

Analisi di campi di precipitazione

Benassi Federica - Ventrucci Chiara

Novembre 2021

1 Introduzione

Lo scopo dell'esperienza è confrontare campi di precipitazione ottenuti con diverse tecniche. Sono stati analizzati, in particolare, dati provenienti da satelliti, rianalisi e misure a terra, utilizzando questi ultimi come riferimento. Sono state confrontate sia le funzioni densità di probabilità e le correlazioni tra i tassi misurati o stimati sia le capacità di previsione di un evento attraverso tabelle di contingenza.

2 Dataset

I quattro dataset utilizzati provengono da istituzioni, missioni o progetti differenti:

1. dataset pluviometrico da Rete Pluviometrica Italiana (PLV), interpolato con kriging;
2. dataset di stima dalla missione NASA/JAXA GPM (GPM);
3. dataset di stima dal progetto EUMETSAT H-SAF (SAT);
4. dataset di stima dalla rianalisi ERA5 di ECMWF (ERA5).

Ciascuno di essi ha diversi tempi di campionamento e spaziature di griglia, riportati nella Tab.1.

Dataset	Precipitazione	Campionamento	Spaziatura di griglia
PLV	Cumulata (mm/h)	1 h	5 km
GPM	Rate istantaneo (mm/h)	30 min	$0.1^\circ \times 0.1^\circ$
SAT	Rate istantaneo ($\text{kg}/(\text{m}^2 \cdot \text{s})$)	15 min	$4 \div 8 \text{ km}$
ERA5	Cumulata (mm/h)	1 h	$0.25^\circ \times 0.25^\circ$

Tabella 1: Parametri caratteristici dei dataset utilizzati

Si è scelto di focalizzare l'analisi dei dati pluviometrici su una regione rettangolare posizionata tra le latitudini 44-45 °N e 9.75-12.25 °E: quest'area copre buona parte dell'Emilia Romagna. Come si può osservare dalla Fig.1, la diversa spaziatura di griglia di ogni dataset rende eterogenea la rappresentatività dell'area considerata ed è, quindi, necessario uniformare tutti i dati a una griglia comune. Ritenendo la griglia di ERA5 troppo rada per avere una riproduzione attendibile del campo di precipitazione, si è scelto un valore di 0.1° . Per ogni dataset, poi, si è proceduto assegnando ai nuovi punti di griglia il valore di tasso di precipitazione corrispondente al punto più vicino della griglia originale, utilizzando la formula della distanza di Haversine. In questo modo, si è ottenuta una griglia di 286 punti, per ogni istante temporale. L'analisi è stata svolta su un arco temporale di 24 ore, affinché fosse considerato un numero di dati sufficiente per avere consistenza statistica. Il giorno scelto è il 14 ottobre 2016. I dataset sono stati uniformati anche temporalmente, assumendo intervalli di un'ora: per GPM e SAT, le cui frequenze di campionamento sono, rispettivamente, di 30 e 15 minuti, il valore associato a una determinata ora è stato stabilito come la media tra i dati compresi tra quell'ora e la successiva. I dati sono stati sottoposti a un filtraggio preliminare di valori anomali: le stazioni pluviometriche a terra presentano, infatti, valori negativi in caso di malfunzionamento.

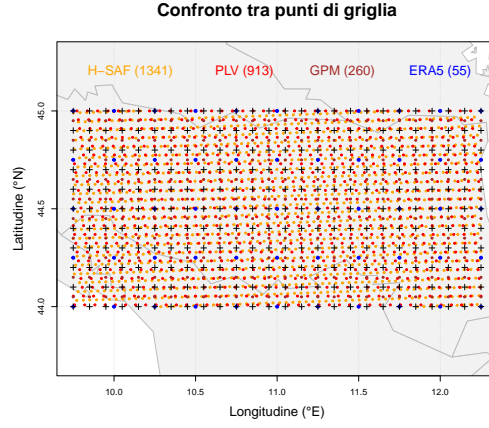


Figura 1: Confronto tra le griglie dei dataset nella regione selezionata, in nero quella comune

