

# Analisi testuale dei bollettini meteo come strumento per validare la classificazione dei tipi di circolazione

Alfonso Crisci<sup>1,\*</sup>, Bernardo Gozzini<sup>2</sup>, Valentina Grasso<sup>1,2</sup>, Gianni Messeri<sup>1,2</sup>, Roberto Vallorani<sup>1,2</sup>, Federica Zabini<sup>1,2</sup>

1 CNR IBIMET, Via G. Caproni 8, Firenze, 50138 - 2 Consorzio LaMMA, Via Madonna del Piano, Seto Fiorentino, 50019

\* Alfonso Crisci, Tel: +055 4483 251, E-mail: a.crisci@ibimet.cnr.it

L'obiettivo di questo lavoro è quello di esplorare nuovi strumenti di indagine nel campo del trattamento dell'informazione meteorologica partendo da diverse tipologie di dati come quelli di natura testuale. Il Consorzio LaMMA (Regione Toscana-CNR), servizio meteo della Regione Toscana, ha elaborato due classificazioni dei principali schemi di circolazione atmosferica giornaliera (tipi di circolazione) su un dominio europeo, utilizzando gli strumenti sviluppati all'interno dell'azione europea COST 733. Entrambe le classificazioni sono finalizzate a spiegare rispettivamente il regime termico [SON-9] e quello pluviometrico [PCT-9] che insistono in un dato giorno sulla Toscana. In questo studio abbiamo preso in analisi la classificazione relativa ai regimi di precipitazione, realizzata applicando il metodo PCT (basato sulle analisi delle componenti principali) applicato sui dati di pressione a livello del mare ottenuti dai dati di Reanalysis 2 NCEP/NCAR per il periodo 1981-2010 [1]. La disponibilità di poter associare un inventario giornaliero del pattern di circolazione dell'atmosfera ed uno storico di prodotti testuali generati dall'attività dei previsori, rappresenta un'opportunità per valutare il grado di coerenza fra prodotti meteo di natura diversa ma orientati verso l'unico scopo di descrivere in modo efficace lo stato del tempo. Questa prima analisi ha mostrato interessanti potenzialità evidenziando bene come tipi di tempo opposti, ciclonici/anticiclonici, diano effettivamente luogo a corpus testuali sensibilmente diversi.

[1] Vallorani et al (2017), Circulation type classifications for temperature and precipitation stratification in Italy, International journal of climatology <https://doi.org/10.1002/joc.5219>

## Dati e metodi

In questo studio abbiamo analizzato le informazioni relative alle precipitazioni usando la classificazione PCT-9 per i regimi pluviometrici. L'analisi testuale ha riguardato la descrizione "stato del cielo" contenuta nei bollettini meteo emessi dal Consorzio LaMMA nel periodo 01/01/2011 - 31/01/2016.

I testi relativi a 1854 bollettini sono stati suddivisi in 9 corpora raggruppando i testi emessi nei giorni con lo stesso tipo di tempo (PCT 1-9). Tra le variabili testuali esaminate si presentano quelle relative a: lunghezza dei testi; ricchezza di vocabolario; frequenza dei termini connessi al dominio semantico della precipitazione; caratteri di unicità di ogni corpus. Sono state poi calcolate le frequenze di occorrenza di due pattern testuali polari rispettivamente al carattere spaziale delle precipitazioni [diffus\*/isolat\*] e della loro intensità [deboli/intense] e su lemmi legati alle precipitazioni solide [grandin\*,nev\*]. Le frequenze relative sono state rappresentate sia su base complessiva che su base stagionale (consultabile tramite QRcode). Le analisi sono state condotte sulla piattaforma <https://voyant-tools.org/>.

## Evidenze

Il corpus totale dei testi analizzati contiene 59.712 parole in totale, con 845 forme di parola uniche. Le parole più frequenti nel corpus sono: nuvoloso (2152); zone (1234); poco (1011); pomeriggio (845); sereno (719); mattinata (665); precipitazioni (635); addensamenti (573); nuvolosità (558); serata (547). Come mostra la tabella il volume testuale relativo allo stato del cielo varia in funzione del tipo di circolazione, fornendo una misura indiretta della complessità della situazione meteo. I grafici delle frequenze relative mostrano abbastanza chiaramente come esiste una coerenza fra la natura dei regimi pluviometrici e la frequenza di alcuni termini specifici utilizzati nei testi dei bollettini. I tipi di circolazione "ciclonica/anticiclonica" confermano e spiegano le distribuzioni delle frequenze terminologiche sia a livello complessivo (annuale), che a livello (con ulteriore accuratezza).

