

## Spazializzazione di dati climatici a livello nazionale tramite modelli regressivi localizzati

Blasi C<sup>(1)</sup>, Chirici G\*<sup>(2)</sup>, Corona P<sup>(3)</sup>, Marchetti M<sup>(2)</sup>, Maselli F<sup>(4)</sup>, Puletti N<sup>(5)</sup>

(1) Dipartimento di Biologia Vegetale, Università "La Sapienza" Roma, P.le Aldo Moro 5 - 00185 Roma; (2) Dipartimento di Scienze e Tecnologie per l'Ambiente e il Territorio, Università del Molise, 86090 Pesche (IS); (3) Dipartimento dell'Ambiente Forestale e delle sue Risorse, Università della Tuscia, Viterbo (VT); (4) CNR - Istituto di Biometeorologia, Via Madonna del Piano 10, 50019 Sesto Fiorentino, Firenze; (5) Dipartimento di Scienze e Tecnologie Ambientali Forestali, Università degli Studi di Firenze. Via S. Bonaventura 13, I-50145 Firenze (Italy). - \*Corresponding author: Gherardo Chirici (gherardo.chirici@unimol.it).

**Abstract:** *Spatialization of climatic data at the Italian national level by local regressive models.* The availability of spatialised climatic data is an essential pre-requisite for the implementation of GIS-based analysis in many application fields. Among the different methodologies for the spatialization of climatic data collected in weather-stations the most used are those based on geostatistical approaches, on parametric correlative models or on neural networks. Within the "Completamento delle Conoscenze Naturalistiche di Base" project, funded by the Italian Ministry for the Environment (Department of Nature Protection) a database of 403 weather-stations distributed across Italy with a time series of thirty years was collected. Data of mean monthly temperature (minimum and maximum) and rainfalls were spatialized by a local linear univariate regressive method based on elevation as independent variable. A total of 36 monthly maps with a geometric resolution of 250 m was generated. The present paper introduces the adopted methodology and the accuracy results estimated by leave-one-out cross validation.

**Keywords:** Climate, Meteorological data, Spatialization, Regressive models, Italy.

*Received: Feb 07, 2007 - Accepted: Apr 30, 2007.*

**Citation:** Blasi C, Chirici G, Corona P, Marchetti M, Maselli F, Puletti N, 2007. Spazializzazione di dati climatici a livello nazionale tramite modelli regressivi localizzati. *Forest@* 4 (2): 213-219. [online] URL: <http://www.sisef.it/>.

### Introduzione

La pianificazione delle risorse territoriali e ambientali necessita, per una corretta impostazione delle più idonee opzioni gestionali, di una rilevante mole di dati georeferenziati che al contempo abbiano elevata qualità, sia tematica che geometrica, e sufficiente livello di omogeneità rispetto agli standard di riferimento esistenti.

L'influenza dei fattori climatici sulla diffusione dei diversi biomi terrestri è fenomeno noto in ecologia, alle diverse scale di riferimento. Nel tempo sono stati proposti vari tentativi di classificazione al fine di inquadrare sistematicamente e di individuare l'estensione geografica dei diversi climi. Con specifico riferimento allo studio delle relazioni tra clima e vegetazione i sistemi di classificazione maggiormente utilizzati sono su base "causale". In questa categoria

rientrano anche le classificazioni fitoclimatiche (Pavari 1916, De Philippis 1937, Rivas-Martinez 1995, Blasi et al. 2001), che distinguono le diverse classi in funzione, prevalentemente, di valori soglia di temperature e precipitazioni o di indici da essi derivati.

Le operazioni di raccolta e standardizzazione di dati utili per la derivazione di informazioni climatiche sono tanto più onerose quanto più dettagliata vuole essere la rappresentazione della variabilità del fenomeno a livello geografico.

La possibilità di spazializzare i dati meteo-climatici acquisiti puntualmente sul territorio per derivarne mappe tematiche permette di ottenere un'immagine sinottica del fenomeno sul territorio stesso. Per questo motivo tali strati informativi risultano particolarmente apprezzati nei più diversi settori applicativi della pianificazione ambientale.