3 Analisi dei campi

Le differenze di rappresentazione del campo di precipitazione emergono anche da una semplice visualizzazione di ciascun dataset sulla griglia comune, come in Fig.2. I tassi di ERA5 sono mediamente più bassi, mentre c'è un buon accordo tra SAT e GPM. Le zone grigie rappresentano precipitazioni al di sotto di 0.1 mm/h o dati non disponibili. Per PLV, questo è attribuibile nel 22% dei casi ad anomalie nel funzionamento degli strumenti.

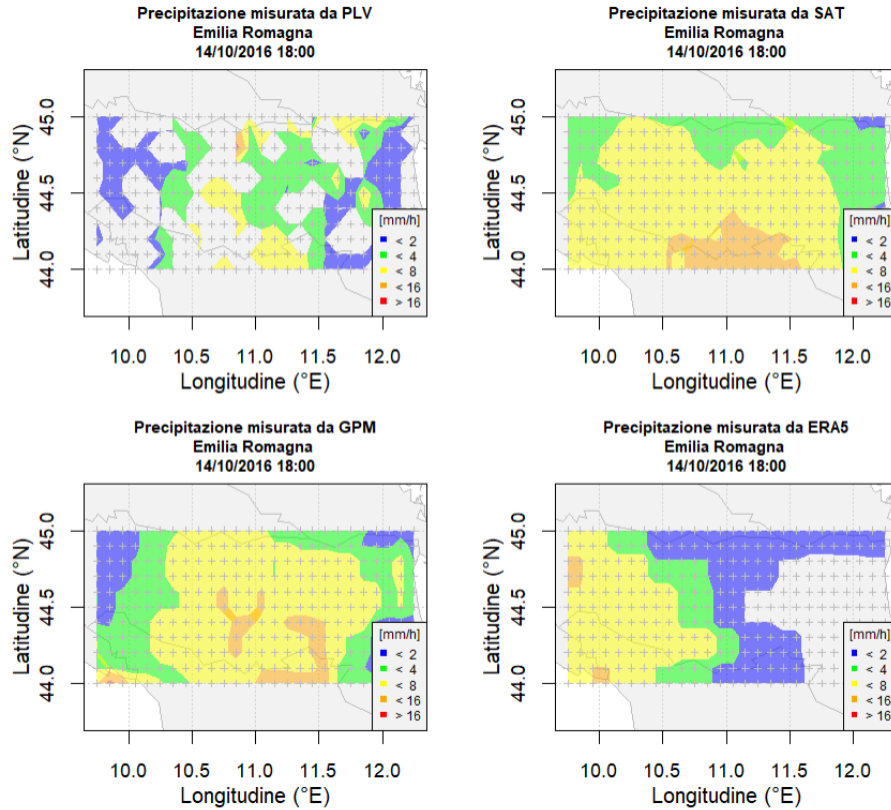


Figura 2: Precipitazione con griglia comune misurata o stimata alle 18:00

In Fig.3 sono riportate le densità di probabilità (pdf) dei dati analizzati. Si può notare che soltanto GPM riesce a coprire tutto il range di tassi di precipitazione misurato dai pluviometri a terra, mentre ERA5 e SAT raggiungono al massimo i 15 mm/h. Per gli eventi di precipitazione intorno ai 5 mm/h, sembra esserci una sovrastima da parte di SAT e sottostima da parte di ERA5. Nonostante ciò, la pdf di ERA5 è la più simile a quella di PLV.

