

UNIVERSITÀ DEGLI STUDI DELL'AQUILA
FACOLTÀ DI SCIENZE MATEMATICHE, FISICHE E NATURALI



CORSO DI LAUREA MAGISTRALE IN FISICA

*Mappatura del potenziale radon geogenico
del territorio della Provincia dell'Aquila,
mediante strumenti avanzati di
modellistica statistica dei dati
sperimentali.*

Relatore interno

Dott. Piero Di Carlo

Laureanda

Sara De Novellis

Relatore esterno

Dott. Sergio Palermi

ANNO ACCADEMICO 2011 - 2012

Ai sensi della legge n.633/1941 la riproduzione integrale o di parti della presente tesi è consentita esclusivamente previo consenso scritto dell'autore

Indice

Introduzione	1
1 Radon: sorgenti, trasporto e effetti sulla salute	7
1.1 Il Radon	7
1.2 Sorgenti e trasporto del radon	10
1.2.1 Sorgenti	12
1.2.2 Trasporto	16
1.2.3 Concentrazione di radon outdoor	23
1.2.4 Ingresso negli edifici	24
1.3 Effetti sulla salute e normativa	29
2 Campagne di misure di concentrazione radon indoor in Abruz- zo.	33
3 Analisi multilevel per le concentrazioni di radon indoor nella provincia aquilana	42
3.1 Analisi multilevel e variabile radon standardizzata	44
3.2 Inferenza classica e bayesiana	46
3.3 Analisi bayesiana dei dati di concentrazione di radon indoor	48
3.4 Multilevel computation	51
3.5 Modello multilevel con predittori e dati non annidati	58

3.6	Distribuzione log-normale delle misure	59
3.7	La standardizzazione delle misure di radon indoor	64
3.8	Applicazione dei modelli con predittori ai dati di concentrazione radon.	68
4	Geographically Weighted Regression e mappa del potenziale radon geogenico	79
4.1	Geographically Weighted Regression	86
4.2	Radon indoor e Geographically Weighted Regression	90
4.3	Il potenziale radon geogenico	104
4.3.1	Mappatura per la Provincia dell’Aquila	106
4.3.2	Considerazioni finali	113
	Conclusioni	115
A	Analisi spaziale per la mappatura delle temperature medie	119
A.1	Variogramma	124
A.1.1	Variogramma sperimentale	126
A.1.2	Caratteristiche del variogramma	127
A.1.3	Fit del variogramma	128
A.2	La stima spaziale	129
A.3	Il kriging	131
A.3.1	Ordinary Kriging	131
A.3.2	Kriging con drift esterno	134
A.4	Cross validation leave-one-out	137
A.5	Temperature medie nella provincia aquilana.	138
A.5.1	Influenza dell’ altitudine.	143
B	Codici in R	149
B.1	Test di normalità per i dati log-trasformati di radon	149
B.2	Analisi multilevel bayesina con R2jags	150
B.2.1	Modello per l’intercetta random e coefficienti fissi	150

B.2.2	Modello a effetti misti	151
B.3	Kriging con drift esterno	152
B.4	Indice I di Moran	153
Bibliografia		154
Ringraziamenti		163

Introduzione

Il radon è un gas radioattivo naturale noto per essere responsabile di circa la metà dell'esposizione della popolazione dovuta alle radiazioni ionizzanti, escluse le esposizioni in ambito medico (UNSCEAR, 2000).

L'Agenzia Internazionale per la Ricerca sul Cancro (IARC/OMS) classifica il radon e la sua progenie come agenti cancerogeni di gruppo 1, per i quali si ha il massimo grado di certezza degli effetti (Ministero della Salute, 2002).

Il radon è un gas nobile, inodore e inerte, è moderatamente solubile in acqua ed è largamente presente in natura. Il gas diventa di particolare interesse sanitario quando si accumula negli ambienti chiusi, infatti in atmosfera, il radon prodotto principalmente dal suolo, si diluisce mantenendo concentrazione in genere basse ($8-12 \text{ Bq m}^{-3}$) (UNSCEAR, 2000). Negli ambienti indoor, invece, tende ad accumularsi e raggiunge livelli tali da rappresentare un rischio per la salute.

L'Organizzazione Mondiale della Sanità, in una recente pubblicazione (WHO, 2009) e a seguito degli ultimi studi epidemiologici, ha proposto l'adozione di un livello nazionale di riferimento per la concentrazione media annua di radon nell'aria indoor delle abitazioni pari a 100 Bq m^{-3} e, quando ciò non sia possibile, che il livello individuato non sia superiore a 300 Bq m^{-3} . Anche in sede europea (Euratom) è in corso di elaborazione una nuova direttiva, che sostituirà la vecchia 96/29/Euratom, nella quale si rivedono, sulla base delle indicazioni WHO e

della pubblicazione n°103 dell'ICRP (International Commission on Radiological Protection), i limiti di azione per la concentrazione di radon indoor sia per gli ambienti di lavoro che di vita in genere (EC 2011). La stessa raccomandazione assegna agli Stati membri il compito di redigere un piano nazionale radon, di fissare dei limiti di riferimento nazionali di concentrazione indoor e di individuare le cosiddette *radon-prone areas* (zone caratterizzate da elevate concentrazioni di radon negli edifici).

In Italia, la normativa vigente in materia di protezione dai rischi derivanti da radiazioni ionizzanti (Decreto Legislativo n. 241/2000), recepisce la raccomandazione 96/29/Euratom e, per i soli ambienti di lavoro, fissa un livello di azione, il cui superamento rende necessarie azioni di rimedio, di 500 Bq m^{-3} di concentrazione media annua di radon. Lo stesso Decreto Legislativo (art. 10-sexies) assegna alle Regioni il compito di individuare le radon prone areas.

L' Agenzia Regionale per la Tutela dell' Ambiente (ARTA), per conto della Regione Abruzzo, ha promosso diverse campagne regionali di misura di concentrazione di radon indoor dal 2005, tuttora in corso, al fine di individuare tali zone.

Il lavoro svolto in questa tesi rientra in questo progetto, anche in vista del futuro recepimento della nuova raccomandazione Euratom (BSS; EC 2011), la cui pubblicazione è prevista per la fine del 2013. Lo scopo è di studiare la relazione tra alcune variabili ambientali e la concentrazione di radon indoor, al fine di mappare il potenziale radon geogenico, cioè una quantità connessa con il rischio radon dipendente solo dalle condizioni geologiche e geofisiche dell'area considerata. Una corretta definizione del potenziale radon, infatti, potrebbe fornire una guida per l'identificazione delle radon-prone areas, in modo particolare quando il numero e/o la qualità dei dati di radon indoor è inadeguata. In linea di principio, le sole informazioni geologiche (cioè i dati radiometrici e geochimici, la permeabilità e l'umidità del suolo, il contenuto di radon nei gas del terreno) potrebbero essere sufficienti per stimare il potenziale radon (Grunderse e Schumann, 1996; Orlando et al., 2000; Kemski et al. 2001; Ielsch et al. 2010). In

studi precedenti, è stato anche esplorato un approccio alternativo, in grado di combinare a vari livelli le misure di radon indoor e i dati geologici (Smethurst et al., 2008; Appleton et al., 2008; 2011). In altri casi (Friedmann, 2005; Bossew et al., 2008) i dati di radon sono stati pre-processati in modo da filtrare, per quanto possibile, la variabilità connessa con l'edificio (dovuta al piano, ai materiali da costruzione, alla presenza di un vespaio etc.).

Generalmente, la valutazione del potenziale radon è stata ottenuta mediante l'utilizzo di tecniche statistiche di stima globale, assumendo che la relazione studiata sia spazialmente stazionaria (Apte et al., 1999; Smith and Field, 2007; Bossew et al., 2008).

Il nostro approccio, invece, analizza le conseguenze connesse a significative variazioni spaziali nella relazione tra le misure di radon indoor, adeguatamente corrette, e le caratteristiche geologiche. Questo ci permette di stimare la concentrazione di radon indoor di un edificio standard considerando le proprietà ambientali locali, in accordo con il potenziale radon geogenico del terreno sottostante, consentendoci, quindi, di mappare con maggior precisione le eventuali aree ad elevata concentrazione di radon.

Per il nostro studio abbiamo utilizzato i dati sperimentali relativi a 757 edifici appartenenti a 96 comuni della Provincia dell'Aquila, raccolti da ARTA nel corso delle varie campagne di misura del radon indoor.

Le variabili ambientali che influenzano la concentrazione di radon in un edificio sono di natura sia geologica/geofisica (litologia, contenuto di uranio/radio nel terreno, permeabilità del suolo, clima etc) che edilizia (Appleton et al., 2008).

Per quanto concerne la valutazione delle variabili edilizie (materiali da costruzione, anno di costruzione etc), considerando la struttura gerarchica dei dati (cioè misure di radon in edifici raggruppati per comuni o altre unità territoriali amministrative, geometriche o di natura geologica), abbiamo eseguito un'analisi multilevel bayesiana con metodo Monte Carlo basato sulle proprietà della catena di Markov (MCMC). Lo scopo è quello di definire una nuova variabile dipendente normalizzata che corrisponda ad una concentrazione di radon indoor riscontra-

bile in un edificio con caratteristiche standard, eliminando in tal modo parte della variabilità di tipologie costruttive che caratterizza il campione di dati sperimentali a disposizione. Il calcolo è svolto associando a ciascun parametro della regressione una distribuzione non informativa a priori, mentre la distribuzione di probabilità a posteriori è ottenuta per integrazione numerica con simulazioni MCMC, campionate con algoritmo di Gibbs.

L'influenza dei parametri geologici e geofisici sulla concentrazione radon indoor standard, così ottenuta, viene analizzata attraverso la Geographically Weighted Regression (GWR) (Fotheringham et al. 2002). Si tratta di una metodologia statistica di recente introduzione che permette di formulare modelli di regressione non stazionari per variabili spaziali, in grado di fittare localmente il dataset, tenendo conto delle condizioni locali.

Le variabili predittive utilizzate nel modello includono:

- i dati radiometrici (contenuto di uranio, torio e potassio nello strato superficiale del suolo, tratti da Bellotti et. al, 2007), interpolati spazialmente con la tecnica del block kriging ordinario su una griglia di 2×2 km (Nissi et al., 2012);
- la temperatura media annuale, stimata con un kriging con drift esterno. A differenza di quanto fatto nel lavoro di Pasculli et al. 2013, la dipendenza climatica della concentrazione di radon indoor è analizzata utilizzando direttamente la temperatura invece che l' altitudine. Ci siamo serviti dei valori registrati da 43 stazioni termometriche sparse sul territorio aquilano e gestite dal Servizio Idrografico della Regione Abruzzo (<http://www.regione.abruzzo.it/xIdrografico/index.asp>);
- le variabili geologiche quali permeabilità, spessore, carsismo, etc., ricavate grazie alle mappe prodotte in precedenti studi (Ghisetti e Vezzani, 1998; Apat, 2006; ISPRA 2012) e rielaborate in relazione al contributo potenziale alla distribuzione di radon (Miccadei e Piacentini in Pasculli et al. 2013). La permeabilità per i depositi continentali quaternari, in partico-

lare, è frutto di un lavoro di revisione di Palermi e Desiderio (non pubblicato, comunicazione personale, 2013) basato su mappe geologiche ISPRA (progetto CARG e in qualche caso vecchia carta geologica 1:10000).

La tesi è organizzata come segue.

Nel primo capitolo viene descritto il gas radon, le sorgenti e le modalità di trasporto, fino all'accumulo negli ambienti chiusi. Nell'ultima parte vengono brevemente introdotti gli effetti sulla salute e la normativa vigente in materia.

Nel secondo capitolo invece si riportano velocemente gli scopi e i risultati ottenuti nelle campagne di misura di concentrazione di radon indoor, svolte sul territorio abruzzese dal 1991 e tutt'ora in corso.

Il lavoro prosegue con la descrizione di quanto fatto per eliminare la variabilità introdotta nei livelli di radon indoor dai fattori edilizi, con una panoramica sull'analisi multilevel bayesiana e il metodo computazionale Monte Carlo basato sulle proprietà delle catene di Markov.

Nel quarto capitolo si affronta l'influenza delle variabili geologiche e geofisiche sulle concentrazioni di radon indoor, si descrive la GWR e si mappano le stime dei coefficienti locali della regressione. La seconda parte invece tratta il tema del potenziale radon geogenico, della stima di questa variabile per i centri abitati per cui non sono attualmente disponibili misure di radon e si presenta la prima mappa del potenziale radon per la Provincia dell' Aquila.

Nella prima appendice viene descritto il metodo di stima della temperatura media annua per i siti di misura di concentrazione radon. Si illustra la tecnica del kriging con drift esterno utilizzato per i nostri fini.

L'analisi statistica dei dati è svolta all'interno di **R**, un ambiente di sviluppo specifico, distribuito con licenza GNU GPL (<http://www.r-project.org>). Per l'analisi multilevel bayesiana ci siamo serviti di JAGS (Just Another Gibbs Sampler), un programma che utilizza simulazioni Markov Chain Monte Carlo, campionate con l'algoritmo di Gibbs, e della libreria *R2jags* che permette di ri-

chiamarlo all'interno di **R** (<http://mcmc-jags.sourceforge.net>). Per il GWR invece abbiamo utilizzato il software GWR4.0 di Nakata et al. (<http://gwr.nuim.ie>). Nella seconda appendice sono riportati i codici scritti in **R**.

Capitolo 1

Radon: sorgenti, trasporto e effetti sulla salute

Nel capitolo si descrivono i fattori e i processi principali che influenzano la dinamica della concentrazione di radon indoor. Viene trattata la generazione, la migrazione dal mezzo sorgente e la capacità di penetrare e di accumularsi del gas all'interno degli edifici. Nell'ultima parte del capitolo si riportano brevemente gli effetti sulla salute e le disposizioni legislative vigenti in materia.

1.1 Il Radon

La radioattività naturale rappresenta la principale fonte di esposizione dell'uomo alle radiazioni ionizzanti. L'UNSCEAR (*United Nations Scientific Committee on the Effects of Atomic Radiation*) stima che la dose efficace annua pro capite dovuta alla radioattività naturale sia circa di 2.4 mSv, ovvero pari a circa l'85% della dose annua totale (UNSCEAR, 2000). La più importante fonte terrestre di radiazioni è costituita dai radionuclidi primordiali, quelli cioè prodotti al momento della formazione della Terra ed ancora radioattivi a causa del lungo tempo di dimezzamento, quali il ^{40}K e le famiglie radioattive dell'uranio e del torio.

Le catene di decadimento dei radionuclidi primordiali, ^{238}U , ^{232}Th e ^{235}U , contengono un radioisotopo comune : il radon (fig. (1.1)). Molti dei radionuclidi intermedi della serie di uranio e torio sono isotopi di metalli pesanti e chimicamente reattivi. Per questa ragione, una volta formati dal decadimento del predecessore, questi tendono a restare nel materiale nel quale sono stati originariamente generati. L'unica eccezione è rappresentata proprio dal radon, il gas nobile più pesante, che non crea legami chimici e tende quindi a migrare all'interno del materiale in cui si è formato, muovendosi in relazione a gradienti di concentrazione, pressione e temperatura. In generale, si può affermare che la concentrazione di uranio nel materiale, le caratteristiche meccaniche dello stesso — come porosità e granulosità — e le condizioni ambientali — come pressione, temperatura e umidità — caratterizzano il processo di esalazione di radon.

L'inalazione del radon e dei suoi prodotti di decadimento contribuisce per il 50% alla dose annua pro capite dovuta alla radioattività naturale, con variazioni dovute alla variabilità dei livelli di radon presenti negli ambienti chiusi e correlate, in ultima analisi, alla quantità di radionuclidi naturali presenti nel suolo. L'isotopo di radon più abbondante in natura è il ^{222}Rn , che ha un'emivita ($t_{1/2}$) di 3.82 giorni, il thoron (^{219}Rn) e l'actinon (^{220}Rn) hanno invece tempi di dimezzamento più brevi, rispettivamente di 3.96 s e 55.6 s. Nella serie dell' ^{238}U , il ^{222}Rn è prodotto dal decadimento del ^{226}Ra , che ha un tempo di dimezzamento di 1600 anni. Il radon a sua volta decade nel piombo stabile ^{206}Pb attraverso una catena di decadimenti intermedi che producono isotopi radioattivi di elementi solidi (cfr. figura 1.2). Tra questi, i prodotti di decadimento a vita breve, ^{218}Po , ^{214}Pb , ^{214}Bi e ^{214}Po , rivestono notevole importanza per quanto riguarda l'esposizione umana. Infatti, questi radionuclidi si legano facilmente all'aereosol presente in aria e, inalati, si depositano lungo le pareti dell'albero bronchiale; quindi decadendo, emettono radiazione ionizzante con range piccolo (decine di μm) e energia sufficiente per colpire le cellule basali dell'epitelio bronchiale, bersagli critici per eventuali modificazioni cancerose.

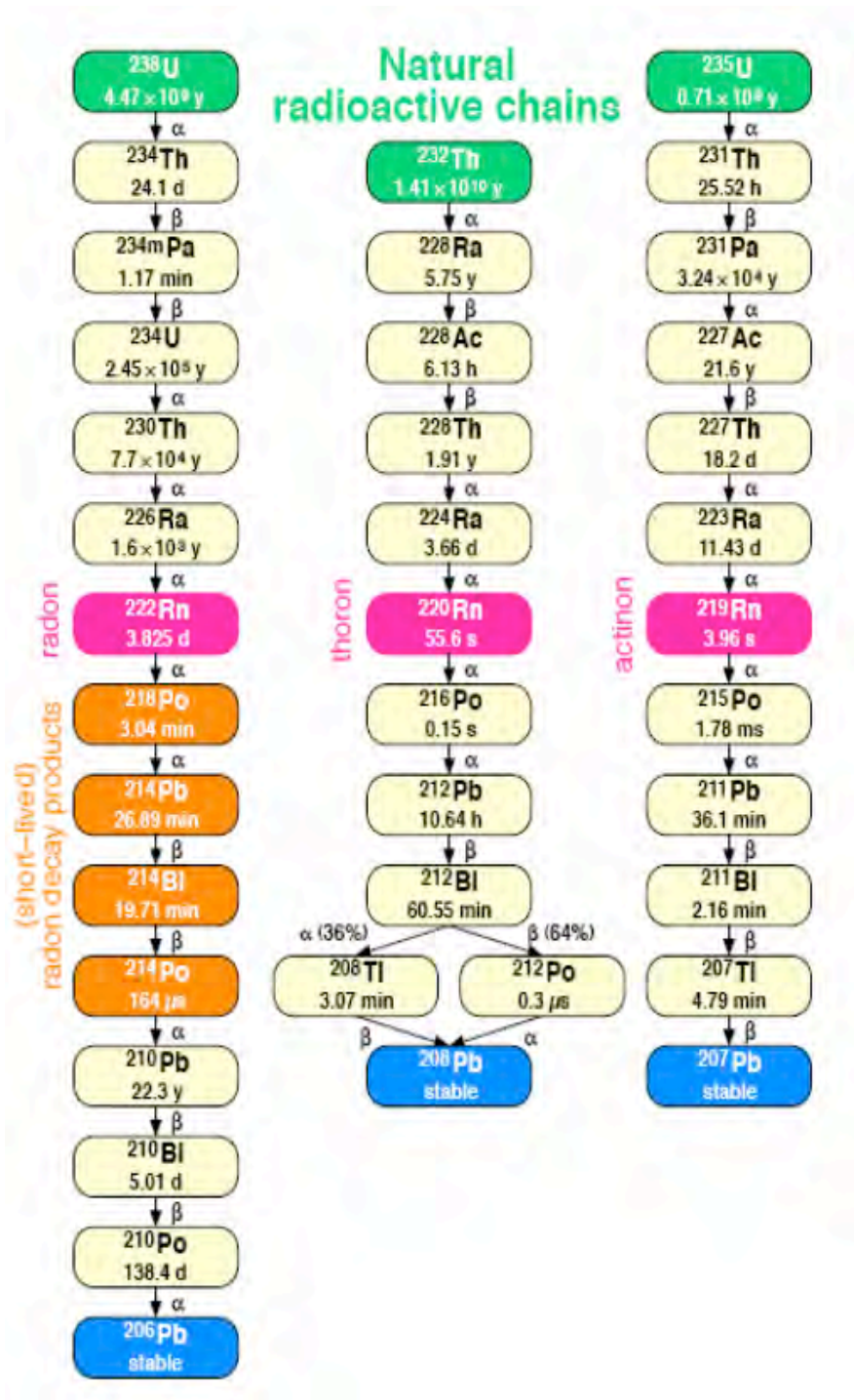


Figura 1.1: Schema di decadimento delle famiglie radioattive di ^{238}U , ^{232}Th e ^{235}U (dall'Istituto Superiore della Sanità).

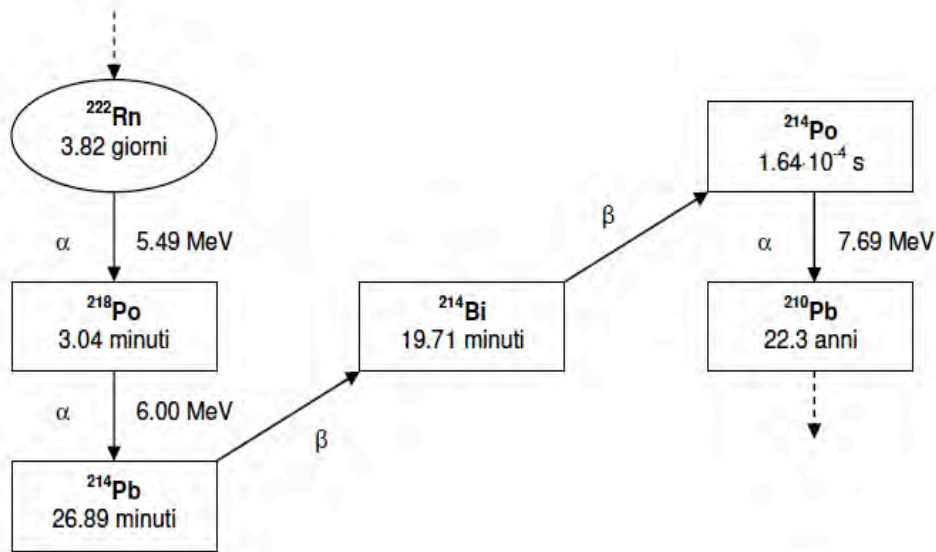


Figura 1.2: Schema di decadimento del ^{222}Rn

1.2 Sorgenti e trasporto del radon

Il radon è generato dal decadimento radioattivo del radio contenuto nella crosta terrestre e la concentrazione indoor del gas dipende dalla sua capacità di penetrare all'interno degli edifici. Le modalità di ingresso del radon nell'ambiente indoor sono molteplici: la principale è costituita da processi di trasporto del gas originatosi nel suolo (o roccia) sottostante l'edificio, mentre i materiali di costruzione giocano generalmente un ruolo secondario e solo in alcuni casi le forniture di acqua e gas naturale possono rappresentare una via di ingresso non trascurabile. L'aria esterna, infine, contribuisce per una quota minima alle concentrazioni indoor, in quanto caratterizzata da livelli di radon bassi, mediamente nel range $5\text{-}15 \text{ Bq m}^{-3}$, grazie grazie all'efficacia dei processi di mescolamento nell'atmosfera (UNSCEAR, 2000). Per comprendere i meccanismi di generazione e trasporto del radon, possiamo individuare quattro fasi nel processo che, partendo dalla generazione del radon, porta al suo accumulo negli edifici (fig. (1.3)):

- generazione di radon nel materiale (sia suolo/roccia che materiale di costruzione dell'edificio): dipende dal contenuto di radio, dalla grandezza dei grani, dalla porosità e dal contenuto di acqua;
- migrazione all'interno del materiale sorgente, riconducibile a due processi indipendenti: la diffusione molecolare in risposta ad un gradiente di concentrazione del gas, e, se la sorgente è il suolo sottostante l'edificio, l'avvezione data dal gradiente di pressione; i parametri che influiscono sul trasporto sono essenzialmente il coefficiente di diffusione e la permeabilità del mezzo;
- ingresso nell'edificio: dipende dalle caratteristiche di trasporto ma anche dalle proprietà di interfaccia tra il mezzo materiale sorgente dei processi di trasporto e l'ambiente interno;
- accumulo all'interno dell' edificio: dipende dal bilanciamento tra il rateo di entrata e i processi di rimozione del gas (essenzialmente la ventilazione degli ambienti e il decadimento radioattivo).

