

Studio statistico dell'impatto della rete MeteoNetwork sulla stima di
parametri meteo superficiali in Emilia-Romagna

Matteo Sartori
CINECA

Luca Jacopo Avaldi
CINECA

Patruno Paolo
Arpae-SIMC Emilia Romagna

3 febbraio 2016

Sommario

Viste le differenti caratteristiche tra le reti osservative meteorologiche istituzionali (Arpae-SIMC) e quelle amatoriali (MeteoNetwork) è utile utilizzare l'intero insieme dei dati per alcuni scopi, quali ad esempio l'analisi regionale di un parametro meteo superficiale? La differente rappresentatività e i differenti errori strumentali potrebbero indurre un peggioramento delle stime conseguente all'introduzione di errori? Questo studio statistico cerca di dare una parziale risposta. Con almeno due parametri sui tre esaminati la qualità delle stime si consolida con l'integrazione delle reti; si può considerare utile l'uso congiunto delle reti per alcune applicazioni.

Indice

0.1	introduzione	1
0.2	Dati utilizzati nell'analisi	1
0.3	Strumenti per l'elaborazione dei dati	2
0.4	Preparazione dei dati	2
0.4.1	Gross error check	3
0.4.2	Detrend lineare	3
0.4.3	Controllo di qualità spaziale e temprale	3
0.5	Metodi di confronto	4
0.5.1	Confronto tra coppie di stazioni	4
0.5.2	Studio dell'errore di previsione delle reti	5
0.6	Risultati	5
0.6.1	Temperatura	5
0.6.2	Temperatura di rugiada	10
0.6.3	Precipitazioni	15
0.7	Conclusioni	17
0.8	Prospettive future	18
0.9	Ringraziamenti	18
0.10	Licenza	18

0.1 introduzione

L'obiettivo di questo studio consiste nel verificare la possibilità di integrare (in applicazioni specifiche) le reti di stazioni meteorologiche **MeteoNetwork**, e **Arpae Emilia Romagna**. L'analisi proposta è riferita al solo territorio regionale emiliano-romagnolo. La densità delle due reti osservative è differente come lo è la strumentazione e spesso anche la rappresentatività delle stazioni. L'associazione MeteoNetwork Onlus è sostenuta da meteo-amatori e ha tra gli scopi sociali quello di sviluppare e diffondere la conoscenza delle scienze meteorologiche, climatologiche e dell'ambiente. L'Agenzia Regionale Prevenzione Ambiente Energia della Regione Emilia-Romagna Servizio Idro-Meteo-Clima, svolge attività osservative e previsionali operative, di supporto alla pianificazione e di ricerca e sviluppo, in meteorologia, climatologia, idrologia, agrometeorologia, radarmeteorologia e meteorologia ambientale. Arpae gestisce 8 sistemi di monitoraggio e valutazione dello stato dell'ambiente, costituiti da oltre 20 reti di monitoraggio per la sorveglianza continua delle principali matrici ambientali (acqua, aria e atmosfera, suolo). Questo studio vuole essere propedeutico all'impiego di strumenti innovativi e a un nuovo tipo di governance basato sulla partecipazione attiva nel campo del monitoraggio ambientale. In questo contesto in cui il cittadino può e deve avere un ruolo attivo, nasce e si sviluppa R-MAP (Rete di Monitoraggio Ambientale Partecipativo), un progetto sostenuto tra gli altri da Arpae e Cineca, basato sulla condivisione di strumenti hardware e software per la raccolta di dati ambientali. La collaborazione tra Arpae e Cineca, all'interno della quale questo studio si colloca, è incentrata su interessi comuni di integrazione delle reti di comunicazione, anche wireless, con sistemi di monitoraggio ambientale.

0.2 Dati utilizzati nell'analisi

Sono stati esaminati tre parametri meteo-superficiali: *temperatura* a 2 metri dal suolo, *umidità relativa* a 2 metri dal suolo e *precipitazione*. Un quarto parametro (*temperatura di rugiada* a 2 metri dal suolo, è stato ricavato dalla combinazione di temperatura e umidità mediante l'*approssimazione di Magnus-Tetens*. Il parametro *umidità relativa* non è stato analizzato singolarmente: il suo comportamento è stato studiato attraverso la *temperatura di rugiada*.

La rete resa disponibile da Arpae-ER denominata RIRER (Rete Integrata Regione Emilia Romagna che comprende le stazioni di proprietà della Regione, di Arpae e altri enti istituzionali quali AIPO, Consorzi di Bonifica, ex Province etc.) comprende 357 centraline, di cui:

- 209 con sensore di temperatura
- 87 con sensore di umidità
- 348 con pluviometro

La rete MeteoNetwork-ER comprende 107 centraline, ognuna dotata di tutti sensori elencati sopra.

Sono state selezionate dalla rete tutte le centraline che hanno trasmesso almeno un dato durante l'intervallo considerato (dal 01/01/2010 00:00 al 30/06/2015 23:00). Tutte le rilevazioni sono state eseguite alle ore esatte (l'*intervallo di campionamento* scelto è **sessanta minuti**) e in modo sincrono. Sono stati quindi analizzati 48168 istanti temporali.

Sono stati generati quattro dataset per Arpae e quattro per MeteoNetwork (uno per ogni parametro in studio).

Un aspetto da non sottovalutare è la presenza di dati mancanti all'interno dei dataset:

- **Arpae** – Per ogni istante temporale, mediamente il 20% delle osservazioni Arpae non è disponibile. Ciò è dovuto a errori di trasmissione, mancato funzionamento o a dati rimossi durante la fase di pre-processing gestita direttamente da Arpae;
- **MNW** – È più difficile fare una valutazione quantitativa sui dati mancanti nei dataset MeteoNetwork. Essendo una rete su base volontaria e in continua espansione, molte centraline sono state installate successivamente alla data di inizio del periodo in analisi. Analogamente, alcune centraline esistenti a tale data non sono sopravvissute fino alla data conclusiva dell'analisi. A prescindere da questo fatto, l'assenza dei dati può essere causata da errori di trasmissione o mancata registrazione. E' inoltre importante segnalare che, a causa di un'avaria dei server MeteoNetwork, sono stati perduti i dati di tutti i parametri da Luglio 2012 a Luglio 2013 circa.

E' importante specificare che da Luglio 2013 la numerosità delle centraline disponibili nella rete MeteoNetwork è aumentata sensibilmente.

0.3 Strumenti per l'elaborazione dei dati

Le analisi sono state eseguite mediante l'utilizzo del software statistico **R**, supportato da **Apache Hive** su piattaforma **HortonWorks** come base dati. Sono stati sfruttati algoritmi e funzioni di *parallelizzazione* specifici per ridurre il tempo macchina necessario al calcolo.

0.4 Preparazione dei dati

Prima di procedere con l'analisi, i dati sono stati sottoposti a una fase iniziale di *pre-processing*, che può essere così riassunta:

- Gross error check
- Detrend lineare
- Rimozione incrociata dei valori anomali

I dataset Arpae e MeteoNetwork sono stati trattati separatamente. È importante specificare che questa fase dello studio non è volta a proporre una serie di algoritmi rigorosi per il controllo di qualità dei dati, bensì a eliminare **euristicamente** le misurazioni più anomale, riducendo così sensibilmente la possibilità di includere nell'analisi *outlier* e grossolani errori di misurazione o registrazione.

0.4.1 Gross error check

Questa fase è volta a rimuovere gli errori più grossolani (di registrazione o misurazione) dai dataset, ed è stata applicata a tutti i parametri in studio. A tale proposito, sono state individuate euristicamente soglie di *plausibilità* per ogni variabile d'interesse prossime ai campi di misura degli strumenti. Le osservazioni esterne ai range proposti sono state quindi eliminate. Data la natura euristica di tale procedura, non sono formulate ipotesi di tipo probabilistico o statistico a riguardo.

- **Temperatura:** sono definite sogliemensili minime e massime oltre le quali una rilevazione è considerata estremamente improbabile. Le osservazioni al di fuori di tali range sono eliminate;
- **Umidità relativa:** essendo questo parametro espresso in percentuale, tutte le rilevazioni esterne al range 5% - 100% sono state eliminate;
- **Temperatura di rugiada:** le soglie descritte per la temperatura sono state convertite in *temperatura di rugiada*, utilizzando un'umidità relativa pari a 10% per la soglia minima e 100% per la soglia massima;
- **Precipitazioni:** tutte le osservazioni maggiori della precipitazione oraria più alta registrata in Italia dal 2000 a oggi sono state eliminate.

A titolo di esempio nel dataset Arpaè relativo alle temperature sono stati scartati in tutto il periodo 9 valori.

0.4.2 Detrend lineare

Questa fase è stata applicata ai soli parametri *temperatura* e *temperatura di rugiada*, e propone di rimuovere dalle osservazioni l'effetto della diversa altitudine delle centraline. Supponendo che la relazione temperatura ~ altitudine sia lineare, per ogni istante temporale è stato stimato un **modello di regressione lineare univariata**. La disciplina insegna che tale assunzione non sempre è corretta, ma analizzando i *coefficienti di determinazione lineare* (R^2) dei 48168 x 4 modelli stimati si può osservare che nell'80% dei casi il modello lineare approssima sufficientemente la realtà (utilizzando $R^2 = 0.3$ come soglia minima di tolleranza). Per semplicità si è quindi scelto di mantenere quest'assunzione, con la consapevolezza di commettere qualche errore in fase di stima. E' importante specificare che la stazione più elevata tra quelle prese in esame è situata a meno di 1200 metri di altitudine.

Inoltre il detrend è stato applicato solo alle stazioni a una quota minore di 400 metri qualora l' (R^2) della regressione su tutte le centraline fosse più basso della soglia stabilita, eliminando quindi i valori delle centraline con quota superiore a 400 metri.

Per ogni istante temporale, il **detrend** così definito sfrutta il *gradiente di temperatura* stimato (il coefficiente angolare della retta) per eliminare l'effetto della diversa altitudine delle centraline attraverso la formula

$$t_0 = t_h - \alpha h$$

definendo:

- α gradiente di temperatura stimato
- h altitudine della centralina considerata
- t_h temperatura misurata dalla centralina
- t_0 stima della temperatura riportata al livello del mare

0.4.3 Controllo di qualità spaziale e temporale

Questa fase è stata applicata ai soli parametri già sottoposti alla procedura di **detrend** appena descritta. La plausibilità di ogni osservazione è valutata mediante un controllo di consistenza **temporale** e un controllo di consistenza **spaziale**. **Le osservazioni che falliscono entrambi i test sono rimosse dai dataset.**

- **CONTROLLO SPAZIALE** Per ogni istante temporale, è valutata la *gaussianità* della distribuzione delle osservazioni mediante il *test non-parametrico di Kolmogorov-Smirnov*. Se il test non è rifiutato, sono segnalate le osservazioni esterne al range $\mu \pm 3\sigma$. Se il test è rifiutato, sono segnalate le osservazioni il cui valore assoluto dello scarto dalla mediana dell'istante temporale supera il 99° percentile. Questa ultima procedura elimina la coda destra della distribuzione a prescindere dal valore delle osservazioni segnalate e quindi alcuni valori plausibili potrebbero essere erroneamente segnalati;

- **CONTROLLO TEMPORALE** Ogni stazione è analizzata separatamente. Per ogni ora del giorno e per ogni mese, è valutata la *gaussianità* della distribuzione delle osservazioni mediante il *test non-parametrico di Kolmogorov-Smirnov*. Per ogni stazione, sono quindi implementati 24x12 test differenti. Se il test non è rifiutato, sono segnalate le osservazioni esterne al range $\mu \pm 3\sigma$. Se il test è rifiutato, sono segnalate le osservazioni il cui valore assoluto dello scarto dalla mediana dell'istante temporale supera il 99° percentile.