## Conclusioni

L'analisi dei termini linguistici connessi alla precipitazione ha mostrato come la classificazione in tipi di tempo delinei frequenze molto diverse dei termini stessi all'interno dei bollettini. Molto frequenti nei corpus relativi a circolazioni cicloniche, i termini connessi alla precipitazione riducono drasticamente la loro frequenza negli altri corpus. L'analisi testuale diventa un metodo per descrivere e fornire ulteriori elementi per validare una classificazione dei tipi di circolazione finalizzata alla descrizione di un clima su un dominio spaziale limitato come quello della Toscana.

### dati testuali

1854 bollettini

9 corpora

59.712 parole

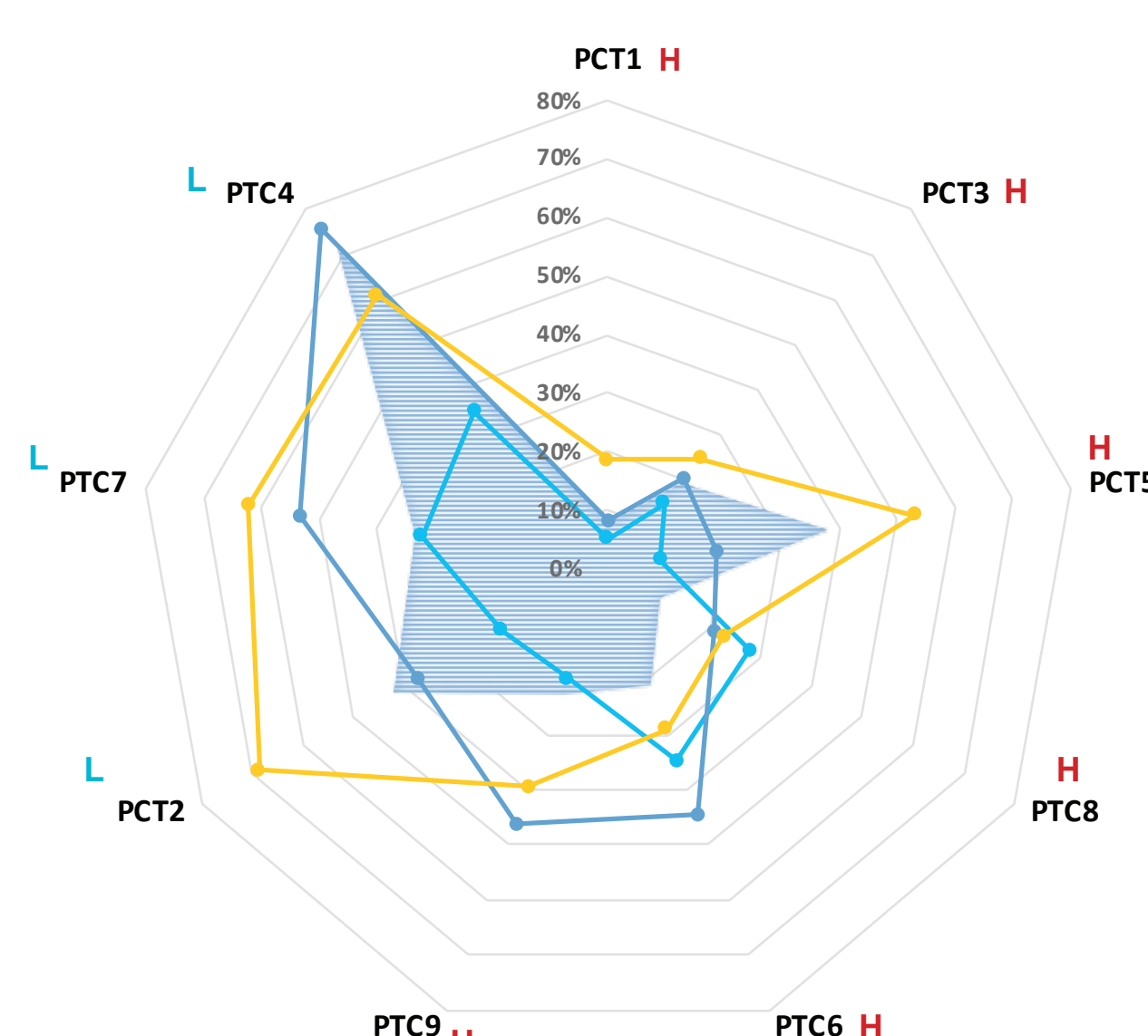
845 parole uniche

Tipi di circolazione	N. bollettini corpus	parole del corpus	densità del vocabolario	Parole connesse alla precipitazione *
PCT1	284	6702	0.067	133
PCT2	249	9112	0.057	478
PCT3	183	5073	0.078	152
PCT4	162	6338	0.069	469
PCT5	177	5471	0.080	217
PCT6	245	8566	0.057	346
PCT7	90	3540	0.110 (+)	198
PCT8	158	4545	0.081	130
PCT9	306	10265	0.053 (-)	516

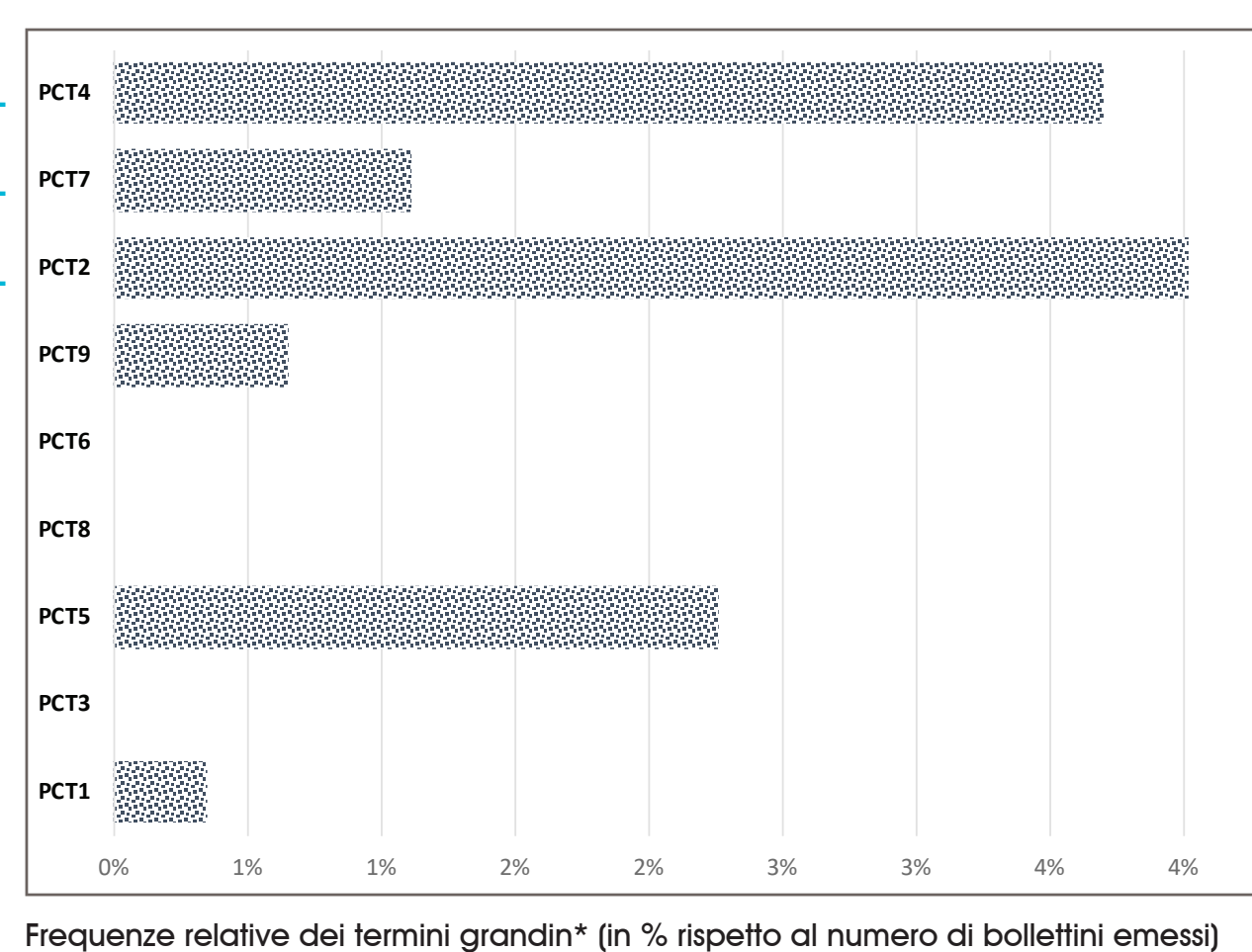
\*Sono state analizzate le parole relative alle precipitazioni con i seguenti prefissi: piogg\*, rovesci\*, temporal\*, precipitazione\*, nev\*, grandin\*