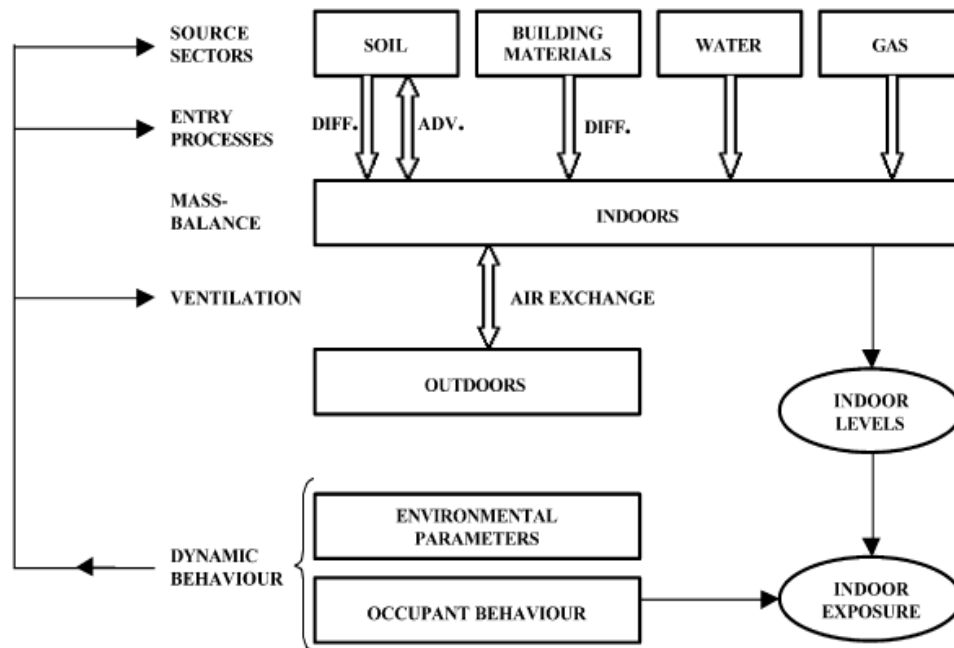


Figura 1.3: La formazione del radon, l'entrata e l'accumulo negli edifici. (da Font e Baixeras, 2003)

1.2.1 Sorgenti

L'uranio e il radio sono elementi presenti in tutti i tipi di rocce e suoli in concentrazioni variabili, generalmente comprese tra 0.5 e 5 mg/kg. La distribuzione varia in funzione del tipo di roccia o di terreno ed in relazione alle modalità ed agli ambienti geologici di formazione degli stessi. Variazioni di concentrazione si possono avere localmente anche nell'ambito di uno stesso tipo di materiale.

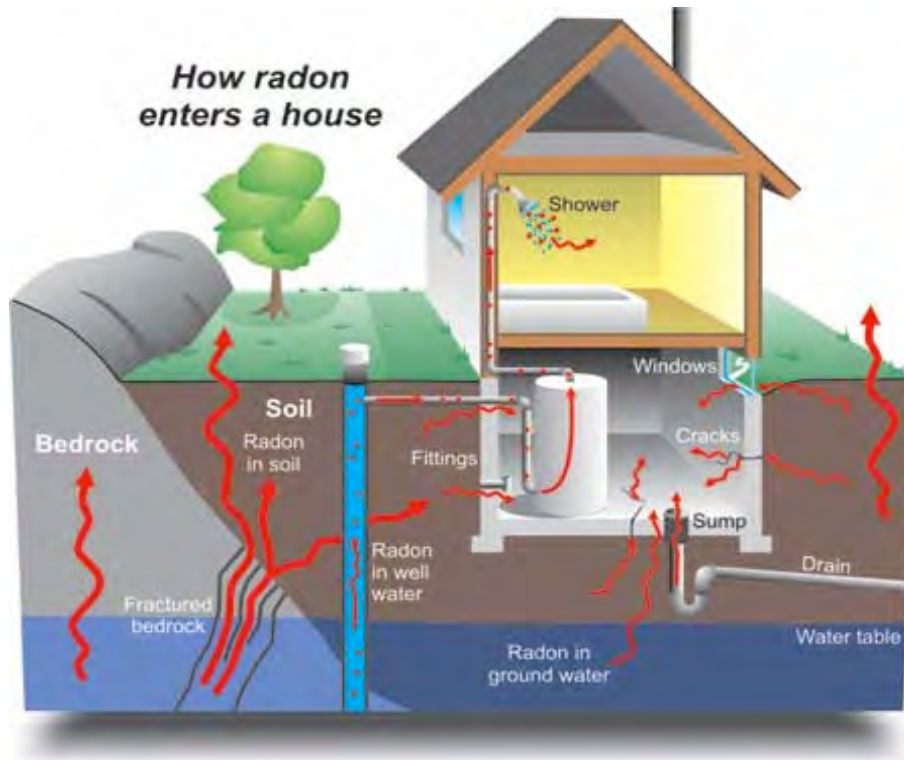


Figura 1.4: Rappresentazione delle possibili sorgenti di radon e vie di ingresso del gas negli edifici.

In genere il valore medio di concentrazione di radio nel suolo, in termini di attività specifica, varia tra circa 10 e 100 Bq/kg (Nazaroff, 1992). Il flusso di radon dal suolo che ne deriva dipende quindi da questo valore ed inoltre da parametri quali il contenuto di umidità del suolo, la porosità, i valori di temperatura e pressione atmosferica al momento della misura (NCRP n.97).

Nazaroff et al (1988) proposero un'equazione che esprime la concentrazione di radon nel gas del suolo (C_{Rn}) in funzione della concentrazione di radio nel terreno (C_{Ra}) e in assenza di trasporto di radon:

$$C_{Rn} = C_{Ra} \frac{f \rho_s (1 - \epsilon)}{\epsilon [m (K_T - 1) + 1]} \quad (1.1)$$

dove f è il coefficiente di emanazione, ρ_s è la densità dei grani del terreno (tipicamente 2700 Kg/m^3 , UNSCEAR 2000), ϵ è la porosità totale, m è la frazione della porosità riferita all'acqua (chiamata anche frazione di saturazione, per il

terreno asciutto m è zero) e K_T che è il rapporto tra la concentrazione di gas radon in acqua e in aria.

L'influenza del suolo sulle concentrazioni di radon rilevabili negli edifici riconducibile alle tipologie litologiche affioranti e profonde in associazione a parametri quali fratturazione, presenza di faglie, sismicità e vulcanismo. In due particolari contesti geologici si possono verificare concentrazioni anomale di uranio e radio e quindi di radon: le zone idrotermali e le aree interessate da un carsismo sviluppato. Nelle aree idrotermali si determinano circolazioni di fluidi che raggiungono la superficie da profondità anche elevate (UNSCEAR, 2000). Tali fluidi, in funzione delle rocce con cui entrano in contatto, possono arricchirsi in uranio e radio ed in radon stesso. Il carsismo è un processo di dissoluzione chimica che si sviluppa generalmente su rocce calcaree povere in uranio e radio e determina lo sviluppo in profondità di un reticolo di condotti, fenditure e cavità, con ampiezza anche molto accentuata, che possono rappresentare vie preferenziali di migrazione per il radon.

Tipo di suolo	^{226}Ra (Bq kg $^{-1}$)	^{222}Rn (Bq m $^{-3}$)
Suoli con contenuto normale di radio	15-65	5000-30000
Suoli con frammenti di granito	130-125	10000-60000
Suoli con frammenti di granito ricco di uranio	125-360	10000-200000
Ghiaia	30-75	10000-150000
Sabbia	5-35	2000-20000
Limo	10-50	5000-60000
Argilla	10-100	10000-100000

Tabella 1.1: Concentrazioni medie ^{222}Rn e ^{226}Ra nei suoli alla profondità di 1 m. (tratte da ANPA,2000.)

I materiali da costruzione possono anch'essi contribuire alle concentrazioni di radon indoor, anche se in maniera meno significativa del suolo.

In genere il contenuto tipico di radioattività nei materiali da costruzione è infatti pari a circa 50 Bq kg^{-1} , dovuto alla presenza di ^{238}U , ^{232}Th e dei relativi prodotti di decadimento. In alcuni materiali particolari, il contenuto di radioattività può discostarsi significativamente da questo valore. In particolare, materiali da costruzione di origine naturale, provenienti da zone particolarmente ricche di ^{238}U e ^{232}Th , possono presentare livelli di radioattività di un ordine di grandezza superiore. In Italia questa situazione si presenta per il tufo originario dell'Italia centro-meridionale. Livelli elevati di radioattività possono caratterizzare anche materiali compositi derivanti da lavorazioni industriali, quali il fosfogesso, sottoprodotto dell'industria dei fosfati.

materiali da costruzione	campioni misurati	$C_{Ra^{226}}$ (Bq m^{-3})	$C_{Th^{226}}$ (Bq m^{-3})	C_{K40} (Bq m^{-3})
tufo	46	209	349	1861
cemento	185	42	66	369
calcestruzzo	23	22	16	237
laterizi	185	29	26	711
sabbia	43	18	22	530
pietra	13	24	37	645

Tabella 1.2: Riepilogo dei dati pubblicati sulla concentrazione di attività di diversi materiali di costruzione usati in Italia (tratti da Raspa et al. 1999).

Il radon è moderatamente solubile in acqua. Il suo coefficiente di solubilità, definito come rapporto tra la concentrazione di radon in acqua e la concentrazione in aria, è pari a 0.51 alla temperatura di 0° C e 0.25 a 20° C (Nazaroff, 1992). La solubilità del radon in acqua dipende dalla temperatura: più fredda è la corrente di acqua, maggiore sarà la sua solubilità. Grazie a questa proprietà, il radon può essere assorbito da correnti di acqua, sotterranee e non, che attraversano suoli contenenti radon, questo fa sì che il trasporto da parte dell'acqua possa costituire un meccanismo non trascurabile per l'ingresso negli edifici.

Alcuni parametri geologici, inoltre, possono incrementare notevolmente la mobilità dei fluidi nel sottosuolo, aumentando così le capacità di migrazione del radon.

La fratturazione, ad esempio, rende gli ammassi rocciosi permeabili ai fluidi, particolarmente se sono presenti fratture di grande estensione o aperture alle quali si associano reticoli di fratture di minore estensione e importanza.

L'acqua contribuisce in maniera minima alla concentrazione di radon all'interno degli edifici. Infatti, comunque, il rapporto tra la concentrazione presente in aria e quella presente nell'acqua è dell'ordine di 10^{-4} , considerando valori tipici per il volume di acqua utilizzato, per il volume dell'edificio e per il rateo di ventilazione (Nazaroff, 1992).

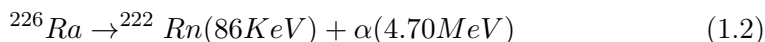
1.2.2 Trasporto

L'emanazione di radon è riferita a quei processi che determinano la fuga degli atomi di radon da un dato materiale. Il *coefficiente di emanazione* ϵ è definito come il rapporto tra il numero di atomi di radon che fuoriescono da una certa quantità di materiale e il numero totale di atomi formatisi per decadimento radioattivo del radio in quel materiale per unità di tempo. A seconda della struttura del materiale e del contenuto di acqua, il coefficiente di emanazione del radon dal suolo varia da 0.05 a 0.7 (Nazaroff, 1992).

Tipo di suolo	Emanazione (%)
Ghiaia	15-40
Sabbia	15-30
Argilla	30-70
<i>Crushed rock</i> (1-8 mm)	5-15
<i>Crushed granite</i> ricco in uranio (1-8 mm)	15-30

Tabella 1.3: Coefficiente di emanazione in percentuale di radon nei suoli (da Nazaroff et al., 1992)

Il potere di emanazione è essenzialmente dovuto a due fenomeni: diffusione dei gas nei solidi e processo di rinculo. Poiché il coefficiente di diffusione del gas nei grani dei solidi è molto basso, si assume che il principale contributo al potere di emanazione è dato dal processo di rinculo degli atomi di radon, dovuto all'emissione della particella α nel decadimento:



Tuttavia elevate concentrazioni di radon nei gas del suolo si riscontrano solo quando questo elemento può essere espulso dal reticolo cristallino dei minerali che lo contengono (vd. figura 1.5).

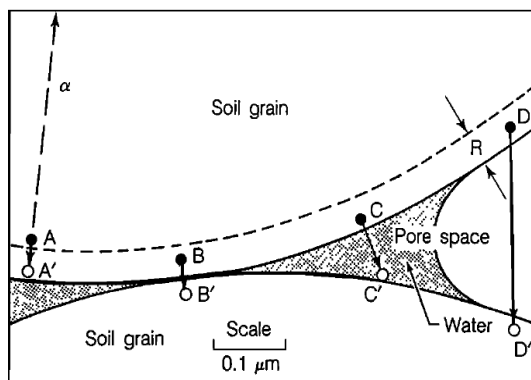


Figura 1.5: illustrazione schematica della traiettoria di rinculo del radon tra grani del terreno. Due grani sferici di $2 \mu\text{m}$ di diametro sono in contatto in B. La porzione dei pori è riempita di acqua. Gli atomi di radio, indicati con il cerchio pieno, decadono e producono una particella α e un atomo di radon rinculato che può finire in uno dei punti indicati dai cerchi vuoti. In C l'atomo di radon termina nell'acqua. (Tanner, 1980)

Nel decadimento del radio (${}^{226}\text{Ra}$) viene emessa una particella alfa e l'atomo di radon appena formato rincula nella direzione opposta. La distanza che il radon può percorrere nel granulo di minerale in cui si è formato varia tra 0.02 e $0.07 \mu\text{m}$.

La posizione dell'atomo di radio nel granulo (ovvero la sua distanza dalla superficie) e la direzione di rinculo dell'atomo (verso la superficie del granulo o verso l'interno) determinano la fuoriuscita o meno del radon verso i pori esistenti tra

i granuli del materiale. Una volta uscito dal granulo, l'atomo di radon potrebbe comunque attraversare il poro e finire racchiuso in un granulo adiacente, se le dimensioni del poro sono troppo piccole. Questo è dovuto al cammino che l'atomo di radon può percorrere nell'aria che riempie i pori e che è pari a circa $63 \mu\text{m}$ (Tanner, 1980).

La temperatura è un fattore che influenza poco il potere di emanazione, il quale risulta invece sensibile alle variazioni di umidità, cioè al contenuto di acqua nel materiale. Infatti lo stopping-power dell'acqua per gli atomi di radon è decisamente maggiore rispetto a quello dell'aria; quindi se i pori del materiale contengono acqua, l'atomo di radon viene rallentato da questa e può con maggiore probabilità terminare il suo cammino nel poro. Infatti il cammino tipico in acqua è inferiore a quella in aria ed è pari a circa $0.1 \mu\text{m}$ (Nazaroff, 1992).

La presenza di acqua nel materiale incrementa perciò il coefficiente di emanazione del radon. Nella maggior parte dei suoli e delle rocce solo il 10-40% del radon effettivamente prodotto fuoriesce dai granuli ed entra nei pori dove è libero di migrare. L'effettiva emanazione di radon da una roccia dipende quindi dalla superficie specifica dei singoli granuli o frammenti: maggiore è la superficie esposta rispetto alla massa della roccia, maggiore è la capacità di emanazione del radon.

Il radon, una volta formato ed emanato dalle rocce e dai minerali, può entrare a far parte del sistema dei gas interstiziali delle rocce e dei suoli, oppure venire disciolto nelle acque di falda. In entrambi i casi il radon comincia un percorso di migrazione la cui estensione dipende in gran parte dal suo tempo di dimezzamento.

I processi che determinano la migrazione del radon, come di qualsiasi altro gas, sono: la diffusione attraverso i pori ed i capillari nel suolo in direzione opposta all'aumento del gradiente di concentrazione; l'avvezione indotta dalla pressione attraverso i capillari nel suolo; il trasporto da parte di un altro fluido

(liquido o gassoso). I primi due processi consentono lo spostamento del radon su distanze dell'ordine di grandezza dei centimetri o dei metri. Il terzo può in alcuni casi determinare migrazioni per distanze molto maggiori. I processi di diffusione e avvezione sono maggiormente favoriti se nel suolo sono presenti spazi per la migrazione del gas, che possono variare da interstizi molecolari a vere e proprie cavità sotterranee e che possono essere isolati o interconnessi tra loro (NCRP n.97). La diffusione, descritta dalla legge di Fick, dipende dai seguenti parametri:

- porosità, che può essere distinta in: primaria o intergranurale, che trae origine dal sedimento ed è quindi legata alle sue modalità di formazione; secondaria, che è determinata da fenomeni successivi, come fratturazione, soluzione, ricristallizzazione, ecc.

Si ha inoltre una porosità totale, che comprende la totalità dei pori, che è definita come rapporto tra volume totale dei vuoti e volume di un mezzo poroso, in altri termini è la percentuale di roccia occupata da vuoto, che a sua volta può essere riempito di aria o acqua (in genere).

La porosità effettiva, invece, si riferisce soltanto ai pori intercomunicanti, utilizzabili dal fluido in movimento, ed è definita come il rapporto tra volume dei vuoti comunicanti e volume totale del mezzo poroso. La porosità effettiva influenza il grado di esalazione del radon in quanto determina la frazione di volume occupato da pori aperti e liberi, disponibili per il trasporto e la diffusione del radon.

La porosità è influenzata oltre che dalla forma, dimensione e disposizione degli elementi, anche dalla loro composizione; per esempio la presenza di cemento in un sedimento, ne diminuisce la porosità fino da annullarla.

- permeabilità: l'unità di misura è il Darcy, pari alla permeabilità di una roccia che si lascia attraversare con la velocità di 1 cm/s da un liquido con viscosità $\nu = 10^{-3}$ (Pa sec), sotto un gradiente di pressione di 1 atm/cm (1darcy=0.987·10⁻¹² m²). Essa è la proprietà di un terreno o di una roccia

di lasciarsi attraversare dai fluidi senza che avvengano in essi alterazioni di struttura o spostamenti parziali. Questa proprietà dipende dalla natura del mezzo, dal fluido che l'attraversa e dalle condizioni dinamiche cui il fluido è sottoposto.

Per quanto riguarda la natura del mezzo, la permeabilità è funzione delle dimensioni e continuità dei vuoti: più i vuoti sono ampi e rettilinei, più facilmente avviene il passaggio del fluido.

La permeabilità secondaria, determinata dalla presenza di faglie e fratture che rappresentano vie preferenziali di trasporto del gas, può facilitare la fuoriuscita del ^{222}Rn in superficie. Questo avviene con più facilità se le faglie sono aperte, se non sono coperte da strati di terreno e se attraverso di esse vi è l'ascensione di gas o fluidi idrotermali che facilitano il trasporto del radon prima del suo decadimento.

- contenuto di acqua, in virtù dell'elevato potere frenante.

Nel trasporto per diffusione molecolare il radon si muove in direzione opposta a quella del gradiente di concentrazione; è possibile derivare l'espressione del flusso di radon espresso in $\text{Bq m}^{-2} \text{ s}^{-1}$, in determinate condizioni geometriche. Assumendo la terra come una massa porosa di materiale omogeneo semi infinita, il rateo di flusso di radon J_D che emerge dalla superficie può essere espresso dalla legge di Fick (UNSCEAR 1988; 1993; 2000):

$$J_D = C_{Ra} \lambda_{Rn} f \rho \sqrt{\frac{D_e}{\lambda_{Rn} \epsilon}} \quad (1.3)$$

C_{Ra} (Bq kg^{-1}) è la concentrazione di radio nel terreno, λ_{Rn} è la costante di decadimento del ^{222}Rn ($2.1 \cdot 10^{-6} \text{ s}^{-1}$); f è il coefficiente di emanazione del campione; ρ è la densità del materiale (kg m^{-3}); D_e è il coefficiente di diffusione effettivo per il materiale ($\text{m}^2 \text{ s}^{-1}$) e ϵ è la porosità; l'espressione sotto radice quadrata è la lunghezza di diffusione, mentre la parte restante costituisce il rateo volumetrico di produzione del radon ($\text{Bq m}^{-3} \text{ s}^{-1}$).

Utilizzando i valori appropriati, si stima un rate di flusso $J_D = 0.033 \text{ Bq m}^{-2} \text{ s}^{-1}$, con una lunghezza di diffusione di circa 1 m (UNSCEAR, 2000).

Tabella 1.4: Coefficienti di diffusione del radon per diversi materiali (UNSCEAR 1982, Nazaroff et al. 1992)

Mezzo	Coefficiente di diffusione D ($\text{m}^2 \text{ s}^{-1}$)
Aria	10^{-5}
Ciottoli - ghiaia grossolana	$10^{-5} - 5 \cdot 10^{-6}$
Sabbia asciutta	10^{-5}
Sabbia umida	$2.5 \cdot 10^{-6}$
Detrito	$5 \cdot 10^{-7} - 2.5 \cdot 10^{-7}$
Detrito argilloso	$8 \cdot 10^{-8}$
Acqua	10^{-9}

I coefficienti di diffusione in tabella 1.4 mostrano come il radon diffonde bene nei terreni composti da ghiaia grossolana e in aria, mentre la diffusione è bassa per l'argilla satura di acqua e in acqua. La distanza di diffusione ($\sqrt{D\tau}$, con τ la vita media del ^{222}Rn , tab. 1.5) in un terreno secco è ridotta di circa due ordini di grandezza quando il terreno è quasi saturo di acqua.

Tabella 1.5: Lunghezza di di diffusione del radon per diversi mezzi (Tanner,1964)

Mezzo	Distanza di diffusione media ^{222}Rn (cm)	Costante di diffusione ^{222}Rn (cm^2/s)
Aria	220	10^{-2}
Suolo poroso	155	$5 \cdot 10^{-2}$
Acqua	2.2	10^{-2}
Suolo poroso saturo	1.55	$5 \cdot 10^{-6}$

Per quanto riguarda i meccanismi di avvezione, essi sono dovuti a diversi

processi:

- emissione di gas generati da regioni calde, per es. vulcani;
- stress nei materiali rocciosi, che può essere generato dall'insorgenza di costrizioni che in genere precedono i terremoti e le eruzioni vulcaniche;
- avvezione fluida, che produce un flusso di radon per permeabilità dei materiali sufficientemente alte;
- flusso indotto da variazioni di pressione nelle vicinanze della superficie terrestre, legate alla meteorologia.