È stato osservato che l'80% dei test di *gaussianità spaziale* implementati ha portato a un non-rifiuto dell'ipotesi nulla, mentre solamente il 69% dei test di *gaussianità temporale* implementati ha portato a un non-rifiuto dell'ipotesi nulla.

Sempre a titolo di esempio nel dataset Arpae relativo alle temperature sono stati segnalati dal controllo spaziale 64.946 dati e 29.263 dati dal controllo temporale; incrociando i due risultati, sono stati eliminati 755 elementi.

0.5 Metodi di confronto

In questo studio sono proposti due diversi approcci per valutare l'integrabilità delle reti Arpae e MeteoNetwork:

- **Confronto tra coppie di stazioni** (parametri *temperatura* e *temperatura di rugiada*);
- **Studio dell'errore globale di previsione delle reti** mediante *leave-one-out cross validation* (parametri *temperatura*, *temperatura di rugiada* e *precipitazioni*). In questo studio il termine previsione è usato in senso statistico e non riguarda l'argomento delle previsioni meteorologiche.

0.5.1 Confronto tra coppie di stazioni

Sono state selezionate dodici coppie di stazioni Arpae e MeteoNetwork (otto per *temperatura*, quattro per *temperatura di rugiada*) in modo da soddisfare i seguenti requisiti:

- Vicinanza (distanza massima tollerata = **3 km**);
- Altitudine compatibile;
- Rappresentatività delle stazioni (esempio: urbane o extraurbane);
- Area climatica omogenea;
- Assenza di ostacoli orografici tra le stazioni.

Per ogni istante temporale è stata calcolata la **differenza** tra il valore della rilevazione Arpae e di quella MeteoNetwork. La distribuzione della nuova variabile così costruita è stata studiata mediante:

- Boxplot;
- Confronto della distribuzione delle differenze (ore notturne e diurne);
- Statistiche descrittive mensili (MBE, RMSE, etc.);
- Differenze massime e minime mensili.

È importante specificare che le statistiche descrittive proposte in questo paragrafo (*root mean square error* e *mean biased error*) sono state originariamente concepite per misurare la discrepanza media fra valori osservati e valori stimati. In questa specifica applicazione invece non vi sono stime o osservazioni, bensì misurazioni effettuate su strumenti differenti. Perciò gli indicatori proposti vanno interpretati come semplici misure sintetiche di distanza tra serie storiche.

L'osservazione dei risultati ha permesso di formulare ipotesi di tipo quantitativo sulla comparabilità di stazioni geograficamente vicine ma appartenenti a reti differenti. Basandosi sulla conoscenza dell'errore complessivo del sistema di misura (sensore e schermo antiradiazioni) per una stazione professionale si è **euristicamente** stabilito che **1.5 °C** (parametro *temperatura*) sia una stima della tolleranza ammessa nella comparazione delle due misure; questo senza pretendere di formulare altre ipotesi di tipo statistico. Oltre a questo bisogna tenere in considerazione il fatto che la distanza tra le due stazioni prese in esame non è mai inferiore a 100 metri: tale soglia è stata perciò inserita nei grafici di output per facilitarne la lettura.

Per quanto riguarda il parametro *temperatura di rugiada*, essendo questo derivato da *tempertaura* e *umidità relativa*, la tolleranza nella comparazione è dato dal contributo di entrambi gli errori che entrano nella relazione usata per derivare la *temperatura di rugiada*. Allo scopo di facilitare la lettura è stata utilizzata una soglia pari a **3 °C**.

0.5.2 Studio dell'errore di previsione delle reti

L'errore di previsione della **rete Arpae** è stato confrontato con l'errore di previsione della **rete congiunta Arpae + MeteoNetwork**: in altre parole, si è voluto determinare come l'aggiunta delle centraline MeteoNetwork modificasse in maniera apprezzabile l'errore complessivo di previsione, calcolato mediante **leave-one-out cross-validation**.

Ogni istante temporale è stato analizzato separatamente. Per ogni parametro in studio, sono stati calcolati 48168x2 indici di dissimilarità (**RMSE**) tra i valori osservati e i valori previsti. È stato scelto l'*algoritmo di interpolazione spaziale di Shepard* nella sua versione *k-nearest neighbours* per stimare le previsioni.

Per ogni parametro, le reti Arpae e Arpae + MeteoNetwork sono quindi state comparate mediante la costruzione della variabile D , definita come la differenza puntuale tra i RMSE delle reti ad un dato istante temporale. Il comportamento di tale variabile può essere osservato attraverso la costruzione di **una serie storica di indici RMSE**.

Se D è maggiore di zero al generico istante t , allora il RMSE della rete Arpae è maggiore del RMSE della rete congiunta: quindi l'aggiunta della rete MeteoNetwork in t contribuisce a diminuire l'errore globale di stima. Una media mobile è stata aggiunta agli output in modo da facilitare l'individuazione di un eventuale trend delle serie.

Il comportamento della variabile D può essere studiato anche attraverso **boxplot**, divisi per stagione (primavera-estate e autunno-inverno) e per fascia oraria (09-15 e 21-03) per osservare eventuali pattern nella distribuzione delle differenze.

Le due reti possono anche essere studiate attraverso la comparazione di curve rappresentanti le **densità stimate degli errori di previsione**, alle quali sono state giustapposte le relative mediane: se la curva più spostata verso lo zero è quella relativa ai RMSE della rete congiunta, è ipotizzabile che l'aggiunta della rete MeteoNetwork tendenzialmente contribuisca alla riduzione dell'errore di previsione.

Per quanto riguarda il parametro *precipitazione* è opportuno precisare che lo studio è stato limitato alle sole **ore di pioggia**, definite come gli istanti temporali nei quali la somma delle precipitazioni della rete congiunta Arpae + MeteoNetwork supera i 2 kg/m^2 . I restanti istanti temporali (assenza di precipitazione) non sono stati presi in considerazione.

Per le *precipitazioni* sono inoltre presentate due ulteriori elaborazioni dei risultati della *cross-validation* che non richiedono il calcolo di indici sintetici per descrivere l'errore di previsione:

- **Boxplot**: le precipitazioni previste sono raggruppate in base ai percentili delle relative precipitazioni osservate, tenendo in considerazione solamente le precipitazioni osservate $> 0.5 \text{ kg/m}^2$. Sono quindi elaborati i rapporti tra le previsioni e le relative osservazioni. La distribuzione di questa nuova variabile è rappresentata mediante boxplot (calcolati su tutto il periodo di osservazione) divisi per percentile delle osservazioni e per stagione di riferimento. Se la rete MeteoNetwork contribuisse a migliorare la precisione delle stime, ci si dovrebbe aspettare che i boxplot della rete congiunta siano *più schiacciati attorno al valore 1* rispetto ai boxplot relativi alla sola rete Arpae;
- **Tavole di contingenza**: previsioni e osservazioni sono *discretizzate* in base alla seguente regola binaria:
 - *Ora di pioggia*: osservazione (previsione) $> 0.2 \text{ kg/m}^2$
 - *Ora non di pioggia*: osservazione (previsione) $\leq 0.2 \text{ kg/m}^2$

Per ogni rete è quindi generata una *tavola di contingenza 2x2* che illustri la concordanza tra valori previsti e osservati. Accuratezza, sensibilità e specificità sono alcuni degli indici che possono essere calcolati su queste tavole. Se la rete MeteoNetwork contribuisse a migliorare l'accuratezza delle previsioni, ci si dovrebbero aspettare valori di questi indicatori più elevati nella tabella riguardante la rete congiunta.

0.6 Risultati

0.6.1 Temperatura

In Figura 1 è riportata la localizzazione delle otto coppie di centraline prese in esame.

Dalla Figura 2 alla Figura 5 sono riportati gli output riguardanti il confronto puntuale tra le centraline **s05444** (Arpa) ed **ero028** (MeteoNetwork), entrambe collocate a ridosso della spiaggia di Cesenatico e situate a 340 metri di distanza l'una dall'altra. La differenza di altitudine tra le due centraline è inferiore a un metro.

Localizzazione delle centraline confrontate (temperatura)

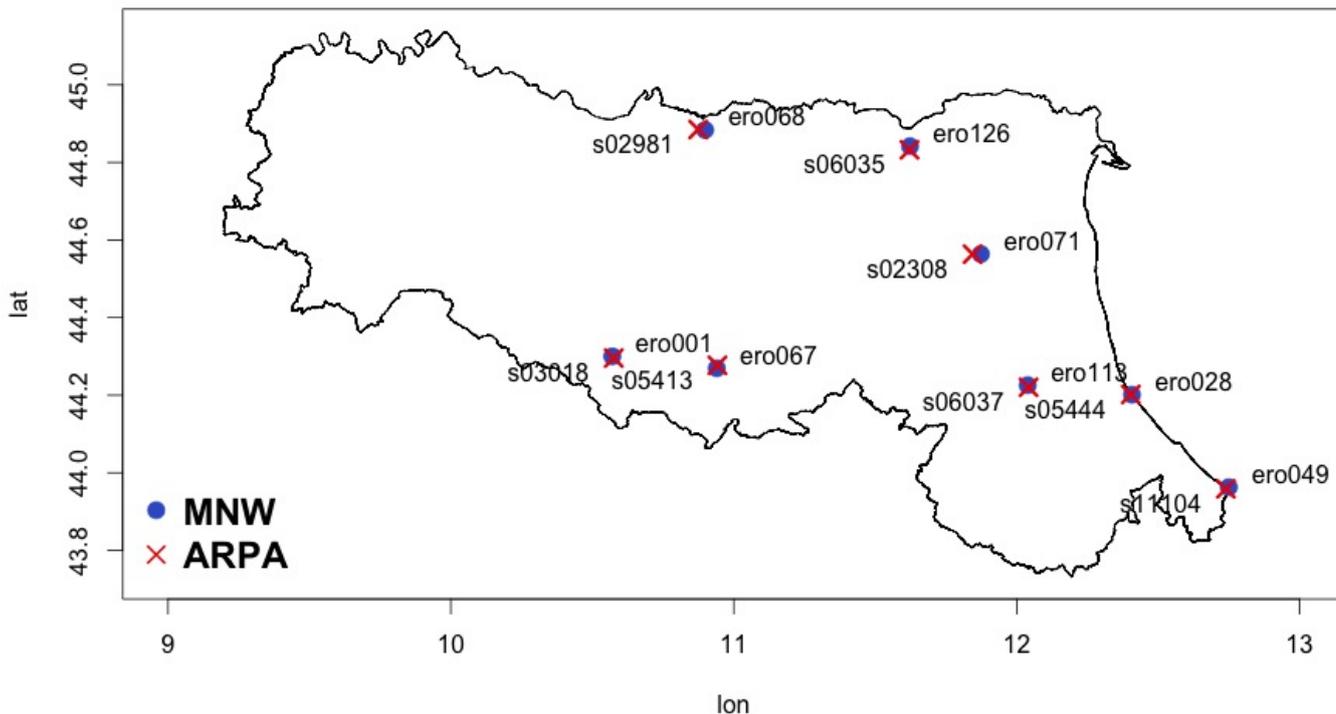


Figura 1

ARPA 's05444' vs. MNW 'ero028' (temperatura) boxplot delle differenze (elementi confrontati: 68.66%)