### Analisi della frequenza dei termini relativi alla pioggia per ciascun tipo di circolazione

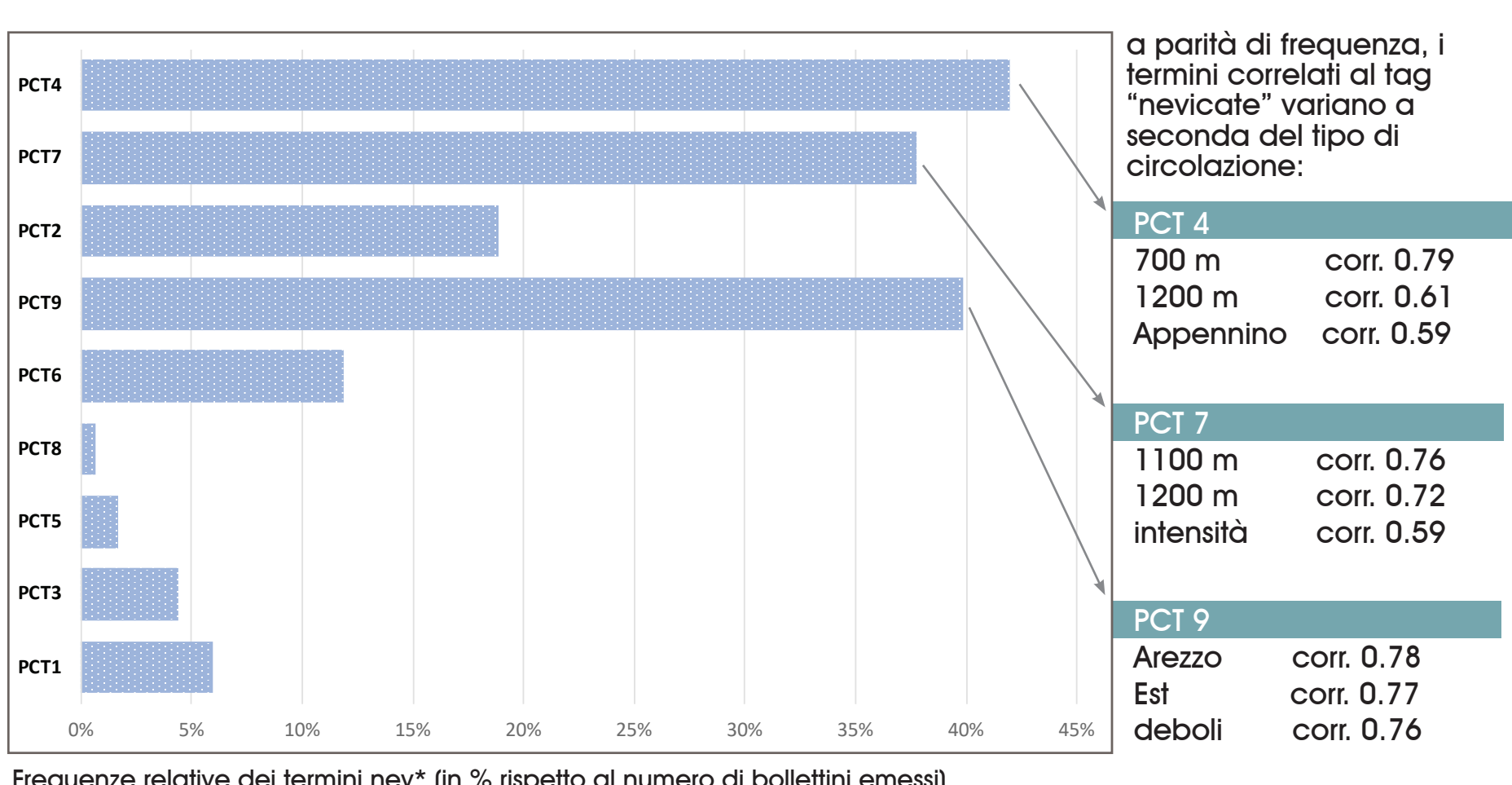
temporal\* // piogg\* // precipitazione\* // rovesci\*