L'avvezione dipende dal gradiente di pressione e dalla permeabilità, K , che dalla legge di Darcy è:

$$K = \eta \frac{dV}{S dt} \left(\frac{dP}{dz} \right)^{-1} \quad (1.4)$$

dove η è la viscosità del fluido, $\frac{dV}{S dt}$ è il volume mosso per unità di massa nell'unità di tempo, S è la sezione del poro e $\frac{dP}{dz}$ è il gradiente di pressione. Si distingue una permeabilità assoluta ed una efficace o relativa per una data roccia: quella assoluta è la permeabilità della roccia rispetto al flusso di un gas secco (in assenza di acqua) o di un liquido unico, mentre la permeabilità relativa si riferisce ad un solo componente di un liquido misto. A differenza della porosità la permeabilità è una grandezza vettoriale ed è una quantità che può assumere un'ampia gamma di valori (da 10^{-8} m^2 per la ghiaia pulita fino a 10^{-16} m^2 per l'argilla, Nazaroff 1992), la presenza di crepe e fessure può aumentare di molto la permeabilità del mezzo.

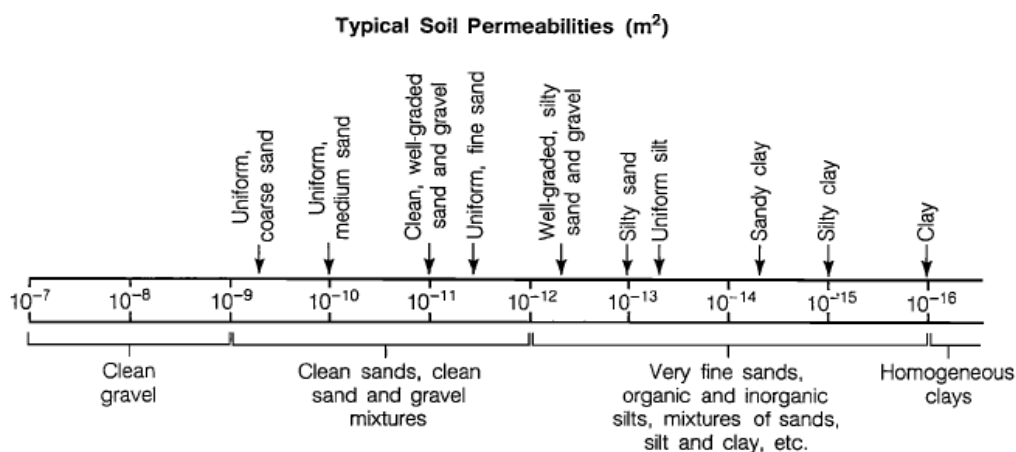


Figura 1.6: Permeabilità tipiche dei terreni (tratte da Nazaroff et al., 1992)

Come accennato in precedenza, infine, le aree ad alto gradiente geotermico, e in particolare le aree idrotermali, sono caratterizzate da movimento di fluidi verso la superficie che possono ulteriormente aumentare la possibilità di migrazione del radon. In presenza di tali parametri geologici si possono, quindi, sviluppare percorsi di migrazione lunghi, facendo sì che anche radon di origine profonda possa raggiungere la superficie.

1.2.3 Concentrazione di radon outdoor

La concentrazione di radon nell'ambiente esterno è influenzata oltre che dal rateo di esalazione del suolo anche da fenomeni di mixing atmosferici (UNSCEAR, 2000).

Il riscaldamento solare durante il giorno introduce delle turbolenze negli stati atmosferici così che il radon viene trasportato più velocemente verso l'alto. Durante la notte e nelle ore iniziali del giorno, le condizioni di inversione atmosferica tendono a bloccare il radon vicino al terreno e questo comporta una variazione giornaliera di radon fino ad un fattore 10 (UNSCEAR, 2000). Si osservano inoltre variazioni stagionali relative all'effetto delle precipitazioni e dei venti.

Risultati recenti delle misure di radon outdoor hanno portato a stimare un valore di concentrazione medio pari a 10 Bq m^{-3} (UNSCEAR, 2000).

Di Carlo et al. (2009) hanno condotto un biennio di misure della concentrazione di radon superficiale outdoor a L'Aquila, con lo scopo di evidenziare la relazione tra variabili meteorologiche e fisiche con le concentrazioni di radon misurate. Il lavoro ha rilevato un' importante anticorrelazione ($R=-0.54$) tra la concentrazione di radon e la velocità del vento superficiale, che evidenzia l'azione dinamica di riduzione dell'abbondanza del radon emesso dal suolo.

La temperatura e l' umidità del terreno incidono sul flusso di radon emesso (rispettivamente con correlazione negativa e positiva).

Radon and ...	Year	Winter	Spring	Summer	Fall
Wind speed	-0.54 ± 0.02	-0.50 ± 0.04	-0.61 ± 0.03	-0.57 ± 0.03	-0.49 ± 0.04
Temperature	-0.47 ± 0.02	-0.49 ± 0.04	-0.52 ± 0.03	-0.72 ± 0.02	-0.43 ± 0.04
Relative humidity	0.37 ± 0.02	0.41 ± 0.04	0.25 ± 0.04	0.58 ± 0.03	0.58 ± 0.03
Wind direction	0.17 ± 0.02	0.09 ± 0.05	0.13 ± 0.04	0.34 ± 0.04	0.02 ± 0.05
Ozone	-0.47 ± 0.02	-0.58 ± 0.03	-0.62 ± 0.03	-0.55 ± 0.03	-0.54 ± 0.03

All the data reported in this table are hourly averaged

Figura 1.7: Coefficienti di correlazione tra radon e parametri meteorologici e fisici (Di Carlo et al., 2009)

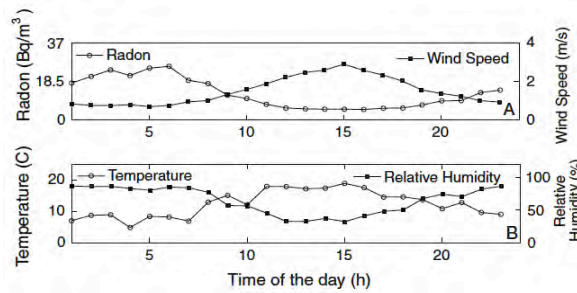


Figura 1.8: Ciclo diurno del radon superficiale e la velocità del vento (pannello a), temperatura e umidità relativa (pannello b) (Di Carlo et al., 2009)

1.2.4 Ingresso negli edifici

La quantità di radon che entra negli edifici dipende, come già detto, principalmente dalla concentrazione del gas negli strati più superficiali del suolo. Il meccanismo responsabile dell' ingresso del radon dalla superficie terrestre agli

ambienti chiusi è la differenza di pressione esistente tra interno ed esterno negli edifici stessi. Nel caso di un edificio in depressione rispetto all'esterno, si instaura un meccanismo che richiama aria dal suolo attraverso le fessure presenti nella struttura dell'edificio. La differenza di pressione tra interno ed esterno è causata dall'effetto camino e dall'effetto vento. L'effetto camino è generato dalla differenza di temperatura tra interno ed esterno, che genera a sua volta una differenza di pressione. L'effetto vento è dovuto invece alla differenza tra la velocità dell'aria all'esterno e quella all'interno, che determina una pressione sulle pareti (Ministero della Salute, 2002).

I meccanismi di ingresso del radon negli edifici sono influenzati da una serie di fattori locali e variabili nel tempo. Particolarmente accentuati sono gli effetti del clima e delle sue variazioni climatiche stagionali e giornaliere, che influenzano la pressione atmosferica, la temperatura, la ventilazione, l'umidità. Questi fattori influenzano a loro volta fortemente la diffusione del radon dal suolo verso la superficie ed il suo ingresso negli edifici.

Tipicamente i valori di concentrazione di radon all'interno di un edificio sono più elevati nel periodo dell'anno caratterizzato da temperature esterne minori, ovvero durante l'autunno e l'inverno.

Intense precipitazioni favoriscono invece il cosiddetto effetto tappo che consiste in una diminuzione del potere di emanazione di radon del suolo. Le piogge infatti infiltrandosi nel terreno tendono a occupare i pori interstiziali del materiale. Il trasporto del radon avviene così all'interno di un mezzo liquido e il flusso di radon dal suolo in atmosfera (esalazione) viene fortemente inibito essendo il coefficiente di diffusione del radon in acqua pari ad 1/10.000 di quello in aria. Soltanto in primavera, quando le temperature più alte avranno favorito l'evaporazione dell'acqua e quindi lo svuotamento dei pori, il suolo si riporterà al suo naturale rateo di emanazione (UNSCEAR, 2000).

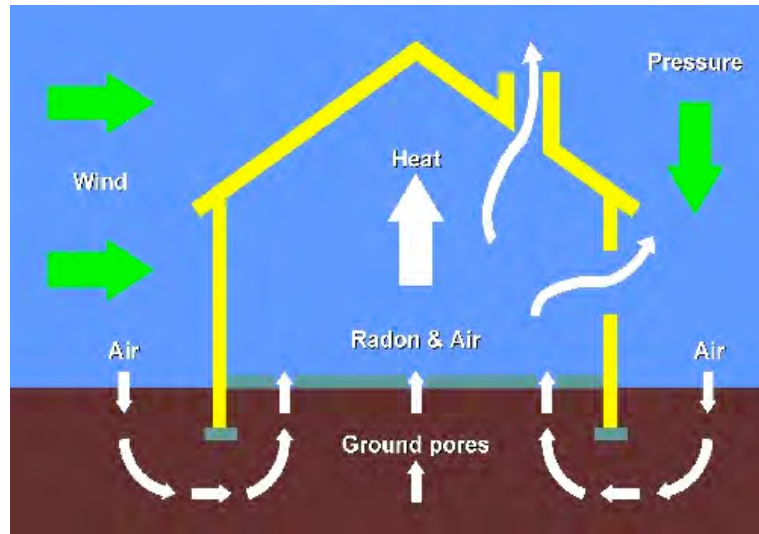
Le alte temperature invece favoriscono l'esalazione in atmosfera ma tale fenomeno è compensato da una conseguente dispersione del gas, questo comporta che la concentrazione media rimanga pressoché inalterata. L'esalazione del gas dal

suolo è inoltre legata alla pressione atmosferica, ma solo relativamente agli strati più superficiali, in cui la bassa pressione può facilitare il flusso di radon. Gli strati più profondi invece non sono soggetti a tale influenza ed inoltre gli stati di bassa pressione sono comunque accompagnati da piogge intense che, esercitando l'effetto tappo, impediscono l'esalazione del gas, compensando così gli effetti legati alla bassa pressione (NCRP, 1988).

Effetto Camino : è dovuto alla differenza di temperatura tra interno ed esterno dell'abitazione, a seguito della quale si forma una differenza di pressione ΔP ; quanto più caldo è l'interno dell'edificio, tanto maggiore sarà il flusso di aria fredda risucchiata dal terreno; il ΔP è dato dalla formula:

$$\Delta P(z) = \mu \left(\frac{1}{t_{est}} - \frac{1}{t_{int}} \right) (z_0 - z) \quad (1.5)$$

dove μ è una costante pari a 3462 Pa K m^{-1} , t_{est} e t_{int} sono le temperature rispettivamente interna ed esterna, misurata in K. In base alla (1.5), con una differenza di temperatura pari a 30 K, la differenza di pressione sarà di 1.3 Pa; questo implica che attraverso una fessura larga 1 mm e lunga qualche metro, possono venir aspirati diversi metri cubi di aria all'ora (Ministero della Salute, 2002). Il funzionamento di una stufa o di sistemi di aspirazione in bagni o cucine possono produrre un effetto di risucchio e far quindi aumentare le concentrazioni di radon all'interno degli edifici;



Effetto Vento : è dovuto alla differenza di velocità dell'aria tra esterno e interno dell'abitazione; la pressione P esercitata su una parete è data dalla relazione:

$$P = P_0 + C_P \left(\frac{1}{2} \rho v^2 \right) \quad (1.6)$$

dove P_0 indica la pressione statica del vento, v la sua velocità, ρ la densità dell'aria e C_P un coefficiente empirico di pressione e dipende da una moltitudine di fattori (direzione del vento, presenza di aperture, forma dell'edificio).

Per fare un esempio, alla velocità di 5 m s^{-1} e con una temperatura di $10 \text{ }^\circ\text{C}$, viene indotta una depressione pari a $\Delta P = -5 \text{ Pa}$. Ci si aspetta che la concentrazione di radon nelle abitazioni sia maggiore con l'incremento della velocità del vento. Tuttavia, se la ventosità è elevata, l'aumento del radon indoor è ridotto a causa dell'incremento a loro volta delle infiltrazioni d'aria nell'ambiente mai propriamente ermetico.

In base a quanto esposto, la concentrazione di radon nelle abitazioni quindi può subire notevoli variazioni, sia giornaliere che stagionali; in generale, valori più elevati si osserveranno al mattino e in inverno — quando cioè la differenza di temperatura con l'esterno è maggiore. Inoltre, si è osservato che la concentrazione aumenta quando il terreno ghiaccia: è ragionevole aspettarsi infatti che lo strato di ghiaccio ostacoli il normale flusso dal terreno di gas, che quindi diffonde

attraverso zone non gelate, come il pavimento delle cantine o attraverso crepe. L'ingresso di radon negli edifici è fortemente influenzato anche dalle caratteristiche di permeabilità del suolo sul quale vengono costruiti. Questo è il caso di edifici le cui fondamenta vengono coperte con uno strato di materiale ad alta permeabilità (ad esempio ghiaia) che può alterare le caratteristiche diffusive locali e permettere un maggior flusso di radon in entrata.

Fratture o crepe di qualsiasi dimensione nel suolo e nei materiali delle fondamenta svolgono un ruolo importante nel determinare la concentrazione di radon indoor. Nei piani interrati, deve essere presa in considerazione anche la diffusione del radon attraverso il cemento delle pareti. Nazaroff e Teichmann (1990) proposero un'equazione di equilibrio in uno stato stazionario da applicare alla concentrazione di radon negli edifici:

$$Rn_{in} = \frac{S_s + S_w + S_b + Qrn_{out}}{Q + \lambda_{Rn}V + rV} \quad (1.7)$$

dove Rn_{in} e Rn_{out} sono rispettivamente le concentrazioni di radon indoor e outdoor espresse in Bq m^{-3} , S_s , S_w , S_b sono i flussi di radon in ingresso, rispettivamente dal suolo, dall'acqua e dal materiale da costruzione in Bq s^{-1} , Q è il flusso di aria esterna entrante nell'edificio ($\text{m}^3 \text{s}^{-1}$), V è il volume dell'edificio in m^3 , e λ_{Rn} è la costante di decadimento del radon ($2.1 \cdot 10^{-3} \text{s}^{-1}$).

I contributi relativi delle varie fonti di radon variano fortemente da un edificio all'altro. UNSCEAR (2000) stima l'importanza relativa di ogni fattore per un'abitazione singola standard in muratura e in un clima temperato; il radon entrante dal terreno sottostante, per lo più da avvezione attraverso delle discontinuità nel pavimento, costituisce il 60% del totale. La diffusione dai materiali da costruzione e le infiltrazioni di radon con l'aria esterna possono giocare un ruolo importante (superiore al 20% e 18% rispettivamente), mentre le altre sorgenti (acqua e gas naturale) sono relativamente meno importanti.

Ci si aspetta che le concentrazioni di radon diminuiscano ai piani alti in cui i contributi maggiori saranno dei materiali di costruzione e dell'aria esterna (nel

territorio aquilano è stato rilevato un valore medio annuo di radon outdoor di circa 12 Bq m^{-3} (Di Carlo, 2002)); a volte il radon entrante dal terreno può raggiungere in misura significativa i piani superiori in virtù di particolari pattern di circolazione dell'aria all'interno dell'edificio, che sfruttano, per esempio, la tromba delle scale, il pozzo dell'ascensore o cavedi per il passaggio di tubazioni etc.

1.3 Effetti sulla salute e normativa

L'esposizione al radon ^{222}Rn e ai suoi prodotti di decadimento è il secondo fattore di rischio, dopo il fumo, per l'insorgenza del tumore polmonare. L'Organizzazione Mondiale della Sanità (WHO), sulla base delle risultanze di indagini epidemiologiche, riguardanti principalmente minatori fortemente esposti al radon, ha classificato il gas come appartenente al gruppo 1 delle sostanze cancerogene, ovvero tra quelle che hanno il massimo livello di certezza dell'effetto. Gli studi epidemiologici sull'esposizione domestica al radon (Darby et al., 2005) hanno mostrato che la soglia di rischio, se esiste, è inferiore a 150 Bq m^{-3} (in termini di concentrazione media annua di radon) e che un incremento di 100 Bq m^{-3} nella concentrazione indoor del gas comporta un aumento del rischio di tumore polmonare del 16%.

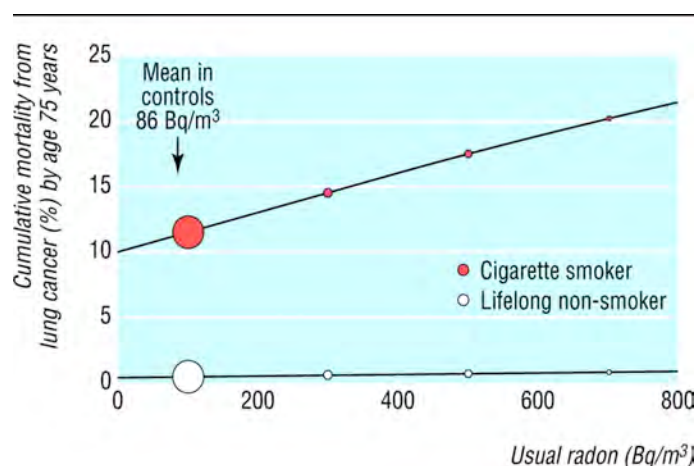


Figura 1.9: Mortalità cumulativa per tumore polmonare all'età di 75 anni in relazione alla concentrazione di radon nelle abitazioni per fumatori e non fumatori. (Darby et al., 2005)

Gli effetti sulla salute sono principalmente dovuti ai prodotti di decadimento a vita breve più che al radon stesso che ha un'emivita di circa 4 giorni e nella maggior parte dei casi se inalato viene esalato prima che possa decadere.

Al contrario, i prodotti di decadimento ^{218}Po (3.05 min), ^{214}Pb (26.8 min), ^{214}Bi (19.9 min) e ^{214}Po (164 μs) sono atomi e ioni metallici che si legano facilmente con gli aerosol ambientali e se inalati si depositano lungo le vie aeree dell'albero bronchiale e decadono rilasciando particelle α (nel processo di decadimento del ^{218}Po e ^{214}Po) che danno il maggior contributo all'esposizione totale. Le particelle β e γ prodotte dal decadimento di ^{214}Pb e ^{214}Bi non incidono in maniera significativa sull'esposizione. Le particelle α , avendo un range piccolo (decine di μm), irradiano le cellule dell'epitelio bronchiale danneggiando il loro DNA e aumentando così il rischio di tumore. Per quanto concerne, invece, il thoron (^{220}Rn) questo è caratterizzato da un tempo di dimezzamento breve (55.6 s) che non gli permette di spostarsi molto da dove è stato prodotto per cui, data la scarsa concentrazione nel suolo, il rischio associato ad esso e ai suoi prodotti di decadimento è basso.

Al fine di avere stime certe del rischio associato al radon, sono stati condotti molti studi a livello internazionale e le maggiori istituzioni scientifiche raccoman-

dano vari livelli di riferimento o di azione (in termini di concentrazione media annua di radon) sopra i quali bisogna intervenire con azioni di rimedio per ridurre la concentrazione di radon e sotto i quali il rischio invece viene considerato accettabile. Negli anni questi valori si modificano con l'aumento delle conoscenze sull'argomento.

Recentemente il WHO (2009) nell' *Handbook on indoor radon* ha proposto un livello di riferimento di 100 Bq m^{-3} come strumento efficace di riduzione del rischio per la popolazione legato al radon. Si tratta di un livello particolarmente restrittivo che potrebbe essere difficile adottare in paesi con determinate caratteristiche; in questi casi il WHO raccomanda un livello di riferimento comunque non superiore a 300 Bq m^{-3} , valore che corrisponde a circa 10 mSv in termini di dose efficace per anno.

Uno studio del WHO su 36 paesi, ha evidenziato che quasi tutti hanno posto dei livelli di riferimento tra i 200 Bq m^{-3} e i 400 Bq m^{-3} per gli edifici esistenti (WHO, 2007). Altri paesi hanno scelto invece livelli differenti per le case nuove e per quelle già esistenti, scegliendone di più alti per le seconde perchè le azioni di rimedio sono più costose e meno efficaci delle azioni preventive.

La normativa europea attraverso la raccomandazione 90/143/Euratom del 1990, sulla base dell'ICRP no.39 del 1984, tutela la popolazione contro l'esposizione al radon negli ambienti residenziali, fissando specifici livelli di riferimento. Per gli edifici esistenti si fissa un livello di riferimento pari a 400 Bq m^{-3} , corrispondente a 20 mSv/anno, superato il quale si raccomanda di effettuare azioni di rimedio per ridurre la concentrazione di radon; per gli edifici da costruire si fissa un obiettivo progettuale di 200 Bq m^{-3} , equivalenti a 10 mSv/anno.

In Italia la legislazione per il rischio radon è limitata al Decreto Legislativo 241 del 2000, che recepisce la Direttiva europea 96/29/Euratom, sulla protezione della popolazione e dei lavoratori dai rischi derivanti dalle radiazioni ionizzanti. Il D.Lgs 241 si occupa del rischio radon per le sole attività lavorative, individuando un livello di azione di 500 Bq m^{-3} . Lo stesso Decreto assegna alle autorità

regionali il compito di individuare eventuali aree ad elevata probabilità di alte concentrazioni di radon indoor nell'ambito del territorio regionale (le cosiddette radon-prone areas).

Entro la fine del 2013, è prevista la pubblicazione di una nuova raccomandazione *Basic Safety Standards* (BSS) Euratom (Commissione Europea per l'Energia Atomica), che, sulla base degli ultimi studi epidemiologici, rivede i limiti di azione per le concentrazioni di radon indoor. Questa diventerà una legge europea e dovrà essere recepita da tutti i Paesi membri, Italia compresa. La proposta di Direttiva indica come massimo livello di riferimento della concentrazione di radon nelle abitazioni il valore di 300 Bq m^{-3} per gli edifici esistenti e di 200 Bq m^{-3} per gli edifici di nuova costruzione.

Nelle stessa proposta BSS si richiama l'esigenza di individuare le radon-prone areas. La Def. 71, dell'Art. 4, riporta: *una radon prone area è un' area geografica o una regione amministrativa, individuata sulla base di indagini che indicano che la percentuale attesa di edifici superiore al livello di riferimento nazionale sia significativamente più alta rispetto ad altre parti del paese* (EC, 2011).

Capitolo 2

Campagne di misure di concentrazione radon indoor in Abruzzo.

In Italia le prescrizioni della normativa nazionale (D.Lgs 241/2000) hanno portato molte Regioni a sviluppare delle campagne di misura di radon nelle abitazioni, mirate all'individuazione delle radon-prone areas nel territorio della provincia dell'Aquila.

I dati utilizzati nel lavoro di tesi sono l'insieme delle misure effettuate in diverse indagini regionali portate a termine nel corso dell'ultimo ventennio. Tutte le informazioni legate alla misura di concentrazione di radon indoor (coordinate geografiche, altitudine, caratteristiche dell'edificio, piano dove viene posizionato il dosimetro, caratteristiche litologiche del suolo sottostante, ecc...) sono raccolte e conservate nel database di ARTA.

