

Roma, 12 maggio 2025

NUOVE FONTI PER LE STATISTICHE
TERRITORIALI.
RISULTATI E PROSPETTIVE

MOBILE NETWORK OPERATORS (MNO) DATA PER IL TERRITORIO: UN RIFERIMENTO ALLE STATISTICHE PER IL TURISMO

LORENZO CAVALLO, MARIA TERESA SANTORO, TIZIANA TUOTO

Introduzione

L'informazione statistica in ambito turistico riveste un ruolo sempre più centrale ed è oggetto di crescente interesse da parte di una pluralità di utenti, tra cui decisori politici, operatori del settore, ricercatori e analisti.

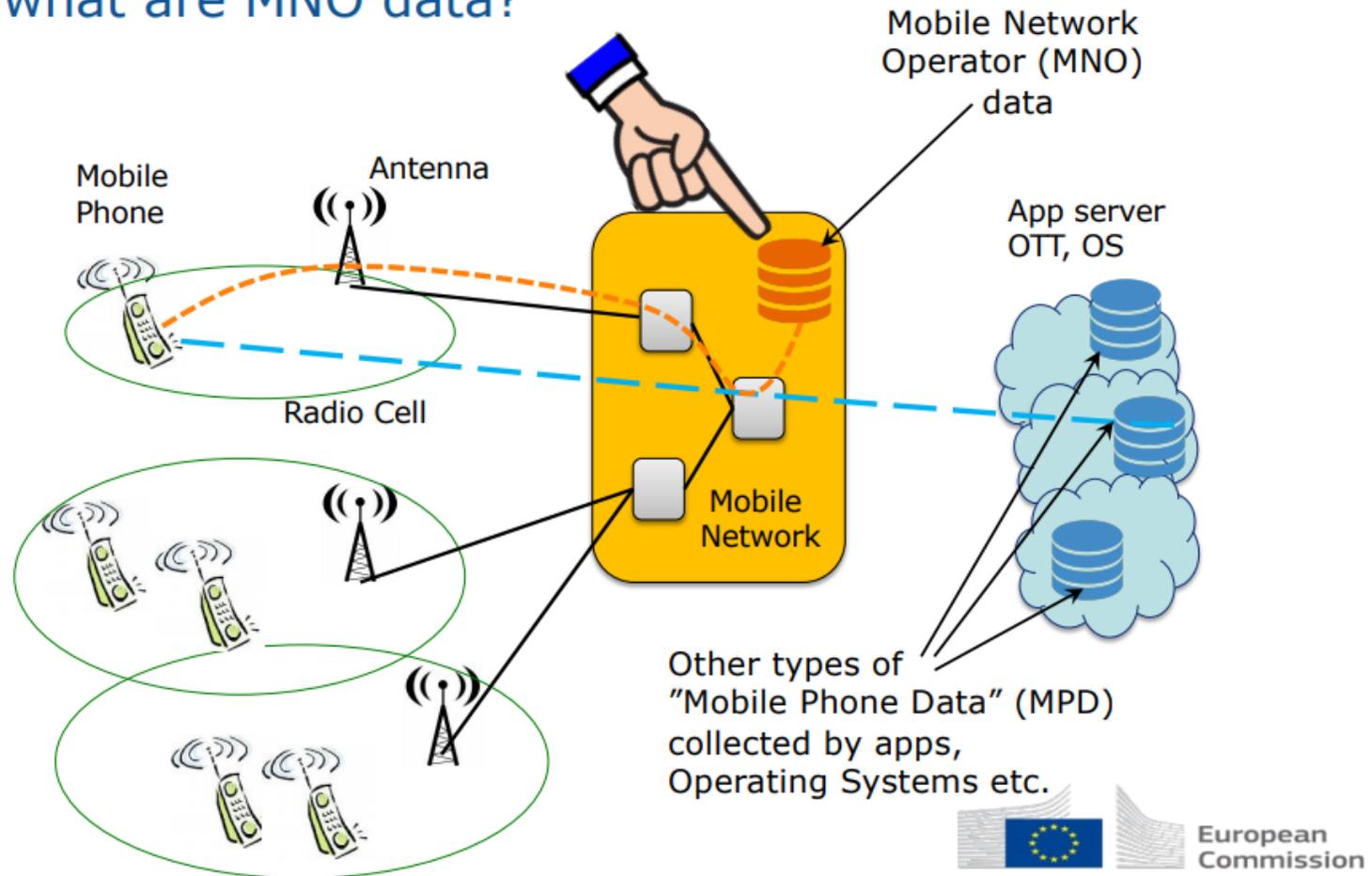
In tale contesto, l'adozione di fonti di dati alternative per la produzione di statistiche ufficiali si configura come una tematica di particolare rilevanza.

In questa presentazione :

- si introducono alcune delle attività in corso a **livello europeo** per l'uso integrato di dati di telefonia nella statistica ufficiale e
- si fornisce una panoramica delle attività svolte e dei principali risultati ottenuti nell'ambito di una collaborazione tra l'**Istituto Nazionale di Statistica (Istat)** e un operatore di telefonia mobile.

Contesto: cosa sono gli MNO data

what are MNO data?