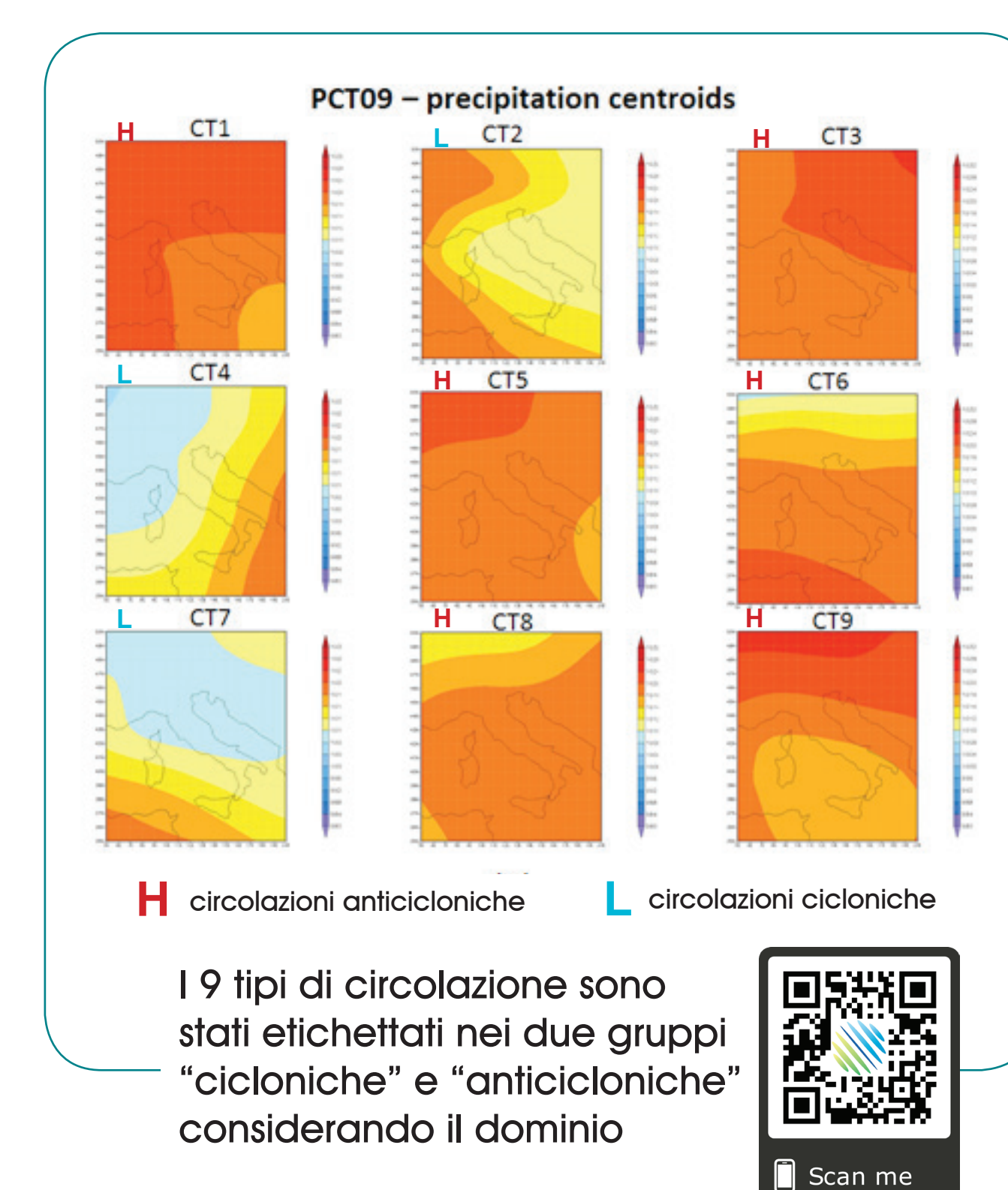
### quando grandina?