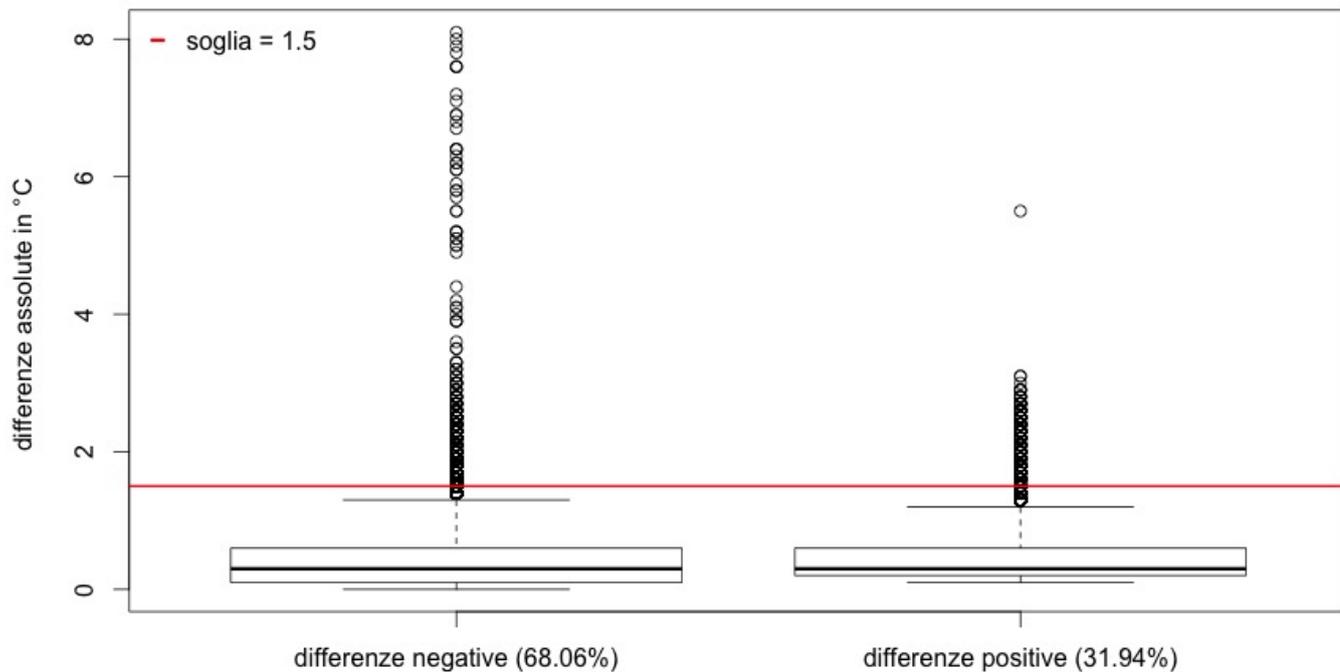


Figura 2

ARPA 's05444' vs. MNW 'ero028' (temperatura)
densità delle differenze diurne (9:00-15:00) e notturne (21:00-3:00)

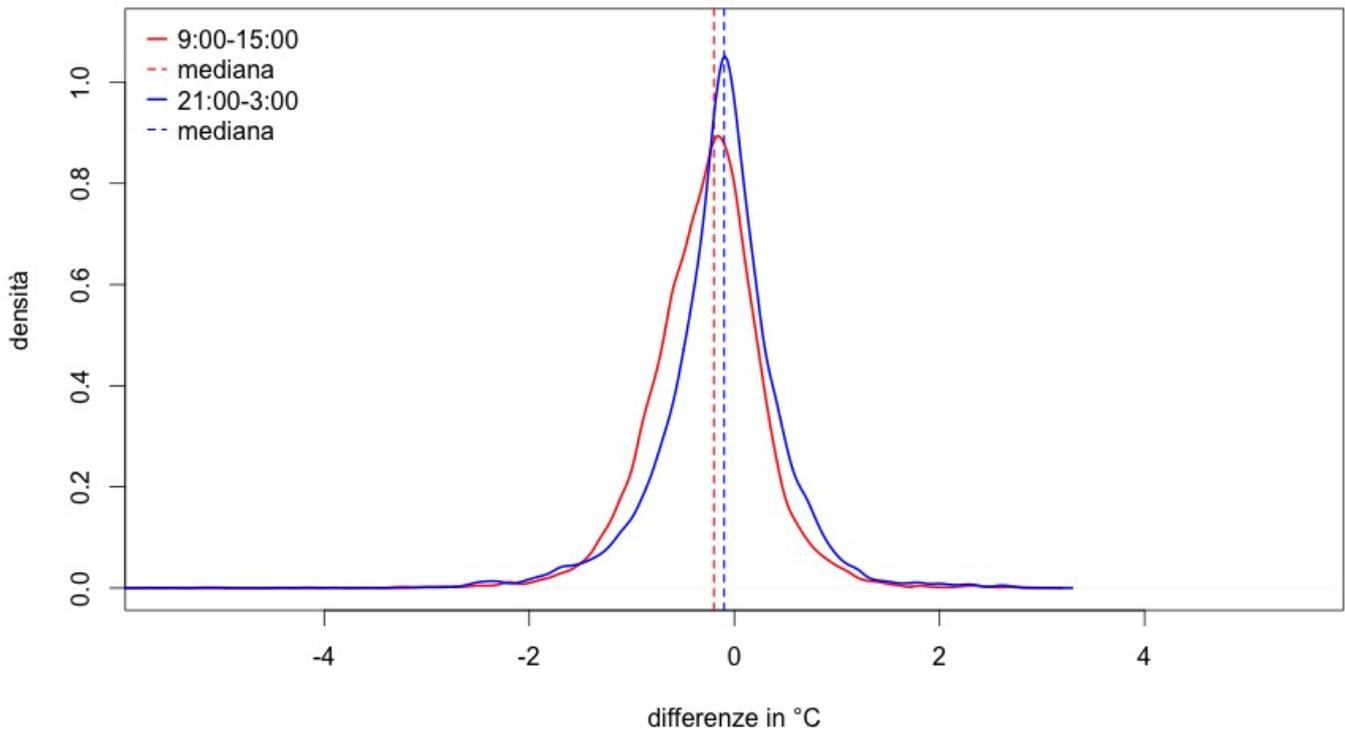


Figura 3

ARPA 's05444' vs. MNW 'ero028' (temperatura)
MBE e differenze massime mensili

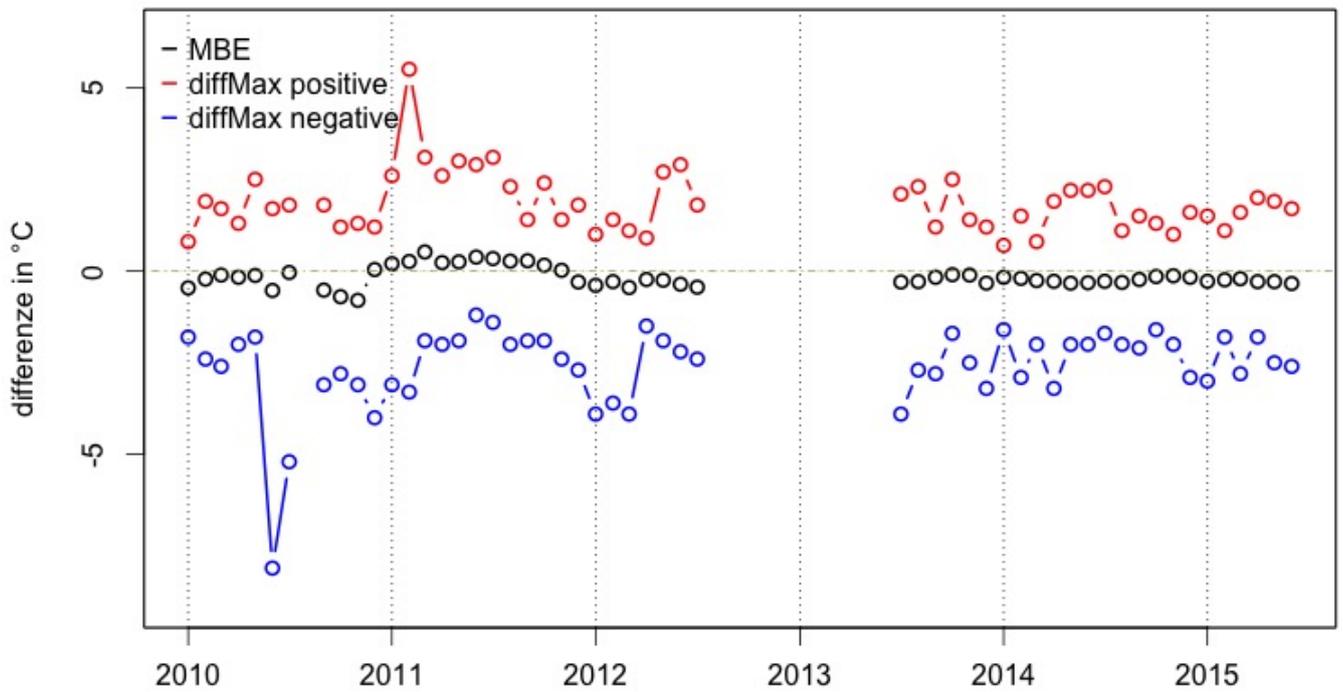


Figura 4

ARPA 's05444' vs. MNW 'ero028'
statistiche mensili (temperatura)