In letteratura sono reperibili molteplici procedure che permettono di inferire a intere superfici valori meteoroclimatici puntuali quali, appunto, la temperatura e la precipitazione acquisite da stazioni meteorologiche (Maracchi et al. 2005, Wackernagel 1995). Da citare sono i metodi stocastici afferenti al settore della geostatistica come *kriging* e *cokriging* (Olea 1999) e quelli basati sull'uso di reti neurali (Gurney 2003).

Nel presente lavoro viene presentata una procedura innovativa per la spazializzazione di variabili climatiche a livello nazionale, basata sull'utilizzo del metodo della regressione lineare pesata localmente (Cleveland & Devlin 1988).

### La base di dati

L'area di indagine interessa tutto il territorio italiano e si estende tra 36° e 47° 30' di latitudine nord e tra 5° 30' e 18° 30' di longitudine est.

La quota s.l.m., parametro che la metodologia adotta come variabile indipendente nel metodo regressivo, fa riferimento a un modello digitale del terreno dell'Italia con pixel di 250 m ottenuto per ricampionamento di un DEM (*Digital Elevation Model*) con passo originario di 75 m. Il sistema di proiezione considerato è UTM fuso 32 nord su datum WGS84.

La base dati meteorologica è costituita da misurazioni giornaliere relative a temperature minime, massime e precipitazioni acquisite per un periodo di circa trenta anni (periodo 1950-1980) in 403 stazioni termopluviometriche dislocate sul territorio nazionale (Fig. 1). La base di dati è stata acquisita nell'ambito del progetto "Completamento delle Conoscenze Naturalistiche di base" (Unità di Ricerca del Dipartimento di Biologia Vegetale dell'Università "La Sapienza" e Istituto di Ecologia ed Idrologia Forestale del Consiglio Nazionale delle Ricerche di Cosenza, responsabili scientifici C. Blasi e F. Iovino - Blasi 2003). L'elaborazione dei dati grezzi ha permesso di colmare i dati mancanti con una procedura di *gap filling* basata sulla correlazione locale tra stazioni meteorologiche vicine (Falge et al. 2001) e quindi di calcolare per ogni stazione i valori climatici medi mensili riferiti alle tre variabili osservate. I dati aggregati mensili sono stati calcolati nel seguente modo: la temperatura media massima mensile come media dei massimi giornalieri; la temperatura media minima mensile come media dei minimi giornalieri; le precipitazioni mensili come somma delle precipitazioni giornaliere; i dati mensili sono stati quindi mediati lungo tutto il periodo trentennale della serie temporale disponibile.

### Metodo di spazializzazione

Il sistema di spazializzazione proposto trova il proprio fondamento sulla calibrazione a livello locale (cioè pixel per pixel) dei parametri di una regressione lineare (medie e matrici di varianza e covarianza) tra le diverse variabili dipendenti (temperature medie minima e massima, precipitazioni mensili) con la variabile indipendente quota estratta dal modello digitale del terreno. A ogni pixel dell'immagine corrisponde quindi una propria specifica equazione che predice la variabile dipendente in funzione della quota.

L'applicazione di procedure di regressione univariata o multivariata su superfici relativamente ampie per la spazializzazione di attributi meteo-climatici è fondata sull'assunto di omogeneità che prevede la sostanziale costanza della relazione che lega la variabile dipendente oggetto della spazializzazione con le variabili indipendenti considerate (Anderson 1984). Tale assunto trova evidenti limiti applicativi sul territorio italiano a causa dell'accentuata eterogeneità spaziale del fenomeno qui indagato (Lacaze et al. 1996, Maselli et al. 1998). Per questi motivi le procedure calibrate localmente permettono di aumentare significativamente l'accuratezza e l'affidabilità statistica delle procedure di spazializzazione (Maselli 2002).

I modelli di regressione localmente calibrata si fondano sullo stesso principio dei metodi stocastici geostatistici come *kriging* e *cokriging* (Cleveland & Devlin 1988, Wackernagel 1998): "le cose tra loro più vicine sono più simili rispetto alle cose tra loro più lontane" (prima legge della geografia di Tobler).

Tradizionalmente le procedure di regressione lineare multivariata descrivono l'interrelazione tra la variabile dipendente e una o più variabili indipendenti determinando i vettori della media e delle matrici di varianza/covarianza (Anderson 1984). Questi parametri derivano dall'analisi di un set di training nei quali sia il valore della variabile dipendente che di quella/e indipendente/i sono noti.