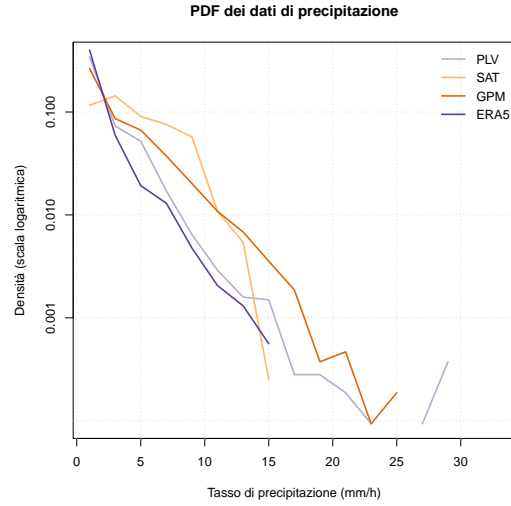


Figura 3: Funzioni densità di probabilità per i tassi di precipitazione

Nonostante non ci siano estreme differenze tra le funzioni densità di probabilità, i dati misurati o stimati nei quattro casi sono poco correlati tra loro. Questo è ben visibile nelle Fig.4 e Fig.5: c'è, infatti, molta dispersione, non solo quando si considera il comportamento rispetto al pluviometro, ma anche tra le altre coppie di dataset.

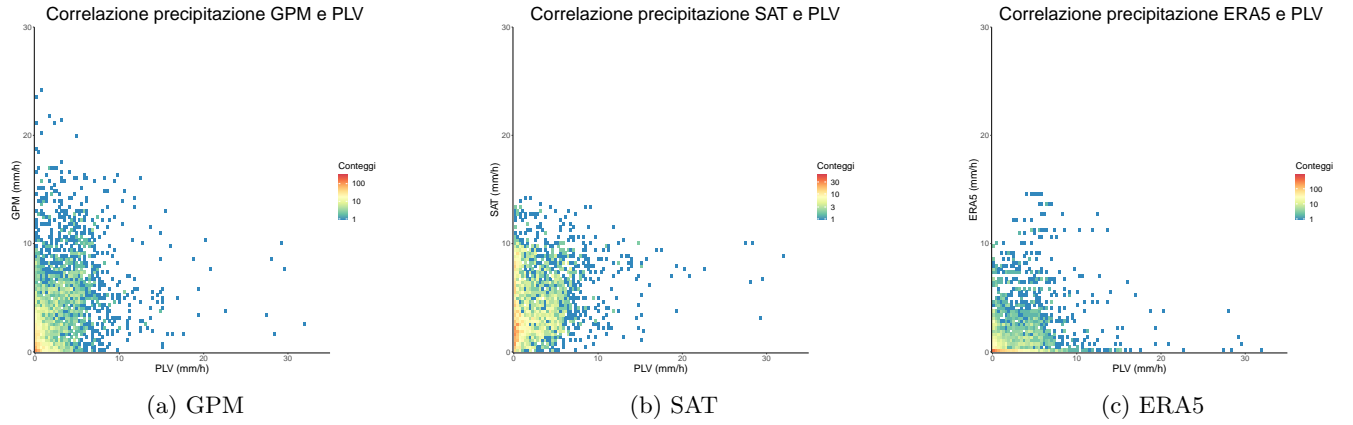


Figura 4: Correlazione tra misure di precipitazione dei pluviometri a terra e degli altri dataset