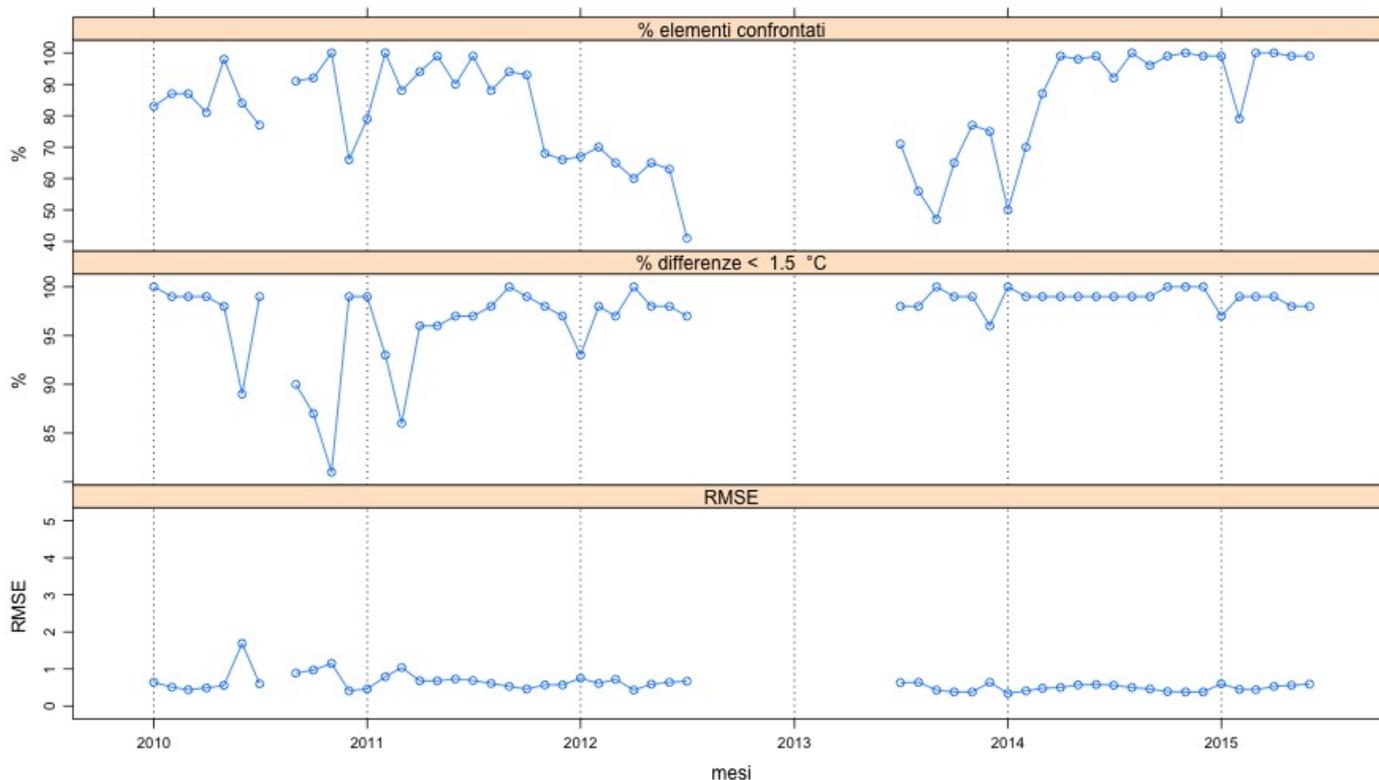


Figura 5

Le centraline prese in esame si comportano in maniera sostanzialmente analoga:

- Esclusi gli *outlier*, la maggior parte delle differenze esaminate è minore (in valore assoluto) della soglia proposta (1.5 °C). Va specificato che nel quasi 70% dei casi il valore rilevato dalla centralina MeteoNetwork sovrastima il valore rilevato da Arpae;
- Gli scostamenti tra le distribuzioni delle differenze notturne e diurne non sono particolarmente accenutati, anche se nelle ore notturne le differenze tendono a diminuire;
- Esclusi alcuni valori anomali (che andrebbero esaminati separatamente) le differenze estreme mensili rientrano nei limiti del plausibile;
- La serie storica dei RMSE mensili non mostra pattern particolari o valori anomali (escluso Giugno 2010, mese nel quale si è verificato il valore anomalo visibile in Figura 4).

Dalla Figura 6 alla Figura 8 sono riportati gli output riguardanti lo studio dell'errore di previsione:

- La distribuzione dei RMSE nelle due reti è molto simile: la curva relativa ai RMSE della rete congiunta è tuttavia lievemente più spostata verso lo zero rispetto alla curva della sola rete Arpae;
- La serie storica delle differenze tra RMSE non lascia intravedere pattern particolari: le differenze non sono mai numericamente molto consistenti, anche se durante alcuni brevi periodi la rete MeteoNetwork sembra aver aumentato l'errore di previsione. Per facilitare la lettura del grafico è stata aggiunta una **media mobile** (ampiezza di quattro settimane) e una **doppia soglia di tolleranza**, centrata sullo zero e di ampiezza pari a un decimo del range delle differenze.
- Stagione e fascia oraria sembrano non influire sulla distribuzione delle differenze tra RMSE.

**Studio dell'errore di previsione delle reti (temperatura)
confronto densità RMSE (ARPA vs. ARPA+MNW)**

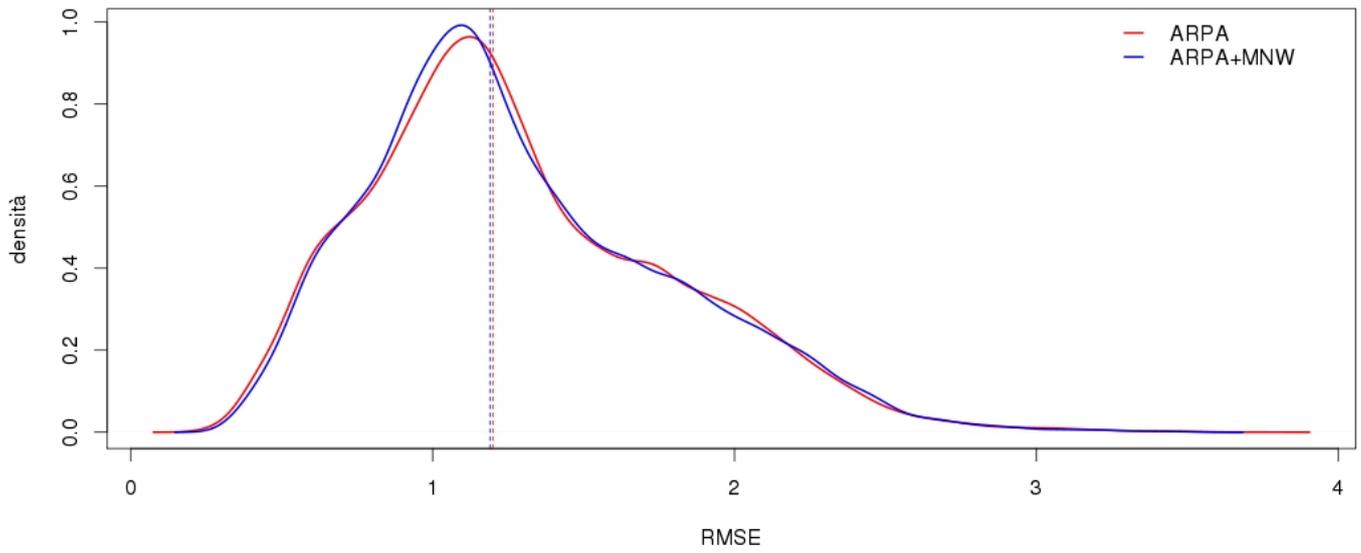


Figura 6

**Studio dell'errore di previsione delle reti (temperatura)
serie storica differenza RMSE (ARPA vs. ARPA+MNW)**

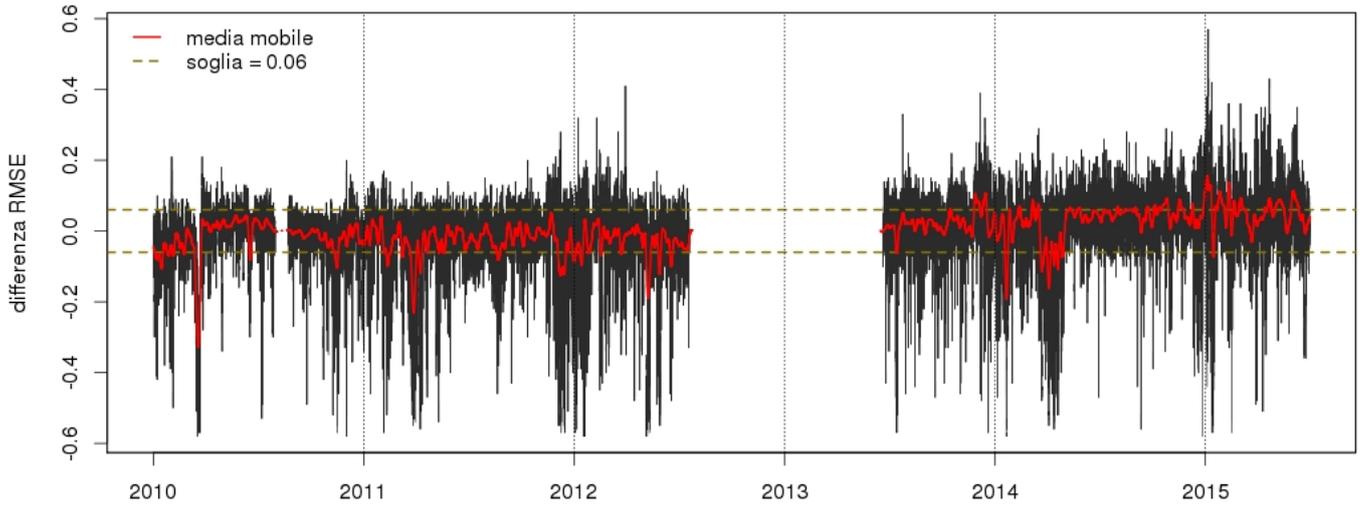


Figura 7

**Studio dell'errore di previsione delle reti (temperatura)
boxplot differenze RMSE (ARPA vs. ARPA+MNW)**

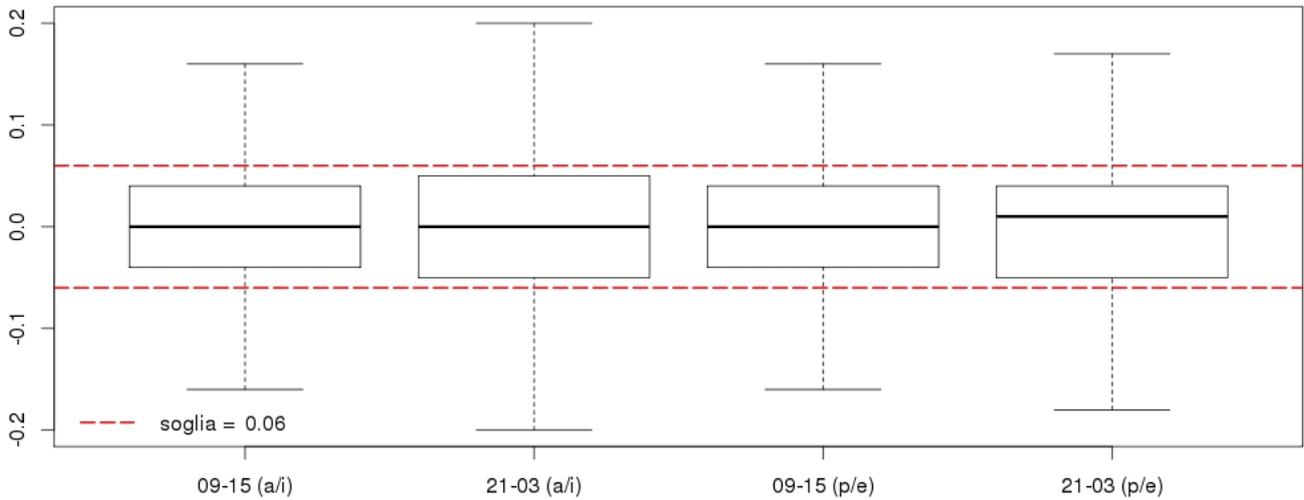


Figura 8

0.6.2 Temperatura di rugiada

In Figura 9 è riportata la localizzazione delle quattro coppie di centraline prese in esame.

