education. Gross enrollment ratio is the ratio of total enrollment, regardless of age, to the population of the age group that officially corresponds to the level of education shown. Secondary education completes the provision of basic education that began at the primary level, and aims at laying the foundations for lifelong learning and human development, by offering more subject- or skill-oriented instruction using more specialized teachers	

I	Gross enrollment ratio. Secondary all programmes male	UNESCO	Number of pupils (male) enrolled in secondary, regardless of age, expressed as a percentage of the population (total, male, female) in the theoretical age group for secondary education.	
I	Gross enrollment ratio. Tertiary (ISCED 5 and 6) female	UNESCO	Number of graduates (female) from programmes at ISCED5 and 6, regardless of age, expressed as a percentage of the population (total, male, female) at the theoretical graduation age for such programmes. Tertiary education, whether or not to an advanced research qualification, normally requires, as a minimum condition of admission, the successful completion of education at the secondary level	
I	Gross enrollment ratio. Tertiary (ISCED 5 and 6) male	UNESCO	Number of graduates (male) from programmes at ISCED5 and 6, regardless of age, expressed as a percentage of the population (total, male, female) at the theoretical graduation age for such programmes.	
	Growth rate refugee by destination country	UNCHR	Difference between the number of refugee in the year T and the consistence of refugee in year T-1, divided by the number of refugee in year T-1 (for countries of origin). Percentage values.	
	Growth rate refugee by origin country	UNCHR	Difference between the number of refugee in the year T and the consistence of refugee in year T-1, divided by the number of refugee in year T-1 (for countries of destination). Percentage values.	
E	Human Development Index (HDI)	UN	<p>The Human Development Index (HDI) is a composite statistic of life expectancy, education, and income indices used to rank countries into four tiers of human development. It was created by the Pakistani economist Mahbub ul Haq and the Indian economist Amartya Sen in 1990 and was published by the United Nations Development Programme.</p> <p>The HDI can be viewed as an index of "potential" human development (or the maximum IHDI that could be achieved if there were no inequality).</p> <p>The 2010 Human Development Index, the HDI, combines three dimensions:</p> <ul style="list-style-type: none"> • A long and healthy life: Life expectancy at birth • Education index: Mean years of schooling and Expected years of schooling • A decent standard of living: GNI per capita (PPP US\$) • <p>In its 2010 Human Development Report, the UNDP began using a new method of calculating the HDI. The following three indices are used:</p> <ol style="list-style-type: none"> 1. Life Expectancy Index (LEI) = $\frac{LE - 20}{82.3 - 20}$ 2. Education Index (EI) = $\frac{\sqrt{MYSI \cdot EYSI}}{0.951}$ <ol style="list-style-type: none"> 2.1 Mean Years of Schooling Index (MYSI) = $\frac{MYS}{13.2}$ 2.2 Expected Years of Schooling Index (EYSI) = $\frac{EYS}{20.6}$ 3. Income Index (II) = $\frac{\ln(GNIpc) - \ln(100)}{\ln(107,721) - \ln(100)}$ <p>Finally, the HDI is the geometric mean of the previous three normalized indices: $HDI = \sqrt[3]{LEI \cdot EI \cdot II}$. LE: Life expectancy at birth MYS: Mean years of schooling (Years that a 25-year-old person or older has spent in schools) EYS: Expected years of schooling (Years that a 5-year-old child will spend with his education in his whole life) GNIpc: Gross national income at purchasing power parity per capita</p>	
E	Imports of goods and services (% of GDP)	World Bank	Imports of goods and services represent the value of all goods and other market services received from the rest of the world. They include the value of merchandise, freight, insurance, transport, travel, royalties, license fees, and other services, such as communication, construction, financial, information, business, personal, and government services. They exclude	

			compensation of employees and investment income (formerly called factor services) and transfer payments.	
H	Improved sanitation facilities (% of population with access)	World Bank	Access to improved sanitation facilities refers to the percentage of the population using improved sanitation facilities. The improved sanitation facilities include flush/pour flush (to piped sewer system, septic tank, pit latrine), ventilated improved pit (VIP) latrine, pit latrine with slab, and composting toilet. (WHO/UNICEF Joint Monitoring Programme (JMP) for Water Supply and Sanitation)	
H	Improved water source (% of population with access)	World Bank	Access to an improved water source refers to the percentage of the population using an improved drinking water source. The improved drinking water source includes piped water on premises (piped household water connection located inside the user's dwelling, plot or yard), and other improved drinking water sources (public taps or standpipes, tube wells or boreholes, protected dug wells, protected springs, and rainwater collection). (WHO/UNICEF Joint Monitoring Programme (JMP) for Water Supply and Sanitation)	
E	Income Gini coefficient	World Bank	The Gini index provides a convenient summary measure of the degree of inequality. Data on the distribution of income or consumption come from nationally representative household surveys. Where the original data from the household survey were available, they have been used to calculate the income or consumption shares by quintile. Otherwise, shares have been estimated from the best available grouped data. The distribution data have been adjusted for household size, providing a more consistent measure of per capita income or consumption. No adjustment has been made for spatial differences in cost of living within countries, because the data needed for such calculations are generally unavailable. Survey year is the year in which the underlying household survey data were collected or, when the data collection period bridged two calendar years, the year in which most of the data were collected. (World Bank Indicators)	
H	Infant mortality rate, Iq0, for both sexes combined (infant deaths per 1,000 live births)	UN Population Division	Probability of dying between birth and exact age 1. It is expressed as average annual deaths per 1,000 births	
S	Intentional homicide, rate per 100,000 population (1995 - 2011)	UN Population Division	Intentional homicide is defined as unlawful death purposefully inflicted on a person by another person. The rate is computed as the quote of intentional homicide per 100.000 inhabitant.	
S	Internal Displaced people (%)	UNCHR	An internally displaced person (IDP) is someone who is forced to flee his or her home but who remains within his or her country's borders. They are often referred to as refugees, although they do not fall within the current legal definition of a refugee.	Internal_displaced/population(total)*100
E	Internet users (per 100 people)	International Telecommunication Union, World Telecommunication/ICT Development Report and database, and World Bank estimates.	The Internet is a world-wide public computer network. It provides access to a number of communication services including the World Wide Web and carries email, news, entertainment and data files, irrespective of the device used (not assumed to be only via a computer - it may also be by mobile phone, PDA, games machine, digital TV etc.). Access can be via a fixed or mobile network.	Internet users are people with access to the worldwide network.

E	Labor force participation rate for ages 15-24, female (%)	ILO	<p>The labor force is the actual number of people available for work. The labor force of a country includes both, the employed and the unemployed.</p> <p>The labor force participation rate is the ratio between the labor force (female 15-24) and the overall size of their cohort (national population of the same age range).</p> <p>(International Labour Organization, Key Indicators of the Labour Market database. Modelled ILO estimates.)</p>	Labor force participation rate for ages 15-24 is the proportion of the population ages 15-24 that is economically active: all people who supply labor for the production of goods and services during a specified period.
E	Labor force participation rate for ages 15-24, male (%)	ILO	<p>The labor force is the actual number of people available for work. The labor force of a country includes both, the employed and the unemployed. The labor force participation rate is the ratio between the labor force (male 15-24) and the overall size of their cohort (national population of the same age range).</p> <p>(International Labour Organization, Key Indicators of the Labour Market database. Modelled ILO estimates.)</p>	Labor force participation rate for ages 15-24 is the proportion of the population ages 15-24 that is economically active: all people who supply labor for the production of goods and services during a specified period.
E	Labor force, female (% of total labor force)	ILO	<p>The labor force is the supply of labor available for producing goods and services in an economy. It includes people who are currently employed and people who are unemployed but seeking work as well as first-time job-seekers. Not everyone who works is included, however. Unpaid workers, family workers, and students are often omitted, and some countries do not count members of the armed forces. Labor force size tends to vary during the year as seasonal workers enter and leave. Usually active population is measured in relation to a long reference period such as a year, and the currently active population (labor force) is measured in relation to a short reference period such as one day or one week.</p> <p>The data are mostly estimated by the ILO, and may differ from national estimates.</p>	The labor force, female (% of total labor force) is the proportion of female population that is economically active over total labour force.
	Land area (sq. km)	Food and Agriculture Organization	Total land area does not include inland water bodies such as major rivers and lakes. Variations from year to year may be due to updated or revised data rather than to change in area.	Land area is a country's total area, excluding area under inland water bodies, national claims to continental shelf, and exclusive economic zones. In most cases the definition of inland water bodies includes major rivers and lakes.
E	Land under cereal production (hectares)	Food and Agriculture Organization	<p>Cereals production includes wheat, rice, maize, barley, oats, rye, millet, sorghum, buckwheat, and mixed grains. Production data on cereals relate to crops harvested for dry grain only. Cereal crops harvested for hay or harvested green for food, feed, or silage and those used for grazing are excluded.</p> <p>A cereal is a grass cultivated for the edible components of their grain, composed of the endosperm, germ, and bran. Cereal grains are grown in greater quantities and provide more food energy worldwide than any other type of crop; cereal crops therefore can also be called staple crops.</p>	Land under cereal production refers to harvested area, although some countries report only sown or cultivated area. Production data on cereals relate to crops harvested for dry grain only. Cereal crops harvested for hay or harvested green for food, feed, or silage and those used for grazing are excluded.
S	Language Fractionalization Index		See: Alesina, A., Devleeschauwer, A., Easterly, W., Kurlat, S., & Wacziarg, R. (2003). Fractionalization. <i>Journal of Economic growth</i> , 8(2), 155-194.	