Il modello per la spazializzazione qui proposto prende spunto dalla teoria del *fuzzy set* (Zadeh 1965) secondo la quale le equazioni di regressione possono essere ottenute assegnando un peso ad ogni elemento del set di training (Wang 1990a, Wang 1990b).

In particolare, la media *fuzzy* ( $Mean_i^*$ ) di ogni  $i$ -esima variabile (dipendente o indipendente), può essere quantificata nel seguente modo (eqn. 1):

$$Mean_i^* = \left( \sum_{x=1}^n W_x V_x \right) / \left( \sum_{x=1}^n W_x \right)$$

**Fig. 1** - Distribuzione spaziale delle 403 stazioni termopluviometriche utilizzate nel progetto.



dove le sommatorie vengono applicate per ogni elemento del campione,  $V_{xi}$  è il valore della variabile per l'elemento campionario  $x$ , mentre  $W_x$  è il peso dato all'elemento campionario. Analogamente, ogni elemento delle matrici di varianza/covarianza fuzzy ( $Cov_{ij}$ ) può essere trovato come segue (Wang 1990a, Wang 1990b - eqn. 2):

$$Cov_{ij} = \frac{\sum_{i=1}^n W_x \cdot (V_{x_i} - Mean_i^*) \cdot (V_{x_i} - Mean_i^*)}{\sum_{i=1}^n W_x}$$

Da un punto di vista operativo per ciascun pixel per il quale debba essere stimato il valore ignoto della variabile dipendente climatica viene calcolato un modello di regressione lineare univariata usando, di tutto il set di training costituito dalle informazioni acquisite nelle 403 stazioni meteorologiche disponibili, preferenzialmente quelle più vicine al pixel in esame (Maselli 2001).

I dati delle stazioni localmente più vicine vengono pesate sulla base di una relazione esponenziale negativa rispetto alla distanza euclidea (eqn. 3):

$$W_x = \exp \frac{-D_x}{R}$$

dove  $W_x$  è il peso fuzzy assegnato a ciascun pixel di training  $x$ ;  $D_x$  è la distanza euclidea (distanza orizzontale espressa in km) fra il pixel in esame e la stazione considerata;  $R$  è il range della funzione (distanza orizzontale, espressa nella stessa unità di misura di  $D_x$ ) che permette all'operatore di definire l'importanza relativa delle stazioni più vicine al pixel incognito rispetto a quelle più lontane (Maselli 2001). Al diminuire del valore di  $R$  aumenta il peso assegnato alle stazioni più vicine rispetto a quelle più lontane: per valori di  $R$  molto grandi il peso delle stazioni è quasi uguale, indipendentemente dalla loro distanza dal pixel in esame.

La scelta del valore di range ( $R$ ) più appropriato è essenziale perché esso determina le stazioni che ven-

**Tab. 1** - Valori mensili del coefficiente di correlazione di Pearson ( $r$ ) e dell'errore medio della stima in termini assoluti (BIAS, media dei valori stimati - media dei valori veri) e percentuali (BIAS% calcolato come BIAS rispetto alla media dei valori veri) ottenuti nella validazione con metodo *leave-one-out* della spazializzazione dei dati climatici considerati. I valori di BIAS per le temperature sono in gradi centigradi, quelli per le precipitazioni in mm.

Mese	Precipitazioni			Temperature minime			Temperature massime		
	$r$	BIAS	BIAS%	$r$	BIAS	BIAS%	$r$	BIAS	BIAS%
Gennaio	0.64	-0.42	-0.30	0.91	-0.06	-4.73	0.92	-0.03	-0.42
Febbraio	0.65	-0.34	-0.29	0.91	-0.05	-2.71	0.90	-0.02	-0.23
Marzo	0.77	-0.07	-0.07	0.88	-0.05	-1.29	0.87	-0.01	-0.11
Aprile	0.82	0.25	0.30	0.86	-0.05	-0.74	0.83	0.00	0.02
Maggio	0.86	0.45	0.59	0.84	-0.04	-0.35	0.82	0.02	0.10
Giugno	0.91	0.42	0.69	0.85	-0.04	-0.25	0.82	0.04	0.14
Luglio	0.92	0.24	0.56	0.82	-0.05	-0.31	0.81	0.03	0.09
Agosto	0.90	0.12	0.21	0.85	-0.05	-0.27	0.83	0.01	0.05
Settembre	0.78	0.12	0.16	0.87	-0.04	-0.31	0.86	0.02	0.07
Ottobre	0.78	-0.04	-0.03	0.89	-0.04	-0.42	0.88	-0.01	-0.05
Novembre	0.78	0.08	0.05	0.91	-0.04	-0.66	0.91	-0.03	-0.20
Dicembre	0.73	-0.04	-0.03	0.92	-0.04	-1.58	0.92	-0.04	-0.39

