

ALGORITMI DI REGIONALIZZAZIONE BASATI SUI FLUSSI DI PENDOLARISMO:
ANALISI E CONFRONTI

Luisa Franconi ¹, Michele D'Alò ²

SOMMARIO

L'obiettivo del presente lavoro è quello di illustrare i risultati preliminari delle analisi condotte dall'ISTAT per l'individuazione dei sistemi locali del lavoro (SLL) 2011. Lo scopo è quello di definire il massimo numero di aree, ottenute come aggregazioni di comuni, che soddisfano opportuni criteri statistici. A tal fine, sono stati analizzati alcuni algoritmi di definizione dei SLL utilizzando i dati del censimento 2001. L'attività svolta finora ha compreso l'analisi di alcuni metodi adottati da altri istituti nazionali di statistica; del metodo, denominato EURO, proposto da Eurostat con l'obiettivo di individuare SLL armonizzati a livello europeo; lo studio di nuove ed avanzate metodologie che adottano approcci probabilistici. La costruzione dei SLL avviene cercando di massimizzare il livello di interazione tra comuni appartenenti allo stesso SLL, espressa dai flussi di pendolarismo giornaliero tra luogo di residenza e luogo di lavoro – rilevati dall'Istat in occasione dei censimenti della popolazione e delle abitazioni. Tutti gli algoritmi implementati danno luogo a risultati differenti rispetto ai SLL 2001. L'analisi si è concentrata sull'applicazione di due varianti del metodo TTWA, i quali, a differenza dell'algoritmo applicato dall'Istat per l'individuazione dei SLL 2001, consentono di soddisfare il vincolo di auto-contenimento minimo.

¹ Istat, Direzione centrale per lo sviluppo dei sistemi informativi e dei prodotti integrati, la gestione del patrimonio informativo e la valutazione della qualità, via A. Depretis, 77, 00184, Roma, e-mail: franconi@istat.it.

² Istat, Dipartimento per l'integrazione, la qualità e lo sviluppo delle reti di produzione e ricerca, via C. Balbo, 16, 00184, Roma, e-mail: dalo @istat.it.

1. Introduzione

I Sistemi locali del lavoro (SLL), nell'accezione proposta dall'Istat fin dal 1981, rappresentano dei luoghi (precisamente identificati e simultaneamente delimitati su tutto il territorio nazionale³) dove la popolazione risiede e lavora e dove quindi indirettamente tende ad esercitare la maggior parte delle proprie relazioni sociali ed economiche. Da un punto di vista tecnico e metodologico i SLL sono costruiti come aggregazione di due o più comuni cercando di massimizzare il livello di interazione tra comuni appartenenti allo stesso SLL, espressa dai flussi di pendolarismo giornaliero tra luogo di residenza e luogo di lavoro – rilevati dall'Istat in occasione dei censimenti della popolazione e delle abitazioni. L'individuazione dei SLL dipende anche dalla soddisfazione di vincoli imposti sul livello di auto-contenimento e sulla loro dimensione.

La necessità di individuare aree territoriali che esulino dalle tradizioni suddivisioni amministrative del territorio è legata all'obiettivo di identificare ed analizzare caratteristiche economiche e sociali di aree specifiche che dipendono dai processi di auto organizzazione della popolazione attiva, misurati mediante i movimenti giornalieri che i singoli individui operano per conciliare l'attività lavorativa con quella sociale.

Gli strumenti spesso utilizzati dagli Istituti di Statistica per la definizione dei SLL sono basati su un algoritmo di regionalizzazione proposto da Coombes et al. (1986) denominato TTWA (Travel-To-Work Areas).

L'obiettivo del presente lavoro è quello di fornire i risultati preliminari delle analisi condotte dall'ISTAT per la predisposizione dei SLL 2011. L'attività svolta finora ha compreso l'analisi e la sperimentazione dei metodi adottati da altri istituti nazionali di statistica per la determinazione di mercati locali del lavoro o da Eurostat per la definizione di sistemi locali del lavoro armonizzati a livello europeo, (metodo denominato EURO), ma anche nello studio di nuove metodologie che adottano approcci considerati di interesse.

Una panoramica dei metodi utilizzati in Europa è fornita dal rapporto di ricerca predisposto per Eurostat da Coombes et al. (2012). Gli algoritmi ivi presentati seguono principi comuni relativamente a:

1. Scopo: aree appropriate per definire policy;
2. Rilevanza: la soglia minima in termini di occupati della zona permette di diffondere una informazione statistica affidabile in particolare relativamente al tasso di disoccupazione locale e alle stime dell'occupazione;
3. Completezza: le zone sono una partizione della superficie totale dello stato;

³

La totale copertura nazionale dei SLL rappresenta una delle più importanti caratteristiche di questa partizione territoriale.

4. Partizione: ciascun comune può appartenere ad una sola zona;
5. Contiguità: i comuni di ciascuna zona sono contigui;
6. Coerenza: ciascuna zona è costituita da un insieme di comuni non frazionati;
7. Conformità: le zone possono non rispettare i confini amministrativi;
8. Omogeneità: le zone non devono essere troppo estese territorialmente o troppo numerose in termini di occupati.

La maggioranza dei paesi che finora hanno predisposto SLL hanno utilizzato algoritmi che sono varianti del metodo proposto Coombes et al. (1986). Tale approccio è stato seguito anche dall'Istat (Istat e IRPET, 1989 e Istat, 1997) nelle precedenti occasioni. Con l'obiettivo di predisporre i nuovi SLL è stata avviata una collaborazione tra l'Istat ed alcuni altri istituti di statistica formalizzata nell'ambito di una task force di Eurostat con la finalità di condividere metodi, strumenti, analisi e problematiche. Gli algoritmi considerati sono quelli utilizzati da altri paesi, in particolare Francia, Gran Bretagna e Paesi Bassi.

Il presente lavoro si articola come segue: nel paragrafo 2 si descrive il metodo utilizzato dall'Insee (Francia) per la determinazione delle zones d'emploi 2010 e si illustrano i risultati ottenuti dalla sperimentazione effettuata con i dati di pendolarismo relativi al Censimento Italiano del 2001. Nel paragrafo 3 si descrive brevemente l'algoritmo descritto in Coombes e Bond (2007) denominato EURO, adottato da ONS e proposto in ambito europeo da Eurostat per la definizione di SLL che possano essere armonizzati a livello di Unione Europea. Nel paragrafo 4 si illustra una variazione del algoritmo EURO proposta invece dal Istituto di statistica olandese CBS. Nel paragrafo 5 si riportano i risultati dell'applicazione dell'algoritmo, implemento in R, che implementa le due versioni del metodo TTWA considerate. Nel paragrafo 6 sono infine riportate le conclusioni di queste prime analisi oltre ad una breve rassegna di alcuni metodi alternativi all'approccio proposto da Coombes.

Al fine di descrivere i metodi di regionalizzazione basati sui flussi di pendolarismo è utile introdurre la seguente notazione ed utilizzata nel resto del documento. Si indicherà con f_{hk} il flusso di occupati che risiedono nella località h e lavorano nella località k ; con $R_h = \sum_k f_{hk} = f_{h\cdot}$ gli occupati che risiedono nella località h (Residents di h); con $W_k = \sum_h f_{hk} = f_{\cdot k}$ gli occupati che lavorano nella località k (Workers in k) e con $RW_k = f_{kk}$ gli occupati che risiedono e lavorano nella località k . Le due funzioni chiave nella definizione dei sistemi locali del lavoro sono le funzioni di auto-contenimento dal lato dell'offerta di posti di lavoro: $SCO = RW_k / R_k = f_{kk} / f_{k\cdot}$ e dal lato della domanda di posti di lavoro: $SCA = RW_k / W_k = f_{kk} / f_{\cdot k}$.