P	Life expectancy at birth, female (years)	UN Population Division	<p>Life expectancy at birth used here is the average number of years a newborn is expected to live if mortality patterns at the time of its birth remain constant in the future. It reflects the overall mortality level of a population, and summarizes the mortality pattern that prevails across all age groups in a given year. It is calculated in a period life table which reflects a snapshot of a mortality pattern of a population at a given time. It therefore does not reflect actual mortality patterns that a person actually goes through during his/her life, which can be calculated in a cohort life table.</p> <p>High mortality in young age groups significantly lowers the life expectancy at birth. But if a person survives his/her childhood of high mortality, he/she may live much longer. For example, in a population with a life expectancy at birth of 50, there may be few people dying at age 50. The life expectancy at birth may be low due to the high childhood mortality so that once a person survives his/her childhood, he/she may live much longer than 50 years.</p>	
P	Life expectancy at birth, male (years)	UN Population Division	<p>Life expectancy at birth is the average number of years a newborn is expected to live if mortality patterns at the time of its birth remain constant in the future. It reflects the overall mortality level of a population, and summarizes the mortality pattern that prevails across all age groups in a given year. It is calculated in a period life table which reflects a snapshot of a mortality pattern of a population at a given time. It therefore does not reflect actual mortality patterns that a person actually goes through during his/her life, which can be calculated in a cohort life table.</p> <p>High mortality in young age groups significantly lowers the life expectancy at birth. But if a person survives his/her childhood of high mortality, he/she may live much longer. For example, in a population with a life expectancy at birth of 50, there may be few people dying at age 50. The life expectancy at birth may be low due to the high childhood mortality so that once a person survives his/her childhood, he/she may live much longer than 50 years.</p> <p>(World Population Prospects; Statistical Division. Population and Vital Statistics Report (various years), Census reports and other statistical publications from national statistical offices; Eurostat: Demographic Statistics; Secretariat of the Pacific Community: Statistics and Demography Programme, and U.S. Census Bureau: International Database.)</p>	
E	Livestock production index (2004-2006 = 100)	Food and Agriculture Organization	<p>Livestock production index includes meat and milk from all sources, dairy products such as cheese, and eggs, honey, raw silk, wool, and hides and skins</p> <p>$LPI = \frac{\text{LivestockProduced}(\text{Year } t) * \text{meanPriceLivestock}(2004-2006)}{\text{meanLivestock_produced}(2004-2006) * \text{meanPriceLivestock}(2004-2006)}$</p>	See general description of FAO production indices..
	Macro Area	United Nations	<p>List of the names of countries or areas in alphabetical order, three-digit numerical codes used for statistical processing purposes by the Statistics Division of the United Nations Secretariat and three-digit alphabetical codes assigned by the International Organization for Standardization (ISO).</p> <p>The composition of geographical regions used by the Statistics Division in its publications and databases is also shown. This list of countries or areas includes those countries or areas for which statistical data are compiled by the Statistics Division of the United Nations Secretariat. These names are based on Terminology Bulletin No. 347/Rev.1: Country Names.</p> <p>Source: United Nations publication, Standard Country or Area Codes for Statistical Use, Revision 4.</p>	Link: http://unstats.un.org/unsd/methods/m49/m49regin.htm
H	Mortality rate male (NORMALIZED)	UN Population Division	<p>Standardized mortality rate (male), considering as weight (standard population) the mondial population by age group.</p> <p>(Project staff estimates on UN Population Division data)</p>	$SDR_t = \frac{[\sum_{\text{age groups}} (M_{ar} P_{as})]}{P_s} \times 1000$ <p>Mar is the age-specific mortality rate for the region. Pas is the number of people in the age group in the standard population. Ps is the total standard population.</p>

H	Malnutrition prevalence, weight for age (% of children under 5)	World Health Organization	<p>Prevalence of child malnutrition is the percentage of children under age 5 whose weight for age is more than two standard deviations below the median for the international reference population ages 0-59 months. The data are based on the WHO's new child growth standards released in 2006.</p> <p>Estimates of child malnutrition, based on prevalence of underweight, stunting, wasting and overweight, are from national survey data. For more information, consult the original sources. Height is measured by recumbent length for children up to two years old and by stature while standing for older children.</p> <p>(WHO, Global Database on Child Growth and Malnutrition. Aggregation is based on UNICEF/WHO Joint Global Malnutrition Analysis Data Set 2011, and additional analysis by UNICEF.)</p>
H	Maternal mortality ratio	World Health Organization	<p>Maternal mortality ratio is the number of women who die from pregnancy-related causes while pregnant or within 42 days of pregnancy termination per 100,000 live births.</p> <p>The data are estimated with a regression model using information on the proportion of maternal deaths among non-AIDS deaths in women ages 15-49, fertility, birth attendants, and GDP.</p>
H	Measles (MCV) immunization coverage among 1-year-olds (%)	World Health Organization	<p>The percentage of children under one year of age who have received at least one dose of measles-containing vaccine in a given year.</p> <p>Reports of vaccinations performed by service providers (e.g. district health centres, vaccination teams, physicians) are used for estimates based on service/facility records.</p> <p>The estimate of immunization coverage is derived by dividing the total number of vaccinations given by the number of children in the target population, often based on census projections.</p> <p>Data on measles immunization are obtained from the Surveys.</p> <p>The principle types of surveys are the Expanded Programme on Immunization (EPI) 30-cluster survey, the UNICEF Multiple Indicator Cluster Survey (MICS), and the Demographic and Health Survey (DHS). The indicator is estimated as the percentage of children ages 12-23 months who received at least one dose of measles vaccine either any time before the survey or before the age of 12 months.</p>
S	Military expenditure (% of central government expenditure)	Stockholm International Peace Research Institute (SIPRI),	<p>SIPRI military expenditure data includes military and civil personnel, including retirement pensions and social services for military personnel; operation and maintenance; procurement; military research and development; and military aid (in the military expenditures of the donor country). Excluded are civil defense and current expenditures for previous military activities, such as for veterans' benefits, demobilization, and weapons conversion and destruction. This definition cannot be applied for all countries, however, since that would require more detailed information than is available about military budgets and off-budget military expenditures (for example, whether military budgets cover civil defense, reserves and auxiliary forces, police and paramilitary forces, and military pensions).</p> <p>SIPRI data for the most recent years include two types of estimate which apply to all countries: (a) figures for the most recent years are for adopted budgets, budget estimates or revised estimates, and are revised, more often than not, in subsequent years; and (b) the deflator used for the latest year in the series is an estimate.</p> <p>SIPRI's primary source of military expenditure data is official data provided by national governments. These data are derived from budget documents, defense white papers, and other public documents from official government agencies, including government responses to questionnaires sent by SIPRI, the UNODA, or the Organization for Security and Co-operation in Europe. Secondary sources include international statistics, such as those of NATO and the IMF's Government Finance Statistics Yearbook. Other secondary sources include country reports of the Economist Intelligence Unit, country reports by IMF staff, and specialist journals and newspapers.</p> <p>The SIPRI military expenditure figures are presented on a calendar-year basis. The only exception is the USA, for which statistics report data on a fiscal-year basis. Calendar-year data are calculated on the assumption of an even rate of expenditure throughout the fiscal year.</p> <p>(Yearbook: Armaments, Disarmament and International Security)</p>

E	Mobile cellular subscriptions (per 100 people)	International Telecommunication Union, World Telecommunication/ICT Development Report and database, and World Bank estimates.	Refers to the subscriptions to a public mobile telephone service and provides access to Public Switched Telephone Network (PSTN) using cellular technology, including number of pre-paid SIM cards active during the past three months. Post-paid and prepaid subscriptions are included. Data on mobile cellular subscribers are derived using administrative data that countries (usually the regulatory telecommunication authority or the Ministry in charge of telecommunications) regularly, and at least annually, collect from telecommunications operators. Data for this indicator are readily available for approximately 90 percent of countries, either through ITU's World Telecommunication Indicators questionnaires or from official information available on the Ministry or Regulator's website. For the rest, information can be aggregated through operators' data (mainly through annual reports) and complemented by market research reports. Mobile cellular subscriptions (per 100 people) indicator is derived by all mobile subscriptions divided by the country's population and multiplied by 100.	
H	Mortality rate, adult 15-35 female	UN Population Division	Number of female deaths (15-35 age group) over a given period divided by the mean of the female population (15-35) at the start and the end of the period. The denominator is multiplied by five years in which subjects are exposed to the risk of death	Female_death(15-35)/Female_population (15-35)*5
H	Mortality rate, adult 15-35 male	UN Population Division	Number of male deaths (15-35 age group) over a given period divided by the mean of the male population (15-35) at the start and the end of the period. The denominator is multiplied by five years in which subjects are exposed to the risk of death	male_death(15-35)/male_population (15-35)*5
H	Mortality rate, adult, female (per 1,000 female adults)	UN Population Division.	Adult female mortality rate is the probability of dying between the ages of 15 and 60--that is, the probability of a 15-year-old dying before reaching age 60, if subject to current age-specific mortality rates between those ages. The main sources of mortality data are vital registration systems and direct or indirect estimates based on sample surveys or censuses. A "complete" vital registration system - covering at least 90 percent of vital events in the population - is the best source of age-specific mortality data. Where reliable age-specific mortality data are available, life tables can be constructed from age-specific mortality data, and adult mortality rate can be calculated from life tables.	
H	Mortality rate, adult, male (per 1,000 male adults)	UN Population Division.	Adult male mortality rate is the probability of dying between the ages of 15 and 60--that is, the probability of a 15-year-old dying before reaching age 60, if subject to current age-specific mortality rates between those ages The main sources of mortality data are vital registration systems and direct or indirect estimates based on sample surveys or censuses. A "complete" vital registration system - covering at least 90 percent of vital events in the population - is the best source of age-specific mortality data. Where reliable age-specific mortality data are available, life tables can be constructed from age-specific mortality data, and adult mortality rate can be calculated from life tables.	
H	Mortality rate, female (NORMALIZED)	UN Population Division	Standardized mortality rate (female), considering as weight (standard population) the mondial population by age group. (Project staff estimates on UN Population Division data)	$SDR_f = \frac{\sum_{age\ groups} (M_{ar} P_{as})}{P_s} \times 1000$ M_{ar} is the age-specific mortality rate for the region. P_{as} is the number of people in the age group in the standard population. P_s is the total standard population.