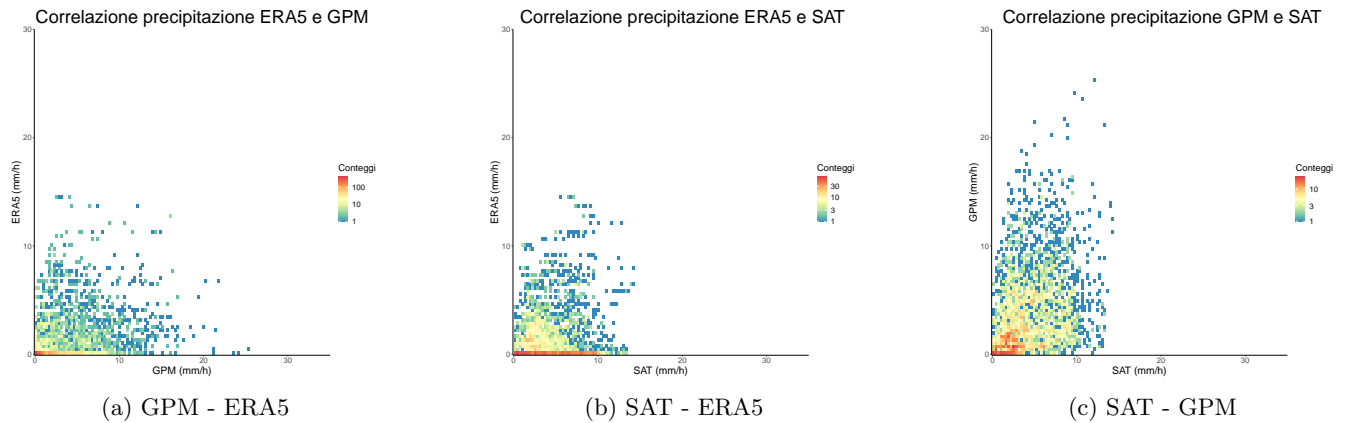


Figura 5: Correlazione tra misure di precipitazione stimate

Si possono riconoscere le tendenze precedentemente individuate: i punti dello scatterplot tra SAT ed ERA5 sono concentrati in prossimità dell'origine, dato che sono stimati soltanto i tassi di precipitazione più bassi, mentre lo scatterplot tra GPM e PLV coinvolge una regione più ampia. La maggiore dispersione dei punti si osserva negli scatterplot in cui è presente il dataset SAT: i conteggi per ciascun punto sono, infatti, mediamente più bassi.

In Fig.6 sono riportati gli indici di correlazione statistica per ogni coppia di dataset. Sono stati calcolati errore medio (ME), errore medio assoluto (MAE), deviazione standard (RMSE), coefficiente di variazione (CV) e coefficiente di correlazione lineare (r). Si può osservare che, rispetto al riferimento a terra (Fig.6a), il dataset che riporta i valori minimi per gli errori medi (ME e MAE) è ERA5; non sono, tuttavia, trascurabili, dato che l'errore medio assoluto è di quasi 2 mm/h e c'è una generale sottostima. GPM e SAT tendono invece a una sovrastima dei valori e a una differenza assoluta maggiore. In tutti i casi, il coefficiente di correlazione lineare è molto basso mentre il CV è piuttosto alto; questo può essere dovuto alla grande dispersione dei dati e ai diversi range di tassi di precipitazione presenti nei dataset. Il valore massimo del coefficiente di correlazione lineare rispetto al riferimento corrisponde a GPM: confrontando questo risultato con le pdf in Fig.3, si può osservare che GPM è l'unico dataset rappresentativo anche degli eventi di precipitazione più intensi, più rari ma rilevati dalle misure a terra. In Fig.6b sono riportati, invece, gli indici statistici tra i campi stimati. La coppia GPM-SAT presenta il valore maggiore di correlazione lineare, seppur basso, ed errori e dispersione sono minori rispetto alle altre due coppie, come già emerso dall'analisi della Fig.2.