2. L'algoritmo Anabel per la definizione delle zone d'emploi

La *zone d'emploi* è definita come lo spazio geografico all'interno del quale la maggior parte della popolazione attiva vive e lavora e nel quale i "soggetti economici" trovano la maggior

parte della mano d'opera a loro necessaria⁴. Il metodo per definire le *zone d'emploi* è fondato unicamente sul criterio degli spostamenti tra il comune di domicilio e il comune di lavoro⁵ e ha l'obiettivo di massimizzare il tasso di stabilità ovvero il numero di occupati che vivono e lavorano nella zona rapportato al numero di occupati che vivono e lavorano nella zona o al di fuori di essa (totale degli occupati residenti). L'algoritmo deterministico utilizzato dall'Insee per aggiornare i sistemi locali del lavoro (*zone d'emploi*) nel 2010 (Insee, Dares, Datar, 2012) è stato implementato nell'applicativo Anabel (Durieux, 2012); una versione semplificata di questo, ttwa (Semecurbe e Timoteo, 2013), è disponibile sul sito CRAN di R: <http://cran.r-project.org/>; tale algoritmo è stato utilizzato dal GdL per effettuare delle sperimentazioni sulla base dei dati del censimento della popolazione del 2001.

Alla fase di individuazione delle aree tramite il software segue una fase di validazione regionale (tre mesi) che, sotto la guida dei Prefetti delle regioni e dello sportello unico per le aziende (DIRECCTE), vede impegnati tutti gli attori locali nella validazione delle *zone d'emploi*.

L'algoritmo Insee è un algoritmo di classificazione gerarchico di tipo agglomerativo caratterizzato dall'impossibilità di separare gruppi formati al passo precedente e basato sulla definizione di regole di arresto per la definizione della soluzione ottimale. I parametri per definire i criteri di arresto sono riportati nella Tabella 1.

La similarità tra due località, ovvero il criterio secondo il quale aggregare due località, è definita come la proporzione di occupati residenti in h che si recano al lavoro in k (*supply side dependence*):

$$\text{Link}(h,k) = \mathbf{f}_{hk} / \mathbf{f}_h = \mathbf{f}_{hk} / \mathbf{R}_h \quad (1)$$

Tabella 1: Parametri per definire le regole di arresto dell'algoritmo Anabel.

| Parametro | Definizione |
|--------------------------|---|
| Legame minimo (min_Link) | Soglia sul valore della funzione da massimizzare al di sotto della quale due località non possono essere aggregate. |
| Isolamento parziale | Soglia sul totale degli occupati residenti di una località/zona al di sopra della quale essa non può essere satellite di (ovvero essere aggregata a) un'altra località/zona. |
| Isolamento totale | Soglia sul totale degli occupati residenti di una località/zona al di sopra della quale essa non può più evolvere: essa quindi non potrà essere né satellite di un'altra località/zona, né accogliere per aggregazione una qualsiasi altra località. Scopo: inibire la creazione di poli di grandi dimensioni. |

⁴ “Une zone d'emploi est un espace géographique à l'intérieur duquel la plupart des *actifs* résident et travaillent, et dans lequel les établissements peuvent trouver l'essentiel de la main d'œuvre nécessaire pour occuper les emplois offerts.” Insee, <http://www.insee.fr/fr/methodes/default.asp?page=definitions/zone-emploi.htm>.

⁵ Idem, pag. 53

I passi dell'algoritmo iterativo sono i seguenti:

1. Al primo passo si considerano tutte le località come sistemi locali;
2. Si individuano le località/sistemi locali h e k che presentano il valore massimo della funzione Link (nel caso di parità si sceglie la prima occorrenza), dove h è tale che: $R_h < \text{min_Size}$ (*soglia di isolamento parziale*);
3. Si verificano le regole di arresto (ed eventualmente la contiguità tra aree):
 - a. se $\text{link}(h,k) > \text{min_Link}$
 - b. se $R_h < \text{soglia di isolamento totale}$;
 - c. se h e k sono adiacenti (nel caso in cui si sia presente una matrice di contiguità);

le località h e k sono aggregate. Le matrici dei flussi e delle contiguità sono ricalcolate e si torna al passo 2., altrimenti l'algoritmo si ferma.

A queste tre fasi principali se ne aggiunge una quarta opzionale detta di pulizia. In tale fase l'applicativo Anabel permette di riassegnare i comuni isolati a sistemi locali già sviluppati.

2.1. Analisi dei risultati

I risultati delle analisi condotte sui dati del Censimento della popolazione del 2001 si sono rivelati molto deludenti e l'approfondimento con i tecnici francesi che si sono occupati dello sviluppo ha messo in luce alcuni aspetti considerati non desiderabili per l'implementazione italiana:

- 1) SLL unitari: A causa della soglia sulla popolazione residente l'algoritmo tende a creare molti SLL costituiti da un solo comune;
- 2) Enclaves: il metodo tende a formare SLL all'interno di SLL;
- 3) Limiti amministrativi: sono imposte delle condizioni restrittive per superare i confini regionali;
- 4) Aree metropolitane: sono trattate separatamente; Mas, S. (2013).
- 5) Il metodo viene applicato distintamente su ciascuna regione con parametri appropriati ai flussi presenti, con un controllo nazionale in caso di dubbi; Semecurbe, F. (2014).

Come si può evidenziare dalle Figure 1 e 2, l'applicazione sperimentale del metodo Anabel ai dati del 2001 ha confermato la non applicabilità del metodo al contesto italiano. Il risultato, oltre ad essere molto distante da quello a suo tempo ottenuto, presenta un numero molto elevato di enclaves, soprattutto in prossimità delle grandi aree urbane (vedi zoom su Milano e Torino).

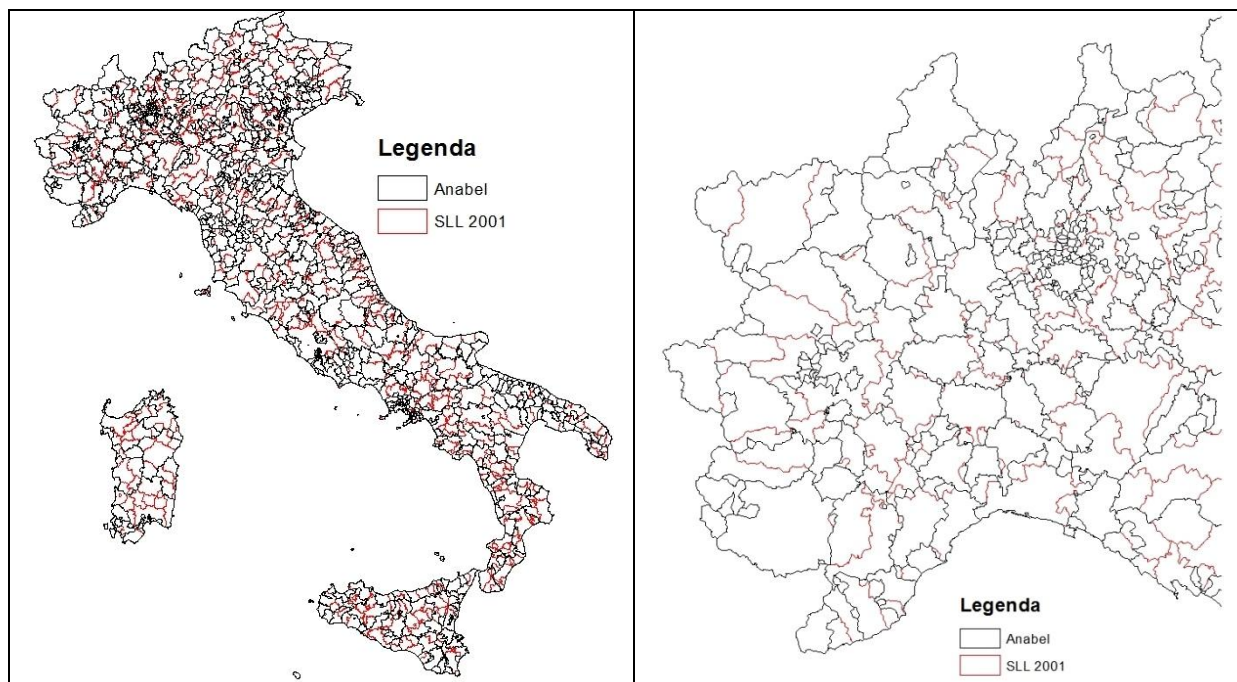


Figura 1 – Confronto tra l'applicazione dell'algoritmo Anabel su dati 2001 e i SLL Istat 2001

Inoltre questo metodo non garantisce sulla costruzione di aree che soddisfino livelli minimi di auto-contenimento, sia dal lato della domanda che dell'offerta. Questo risultato è ben illustrato nella successiva Figura 2 che confronta gli indici di autocontenimento dei due approcci. Inoltre il metodo Anabel tende a creare molti sistemi locali costituiti da un solo comune.