gono maggiormente considerate nella stima di  $Mean_i^*$ : in altre parole vengono considerate localmente solo le stazioni che si trovano entro una distanza pari o inferiore a  $R$ .

Nella sperimentazione il valore ottimale di  $R$  è stato determinando verificando l'accuratezza dei risultati tramite *leave-one-out cross validation* (Efron & Tibshirani 1993). Questo metodo consiste nell'escludere un punto di riferimento dalla costruzione di un modello di stima, che è poi applicato proprio sul punto escluso. La ripetizione di tale processo per tutti i punti di riferimento (le 403 stazioni termopluviometriche) consente di ottenere una serie completa di stime che possono poi essere confrontate con i dati misurati. Nel caso specifico la procedura è stata applicata con diversi valori di  $R$  delle regressioni localizzate, fino ad individuare quello che mediamente produceva il minor errore.

Una volta quantificato  $R$  si è passati alla stima delle variabili climatiche per tutti i *pixel* dell'immagine. Sono state quindi generate 12 immagini mensili per ognuna delle tre variabili climatiche considerate

## Risultati

Nel caso di studio il valore ottimale di  $R$  per il territorio italiano è risultato pari a 20 km. L'accuratezza della spazializzazione delle 36 variabili climatiche valutata tramite *leave-one-out cross validation* sulla base delle 403 stazioni meteorologiche utilizzate è presentata in Tab. 1. I risultati della spazializzazione

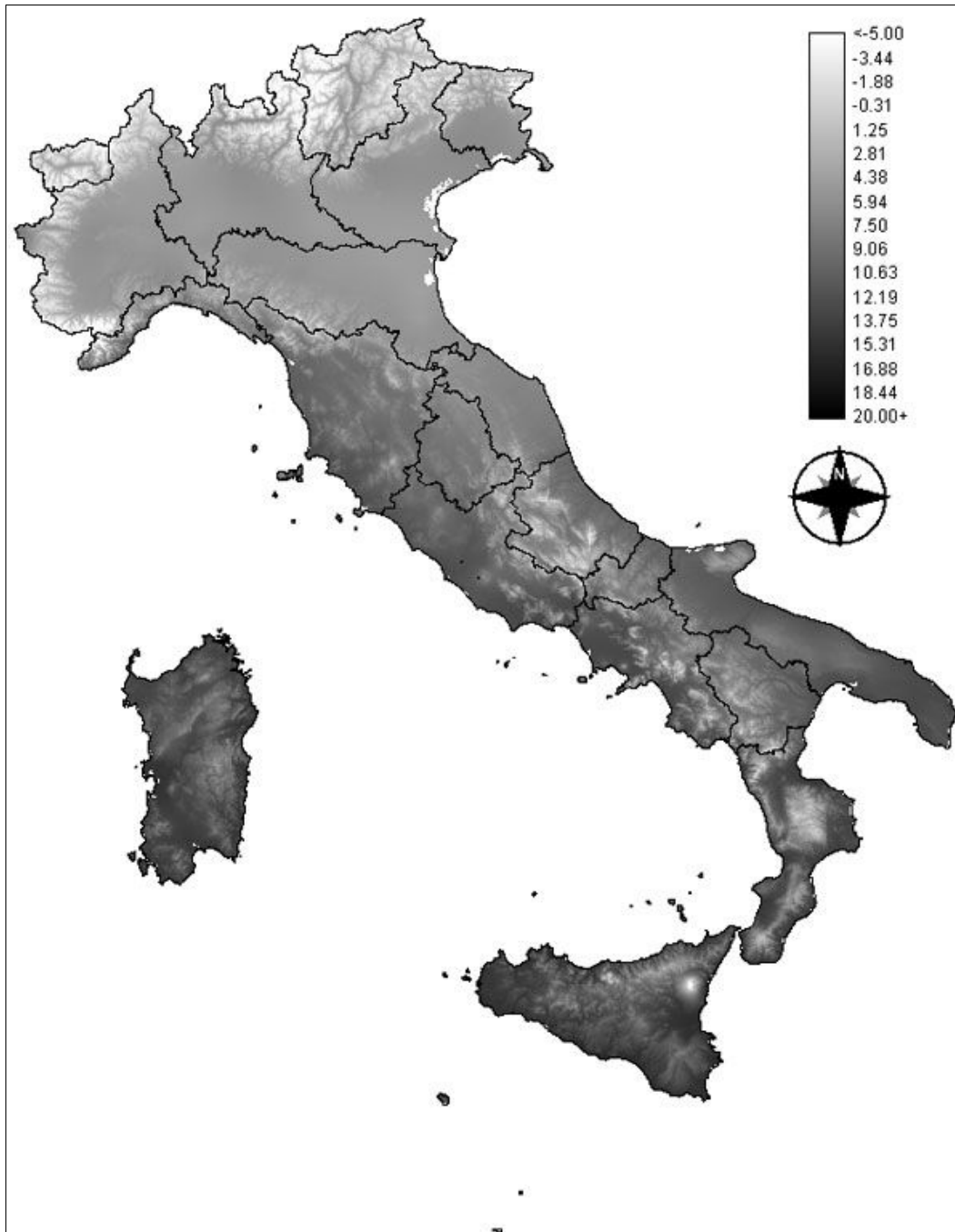
delle minime mensili variano, in termini di accuratezza espressa come errore quadratico medio, tra 1.51 e 1.89 °C per le temperature massime, tra 1.49 e 1.84 °C per le minime e tra 14.2 e 66.4 mm per le precipitazioni.