Nel capitolo si riassumono brevemente gli scopi e risultati delle varie indagini di misura susseguitesì dal 1990 a oggi.

Ancor prima delle disposizioni legislative (D.Lgs 241/2000), negli anni novanta in Abruzzo, come in altre regioni italiane, vennero avviate due campagne

di misure: una nazionale, che intendeva stimare la distribuzione delle concentrazioni annuali di radon nelle abitazioni italiane (Bochicchio et al. 1996), l'altra invece negli asili regionali.

- *Indagine Nazionale Radon 1992 (INR)*

I dati relativi a questa campagna, che costituì il contributo abruzzese all'indagine nazionale svolta nel 1990-92 sotto il coordinamento di ISS e ENEA-DISP (Bochicchio et al., 1996), interessò 7 comuni (6 comuni estratti a caso più Pescara, che partecipava di diritto all'indagine in quanto unico comune abruzzese con popolazione superiore a 100.000 abitanti), e la maggior parte dei quali ricadenti nella provincia dell'Aquila.

In ciascun comune vennero estratte a sorte un certo numero di abitazioni nelle quali furono collocati dosimetri per due semestri consecutivi, al fine di ottenere stime della concentrazione media annua di attività di radon. La media aritmetica dei valori misurati si attestò sui 61 Bq m^{-3} , leggermente inferiore alla media nazionale stimata da Bochicchio et al. di 70 Bq m^{-3} .

Provincia	Comuni interessati	n. siti (abitazioni)	media geometrica (GM) Bq/m^3	dev.standard geometrica (GSD) Bq/m^3	media aritmetica (AM) Bq/m^3	dev.std. (σ) Bq/m^3
PE	Pescara	11	23,0	1,4	24,3	9,8
CH	Crecchio	11	63,6	2,0	83,8	89,3
TE	---	0
AQ	Barisciano	36	37,4	1,70	43,3	34,0
	Lecce nei Marsi	13	69,7	2,20	91,8	70,9
	Pratola Peligna	10	40,8	1,50	46,1	28,9
	Raiano	14	48,9	2,00	64,8	115,8
	S. Vincenzo Valle Roveto	8	63,9	2,50	100,1	78,5
ABRUZZO		103	45,2	1,9	60,8	67,6

Figura 2.1: Indagine nazionale radon, 1991-92: comuni abruzzesi interessati e statistiche relative. (ARTA, 2008)

- *Indagine Regionale Asili 1993-1995 (AS)*

Campagna di misura effettuata negli asili della Regione negli anni 93-95 (ARTA, 2008). Tali edifici furono ritenuti di particolare interesse per la collocazione in prossimità del suolo (generalmente i locali destinati a tali

strutture sono situati ai piani bassi), idonea per una valutazione più diretta, rispetto alle campagne condotte nelle normali abitazioni, dell'influenza del tipo di suolo sulla concentrazione di radon indoor.

La media aritmetica dei valori misurati si attesta su 66 Bq m^{-3} . La differenza rispetto al valore trovato nell'Indagine Nazionale Radon (INR), pari a 61 Bq m^{-3} , si spiega, principalmente, osservando che le misure svolte nell'ambito della INR riguardano abitazioni poste a vari piani, mentre gli asili si trovano in maggioranza a piano terra, dove ci si aspetta di trovare concentrazioni più elevate di radon. La concentrazione media risultò massima nella zona appenninica (coincidente con la provincia dell'Aquila e con i territori delle altre province posti ad altitudini superiori ai 500 m), con un valore medio annuo di $87,2 \text{ Bq m}^{-3}$, mentre nella fascia costiera e sublitoranea si attestò su un valore sensibilmente inferiore ($57,1 \text{ Bq m}^{-3}$).

Provincia	n. siti (asili)	media geometrica (GM) - Bq/m^3	dev.standard geometrica (GSD) - Bq/m^3	media aritmetica (AM) - Bq/m^3	dev.std. (σ) Bq/m^3
AQ	104	69,9	2,1	95,1	87,7
TE	124	45,8	2,2	62,2	58,1
PE	108	52,0	1,8	61,3	38,2
CH	163	41,5	2,1	55,0	51,0
ABRUZZO	499	49,8	2,1	66,5	61,8

Figura 2.2: Indagine Regionale Asili 1993-1995: statistiche relative alle provincie (Arta, 2008)

Le misure coprirono l'arco di tempo di un anno, suddiviso in due semestri (primavera-estate e autunno-inverno), al fine di valutare la variabilità stagionale.

Abruzzo Radon Project, ARP1.

Con l'intento di rappresentare in maniera più comprensiva la distribuzione di concentrazione di radon indoor, l'ARTA Abruzzo intraprese una prima campagna di misure nel 2005-2008 (chiamata Abruzzo Radon Project, ARP1) (Palermi e Pasciulli, 2008). In mancanza infatti di criteri normativi specifici, ciascuna Re-

gione attuò un proprio piano di campionamento e la maggior parte furono basati sulla suddivisione del territorio in unità geografiche di tipo geometrico (p.es. le aeree definite da una griglia di passo costante o variabile) o amministrativo (tipicamente i territori comunali), (Ministero della Salute, 2002).

L'indagine ARP1 era basata su un campionamento di tipo spaziale su unità geografiche di natura geometrica nelle quali era stato suddiviso il territorio abruzzese (maglie quadrate 6 km x 6 km e sottomaglie 2 km x 2 km) In ciascuna sottomaglia o cella andava effettuata la scelta casuale di almeno un sito di misura (abitazioni/luoghi pubblici/luoghi di lavoro). Nel caso in cui una o più celle 2 x 2 risultassero prive di abitazioni (cosa non infrequente nella regione appenninica), occorreva infittire il campionamento nelle celle adiacenti.

PIANO	AQ		TE		PE		CH		ABRUZZO	
	N. SITI	MEDIA RADON (Bq/m3)	N. SITI	MEDIA RADON (Bq/m3)	N. SITI	MEDIA RADON (Bq/m3)	N. SITI	MEDIA RADON (Bq/m3)	N. SITI	MEDIA RADON (Bq/m3)
-1	31	103,53	11	46,33	8	36,15	3	34,88	53	77,60
0	287	89,17	258	61,37	313	49,57	321	64,02	1179	65,73
1	46	56,37	55	30,43	88	25,50	139	39,48	328	36,58
2	28	37,59	9	17,53	5	34,91	26	24,01	68	29,54
3	2	38,70	2	14,49	1	10,74	6	17,99	11	20,46
4	1	30,20	0	-	1	20,20	1	10,48	3	20,29
5	2	54,15	0	-	3	20,90	0	-	5	34,20

Figura 2.3: Abruzzo Radon Project 2005-2008: valori medi di concentrazione di radon indoor per provincia e piano (-1: sotterraneo/seminterrato) L'analisi comprende anche i dati delle precedenti campagne di misura. (fonte: *Quaderno Arta, 2008*)

Abruzzo Radon Project 2, ARP2.

Una riflessione critica sui risultati di questa campagna, non completamente soddisfacenti perché poco rappresentativi del patrimonio edilizio di tutta la regione, portò alla progettazione di una seconda indagine (Abruzzo Radon Project 2, ARP2, dal 2009 ad oggi non ancora conclusa) mirata ad assicurare un campionamento più omogeneo del territorio (Palermi et al., 2012). L'obiettivo principale era giungere ad una caratterizzazione della distribuzione territoriale della concentrazione di radon nelle abitazioni, che sia rappresentativa dell'esposizione della

popolazione abruzzese, al fine di consentire una valutazione adeguata dei rischi sanitari connessi. Si propose, inoltre, di elaborare una previsione quantitativa della distribuzione sull'intero territorio abruzzese di valori superiori a determinati livelli di soglia, in ottemperanza al art. 10-sexies del D.Lgs 241/2000 (individuazione delle zone della regione e delle tipologie edilizie in cui vi sia maggiore probabilità di osservare elevate concentrazioni di radon).

Nel progettare un'indagine sperimentale idonea al conseguimento dei questi obiettivi, nell' ARP2 si è adottato uno schema multistadio e stratificato seguendo l'esempio di numerosi altri studi (come in Bochicchio et al., 1996).

La stratificazione è un'operazione preliminare al campionamento e consiste nel raggruppamento dei membri di una popolazione in sottogruppi relativamente omogenei, rispetto ad una o più caratteristiche ritenute critiche ai fini dell'indagine. Si tratta di un approccio che assicura che membri di tutti gli strati siano presenti all'interno del campione ed è inoltre da preferire nei casi in cui si assume che i vari strati abbiano valori attesi della variabile di interesse molto diversi tra loro (Malec, 2000).

Nei piani di campionamento multistadio si procede prima all'estrazione di un campione casuale di cosiddette unità primarie di campionamento (Primary Sampling Units, PSU), che nel nostro caso coincidono con i comuni, all'interno delle quali vengono poi estratti campioni di unità secondarie (Listing Units, LSU) (Shimizu, 2000), le abitazioni delle famiglie residenti nei comuni stessi.

La scelta di utilizzare i comuni si basa sulla disponibilità di registri anagrafici a livello comunale, mentre la possibilità di concentrare i campionamenti in poche PSU adeguatamente selezionate riduce considerevolmente costi e tempi di attuazione dell'indagine.

Tra i risultati, emergono valori medi comunali che si attestano prevalentemente tra 34 e 55 Bq m⁻³ nella zona appenninica e tra 25 e 35 Bq m⁻³ nella zona litoranea, dati che inducono a stime preliminari del valore medio regionale non superiori a 40 Bq m⁻³, dunque significativamente inferiori ai 60 Bq m⁻³ stimati dall'Indagine nazionale radon dei primi anni 90. Questa conclusione non deve

sorprendere, in quanto quell'indagine, finalizzata a stimare un valore medio rappresentativo a livello nazionale, non era idonea a fornire stime accurate a livello regionale, soprattutto nei casi in cui la numerosità complessiva del campione era ridotta (in Abruzzo circa 100 abitazioni distribuite in soli 7 comuni, quasi tutti appartenenti all'area interna appenninica) (Palermi et al, 2012).

Ora, con lo scopo di definire la nuova variabile del potenziale radon geogenico, abbiamo raggruppato tutti i dati disponibili dalle varie campagne regionali e nazionali svolte fino ad oggi. Ponendo l'attenzione sulla sola provincia aquilana, che ha una superficie di 5035 km², abbiamo un totale di 757 misure, effettuate in edifici appartenenti a 96 comuni (su 108 totali). Nella tab. (2.1) e nella fig. (2.4) sono riportate le statistiche delle misure per alcuni dei comuni della provincia. La distribuzione dei valori misurati di concentrazione di radon in fig.(2.5) è consistente con una distribuzione lognormale, caratteristica propria dei dataset di radon indoor (Miles et al., 1997).

Comune	n	AM	SD	GM	GSD	MAX	MIN
Sulmona	56	56.9	67.5	39.8	2.1	430	10
L'Aquila	54	68.6	71.6	47.9	2.3	393	11.6
Pratola Peligna	40	47.0	33.7	41.3	1.6	187.2	18.5
Celano	35	60.8	65.4	45.4	2.0	340	20
Magliano de Marsi	28	62.4	65.9	46.1	2.1	350.5	18
Castel di Sangro	25	43.4	18.3	39.6	1.6	79.1	17
Avezzano	23	46.6	26.5	38.9	1.9	95.8	11.3
Bugnara	22	56.8	37.8	47.5	1.8	162	20
...
tot	757	76.3	90.4	52.5	2.2	772.4	10.0

Tabella 2.1: Statistiche delle misure di concentrazione di radon indoor (Bq m⁻³) per alcuni dei 96 comuni. n: il numero di dati acquisiti; AM: la media aritmetica; SD: deviazione standard; GM: media geometrica; GSD: deviazione standard geometrica; MAX e MIN: valori massimi e minimi.

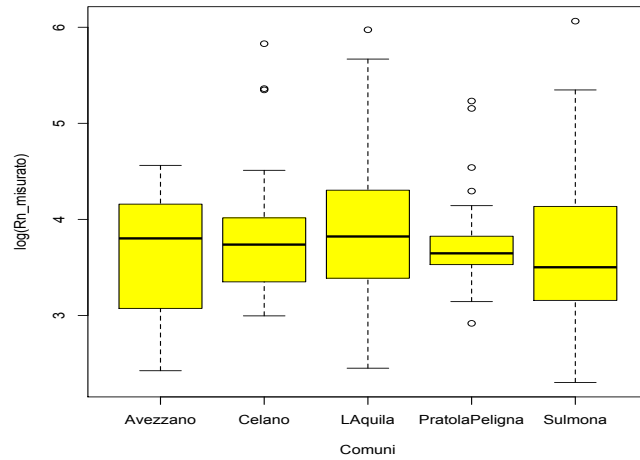


Figura 2.4: Boxplot del logaritmo della misura di concentrazione radon per i 5 comuni più grandi della provincia aquilana. Il rettangolo è delimitato dal primo e dal terzo quartile, $q_{1/4}$ e $q_{3/4}$, e diviso al suo interno dalla mediana, $q_{1/2}$. I segmenti sono delimitati dal minimo e dal massimo dei valori.

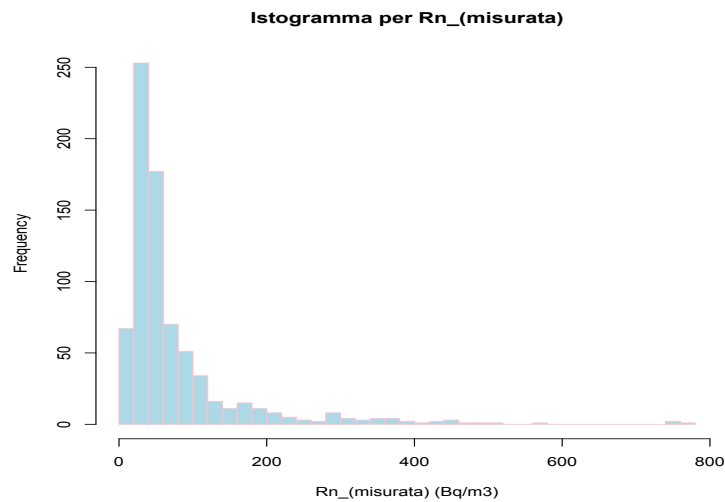


Figura 2.5: Distribuzione delle misure di concentrazione radon indoor.

Ciascuna misura raccolta nel database di ARTA è corredata da un insieme di informazioni necessarie per la corretta valutazione del dato stesso. Oltre ad essere specificate le coordinate geografiche dell'edificio, viene annotata la tipologia, il materiale con cui è stato edificato, il periodo in cui è stato costruito, l'altitudine,

Capitolo 2. Campagne di misure di concentrazione radon indoor in Abruzzo.

il piano in cui viene fatta la misura etc. Ad ogni dato successivamente vengono aggiunte le informazioni geologiche del territorio che possono incidere sul risultato della misura. Infatti il dato di concentrazione di radon indoor misurato, come visto nel precedente capitolo, dipende da una moltitudine di fattori di natura sia geologica (contenuto di uranio nel suolo, permeabilità etc), sia edilizia (tipologia costruttiva, materiali impiegati etc), sia comportamentale (tasso di ventilazione negli ambienti, differenza di temperatura con l'esterno etc). Nella tabella (2.2) vengono riassunte le misure di concentrazione del gas, raggruppandole in base alle caratteristiche degli edifici dove sono state effettuate.

	caratteristiche	n	AM	SD	GM	GSD	MAX	MIN
	0CT	298	102.45	116.12	68.02	2.36	772.40	10.80
	0NCT	130	80.59	90.85	53.96	2.31	485.00	11.30
piano	1° piano	203	44.97	27.25	38.85	1.70	170.60	10.00
	piano>1	82	35.11	17.46	32.28	1.48	147.00	14.00
	sotterraneo	44	107.74	89.69	82.20	2.09	431.00	18.50
tipologia	abitazione	496	65.09	79.26	46.19	2.08	753.00	10.00
	luogo di lavoro	89	100.48	120.86	62.32	2.55	772.40	13.20
	pubblico	45	79.14	76.28	58.88	2.10	431.00	15.20
	scuola	127	102.11	102.66	73.42	2.18	749.00	19.00
anno di costr.	post 1960	430	68.22	77.40	48.56	2.12	749.00	10.00
	prima del 1919	142	68.75	77.96	49.63	2.09	578.00	10.80
	tra 1919 e 1960	185	100.87	118.94	65.54	2.38	772.40	13.20
mat. di costr.	lat-cem	505	75.57	86.67	52.42	2.19	753.00	10.00
	pietra	225	82.58	102.07	55.49	2.24	772.40	10.80
	altro	27	37.46	18.36	33.38	1.64	83.00	14.00

Tabella 2.2: Statistiche per le misure di radon indoor in Bq m^{-3} . *0CT*: misura effettuata al piano terra di un edificio in contatto con il suolo; *0NCT*: piano terra non in contatto con il terreno (per es. è presente un locale sotterraneo o un'intercapedine (p.es. un vespaio)); *lat-cem*: laterizi o cemento; *altro*: per es. legno.

I dosimetri

I dosimetri utilizzati per le campagne di misure condotte fino ad oggi sono dei rivelatori a tracce passivi sensibili a radiazioni alfa. I rivelatori a tracce hanno caratteristiche eccellenti di basso costo, semplicità, robustezza e risposta sostanzialmente indipendente dalle condizioni ambientali.

Nelle indagini condotte negli anni '90 vennero impiegati dei dosimetri LR-115, in cui l'elemento sensibile è un film sottile di nitrato di cellulosa. La radiazione α emessa durante il decadimento del ^{222}Rn e dei suoi prodotti di decadimento, adeguatamente rallentata, impressiona la pellicola e una volta sviluppata fornisce informazioni sulla concentrazione del gas.

Nelle campagne di misura condotte dall'ARTA dal 2005 a oggi, vengono impiegati invece i CR-39 (Columbia Resin 1939), anch'essi rivelatori passivi. Il CR-39 è un polimero plastico sensibile a particelle α di energia compresa tra 0.1 MeV e 60 MeV. La radiazione α colpendo il CR-39 deposita energia provocando un canale di danno sulle molecole, nota come *traccia latente*, visibile al microscopio dopo un processo di lavorazione chimica. Al contrario dei LR-115, i CR-39 sono depositati all'interno di un apposita camera diffusiva in prolipropilene affinché la valutazione quantitativa delle tracce sia attribuibile al solo ^{222}Rn e ai suoi prodotti di decadimento.

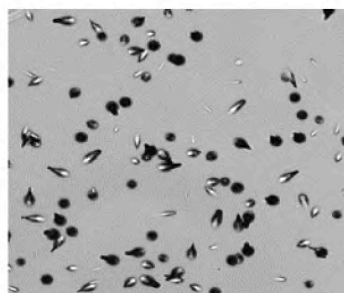


Figura 2.6: Dosimetro passivo utilizzato da ARTA e particolare dell'immagine fornita dal microscopio delle singole tracce delle particelle α (ARTA, 2008)

Capitolo 3

Analisi multilevel per le concentrazioni di radon indoor nella provincia aquilana

Nel capitolo si analizza la dipendenza della misura di concentrazione radon indoor dalle peculiarità dell'edificio dove viene svolta. Lo scopo è di ridefinire una variabile standardizzata al netto dalle caratteristiche dell'edificio. Questo nuovo valore di concentrazione del gas sarà poi utilizzato nell'analisi spaziale. Si studia anche la possibile relazione tra la classe litologica di appartenenza del sito in cui è posto l'edificio e la misura di radon.

Per predire il ruolo delle caratteristiche degli edifici e delle informazioni geologiche sulla misura di concentrazione di radon indoor, utilizziamo, come fatto da altri autori (Apte et. al, 1999; Andersen et al., 2001), una tecnica statistica conosciuta come analisi bayesiana multilevel. Come menzionato nel capitolo precedente, le indagini radon non sono basate su un campione random di una popolazione omogenea, ma sono spesso misure annidate in sottogruppi eterogenei. Nel nostro studio le unità più grandi sono i comuni e all'interno di queste ci sono le unità individuali (l'edificio) dove viene svolta la misura. Il modello multilevel,

in questo caso, diversamente dall'approccio classico, permette di esplorare i dati di radon a livello comunale, facilitando la valutazione delle sorgenti di variabilità (tra e all'interno dei gruppi).

Il modello completo per predire il logaritmo della concentrazione di radon dell'abitazione i -esima nel comune j -esimo, y_{ij} , è dato dalla relazione (Apte, 1999):

$$y_{ij} = X_{ij}\beta_j + \sum_k f_{jk}\gamma_k + \alpha_j + \epsilon_{ij} \text{ per } i = 1, \dots, n \quad (3.1)$$

$$\text{o } y_{ij} \sim N(\mu + \alpha_j + \sum_k f_{jk}\gamma_k, \sigma_y^2) \text{ per } i = 1, \dots, n \quad (3.2)$$

$$\alpha_j \sim N(0, \sigma_\alpha^2) \text{ per } j = 1, \dots, J \quad (3.3)$$

$$\gamma_k \sim N(0, \sigma_\gamma^2) \text{ per } k = 1, \dots, K \quad (3.4)$$

$$\epsilon_{ij} \sim N(0, \sigma_y^2) \quad (3.5)$$

in cui X_i è il vettore di variabili esplicative per l'edificio i e nel comune j . Le variabili in X che influenzano la misura di concentrazione radon indoor possono essere: il piano nel quale viene fatta la misura, la tipologia di abitazione (scuola, abitazione, edificio pubblico ecc...), l'anno di costruzione dell'edificio e i materiali utilizzati per edificarlo.

Il vettore β è composto dai coefficienti associati con le variabili esplicative X , essi possono variare in ciascun gruppo (come nell' eq. (3.2)) o essere costanti.

La variabile α_j è l' *effetto random del comune* ed è il valore di cui differisce il logaritmo della media geometrica vera della concentrazione di radon indoor per il comune j dalla stessa quantità predetta utilizzando il modello con le variabili esplicative.

I parametri $\{\gamma_k\}$ sono gli *effetti random dell'unità geologica* sulla concentrazione di radon che può essere diversa a seconda dell' unità litologia a cui appartiene l'abitazione. Questo è un ulteriore raggruppamento, *livello*, nel modello statistico, insieme al comune, specificato dalla variabile k (come in Apte et al.); f_{jk} è

la frazione di comuni j che appartengono alla k – esima classe litologica. Approfondiremo nei successivi paragrafi l’influenza di ciascuna variabile.

Il termine predetto della concentrazione di radon generalmente differisce dal valore vero e l’errore residuo è espresso da ϵ_{ij} (eq. 3.5), di cui si assume una distribuzione normale (nel log-spazio) con media nulla e varianza σ_y^2 stimata dai dati.

3.1 Analisi multilevel e variabile radon standardizzata

I modelli multilevel nascono dalla necessità di analizzare dati con struttura gerarchica, ovvero dati relativi a popolazioni in cui le unità individuali sono contenute in aggregati di livello superiore, i quali, a loro volta, possono essere contenuti in ulteriori unità sempre più ristrette in numero e più ampie in dimensione.

Le variabili possono essere definite ad ogni livello della gerarchia: alcune possono essere misurate direttamente al loro livello naturale, mentre altre possono essere spostate da un livello all’altro mediante procedure di aggregazione o disaggregazione.