- Sono dati posseduti da privati, privately held data, ottenuti dall'interazione tra il dispositivo mobile e la rete mobile
- raccolti principalmente per la fatturazione e la manutenzione della rete
- Il nostro intento è il loro ri-uso a fini statistici

Caratteristiche dei dati



I dati MNO, nella forma <SIM, timestamp, cell ID>:

COSA: Codice SIM (Subscriber Identity Module)

QUANDO: Giorno/Ora/Minuto

DOVE: Area coperta dall'antenna

*Cosa **NON** dicono:*

CHI ha con se' la SIM

PERCHÉ la SIM è in un'area specifica o si è spostata in quel luogo

- ✓ I dati MNO contengono informazioni spazio-temporali sui telefoni, non sulle persone
- ✓ Possono essere utilizzati per misurare la **presenza umana** (dove si trovano le persone) e la **mobilità** (dove e quando si spostano da/verso) ... prendendo i «telefoni» come proxy per gli «esseri umani».
- ✓ **Vantaggi:** tempestività, granularità temporale, continuità temporale, copertura spaziale - tutto il paese, tutta l'UE, copertura della popolazione - quasi tutti possiedono uno o più telefoni
- ✓ **Limiti e sfide:** incertezza dell'unità statistica (telefoni:persone \neq 1:1), incertezza spaziale (risoluzione spaziale variabile e possibilmente bassa), mancanza di informazioni sullo "scopo" e di variabili sociodemografiche, problemi di accesso ai dati (commerciali, legali), incertezza temporale (osservazioni discontinue)

La prospettiva europea sui dati MNO

Il Sistema delle Statistiche Europee e la Commissione Europea stanno esplorando il riutilizzo dei dati MNO per muovere da casi di studio una tantum e progetti di ricerca esplorativi a produzione di statistiche ufficiali:

- ❖ **Algoritmi e metodologie aperti e trasparenti**
- ❖ **Standardizzazione** di definizioni, dati e metodi, per garantire la comparabilità dei risultati tra i diversi fornitori di dati, tra i vari Paesi e con altri prodotti statistici regolari.
- ❖ **Integrazione** con altre fonti di dati, per superare le limitazioni dei dati di MNO.
- ❖ **Nuovi modelli di partnership**: salvaguardare la privacy delle persone e le informazioni sensibili delle imprese, garantendo al contempo apertura e trasparenza.
- ❖ Concentrarsi sulle **esigenze delle statistiche ufficiali**, migliorare e integrare l'attuale produzione statistica.

La prospettiva europea sui dati MNO

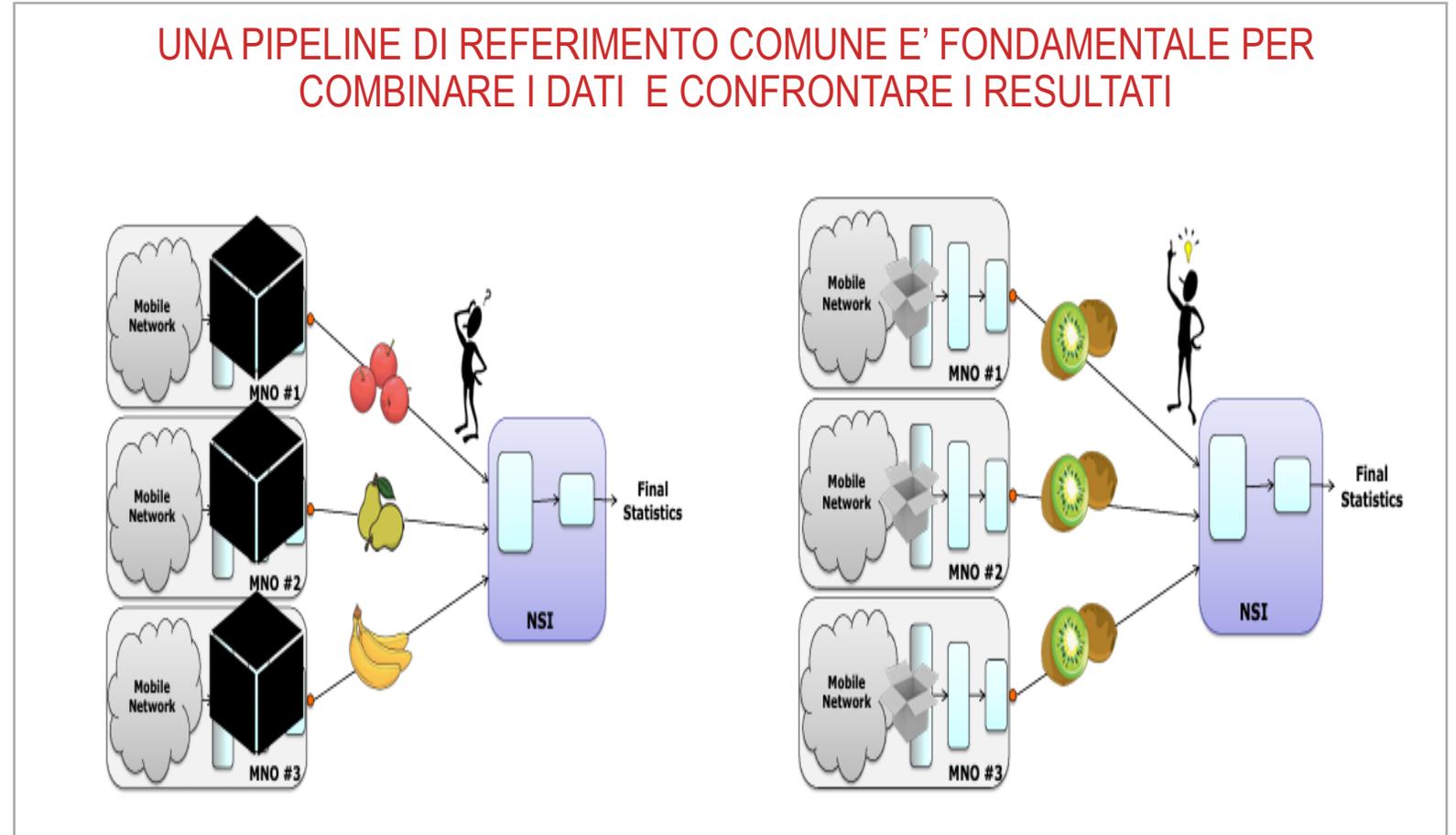
- **Standardizzazione e trasparenza:** identificare le principali fasi del processo produttivo e definire una pipeline di riferimento end-to-end dall'elaborazione dei dati elementari grezzi dell'ORM agli indicatori statistici finali.
- **Integration** con altre fonti di dati: sviluppare **metodologie statistiche rigorose** per rendere i dati MNO «adatti» e utili per le statistiche ufficiali, attraverso l'integrazione con altre fonti di dati

