



Previsioni numeriche della concentrazione pollinica in Emilia-Romagna

Il complesso sistema di monitoraggio dei pollini allergenici in funzione da anni nel territorio della nostra regione ha consentito di intraprendere un'attività previsionale altrettanto complessa e importante. Con la previsione numerica si cerca di individuare la classe di presenza giornaliera dei pollini d'interesse. È ovvio che questo risultato è alla base di ogni efficace azione di prevenzione.

32

La sempre maggiore incidenza di malattie respiratorie legate alla presenza di pollini nell'aria durante tutto il corso dell'anno rappresenta un aspetto molto rilevante di sanità pubblica. Il monitoraggio qualitativo e quantitativo della presenza pollinica in un dato luogo e la comunicazione dell'informazione a tutti i soggetti interessati sono tematiche ben radicate nelle attività che Arpa Emilia-Romagna (Arpa-ER) dedica a questi argomenti. In questo ambito, anche l'attività legata alla previsione dell'andamento delle concentrazioni polliniche giornaliere può rappresentare il punto di partenza per un contributo alla prevenzione delle patologie inerenti alla presenza dei pollini in atmosfera.

L'attività di monitoraggio pollinico intrapresa da Arpa-ER ormai da numerosi anni consente di avere a disposizione un database omogeneo e abbastanza lungo di dati di concentrazione pollinica, che consentono di implementare un *modello statistico* per la previsione pollinica: si tratta sostanzialmente di un modello basato su un'analisi statistica dei dati, in

grado di fornire una stima del dato pollinico giornaliero partendo da serie temporali di osservazioni della concentrazione pollinica giornaliera e del dato meteorologico corrispondente su un periodo temporale abbastanza lungo in modo da poter costruire una relazione statistica tra le due tipologie di dato sufficientemente robusta. Per la nostra analisi sono disponibili 6 anni solari completi (dal 1/1/2000 al 31/12/2005) dei dati di 4 famiglie botaniche che hanno una notevole diffusione nel nostro ambito territoriale (Composite, Cupressacee-Taxacee, Graminacee, Urticacee) riferiti a 12 stazioni di monitoraggio afferenti alla rete regionale dell'Emilia-Romagna. L'informazione meteorologica relativa alla stessa località in cui è presente la stazione di monitoraggio pollinico comprende le serie giornaliere di temperatura dell'aria a 2 metri (minima, media e massima), di umidità relativa (minima, media e massima) e di precipitazione cumulata nell'arco della giornata. A queste variabili meteorologiche "primarie" vengono affiancate

serie giornaliere di altre grandezze, calcolate a partire da queste osservabili di base. Nello specifico, le grandezze "secondarie" comprendono la sommatoria termica dei gradi in eccesso rispetto a varie soglie, il numero di giorni che presentano una precipitazione superiore a una certa soglia, il numero di ore giornaliere in cui la temperatura massima eccede una certa soglia (calcolato usando una formula empirica che dipende dalle temperature massima e minima), l'evapotraspirazione potenziale giornaliera e cumulata (basata anch'essa su una formula empirica) e il deficit idrico, che rappresenta la differenza giornaliera tra l'evapotraspirazione potenziale e la precipitazione. Le sommatorie e i conteggi di queste grandezze vengono fatte partire dal 1 gennaio.

L'approccio statistico che è stato scelto è basato su una *rete neurale*, modello di tipo non-parametrico che ha raggiunto un notevole sviluppo in particolare negli ultimi anni in svariati campi di applicazione, in quanto si tratta di uno strumento abbastanza robusto

dal punto di vista statistico (purché si abbiano a disposizione osservazioni in quantità sufficiente) specialmente quando le relazioni tra le variabili di input e di output non sono lineari. Lo svantaggio principale consiste nel comportamento tipo *black-box* del modello, che causa una certa difficoltà nel determinare la topologia ottimale della rete. La struttura della rete, infatti, distribuisce l'informazione fra le unità che la costituiscono e modula i vari contributi modificando i pesi relativi alle connessioni tra le unità in modo interattivo, fino a che non si supera una prescritta tolleranza: in questo modo, le variazioni dei pesi amplificano o riducono l'importanza dei predittori osservati sul risultato previsionale della rete.

Le serie storiche delle osservazioni disponibili vengono suddivise in un *periodo di addestramento* della rete, in cui vengono calcolati i pesi che collegano le varie unità della rete neurale, e il *periodo di test*, in cui vengono applicati i pesi alle configurazioni delle variabili che non sono state utilizzate per l'addestramento



della rete. Per l'addestramento della rete si usa un sottoinsieme dato da un campione dell'80% del numero totale dei giorni disponibili nel periodo dal 2000 al 2004 (per ridurre problemi di over-fitting); l'anno 2005 viene usato come periodo di test della previsione vera e propria, supponendo che la serie giornaliera dei dati meteorologici dell'anno sia una previsione, in un approccio di tipo *perfect-prog*. Tutti i predittori introdotti nella rete neurale sono normalizzati rispetto all'intervallo definito dalla differenza tra il valore massimo e il valore minimo del predittore stesso, in modo da effettuare l'addestramento della rete con grandezze non-dimensionali; un'altra precauzione adottata riguarda il fatto che alle serie giornaliere di concentrazione pollinica viene applicata una media mobile su 7 giorni in modo da ridurre l'impatto negativo dovuto a eventuali variazioni troppo pronunciate su giorni vicini, specialmente in presenza di serie storiche relativamente brevi. Il modello prevede anche l'inserimento di una componente di autocorrelazione, che è rappresentata dalla concentrazione pollinica osservata nell'arco dei 7 giorni della settimana precedente a quella cui si riferisce la previsione, sulla base della tempistica legata alla produzione operativa del bollettino pollinico di Arpa-ER.