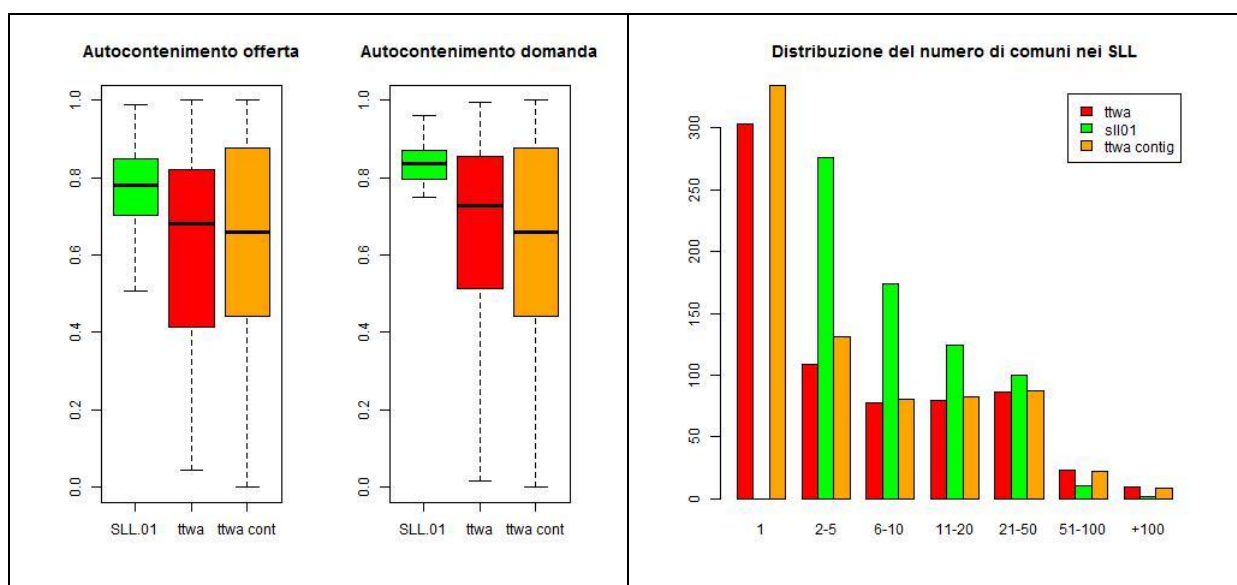


Figura 2 Boxplot autocontenimento e distribuzione dei comuni nei sistemi locali⁶

⁶ ttwa cont” e “ttwa” sono i risultati dell'applicazione di Anabel con e senza vincolo di contiguità tra i comuni

3. EURO – metodo ONS

Nel 2007 l'istituto nazionale di statistica britannico rivede il dettaglio geografico delle località di definizione dei flussi, non ricorre al campionamento dei dati e semplifica la metodologia "classica" (Coombes e Bond, 2007) per la predisposizione dei propri sistemi locali del lavoro, definiti "Travel-to work-areas" (TTWA). I criteri per la definizione dei TTWA sono così riassunti nel sito dell'Istituto :

- almeno il 75% della forza lavoro dell'area lavora nell'area stessa e almeno il 75% degli occupati che vivono nell'area lavorano nell'area stessa;
- l'area deve avere una dimensione minima di 3,500 occupati;
- per aree con una popolazione di occupati maggiore di 25,000 è accettato un livello di auto-contenimento (offerta e domanda) di 66.7%;
- per una necessità di consistenza Coombes e Bond (2007) applicano un unico set di parametri per l'intero paese (e non ottimizzati su base regionale come Anabel).

Questa metodologia è proposta da Eurostat per la creazione di sistemi locali del lavoro armonizzati a livello europeo e denominata EURO.

L'idea alla base dell'algoritmo EURO è quella di eliminare la soglia unica sulle funzioni di auto-contenimento e sul numero di posti di lavoro tramite la definizione di un trade-off tra posti di lavoro e auto-contenimento. A fronte di valori elevati delle funzioni di auto-contenimento (superiori ad soglia minima prefissata) si accettano sistemi locali del lavoro di dimensioni ridotte (numero di occupati residenti superiore ad una soglia minima prefissata), mentre per sistemi locali del lavoro di dimensioni maggiori (con un numero di occupati residenti superiore ad un valore target prefissato) la soglia di accettazione dell'auto-contenimento diminuisce (soglia target dell'auto-contenimento). L'algoritmo dipende quindi da quattro parametri: un valore minimo e un valore target sia per il numero di occupati residenti ($size_{Min}$ e $size_{Target}$) sia per le funzioni di auto-contenimento (sc_{Min} e sc_{Target}).

Pertanto, i risultati ottenuti devono soddisfare due proprietà legate all'auto-contenimento e al numero di persone economicamente attive residenti. Per ognuno di questi due requisiti devono essere definiti un valore minimo ed uno target pertanto l'algoritmo dipende da quattro parametri dati da

1. Autocontenimento minimo ($minSC$);
2. Autocontenimento target ($tarSC$);
3. Dimensione minima ($minSZ$);
4. Dimensione target ($tarSZ$).

Inoltre, l'auto-contenimento è misurato sia dal lato della domanda ed è definito come il rapporto tra il numero di persone che vivono e lavorano all'interno dell'area ed il numero di lavoratori, sia dal lato dell'offerta, misurato come il rapporto tra il numero di persone che lavorano e vivono e il numero di lavoratori residenti.

Il metodo permette di raggruppare i comuni mediante la seguente funzione:

$$\frac{\min SC}{\text{tar} SC} \leq \left(1 - \left(1 - \frac{\min SC}{\text{tar} SC} \right) \cdot \text{MAX} \left(\frac{\text{tar} SZ - SZ}{\text{tar} SZ - \min SZ}, 0 \right) \right) \cdot \left(\frac{\text{MIN}(SC, \text{tar} SC)}{\text{tar} SC} \right)$$

Pertanto un raggruppamento di comuni possono costituire un SLL soltanto se la condizione precedente è soddisfatta.

Operativamente, l'algoritmo considera inizialmente ciascun comune come un proto-sistema locale. Per ciascun proto-sistema locale h calcola l'indice *validity* che esprime il trade-off tra i valori del numero di occupati che risiedono nel proto-sistema e delle funzioni di auto-contenimento nel proto-sistema stesso in relazione alle soglie minime e target definite in input. In particolare, indicando con $\text{msc}_h = \min(RW_h / R_h, RW_h / W_h)$, si ha:

$$\begin{aligned} \text{validity} &= 1 \text{ se } R_h \geq \text{size}_{\text{Target}} \text{ e } \text{msc}_h \geq \text{sc}_{\text{Target}} \\ \text{validity} &= 1 \text{ se } \text{size}_{\text{Target}} > R_h \geq \text{size}_{\text{Min}} \text{ e } \text{msc}_h \geq m * R_h + n \\ &\quad \text{dove } m = (\text{sc}_{\text{Target}} - \text{sc}_{\text{Min}}) / (\text{size}_{\text{Min}} - \text{size}_{\text{Target}}) \\ &\quad \text{e } n = (\text{sc}_{\text{Min}} * \text{size}_{\text{Min}} - \text{sc}_{\text{Target}} * \text{size}_{\text{Target}}) / (\text{size}_{\text{Min}} - \text{size}_{\text{Target}}); \\ &\quad \text{altrimenti} \\ \text{validity} &= \min(R_h / \text{size}_{\text{Target}}, 1) * \min(\text{msc}_h / \text{sc}_{\text{Target}}, 1). \end{aligned}$$