H	Mortality rate, under-5 (per 1,000 live births)	United Nation	<p>Under-five mortality rate is the probability per 1,000 that a newborn baby will die before reaching age five, if subject to current age-specific mortality rates</p> <p>The main sources of mortality data are vital registration systems and direct or indirect estimates based on sample surveys or censuses. A "complete" vital registration system - covering at least 90 percent of vital events in the population - is the best source of age-specific mortality data.</p> <p>Estimates of neonatal, infant, and under-five mortality tend to vary by source and method for a given time and place. Years for available estimates also vary by country, making comparison across countries and over time difficult.</p> <p>To make neonatal, infant, and under-five mortality estimates comparable and to ensure consistency across estimates by different agencies, the United Nations Inter-agency Group for Child Mortality Estimation (IGME), which comprises the United Nations Children's Fund (UNICEF), the World Health Organization (WHO), the United Nations Population Division, the World Bank, and other universities and research institutes, developed and adopted a statistical method that uses all available information to reconcile differences.</p> <p>The method uses statistical models to obtain a best estimate trend line by fitting a country-specific local regression model of mortality rates against their reference dates.</p> <p>Neonatal, infant, and under-five mortality rates are higher for boys than for girls in countries in which parental gender preferences are insignificant. Under-five mortality rates capture the effect of gender discrimination better than neonatal and infant mortality rates do. Where female child mortality is higher, girls probably have unequal access to resources.</p> <p>(Estimates developed by the UN Inter-agency Group for Child Mortality Estimation)</p>	Under-five mortality rate is the probability per 1,000 that a newborn baby will die before reaching age five, if subject to current age-specific mortality rates.
E	National poverty line	World Bank	National poverty rate is the percentage of the population living below the national poverty line. National estimates are based on population-weighted subgroup estimates from household surveys.	
H	Neonatal mortality rate (per 1000 live births)	World Health Organization	Neonatal mortality rate is the number of neonates dying before reaching 28 days of age, per 1,000 live births in a given year.	See notes to mortality rate under 5

E	Net ODA received (% of central government expense)	OECD	<p>Net official development assistance (ODA) per capita consists of disbursements of loans made on concessional terms (net of repayments of principal) and grants by official agencies of the members of the Development Assistance Committee (DAC), by multilateral institutions, and by non-DAC countries to promote economic development and welfare in countries and territories in the DAC list of ODA recipients. It includes loans with a grant element of at least 25 percent (calculated at a rate of discount of 10 percent).</p> <p>Central government expense is cash payments for operating activities of the government in providing goods and services. It includes compensation of employees (such as wages and salaries), interest and subsidies, grants, social benefits, and other expenses such as rent and dividends.</p> <p>The flows of official and private financial resources from the members of the Development Assistance Committee (DAC) of the Organisation for Economic Co-operation and Development (OECD) to developing economies are compiled by DAC, based principally on reporting by DAC members using standard questionnaires issued by the DAC Secretariat.</p> <p>The ODA excludes nonconcessional flows from official creditors, which are classified as "other official flows," and aid for military and anti-terrorism purposes. Transfer payments to private individuals, such as pensions, reparations, and insurance payouts, are in general not counted. In addition to financial flows, ODA includes technical cooperation, most expenditures for peacekeeping under UN mandates and assistance to refugees, contributions to multilateral institutions such as the United Nations and its specialized agencies, and concessional funding to multilateral development banks.</p> <p>Flows are transfers of resources, either in cash or in the form of commodities or services measured on a cash basis. Short-term capital transactions (with one year or less maturity) are not counted. Repayments of the principal (but not interest) of ODA loans are recorded as negative flows. Proceeds from official equity investments in a developing country are reported as ODA, while proceeds from their later sale are recorded as negative flows.</p> <p>The official development assistance estimates are published annually at the end of the calendar year in International Development Statistics (IDS) database. Net ODA received as a percent of central government expense is calculated using values in U.S. dollars converted using the DEC alternative conversion factor which is the underlying annual exchange rate used for the World Bank Atlas method.</p>	<p>Development Assistance Committee of the Organisation for Economic Co-operation and Development, Geographical Distribution of Financial Flows to Developing Countries, Development Co-operation Report, and International Development Statistics database. Data are available online at: www.oecd.org/dac/stats/ids online. IMF central government expense estimates are used for the denominator.</p>
E	Net ODA received per capita (current US\$)	OECD	<p>Net official development assistance (ODA) per capita consists of disbursements of loans made on concessional terms (net of repayments of principal) and grants by official agencies of the members of the Development Assistance Committee (DAC), by multilateral institutions, and by non-DAC countries to promote economic development and welfare in countries and territories in the DAC list of ODA recipients. It includes loans with a grant element of at least 25 percent (calculated at a rate of discount of 10 percent).</p> <p>Total population is based on the de facto definition of population. The values shown are midyear estimates. Net official development assistance per capita is net ODA divided by midyear population.</p>	
P	Net reproduction rate (surviving daughters per woman)	UN Population Division	<p>The average number of daughters a hypothetical cohort of women would have at the end of their reproductive period if they were subject during their whole lives to the fertility rates and the mortality rates of a given period. It is expressed as number of daughters per woman.</p>	
H	Number of people (all ages) living with HIV per 1000 pop.	World Health Organization	<p>The number of people with HIV infection, whether or not they have developed symptoms of AIDS, estimated to be alive at the end of a specific year. Method of estimation Countries produce national estimates of the number of people living with HIV, which are compiled and published annually by UNAIDS and WHO. Standard methods and tools for HIV estimates that are appropriate to the pattern of the HIV epidemic are used. However, to obtain the best possible estimates, judgement needs to be used as to the quality of the data and how representative it is of the population.</p>	

E	Openness Indicator 100* (Exp+Imp)/GDP \$2005		Exportation plus importation divided by GDP \$2005, expressed as percentage	
E	Permanent cropland (% of land area)	Food and Agriculture Organization	The data on Permanent cropland and land area are collected by the Food and Agriculture Organization (FAO) of the United Nations from official national sources through the questionnaire are supplemented with information from official secondary data sources. The secondary sources cover official country data from websites of national ministries, national publications and related country data reported by various international organizations.	Km2(crop_area)/total land area (Km2)*100
E	Personal remittances, received (% of GDP)	World Bank	Personal transfers consist of all current transfers in cash or in kind made or received by resident households to or from nonresident households. Personal transfers thus include all current transfers between resident and nonresident individuals. Compensation of employees refers to the income of border, seasonal, and other short-term workers who are employed in an economy where they are not resident and of residents employed by nonresident entities. Data are the sum of two items defined in the sixth edition of the IMF's Balance of Payments Manual: personal transfers and compensation of employees. Data , in current U.S. dollars, are divided by GDP. (WB estimates based on IMF balance of payments data, and World Bank and OECD GDP estimates)	
S	Political Rights	Freedom House	The Freedom in the World survey provides an annual evaluation of the progress and decline of freedom in 195 countries and 14 related and disputed territories. The survey, which includes both analytical reports and numerical ratings, measures freedom according to two broad categories: political rights and civil liberties. Political rights ratings are based on an evaluation of three subcategories: electoral process, political pluralism and participation, and functioning of government. Civil liberties ratings are based on an evaluation of four subcategories: freedom of expression and belief, associational and organizational rights, rule of law, and personal autonomy and individual rights. Each country is assigned a numerical rating from 1 to 7 for both political rights and civil liberties, with 1 representing the most free and 7 the least free. The ratings are determined by the total number of points (up to 100) each country receives on 10 political rights questions and 15 civil liberties questions; countries receive 0 to 4 points on each question, with 0 representing the smallest degree and 4 the greatest degree of freedom. The average of the political rights and civil liberties ratings, known as the freedom rating, determines the overall status: Free (1.0 to 2.5), Partly Free (3.0 to 5.0), or Not Free (5.5 to 7.0). Freedom House also assigns upward or downward trend arrows to countries which saw general positive or negative trends during the year that were not significant enough to result in a ratings change.	The survey assigns the designation of electoral democracy to countries that have met certain minimum standards. The numerical benchmark for a country to be listed as an electoral democracy is a total of 7 points or more (out of a possible 12) for the 3 political rights subcategory questions on electoral process, as well as a total of 20 points or more (out of a possible 40) for all 10 political rights questions. Freedom House does not maintain a culture-bound view of freedom. The methodology of the survey is grounded in basic standards of political rights and civil liberties, derived in large measure from relevant portions of the Universal Declaration of Human Rights. These standards apply to all countries and territories, irrespective of geographical location, ethnic or religious composition, or level of economic development.
E	Phosphate rocks production	British Geological survey, National Environment Research Council	Dummy variable=1 producer country; 0 otherwise	BritMinSurv
P	Population 15-35 female (%)	UN Population Division	Percentage of female Population by 15-35 Age Groups. The population in the year is calculated as the average of the values at the start and end of the period (De facto population as of 1 July of the year indicated). Figures are expressed per 100 total population.	