Nell'arco dell'anno, gli errori di stima delle temperature massime tendono ad aumentare nel periodo estivo (giugno-agosto). Le temperature minime presentano una distribuzione dell'errore più omogenea e costante, pur presentando un picco nei mesi di luglio-agosto. Per le precipitazioni, invece, esiste un trend che vede i valori massimi di errore nei mesi invernali (dicembre-febbraio). Distribuzioni di questo tipo sono probabilmente dovute all'accentuata variabilità spaziale propria delle variabili climatiche.

Per quanto invece riguarda la distribuzione geografica, i valori massimi di errore si hanno, per tutte le tre variabili considerate, nelle zone più elevate. La presenza di errori maggiori nelle zone montuose è spiegabile considerando come forti dislivelli di quota creino gradienti significativi nelle variabili meteorologiche, che non sempre possono essere accuratamente modellizzati tramite procedure di spazializzazione statistica quali quella applicata. Anche l'effetto longitudinale sugli errori è probabilmente attribuibile alla presenza di montagne più alte nel Nord Italia.

## Conclusioni

Seguendo l'approccio della regressione lineare univariata calibrata a livello locale la spazializzazione di



**Fig. 2** - Risultato della spazializzazione dei dati delle temperature massime nel mese di gennaio (valori in °C).

un database meteorologico contenente i valori mediati su una serie temporale di trenta anni riferiti alle temperature minime, a quelle massime e alle precipitazioni mensili ha permesso di realizzare 36 strati informativi raster di 250 m di risoluzione geometrica. Immagini di esempio del risultato del processo di spazializzazione sono riportate nelle Fig. 2 e 3 (Fig. 2, Fig. 3).

La scelta della quota come unica variabile indipen-

dente determina una opportuna semplicità dei modelli generati, permettendo di ottenere risultati robusti sotto il profilo statistico e soddisfacenti in termini di accuratezza delle stime, anche in confronto a risultati analoghi ottenuti con tecniche di *kriging* e *cokriging* (Maracchi et al. 2005). In questo senso, l'efficacia applicativa del modello ha permesso di generare strati informativi di accuratezza e risoluzione geometrica tale da poter essere impiegati operativamen-