Dalla Figura 10 alla Figura 13 sono riportati gli output riguardanti il confronto puntuale tra le centraline **s02981** (Arpae) ed **ero068** (MeteoNetwork), situate a Rolo (territorio pianeggiante) e collocate a 2.15 km di distanza l'una dall'altra. La differenza di altitudine tra le due centraline è inferiore a un metro.

Dall'osservazione degli output si può evincere che:

- Esclusi gli *outlier*, la stragrande maggioranza delle differenze esaminate è minore (in valore assoluto) della soglia proposta (3 °C). Va specificato che nell'80% dei casi i valori ricavati dalle misurazioni della centralina MeteoNetwork sovrastimano i valori ricavati da quella Arpae;
- Le curve di densità delle differenze notturne e diurne sono simili, anche se le differenze diurne hanno code più pesanti. È evidente che in generale i valori di MeteoNetwork tendono a essere più elevati dei valori Arpae (entrambe le mediane sono molto lontane dallo zero);
- Come pronosticato, le differenze estreme sono elevate perché due variabili diverse contribuiscono alla stima della *temperatura di rugiada*. Anche il MBE mensile conferma la tendenza alla negatività delle differenze.

Localizzazione delle centraline confrontate (punto di rugiada)

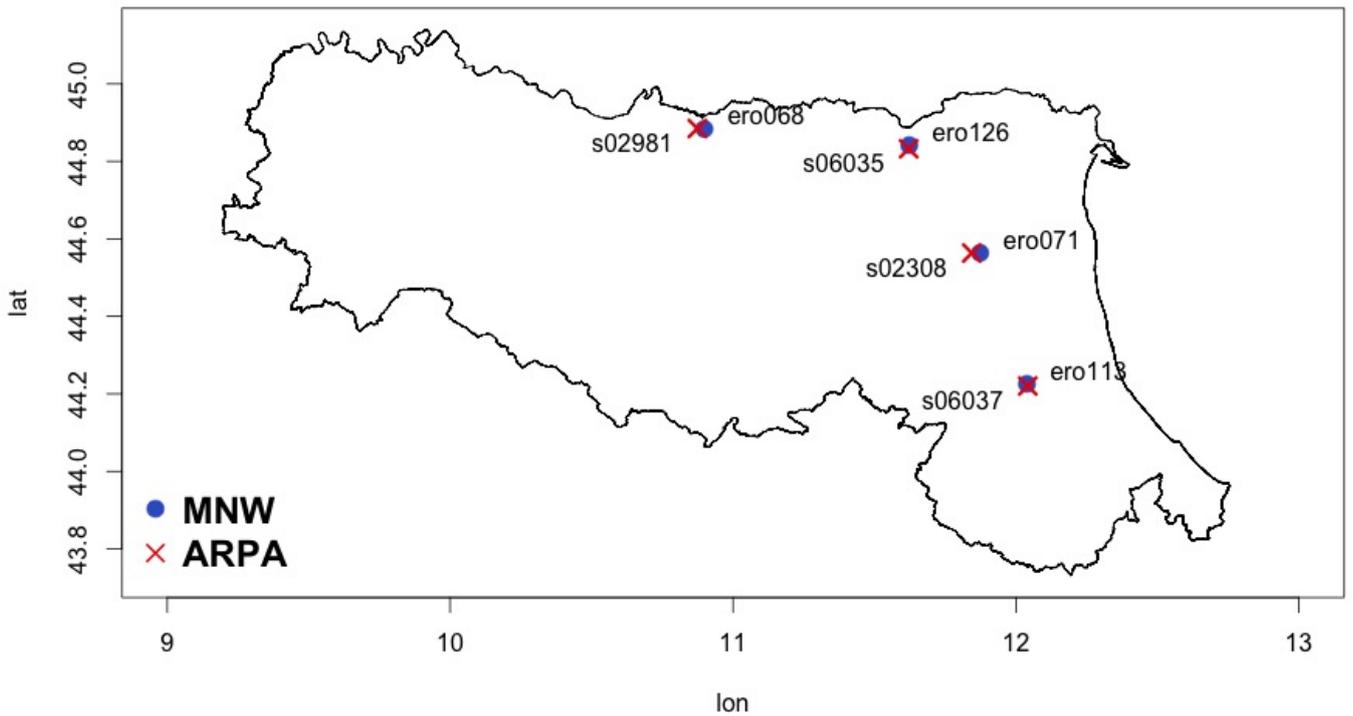


Figura 9

ARPA 's02981' vs. MNW 'ero068' (temperatura di rugiada) boxplot delle differenze (elementi confrontati: 49.4%)

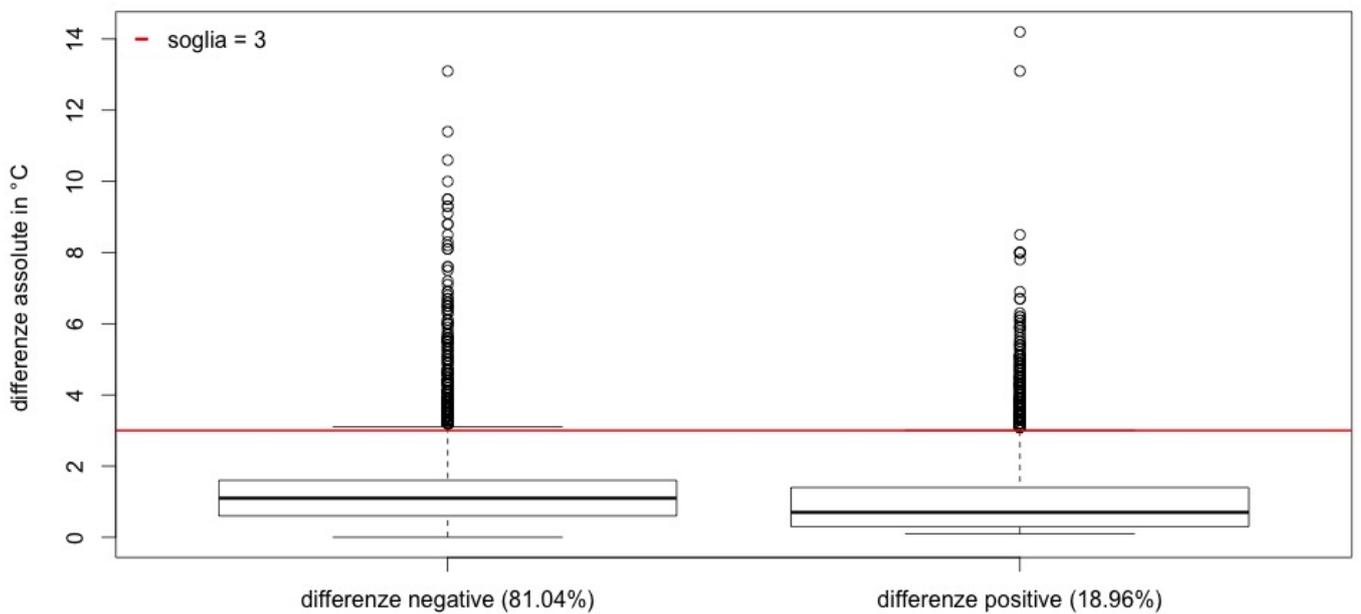


Figura 10

ARPA 's02981' vs. MNW 'ero068' (temperatura di rugiada)
densità delle differenze diurne (9:00-15:00) e notturne (21:00-3:00)

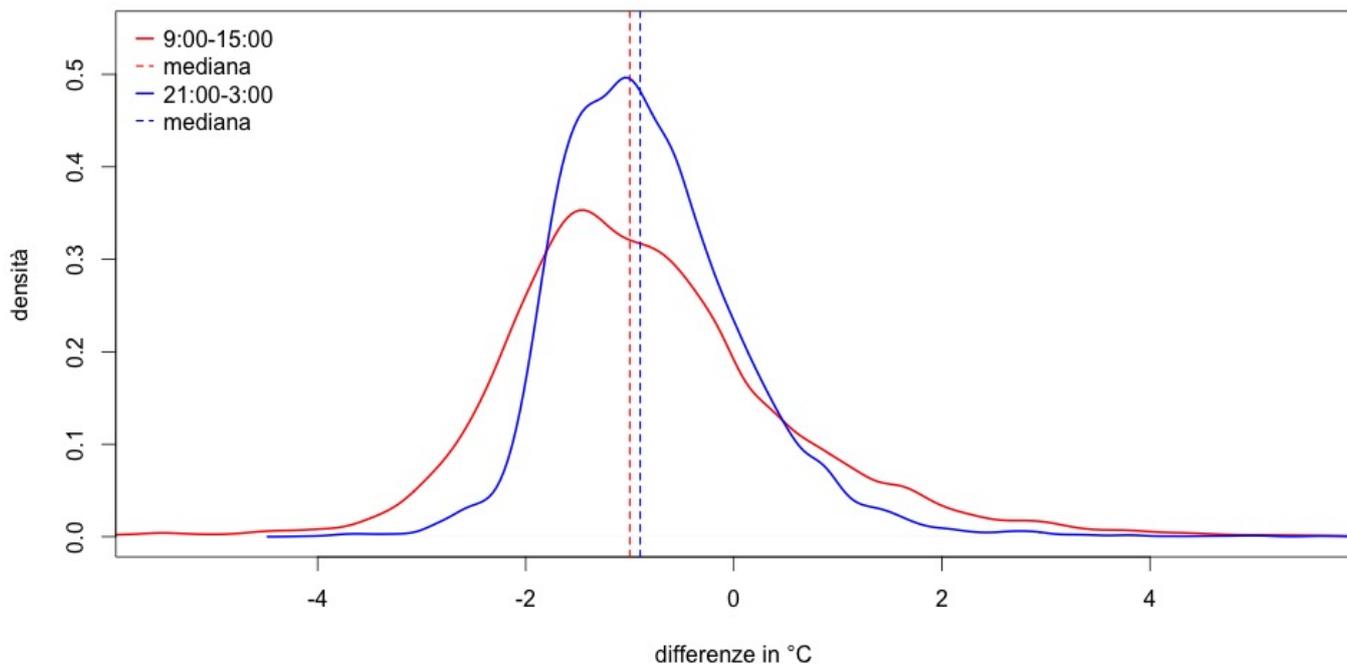


Figura 11

ARPA 's02981' vs. MNW 'ero068' (temperatura di rugiada)
MBE e differenze massime mensili