In Coombes and Bond (2007) gli autori delineano il seguente algoritmo:

- Passo 1: ordinare in ordine decrescente tutti i proto-sistemi locali secondo l'indice *validity*;
- Passo 2: se $\min(\text{validity}) = 1$ l'algoritmo si arresta, altrimenti proseguire con il Passo 3;
- Passo 3: sciogliere nelle singole località costituenti il proto-sistema locale h che nel passo 2 ha ottenuto il minimo dell'indice *validity*;
- Passo 4: allocare le singole località costituenti h del passo 3 al proto-sistema locale k che massimizza la seguente funzione coesione o interazione:
$$L_{hk} = [(f_{hk})^2 / (f_h \cdot f_k)] + [(f_{kh})^2 / (f_k \cdot f_h)] \quad (1)^7$$
- Passo 5: aggiornare la lista dei proto-sistemi locali, ricalcolare il numero degli occupati residenti e le funzioni di auto-contenimento per il proto-sistema locale k ottenuto dall'aggregazione con h ;
- Passo 6: ritornare al Passo 1.

I risultati che si ottengono applicando tale algoritmo è schematizzato nella figura 3, da cui si evince che i SLL più piccoli devono avere un auto-contenimento maggiore (pari a quello target), mentre all'aumentare della loro dimensione il livello di auto-contenimento può ridursi

⁷

Per convenzione, se la funzione restituisce una qualsiasi forma indeterminata questa viene posta pari a 0.

fino a quando raggiunta la dimensione target il livello di auto-contenimento può essere pari al minimo fissato nell'algoritmo.

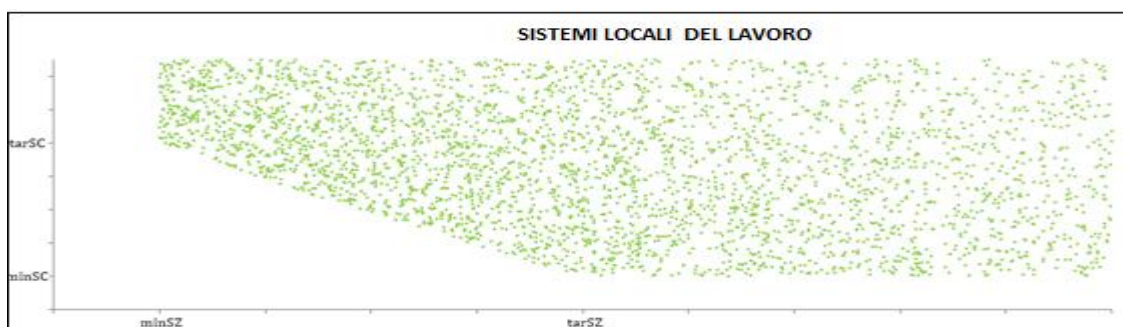


Figura 3. Schematizzazione del metodo TTWA-ONS

4. Variazione di EURO utilizzata da CBS

Le analisi effettuate da CBS utilizzando l'algoritmo di Coombes e Bond (2007) non hanno portato a risultati soddisfacenti con i dati dei flussi di pendolarismo olandesi. Per ovviare ai problemi riscontrati l'Istituto ha modificato l'area di accettazione nel piano cartesiano definito dal numero di posti di lavoro e dall'auto-contenimento, che individua un'ellisse e si considera un unico livello di auto-contenimento. Tuttavia, l'algoritmo CBS ricalca i passi 1-5 dell'algoritmo EURO (sostituendo la funzione che deve essere rispettata affinché un raggruppamento di comuni possa essere definito SLL). L'indice in questo caso dipende da un unico parametro per l'auto contenimento (sc_{Min}) e da due parametri per l'ampiezza ($size_{Min}$ e $size_{Target}$). Dopo aver operato una standardizzazione delle ascisse

$$x_h = \min(size_h, size_{Target}) / \min(size_{tot}, size_{Target})$$

dove $size_{tot}$ è la somma del numero di residenti di tutti i SLL, in modo tale che entrambe le coordinate varino tra zero e uno, l'indice, r , di interesse in questo caso è definito come la distanza tra i fuochi dell'ellisse, ovvero:

$$r = \sqrt{\left\{ \left[\frac{(1 - x_h)}{1 - (size_{Min} / \min(size_{tot}, size_{Target}))} \right]^2 + \left[\frac{(1 - sc_h)}{(1 - sc_{Min})} \right]^2 \right\}}$$

Se $r \leq 1$ il SLL h è contenuto nell'ellisse ed è considerato come ammissibile rispetto ai parametri definiti. Se $r > 1$ occorre identificare un altro SLL k da unire al SLL h in esame.

Nella figura 4 si riporta la schematizzazione dei risultati attesi dall'applicazione dell'algoritmo TTWA versione CBS. Anche in questo caso l'auto-contenimento dei raggruppamenti più piccoli è più alto rispetto a quelli di dimensione maggiore.

Nella figura 5 è riportato il confronto tra i risultati ottenuti sulla base delle due funzioni utilizzati nelle due varianti del metodo TTWA. Nell'ipotesi che l'auto-contenimento fissato nell'algoritmo CBS sia 65% e il minimo e il massimo dell'auto-contenimento fissato per l'algoritmo ONS siano pari al 66,7% e 75%.

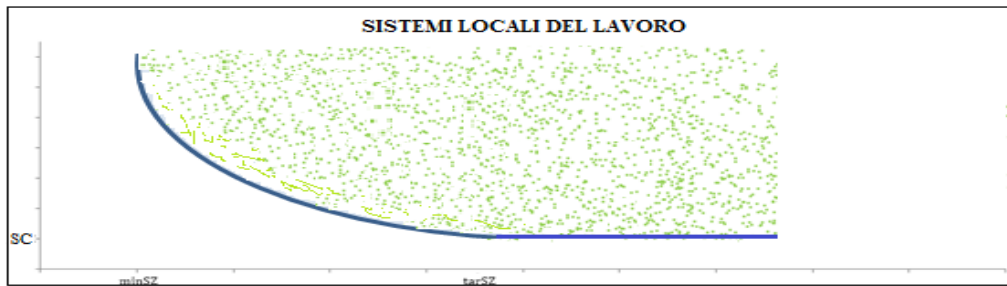


Figura 4. Schematizzazione del metodo TTWA-CBS

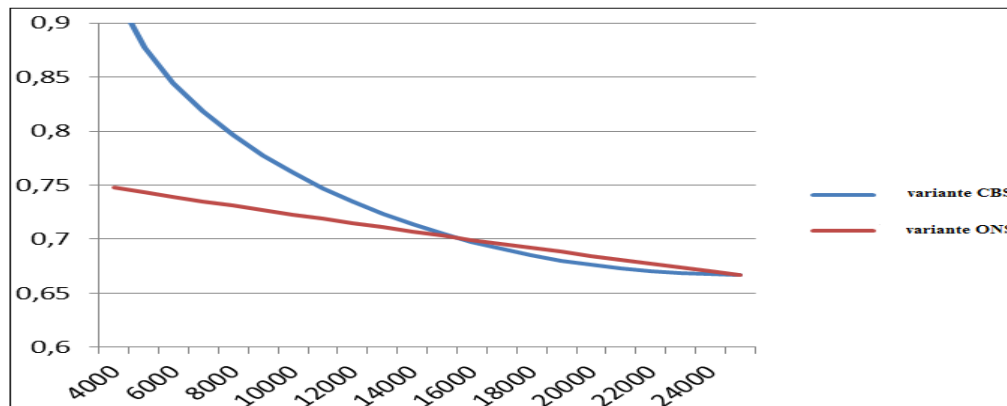


Figura 5. Confronto tra le due varianti del metodo TTWA

5. Analisi dei risultati

Le due varianti del metodo sono stati applicati ai dati di pendolarismo del 2001 con lo scopo di confrontare i risultati ottenuti con i SLL 2001 pubblicati dall'Istat. I parametri dell'algoritmo ONS e CBS sono stati fissati in modo da essere uguali a quelli definiti per l'individuazione dei SLL del 2001, ossia una dimensione minima pari a 1000 ed un auto-contenimento pari al 75%. L'algoritmo di implementazione scritto in R è stato fornito dall'Istituto di statistica olandese (CBS)⁸, è attualmente in fase di revisione ed analisi. Nelle figure seguenti si riportano i cartogramma relativi ai 684 SLL del 2001, ai 452 SLL ottenuti con l'algoritmo TTWA versione ONS ed ai 435 ottenuti con la variante CBS. Dall'analisi delle tre mappe risulta evidente che ha parità di parametri le due varianti del metodo TTWA danno luogo a SLL di maggiore ampiezza rispetto a quelli originali, soprattutto in corrispondenza dei grandi centri urbani. Evidente è la maggiore ampiezza del SLL di Roma.