progetto MULTI-MNO

progetto MNO-MINDS
Methods for Integrating New Data
Sources project

Il progetto MULTI-MNO: una pipeline metodologica di riferimento

- La standardizzazione metodologica comprende una chiara e omogenea **specificazione** degli **indicatori statistici**
- Vengono proposti 13 casi d'uso: popolazione presente, mobilità, **turismo**, esposizione al rischio, accesso ai servizi,...
- Viene introdotto il concetto di **usual environment** accanto a quello di “**residenza abituale**”

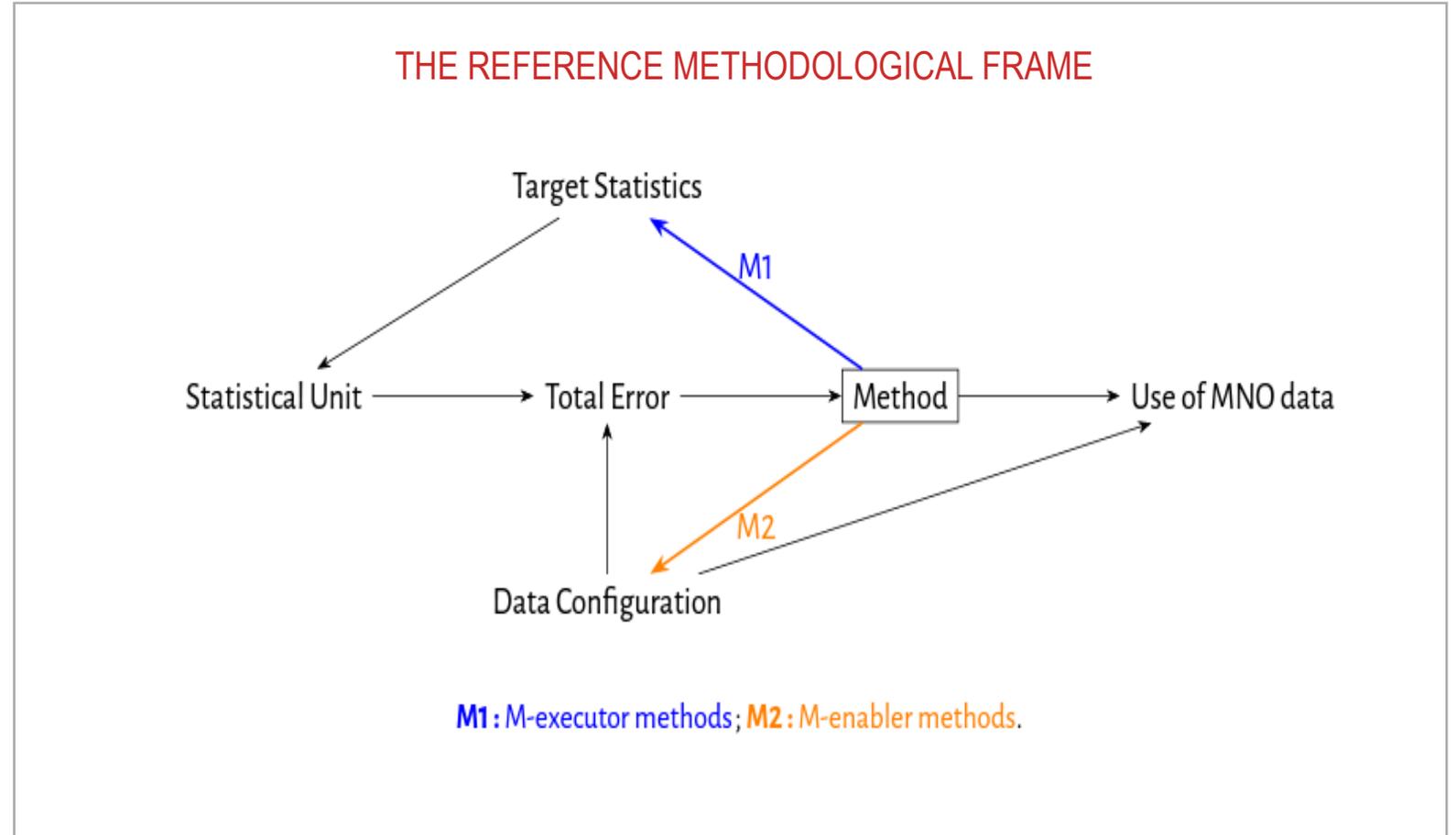


Il progetto MNO-MINDS per lo sviluppo di metodi di integrazione

La fonte di errore, **incertezza**, è al centro e guida nella scelta dei metodi

Tre approcci principali per l'inferenza con dati MNO:

- Randomisation
- Quasi-randomisation
- Super-population modelling



L'esperienza italiana per le statistiche sul turismo

Dal 2020 l'Istat, in collaborazione con uno dei principali fornitori di telefonia mobile in Italia, ha cominciato ad esplorare il potenziale dei dati MNO nel fornire stime e tendenze relative alle principali variabili delle rilevazioni sul turismo, coprendo sia il settore del turismo nazionale sia quello *inbound*.

Tale collaborazione è proseguita con il **Project Foundation** (2023-2024) che ha approfondito vari aspetti, concentrandosi in particolare sul miglioramento della convergenza delle definizioni concettuali e delle classificazioni.

Alcuni dei temi esplorati includevano:

- ✓ Il potenziale utilizzo dei dati MNO per **disaggregare ulteriormente i flussi turistici** a livello territoriale e/o temporale.
- ✓ La possibilità di sfruttare la **tempestività dei dati MNO**.
- ✓ L'esplorazione dell'utilizzo dei dati MNO per la **stima, la previsione o la calibrazione di stime**.
- ✓ L'identificazione di possibili **nuove aree di interesse**.

Dallo *Sprint* al *Project Foundation*

Sprint

Perimetro dell'analisi: Provincia di Rimini e comune di Roma (2 aree a forte vocazione turistica)

Periodi temporali: 3 periodi - agosto 2019 e 2020 e aprile 2020 (periodo di lockdown con flussi turistici pari a 0)

Caso di studio (lato offerta): arrivi e presenze di turisti residenti in Italia (nazionali) e provenienti dall'estero (*inbound*)



Project Foundation

Perimetro dell'analisi: Regioni Emilia Romagna e Lazio (dati di dettaglio a livello comunale, provinciale e regionale)

Periodi di tempo: Da giugno 2022 a ottobre 2023

Caso di studio (lato offerta): arrivi e presenze per residenza dei clienti (per Regione di provenienza per i turisti nazionali, per Paesi estero di provenienza per i turisti internazionali)



Da **Giugno 2022** a **Ottobre 2023**

Una delle prime sfide incontrate è stata stabilire una **relazione tra la popolazione residente e utenti MNO**.

È essenziale che tutte le informazioni possano essere estese all'intero universo di riferimento.

Per superare questo problema, è stato utilizzato un algoritmo proprietario sviluppato dal MNO *(non condiviso)*.

Questo algoritmo è progettato per rappresentare l'intero universo degli utenti, estendendosi oltre ai soli possessori di SIM o stranieri connessi alla rete italiana.

L'algoritmo è caratterizzato da *(informazioni condivise)*:

- ✓ Analisi della quota di mercato locale dell'operatore, considerando le SIM italiane per provincia e fasce d'età;
- ✓ Quota di mercato nazionale delle SIM estere;
- ✓ Quota di mercato in base al tipo di SIM da studi di mercato e rapporti ufficiali (AGCOM);
- ✓ Caratteristiche socio-demografiche degli utenti derivate da dati proprietari.

Una seconda serie importante di questioni è stata affrontata inizialmente durante lo ***Sprint*** e approfondite durante il ***Project Foundation***:

- ✓ **Incerteza nella localizzazione amministrativa**: difficoltà nell'assegnare con esattezza le utenze SIM a un Comune specifico, sia per la copertura territoriale delle celle, non coincidente con i confini amministrativi, sia per l'ambiguità nella determinazione della **residenza abituale**.
- ✓ **Armonizzazione definitoria**: allineamento tra le definizioni delle statistiche ufficiali del turismo (arrivi, presenze, residenza) e le informazioni ricavate dai dati di telefonia mobile, al fine di garantirne la coerenza interpretativa.
- ✓ **Eterogeneità nella risoluzione spaziale**: precisione geografica variabile in relazione alla densità infrastrutturale, più elevata nei contesti urbani, dove la concentrazione delle antenne è maggiore, più limitata nelle aree rurali.
- ✓ **Differenti campi di osservazione**: i dati telefonici riguardano tutti i tipi di alloggio, incluse seconde case, case di parenti e amici, etc., mentre Istat considera solo le strutture ricettive.
- ✓ **Distinzione dei flussi dei clienti domestici e inbound**: difficoltà nel distinguere i residenti nazionali dagli utenti provenienti dall'estero, a prescindere dalla nazionalità della SIM.