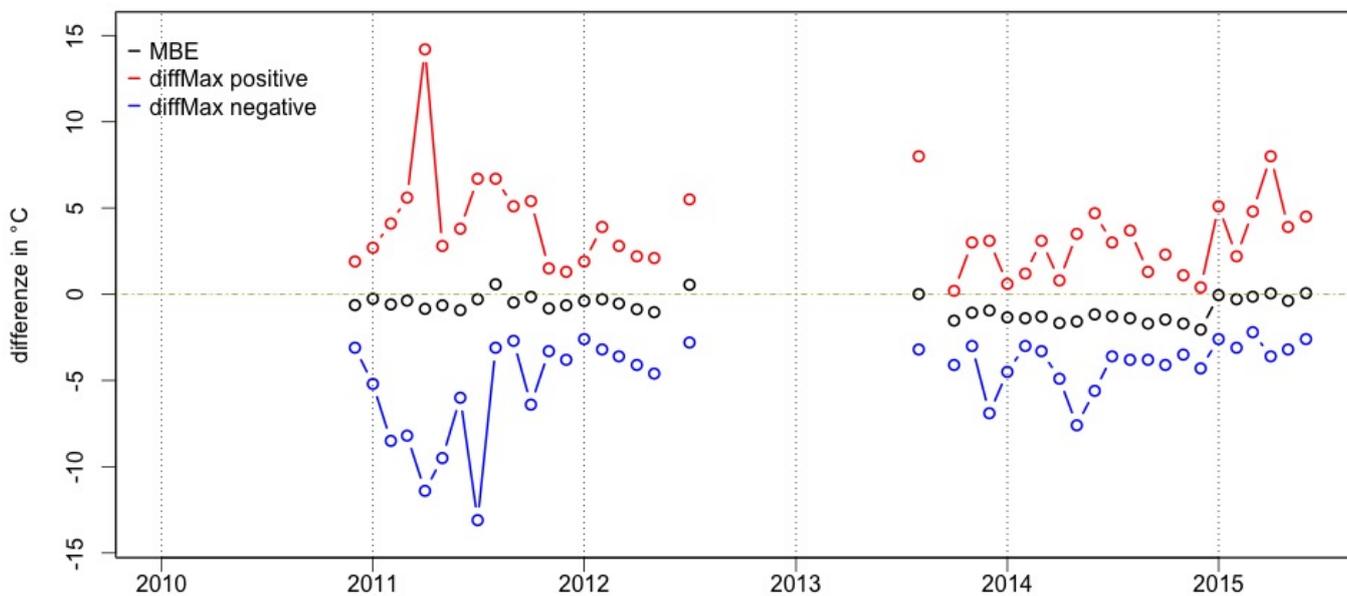


Figura 12

ARPA 's02981' vs. MNW 'ero068'
statistiche mensili (temperatura di rugiada)

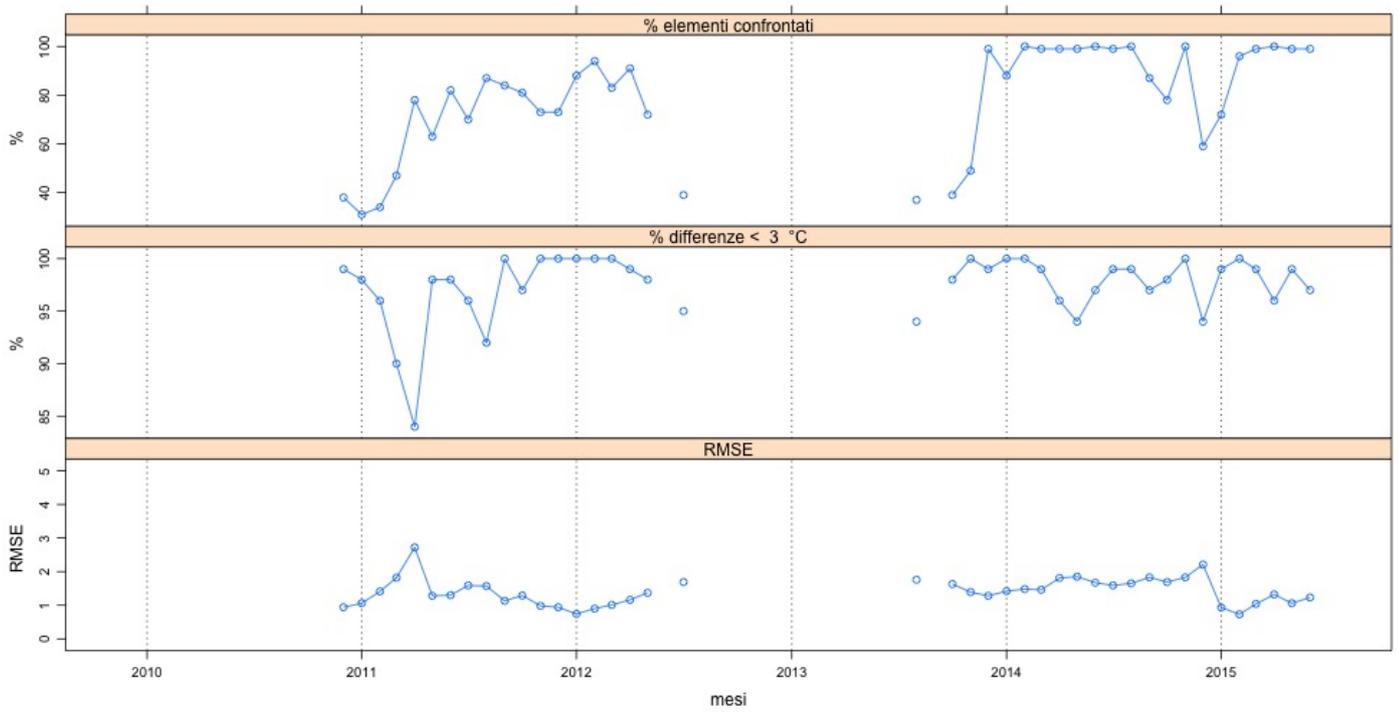


Figura 13

Studio dell'errore di previsione delle reti (temperatura di rugiada)
confronto densità RMSE (ARPA vs. ARPA+MNW)