P	Population 15-35 male (%)	UN Population Division	Percentage of male Population by 15-35 Age Groups. The population in the year is calculated as the average of the values at the start and end of the period (De facto population as of 1 July of the year indicated). Figures are expressed per 100 total population.	
P	Population ages 0-14 (% of total)	UN Population Division	Percentage of Total Population by 0-14 Age Groups. The population in the year is calculated as the average of the values at the start and end of the period Total population is based on the de facto population including all residents regardless of legal status or citizenship - except for refugees not permanently settled in the country of asylum, who are generally considered part of the population of their country of origin. The values shown are midyear estimates. (World Population Prospects)	
P	Population density (people per sq. km of land area)	Food and Agriculture Organization and World Bank population estimates.	Population density is midyear population divided by land area in square kilometers. This ratio can be calculated for any territorial unit for any point in time, depending on the source of the population data. Population estimates are prepared by World Bank staff from variety of sources. They are based on the de facto definition of population and include all residents regardless of legal status or citizenship, within the physical boundaries of a country and under the jurisdiction of that country's political control. Refugees not permanently settled in the country of asylum are considered part of the population of their country of origin. Population numbers are either current census data or historical census data extrapolated through demographic methods. The count also excludes visitors from overseas. Land area is a country's total area, excluding area under inland water bodies, national claims to continental shelf, and exclusive economic zones. In most cases the definition of inland water bodies includes major rivers and lakes.	
P	Population in the largest city (% of urban population)	United Nations	Urban population refers to people living in urban areas as defined by national statistical offices. The indicator is calculated using World Bank population estimates and urban ratios from the United Nations World Urbanization Prospects. The United Nations Population Division and other agencies provide current population estimates for developing countries that lack recent census data and pre- and post-census estimates for countries with census data. The cohort component method - a standard method for estimating and projecting population - requires fertility, mortality, and net migration data, often collected from sample surveys, which can be small or limited in coverage. Population estimates are from demographic modeling and so are susceptible to biases and errors from shortcomings in the model and in the data. Because the five-year age group is the cohort unit and five-year period data are used, interpolations to obtain annual data or single age structure may not reflect actual events or age composition. Countries differ in the way they classify population as "urban" or "rural." Typically, a community or settlement with a population of 2,000 or more is considered urban, but national definitions are most commonly based on size of locality. Eurostat defines urban areas as clusters of contiguous grid cells of 1 km ² with a density of at least 300 inhabitants per km ² and a minimum population of 5,000. Further it defines high-density cluster as contiguous grid cells of 1 km ² with a density of at least 1,500 inhabitants per km ² and a minimum population of 50,000. The population of a city or metropolitan area depends on the boundaries chosen. For example, in 1990 Beijing, China, contained 2.3 million people in 87 square kilometers of "inner city" and 5.4 million in 158 square kilometers of "core city." The population of "inner city and inner suburban districts" was 6.3 million and that of "inner city, inner and outer suburban districts, and inner and outer counties" was 10.8 million. (Most countries use the last definition.) (World Urbanization Prospects.)	
P	Population in urban agglomerations of more than 1 million (% of total population)	World Bank	Population in urban agglomerations of more than one million is the percentage of a country's population living in metropolitan areas that in 2000 had a population of more than one million people. (World Bank Staff estimates based on United Nations, World Urbanization Prospects)	

P	Population, total	UN Population Division	<p>Population projections, starting from a base year (most recent year that ends with 0 or 5), are projected forward using five-year period assumptions of mortality, fertility, and migration by age and sex through 2050, based on the UN Population Division's World Population Prospects database medium variant.</p> <p>Population estimates are usually based on national population censuses. Estimates for the years before and after the census are interpolations or extrapolations based on demographic models.</p> <p>Errors and undercounting occur even in high-income countries; in developing countries errors may be substantial because of limits in the transport, communications, and other resources required to conduct and analyze a full census.</p> <p>The quality and reliability of official demographic data are also affected by public trust in the government, government commitment to full and accurate enumeration, confidentiality and protection against misuse of census data, and census agencies' independence from political influence. Moreover, comparability of population indicators is limited by differences in the concepts, definitions, collection procedures, and estimation methods used by national statistical agencies and other organizations that collect the data.</p> <p>The currentness of a census and the availability of complementary data from surveys or registration systems are objective ways to judge demographic data quality. Some European countries' registration systems offer complete information on population in the absence of a census. International migration is the only other factor besides birth and death rates that directly determines a country's population growth. Estimating migration is difficult. At any time many people are located outside their home country as tourists, workers, or refugees or for other reasons. Standards for the duration and purpose of international moves that qualify as migration vary, and estimates require information on flows into and out of countries that is difficult to collect.</p>	
S	Population: Christian %	PEW	Percentage of Christian population on the total population in the year. Figures are expressed per 100 total population.	Cristian population/Total population*100
S	Population: Muslim %	PEW	Percentage of Muslim population on the total population in the year. Figures are expressed per 100 total population.	Muslim population/Total population*100
E	Poverty gap at \$1.25 a day (PPP) (%)	World Bank	Poverty gap is the mean shortfall from the poverty line (counting the nonpoor as having zero shortfall), expressed as a percentage of the poverty line. This measure reflects the depth of poverty as well as its incidence. (WB-Human development)	
H	Prevalence of tuberculosis (per 100000 population)	WHO	<p>The number of cases of tuberculosis (all forms) in a population at a given point in time (the middle of the calendar year), expressed as the rate per 100 000 population. It is sometimes referred to as "point prevalence". Estimates include cases of TB in people with HIV.</p> <p>See Annex 1 of the WHO global tuberculosis control report</p>	
H	Prevalence of undernourishment (% of population)	Food and Agriculture Organization	Data on undernourishment are from the Food and Agriculture Organization (FAO) of the United Nations and measure food deprivation based on average food available for human consumption per person, the level of inequality in access to food, and the minimum calories required for an average person. (The State of Food Insecurity in the World)	
H	Prevalent Ethnic group (% total population)	CIA The World Factbook	This entry provides a list of prevalent ethnic groups by countries as percentage of total population.	
E	Producing more than 1% of world oil production	British Geological survey, National Environment Reserach Council	<p>Total oil produced (tonnes (metric)) fdivided by total produced in the world (%), considering only a percentage grather than 1 %. (percentage)</p> <p>Dummy variable=1 producer country with a production >1% world production; 0 otherwise</p>	British Min. survey

S	Proportion of seats held by women in national parliaments (%)	Inter-Parliamentary Union (IPU)	<p>The proportion of seats held by women in national parliaments is the number of seats held by women members in single or lower chambers of national parliaments, expressed as a percentage of all occupied seats; it is derived by dividing the total number of seats occupied by women by the total number of seats in parliament.</p> <p>National parliaments can be bicameral or unicameral. This indicator covers the single chamber in unicameral parliaments and the lower chamber in bicameral parliaments. It does not cover the upper chamber of bicameral parliaments. Seats are usually won by members in general parliamentary elections. Seats may also be filled by nomination, appointment, indirect election, rotation of members and by-election. Seats refer to the number of parliamentary mandates, or the number of members of parliament.</p>	
I	Public spending on education, total (% of GDP)	UNESCO	Public expenditure on education as % of GDP is the total public expenditure (current and capital) on education expressed as a percentage of the Gross Domestic Product (GDP) in a given year. Public expenditure on education includes government spending on educational institutions (both public and private), education administration, and transfers/subsidies for private entities (students/households and other private entities).	
I	Public spending on education, total (% of government expenditure)	UNESCO	Public expenditure on education as % of total government expenditure is the total public education expenditure (current and capital) expressed as a percentage of total government expenditure for all sectors in a given financial year. Public education expenditure includes government spending on educational institutions (both public and private), education administration, and subsidies for private entities (students/households and other private entities).	
I	Pupil-teacher ratio. Primary	UNESCO	Pupil-teacher ratio. Primary is the number of pupils enrolled in primary school divided by the number of primary school teachers.	
I	Pupil-teacher ratio. Secondary	UNESCO	Pupil-teacher ratio. Secondary is the number of pupils enrolled in secondary school divided by the number of primary school teachers.	
	Refugee in camps (% total arrival country population)	UNCHR	Consistence of refugee (in camp) divided by population of the country of arrival (percentage)	

	Refugee population by destination country (% destination country population)	UNHCR	<p>The United Nations High Commissioner for Refugees (UNHCR) collects and maintains data on refugees in their Statistical Online Population Database. The refugee data does not include Palestinian refugees residing in areas under the mandate of the United Nations Relief and Works Agency for Palestine Refugees in the Near East (UNRWA). However, the Palestinian refugees living outside the UNRWA areas of operation do fall under the responsibility of UNHCR and are thus included in the Statistical Online Population Database.</p> <p>Refugees are an important part of migrant stock. The refugee data refer to people who have crossed an international border to find sanctuary and have been granted refugee or refugee-like status or temporary protection. There are three main providers of refugee data: governmental agencies, UNHCR field offices and NGOs. Registrations, together with other sources - including estimates and surveys - are the main sources of refugee data. In the absence of Government estimates, UNHCR has estimated the refugee population in most industrialized countries, based on recognition of asylum-seekers. Prior to 2007, resettled refugees were included in these estimates.</p> <p>Up to and including 2006, to ensure that the refugee population in countries that lack a refugee registry is reflected in the global statistics, the number of refugees was estimated by UNHCR based on the arrival of refugees through resettlement programmes and the individual recognition of refugees over a 10-year (Europe and, since 2006, the United States) or 5-year (the United States before 2006, Canada and Oceania) period. Starting with the 2007 data, the cut-off period has been harmonized and now covers a 10-year period for Europe and non-European countries. Resettled refugees, however, are excluded from the refugee estimates in all countries.</p> <p>The 2007-2011 refugee population category includes people in a refugee-like situation, most of who were previously included in the Others of concern group. This sub-category is descriptive in nature and includes groups of persons who are outside their country or territory of origin and who face protection risks similar to those of refugees, but for whom refugee status has, for practical or other reasons, not been ascertained.</p> <p>Asylum seekers - people who have applied for asylum or refugee status and who have not yet received a decision or who are registered as asylum seekers - and internally displaced people - who are often confused with refugees - are not included in the data. Unlike refugees, internally displaced people remain under the protection of their own government, even if their reason for fleeing was similar to that of refugees.</p> <p>Palestinian refugees are people (and their descendants) whose residence was Palestine between June 1946 and May 1948 and who lost their homes and means of livelihood as a result of the 1948 Arab-Israeli conflict.</p>	Statistical Yearbook and data files, complemented by statistics on Palestinian refugees under the mandate of the UNRWA as published on its website. Data from UNHCR are available online at: www.unhcr.org/statistics/populationdatabase .)
	Refugee population by origin country (% origin country population)	UNHCR	See description of Refugee population by destination country	
S	Religion Fractionalization Index		See: Alesina, A., Devleeschauwer, A., Easterly, W., Kurlat, S., & Wacziarg, R. (2003). Fractionalization. <i>Journal of Economic growth</i> , 8(2), 155-194.	
E	Road density (km of road per 100 sq. km of land area)	International Road Federation	National road associations are the primary source of International Road Federation (IRF) data.	Road density is the ratio of the length of the country's total road network to the country's land area. The road network includes all roads in the country: motorways, highways, main or national roads, secondary or regional roads, and other urban and rural roads.