La previsione giornaliera dell'andamento della concentrazione pollinica di una data famiglia per-

mette di ricavare automaticamente la classe di concentrazione giornaliera corrispondente (assente, bassa, media e alta, sulla base della classificazione elaborata da ISAO-CNR, con diversi estremi a seconda della famiglia). In questo modo è possibile introdurre anche una tabella di contingenza previsto-osservato tra le varie classi, sia per mettere in evidenza la capacità o meno della rete neurale di ripartire correttamente i giorni fra le diverse classi di concentrazione, sia per evidenziare eventuali comportamenti sistematici (*bias*) della previsione rispetto all'osservazione. Complessivamente, i risultati ottenuti mostrano una tendenza alla sottostima del valore osservato da parte del valore previsto con la rete neurale (comportamento riscontrabile soprattutto per la famiglia delle Composite), sia per il valore di picco della stagione pollinica, che per gli altri massimi relativi eventualmente presenti nel corso della stagione in alcune località. D'altra parte, va sottolineato che dal punto di vista dell'impatto sulla salute umana l'informazione relativa al picco stagionale risulta meno essenziale rispetto alla distribuzione dei giorni nelle classi di concentrazione: soprattutto gli episodi acuti (serie di giorni consecutivi caratterizzati da una classe di concentrazione elevata) nell'arco della stagione rivestono una notevole importanza come informazione da segnalare in anticipo ai soggetti interessati. La rete

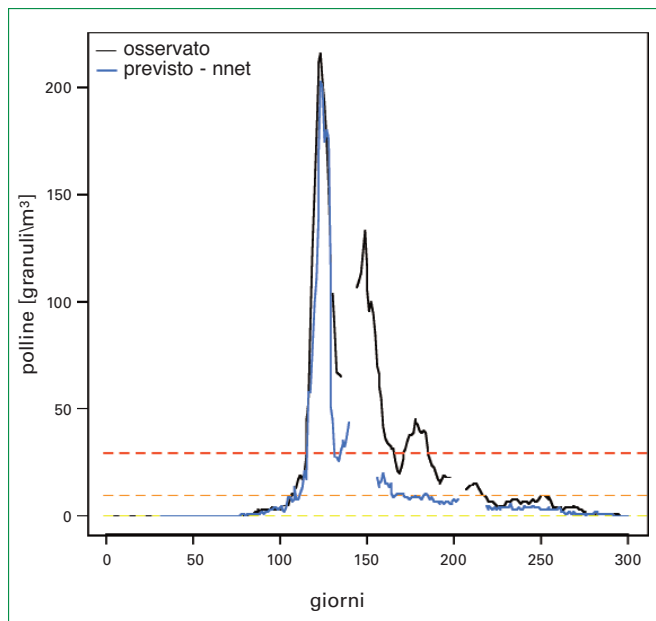


Fig. 1 - Andamento della concentrazione pollinica delle Graminacee nella stazione di Modena (mo1) nel corso della stagione pollinica 2005. L'asse delle x indica il giorno progressivo a partire dal 1 gennaio e le linee tratteggiate indicano rispettivamente l'estremo inferiore della concentrazione bassa (in giallo), media (in arancione) e alta (in rosso)

neurale mostra incoraggianti segnali nell'ottica della rappresentazione di queste situazioni, con particolare riferimento a quelle che avvengono dopo la fase di "innesco" della stagione pollinica; per la famiglia delle Graminacee si hanno buoni risultati durante l'intero arco della stagione (vedi figura). Un ulteriore elemento da sottolineare riguarda il fatto che la famiglia delle Cupressacee/Taxacee mostra risultati abbastanza diversi a seconda della località in esame: una possibile spiegazione può essere legata al fatto che le piante di questa famiglia fioriscono all'inizio dell'anno solare, per cui sarebbe più indicato introdurre variabili meteorologiche primarie e secondarie a partire dall'autunno precedente alla fioritura. Va ricordato, comunque, che, stante la disponibilità dei dati dal 1/1/2000, questo criterio rende di fatto non utilizzabile un'intera stagione, riducendo, di conseguenza, anche la robustezza statistica della procedura.

Inoltre, la disponibilità della classe di concentrazione giornaliera prevista e osservata consente di definire un criterio di inizio, picco e fine della stagione

pollinica; in questo modo si può determinare il ritardo o l'anticipo della stagione "prevista" rispetto a quella "osservata". Per la famiglia delle Composite (che presenta valori relativamente bassi di concentrazione anche per le classi di concentrazione media e alta), in alcune località l'inizio della stagione pollinica viene notevolmente anticipato, cosa che dipende con molta probabilità dalla presenza di instabilità numerica della rete neurale, specialmente prima dell'innesco della stagione; questa famiglia, infatti, presenta i valori più elevati in estate, piuttosto in avanti rispetto al punto di partenza del modello il 1 gennaio. Migliori risultati si ottengono sia per la famiglia delle Urticacee che, in particolar modo, per quella delle Graminacee: in numerose stazioni, infatti, si riescono a ottenere differenze minime sia sull'inizio che sul picco della stagione pollinica.

*Stefano Marchesi
Andrea Ranzi
Paolo Lauriola
Arpa Emilia Romagna*



FOTO ARCH. APPARIVISTA