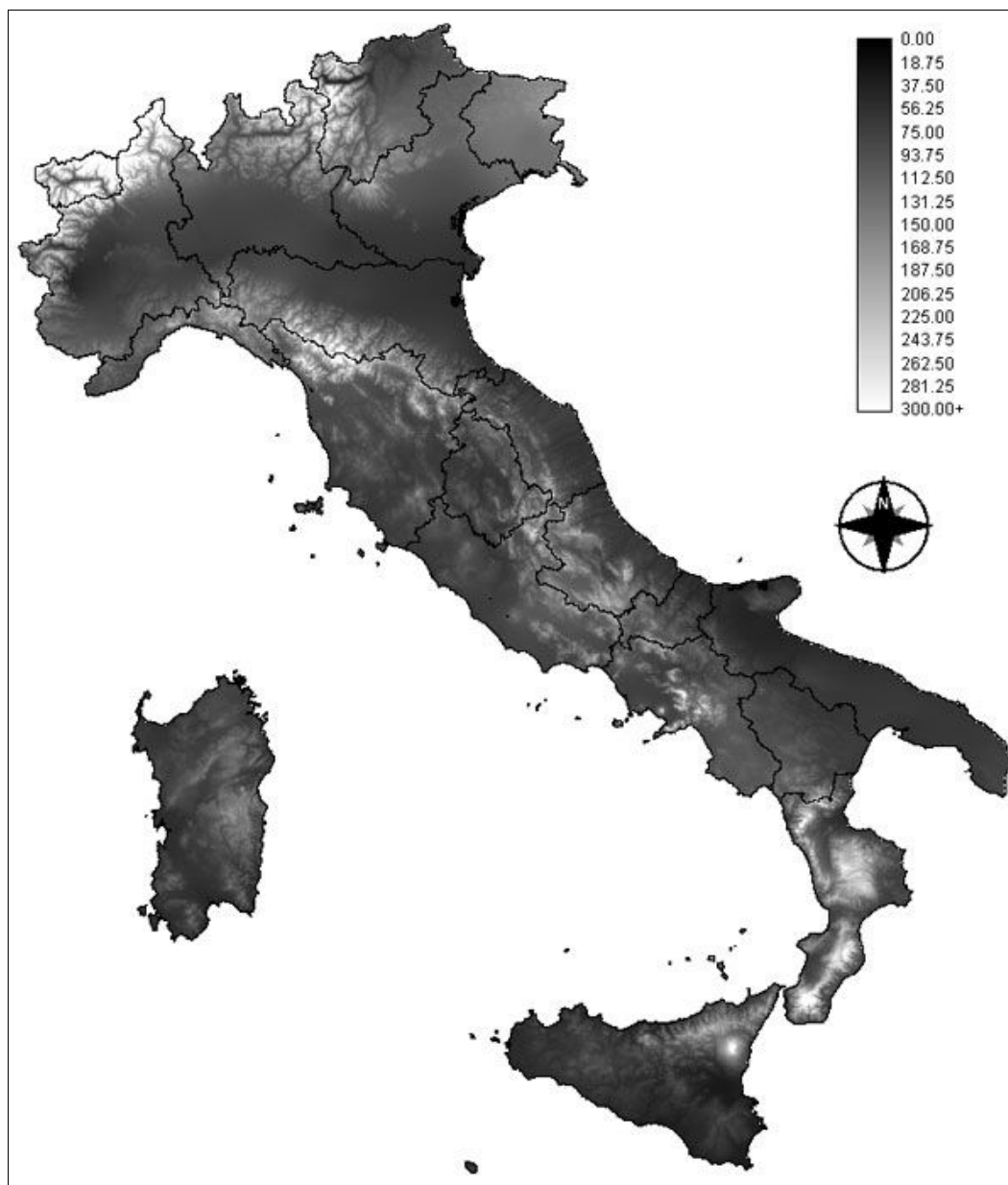


Fig. 3 - Risultato della spazializzazione dei dati delle precipitazioni nel mese di marzo (valori in mm).

te in applicazioni a supporto della pianificazione territoriale e ambientale.

### Riconoscimenti

Il lavoro è stato svolto dagli Autori in parti uguali.

### Bibliografia

Anderson TW (1984). An introduction to Multivariate Statistical Analysis. 2<sup>nd</sup> edition. John Wiley & Sons, New York, USA.