P	Rural population (% of total population)	World Bank	<p>Rural population is calculated as the difference between the total population and the urban population. With more than 3 billion people, including 70 percent of the world's poor people, living in rural areas, adequate indicators to monitor progress in rural areas are essential. However, few indicators are disaggregated between rural and urban areas. Rural population is approximated as the midyear non-urban population. While a practical means of identifying the rural population, it is not a precise measure.</p> <p>The United Nations Population Division and other agencies provide current population estimates for developing countries that lack recent census data and pre- and post-census estimates for countries with census data.</p> <p>World Bank Staff estimates based on United Nations, World Urbanization Prospects.</p>	Rural population refers to people living in rural areas as defined by national statistical offices. It is calculated as the difference between total population and urban population.
S	Status NotFree=1; PartiallyF=2; Free=3	Freedom House	<p>The Freedom in the World survey provides an annual evaluation of the progress and decline of freedom in 195 countries and 14 related and disputed territories. The survey, which includes both analytical reports and numerical ratings, measures freedom according to two broad categories: political rights and civil liberties. Political rights ratings are based on an evaluation of three subcategories: electoral process, political pluralism and participation, and functioning of government. Civil liberties ratings are based on an evaluation of four subcategories: freedom of expression and belief, associational and organizational rights, rule of law, and personal autonomy and individual rights. Each country is assigned a numerical rating from 1 to 7 for both political rights and civil liberties, with 1 representing the most free and 7 the least free. The ratings are determined by the total number of points (up to 100) each country receives on 10 political rights questions and 15 civil liberties questions; countries receive 0 to 4 points on each question, with 0 representing the smallest degree and 4 the greatest degree of freedom.</p> <p>The average of the political rights and civil liberties ratings, known as the freedom rating, determines the overall status: Free (1.0 to 2.5), Partly Free (3.0 to 5.0), or Not Free (5.5 to 7.0).</p> <p>Freedom House also assigns upward or downward trend arrows to countries which saw general positive or negative trends during the year that were not significant enough to result in a ratings change.</p>	See Political rights, Civil liberties
E	Total debt service (% of exports of goods, services and primary income)	World Bank	<p>Data on external debt are gathered through the World Bank's Debtor Reporting System (DRS). Long term debt data are compiled using the countries report on public and publicly guaranteed borrowing on a loan-by-loan basis and private non guaranteed borrowing on an aggregate basis. These data are supplemented by information from major multilateral banks and official lending agencies in major creditor countries. Short-term debt data are gathered from the Quarterly External Debt Statistics (QEDS) database, jointly developed by the World Bank and the IMF and from creditors through the reporting systems of the Bank for International Settlements. Debt data are reported in the currency of repayment and compiled and published in U.S. dollars. End-of-period exchange rates are used for the compilation of stock figures (amount of debt outstanding), and projected debt service and annual average exchange rates are used for the flows. Exchange rates are taken from the IMF's International Financial Statistics. Debt repayable in multiple currencies, goods, or services and debt with a provision for maintenance of the value of the currency of repayment are shown at book value.</p>	Total debt service is the sum of principal repayments and interest actually paid in foreign currency, goods, or services on long-term debt, interest paid on short-term debt, and repayments (repurchases and charges) to the IMF. (WB, International Debt Statistics)
E	Total debt service (% of GNI)	World Bank	Total debt service is divided by GNI and expressed as percentage	Total debt service is the sum of principal repayments and interest actually paid in foreign currency, goods, or services on long-term debt, interest paid on short-term debt, and repayments (repurchases and charges) to the IMF.

P	Total density per 100000 population: Provincial hospitals	WHO	Number of provincial hospitals, from the public and private sectors, per 100,000 population.	Information collected directly from ministries of health through the baseline country survey on medical devices 2010, conducted by HQ/HSS/EHT/DIM. The population data was obtained from the United Nations Statistics Division.
P	Total dependency ratio (ratio of population aged 0-14 and 65+ per 100 population 15-64)	UN Population Division	Total Dependency Ratio ((Age 0-14 + Age 65+) / Age 15-64). De facto population as of 1 July of the year indicated.	
P	Total fertility (children per woman)	UN Population Division	The average number of children a hypothetical cohort of women would have at the end of their reproductive period if they were subject during their whole lives to the fertility rates of a given period and if they were not subject to mortality. It is expressed as children per woman.	Total Fertility Rate (TFR) = (The sum of the Age Specific Fertility Rates x The number of years in each age group) / 1000
	UN Classification Countries (developed-developing regions)	United Nations	Classification used by the Statistics Division in its publications and databases with reference to economic development. Link: http://unstats.un.org/unsd/methods/m49/m49regin.htm It identifies developed and developing countries.	
E	Unemployment, total (% of total labor force)	ILO	The standard definition of unemployed persons is those individuals without work, seeking work in a recent past period, and currently available for work. Persons who did not look for work but have an arrangements for a future job are counted as unemployed. It is the labour force or the economically active portion of the population that serves as the base for this indicator, not the total population. (International Labour Organization, Key Indicators of the Labour Market database)	Unemployment refers to the share of the labor force that is without work but available for and seeking employment. Definitions of labor force and unemployment differ by country.
E	Use of IMF credit (DOD, current US\$)	World Bank	Data related to the operations of the IMF come from the IMF Treasurer's Department and are converted from special drawing rights (SDRs) into dollars using end-of-period exchange rates for stocks and average over the period exchange rates for converting flows. DOD refers to disbursed and outstanding debt; data are in current U.S. dollars. Data on external debt are gathered through the World Bank's Debtor Reporting System (DRS). Long term debt data are compiled using the countries report on public and publicly guaranteed borrowing on a loan-by-loan basis and private non guaranteed borrowing on an aggregate basis. These data are supplemented by information from major multilateral banks and official lending agencies in major creditor countries. Short-term debt data are gathered from the Quarterly External Debt Statistics (QEDS) database, jointly developed by the World Bank and the IMF and from creditors through the reporting systems of the Bank for International Settlements. Debt data are reported in the currency of repayment and compiled and published in U.S. dollars. End-of-period exchange rates are used for the compilation of stock figures (amount of debt outstanding), and projected debt service and annual average exchange rates are used for the flows. Exchange rates are taken from the IMF's International Financial Statistics. Debt repayable in multiple currencies, goods, or services and debt with a provision for maintenance of the value of the currency of repayment are shown at book value.	Use of IMF credit denotes members' drawings on the IMF other than amounts drawn against the country's reserve tranche position. Use of IMF credit includes purchases and drawings under Stand-By, Extended, Structural Adjustment, Enhanced Structural Adjustment, and Systemic Transformation Facility Arrangements as well as Trust Fund loans. SDR allocations are also included in this category. (International Debt Statistics)
H	Women aged 15+ living with HIV (per 1000 women aged 15+)	UNAIDS WHO	Percentage of the total female population ages 15+ who are living with HIV. HIV prevalence rates reflect the rate of HIV infection in each country's population. Low national prevalence rates can be misleading, however. They often disguise epidemics that are initially concentrated in certain localities or population groups and threaten to spill over into the wider population. In many developing countries most new infections occur in young adults, with young women especially vulnerable.	

			<p>Data on HIV are from the Joint United Nations Programme on HIV/AIDS (UNAIDS). Changes in procedures and assumptions for estimating the data and better coordination with countries have resulted in improved estimates of HIV and AIDS. For example, improved software was used to model the course of HIV epidemics and their impacts, making full use of information on HIV prevalence trends from surveillance data as well as survey data. The software explicitly includes the effect of antiretroviral therapy when calculating HIV incidence and models reduced infectivity among people receiving antiretroviral therapy, which is having a larger impact on HIV prevalence and allowing HIV-positive people to live longer. The software also allows for changes in urbanization over time - important because prevalence is higher in urban areas and because many countries have seen rapid urbanization over the past two decades. The estimates include plausible bounds, not shown in the Indicator data, which reflect the certainty associated with each of the estimates.</p> <p>(UNAIDS and the WHO's Report on the Global AIDS Epidemic)</p>	
	World Bank classification countries	World Bank	<p>Economies are divided according to 2012 GNI per capita, calculated using the World Bank Atlas method. The groups are: low income, \$1,035 or less; lower middle income, \$1,036 - \$4,085; upper middle income, \$4,086 - \$12,615; and high income, \$12,616 or more.</p>	
I	Youth (15-24) literacy rate (%). Female	WB UNESCO	<p>Youth literacy rate is the percentage of female population ages 15-24 who can, with understanding, read and write a short, simple statement on their everyday life.</p> <p>Literacy statistics for most countries cover the population ages 15 and older, but some include younger ages or are confined to age ranges that tend to inflate literacy rates. The youth literacy rate for ages 15-24 reflects recent progress in education. It measures the accumulated outcomes of primary education over the previous 10 years or so by indicating the proportion of the population who have passed through the primary education system and acquired basic literacy and numeracy skills. Generally, literacy also encompasses numeracy, the ability to make simple arithmetic calculations.</p>	<p>The UIS collects data on youth and adult literacy through its annual survey on literacy and educational attainment. Adult literacy rates concern the population aged 15 years and older, while youth literacy rates cover the population between the ages of 15 to 24 years.</p>
I	Youth (15-24) literacy rate (%). Male	WB UNESCO	<p>Youth literacy rate is the percentage of male population ages 15-24 who can, with understanding, read and write a short, simple statement on their everyday life.</p> <p>Literacy statistics for most countries cover the population ages 15 and older, but some include younger ages or are confined to age ranges that tend to inflate literacy rates. The youth literacy rate for ages 15-24 reflects recent progress in education. It measures the accumulated outcomes of primary education over the previous 10 years or so by indicating the proportion of the population who have passed through the primary education system and acquired basic literacy and numeracy skills. Generally, literacy also encompasses numeracy, the ability to make simple arithmetic calculations.</p>	

Legenda: E=Economics; I=Education; P=Population; H=Health; S=Society/politics. Variables without label refer to classifications or to statistical indicators which have not been used in the estimation step of the AFM analysis. These variables may be used as supplementary information in the graphical representation.