Definizioni adottate (*Sprint e Project Foundation*)

CONCETTO	DEFINIZIONE UFFICIALE	DEFINIZIONE ADOTTATA
Residenza abituale	Comune di residenza del turista	Cella notturna prevalente: cella più frequente tra le 20 e le 8 Comune prevalente: dove l'utente trascorre più ore (min. 6h per notte) Residenza abituale: comune della Cella notturna prevalente nei 12 mesi precedenti Comune di destinazione: comune diverso dalla Residenza abituale
Turista	Viaggiatore con pernottamento in un comune diverso dalla Residenza abituale	Utente in una Cella notturna prevalente diversa dalla Residenza abituale (cioè in un Comune di destinazione)
Arrivi	Clients che effettuano il <i>check-in</i> in strutture ricettive nel periodo di riferimento	Il Turista è registrato come Arrivo nel Comune di destinazione il primo giorno del viaggio
Presenze	Notti trascorse in strutture ricettive nel periodo di riferimento	Somma giornaliera dei Turisti nel Comune di Destinazione durante il periodo di riferimento
Turisti Residenti	Turisti residenti in Italia	Utenti con SIM italiana <i>o SIM estera ma con almeno 90 giorni di permanenza in Italia nei 12 mesi di riferimento</i>
Turisti Non residenti	Turisti non residenti in Italia	Utenti con SIM estera <i>e meno di 90 giorni di permanenza in Italia nei 12 mesi di riferimento</i>

I primi test sperimentali delle definizioni

I primi test sulle definizioni adottate hanno evidenziato una significativa discrepanza tra i dati turistici stimati da MNO e quelli provenienti dalla rilevazione Istat «Movimento dei clienti negli esercizi ricettivi» (in parte attesa visto il differente campo di osservazione che per Istat include solo gli esercizi ricettivi ed esclude seconde case, alloggi privati, ospiti di parenti e amici, ecc.).

Per testare la qualità delle definizioni adottate e le ragioni delle ampie differenze, nello **Sprint** è stato analizzato il mese di aprile 2020 (mese caratterizzato dal *lockdown*). Essendo un periodo di flussi turistici prossimi allo zero, ci si attendeva un sostanziale allineamento tra le fonti.

I dati di aprile 2020, tuttavia, hanno mostrato una **disparità maggiore** rispetto ai mesi estivi.

Questo test ha tuttavia evidenziato alcuni punti da affrontare:

- ❖ considerare **un solo comune nella residenza abituale sovrastimava i flussi domestici** (per la presenza di luoghi di lavoro, seconde case, destinazioni frequentate, ecc.), considerati tra i comuni di destinazione turistica;
- ❖ presenza di **provenienze anomale di turisti stranieri** da alcune nazioni (ad es. presenza anomala di SIM estere ad aprile 2020 di nazionalità tipicamente presenti in Italia per motivi di lavoro piuttosto che per svago).

Miglioramenti del *Project Foundation*

Sulla base delle evidenze emerse nei primi test sperimentali sono stati introdotti dei cambiamenti metodologici, definatori e interpretativi per fare dei passi avanti nelle analisi.

A titolo di esempio alcune delle modifiche hanno riguardato:

- ✓ Miglioramento nella definizione di **Residenza abituale**: da **Comune prevalente** a **Comune prevalente NOTTURNO** (ovvero il comune in cui l'utente ha trascorso il maggior numero di notti nei dodici mesi di riferimento).
- ✓ Miglioramento della definizione di **Turista Non residente**: utenti che non utilizzano una SIM italiana e che hanno trascorso meno di 90 giorni in Italia nei 12 mesi di riferimento (per escludere i lavoratori stranieri residenti in Italia dal conteggio dei turisti).
- ✓ Introduzione del concetto di «**Usual Environment**» o «**Ambiente abituale**» per una migliore stima dei turisti residenti in Italia, considerando anche un «cuscinetto» del comune di provenienza.

Usual Environment (Ambiente abituale)

Lo «**Usual Environment**» o «**Ambiente abituale**» è un concetto presente nelle definizioni internazionali di turismo. Comprende tutti i comuni in cui una persona trascorre abitualmente del tempo: la presenza di una persona in questi comuni **non è indice di comportamento turistico**.

Un comune è considerato parte del proprio **ambiente abituale** se l'utente supera una determinata soglia oraria per un determinato numero di mesi in quel comune.

L'introduzione di **ambiente abituale** riduce il fenomeno turistico includendo più comuni in cui l'utente non è considerato un **turista**.

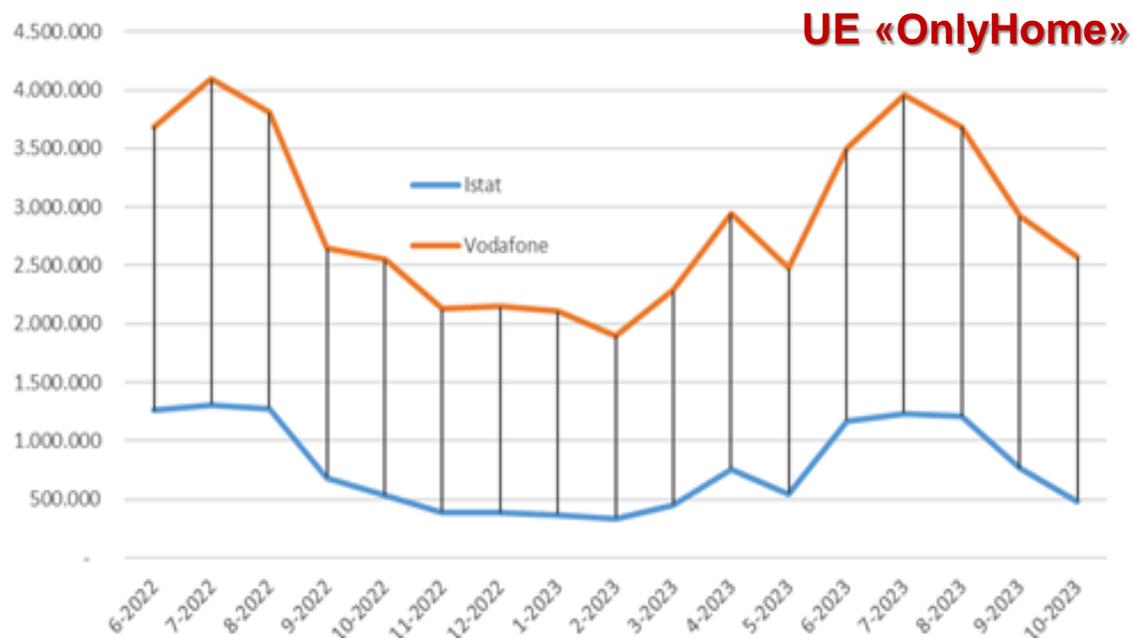
Sono state testate diverse definizioni **Usual Environment** ciascuna con una versione con «buffer», ossia aggiungendo i comuni limitrofi all'insieme ottenuto.