⁸ Eurostat ha reso disponibile anche un analogo script scritto in linguaggio Java. Trattandosi però di un algoritmo computazionalmente meno efficiente, il GDL lo sta utilizzando come benchmark.

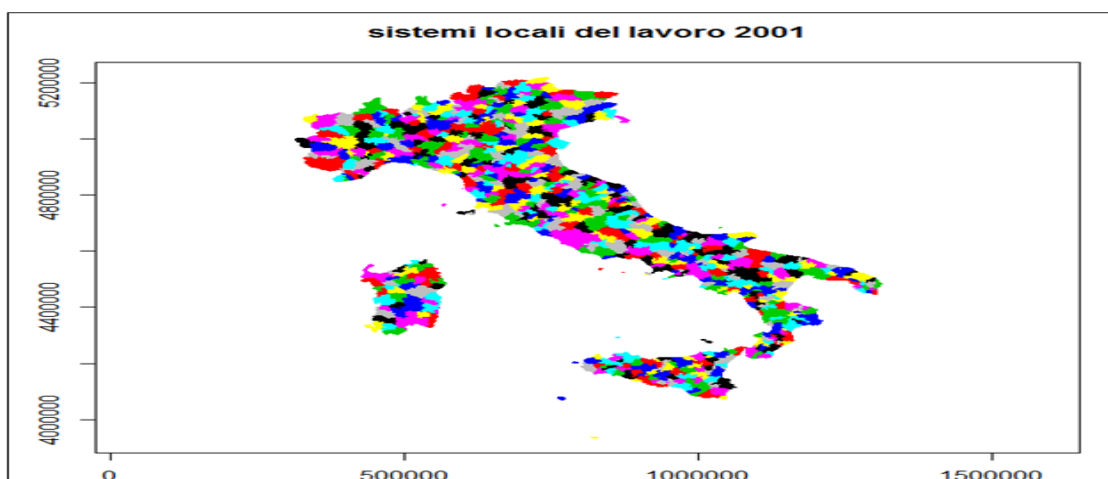


Figura 6. Cartogramma dei SLL 2001

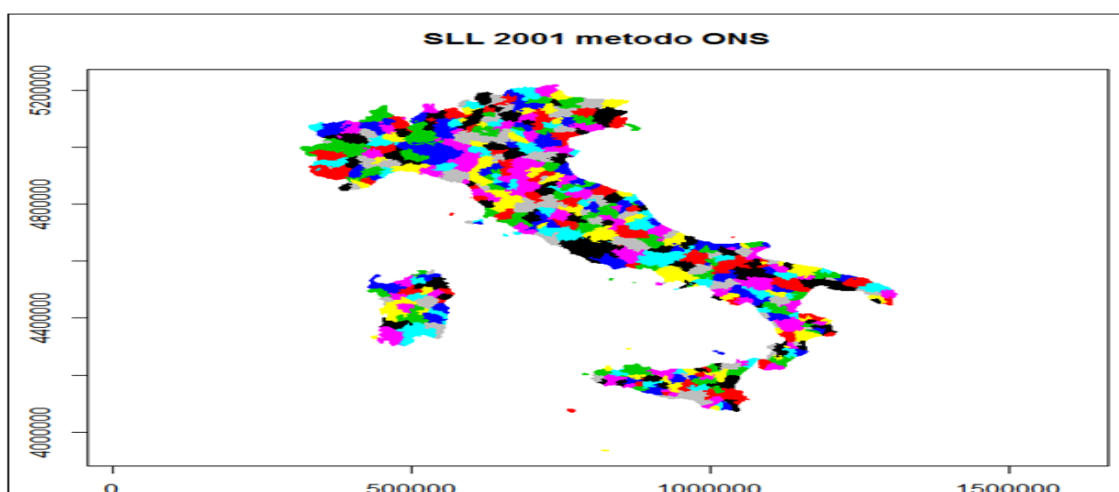


Figura 7. Cartogramma dei SLL 2001 metodo ONS

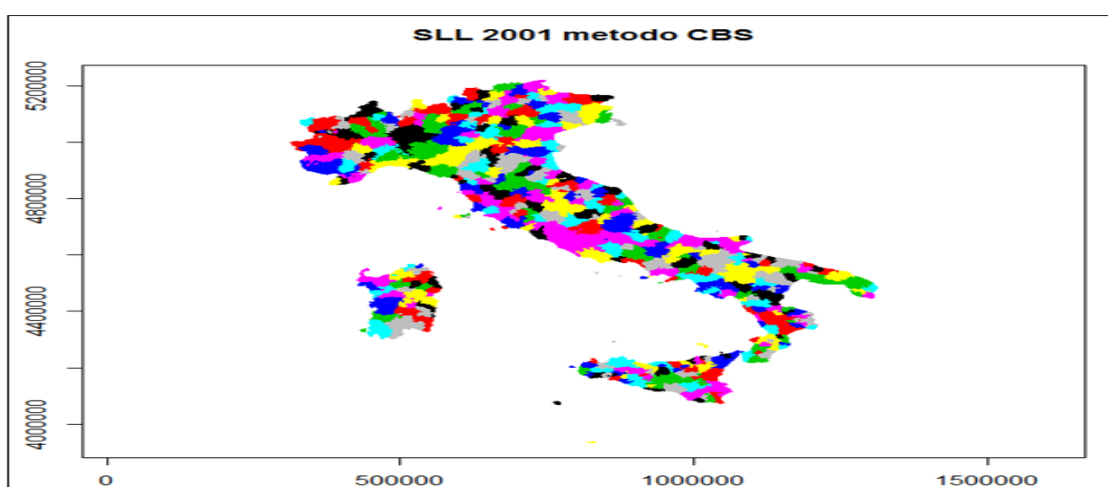


Figura 8. Cartogramma dei SLL 2001 metodo CBS

Quanto detto in precedenza si evidenzia anche nella figura 9 in cui si riporta l'ampiezza dei SLL in base al numero dei comuni. La conclusione è che l'algoritmo implementato nel 2001 ha generato un numero elevato di SLL molto piccoli, mentre le due varianti del metodo TTWA in fase di sperimentazione danno luogo ad una distribuzione di SLL più uniforme rispetto alla loro ampiezza espressa in termini di numero di comuni costituenti i SLL. Tale diversità ha ovviamente un notevole impatto sulla interpretazione dei risultati finali.