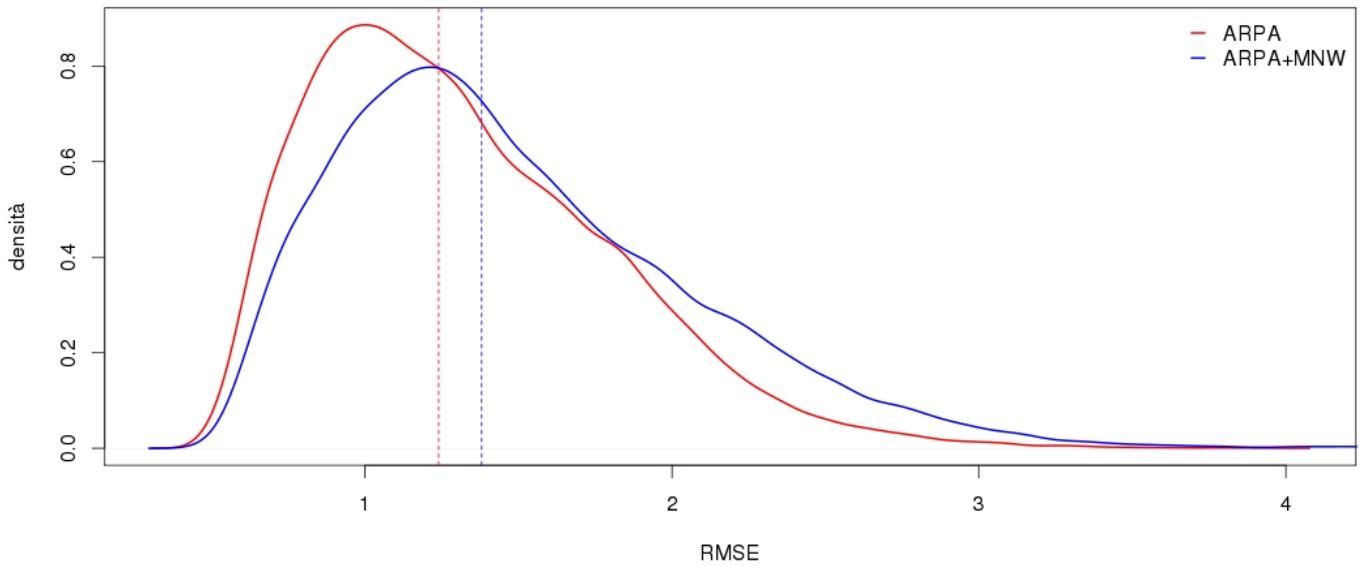


Figura 14

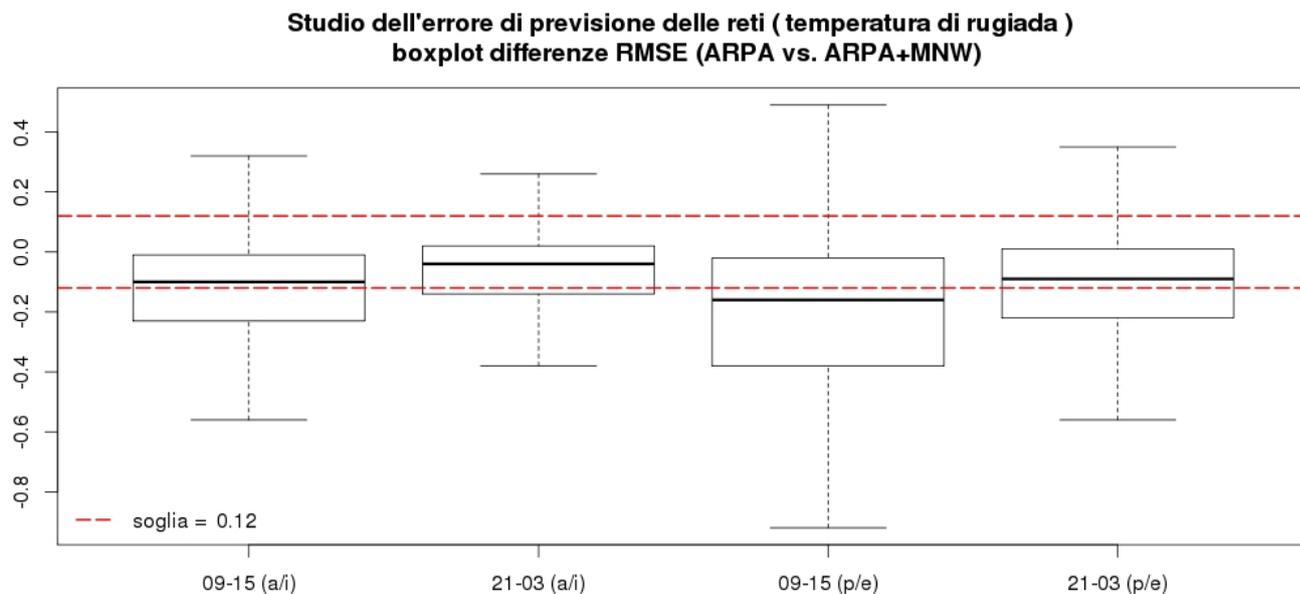


Figura 15

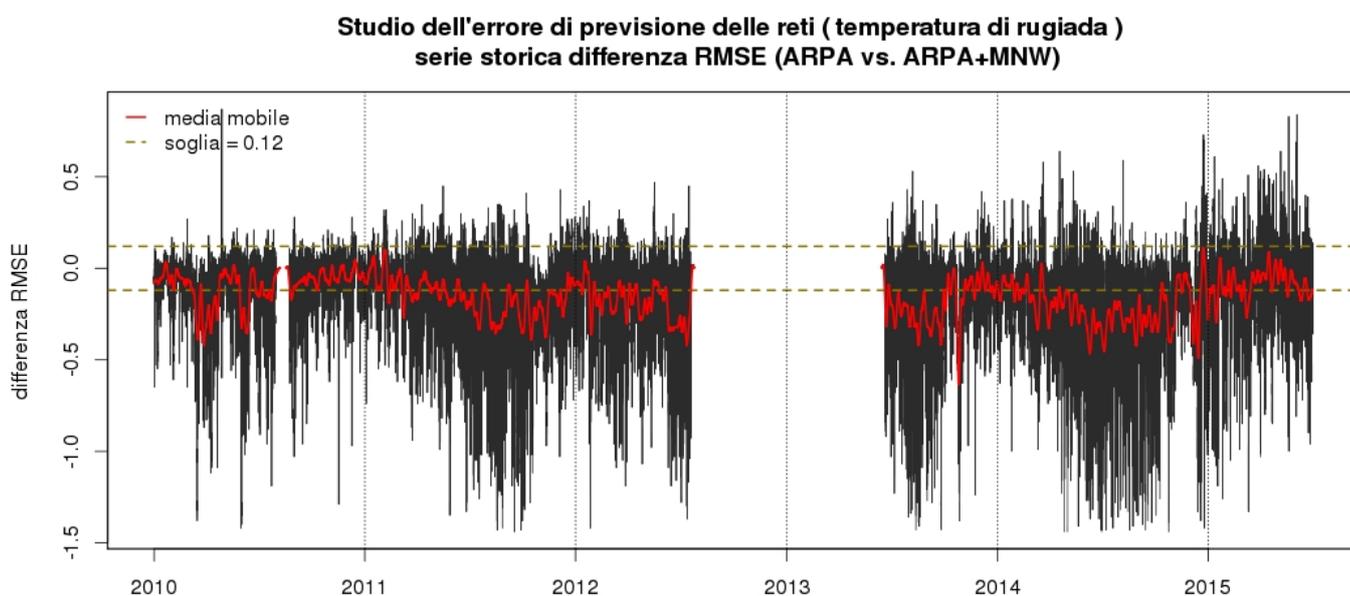


Figura 16

Dalla Figura 14 alla Figura 16 sono riportati gli output riguardanti lo studio dell'errore di previsione delle reti:

- La distribuzione dei RMSE nelle due reti è molto differente: pare che l'aggiunta della rete MeteoNetwork peggiori in maniera significativa la precisione delle stime della rete congiunta;
- Nella serie storica delle differenze tra RMSE vi è una ciclicità piuttosto evidente: nei mesi estivi l'aggiunta della rete MeteoNetwork danneggia la previsione delle stime in maniera accentuata. Per facilitare la lettura del grafico è stata aggiunta una **media mobile** (ampiezza di quattro settimane) e una **doppia soglia di tolleranza**, centrata sullo zero e di ampiezza pari a un decimo del range delle differenze;
- Come osservato al punto precedente **stagione** e **fascia oraria** sembrano influire sulla distribuzione delle differenze tra RMSE: nella fascia diurna estiva le differenze negative sono preponderanti. È stata formulata un'ipotesi per spiegare questo fatto: probabilmente i rilevatori di umidità delle centraline MeteoNetwork tendono a sovrastimare l'umidità relativa reale. Questa distorsione è più evidente nella stagione meno umida (i mesi estivi per l'appunto).

0.6.3 Precipitazioni

Dalla Figura 17 alla Figura 19 sono riportati gli output riguardanti lo studio dell'errore di previsione delle reti:

- La distribuzione dei RMSE nelle due reti è quasi identica: ciò significa che l'aggiunta della rete Meteo-Network non influenza in modo particolare la precisione delle stime. I RMSE sono generalmente piuttosto contenuti, con qualche sporadico valore molto alto, dovuto probabilmente a fenomeni locali molto intensi e difficilmente prevedibili;
- Nella serie storica delle differenze tra RMSE non sono osservabili pattern particolari. Per facilitare la lettura del grafico è stata aggiunta una **media mobile** (ampiezza di quattro settimane) e una **doppia soglia di tolleranza**, centrata sullo zero e di ampiezza pari a un decimo del range delle differenze. Se si escludono gli *outlier* (come in Figura 18) le differenze sono talmente esigue da essere interamente comprese tra le soglie proposte;
- **Stagione e fascia oraria** sembrano non influire sulla distribuzione delle differenze tra RMSE: la maggior parte delle differenze ha valore assoluto molto basso.