- **onlyhome:** solo il comune di residenza abituale;
- **6m-50h:** comuni con almeno 6 mesi di presenza e un numero di ore ≥ 50 , più il comune di residenza abituale;
- **4m-50h:** comuni con almeno 4 mesi di presenza e un numero di ore ≥ 50 , più il comune di residenza abituale;
- **4m-25h:** comuni con almeno 4 mesi di presenza e un numero di ore ≥ 25 , più il comune di residenza abituale;
- **6m-0h:** comuni con almeno 6 mesi di presenza oppure comuni con 3 mesi di presenza e ≥ 125 ore, 4 mesi e ≥ 100 ore, o 5 mesi e ≥ 75 ore, più il comune di residenza abituale.

Project Foundation: risultati delle analisi dei dati

- ✓ **Innovazione metodologica:** l'introduzione del concetto di **ambiente abituale** rappresenta un elemento di novità nel confronto tra dati ufficiali e informazioni derivanti da fonti MNO.
- ✓ **L'introduzione dell'ambiente abituale ha rafforzato la convergenza con le statistiche ufficiali :** le stime MNO dei flussi domestici tenendo conto dell'**ambiente abituale**, risultano significativamente più vicine a quelle delle fonti statistiche ufficiali. Tuttavia, la definizione dei parametri che delimitano lo **Usual Environment (UE)** rimane attualmente soggettiva e necessita ulteriori test.
- ✓ Gli **andamenti temporali risultano sostanzialmente coerenti tra loro**, indipendentemente dalla variante di UE adottata.
- ✓ **Eterogeneità territoriale:** le differenze tra i territori analizzati (Emilia-Romagna e Lazio) sono marcate. In Emilia-Romagna si rileva una maggiore coerenza tra le due fonti (copertura più ampia delle strutture ricettive nella statistica ufficiale). Nel Lazio, invece, la distanza tra le due fonti è più pronunciata.
- ✓ **Limiti nella disaggregazione territoriale:** il confronto tra le fonti risulta ulteriormente complesso quando si scende a livelli territoriali inferiori alla Regione. L'assenza di una chiara spiegazione delle divergenze osservate, sia nei contesti urbani di grandi dimensioni sia in quelli più piccoli, rende difficoltosa l'interpretazione dei dati disaggregati.

Arrivi dei «Residenti in Italia» in Emilia-Romagna



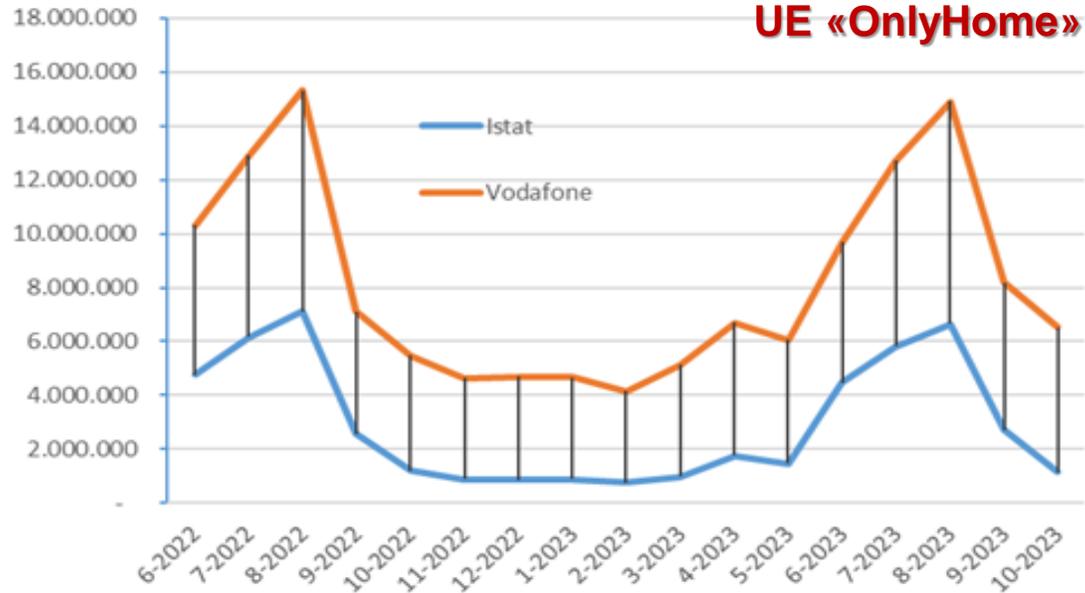
Parallelamente, si osserva una maggiore convergenza nelle dinamiche temporali delle due fonti: le traiettorie evolutive, già caratterizzate da una certa affinità, tendono a sovrapporsi in misura maggiore, con un'attenuazione dei picchi e una riduzione delle divergenze più marcate in corrispondenza di determinati periodi (ad es. ottobre 2022, luglio 2023, ecc.).