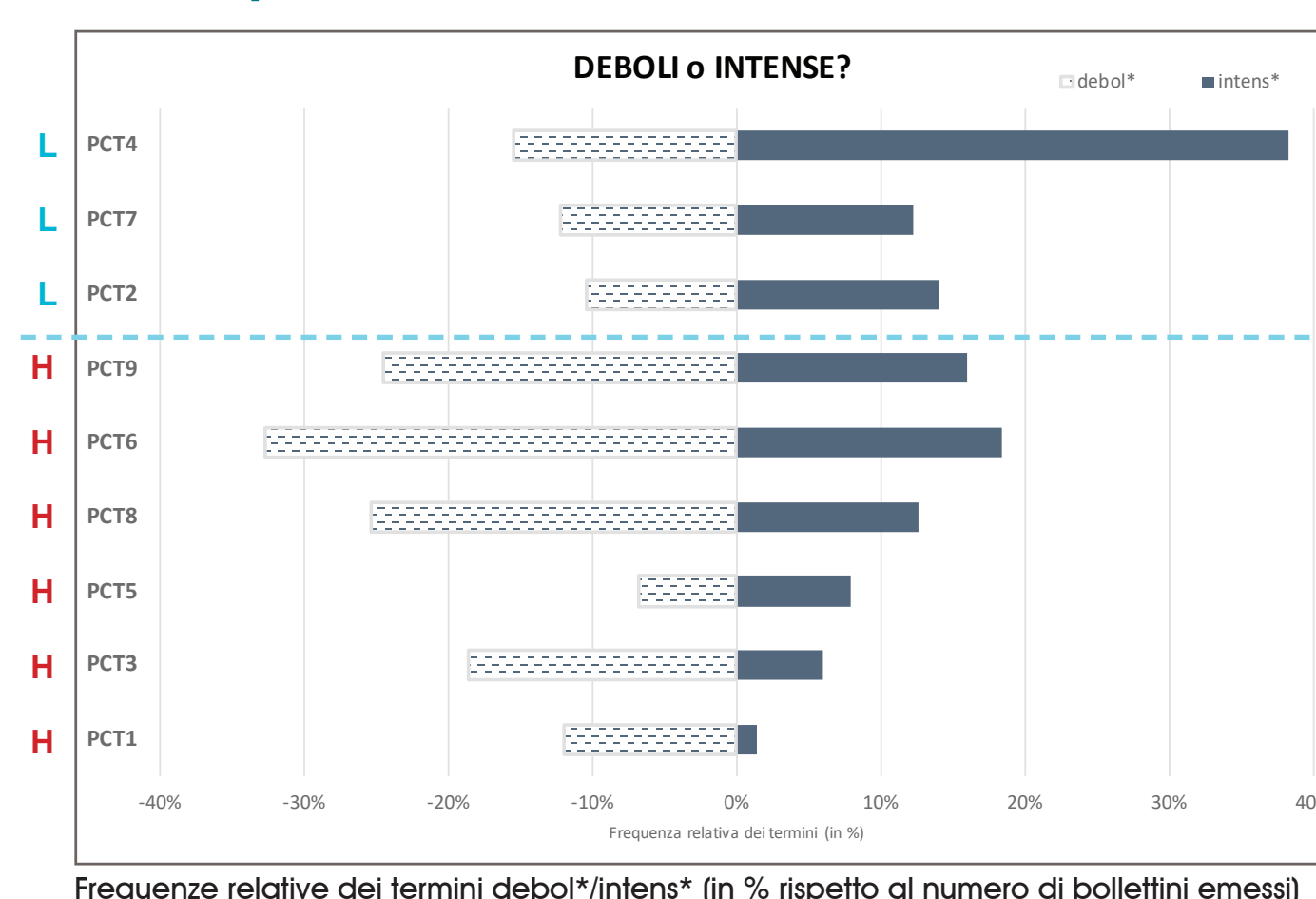
### quando nevica?