Per aggregazione si intende che le variabili definite ad un livello più basso della gerarchia sono trasferite al livello immediatamente superiore; contrariamente la disaggregazione prevede che si assegnino, alle unità di livello inferiore, variabili osservate a livello superiore. Esempi di questa struttura gerarchica riguardano studenti raggruppati in classi e per scuola di appartenenza o distretto scolastico, oppure aggregazioni di cittadini per quartiere di residenza inserite in unità sempre più ampie quali il comune, la regione o lo stato.

Nelle varie campagne di misura di radon indoor svolte fino ad ora si è scelto di analizzare i dati raccolti nelle abitazioni raggruppati in comuni e poi in zone della regione, in cui i livelli di radon attesi sono sostanzialmente diversi (area interna appenninica e area collinare-costiera).

I modelli multilevel, conosciuti come modelli a effetti misti o a effetti random,

sono una generalizzazione della regressione lineare. Questi hanno lo scopo di spiegare o descrivere il valore atteso di una variabile dipendente, condizionato ad un insieme di variabili esplicative, attraverso un'opportuna funzione. Inoltre, in questi modelli la funzione è specificata in modo tale da considerare esplicitamente la struttura gerarchica dei dati, cercando di integrare l'analisi a livello individuale con quella a livello aggregato.

L'idea che sta alla base dei modelli multilevel è quella di considerare non un'unica equazione per tutti i gruppi in cui può essere suddivisa la popolazione, ma un sistema di equazioni con parametri diversi per ogni gruppo di individui. Le possibili variabili esplicative provengono dai diversi livelli della struttura gerarchica della popolazione. Le ipotesi generiche che stanno alla base dei modelli multilevel sono le seguenti:

- un dataset strutturato gerarchicamente;
- una variabile dipendente misurata al livello più basso;
- delle variabili esplicative misurate sui diversi livelli della gerarchia.

I coefficienti di regressione per i modelli multilevel variano a seconda del gruppo, sono indicati come *effetti random* e hanno a loro volta una distribuzione normale con una certa varianza che rappresenta la variabilità della popolazione. Al contrario, gli *effetti fissi* sono parametri costanti (Gelman e Hill, 2007).

Si può pensare al multilivello come una regressione che include delle variabili categoriali le quali descrivono l'appartenenza o meno di ogni unità ad un gruppo. In questa prospettiva si utilizza un fattore con j livelli (j è il numero di gruppi corrispondenti a j predittori nel modello di regressione). In questo modo vengono aggiunti al modello $j-1$ predittori lineari.

Il modello gerarchico è caratterizzato dal fatto che i j coefficienti sono anch'essi stimati da un modello, o più semplicemente, da una comune distribuzione; nel caso del modello descritto dall'equazione (3.1), per esempio, si imposta un secondo modello di regressione per gli α_j usando come predittori le variabili riferite al livello di gruppo. I coefficienti di questo modello vengono chiamati

iperparametri.

Questo secondo modello è stimato simultaneamente con la regressione al primo livello.

I modelli gerarchici sono un compromesso tra due estremi: *complete pooling*, nel quale gli indicatori di gruppo non sono inclusi nel modello, e *no pooling*, nel quale viene stimato un modello per ogni gruppo.

3.2 Inferenza classica e bayesiana

I due approcci principali all'analisi multilevel sono basati, rispettivamente, sul principio della ML (inferenza statistica classica) e sull'inferenza bayesiana. L'inferenza classica, di tipo frequentista, assume la probabilità come una caratteristica intrinseca e oggettiva degli eventi. I dati infatti sono considerati come realizzazioni di variabili aleatorie mentre i parametri hanno carattere deterministico. Il principio fondamentale è la funzione di verosimiglianza (*likelihood*) che racchiude tutte le informazioni che un generico campione \mathbf{X} può fornire. Se si interpreta la $f(x_1, x_2, \dots, x_n | \theta)$ ¹ come la verosimiglianza o la plausibilità che si realizzi la n -upla di dati x_1, x_2, \dots, x_n quando θ è il valore vero assunto dal parametro, sembra ragionevole adottare come stima di θ quel valore che rende massima la verosimiglianza per i dati osservati. In altre parole, la stima di massima verosimiglianza $\hat{\theta}$ è definita come il valore di θ che rende massima $f(x_1, x_2, \dots, x_n | \theta)$, quando i valori osservati sono proprio x_1, x_2, \dots, x_n . Si rende massima quindi la probabilità di osservare quel dato campionamento condizionato ai valori assunti dai parametri oggetti di stima.

Per i modelli multilevel i coefficienti di regressione vengono stimati iterativamente. Al primo passo vengono generati i valori iniziali dei diversi parametri

¹funzione di massa congiunta o densità congiunta di X_1, X_2, \dots, X_n a seconda che le variabili siano discrete o continue. Date le variabili aleatorie $X \in A \subset \mathbb{R}^n$, la probabilità che $X \in A$ è data da:

$$P(X \in A) = \int_A f_{X_1, \dots, X_n}(x_1, \dots, x_n) dx_1 \dots dx_n \quad (3.6)$$

come se la regressione riguardasse un singolo livello gerarchico. Nei passaggi successivi vengono migliorate le stime iniziali iterativamente fino al momento in cui la procedura di stima converge, cioè il programma controlla se la stima è effettivamente cambiata dal passo precedente altrimenti si ferma (convergenza raggiunta). Un problema comune a programmi che utilizzano una procedura iterativa di ML per analisi multilivello non banali è che il processo iterativo non riesce a fermarsi non arrivando mai a convergenza (Hox, 2010). Questo avviene infatti anche nel nostro modello quando inseriamo molte variabili random.

Nella statistica bayesiana, invece l'incertezza sui valori dei parametri del modello multilevel (coefficienti di regressione e componenti della varianza) viene espressa assegnando ai parametri stessi una distribuzione di probabilità. Tale distribuzione viene generalmente chiamata distribuzione a priori, dal momento che viene specificata indipendentemente dai dati campionari. L'analisi multilevel bayesiana fonda le sue radici teoriche nel teorema di Bayes (Bayes,1763), che fornisce una relazione per la probabilità condizionata di Θ dato X :

$$f(\Theta|X) = \frac{f(X|\Theta)f(\Theta)}{f(X)} \quad (3.7)$$

Dopo l'osservazione dei risultati campionari $X = \{x_1, x_2, \dots, x_n\}$, la distribuzione a priori $f(\Theta)$ viene combinata con la verosimiglianza $f(X|\Theta)$ dei dati al fine di produrre una distribuzione a posteriori $f(\Theta|X)$ che esprime l'incertezza sui parametri di popolazione dopo l'osservazione dei dati campionari. Generalmente, la varianza della distribuzione a posteriori è inferiore rispetto alla varianza della distribuzione a priori, in quanto l'osservazione dei risultati campionari riduce l'incertezza sui parametri da stimare.

Per la scelta della distribuzione a priori, esistono due principali alternative: a priori informative e a priori non informative. Le distribuzioni a priori informative sono caratterizzate da una piccola varianza, esprimono una forte credenza a priori sui parametri non noti della popolazione e influenzano notevolmente la distribuzione a posteriori.

Le distribuzioni a priori non informative hanno, per contro, una scarsissima influenza sulla distribuzione a posteriori ed hanno la sola funzione di produrre, combinate con la verosimiglianza, la distribuzione a posteriori. Un esempio di a priori non informativa è costituito dalla distribuzione uniforme definita su tutto l'asse reale, con la quale si afferma semplicemente che il parametro non noto della popolazione può variare tra meno infinito e più infinito, con tutti i valori ugualmente probabili. Un altro esempio di a priori non informativa è costituito dalla distribuzioni normale con varianza infinita.

Una volta determinata la distribuzione a posteriori, nei casi in cui questa abbia una forma matematicamente semplice, è possibile utilizzare le caratteristiche della distribuzione stessa per calcolare una stima puntuale e un intervallo di confidenza (bayesiano) per il parametro di popolazione sconosciuto. Nel caso di una distribuzione a posteriori normale sarà, pertanto, possibile utilizzare la media come stima puntuale del parametro incognito e basare la costruzione dell'intervallo di confidenza sulla deviazione standard della distribuzione a posteriori².

3.3 Analisi bayesiana dei dati di concentrazione di radon indoor

L'analisi dei dati mediante inferenza bayesiana consente di stimare i parametri statistici della distribuzione dei valori di concentrazione di radon nei vari comuni minimizzando l'effetto della ridotta dimensione del campione in alcuni di essi. L'approccio bayesiano consiste nel fare uso dei logaritmi delle medie geometriche campionarie (osservate) (α_j^{obs}) per predire i valori "veri" delle stesse per ogni gruppo (α_j^{pred}), assumendo che queste seguano una distribuzione di probabilità, i cui parametri possono essere ricavati direttamente dai dati. Se in un gruppo, ad esempio, non sono state effettuate misure di concentrazione di radon, il logaritmo della media geometrica del gruppo α_j^{obs} è sconosciuto. Ma

²i percentili 2.5% e 97.5% forniscono gli estremi dell'intervallo di confidenza bayesiano (o intervallo di credibilità) al 95%.

poiché è possibile conoscere, dai dati relativi ai comuni in cui è stato effettuato un numero elevato di misure, la distribuzione delle medie di gruppo, si può da questa risalire alla conoscenza dell'intervallo di maggiore probabilità per il valore della media geometrica del comune privo di misure.

Il teorema di Bayes, come visto nell'eq.(3.7), fornisce una stima per la distribuzione di probabilità della media geometrica (o meglio del logaritmo della media) vera, cioè quella che si otterrebbe se per tutti gli edifici del comune j –esimo si disponesse di una misura della concentrazione di radon:

$$p(\alpha_j|y_{ij}) = \frac{p(\alpha_j) \cdot p(y_{ij}|\alpha_j)}{p(y_{ij})} \quad (3.8)$$

Dall' eq.(3.8) la probabilità dell'evento α_j condizionato a y_{ij} (cioè il set di osservazioni) dipende non solo dalla relazione tra α_j e y_{ij} ma anche dalla probabilità assoluta (a priori) di α_j indipendentemente da y_{ij} .

- La $p(\alpha_j)$ è una probabilità a priori, nel senso che non tiene conto delle informazioni di y_{ij} ed è una distribuzione (normale) informativa $p(\alpha_j) \sim N(\mu, \sigma^2)$, sarebbe non informativa se scegliessimo $\sigma \rightarrow \infty$;
- $p(\alpha_j|y_{ij})$ è la probabilità condizionata di α_j data y_{ij} , è una probabilità a posteriori poiché è derivata da un valore specifico di y_{ij} ed è la probabilità che il valore atteso della media del comune j sia α_j ;
- $p(y_{ij}|\alpha_j)$ è la probabilità condizionata di y_{ij} data α_j , anche chiamata verosimiglianza. Questa rappresenta la probabilità di ottenere le misure y_{ij} se il valore vero della media del comune fosse α_j ;
- $p(y_{ij})$ è la distribuzione di marginale di y_{ij} ed è un fattore di normalizzazione.

Se $p(\alpha_j)$ e $p(y_{ij}|\alpha_j)$ sono entrambi distribuzioni gaussiane, si dimostra che anche la distribuzione a posteriori sarà gaussiana (Carlin e Louis, 2000). Sviluppando il calcolo dall'eq.(3.8) per la stima (*puntuale*) a posteriori di α_j si ottiene:

$$\hat{\alpha}_j = \frac{\frac{\mu_\alpha}{\sigma_\alpha^2} + \frac{n_j}{\sigma_y^2} \bar{y}_j}{\frac{1}{\sigma_\alpha^2} + \frac{n_j}{\sigma_y^2}} = \frac{\sigma_y^2 \mu_\alpha + n_j \sigma_\alpha^2 \bar{y}_j}{\sigma_y^2 + n_j \sigma_\alpha^2} \quad (3.9)$$

la cui varianza è data da:

$$\text{var}(\hat{\alpha}_j) = \left(\frac{1}{\sigma_\alpha^2} + \frac{n_j}{\sigma_y^2} \right)^{-1} = \frac{\sigma_\alpha^2 \sigma_y^2}{\sigma_y^2 + n_j \sigma_\alpha^2} \quad (3.10)$$

cioè il più probabile valore vero di α_j per un dato comune j è la media pesata tra il valore osservato della media nel gruppo j (la stima unpooled \bar{y}_j) e la media della distribuzione dei valori di α_j di ogni comune (pooling completo, μ_α). I pesi sono dati dal numero di osservazioni n_j nel comune e dalle varianze: σ_y^2 è la varianza entro ciascun comune del logaritmo delle misure del radon e σ_α^2 è la varianza tra le quantità delle medie dei logaritmi dei livelli di radon indoor di differenti comuni.

Un modo più compatto di scrivere le relazioni (3.9) e (3.10) è:

$$\hat{\alpha}_j = \omega_j \mu_\alpha + (1 - \omega_j) \bar{y}_j \quad (3.11)$$

$$\text{var}(\hat{\alpha}_j) = (1 - \omega_j) \sigma_y^2 / n_j \quad (3.12)$$

con $\omega_j = \frac{\sigma_y^2}{\sigma_y^2 + n_j \sigma_\alpha^2}$ definito come il *fattore di pooling*.

La media pesata rispecchia la quantità di informazioni disponibili per ciascun comune, da una parte, e la media di tutti i comuni dall'altra:

- le medie campionarie di comuni con poche misure portano meno informazioni, e dalla stima pesata si ottiene un valore di $\hat{\alpha} \rightarrow \mu_\alpha$ per $n_j \rightarrow 0$, che con il fattore di pooling equivale a $\omega_j \rightarrow 1$. Da notare che la stessa condizione su ω_j si ottiene con ω_α^2 piccolo comparato a ω_y^2 , cioè con una distribuzione altamente informativa a priori.

- Le medie di comuni con una grande quantità di dati daranno più informazioni e le stime multilevel corrispondenti saranno più vicine alle medie del singolo comune α_j . Nel limite con $n_j \rightarrow \infty$ si avrà $\hat{\alpha} \rightarrow \alpha_j$. $\omega_j \rightarrow 0$ anche quando ω_α^2 è grande rispetto a ω_y^2 , cioè in presenza di un'incerta informazione a priori.

- Nel caso intermedio, la stima del multilevel è tra i due estremi.

Una caratteristica chiave delle analisi multilevel è *la correlazione intraclasse*:

$$\rho = \sigma_{\alpha}^2 / (\sigma_{\alpha}^2 + \sigma_y^2)$$

ρ è il rapporto fra la varianza tra i gruppi e la varianza totale della distribuzione di misure; questa tiene conto del fatto che, trattandosi di dati gerarchicamente organizzati, in genere le osservazioni individuali non sono del tutto indipendenti. La correlazione intraclasse misura il grado di omogeneità dei risultati y_{ij} entro un gruppo j , ρ oscilla da 1 a 0, quanto più si avvicina all'unità, tanto più alta è la porzione di varianza tra i comuni sulla varianza totale.

Nel nostro modello le misure fatte in abitazioni della stesso comune tendono ad essere spazialmente correlate. I test statistici tradizionali sono fortemente basati sull'assunto di indipendenza tra le osservazioni. Se questa ipotesi risulta violata, le stime degli errori standard prodotte dai test statistici convenzionali sono troppo piccole e, di conseguenza, i risultati che si ottengono appaiono "impropriamente" significativi.

3.4 Multilevel computation

La statistica bayesiana quindi differisce da quella classica dal fatto che ora tutti i parametri non noti sono considerati come variabili random indipendenti. Per questa ragione, la distribuzione a priori deve essere definita all'inizio, cioè questa esprime le informazioni disponibili prima che i dati siano inseriti nell'analisi statistica. In accordo con il teorema di Bayes (eq. 3.8), la distribuzione a posteriori $p(\alpha_j | y_{ij})$ può essere scritta come:

$$p(\alpha_j | y_{ij}) \propto p(\alpha_j) \cdot p(y_{ij} | \alpha_j) \tag{3.13}$$

la $p(\alpha_j|y_{ij})$ ingloba sia le informazioni a priori sia quelle delle misure osservate, rispettivamente con $p(\alpha_j)$ e con la verosimiglianza $p(y_{ij}|\alpha_j)$.

Una delle maggiori difficoltà dell'approccio bayesiano è l'aspetto computazionale, infatti solo in rarissimi casi si riesce ad ottenere analiticamente una distribuzione a posteriori esplicita e che permetta di calcolare semplicemente la stima puntuale e gli intervalli di confidenza per i parametri della popolazione. Nel caso per esempio di una distribuzione a posteriori normale, la media sarà la stima puntuale e la deviazione standard della curva gaussiana sarà l'intervallo di confidenza .

Solo nell'ultimo ventennio lo sviluppo di particolari metodi di integrazione numerica attraverso simulazione, come i metodi Monte Carlo basati sulle catene di Markov, hanno permesso l'applicazione dell'inferenza bayesiana a problemi più complessi.

Le *Markov chains Monte Carlo (MCMC)* sono tecniche di simulazione che generano campionamenti random dalla distribuzione a posteriori. Attraverso un elevato numero di campioni casuali è, infatti, possibile approssimare la reale forma delle distribuzioni. Le distribuzioni a posteriori simulate vengono successivamente utilizzate per calcolare le stime puntuali e gli intervalli di confidenza. Solitamente, viene impiegata la distribuzione marginale di ciascun parametro. La moda della distribuzione a posteriori marginale viene frequentemente utilizzata come stima puntuale del parametro incognito. Questa soluzione costituisce l'equivalente bayesiano dello stimatore di massima verosimiglianza. In alcune situazioni, tuttavia, si preferisce utilizzare la media della distribuzione a posteriori poiché più semplice da calcolare, o la mediana, nel caso di distribuzioni a posteriori asimmetriche.

Una catena di Markov è un processo stocastico³ che riguarda un set di varia-

³Si definisce processo stocastico una famiglia di variabili aleatorie $\{\theta(t), t \in T\}$ dipendenti dal tempo, definite su un unico spazio campione Ω finito e che assumono valori in un insieme definito *spazio degli stati del processo*. Un processo stocastico è quindi un insieme di funzioni che evolvono nel tempo (funzioni campione o realizzazioni), ognuna delle quali è associata ad un determinato elemento dello spazio campione, così che il risultato di un esperimento casuale corrisponde di fatto all'estrazione di una di queste funzioni.

bili random $\{\theta(t)\}$, dove t è un parametro discreto o continuo definito “tempo”, così che

$$f(\Theta^{(t+1)}|\Theta^{(t)}, \dots, \Theta^{(1)}) = f(\Theta^{(t+1)}|\Theta^{(t)}) \quad (3.14)$$

dunque la distribuzione di Θ al tempo $t + 1$ dati tutti i precedenti valori Θ (per $t, t - 1, \dots, 1$) dipende solo dal valore di $\Theta^{(t)}$. Tuttavia $f(\Theta^{(t+1)}|\Theta^{(t)})$ è indipendente al tempo t .

Sotto alcune condizioni, la sequenza di campionamenti che costituiscono la catena di Markov convergerà ad una distribuzione target o *stazionaria*. Al contrario dei metodi di simulazione diretta che forniscono un campionamento direttamente dalla distribuzioni a posteriori, l’algoritmo di MCMC genera catene di Markov che potrebbero convergere alla nostra distribuzione a posteriori.

Il software *JAGS* (Just Another Gibbs Sampler), utilizzato in questa tesi, è in grado di determinare la matrice di transizione per la catena di Markov per generare campioni dalla distribuzione a posteriori di quantità non note di un modello⁴. L’algoritmo iterativo parte da valori iniziali prodotti da generatori di numeri random da distribuzioni, per esempio, normali $N(0,1)$ o uniformi tra $[0,1]$. Il software permette di definire il numero di iterazioni complessive dell’algoritmo e ogni iterazione genera un set completo di parametri dalla distribuzione a posteriori, quindi $\Theta^{(5)}$ e $\Theta^{(t)}$ rispettivamente indicano i valori del vettore random Θ generato alla quinta e alla t -esima iterazione. Il risultato finale è il riassunto numerico e grafico delle distribuzioni marginali (univariate) a posteriori di ciascun parametro non noto.

Jags utilizza il campionamento di Gibbs per generare le catene di Markov. L’algoritmo può essere schematizzato nel modo seguente:

- selezione dei valori iniziali $\Theta^{(0)}$;
- generazione di T valori fino a che non sia raggiunta la distribuzione di equilibrio;

⁴Plummer, Martyn (2003). *JAGS: A Program for Analysis of Bayesian Graphical Models Using Gibbs Sampling*, Proceedings of the 3rd International Workshop on Distributed Statistical Computing (DSC 2003), March 20–22, Vienna, Austria. ISSN 1609-395X.

- controllo della convergenza dell' algoritmo usando delle diagnostiche ad hoc e se i test falliscono si generano più osservazioni;
- eliminazione delle prime B osservazioni;
- si considera $\{\Theta^{(B+1)}, \Theta^{(B+2)}, \dots, \Theta^T\}$ come il campione per l'analisi posteriore;
- raffigurazione della distribuzione a posteriori (di solito l'obbiettivo è la distribuzione univariata marginale);
- infine, si ottiene il riepilogo delle caratteristiche della distribuzione a posteriori (media, mediana, deviazione standard, quantili e correlazioni).

La particolarità dei metodi MCMC è che due successivi campionamenti sono dipendenti. A seconda delle distribuzioni considerate e della quantità di dati disponibili, la correlazione può essere più o meno forte. Sarebbe preferibile che i campionamenti fossero indipendenti e un modo per ovviare a questo problema è quello di definire un intervallo di diradamento, *thinning*, grazie al quale si elimina un numero $n.thin$ di stime successive in modo da ridurre l'autocorrelazione.

Per determinare la convergenza dell'algoritmo, in Jags, viene utilizzata la diagnostica Gelman Rubin (Gelman et. al., 2004). Gelman e Rubin (1992) proposero di controllare la convergenza dell'algoritmo confrontando le conclusioni sui parametri di interesse ottenute da un numero $n.chain$ di catene parallele con differenti valori iniziali. Considerando quindi m catene di Markov, ognuna con i propri parametri iniziali, si comparano le varianze tra ed entro le varie catene. La varianza entro la catena è

$$W = \frac{1}{m} \sum_{i=1}^m s_j^2$$

dove

$$s_j^2 = \frac{1}{n-1} \sum_{i=1}^n (\theta_{ij} - \bar{\theta}_j)^2$$

s_j^2 è la varianza della j -esima catena e W così è la media delle varianze di ogni catena.

La varianza tra le catene è invece

$$B = \frac{n}{m-1} \sum_{j=1}^m m(\bar{\theta}_j - \hat{\theta})^2$$

dove

$$\hat{\theta} = \frac{1}{m} \sum_{j=1}^m \bar{\theta}_j$$

Può così essere stimata la varianza della distribuzione stazionaria come la media pesata di W e B :

$$V(\hat{\theta}) = \left(1 - \frac{1}{n}\right)W + \frac{1}{n}B$$

Per ogni parametro viene calcolata la radice del rapporto tra la varianza a posteriori stimata ($V(\hat{\theta})$) e la varianza entro la catena (W):

$$\hat{R} = \sqrt{\frac{V(\hat{\theta})}{W}} \tag{3.15}$$

\hat{R} è chiamato *fattore di riduzione di scala potenziale*. Quando $\hat{R} < 1.1$ o 1.2 , le catene sono mescolate perfettamente e la convergenza è massima, altrimenti si aumenta il numero di iterazioni permesse e si riesegue il test fino ad ottenere valori di \hat{R} prossimi a 1.