Il volume degli **arrivi domestici**, ossia i clienti **residenti in Italia**, registrati da MNO nel passaggio dal criterio di classificazione dell'**ambiente abituale** più restrittivo («**OnlyHome**») a quello maggiormente inclusivo («**6m-0h**»), mostra una contrazione significativa - pari a circa la metà.

Nonostante tale contrazione, i valori assoluti MNO rimangono costantemente superiori a quelli Istat.



Presenze dei «Residenti in Italia» in Emilia-Romagna

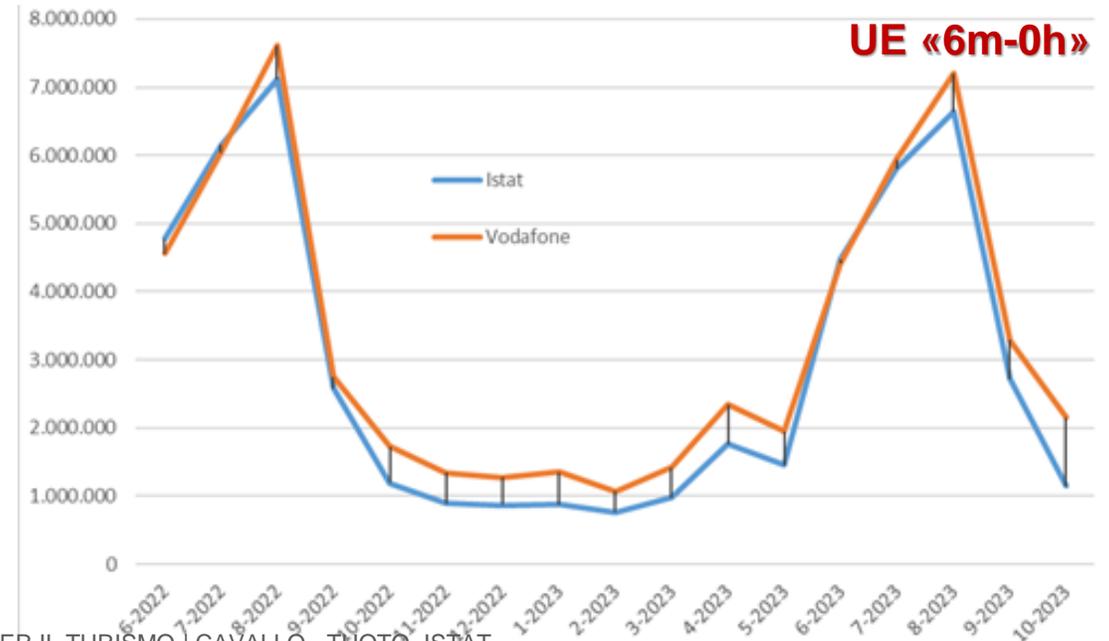


Anche il volume delle **presenze** registrate da MNO e riferite ai **residenti in Italia** evidenzia una riduzione significativa, approssimativamente del 50%, in seguito alla modifica del criterio di classificazione dell'**ambiente abituale** da **«OnlyHome»** a **«6m-0h»**. Tale rimodulazione determina un avvicinamento sostanziale alle stime ufficiali Istat.

La stima delle presenze con opzione **«6m-0h»** risulta pressoché allineata ai valori Istat lungo l'intero periodo considerato, con un rapporto medio mensile tra le due fonti pari a circa 1,1.

In particolare, nei mesi iniziali dell'analisi (giugno e luglio 2022), le presenze MNO risultano addirittura leggermente inferiori rispetto ai valori ufficiali. Scostamenti più marcati si osservano invece nel periodo compreso tra ottobre 2022 e marzo 2023, nonché nel mese di ottobre 2023.

Nel complesso, l'evoluzione temporale delle presenze rilevate attraverso le due fonti presenta un profilo estremamente simile, caratterizzato da un'elevata coerenza nei pattern mensili.



Arrivi e Presenze dei «Residenti all'estero» in Emilia Romagna

- Per i turisti «**non residenti in Italia**» la differenza di valori tra arrivi e presenze delle due fonti è meno evidente rispetto ai turisti domestici.
- In molti casi la stima di MNO è inferiore a quella Istat.
- Gli andamenti sono sostanzialmente gli stessi.



Principali risultati

Nel ***Project Foundation*** ci sono stati passi in avanti rispetto allo ***Sprint***.