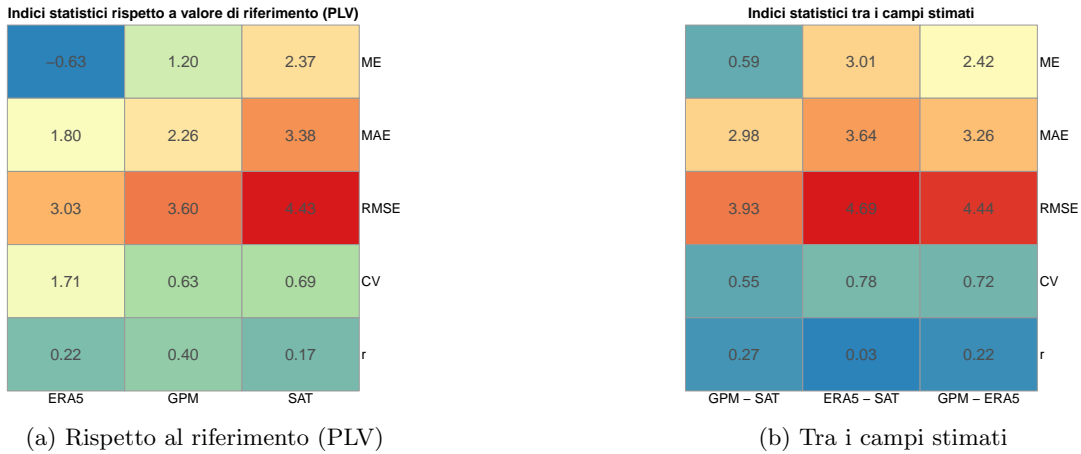


Figura 6: Tabelle di indici statistici tra i vari dataset

Il confronto tra i quattro dataset può essere esteso valutando le capacità di SAT, GPM ed ERA5 di rilevare un evento di precipitazione, potenzialmente diverse a seconda dell'intensità dell'evento stesso. È utile, quindi, costruire tabelle di contingenza, come la Tab.2, considerando valori di soglia differenti. In generale, la stima sarà buona se il numero di eventi *hit* e *correct rejection* è alto, mentre è minimo il numero di *miss* e *false alarm*.

	Evento osservato (SÌ)	Evento osservato (NO)
Evento previsto (SÌ)	3158 (<i>hit</i>)	1242 (<i>false alarm</i>)
Evento previsto (NO)	428 (<i>miss</i>)	524 (<i>correct rejection</i>)

Tabella 2: Tabella di contingenza con soglia 0.1, numero totale di punti = 5352

A partire dalle tabelle di contingenza, sono stati calcolati, quindi, i valori della coppia di indici *Probability of Detection* (POD) e *False Alarm Rate* (FAR) per ogni valore di soglia, riportati in Fig.7. L'andamento del FAR è molto simile per i tre dataset, intorno a 0.3-0.4 per precipitazioni deboli, mentre già da eventi con tassi intorno a 2.5 mm/h diventa maggiore di 0.5, fino a tendere a 1 per tassi maggiori di 10 mm/h, range nel quale ci sono molti meno dati a disposizione, sia perché gli eventi di precipitazione sono più rari sia perché né SAT né ERA5 sono in grado di segnalarli. In corrispondenza della soglia di 2.5 mm/h si osserva, invece, un crollo nell'andamento della POD per ERA5, che passa da 0.4-0.5 per tassi piccoli a poco più di 0.2, mentre le stime di SAT e GPM si mantengono buone fino a tassi di circa 5 mm/h. Per tassi maggiori o uguali a 10 mm/h, tutti i dataset tendono, quindi, a non rilevare gli eventi di precipitazione veri, ma soltanto falsi allarmi.

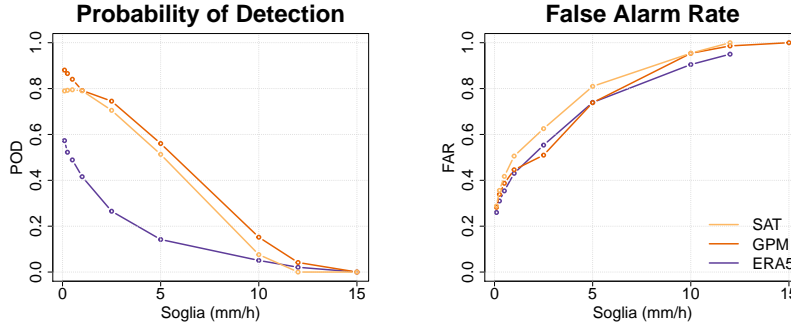


Figura 7: Andamento di POD e FAR al variare della soglia di precipitazione considerata