A.2 Descrizione dell'archivio EM-DAT del CRED

(estratto dal sito: www.emdat.be)

• CRITERIA

For a disaster to be entered into the database at least one of the following criteria must be fulfilled:

- Ten (10) or more people reported killed.
- Hundred (100) or more people reported affected.
- Declaration of a state of emergency.
- Call for international assistance.

• DEFINITION AND CLASSIFICATION

Natural Disaster Categories, Types, and Subtypes

	Hydrometeorological		
Biological	Geophysical	Hydrological	Meteorological
Epidemic	Earthquake	Flood	Storm
Viral infectious disease	Volcano	General flood	Tropical cyclone
Bacterial infectious disease	Mass movement (dry)	Storm surge/coastal flood	Extra-tropical cyclone
Parasitic infectious disease	Rockfall	Mass movement (wet)	Local storm
Fungal infectious disease	Landslide	Rockfall	Climatological
Prion infectious disease	Avalanche	Landslide	Extreme temperature
Insect infestation	Subsidence	Avalanche	Heat wave
Animal stampede		Subsidence	Cold wave
			Extreme winter condition
			Drought/wildfire
			Forest fire
			Land fire

Source: UCL, "EM-DAT: The OFDA/CRED International Disaster Database," UCL, <http://www.emdat.be>.

EM-DAT data include the main following information:

Disaster number: A unique disaster number for each event (8 digits: 4 digits for the year and 4 digits for the disaster number - i.e.: 19950324).

Country: Country(ies) in which the disaster has occurred.

Disaster group: Three groups of disasters are distinguished in EM-DAT: natural disasters, technological disasters and complex emergencies.

Disaster type: Description of the disaster according to a pre-defined classification.

Date: When the disaster occurred. The date is entered as follow: Month/Day/Year.

Killed: Persons confirmed as dead and persons missing and presumed dead (official figures when available).

Injured: People suffering from physical injuries, trauma or an illness requiring medical treatment as a direct result of a disaster.

Homeless: People needing immediate assistance for shelter.

Affected: People requiring immediate assistance during a period of emergency; it can also include displaced or evacuated people.

Total affected: Sum of injured, homeless, and affected.

Estimated Damage: Several institutions have developed methodologies to quantify these losses in their specific domain. However, there is no standard procedure to determine a global figure for economic impact. Estimated damage are given (000') US\$.

- **Recording disasters**

When compiling disaster data, we are sometime confronted with methodological dilemma that can influence the analysis of the data. There are different ways to compile disasters.

When a disaster occurs, it can affect several countries. A hurricane that affects 5 countries can be entered as one disaster or as 5 different disasters.

EM-DAT has historically entered disasters at the country level. This means that a hurricane that affected 5 countries was entered into the database as 5 different entries, one for each of the affected countries (with 5 different Disaster Number).

In 2003, the EM-DAT team decided to improve the methodology for entering disasters into the database. Disasters are currently entered by event. This means that a multi-country disaster will have one identifier (one unique identifier number = Disaster Number), but will have the same number of entries as the number of countries that are affected.

For example, the December 26, 2004 tsunami affected 13 different countries. In EM-DAT, the tsunami has been entered as one event (with one unique Disaster Number) with 13 different country-level disasters.

Main Sources used in EM-DAT. (non exhaustive)

Source Type	Source Information	Type of Disasters Covered
United Nations	OCHA IRIN WFP UNEP WMO WHO FAO	Natural disasters Natural and technological disasters (Africa) Droughts/Famine Natural and technological disasters Natural disasters Epidemics Droughts/Famine
National Governements	National Governments ADRC CDERA	Natural and technological disasters Natural disasters
US Governments	FEMA NOAA OFDA USGS Smithsonian DFO CDC	Natural disasters (America) Natural disasters Natural and technological disasters Earthquakes Volcanoes Floods, slides and windstorms Epidemics
NGO's	IFRC	---
Inter-Governmental Organizations	World Bank	Major natural disasters
ReInsurrance Compagnies	Lloyd Casualty Week SwissRe MünichRe	Natural and some major technological disasters Natural and technological disasters Major natural and technological disasters
Press	AFP	Natural and technological disasters
Others	AirDisaster Emerg.Manag. etc...	Air accidents Natural and technological disasters

A.3 Statistiche descrittive e dati mancanti

EDUCATION	Mean	Q1	Median	Q3	Standard Deviation	Missing	%
Education attain % completed Primary-TOTAL	17,37	8,7	16,29	24,19	10,90	34	19,32
Education attain % completed Primary-FEMALE	16,76	7,95	15,22	24,55	11,11	34	19,32
Education attain % completed Secondary - TOTAL	23,77	11,28	23,16	35,12	14,59	34	19,32
Education attain % completed Secondary-FEMALE	23,08	10,71	23,42	33,7	14,58	34	19,32
Education attain % completed Tertiary-TOTAL	7,38	2,35	6,06	10,63	6,16	34	19,32
Education attain % completed Tertiary-FEMALE	7,48	1,9	6,08	11,09	6,57	34	19,32
Gross intake ratio. Primary. Female	106,78	96,15	101,37	108,74	21,11	51	28,98
Gross intake ratio. Primary. Male	109,64	97,57	102,9	113,89	22,74	51	28,98
Gross enrolment ratio. Secondary. Female	80,3	61,86	90,66	98,94	28,44	53	30,11
Gross enrolment ratio. Secondary. Male	81,56	66,18	87,51	99,51	25,26	54	30,68
Gross enrolment ratio. Tertiary (ISCED 5 and 6). Female	46,04	19,16	44,2	69,87	32,11	66	37,50
Gross enrolment ratio. Tertiary (ISCED 5 and 6). Male	36,65	16,41	35,87	56,56	23,49	66	37,50
Gender parity index for gross intake ratio. Primary	0,98	0,97	0,99	1,01	0,54	51	28,98
Gender parity index for gross enrolment ratio. Secondary	1,86	0,95	1	1,04	9,78	54	30,68
Gender parity index for gross enrolment ratio. Tertiary	1,18	0,89	1,22	1,41	0,57	67	38,07
Pupil-teacher ratio. Primary	26,05	14,29	20,92	33,65	15,50	53	30,11
Pupil-teacher ratio. Secondary	17,53	10,8	14,5	21,87	9,04	91	51,70
Youth (15-24) literacy rate (%). Female	89,12	81,3	98	99,61	15,60	76	43,18
Youth (15-24) literacy rate (%). Male	91,17	85,88	97,42	99,49	11,31	76	43,18
Adult (15+) literacy rate (%). Female	79,71	61,97	90,09	97,08	21,69	78	44,32
Adult illiterate population. % female >15	62,07	56,6	63,74	68,44	10,95	78	44,32
Public expenditure on education as % of total government expenditure	15,68	11,48	15,13	18,52	5,29	98	55,68
Public expenditure on education as % of GDP	4,87	3,66	4,71	5,82	1,89	79	44,89

POLITICS	Mean	Q1	Median	Q3	Standard Deviation	Missing	%
Language Fraction Index	0,39	0,11	0,35	0,63	0,28	6	3,41
Ethnic Fraction Index	0,46	0,24	0,49	0,67	0,26	10	5,68
Religion Fraction Index	0,43	0,21	0,44	0,62	0,22	6	3,41
Prevalent Ethnic group (% total)	61,78	40	70,5	90	31,99	0	0,00
% Population-Christian	52,26	7,55	64,3	87,8	37,72	0	0,00
% Population-Muslim	28,29	0,2	5,45	62,65	38,16	0	0,00
Civil Liberties	3,45	2	3	5	1,80	6	3,41
Political Rights	3,62	1	3	6	2,15	6	3,41
Corruption Index (100=nocorruption)	42,56	28	37	54	19,65	7	3,98
Womens suffrage	61,83	50	60	74	19,92	14	7,95
Proportion of seats held by women in national parliaments (%)	18,08	10	16,75	23,85	10,75	12	6,82
Military expenditure (% of central government expenditure)	10,29	3,55	5,95	10,75	22,68	79	44,89
Armed forces personnel (% of total labor force)	1,41	0,47	0,84	1,56	1,72	15	8,52
Volume arms import-SIPRI trend indicator values (TIVs)	225,02	15	51	217	438,32	63	35,80