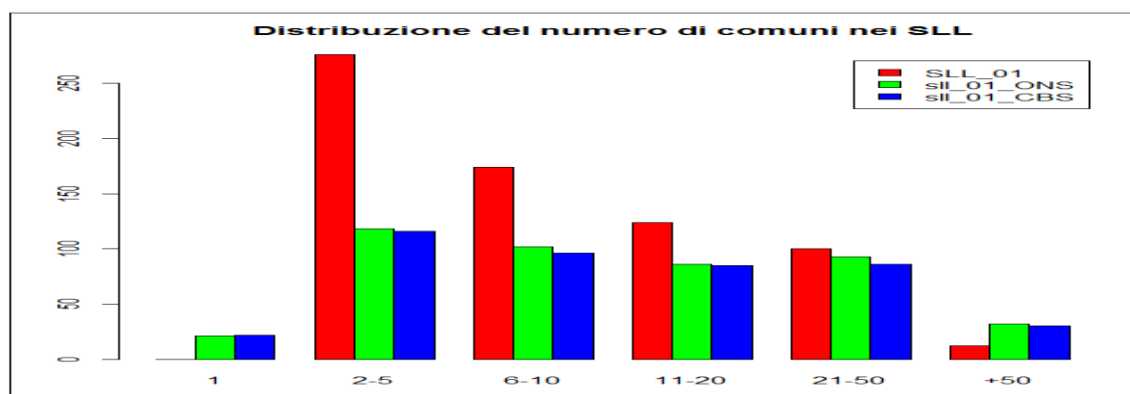


Figura 9. Ampiezza dei SLL in base al numero di comuni.

Nei cartogrammi successivo sono riportati nella stessa mappa i SLL locali ottenuti con il rispettivamente con il metodo ONS e quello CBS e sono identificate dalle aree colorate e messi in relazione con i SLL del 2001 individuati dai bordi. Per una lettura più agevole nella figura 11 si riporta un focus sulle differenze relative alle soluzioni ottenute nella regione Sardegna sia con il TTWA versione ONS che quella CBS.

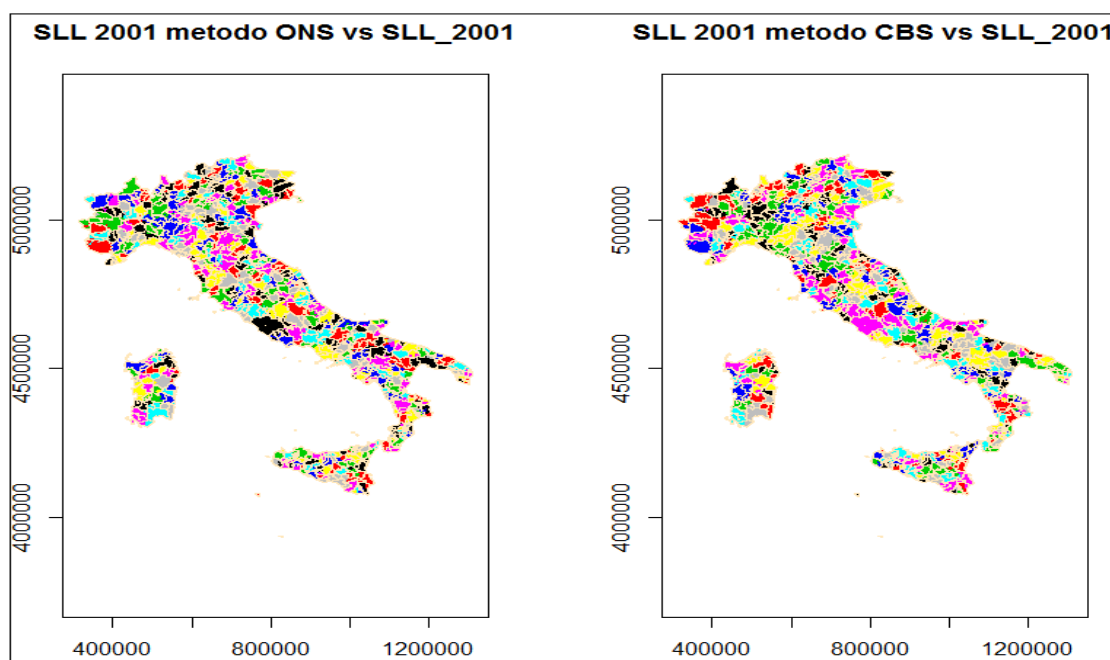


Figura 10. Confronto SLL_2001 ONS e CBS rispetto SLL 2001

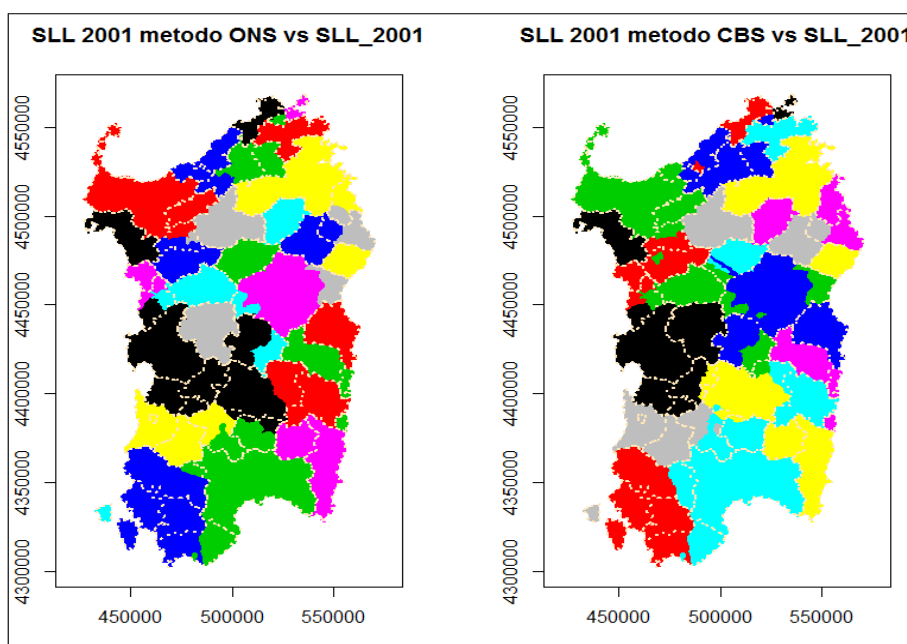


Figura 11. Confronto varianti TTWA rispetto SLL 2001

Nella figura 12 si riportano i box-plot dell'autocontenimento dal lato della domanda e del'offerta dei SLL del 2001 e quelli ottenuti con le due varianti del metodo TTWA.

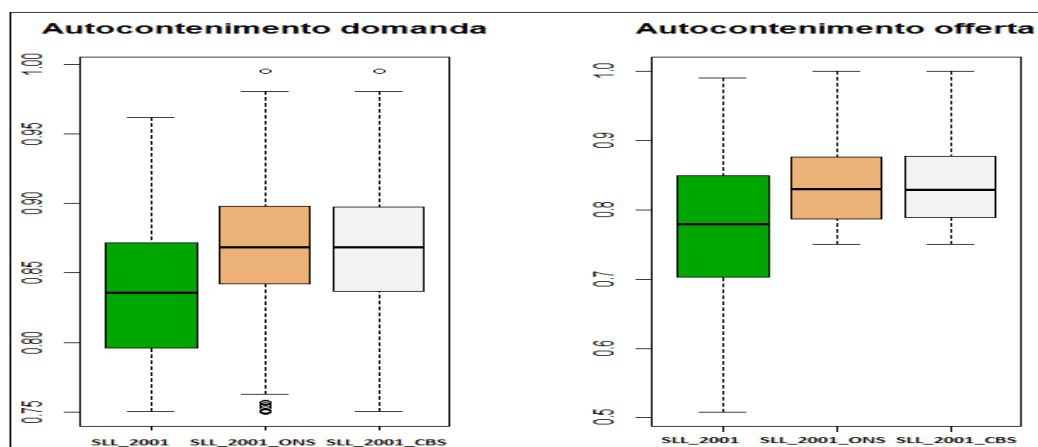


Figura 12. Confronto varianti TTWA rispetto SLL 2001

Dal confronto si evince che per quanto riguarda l'auto-contenimento dal lato della domanda i tre algoritmi rispettano la condizione che il livello di auto-contenimento sia superiore al 75%. Tuttavia, le due varianti del metodo TTWA (soprattutto quello ONS) danno luogo a SLL con auto-contenimento dal lato della domanda migliori. Per quanto riguarda l'auto-contenimento dal lato dell'offerta, rappresentato nei box plot a destra, si evince chiaramente che i SLL del 2001 non rispettano la condizione di minimo dato dal 75%, mentre le due varianti dell'algoritmo TTWA danno luogo a risultati molto simili. A tal proposito per un'analisi più

dettagliata delle soluzioni si deve tener conto anche del numero dei SLL ottenuti. Nell'ipotesi che l'auto-contenimento sia legato anche al numero di SLL ottenuti (minore è il numero più alto dovrebbe essere il livello dell'auto-contenimento) si evince come in definitiva che, a parità di parametri, il metodo ONS, che produce 452 rispetto a 435 della versione CBS, sia l'algoritmo da preferire in base alla sperimentazione condotta sui dati del 2001.