Modello multilevel a singola intercetta.

Il più semplice modello per analizzare i dati radon indoor è un modello di regressione multilevel a singola intercetta:

$$y_{ij} = \alpha_j + \epsilon_{ij} \tag{3.16}$$

$$\alpha_j \sim N(\mu_\alpha, \sigma_\alpha^2) \tag{3.17}$$

$$\epsilon_{ij} \sim N(0, \sigma_y^2) \tag{3.18}$$

dove y_{ij} è la concentrazione di radon (o meglio il suo logaritmo) misurato nell' i -esimo edificio del j -esimo comune; ϵ nell' eq. (3.18) è il termine di errore casuale che si suppone essere una variabile gaussiana con media nulla; nell' eq.(3.17) assegnamo una distribuzione normale di probabilità ai valori che può assumere la media comunale α con media μ_α e varianza σ_α^2 (deveno essere stimati sulla base dei dati sperimentali).

Il predittore α_j è chiamato usualmente “intercetta” e rappresenta la media delle risposte (log-concentrazione radon) su ogni singolo gruppo (nel nostro caso, il comune); mentre μ_α , σ_α^2 , e σ_y^2 , noti come iperparametri del modello, riguardano il livello gerarchicamente più alto del modello stesso e devo essere calcolati dalle misure.

L'equazione (3.16) è un caso speciale di regressione con la sola intercetta e senza coefficienti di regressione.

Il modello può essere anche scritto con un'altra formulazione:

$$y_{ij} = \mu + \alpha_j + \epsilon_{ij} \quad \alpha_j \sim N(0, \sigma_\alpha^2) \quad \epsilon_{ij} \sim N(0, \sigma_y^2) \quad (3.19)$$

L'effetto dell'ipotesi distributiva nell' eq.(3.17) è quello di avvicinare le stime delle intercette α_j alla media globale μ in maniera tale che, in ciascun j -esimo comune si ottenga un compromesso (pooling) tra μ_α e la media campionaria dei n_j dati sperimentali disponibili:

$$\bar{y}_j = \frac{1}{n_j} \sum_{i=1}^{n_j} y_{ij} \quad (3.20)$$

nel limite $\sigma_\alpha^2 \rightarrow \infty$, il vincolo dell'eq.(3.17) non ha effetto e dunque non si realizza nessun *pooling*; nel limite $\sigma_\alpha^2 \rightarrow 0$, invece, si ha una stima di *pooling completo*, cioè $\alpha_i = \mu_\alpha$ per ciascun comune.

Campionamento di Gibbs per un modello multilevel senza predittori

Il campionamento di Gibbs generalmente partiziona il set di parametri non noti

e stima un parametro alla volta, o un gruppo alla volta, condizionato a tutti gli altri. Nel caso semplice di un modello senza predittori, l'algoritmo di Gibbs utilizza, come valori iniziali per α , J numeri random generati da una distribuzione normale $N(0,1)$ e i passi successivi sono: (Gelman e Hill, 2007):

- aggiornare il valore di α : per $j = 1, \dots, J$, calcola $\hat{\alpha}$ e V_j rispettivamente dall'eq. (3.9) e (3.10) e campiona α_j da una distribuzione normale con media $\hat{\alpha}$ e varianza V_j .

- aggiornare μ_α : calcola $\hat{\mu}_\alpha$ come media dei valori dell'intercetta α_j sui comuni

$$\hat{\mu}_\alpha = \frac{1}{J} \sum_{j=1}^J \alpha_j$$

e campiona μ_α da una distribuzione normale con media $\hat{\mu}_\alpha$ e varianza σ_α^2/J .

- aggiornare σ_y : calcola $\hat{\sigma}_y^2$ con

$$\hat{\sigma}_y^2 = \frac{1}{n} \sum_{i=1}^n (y_i - \alpha_{j[i]})^2$$

e campiona $\sigma_y^2 = \hat{\sigma}_y^2/X_{n-1}^2$, dove X_{n-1}^2 è un campionamento random dalla distribuzione χ^2 con $J - 1$ gradi di libertà.

-aggiornare σ_α : calcola $\hat{\sigma}_\alpha^2$

$$\hat{\sigma}_\alpha^2 = \frac{1}{J} \sum_{j=1}^J (\alpha_j - \mu_\alpha)^2$$

e così campiona $\sigma_\alpha^2 = \hat{\sigma}_\alpha^2/X_{J-1}^2$ dove X_{J-1}^2 è il campionamento random dalla distribuzione χ^2 con $J - 1$ gradi di libertà.

Iterando la procedura sopra per T volte si ottiene una catena di campionamenti simulati, quindi una sequenza per ciascun parametro

$$\alpha_1^{(1)}, \dots, \alpha_J^{(1)}; \sigma_y^{2(1)}; \mu_\alpha^{(1)}; \dots; \alpha_1^{(T)}, \dots, \alpha_J^{(T)}; \sigma_y^{2(T)}; \mu_\alpha^{(T)}$$

Se la catena è abbastanza lunga, questa fornisce l'intervallo di incertezza nella

stima del parametro.

3.5 Modello multilevel con predittori e dati non annidati

Nel semplice modello visto nel paragrafo precedente possono essere incluse una o più variabili esplicative, come il piano (x_i) nel quale viene fatta la misura di radon, la tipologia di edificio, i materiali con cui è stato edificato e l'anno di costruzione. I coefficienti di regressione β possono essere fissi o variare su ciascun gruppo (eq. 3.21).

Usando la notazione $j[i]$ per indicare la variabile riferita all'abitazione i del comune j , abbiamo il modello per un singolo coefficiente x (Gelman e Hill, 2007):

$$y_i = \alpha_{j[i]} + \beta_{j[i]}x_i + \epsilon_{ij} \quad (3.21)$$

con

$$\begin{pmatrix} \alpha_j \\ \beta_j \end{pmatrix} \sim N \left(\begin{pmatrix} \mu_\alpha \\ \mu_\beta \end{pmatrix}, \begin{pmatrix} \sigma_\alpha^2 & \rho\sigma_\alpha^2\sigma_\beta^2 \\ \rho\sigma_\alpha^2\sigma_\beta^2 & \sigma_\beta^2 \end{pmatrix} \right) \text{ per } j = 1, \dots, J \quad (3.22)$$

dove al vettore dei predittori è assegnata una distribuzione normale multivariata e ρ è il coefficiente di correlazione intraclasse tra gruppi.

Generalizzando a K coefficienti di regressione, la matrice di covarianza Σ nell'eq.(3.22) ha elementi

$$\begin{cases} \sigma_k^2 \text{ per } \Sigma_{kk} \\ \rho_{kl}\sigma_k\sigma_l \text{ per } \Sigma_{kl} \end{cases} \quad (3.23)$$

A ciascun elemento deve essere assegnata una distribuzione a priori non informativa, ma quando abbiamo più di due effetti random correlati, la formulazione diventa più complessa ed è necessario usare una distribuzione multivariata la quale garantisca che la matrice risultante sia ben definita. La difficoltà infatti viene dal fatto che non sia più possibile definire la distribuzione per un parametro alla volta. La matrice di covarianza non è un insieme di numeri random, ma

deve essere simmetrica e semi-definita positiva che sono proprietà della matrice più che dei singoli elementi che la compongono. Di conseguenza è necessario definire direttamente una matrice non informativa a priori. Una distribuzione di probabilità che viene spesso utilizzata come a priori per matrici di covarianza è la distribuzione Wishart inversa (Gelman e Hill, 2007), il cui approfondimento tuttavia è oltre gli scopi di questa tesi.

3.6 Distribuzione log-normale delle misure

Le concentrazioni di radon indoor, come abbiamo visto, sono influenzate da una moltitudine di fattori quali, per esempio, le caratteristiche proprie del tipo di suolo sul quale si trova l'edificio (permeabilità, porosità, contenuto di uranio e radio ecc...) e le proprietà relative alla costruzione stessa (il materiale usato per edificate, il rateo di ventilazione, la differenza di pressione tra abitazione ed esterno, ecc...). Tutti questi fattori possono essere considerati indipendenti l'uno dall'altro e moltiplicativi, così da giustificare l'usuale distribuzione log-normale che si trova per le misure di concentrazioni radon nelle abitazioni.

La quantità di radon misurata all'interno degli edifici, sottraendo la concentrazione di radon esterna, può essere espressa, pertanto, come combinazione lineare delle differenti distribuzioni lognormali di ciascun fattore esplicativo (Nero et al. 1986; Miles, 1994):

$$\ln(Rn_{in} - Rn_{out}) = \ln(A) + \ln(B) + \ln(C) + \dots \quad (3.24)$$

Rn_{in} è la concentrazione di radon all'interno dell'edificio e Rn_{out} è la concentrazione all'esterno; le A , B ecc... sono le funzioni di fattori come il contenuto di radio del terreno, la permeabilità, il numero e le dimensioni delle vie di ingresso del radon, la pressione nell'edificio, la ventilazione e così via. L'eq.(3.24) può risciversi :

$$Rn_{in} = Rn_{out} + A \times B \times C \times \dots \quad (3.25)$$

Per Rn_{out} l' UNSCEAR nel 1993 ha stimato un valore di 10 Bq/m^3 per le aree continentali; in Abruzzo è stato misurato un valore annuale medio di 12 Bq/m^3 nell'area aquilana (Di Carlo et.al, 2007), mentre i valori ottenuti sull'area costiera sono più bassi, $Rn_{out}^{cost} = 4 \text{ Bq/m}^3$.

I dati ottenuti nelle varie campagne di misure nella regione Abruzzo non hanno una distribuzione puramente lognormale ed è fatto tutt'altro che insolito (Miles, 1994 con i dati dell'indagine inglese). Le possibili ragioni di questo scostamento sono da imputarsi all'incertezza della misura stessa, alla sottrazione del fondo che porta ad un eccesso di valori bassi e a problemi legati alla rappresentatività del campionamento (Miles, 1994; Price et al., 1996).

Dal momento che valori estremamente bassi causano problemi alla coda inferiore della distribuzione, si è scelto di correggere empiricamente il valore di concentrazione radon misurato, shiftando i valori più bassi e lasciando praticamente inalterati i valori più alti (Price et al., 1996):

$$Rn_{in}^{new} = \frac{Rn_{in} - Rn_{out}}{2} + \sqrt{\left(\frac{Rn_{in} - Rn_{out}}{4}\right)^2 + D^2} \quad (3.26)$$

dove con D indichiamo il fattore correttivo.

La correzione per ottenere il miglior fit di una distribuzione lognormale è stata scelta utilizzando i test di normalità di D'Agostino e Shapiro-Wilk. Si è calcolato infatti il p-value al variare del D ($0.01 \text{ Bqm}^{-3} \leq D \leq 5 \text{ Bqm}^{-3}$) e dei valori plausibili della concentrazione di radon outdoor R_{out} ($8 \text{ Bqm}^{-3} \leq R_{out} \leq 12 \text{ Bqm}^{-3}$), scegliendo i fattori correttivi che lo massimizzano.

I test di normalità sono eseguiti su **R** con funzioni del pacchetto *fBasics*.

Utilizzando i dati disponibili ottenuti dalle varie campagne di indagine radon, il p-value massimo del test di D'Agostino si è trovato con i valori di $Rn_{out} = 12 \text{ Bqm}^{-3}$ e la correzione $D = 2 \text{ Bq m}^{-3}$ (tab. 3.2). Il codice che abbiamo scritto in R è in appendice.

Capitolo 3. Analisi multilevel per le concentrazioni di radon indoor nella provincia aquilana

caratteristiche	n	AM	SD	GM	GSD	MAX	MIN
post1960	430	3.48	1.04	3.30	1.42	6.60	0.21
prima del 1919	142	3.52	1.05	3.31	1.50	6.34	0.40
tra 1919 e 1960	185	3.88	1.09	3.72	1.35	6.63	0.99
piano contatto terra	298	3.92	1.12	3.72	1.43	6.63	0.40
piano terra NCT	130	3.60	1.12	3.41	1.43	6.16	0.52
1° piano	203	3.21	0.84	3.07	1.42	5.23	0.21
piano>1	82	2.95	0.65	2.87	1.28	4.91	1.17
Sotterraneo	44	4.16	0.94	4.05	1.28	6.04	1.96
lat-cem	505	3.58	1.05	3.40	1.42	6.61	0.21
pietra	225	3.66	1.08	3.47	1.44	6.63	0.40
altro	27	2.96	0.85	2.81	1.42	4.26	1.17
abitazione	496	3.42	1.02	3.24	1.44	6.61	0.21
luogo di lavoro	89	3.77	1.25	3.53	1.47	6.63	0.99
pubblico	45	3.75	1.00	3.60	1.36	6.04	1.43
scuola	127	4.04	0.97	3.91	1.29	6.60	1.92

Tabella 3.1: statistica per la variabile standardizzata Rn_{in}^{new} in $Bq\ m^{-3}$.

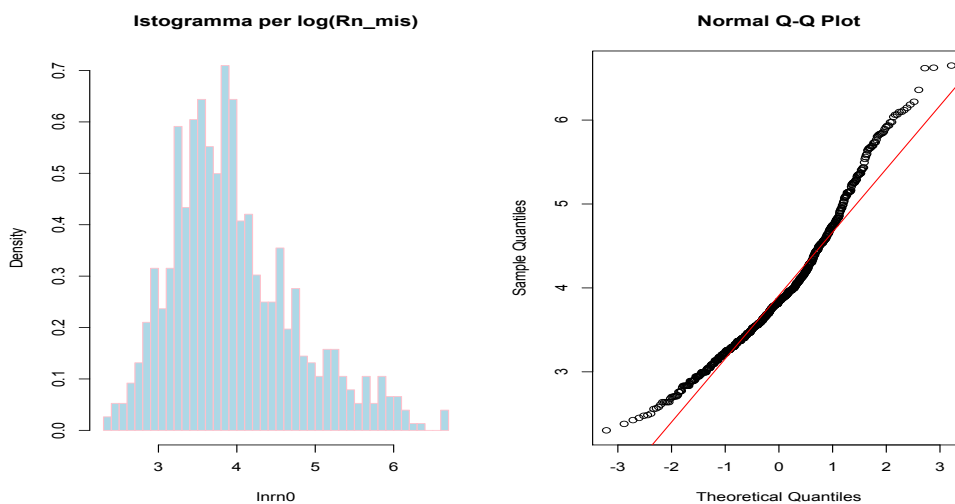


Figura 3.1: *a sinistra*: Istogramma del logaritmo di $Rn_{misurata}$, senza correzione per il fondo; *a destra*: il grafico quantile-quantile per la distribuzione normale teorica $N(0,1)$ e $\log(Rn_{misurata})$.

STATISTIC:	P VALUE:
Chi2 Omnibus: 2.6998	Omnibus Test: 0.2593
Skewness: 0.9185	Skewness Test: 0.3584
Kurtosis: 1.3624	Kurtosis Test: 0.1731

Tabella 3.2: Test di normalità di D'Agostino per $lnrn0 = ln(Rn_{in}^{new})$

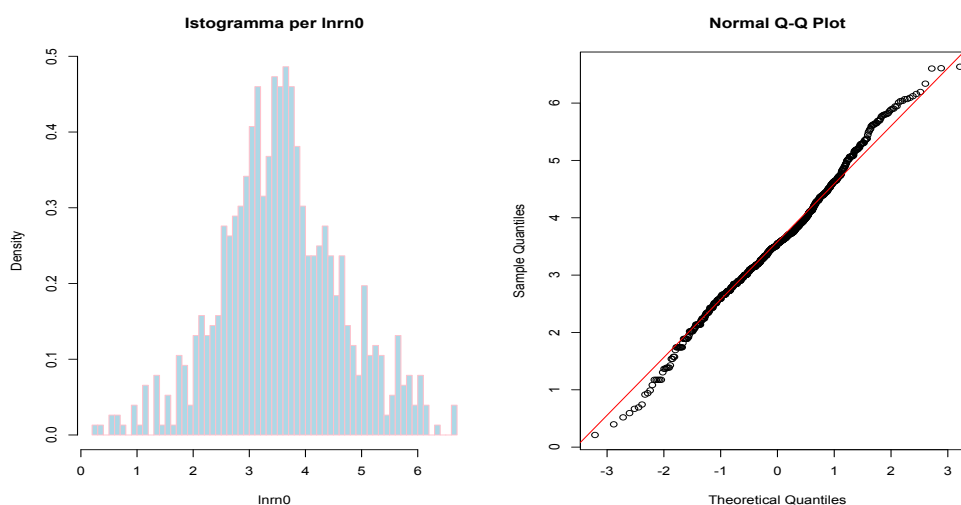


Figura 3.2: *a sinistra*: Istogramma del logaritmo di Rn_{in}^{new} dell'(eq.3.26), per $Rn_{out} = 12Bqm^{-3}$ e $D = 2 Bq m^{-3}$. *a destra*: il grafico quantile-quantile per la distribuzione normale teorica $N(0,1)$ e $ln rn0$.

*Test di Shapiro-Wilk*⁵

Il test di Shapiro-Wilk è considerato uno dei test più potenti per la verifica della normalità, soprattutto per piccoli campioni. La verifica della normalità avviene confrontando due stimatori alternativi della varianza σ^2 : uno stimatore non parametrico basato sulla combinazione lineare ottimale della statistica d'ordine di una variabile aleatoria normale al numeratore, e uno stimatore parametrico, ossia la varianza campionaria, al denominatore. I pesi per la combinazione lineare (a_i) sono disponibili su apposite tavole. La statistica W può essere interpretata co-

⁵L. SOLIANI Statistica univariata e bivariata parametrica e non-parametrica per le discipline ambientali e biologiche.

me il quadrato del coefficiente di correlazione in un diagramma quantile-quantile (QQ plot).

$$W = \frac{(\sum_i^n a_i x_{(i)})^2}{\sum_i^n (x_i - \bar{x})^2} \quad (3.27)$$

La funzione per effettuare il test di normalità in questione in ambiente R è `shapiro.test()`: esso restituisce come risultato il valore della statistica W e il relativo p-value.

*Test di D'Agostino*⁶

Il test D'Agostino, che richiede il calcolo del parametro statistico indicato in seguito con D_A , è un test molto potente per la valutazione dell'adattamento di dati alla distribuzione normale. Viene in genere applicato come alternativa al test di Shapiro e Wilk, soprattutto quando il numero dei dati è maggiore di 50. La procedura prevede che gli n dati vengano preliminarmente ordinati in maniera crescente o decrescente. Quindi è possibile calcolare D_A , definito attraverso la seguente espressione:

$$D_A = \frac{T}{\sqrt{n^3 SS}}$$

Il termine SS è la somma dei quadrati degli scostamenti dei dati X_i dal valore medio

$$SS = \sum_{i=1}^n (X_i - \bar{X})^2$$

il termine T è invece calcolato come:

$$T = \sum_{i=1}^n \left(i - \frac{n+1}{2} \right) X_i$$

in cui i è l'indice progressivo associato ai dati ordinati.

Il parametro D_A così ottenuto viene poi confrontato con la tabella dei valori critici (superiori ed inferiori) relativi al numero di dati testato. Da tale confronto si ricava il livello di significatività dell'adattamento dei dati alla distribuzione normale.

⁶Zar J. H., "Biostatistical Analysis", Prentice-Hall International, 1984.

La funzione per effettuare il test di normalità in R è `dagoTest()`: esso restituisce come risultato il valore della statistica D_A e il relativo p-value.

3.7 La standardizzazione delle misure di radon indoor

Lo scopo della costruzione e dell'utilizzo di un modello statistico è di ottenere il miglior fit del set di dati. Nell'analisi multilevel i coefficienti di regressione, raccolti nel vettore β , indicano di quanto il livello di radon in un certo edificio, con determinate caratteristiche, differisca da una situazione standard o di base. Si scelgono infatti delle specifiche che saranno quelle standard, in modo da riportare tutte le misure a questa condizione. Quindi per esempio se il coefficiente β è positivo per un luogo pubblico significa che la concentrazione di radon sarà maggiore rispetto alla tipologia standard che può essere quella dell'abitazione. Utilizzando i regressori stimati dall'analisi multilevel possiamo definire una nuova variabile standardizzata $Rnstd_i$:

$$Rnstd_i = exp(y_i - \beta X_i) + Rn_{out} \quad (3.28)$$

I coefficienti β sono le mediane delle distribuzioni marginali a posteriori dei relativi parametri.

Le variabili esplicative vengono aggiunte al modello base (eq. 3.21) con una tecnica *stepwise*⁷, che permette di includerle solo se superano un test di significatività (ad esempio, il coefficiente di regressione non è più significativo ad un livello $\alpha = 0.10$ o la statistica di best fitting peggiora rispetto al modello precedente).

Per valutare se il modello implementato descriva meglio il set di dati, nella sta-

⁷*Stepwise selection*: l'analisi comincia con l'equazione di regressione vuota e le variabili indipendenti sono aggiunte una alla volta se soddisfano un criterio statistico (ad esempio, coefficiente di regressione significativo ad un livello $\alpha = 0.05$ o una statistica di best fitting che confronta i vari modelli), ma possono essere eliminate in uno qualunque degli stadi di selezione successivi se non contribuiscono più in modo significativo alla spiegazione della variabilità della variabile dipendente.

tistica frequentista, si confrontano i valori dell' *AIC*, *Akaike's Information Criterion*, o i coefficienti di determinazione aggiustati R_{adj}^2 relativi ai vari modelli (Apte et al, 1999; Gelman e Hill, 2007). Per l'inferenza bayesiana, il software Jags fornisce in output la statistica *Deviance Information Criterion (DIC)* (Spiegelhalter et al., 2002) che è la naturale generalizzazione dell'AIC per un modello multilevel.

R^2 e R_{adj}^2 sono parametri comunemente usati per stabilire l'adeguatezza di un modello multilevel. R^2 è il rapporto tra la variabilità spiegata e la variabilità totale, il coefficiente restituisce la proporzione della variabilità della risposta spiegata dalla variabile esplicativa. Per il modello di regressione lineare $y = \beta X + \epsilon$ possiamo scrivere:

$$R^2 = 1 - \frac{RSS}{Dev(y)}$$

con

$$RSS = \sum_{i=1}^n (residui)^2$$

$$Dev(y) = \sum_{i=1}^n (y_i - \bar{y})^2$$

R^2 è un coefficiente che varia tra 0 ed 1: quando è 0 il modello utilizzato non spiega per nulla i dati; quando è 1 il modello spiega perfettamente i dati.

Nel modello multilevel i predittori spiegano i dati a livelli differenti (m). Seguendo l'approccio di Gelman e Hill (2007), un modello lineare $y_i^{(m)} = X_i^{(m)}\beta + \epsilon_i^{(m)}$ con $i = 1, \dots, n^{(m)}$ si può scrivere per ogni livello m :

$$\theta_i^{(m)} = \hat{\theta}_i^{(m)} + \epsilon_i^{(m)}, \text{ per } i = 1, \dots, n^{(m)}$$

dove i $\hat{\theta}_i^{(m)}$ sono la combinazione lineare dei coefficienti e dei predittori per ciascun livello, e gli errori $\epsilon_i^{(m)}$ sono dati da una distribuzione con media zero e

deviazione standard σ^m . Si definisce R_m^2 :

$$R_m^2 = 1 - \frac{E(V_{i=1}^n \epsilon_i^m)}{E(V_{i=1}^n \theta_i^m)}$$

dove l'operatore E indica si stanno calcolando i valori di aspettazione mediando su l'incertezza dei coefficienti di regressione $\theta_i^{(m)}$ del modello, quindi mediando su una simulazione a posteriori. Al livello più basso (quello dei dati), i $\theta_i^{(m)}$ corrispondono alle singole misure y_i osservate.

$$R_m^2 = 1 - \frac{E(V_{i=1}^n \epsilon_i)}{E(V_{i=1}^n y_i)}$$

Ai livelli superiori del modello, i $\theta_i^{(m)}$ rappresentano un gruppo di coefficienti di regressione (l'intercetta α_j e la pendenza β_j), quindi ci saranno più di un R^2 , uno per il livello delle misure e uno per ogni gruppo dei parametri. Nell'approccio bayesiano con le simulazione, i valori attesi al numeratore e denominatore sono calcolati mediando su campioni simulati a posteriori.