In Fig.8 sono riportati, invece, gli andamenti di *Percent Correct* (PC), *Critical Success Index* (CSI) e *Equitable Threat Score* (ETS). Dato che gli eventi di precipitazione hanno prevalentemente tassi piccoli, gli alti valori di PC per le soglie più grandi analizzate non sono particolarmente indicativi della bontà della stima, dato che la maggior parte degli eventi sarà un *correct rejection*. Si osserva, tuttavia, un valore per le soglie minori intorno a 0.6-0.7 per tutti i dataset. La stima di SAT tende a peggiorare per precipitazioni intorno a 1 mm/h, mentre quella di ERA5 migliora. Data la scarsa frequenza degli eventi di precipitazione intensi, sono ben giustificati i valori di CSI vicini a zero per tassi maggiori di 10 mm/h. La stima migliore appare quella di GPM, con CSI massimo pari a 0.65, almeno fino ai 5 mm/h, anche se gli andamenti di SAT ed ERA5 sono molto simili, con differenze che si mantengono quasi costanti al variare della soglia. I valori di ETS sono, invece, sempre positivi, ma molto piccoli, intorno a 0.1; il massimo valore è raggiunto da GPM per la soglia di 2.5 mm/h ed è pari a 0.25. Ci si aspetta, dunque, un comportamento migliore per eventi di precipitazione debole, dell'ordine di qualche mm/h.

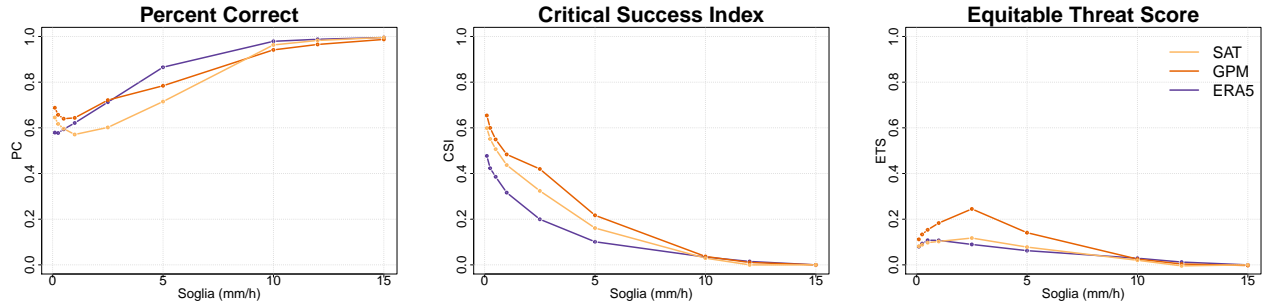


Figura 8: Andamento di PC, CSI e ETS al variare della soglia di precipitazione considerata

Un'analisi simile svolta su intervalli temporali di 3 ore mostra, invece, un comportamento nettamente peggiore di ERA5 rispetto a SAT e GPM: la POD ha, infatti, valori inferiori a 0.5 già con una soglia di 1.5 mm/h, mentre per SAT e GPM è intorno a 0.9. L'andamento del FAR è, invece, anche in questo caso, molto simile tra le tre stime, intorno a 0.3 per 1.5 mm/h.

4 Conclusioni

Nessuna delle tre stime analizzate riesce a rappresentare un evento di precipitazione in completo e fedele accordo con le misure della rete pluviometrica a terra. Dallo studio delle distribuzioni e degli errori statistici emerge, infatti, che, per precipitazioni fino a 15 mm/h, il dataset dalla rianalisi ERA5 è il più simile al riferimento pluviometrico tra le tre stime analizzate. Questo dataset non riporta, tuttavia, dati di precipitazione più intensa; il dataset dalla missione GPM, al contrario, presenta differenze maggiori con le misure a terra ma copre quasi lo stesso range. Quest'ultimo è, inoltre, il più abile a rilevare un evento di precipitazione, soprattutto con un tasso compreso tra 0 e 5 mm/h. Nonostante anche il dataset dal progetto H-SAF presenti una buona capacità di identificazione degli eventi, i valori registrati sono i più dissimili dal riferimento.