

BREVE NOTA ESPLICATIVA

Modello Previsionale Lineare Multiregressivo – CIMAT-CLIMATEMONITOR

Con questo approccio si sfrutta l'opportunità di determinare la dipendenza di una variabile Y rispetto ad un insieme di n variabili indipendenti dette regressori definendo la seguente funzione:

$$Y = f(X_1 \dots X_n)$$

Quindi la variabile dipendente Y è funzione delle variabili indipendenti $(X_1 \dots X_n)$ e quindi dell'influenza di ogni singolo regressore.

Possiamo quindi esprimere l'equazione come:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n$$

Dal momento che possono esserci infinite altre variabili che possono influenzare Y dobbiamo aggiungere una variabile definita ε che riassume tutte le variabili che possono influire su Y ma non sono descritte in $Y = f(X_1 \dots X_n)$

L'equazione del modello quindi diviene:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + \varepsilon$$

dove:

β_0 = intercetta

$\beta_1, \beta_2, \beta_n$ = coefficienti di regressione

ε = rappresenta l'effetto di tutte le variabili non espresse in $(X_1 \dots X_n)$ e può essere positivo o negativo ed è indipendente dai valori dei regressori.

Nel caso di previsione applicata alle variazioni climatiche dobbiamo assumere che tra la variabile dipendente e ciascuna variabile esplicativa vi sia una relazione lineare. Ciò semplifica la nostra funzione introducendo però errori indotti da una pseudo non linearità del sistema ma che comunque possiamo assumere nella variabile di errore ε .

Quindi definiamo che $(X_1 \dots X_n)$ siano variabili deterministiche e per scelta o per obbligo siano prive di errore.

La difficoltà è poi quella di individuare le variabili che influenzano più direttamente Y .

Per esempio se associamo ad Y l'indice AO dobbiamo a priori individuare le variabili, quali il vento zonale a varie quote piuttosto che la pressione sul livello del mare o altri indici teleconnettivi, che nel loro insieme e nei rispettivi lag temporali fungano da variabili indipendenti dell'equazione di regressione lineare multipla.

Ne consegue che il lavoro di rappresentazione numerica deve essere necessariamente preceduto da un intenso lavoro di ricerca tra i vari dataset delle varie variabili e confrontati tra loro.