L'operatore V è la varianza di un campione finito:

$$V_{i=1}^n x_i = (n-1)^{-1} \sum_{i=1}^n (x_i - \bar{x})^2$$

R_{adj}^2 è una variante di R^2 semplice ed è R^2 corretto a seconda del numero di variabili indipendenti. Il motivo della necessità dell'aggiustamento di R^2 è che, per ragioni legate alla particolare procedura di calcolo, al ridursi della differenza fra numero di soggetti n e numero di predittori p , il valore di R^2 tende a "gonfiarsi" indipendentemente dal fatto che i predittori del modello rappresentino effettivamente le maggiori fonti di variazione della variabile dipendente, finché, quando il numero di predittori è uguale al numero di soggetti meno uno, il valore di R^2 è in ogni caso uguale a 1. Quindi R_{adj}^2 è:

$$R_{adj}^2 = 1 - \frac{(n-1)}{(n-p-1)}(1 - R^2)$$

L'analogo bayesiano di R^2 è

$$R_B^2 = 1 - \frac{\sigma^2}{\sigma_y^2}$$

con σ^2 è la varianza del parametro stimato. Questa quantità può essere interpretata come la riduzione proporzionale dell'incertezza sulla variabile risposta Y ottenuta incorporando le variabili X_j . Se il modello migliora con l'aggiunta di predittori, ci si aspetta che tutti gli R^2 aumentino a tutti i livelli del modello⁸.

Il criterio AIC viene preferito per confrontare due modelli che agiscono su uno stesso set di dati (Gelman e Hill, 2007) ed è basato sui logaritmi della verosimiglianza. Il coefficiente AIC è la somma pesata delle log-verosimiglianze del modello e del numero di coefficienti fittati. Si considerino n osservazioni e p parametri fittati, se denotiamo con $f(X|\theta_m, m)$ la verosimiglianza, cioè la probabilità di ottenere un certo set di dati con i parametri stimati nel modello, la statistica AIC è:

$$AIC = -2\log(f(X|\theta_m, m)) + 2p$$

e in accordo con il criterio, il modello migliore avrà il più basso valore di AIC.

Il Deviance Information Criterion (DIC) è definito come

$$DIC(m) = 2\overline{D(\theta_m, m)} - D(\overline{\theta_m}, m) = D(\overline{\theta_m}, m) + 2p_m$$

dove $D(\theta_m, m)$ è la devianza, che è proporzionale alla log-verosimiglianza:

$$D(\theta_m, m) = -2\log f(X|\theta_m, m)$$

questa è la misura della differenza tra la stima ottenuta con il modello e i dati (più bassa è la devianza meglio sarà il fit dei dati).

$\overline{D(\theta_m, m)}$ è la media a posteriori della log-verosimiglianza e p_m è il numero

⁸Gelman and Pardoe *Bayesian Measures of Explained Variance and Pooling in Multilevel (Hierarchical) Models*, 2006

”effettivo” dei parametri per il modello m dato da

$$p_m = \overline{D(\theta_m, m)} - D(\overline{\theta_m}, m)$$

con $\overline{\theta_m}$ che è la media a posteriori dei parametri coinvolti nel modello m .

Nei modelli normali, p_m è approssimativamente il numero vero di parametri p , quindi in questi casi il DIC e l’AIC quasi coincidono. Il valore più piccolo di DIC indica il modello best-fitting.

Per ottenere questa statistica test, *Jags* calcola la log-likelihood ad ogni iterazione del campionamento, utilizzando i valori dei parametri corrispondenti a questa iterazione.

3.8 Applicazione dei modelli con predittori ai dati di concentrazione radon.

Effetto random sull’intercetta Ora utilizziamo e confrontiamo i modelli multilevel descritti finora per fittare i dati di concentrazione radon indoor. Partiamo con un modello con un solo predittore (*mod1*), il piano, e manteniamo come unico effetto random l’intercetta per poi aggiungere uno alla volta gli altri K predittori:

$$y_i \sim N(\alpha_{j[i]} + \beta X_i, \sigma_y^2), \text{ per } i = 1, \dots, n \quad (3.29)$$

$$\alpha_j \sim N(\mu_\alpha, \sigma_\alpha^2), \text{ per } j = 1, \dots, J \quad (3.30)$$

La variabile dipendente y_i è il logaritmo della concentrazione di radon indoor sottratto il fondo e corretto per il fattore D (eq. 3.26). X è la matrice dei predittori $n \times K$, senza il termine costante, n il numero di misure disponibili, e β è il vettore dei K coefficienti.

I modelli studiati sono:

- *mod2*) effetto random: l'intercetta α_j , effetti fissi β_k : il piano e la categoria dell'edificio (luogo pubblico, luogo di lavoro, scuola o abitazione);
- *mod3*) effetto random: l'intercetta α_j , effetti fissi β_k : il piano, la categoria dell'edificio e il periodo in cui è stato costruito (dopo il 1960, prima del 1919 e tra il 1919 e il 1960);
- *mod4*) effetto random: l'intercetta α_j , effetti fissi β_k : il piano, la categoria dell'edificio, l'anno in cui è stato costruito e il materiale di cui è composto (laterizio-cemento, pietra e altro).

Per ciascun parametro del modello deve essere specificata la distribuzione a priori. Per l'effetto random dell'intercetta, associamo ad ogni α_j una distribuzione normale $N(\mu_\alpha, \sigma_\alpha^2)$, dove gli *iperparametri* sono definiti, come si fa generalmente (Gelman e Hill, 2007), attraverso una distribuzione a priori non informativa normale per μ_α : $N(0,1000)$, con un'ampia varianza, e una distribuzione uniforme nel range $[0,100]$ per le componenti della varianza σ_α e σ_y^2 .

Ai coefficienti della regressione β vengono associate delle distribuzioni normali non informative $N(0,1000)$, anch'esse con grande varianza.

Abbiamo 757 dati (n) disponibili dal database di ARTA, raccolti negli edifici di 96 comuni (J) della provincia dell'Aquila (tab. 3.1), dall'analisi multilevel bayesiana otteniamo i risultati riassunti nelle seguenti tabelle.

Possiamo confrontare il valore del DIC e di R_y^2 per ciascun modello (tab. 3.7).

**Capitolo 3. Analisi multilevel per le concentrazioni di radon indoor nella
provincia aquilana**

	mean	sd	2.50%	25%	50%	75%	97.50%	\hat{R}	n.eff
β_1	-0.72	0.09	-0.89	-0.78	-0.72	-0.66	-0.54	1.001	3000
β_2	-0.91	0.13	-1.17	-1.00	-0.91	-0.83	-0.65	1.002	1000
β_S	0.34	0.16	0.02	0.23	0.34	0.45	0.66	1.001	3000
β_{NCT}	-0.35	0.10	-0.55	-0.42	-0.35	-0.28	-0.15	1.002	1400
α_1	3.95	0.26	3.45	3.77	3.95	4.12	4.48	1.001	3000
α_2	3.63	0.13	3.38	3.54	3.63	3.72	3.89	1.002	1200
α_3	3.55	0.18	3.18	3.43	3.55	3.67	3.90	1.001	3000
α_4	4.07	0.18	3.71	3.94	4.07	4.19	4.43	1.001	3000
α_5	4.07	0.19	3.70	3.95	4.07	4.19	4.45	1.001	3000
α_6	3.99	0.29	3.43	3.79	3.99	4.19	4.59	1.001	3000
α_7	4.09	0.30	3.53	3.89	4.08	4.28	4.69	1.002	1000
μ_α	3.98	0.07	3.84	3.93	3.98	4.03	4.12	1.003	740
σ_α	0.32	0.06	0.20	0.28	0.32	0.36	0.45	1.005	560
σ_y	0.97	0.03	0.92	0.95	0.97	0.99	1.03	1.003	950
devianza	2115	13	2090	2106	2115	2124	2143	1.001	3000

Tabella 3.3: Modello (3.30) fittato con Jags, variabile dipendente $\ln(Rn_{in}^{new})$, 3 catene e 10000 interazioni, le prime 5000 non considerate; $\beta_1 = 1^\circ$ piano, $\beta_2 \geq 2^\circ$ piano, $\beta_S =$ sotterraneo; $\beta_{NCT} =$ piano terra senza contatto con il terreno; α_j le intercette per i primi 7 comuni.

**Capitolo 3. Analisi multilevel per le concentrazioni di radon indoor nella
provincia aquilana**

	mean	sd	2.50%	25%	50%	75%	97.50%	\hat{R}	n.eff
β_1	-0.68	0.10	-0.86	-0.74	-0.68	-0.61	-0.48	1.001	2600
β_2	-0.83	0.13	-1.10	-0.92	-0.83	-0.75	-0.57	1.002	1600
β_S	0.36	0.16	0.05	0.25	0.36	0.46	0.66	1.002	1700
β_{NCT}	-0.38	0.11	-0.58	-0.45	-0.38	-0.31	-0.17	1.001	3000
β_{scuola}	0.39	0.11	0.18	0.32	0.39	0.46	0.59	1.002	1800
β_{pubb}	0.00	0.16	-0.32	-0.11	0.00	0.11	0.32	1.003	840
β_{lavoro}	-0.03	0.13	-0.27	-0.11	-0.03	0.06	0.22	1.001	3000
α_1	3.88	0.26	3.38	3.71	3.87	4.05	4.40	1.001	3000
α_2	3.53	0.14	3.26	3.44	3.53	3.63	3.79	1.002	1600
α_3	3.41	0.20	3.02	3.28	3.41	3.55	3.79	1.003	910
α_4	3.99	0.19	3.62	3.86	3.98	4.11	4.36	1.001	3000
μ_α	3.90	0.08	3.73	3.84	3.90	3.95	4.07	1.005	500
σ_α	0.33	0.06	0.21	0.29	0.33	0.37	0.45	1.004	1400
σ_y	0.96	0.03	0.91	0.95	0.96	0.98	1.02	1.001	3000
devianza	2100	14	2074	2090	2099	2110	2131	1.002	2000

Tabella 3.4: Modello *mod2* fittato con Jags, variabile dipendente $\ln(Rn_{in}^{new})$, 3 catene e 10000 interazioni, le prime 5000 non considerate; $\beta_1 = 1^\circ$ piano, $\beta_2 \geq 2^\circ$ piano, $\beta_S =$ sotterraneo; $\beta_{NCT} =$ piano terra senza contatto con il terreno; β_{scuola} ; $\beta_{pubb} =$ ufficio pubblico, $\beta_{lavoro} =$ luogo di lavoro; α_j le intercette per i primi 4 comuni.

**Capitolo 3. Analisi multilevel per le concentrazioni di radon indoor nella
provincia aquilana**

	mean	sd	2.50%	25%	50%	75%	97.50%	\hat{R}	n.eff
β_1	-0.67	0.10	-0.85	-0.73	-0.67	-0.60	-0.48	1.002	2100
β_2	-0.79	0.13	-1.06	-0.88	-0.79	-0.70	-0.53	1.001	3000
β_S	0.40	0.15	0.10	0.30	0.41	0.50	0.71	1.002	2000
β_{NCT}	-0.37	0.11	-0.58	-0.44	-0.37	-0.30	-0.16	1.002	1200
β_{scuola}	0.38	0.11	0.17	0.31	0.38	0.45	0.59	1.001	3000
β_{pubb}	-0.05	0.16	-0.35	-0.16	-0.05	0.06	0.26	1.001	2100
β_{lavoro}	-0.05	0.12	-0.28	-0.13	-0.05	0.03	0.19	1.001	3000
β_{1919}	0.09	0.10	-0.10	0.02	0.09	0.16	0.29	1.001	3000
$\beta_{1919/1960}$	0.36	0.09	0.18	0.30	0.35	0.41	0.53	1.001	3000
α_1	3.78	0.25	3.30	3.61	3.79	3.95	4.29	1.002	1500
α_2	3.47	0.14	3.20	3.37	3.47	3.56	3.75	1.002	1400
α_3	3.29	0.20	2.90	3.16	3.29	3.43	3.68	1.001	3000
α_4	3.87	0.19	3.51	3.74	3.86	3.99	4.24	1.003	970
μ_α	3.78	0.09	3.59	3.72	3.78	3.84	3.96	1.005	440
σ_α	0.32	0.06	0.20	0.27	0.31	0.36	0.45	1.004	820
σ_y	0.95	0.03	0.90	0.94	0.95	0.97	1.01	1.001	2400
devianza	2088	15	2060	2078	2088	2098	2118	1.002	1300

Tabella 3.5: Modello *mod3* fittato con Jags, variabile dipendente $\ln(Rn_{in}^{new})$, 3 catene e 10000 interazioni, le prime 5000 non considerate; $\beta_1 = 1^\circ$ piano, $\beta_2 \geq 2^\circ$ piano, $\beta_S =$ sotterraneo; $\beta_{NCT} =$ piano terra senza contatto con il terreno; β_{scuola} ; $\beta_{pubb} =$ ufficio pubblico, $\beta_{lavoro} =$ luogo di lavoro; $\beta_{1919} =$ edifici costruiti prima del 1919; $\beta_{1919/1960} =$ costruiti tra il 1919 e 1960; α_j le intercette per i primi 4 comuni.

**Capitolo 3. Analisi multilevel per le concentrazioni di radon indoor nella
provincia aquilana**

	mean	sd	2.50%	25%	50%	75%	97.50%	\hat{R}	n.eff
β_1	-0.70	0.09	-0.89	-0.76	-0.70	-0.64	-0.52	1.001	3000
β_2	-0.80	0.13	-1.06	-0.89	-0.80	-0.72	-0.55	1.002	3000
β_S	0.34	0.16	0.03	0.24	0.35	0.44	0.64	1.001	3000
β_{NCT}	-0.40	0.10	-0.61	-0.47	-0.40	-0.33	-0.20	1.001	3000
β_{scuola}	0.33	0.11	0.13	0.26	0.33	0.40	0.54	1.001	3000
β_{pubb}	-0.07	0.16	-0.39	-0.18	-0.07	0.04	0.26	1.003	950
β_{lavoro}	-0.08	0.12	-0.31	-0.16	-0.07	0.01	0.16	1.001	3000
β_{pietra}	-0.03	0.11	-0.24	-0.10	-0.03	0.04	0.18	1.001	3000
β_{altro}	-0.79	0.19	-1.17	-0.92	-0.79	-0.66	-0.44	1.001	3000
β_{1919}	0.07	0.13	-0.19	-0.01	0.07	0.16	0.32	1.001	2600
$\beta_{1919/1960}$	0.33	0.10	0.14	0.27	0.33	0.40	0.52	1.001	3000
α_1	3.85	0.25	3.35	3.68	3.85	4.02	4.35	1.001	3000
α_2	3.58	0.14	3.32	3.49	3.58	3.67	3.85	1.002	1300
α_3	3.35	0.20	2.95	3.21	3.35	3.48	3.72	1.001	3000
α_4	3.91	0.19	3.54	3.79	3.91	4.04	4.29	1.001	3000
μ_α	3.87	0.09	3.69	3.80	3.86	3.93	4.05	1.001	2600
σ_α	0.33	0.06	0.21	0.29	0.33	0.37	0.46	1.005	410
σ_y	0.94	0.03	0.89	0.93	0.94	0.96	0.99	1.001	3000
devianza	2069	15	2042	2060	2068	2079	2100	1.002	1300

Tabella 3.6: Modello *mod4* fittato con Jags, variabile dipendente $\ln(Rn_{in}^{new})$, 3 catene e 10000 interazioni, le prime 5000 non considerate; $\beta_1 = 1^\circ$ piano, $\beta_2 \geq 2^\circ$ piano, $\beta_S =$ sotterraneo; $\beta_{NCT} =$ piano terra senza contatto con il terreno; β_{scuola} ; $\beta_{pubb} =$ ufficio pubblico, $\beta_{lavoro} =$ luogo di lavoro; $\beta_{pietra}; \beta_{altro} =$ altri materiali come il legno; $\beta_{1919} =$ edifici costruiti prima del 1919; $\beta_{1919/1960} =$ costruiti tra il 1919 e 1960; α_j le intercette per i primi 4 comuni.

	Df	AIC	BIC	logLik	χ^2	Df $_\chi$	Pr($> \chi^2$)		DIC	R_y^2
mod1	7	2170.2	2202.6	-1078.1	95.102	4	2.20E-16	***	2206.9	0.190
mod2	10	2160.5	2206.8	-1070.2	15.705	3	1.30E-03	**	2202.5	0.206
mod3	12	2148.1	2203.7	-1062	16.409	2	2.74E-04	***	2197.5	0.218
mod4	14	2142.1	2206.9	-1057	10.032	2	6.63E-03	**	2180.7	0.237

Tabella 3.7: Diagnostiche test a confronto per i modelli, fit eseguito con Jags

La tab.(3.7) mostra che il modello migliore per la descrizione del nostro set di dati è il *mod4*. Tutte le diagnostiche test sono minimizzate tranne che per il *BIC*. Il coefficiente di determinazione cresce linearmente con l'aumento delle variabili esplicative considerate e conferma comunque che il modello preferibile è il *mod4*.

Varying intercept and varying slope. Ora invece ci occupiamo di modelli più realistici a effetti misti dove alcuni coefficienti della regressione multilevel variano tra i gruppi e altri rimangono fissi. Avremo, quindi, una matrice B di $1 + K$ colonne quanti sono i predittori *random* più una colonna per l'intercetta e con J righe quanti sono i comuni, cioè i raggruppamenti delle n misure di concentrazione del gas. Il vettore EF raccoglie invece gli effetti fissi, ha lunghezza T quanti sono i coefficienti fissi del modello.

$$y_i \sim N(B_{j[i]}X_i + EFZ_i, \sigma_y^2), \text{ per } i = 1, \dots, n \quad (3.31)$$

$$B_j \sim N(M_B, \Sigma_B), \text{ per } j = 1, \dots, J \quad (3.32)$$

$M_B = (\mu_\alpha, \mu_{B_1}, \dots)$ è un vettore di lunghezza K che raccoglie le medie delle distribuzioni; Σ_B è la matrice di covarianza $(K + 1) \times (K + 1)$. X e Z sono le matrici dei predittori di elementi rispettivamente $n \times (K + 1)$ e $n \times T$.

Riassumiamo le statistiche per alcuni modelli analizzati servendoci della funzione *jags* in **R**.

Abbiamo utilizzato i quattro predittori: piano (*0CT*, 1° , ≥ 2 , *0NCT*, *S*), categoria edificio (*abitazione*, *scuola*, *luogo pubblico*, *luogo di lavoro*), materiale di costruzione (*laterizio-cemento*, *pietra*, *altro*) e periodo di edificazione (*post 1960*, *prima del 1919*, *tra il 1919 e 1960*) e di volta in volta abbiamo definito le caratteristiche del modello.

	DIC	R_y^2
<i>jags.mat</i>	2184.796	0.248
<i>jags.anno</i>	2171.856	0.239
<i>jags.cated</i>	2222.969	0.246
<i>jags.piano</i>	2192.363	0.249
<i>mat + anno</i>	2184.973	0.253
<i>piano + mat</i>	2188.278	0.262
<i>tutti random</i>	2236.855	0.272

Tabella 3.8: Statistiche test per i modelli *varying intercept and varying slope*.

- *jags.mat* abbiamo lasciato variare sui comuni, oltre all'intercetta, solo il coefficiente regressivo per il materiale da costruzione;
- *jags.anno* effetto random: intercetta e anno di edificazione dell'edificio;
- *jags.cated* effetto random: intercetta e tipologia dell'edificio;
- *jags.piano* effetto random: intercetta e piano dove viene posizionato il dosimetro;
- *materiale + anno* effetto random: intercetta, anno e materiale;
- *piano + materiale* effetto random: intercetta, piano e materiale;
- *tutti random* tutti i predittori random.

Dalle statistiche test riportate in tab.(3.8) si nota che R^2 cresce, questo farebbe pensare ad un fitting migliore, ma osservando anche il *DIC* è probabile che questo aumento sia dovuto solo al maggior numero dei parametri e non ad un effettivo miglioramento del modello. Successivamente servendoci anche dei risultati dell'analisi GWR (vedi cap. [4]) stabiliremo il modello che meglio descrive il set di dati.

Raggruppamento per classe litologica. Analizziamo, adesso, gli effetti dell'applicazione di un ulteriore raggruppamento dei dati oltre quello fatto su base

comunale, ottenuto organizzando i dati alla classificazione litologica del sito su cui si trovano gli edifici oggetto di misura.

Il modello con due raggruppamenti non annidati può essere scritto:

$$y_i = \mu + \alpha_{j[i]} + \beta_{j[i]}x_i + \gamma_{k[j]} + \epsilon_{ij} \text{ per } i = 1, \dots, n; \quad (3.33)$$

$$\alpha_j \sim N(0, \sigma_\alpha^2) \quad j = 1, \dots, J; \quad (3.34)$$

$$\gamma_k \sim N(0, \sigma_\gamma^2) \quad k = 1, \dots, K \quad (3.35)$$

$$\epsilon_{ij} \sim N(0, \sigma_y^2) \quad (3.36)$$

dove μ è l'intercetta complessiva fissa, che racchiude l'effetto dello shift sulla distribuzione dovuto sia al raggruppamento geologico che comunale. I boxplot in fig.(3.4) riassumono le statistiche dei dati appartenenti alle varie classi litologiche. Oltre alla categorizzazione fatta in base alla carta in fig.(3.3), abbiamo considerato anche una nuova macro-variabile (*gross_lito*) che accorpa in quattro macrocategorie le classi litologiche: QCD: depositi quaternari; LIM: calcari e dolomie; SDS: arenarie e CS: argille e sabbie. Si nota che nessuno dei raggruppamenti ha grande influenza sui livelli di radon, in quanto i valori medi risultano poco differenti tra loro.