ECONOMICS	Mean	Q1	Median	Q3	Standard Deviation	Missing	%
GDP per capita (constant 2005 US\$)	10415,56	854,32	3362,74	14189,06	15113,71	10	5,68
GDP per capita growth (annual %)	2,93	0,8	3,03	5,06	3,96	8	4,55
GNI per capita (constant 2005 US\$)	10936,71	915,25	3619,01	14692,2	14949,90	45	25,57
Income Gini coefficient	39,99	33,67	38,77	45,42	8,84	44	25,00
Human Development Index (HDI)	0,67	0,52	0,7	0,8	0,17	6	3,41
Exports of goods and services (% of GDP)	44,14	25,5	35,96	54,47	30,48	16	9,09
Imports of goods and services (% of GDP)	47,9	31,41	43,04	59,85	26,06	16	9,09
Openness Indicator	92,04	60,75	79,14	113,96	53,62	16	9,09
Total debt service (% of GNI)	4,13	0,95	2,49	4,63	5,94	62	35,23
External debt stocks (% of exports of goods, services and primary income)	121,08	75,73	109,11	161,83	67,37	74	42,05
Total debt service (% of exports of goods, services and primary income)	11,35	3,46	6,96	15	12,43	74	42,05
Use of IMF credit (2010, million\$)/Quota(SRD, million)	1,9	0	1,47	2,35	2,41	6	3,41
Personal remittances received (% of GDP)	4,35	0,31	1,58	5,49	6,60	30	17,05
Net ODA received (% of central government expense)	646,27	1,06	6,21	32,86	4823,08	116	65,91
Net ODA received per capita (current US\$)	82,59	14,52	52,28	90,5	117,65	51	28,98
National poverty line	34,52	21,4	33,5	50,1	19,06	91	51,70
Poverty headcount ratio at \$1.25 a day	23,17	0,92	14,52	40,63	24,87	86	48,86
Labour force, female (% of total labor force)	40,78	37,54	44,39	47,11	9,39	4	2,27
Labor force participation rate for ages 15-24, female (%)	40,8	30,15	37,55	51,3	16,67	4	2,27
Labor force participation rate for ages 15-24, male (%)	53,25	43,05	53,6	61,7	13,40	4	2,27
Unemployment, total (% of total labor force)	8,77	4,9	7,5	10,8	5,67	82	46,59
CO2 emissions (kt)	179914,08	2882,26	13560,57	72258,24	781191,89	3	1,70
Electric power consumption (kWh per capita)	4279	830,51	2453,65	5889,53	5955,08	42	23,86
Road density (km of road per 100 sq.km of land area)	69,04	10,29	22,49	65,23	153,46	4	2,27
Internet users (per 100 inhab.)	33,21	8,37	28,33	53	27,52	5	2,84
Mobile cellular subscriptions (per 100 people)	88,3	57,14	91,56	117,19	43,37	2	1,14

DEMOGRAPHIC INDICATORS	Mean	Q1	Median	Q3	Standard deviation	Missing	%
Population, total	38839848	3467022	9434063	27417194	11814,1823	0	0,00
Average annual rate of population change (percentage)	1,63	0,58	1,39	2,45	1,37167371	2	1,14
Population density (people per sq.km of land area)	322,73	27,43	71,88	143,42	40,2307966	1	0,57
Population in the largest city (% of urban population)	30,86	19,82	26,49	40,79	4,18145485	54	30,68
% Population in urban agglomerations of more than 1million	15,54	0	12,33	22,7	4,25603182	0	0,00
Rural population (% of total population)	42,99	24,98	42,46	62,47	4,84702137	1	0,57
Feminity ratio	100,5	98,86	101,55	103,98	3,11720956	1	0,57
Population ages 0-14 (% of total)	29,15	18,05	29,04	39,94	3,33274735	0	0,00
Population ages 15-35 female (% of total)	0,33	0,29	0,34	0,36	0	1	0,57
Population ages 15-35 male (% of total)	0,28	0,26	0,28	0,3	0	1	0,57
Mortality rate (NORMALIZED) female	8,33	4,38	6,12	11,59	2,28203563	1	0,57
Mortality rate (NORMALIZED) male	9,82	6,13	8,27	12,82	2,1920779	1	0,57
Maternal mortality ratio	174,26	20	65,5	255	15,1199093	4	2,27
Neonatal mortality rate (per 1000 live births)	16	4,6	11,6	26,3	3,61955038	5	2,84
Infant mortality rate, 1q0, for both sexes combined (infant deaths per 1.000 live births)	33,27	7,02	21,08	52,68	5,53274547	1	0,57
Mortality rate, under-5 (per 1.000 live births)	41,96	8,3	21,7	66,7	6,62016541	4	2,27
Mortality rate, adult, female (per 1.000 female adults)	171,63	79,32	129,54	247,5	11,1731095	26	14,77
Mortality rate, adult,male (per 1.000 male adults)	235,77	155,59	215,99	297,1	10,7274726	26	14,77
Ages 15-35 mortality rate female	2,11	0,38	0,91	3,02	1,63807252	1	0,57
Ages 15-35 mortality rate male	2,59	0,9	2,04	3,58	1,44749186	1	0,57
Total fertility (children per woman)	3,03	1,75	2,5	4,22	1,27225839	1	0,57
Net reproduction rate (surviving daughters per woman)	1,3	0,83	1,13	1,74	0,7636067	1	0,57
15-19 Age-specific fertility rates (births per 1.000 women)	56	16,78	44,55	84,59	6,80089223	1	0,57
20-24 Age-specific fertility rates (births per 1.000 women)	139,44	76,39	130,23	201,08	8,75065433	1	0,57

DEMOGRAPHIC INDICATORS	Mean	Q1	Median	Q3	Standard deviation	Missing	%
25-29 Age-specific fertility rates (births per 1.000 women)	158,45	103,06	134,75	216,32	8,26874497	1	0,57
30-34 Age-specific fertility rates (births per 1.000 women)	129,58	83,4	109,87	176,06	7,83797594	1	0,57
Child dependency ratio (ratio of population aged 0-14 per 100 population 15-64)	48,61	26,1	44,11	68,94	4,9089849	1	0,57
Total dependency ratio (ratio of population aged 0-14 and 65+ per 100 population 15-64)	60,28	46,12	53,47	77,61	4,41738936	1	0,57
Intentional homicide, rate per 100.000 population (1995-2011)	9,99	1,57	4,6	13,79	3,56094084	17	9,66

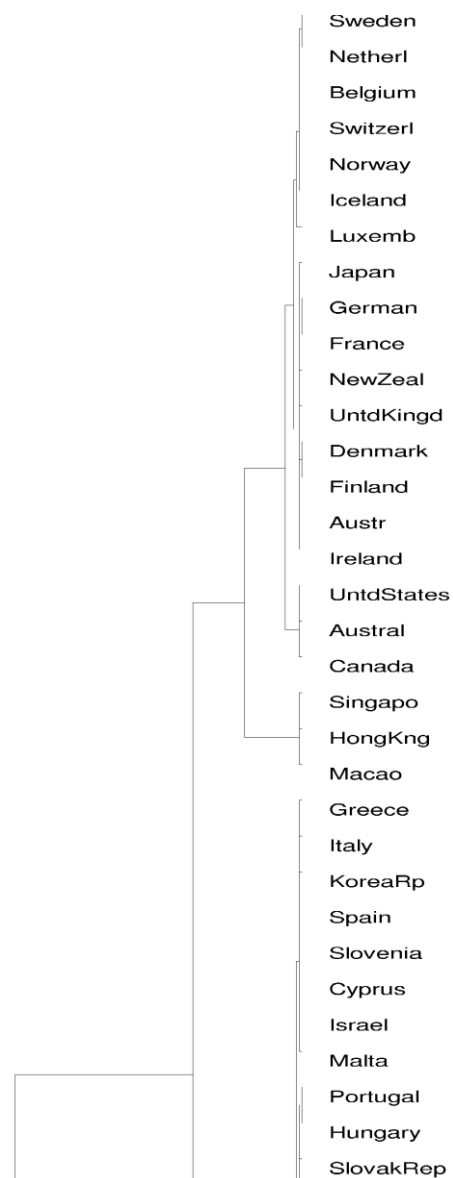
LAND and PRODUCTION	Mean	Q1	Median	Q3	Standard Deviation	Missing	%
Agricultural land (% of land area)	40,47	22,39	41,46	57,97	22,58	1	0,57
Arable land (% of land area)	15,07	4,67	10,67	21,93	13,88	2	1,14
Forest area (% of land area)	30,04	10,1	30,49	44,72	22,48	4	2,27
Permanent crop land (% of land area)	3,02	0,28	1,09	3,66	4,54	6	3,41
Arable land (hectares per person)	0,23	0,09	0,16	0,28	0,24	2	1,14
Cereal yield (kg per hectare)	3534,31	1621,31	2686,98	4239,42	5852,61	8	4,55
Crop production index (2004-2006=100)	110,93	96,67	108,11	124,01	22,13	4	2,27
Food production index (2004-2006=100)	112,33	101,73	111,13	123,42	16,43	4	2,27
Livestock production index (2004-2006=100)	112,92	102,62	112,68	121,24	15,23	4	2,27

REFUGEES	Mean	Q1	Median	Q3	Standard Deviation	Missing	%
Refugee by destination country (Refugee/tot local pop.)x100	0,83	0,17	0,34	0,64	4,02	0	0,00
Refugee by origin country (refugee/pop. origin country)x100	0,39	0	0,01	0,1	1,42	2	1,14
Growth rate refugee by origin country (yearly)x100 base=2009	3,98	-3,68	1,49	10,13	17,89	5	2,84
Internal Displaced people (% Total Pop.)	0,42	0	0	0	1,65	0	0,00
Refugee in camps per 100.000 inhab.	81,95	0	0	0	408,92	0	0,00