Blasi C, Ciancio O, Iovino F, Marchetti M, Michetti L, Di Marzio P, Ercole S, Anzellotti I (2001). Il contributo delle conoscenze fitoclimatiche e vegetazionali nella definizione della rete ecologica d'Italia. In: Atti del Convegno "La

conoscenza zoologica e botanica in Italia: dagli inventari al monitoraggio" (Blasi C, D'Antoni S, la Posta A eds), Roma, Dicembre 2001. Quaderni di Conservazione della Natura 14: 161-180.

Blasi C (2003). Conoscenze naturalistiche in Italia. Ministero dell'ambiente e della Tutela del Territorio, Direzione per la Conservazione della Natura, Società Botanica Italiana, Commissione per la promozione della ricerca botanica. Roma.

Cleveland WS, Devlin SJ (1988). Locally weighted regression: an approach to regression analysis by local fitting. *Journal of the American Statistical Association* 83: 596-610.

- De Philippis A (1937). Classificazioni ed indici del clima in rapporto alla vegetazione forestale italiana. Ricci, Firenze.
- Efron B, Tibshirani RJ (1993). An introduction to the bootstrap. Chapman & Hall, London, UK.
- Falge E, Baldocchi D, Olson RJ, Anthoni P, Aubinet M, Bernhofer C, Burba G, Ceulemans R, Clement R, Dolman H, Granier A, Gross P, Grünwald T, Hollinger D, Jensen NO, Katul G, Keronen P, Kowalski A, Ta Lai C, Law BE, Meyers T, Moncrieff J, Moors E, Munger JW, Pilegaard K, Rannik U, Rebmann C, Suyker A, Tenhunen J, Tu K, Verma S, Vesala T, Wilson K, Wofsy S (2001). Gap filling strategies for defensible annual sums of net ecosystem exchange. *Agricultural Forest and Meteorology* 107: 43-69.
- Gurney K (2003). An introduction to neural networks. Taylor and Francis, London, UK.
- Lacaze B, Caselles V, Coll C, Hill H, Hoff C, De Jong S, Mehl W, Negendank JFW, Riesebois H, Rubio E, Sommer S, Teixeira Filho J, Valor E (1996). DeMon - "Integrated approaches to desertification mapping and monitoring in the Mediterranean basin", Final Report of DeMon-1 Project, Joint Research Centre of European Commission, Ispra (VA), Italy.
- Maracchi G, Genesio L, Magno R, Ferrari R, Crisci A, Bottai L (2005). Progetto DesertNet - Programma Interreg III B - MEDOCC - Asse 4 Misura 4, WP A10 - Azione pilota in Toscana "I diagrammi del clima in Toscana". Rapporto Finale Giugno 2005, LaMMA-CRES in collaborazione con il LAMMA, coordinati dal CNR-IBIMET.
- Maselli F (2001). Definition of spatially variable spectral endmembers by locally calibrated multivariate regression analyses. *Remote Sensing of Environment* 75: 29-38.
- Maselli F (2002). Improved estimation of environmental parameters through locally calibrated multivariate regression analysis. *Photogrammetric Engineering and Remote Sensing* 75: 29-38.
- Maselli F, Petkov L, Maracchi G (1998). Extension of climate parameters over the land surface by the use of NOAA-AVHRR and ancillary data. *Photogrammetric Engineering and Remote Sensing* 64: 199-206.
- Olea RA (1999). *Geostatistics for Engineers and Earth Scientists*. Kluwer Academic Publisher.
- Pavari A (1916). L'influenza dei boschi mediterranei sul clima. *L'Alpe* 24: 43-57.
- Rivas-Martinez S (1995). Bases para una nueva clasificación bioclimática de la Tierra. *Folia Botanica Matritensis* 16.
- Wang F (1990a). Fuzzy supervised classification of remote sensing images. *IEEE Transact. Geosci. Remote Sens.* 28: 194-201.
- Wang F (1990b). Improve remote sensing image analysis through fuzzy information representation. *Photogramm. Eng. Remote Sens.* 56: 1163-1169.
- Wackernagel H (1995). *Multivariate geostatistics: an introduction with applications*. Springer-Verlag, Berlin, Germany.
- Wackernagel H (1998). Kriging, cokriging and external drift. In: "Dealing on spatialisation" (Gozzini B, Hims M eds). EUR 18473, COST 79, European Commission, Luxembourg.
- Zadeh LA (1965). Fuzzy set. *Inf. and Cont.* 8: 338-353.