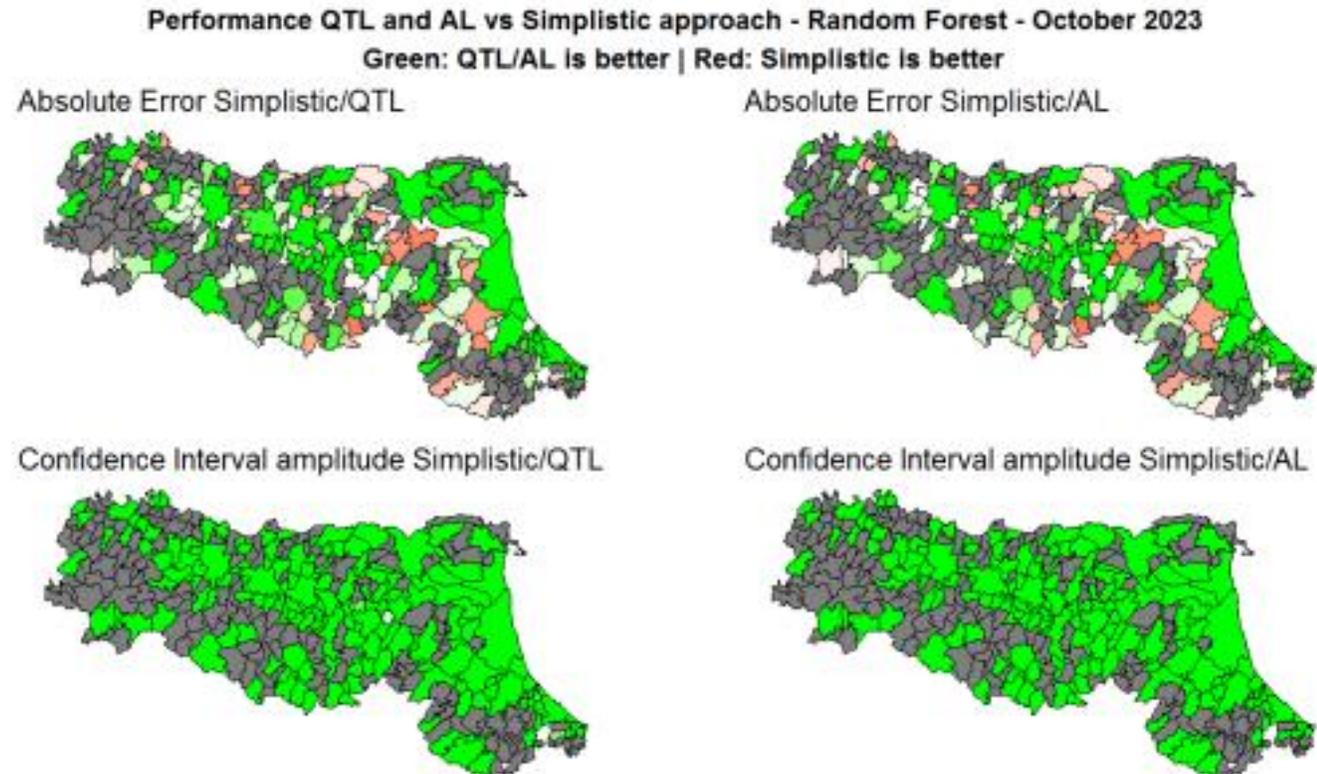
- Le presenze calcolate con i dati MNO sono superiori a quelle rilevate dall'Istat (*i dati MNO non permettono di distinguere dove/in quali strutture hanno soggiornato i turisti*).
- Le stime sembrano migliorate, ma sono ancora necessari ulteriori miglioramenti per arrivare a una convergenza definitiva (ad esempio per gli **arrivi**).
- Lo studio sullo **Usual Environment** ha permesso di migliorare le stime del turismo domestico (*ma la scelta dei parametri da adottare va testata*).
- Un miglioramento significativo è nella definizione di **Turisti Non residenti**, che ha permesso di eliminare molte presenze anomale (turisti provenienti da paesi tipicamente presenti in Italia per motivi di lavoro e non per svago).

I risultati riguardano solo due territori: ***l'analisi nazionale potrebbe produrre risultati diversi***.

Un'applicazione di metodi di integrazione

Finalità: provare a predire le presenze, in anticipo rispetto al dato di indagine, attraverso il dato MNO, disponibile più tempestivamente.

Usiamo i dati di indagine dei mesi precedenti, i dati MNO, e alcune caratteristiche territoriali dei comuni, attraverso metodi di transfer learning.



Valutazioni e prospettive

- ✓ L'esperienza ha confermato le **potenzialità** offerte dai dati provenienti dagli operatori MNO nell'ambito dell'analisi turistica, ma anche i limiti **nell'utilizzo di tali dati in modo affidabile per la produzione di statistiche ufficiali sul turismo**.
- ✓ Le evidenze raccolte indicano che, da subito, i dati MNO potrebbero rivelarsi utili per **stimare le dinamiche territoriali del fenomeno turistico**.
- ✓ Persiste la necessità di **perfezionare ulteriormente la convergenza delle definizioni adottate**, al fine di garantirne una maggiore coerenza e comparabilità.
- ✓ Dato MNO con potenziale significativo per l'apertura di nuove prospettive analitiche, per l'elevata **granularità spaziale** e la **tempestività** di acquisizione. Tuttavia, **necessitano di essere integrati con le fonti statistiche tradizionali per la misurazione del turismo**.
- ✓ Infine, **per un utilizzo sistematico a fini statistici ufficiali dei dati MNO** è necessaria una collaborazione continuativa con gli operatori mobili per testare le definizioni operative, sviluppare algoritmi adeguati e valutare con rigore la qualità complessiva dei dati.

grazie

LORENZO CAVALLO | cavallo@istat.it

MARIA TERESA SANTORO | masantor@istat.it

TIZIANA TUOTO | tuoto@istat.it