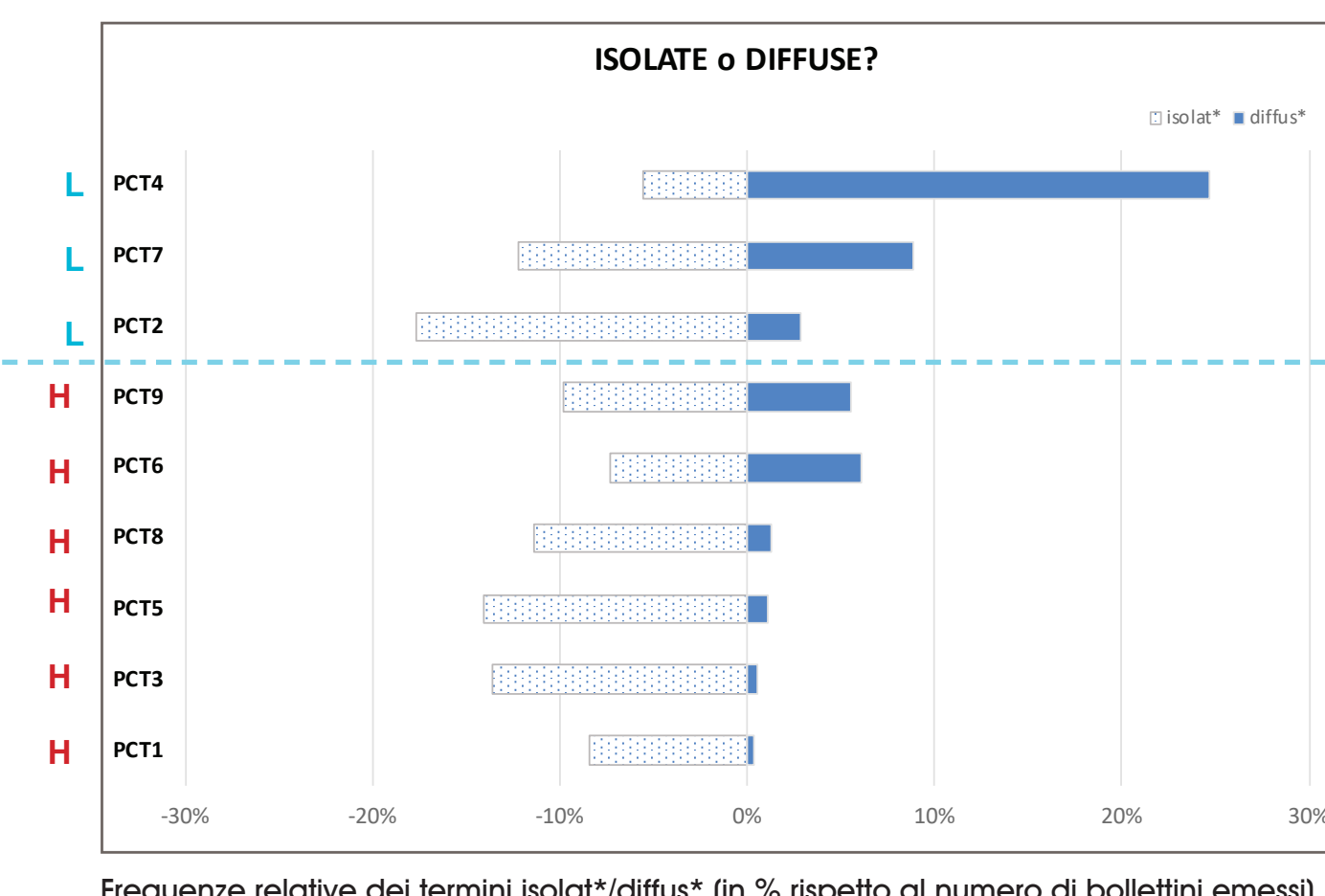
### dati climatici



### Precipitazioni deboli/intense



### Precipitazioni isolate/diffuse



### WEB REPOSITORY e DATI

I dati e i codici di lavoro sono disponibili in repository su GitHub ConsorzioLaMMA

Consulta i grafici di frequenza per ogni stagione