HEALTH	Mean	Q1	Median	Q3	Standard Deviation	Missing	%
Depth of the food deficit (kilocalories per person per day)	95,22	10	48,5	143	117,50	18	10,23
Malnutrition prevalence, weight for age (% of children under 5)	11,39	3,4	11,7	16,2	8,16	165	93,75
Prevalence of undernourishment (% of population)	17,93	5,25	12,45	28,6	14,36	64	36,36
Measles (MCV) immunization coverage among 1-year-olds (%)	87,66	80	93,5	97	12,99	6	3,41
Number of people (all ages) living with HIV per 1000 pop.	12,87	1,1	3,07	8,07	30,62	38	21,59
Number of women aged 15 and over living with HIV per 1000 women aged 15 and over	21	1,01	2,91	15,35	47,58	48	27,27
Deaths due to HIV/AIDS (per 100.000 population)	53,6	1,25	11	53	106,82	40	22,73
Prevalence of tuberculosis (per 100.000 population)	175,57	24	91,5	251	215,04	6	3,41
Deaths due to malaria (per 100.000 population)	20,66	0	0	11	42,91	30	17,05
Improved sanitation facilities (% of population with access)	71	45,4	83,4	98,6	30,57	15	8,52
Improved water source (% of population with access)	86,34	78,2	93,4	99,5	16,25	13	7,39
Total density per 100.000 population: Provincial hospitals	0,62	0,12	0,33	0,65	0,93	70	39,77

A.4 Grafico albero classificazione

Nonostante il grafico dell'albero di classificazione non possa essere rappresentato agevolmente per il gran numero di Paesi coinvolto, abbiamo riportato di seguito le parti che lo compongono per poter apprezzare il modo con cui i Paesi a mano a mano si saldano assieme nei gruppi. Ciò consente, allargando il numero di gruppi, rispetto ai 9 prescelti, di raffinare ulteriormente l'analisi. Si coglie immediatamente ad esempio che laddove Paesi africani e asiatici si ritrovino nello stesso cluster, quelli geograficamente vicini si aggregano per primi.



Cuba
BosniaHe
Georgia
Moldova
Belarus
Armen
Mongolia
Thailand
Iran
Alger
SaudiArab
Libya
Jordan
Uzbekist
Turkmenist
Kyrgyz
Tajikist
Egypt
Morocco
Syria
Iraq
Tunisia
Leban
Turkey
Azerbaij
SriLanka
Alban
Maldives
WBankGaza
Brunei
Malaysia
Vietnam
KoreaDR
Kuwait
Bahrain
Oman
Qatar
UntArabEm
Venez
Colomb
Brazil
Mexico
Peru
China
Kazakhst
TrinidadTob
Belize
Jamaic

Dominican Rep
Ecuador
Paraguay
Hondur
Philippines
Bolivia
Nicarag
Indonesia
Guatem
Gabon
Ghana
Lesotho
Swazil
Zimbabwe
SthAfrica
Botsw
Namibia
Chad
Angol
SierraLeo
CntrAfrRep
Nigeria
Camer
Benin
BurkinaF
Ethiopia
Guinea
GuineaBiss
Liberia
Niger
Mali
Kenya
CongoRp
Cotelvoir
Tanzan
Uganda
SthSudan
Togo
Madagas
CongoDR
Malawi
Mozamb
Zambia
Burundi
EqGuinea
PapuaNwGui
Haiti
Rwanda
TimorL
Solomon
Bhutan
Cambod
Myanmar
LaoPDR
Banglad



Comoros
Yemen
Djibouti
Mauritania
India
Nepal
Pakistan
Eritrea
Sudan
Senegal
Gambia
Somalia
Afghan

A.5 Timeline crisi siriana

(The Cairo Review of Global Affairs, American University, il Cairo)

- March 6, 2011: In the southern town of Daraa, police arrest some fifteen young boys for spray-painting buildings with the slogan, "The people want the downfall of the regime."
- March 15, 2011: On Day of Rage, small protests against the regime of President Bashar Al-Assad take place in Damascus and Aleppo.
- March 18, 2011: Anti-regime demonstrations take place in Damascus, Homs, Baniyas, and Daraa; security forces shoot and kill six protesters in Daraa.
- March 23, 2011: Security forces kill dozens of protesters in Daraa in attempt to end growing uprising.
- March 24, 2011: Al-Assad advisor Butheina Shaaban announces measures and reforms including government salary increases, licensing of new political parties, and studying a possible end to emergency law.
- March 29, 2011: Large pro-regime demonstrations are held in Damascus and elsewhere in Syria.
- March 30, 2011: Bashar Al-Assad addresses parliament in first speech since start of protests; blames foreign conspirators and satellite channels for unrest; promises reforms and dismisses government of Prime Minister Mohammed Al-Otari (1944-); Adel Safar (1953-), former agriculture minister, is later appointed to form a new cabinet.
- April 6, 2011: Bashar Al-Assad announces the granting of Syrian citizenship to tens of thousands of Kurds.
- April 12, 2011: Security forces with tanks and heavy weapons begin mobilizing in Syrian cities to quell spreading unrest.
- April 16, 2011: In a televised address, Bashar Al-Assad announces the repeal of emergency law and the granting of permission to hold demonstrations under certain conditions.
- April 22, 2011: Bloody Friday protests leave more than 120 people dead in Daraa, Damascus and elsewhere in Syria; President Barack Obama and UN Secretary-General Ban Ki-moon issue statements condemning the regime's crackdown on demonstrators.
- April 29, 2011: Syrian refugees begin crossing border into Turkey; first of twenty-one Syrian refugee camps is established in the country; President Obama issues U.S. Executive Order 13572 blocking property of three Syrian officials because of human rights abuses, including Maher Al-Assad (1968-), the Syrian president's younger brother, who is commander of the Republican Guard and the army Fourth Armored Division.
- May 9, 2011: European Union imposes arms embargo on Syrian regime and freezes assets of and applies travel ban on senior regime officials.
- May 18, 2011: U.S. Executive Order 13573 blocks U.S. property owned by Bashar Al-Assad and six other senior Syrian officials.
- June 3, 2011: Security forces reportedly kill thirty-four protesters in Hama during the largest anti-regime demonstration since the unrest began.
- June 6, 2011: Syrian government reports that more than 120 soldiers were killed by "armed gangs" in Jisr Al-Shughour; some Western media reports suggest that residents took up arms to repel assault by regime security forces.
- June 14, 2011: Arab League condemns regime crackdown on protesters; Secretary-General Amr Moussa says Arab states are "worried, angry and actively monitoring" the crisis.
- June 17, 2011: Syrian conflict spreads to Lebanon with the outbreak of clashes in Tripoli between Lebanese supporters and opponents of the Syrian regime.
- July 8, 2011: Tens of thousands of Syrians, many carrying roses and olive branches, stage a demonstration in Hama.
- July 10, 2011: Syrian government holds a two-day national dialogue; opposition factions boycott the meeting.
- July 11, 2011: Secretary of State Hillary Clinton says that Bashar Al-Assad has "lost legitimacy" and that "we have absolutely nothing invested in him remaining in power."
- July 29, 2011: Defectors from the Syrian army announce the formation of the Free Syrian Army, led by former Colonel Riad Al-Asaad (1961-).
- July 31, 2011: Security forces reportedly kill more than 100 protesters in Hama.
- August 3, 2011: UN Security Council issues a statement condemning "the widespread violations of human rights and the use of force against civilians by the Syrian authorities."
- August 8, 2011: Saudi Arabian King Abdullah bin Abdulaziz Al-Saud condemns Bashar Al-Assad's "killing machine"; Saudi Arabia, Kuwait and Bahrain recall their ambassadors from Syria.

- August 17, 2011: U.S. Executive Order 13582 bans Syrian oil imports and new U.S. investments in Syria, and blocks Syrian government property in the United States.
- August 18, 2011: Leaders of the United States, France, Germany and Britain issue coordinated statements calling on Bashar Al-Assad to resign.
- August 23, 2011: Syrian factions inside and outside of Syria form the Syrian National Council to oppose the Al-Assad regime.
- September 2, 2011: European Union bans imports of Syrian oil.
- September 21, 2011: Turkish Prime Minister Recep Tayyip Erdoğan says that his government has cut off contacts with Syria.
- September 27, 2011: Security forces and military defectors engage in heavy fighting in the city of Rastan.
- October 4, 2011: Russia and China veto draft UN Security Council (UNSC) resolution that would condemn “grave and systematic” human rights violations in Syria.
- October 14, 2011: UN High Commissioner for Human Rights Navi Pillay warns that the Syrian crisis “is already showing worrying signs of descending into an armed struggle.”
- October 29, 2011: Security forces shell Homs after clashes between Al-Assad loyalists and military defectors in the city.
- November 2, 2011: Arab League announces that the Syrian government has accepted an Arab League proposal to immediately halt violence, free prisoners and open dialogue with the opposition.
- November 12, 2011: Arab League votes to suspend Syria from membership and impose political and economic sanctions until the Al-Assad regime adheres to Arab League peace plan.
- November 27, 2011: Arab League votes sanctions on Syria, terminating transactions with the Syrian Central Bank, imposing a travel ban on Syrian officials, suspending commercial flights between Syria and Arab countries, and freezing assets related to the Al-Assad regime.
- December 7, 2011: Bashar Al-Assad tells ABC News: “I did my best to protect the people, so I cannot feel guilty.”
- December 12, 2011: UN High Commissioner for Human Rights Navi Pillay says more than 5,000 people have been killed in the Syrian conflict.
- December 19, 2011: Syria signs new Arab League agreement requiring the regime to withdraw security forces and heavy weapons from civilian areas, commence talks with the opposition, and allow human rights workers and journalists into the country.
- December 19-20, 2011: Security forces reportedly kill nearly 200 people including army defectors in a massacre in Idlib province.
- December 23, 2011: Two suicide bombings kill at least forty people outside the State Security Directorate and another security building in Damascus.
- December 26, 2011: Arab League representatives arrive in Syria to monitor Al-Assad regime’s compliance with Arab League agreement.
- December 28, 2011: Syrian regime announces release of more than 750 prisoners detained during protests.
- December 30, 2011: Large anti-regime protests take place across the country.