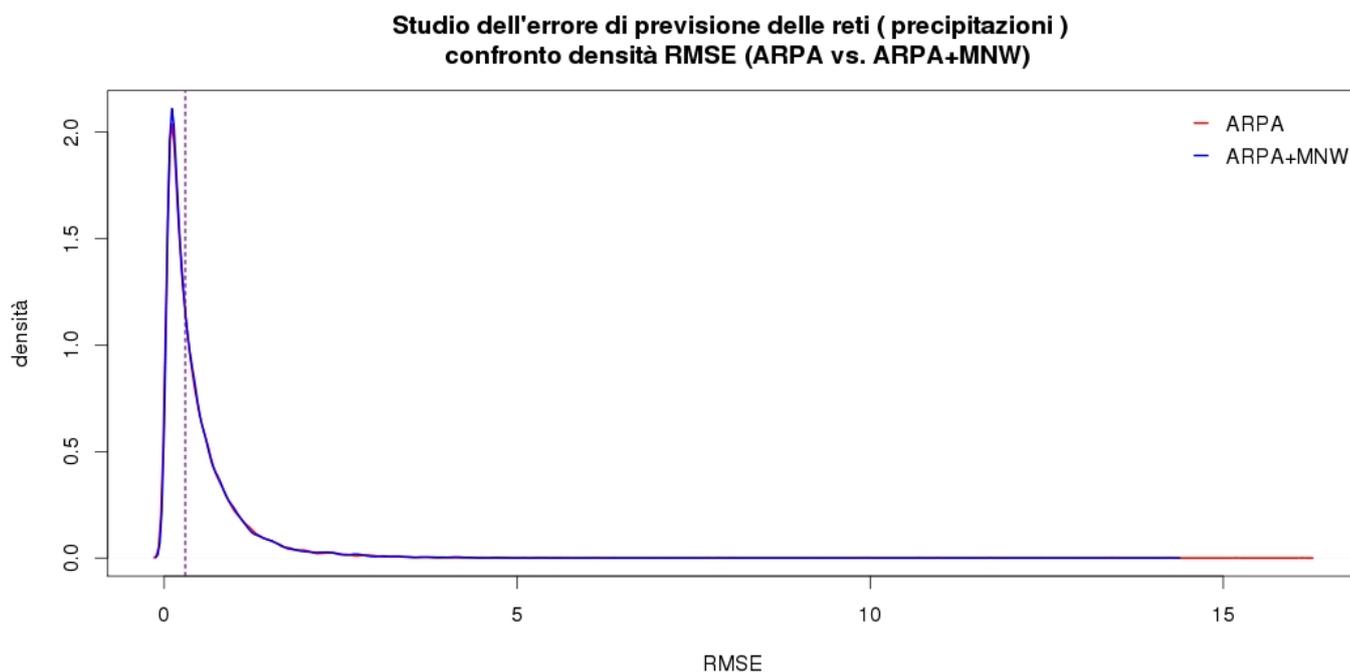


Figura 17

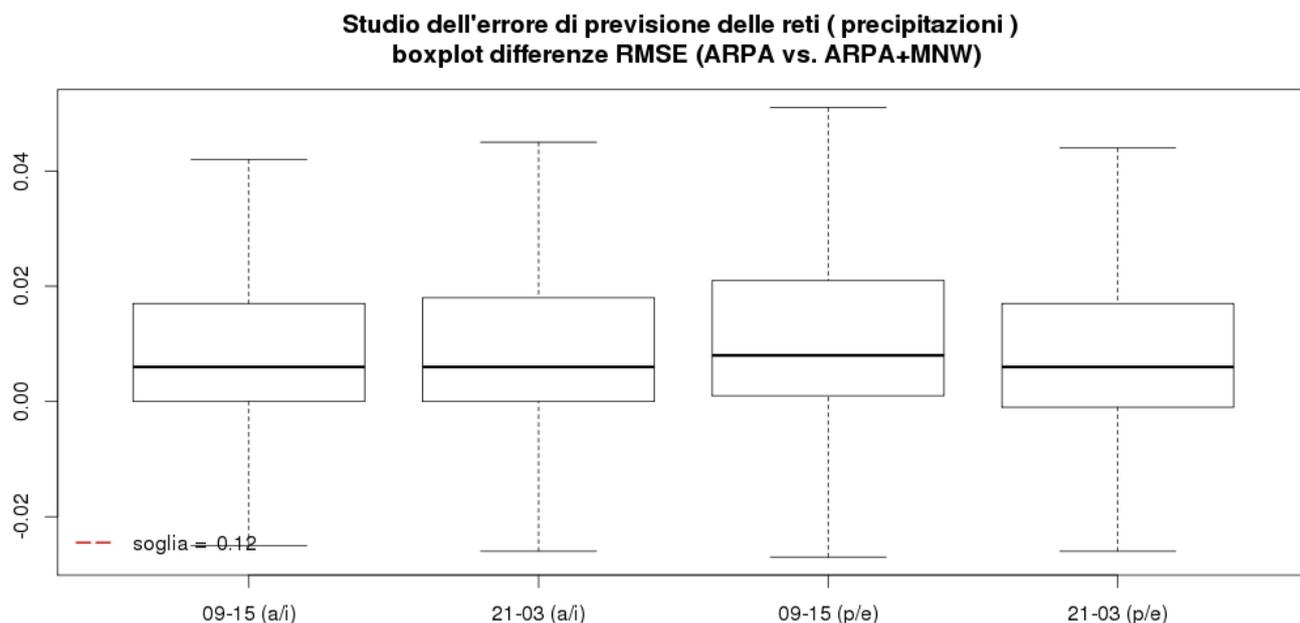


Figura 18

**Studio dell'errore di previsione delle reti (precipitazioni)
serie storica differenza RMSE (ARPA vs. ARPA+MNW)**

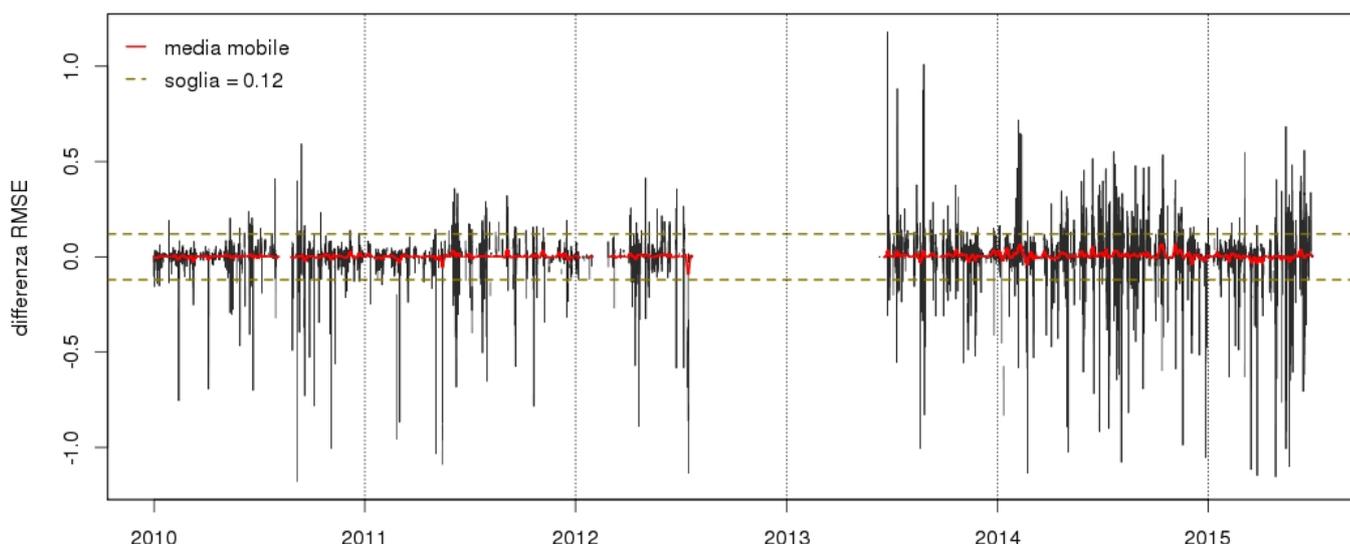


Figura 19

In Figura 20 e Figura 21 sono illustrati i boxplot dei rapporti previsioni/osservazioni delle due reti, divisi per stagione e per percentili delle osservazioni. L'aggiunta della rete MeteoNetwork non modifica sostanzialmente l'output: come ipotizzabile, le previsioni tendono a sottostimare le osservazioni, ma la rete congiunta non mostra evidenti segni di miglioramento rispetto alla sola rete RIRER.

In Tabella 1 e in Tabella 2 sono illustrate le tavole di contingenza descritte nel capitolo precedente. L'aggiunta della rete MeteoNetwork contribuisce a migliorare l'*accuratezza* globale e la *sensibilità*, mentre la *specificità* cala leggermente: in sostanza, la frazione di previsioni corrette di ore di pioggia diminuisce, mentre la frazione di previsioni corrette di ore non di pioggia aumenta. La variazione è comunque molto contenuta. Va specificato che ogni osservazione si riferisce a una specifica centralina e a un preciso intervallo temporale, su cui è osservato e in seguito stimato un valore di *precipitazione*.

Rapporti previsioni/osservazioni (autunno/inverno)

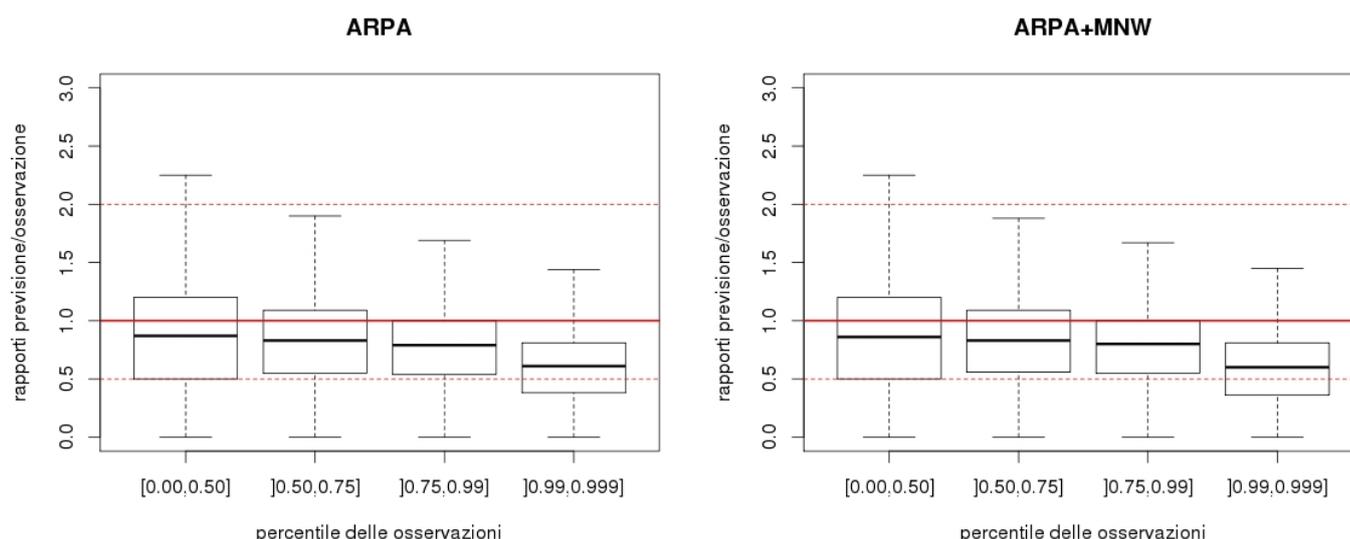


Figura 20

Rapporti previsioni/osservazioni (primavera/estate)

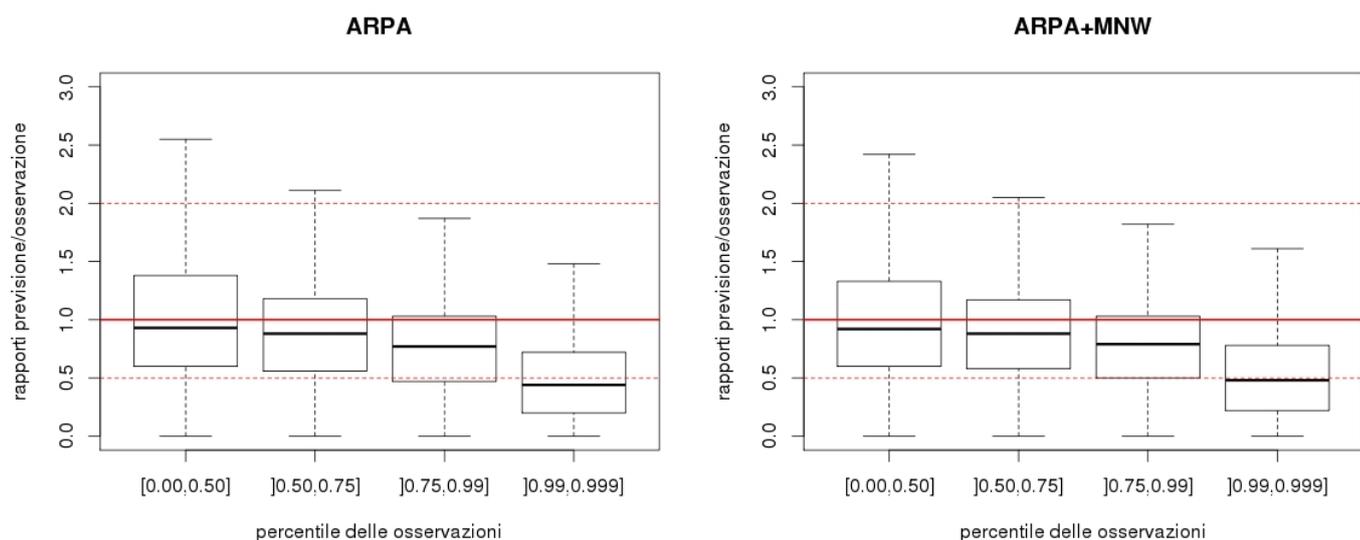


Figura 21

ARPA	Osservazione: SI PIOGGIA	Osservazione: NO PIOGGIA
Previsione: SI PIOGGIA	810103	474284
Previsione: NO PIOGGIA	98430	2878653
Accuracy	Specificity	Sensitivity
0.866	0.892	0.859

Tabella 1

ARPA+MNW	Osservazione: SI PIOGGIA	Osservazione: NO PIOGGIA
Previsione: SI PIOGGIA	909004	514719
Previsione: NO PIOGGIA	120164	3440171
Accuracy	Specificity	Sensitivity
0.873	0.883	0.870

Tabella 2

0.7 Conclusioni

In generale, le stazioni MeteoNetwork in Emilia-Romagna sembrano rappresentare coerentemente i parametri meteo superficiali *temperatura* e *precipitazioni*.

Per quanto riguarda il parametro *umidità relativa* non è invece possibile esprimere un giudizio univoco: il confronto puntuale tra stazioni ha generato risultati abbastanza soddisfacenti, ma lo studio dell'errore di previsione ha evidenziato una possibile distorsione sistematica delle osservazioni di umidità tra le stazioni MeteoNetwork e Arpae.

I **confronti tra coppie di stazioni vicine** hanno evidenziato differenze generalmente paragonabili con il solo errore complessivo di misura.

La collocazione della stazione e quindi la sua rappresentatività pare essere sostanziale e determinante rispetto all'errore di misura strumentale.

Qualità e quantità delle osservazioni MeteoNetwork sono significativamente migliorate negli ultimi due anni.

L'uso congiunto delle due reti per temperatura e precipitazione è generalmente migliorativo o al peggio neutro rispetto all'utilizzo della sola rete *Arpae*, specialmente nell'ultimo biennio osservato. Ulteriori verifiche sono necessarie per l'umidità per la quale l'uso è consigliato solo in situazioni particolari.

L'integrazione futura delle due reti è quindi consigliabile rispetto agli scopi dichiarati in questo studio.

0.8 Prospettive future

- Estensione dello studio al territorio della Regione Veneto.
- Analisi di casi studio.
- Eliminazione di errori casuali e sistematici della rete MeteoNetwork.
- Favorire la collocazione delle stazioni MeteoNetwork in posizioni strategiche e/o rappresentative.

0.9 Ringraziamenti

Si ringraziano per la collaborazione:

- CINECA e Kion:
 - Stefano Roselli
 - Virgilio Pierini
 - Cesare Bandirali
 - Alessandro Cocilova
 - Michele Cicoria
- MeteoNetwork
 - Gianandrea Peressutti
 - Giovanni Tesauro
- SIMC Arpae-Emilia Romagna:
 - Paolo Mezzasalma
 - Gabriele Antolini

0.10 Licenza

Quest'opera è distribuita con Licenza Creative Commons Attribuzione - Non commerciale - Non opere derivate 4.0 Internazionale



<http://creativecommons.org/licenses/by-nc-nd/4.0/>