5.1. Analisi di sensitività dei risultati ottenuti con il metodo ONS

Al fine di verificare l'impatto che i parametri dell'algoritmo ONS hanno sui parametri è stata effettuata una analisi di sensitività dell'algoritmo al variare del loro valore. Tale verifica è stata effettuata sulla regione Sardegna variando singolarmente i parametri dell'algoritmo rispetto ad una loro definizione di riferimento data da:

1. minSZ=3500;
2. tarSZ=25000;
3. minSC=66.7%;
4. tarSC=75%.

Nella tabella 1 sono stati riportati le diverse impostazioni dati ai parametri:

Tabella 1: Parametri dell'algoritmo ONS sulla Sardegna

| varianti | minSZ | tarSZ | minSC | tarSC |
|------------|-------|-------|-------|-------|
| Variante 1 | 1000 | 10000 | 55% | 70% |
| Variante 2 | 1500 | 15000 | 60% | 80% |
| Variante 3 | 2500 | 25000 | 70% | 85% |
| Variante 4 | 5000 | 30000 | 75 % | 90% |

I risultati sono riassunti nei prossimi grafici. In particolare nella figura 13 si rappresentano le variazioni in termini di dimensione media dei SLL, in termini di numero medio di comuni all'interno di ciascun SLL ed il minimo valore di auto-contenimento ottenuto, qualora si vari soltanto il parametro relativo alla dimensione minima; nella figura 14 le variazioni ottenute sulla base di modifiche del parametro dimensione target; nella figura 15 si rappresenta la sensibilità dell'algoritmo a variazioni dell'auto-contenimento minimo ed infine nella figura 16 si rappresenta la sensitività dell'algoritmo rispetto a variazioni dell'auto-contenimento target. Dall'analisi dei risultati ottenuti si evince le variazioni dei singoli parametri hanno un differente impatto sui risultati ottenuti dall'algoritmo.

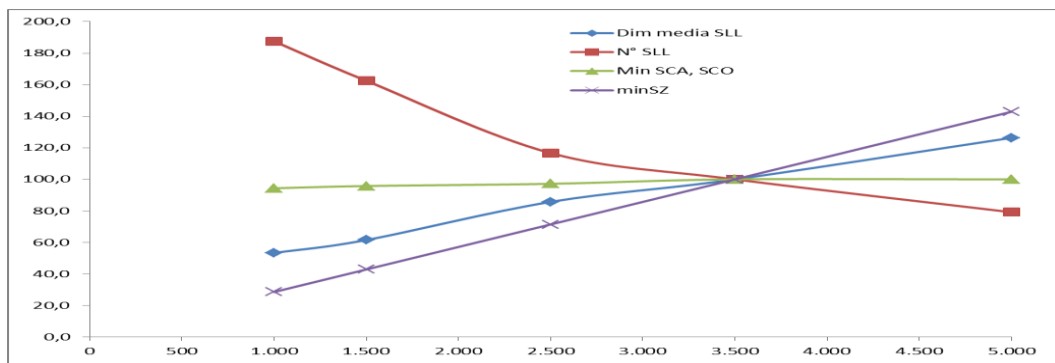


Figura 13. Sensitività rispetto alla dimensione minima

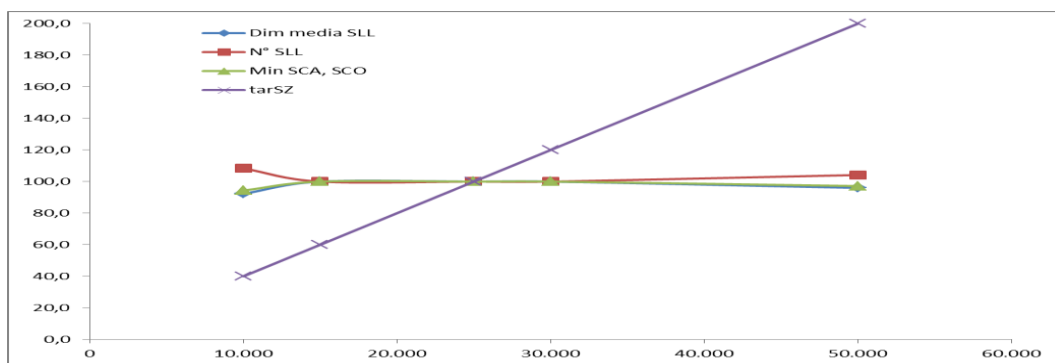


Figura 14. Sensitività rispetto alla dimensione target

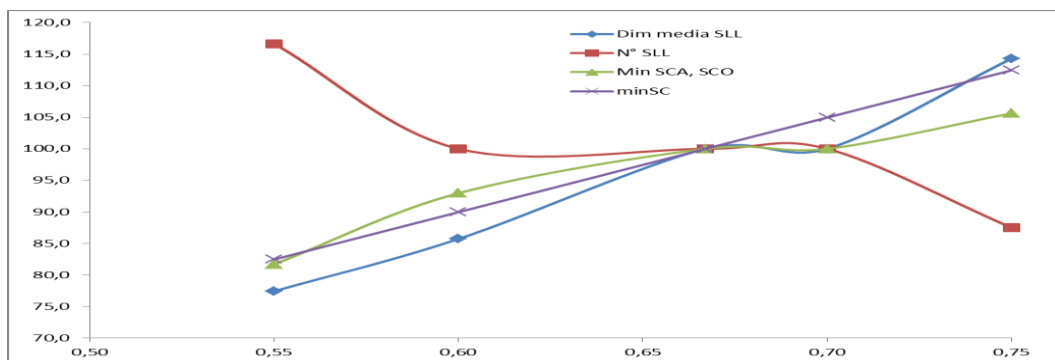


Figura 15. Sensitività rispetto all'aut-contenimento minimo

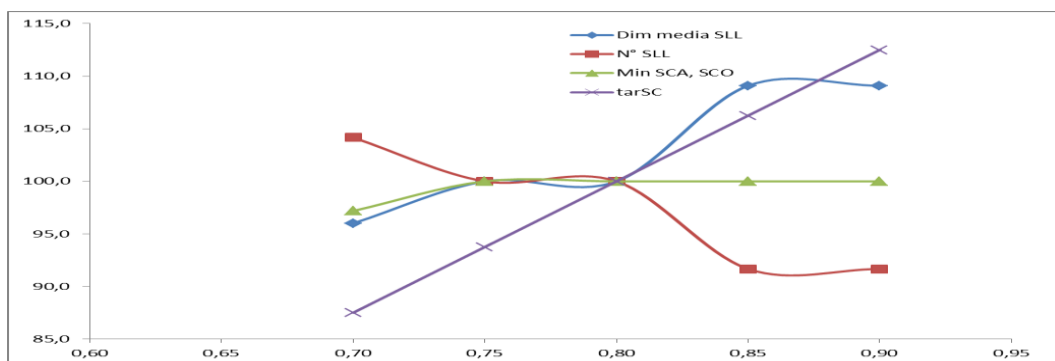


Figura 16. Sensitività rispetto all'auto-contenimento target

Per quanto riguarda la i dati relativi alla sola Sardegna l'algoritmo è sensibile soprattutto a variazioni della dimensione minima ed in misura minore rispetto all'auto-contenimento minimo. Pochissime variazioni nei risultati si hanno invece qualora l'auto-contenimento target, mentre ininfluente pare essere la dimensione target. Tuttavia è utile sottolineare che la struttura del pendolarismo in Sardegna è relativamente semplice, pertanto conclusioni più robuste rispetto alla sensibilità saranno date quando l'analisi sarà condotta su l'intero territorio nazionale.

5. Conclusioni e sviluppi futuri

Scopo del presente lavoro è quello di presentare i risultati preliminari dell'analisi su alcuni algoritmi per l'individuazione di un metodo utile alla definizione dei prossimi SLL del lavoro basati sui flussi di pendolarismo 2011, rilevati tramite l'ultimo Censimento della Popolazione. Al fine di confrontare i risultati, ottenuti con i diversi algoritmi, con i SLL 2001 pubblicati dall'Istat l'analisi è stata condotta sulla base dei dati relativi al 2001.

Tutti gli algoritmi implementati danno luogo a risultati sostanzialmente differenti rispetto ai SLL 2001 in termini di numero di SLL e di conseguenza in termini di dimensione dei SLL. Tuttavia miglioramenti evidenti si hanno dall'applicazione delle due varianti del metodo TTWA, soprattutto se tiene conto che, a differenza dell'algoritmo applicato dall'Istat per l'individuazione dei SLL 2001, quest'ultimi consentono di soddisfare il vincolo di auto-contenimento minimo dal lato dell'offerta. Anche la distribuzione dei SLL in termini di ampiezza appare differente, con possibili effetti sull'interpretazione dei risultati.

Gli algoritmi considerati si basano sul metodo TTWA, tuttavia si ritiene interessante esplorare la possibilità di utilizzare altri metodi basati sulla definizione di un modello statistico. In particolare Roy e Thill (2004) considerando la matrice di pendolarismo hanno proposto un metodo basato su un modelli con interazioni spaziali (spatial interaction models). Lo scopo è quello di descrivere i flussi di pendolarismo attraverso modelli atti a descrivere i processi attraverso i quali entità localizzate in diverse aree interagiscono, individuando e spiegando le modalità di ciascun individuo di auto organizzarsi sul territorio.

Chakraborty et al. (2013) propongono l'utilizzo di modelli bayesiani gerarchici per la determinazione di sistemi locali del lavoro. Il processo di determinazione è diviso in due parti: nel primo passo si modellizzano i dati individuali del pendolarismo (con la possibilità di includere covariate a livello individuale di origine e destinazione e rappresentanti la struttura spaziale). Questa modellizzazione tramite modelli bayesiani permette la stima di inferenze a posteriori sulla matrice del pendolarismo. Nel secondo passo, si utilizzano i risultati della stima del modello e dell'inferenza per la creazione di SLL. In questo secondo passo si definisce una funzione di utilità rispetto alla performance di una partizione e, tramite campioni a posteriori, si ottiene una distribuzione a posteriori dell'utilità per ciascun SLL.

L'utilità è utilizzata sia nell'algoritmo di ricerca di partizioni sia per la valutazione delle singole soluzioni trovate. La funzione di utilità premia la concentrazione di occupati che vivono e lavorano nell'area e disincentiva aree con poche località di base (comuni). La concentrazione è catturata tramite una diagonalizzazione a blocchi delle matrici di interazione spaziale.

Bibliografia

- Chakraborty, A., Beamonte, M.A., Gelfand, A.E., Alonso, M.P., Gargallo, P. e Salvador, M. (2013). Spatial interaction models with individual-level data for explaining labor flows and developing local labor markets. *Comp. Stat. and Data Anal.*, 58, 392-307.
- Casado Díaz, J., Coombes, M., (2011). The delineation of 21st century local labour market areas: a critical review and a research agenda. *Boletín de la Asociación de Geógrafos Españoles* 57, 7–32.
- Coombes, M. and Bond, S. (2007). Travel-to-Work Areas: the 2007 review. London: Office for National Statistics, 2008. (<http://www.ons.gov.uk/ons/guide-method/geography/beginner-s-guide/other/travel-to-work-areas/index.htm>)
- Coombes, M., Casado-Diaz, J.M., Martínez-Bernabeu, L. e Carausu, F. (2012). Study on comparable labour market areas: final research report. 17 October 2012. Eurostat-Framework contract n° :6001. 2008.001 - 2009.065, Specific contract n°: 50405.2010.004 – 2011.325.
- Coombes, M.G., Green, A.E. e Openshow, S. (1986). An efficient algorithm to generate official statistics report areas: the case of the 1984 Travel-to-Work Areas in Britain. *The Journal of Operational Research Society*, Vol. 37, No. 10, pp.943-953. (<https://collaborazione.istat.it/siti/SLL2011/Documentazione%20per%20paese/Forms/AllItems.aspx>) .
- Durieux, E. (2012). Application ANABEL (Analyse Bilocalisée pour les Études Locales): Méthode et guide d'utilisation. PSAR Analyse Territoriale.
- Insee, Dares, Datar, (2012). Atlas des zones d'emploi 2010. <http://www.insee.fr/fr/ppp/bases-de-donnees/donnees-detaillees/atlas-zone-emploi-2010/pdf/atlas-ze-2010.pdf> (visitato febbraio 2014).
- Istat (1997). I sistemi locali del lavoro 1991. Roma, Istituto Poligrafico e Zecca dello Stato, pag 235-242. (<https://collaborazione.istat.it/siti/SLL2011/SLL%20Italia/Forms/AllItems.aspx>)
- Istat e IRPET (1989). I mercati locali del lavoro. Franco Angeli, Milano, Italia. (<https://collaborazione.istat.it/siti/SLL2011/SLL%20Italia/Forms/AllItems.aspx>)

- Mas, S. (2013). French Labour Market Area. Presentazione alla Task Force “Harmonised Labour Market Areas” di Eurostat, 18 Novembre 2013.
- Roy, J. e Thill, J. (2004). Spatial interaction modelling. Papers in regional science. 48 (1), 165-176.
- Semecurbe, F. e Timoteo, J. (2013). Package ttwa. Version 0.8.5.1. (<http://cran.r-project.org/web/packages/ttwa/ttwa.pdf>) .
- Semecurbe, F. (2014). Personal communication.

ABSTRACT

Local Labour Market Areas (LMAs) are geographical areas where most of the people both live and work, i.e. that areas in which the labor demand is mostly satisfied by domestic supply. LMAs have long been recognised as meaningful zonal entities designed to have an explicit validity that is relevant to presenting information and statistics on employment and socio-economic structures, as well as for assessing the effectiveness of regional policy decisions and local government reorganisation. Eurostat is currently considering the creation of harmonised European LMAs as this alternative approach to territorial breakdowns is gaining in importance for analytical purposes and the demand for such data on the part of policy-makers is increasing.

The mapping of functional labour markets on the basis of the commuting patterns of workers has a long history and certainly the "travel to work areas" algorithm proposed by Coombes, et al. (1986) is by far the most used. The method, based on community flows, has been widely adopted both inside and outside Europe, although the choice of threshold values, that great influence has on the determination of the results of the regionalisation, has been severely questioned. A recent revision of such method (Coombes and Bond, 2008) has been adopted by ONS for the production of the new LMAs and is considered by Eurostat for a European implementation. The French Institute of Statistics Insee, has adopted a hierarchical clustering approach, Durieux (2012) for the creation of the zone d'emploi, and little is known on the different results produced by these algorithms.

This paper is aimed to show preliminary results of comparisons between these mathematical algorithms for regionalisation obtained using commuting data from previous population Censuses and further approaches using more complex statistical techniques such as spatial interaction models. Finally, the paper aims at investigating possible ways to tackle the process of understanding and explaining differences between the different methods. In such process we might take into account the spatial configuration of LMAs as a way to explain the geography of a region through the behaviour of the people working in such area.