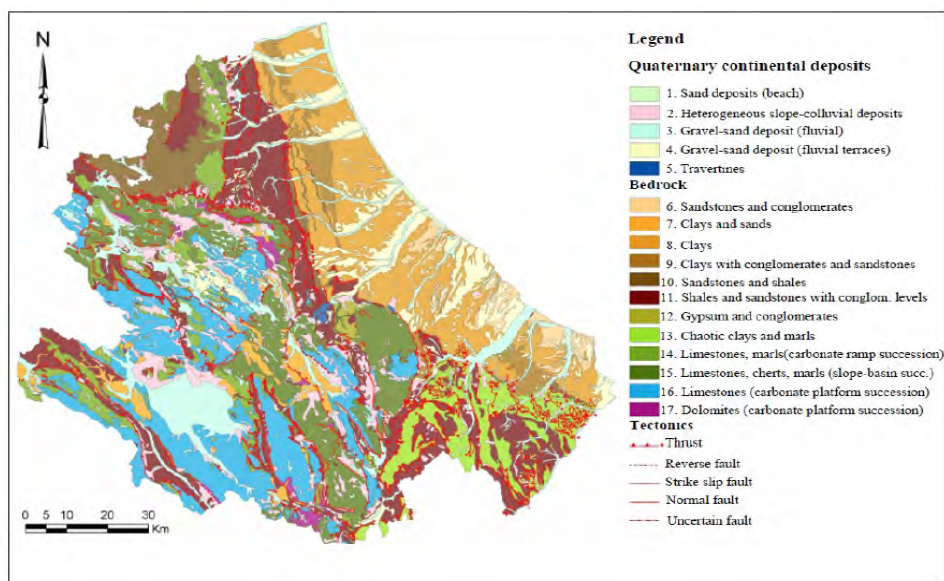


Figura 3.3: Mappa litologica della regione Abruzzo; le litologie sono raggruppate in un numero limitato di unità definite in funzione della loro influenza sulla distribuzione di radon (Pasculli et al., 2013 e referenze in esso).

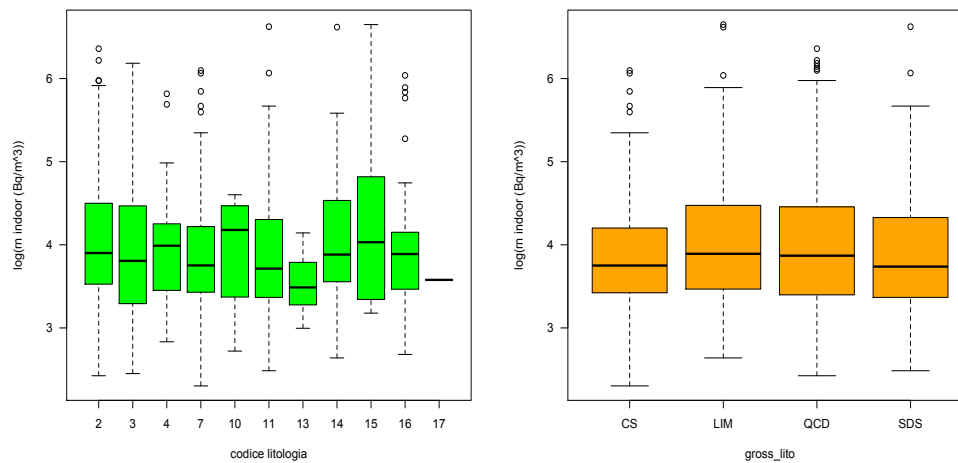


Figura 3.4: Boxplot della concentrazione log-trasformata di radon, raggruppata secondo le classi litologiche a destra per la variabile *lito* (fig. 3.3) e a sinistra per la *gross_lito*.

**Capitolo 3. Analisi multilevel per le concentrazioni di radon indoor nella
provincia aquilana**

mod	μ	st.err	t-value	$\hat{\sigma}_\alpha^2$	$\hat{\sigma}_\beta^2$	$\hat{\sigma}_y^2$	AIC	DIC
mod(lito)	3.76	0.06	67.11	0.33	0.00	0.99	2195	2287
mod(gross_lito)	3.75	0.06	67.11	0.30	1.51e-8	0.99	2195	2286

Tabella 3.9: Riassunto risultati dei modelli con gli effetti geologici.

Dalla tab.(3.9) si evince che la varianza spiegata dagli effetti geologici è nettamente inferiore rispetto a quella dovuta all' appartenenza ad un certo comune e possiamo considerarla di fatto irrilevante per la nostra indagine. D'altre parte ad una conclusione analoga si è giunti nella tesi di Palermo (2010) analizzando un campione più numeroso rappresentativo di tutto il territorio regionale.

Capitolo 4

Geographically Weighted Regression e mappa del potenziale radon geogenico

La distribuzione di radon nel terreno e conseguentemente nell'ambiente interno degli edifici è profondamente connessa con le caratteristiche geologiche del territorio (UNSCEAR, 2000) . Con l'analisi multilevel siamo riusciti a definire una nuova variabile Rn_{std} per la misura di concentrazione di radon indoor filtrata, per quanto possibile, della variabilità dovuta alle caratteristiche dell'edificio. Ora, attraverso la Geographically Weighted Regression (GWR), studiamo la relazione tra la concentrazione di gas indoor e le proprietà geologiche e geofisiche dell'ambiente circostante l'edificio in cui è stata effettuata la misura. Lo scopo è mappare una quantità il più possibile rappresentativa del rischio radon, vale a dire una misura di "ciò che la Terra eroga" in termini di radon indoor in un edificio con caratteristiche standard, indipendente, per quanto possibile, da fattori antropici. Questa quantità, o variabile spaziale, è detta potenziale radon geogenico (Dubois et al.,2010). Nella prima parte del capitolo viene descritta l'applicazione del GWR al nostro set di dati, con i relativi risultati; nella seconda parte invece si costruisce la mappa del potenziale radon geogenico per la provincia dell'Aquila.

L'analisi multilevel ha permesso di definire un nuova variabile Rn_{std} rappre-

sentativa della concentrazione di radon indoor misurata in un edificio standard, cioè in condizioni fisse di piano (piano con contatto a terra), di tipologia (abitazione), di materiale (laterizio e cemento) e di anno di costruzione (dopo il 1960). Ora utilizziamo questa variabile per definire la relazione diretta tra concentrazione di radon indoor e fattori geologici, o in generale ambientali, del sito dell'edificio oggetto di misurazione. Le variabili esplicative del gas indoor, come visto nel primo capitolo, possono suddividersi in tre gruppi: contenuto radiometrico del suolo, caratteristiche geologiche e geomorfologiche del terreno sottostante e attorno all'edificio e variabili climatiche.

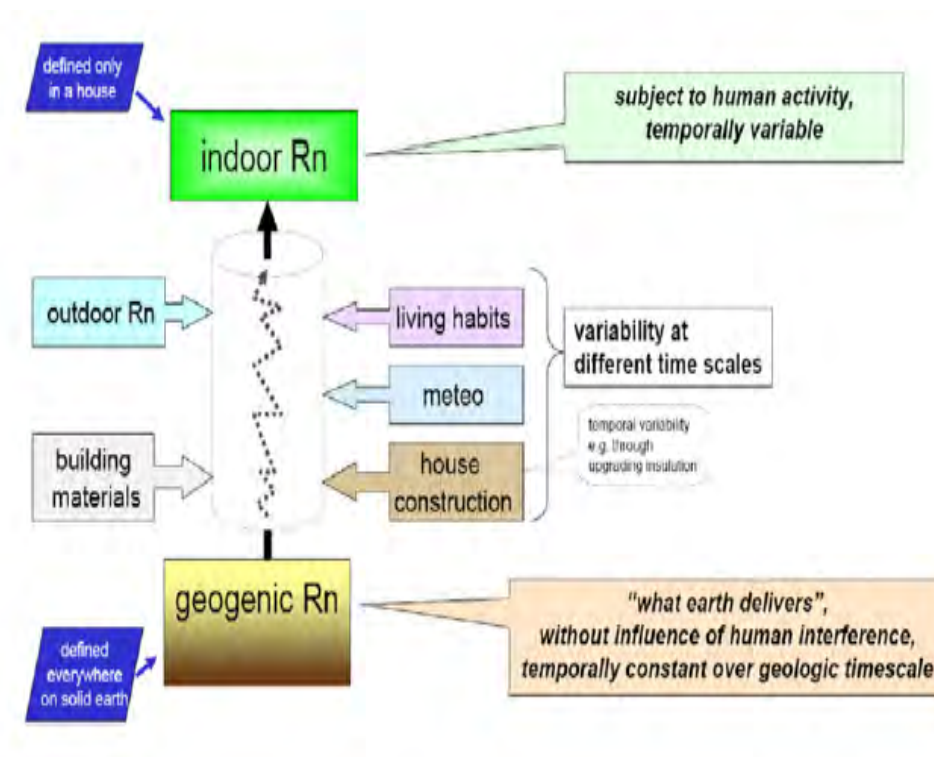


Figura 4.1: Dal radon geogenico alla concentrazione indoor (Gruber et al., 2013).

- *Dati geologici.*

Per comprendere l'incidenza delle caratteristiche geologiche del sito di misura sulla variabilità della concentrazione di radon indoor, nella provincia dell'Aquila o più in generale in Abruzzo, i parametri principali da considerare sono: la classificazione litologica, lo spessore, proprietà quali porosità

e permeabilità, lo stato di fratturazione e il grado di carsismo. Altri parametri come il gradiente geotermico o le sorgenti di acqua calda, non sono inclusi nella nostra analisi, perché sono fenomeni irrilevanti nel territorio (Pasculli et al., 2013).

Le litologie affioranti sono raggruppate in un numero limitato di unità e la mappatura geologica è basata su *unità di roccia*. Un'unità di roccia è un regione circoscritta (con dettaglio spaziale di 100 m) rappresentante una massa omogenea rocciosa a cui associare un valore univoco per ognuno dei parametri geologici considerati. L'analisi della regione Abruzzo ha portato alla costruzione di sei differenti mappe che mostrano la distribuzione spaziale delle caratteristiche litologiche e delle variabili connesse (fig. 4.3).

La permeabilità è direttamente influenzata dalla fratturazione, dalla porosità e dalla granulometria delle rocce e dei suoi componenti, pertanto la mappa è acquisita combinando tutte queste informazioni, come una sorta di variabile derivata (analisi condotta da Miccadei e Piacenti in Pasculli et al., 2013). Rispetto al lavoro di Pasculli et al. 2013, la permeabilità dei depositi quaternari (con particolare riferimento alle conche intermontane di Sulmona, L'Aquila e Avezzano) è stata rivista da Palermi e Desiderio utilizzando, per la codifica, direttamente le informazioni riportate nella mappa geologica dell'ISPRA (progetto CARG, 2012) e dove non disponibile, la vecchia carta geologica 1:10000. La nuova classificazione ha permesso di correggere alcune incongruenze e inesattezze, rivelando la necessità di un'analisi quanto più precisa e puntuale per la corretta codifica del parametro.

Thickness (m)	Class	Average	Variability			
			1	2	3	4
low (0-50)	1	25	±5	±15	±25	±75
medium (50-500)	2	275	±100	±250	±300	±400
high (500-1000)	3	750	±150	±350	±500	±700
very high (>1500)	4	2000	±300	±500	±900	±1200
Fracturation (f/m)	Class	Average	Variability			
			1	2	3	4
not fractured	0	-	-	-	-	-
low (1-5)	1	2	±2	±3	±4	±5
medium (5-10)	2	8	±4	±6	±8	±10
high (10-40)	3	25	±15	±20	±25	±30
very high (40-100)	4	70	±30	±40	±50	±70
Porosity (%)	Class	Average	Variability			
			1	2	3	4
not porous	0	-	-	-	-	-
low (0.01-0.1)	1	0.05	0	±0.05	±0.1	±0.15
medium (0.1-0.3)	2	0.2	0	±0.05	±0.1	±0.15
high (0.3-0.5)	3	0.4	0	±0.1	±0.2	±0.3
very high (0.5-0.7)	4	0.6	0	±0.1	±0.2	±0.3
Permeability (k = m ²)	Class	Average	Variability			
			1	2		
low	1	10 ⁻⁸	10 ⁻⁷ -10 ⁻⁹	10 ⁻⁶ -10 ⁻¹⁰		
medium	2	10 ⁻⁴	10 ⁻³ -10 ⁻⁵	10 ⁻² -10 ⁻⁶		
high	3	10 ⁻¹	10 ⁰ -10 ⁻²	10 ¹ -10 ⁻³		
very high	4	10 ¹	10 ² -10 ⁰	10 ² -10 ⁻¹		
Karst (%)	Class					
null	0					
low	1	based on a qualitative description				
medium	2					
high	3					

Figura 4.2: Classificazione dei parametri geologici e geomorfologici delle unità di roccia mappate nella regione Abruzzo (Pasculli et al., 2013).

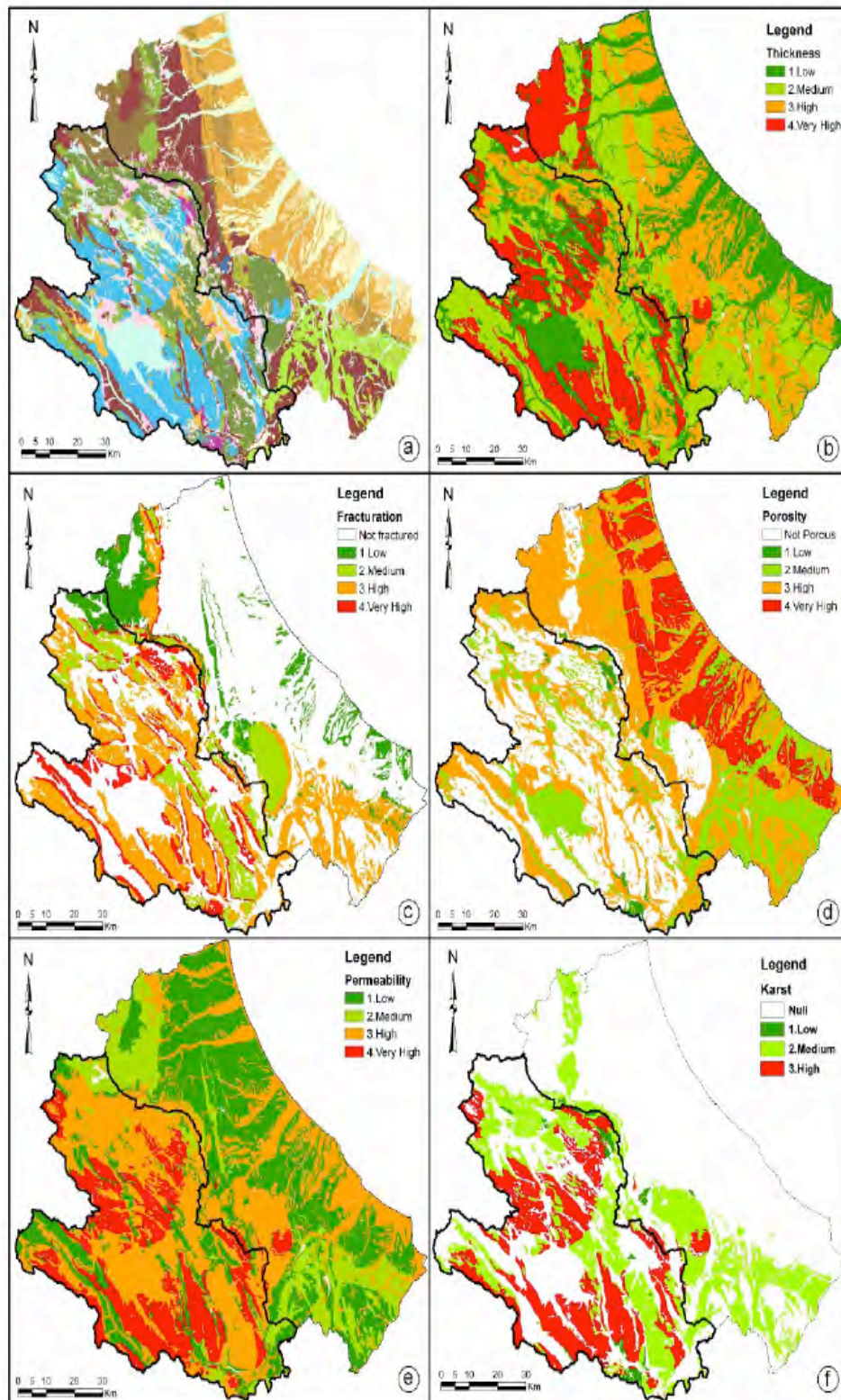


Figura 4.3: Mappa litologica e dei parametri geologici per la regione Abruzzo (Pasculli et al., 2013); la linea nera delimita la provincia dell' Aquila. a) litologia; b) spessore; c) frantumazione; d) permeabilità; e) porosità; f) carsismo.

- *I dati radiometrici.*

I dati radiometrici relativi alle concentrazioni nel suolo di potassio (K), uranio equivalente (eU) e torio equivalente (eTh) sono determinati dal segnale γ emesso rispettivamente dal ^{40}K , ^{214}Bi e ^{208}Tl , presenti nel cosiddetto topsoil (strato superficiale del suolo, fino ad una profondità di circa 30 cm). Bellotti et al. (2007) condussero delle misure di spettrometria gamma su 200 campioni di suolo raccolti nel territorio della Provincia dell'Aquila secondo un disegno campionario basato sui nodi di una griglia regolare di passo 5 km (dati riportati anche nella tesi non pubblicata di S. Fattori, 2006).

Il ^{214}Bi e ^{208}Tl sono prodotti di decadimento a vita breve, rispettivamente, del ^{222}Rn e ^{220}Rn (isotopo del radon dovuto al ^{232}Th). Le concentrazioni equivalenti di eU e eTh rappresentano realmente la quantità di uranio e torio nel campione solo in condizione di perfetto equilibrio secolare nella catena radioattiva. In generale, questo stato è facilmente ottenibile per il torio e non per l'uranio, che è più solubile e mobile del primo rendendo difficile il verificarsi della condizioni di equilibrio con gli atomi "figli". Seguendo Scheib et al. 2006, includiamo come potenziale regressore del radon anche il potassio (K), infatti questo è un buon indicatore del contenuto di argilla e conseguentemente della permeabilità dell'unità di roccia del sito di misura. L'isotopo radioattivo più abbondante del K è il ^{40}K con un'emivita di 1.25×10^9 anni. Il segnale γ emesso dal ^{40}K quando decade in ^{40}Ar può essere utilizzato direttamente per stimare la quantità totale di K nel suolo (espressa in %), grazie alla costanza del rapporto isotopico

Il dataset è stato ampliato con misure fatte, non ancora pubblicate, da ARTA seguendo la stessa procedura di Bellotti in numero comunque non sufficiente a permetterci di associare i dati puntuali radiometrici ad ogni misura di concentrazione di radon indoor presente nel database a disposizione. Si è provveduto, quindi, ad una interpolazione spaziale dei dati con

il block kriging (Olea, 1999). Dividendo infatti il territorio regionale in celle regolari di 2×2 km, il block kriging permette di stimare la media di ciascuna variabile in esame in corrispondenza di ogni centroide della griglia (si veda Nissi et al. 2012).

- *Dati climatici.*

Le concentrazioni di radon indoor sono fortemente condizionate dalla temperatura esterna. Lo stile di vita della popolazione cambia nei periodi freddi, riducendo la ventilazione e aumentando il gradiente termico tra l'interno e l'esterno, incrementando pertanto la capacità di ingresso e di accumulo di radon indoor (cfr. il paragrafo [1.2.4]). Per tenere conto dei parametri climatici abbiamo considerato come variabile esplicativa la temperatura media annua del sito di misura. Questa è stata stimata grazie ai dati raccolti, dal 1991 al 2011, da 43 stazioni termometriche del Servizio Idrografico e Mareografico Italiano (S.I.M.I) sparse sul territorio aquilano. Le temperature medie annue per i siti di misura della concentrazione di radon indoor sono stati predette con un kriging puntuale con drift esterno (Hudson e Wackernagel, 1994), che utilizza l'altitudine del luogo come variabile esplicativa per la stima (si veda l'app. A par. [A.5]).

Negli studi condotti fino ad ora per descrivere la variabilità delle concentrazioni di radon indoor e l'influenza di possibili variabili esplicative, si utilizza spesso la regressione lineare multivariata (Appleton et al., 2008) o analisi bayesiane multilevel (Price et al. 1996; Apte et al. 1999; Bossew et al. 2008) che considerano il fenomeno spazialmente stazionario.

La differenza sostanziale nell'applicazione della Geographically Weighted Regression è che questa tiene conto della variabilità locale della relazione tra le variabili in studio e della possibile autocorrelazione spaziale dei predittori (Fotheringham et al., 2002). La presenza di autocorrelazione viola l'assunzione statistica di indipendenza. Questo implica che le tecniche usuali di regressione globale (OLS) siano inappropriate per modellare il fenomeno. La non stazionarietà spaziale

può verificarsi almeno in due condizioni: quando la relazione tra le variabili sullo spazio differisce o quando l'equazione di regressione potrebbe includere e/o escludere fattori rilevati (errori di specificazione). Qualunque sia la causa dell'eterogeneità, tale modellizzazione non stazionaria migliora la precisione della stima in quanto il modello, fittato localmente, è influenzato dalle sole condizioni locali. Inoltre, il GWR permette di generare delle mappe che integrano diverse informazioni spaziali e mostrano come l'influenza dei fattori del potenziale radon cambia nello spazio.

4.1 Geographically Weighted Regression

La Geographically Weighted Regression (GWR) (Brunsdon, Fotheringham, and Charlton, 1996; Fotheringham, Charlton, and Brunsdon, 1998, 2001) è un'utile tecnica di modellizzazione locale per le relazioni spaziali. Un modello misto (locale-globale) di GWR può essere scritto come:

$$y_i = \sum_l^L \gamma_l z_{l,i} + \sum_k^M \beta_k(u_i, v_i) x_{ik} + \epsilon_i \quad (4.1)$$

dove y_i è la variabile dipendente nel punto i di coordinate (u_i, v_i) ; $\beta_k(u_i, v_i)$ con $k = 1, \dots, M$, sono i coefficienti di regressione per ogni punto i e per ogni variabile k ; x_{ik} è il valore della k -esima covariante predittiva in i ; z_{li} con $l = 1, \dots, L$ è l' l -esima variabile indipendente con coefficienti fissi γ_l ; ϵ_i è il termine di errore in i che segue una distribuzione normale con media nulla e varianza omogenea. L'equazione (4.1) crea una superficie continua di parametri stimati localmente.

In notazione matriciale il modello è:

$$y = \gamma Z + (\beta \cdot X)1 + \epsilon \quad (4.2)$$

con y vettore $n \times 1$ vettore della variabile dipendente. Z è la matrice $n \times L$ delle covarianti che hanno coefficienti di regressione globale (γ). X è la matrice $n \times M$ delle variabili esplicative i cui coefficienti variano spazialmente; β , infatti, è una

matrice $n \times M$ e la i -esima riga è data da:

$$\hat{\beta}(i) = (X^T W(i) X)^{-1} X^T W(i) y \quad (4.3)$$

con $W(i) = \text{diag}[w_1(i), \dots, w_n(i)]$ che è una matrice diagonale che varia per ogni punto di calibrazione i e applica i pesi a ciascuno delle n osservazioni, quindi w_{ni} è il peso attribuito alla regione n nel processo di calibratura nell'intorno del punto i . Alle misure vicine saranno assegnati pesi più grandi rispetto a quelle via via più lontane.

Quindi un importante step nella stima GWR è la scelta dello schema con cui si definiscono questi pesi: è necessario definire una funzione peso (*kernel shape*) e una ampiezza di banda (*bandwidth*). Possono essere considerate e calibrate diverse funzioni peso ("kernels"), sebbene tendono ad essere sempre di tipo Gaussiane o simili a queste, riflettendo la dipendenza trovata in molti processi spaziali. Qualsiasi funzione peso venga utilizzata, comunque, il risultato è fortemente determinato dalla scelta del bandwidth. Questa è la misura della larghezza di banda, o distanza di decadimento, della funzione peso. Si può scegliere un bandwidth costante intorno ad ogni punto in cui viene applicato il modello di regressione, o flessibile (adattivo) mantenendo fisso il numero di punti da considerare nella regressione.

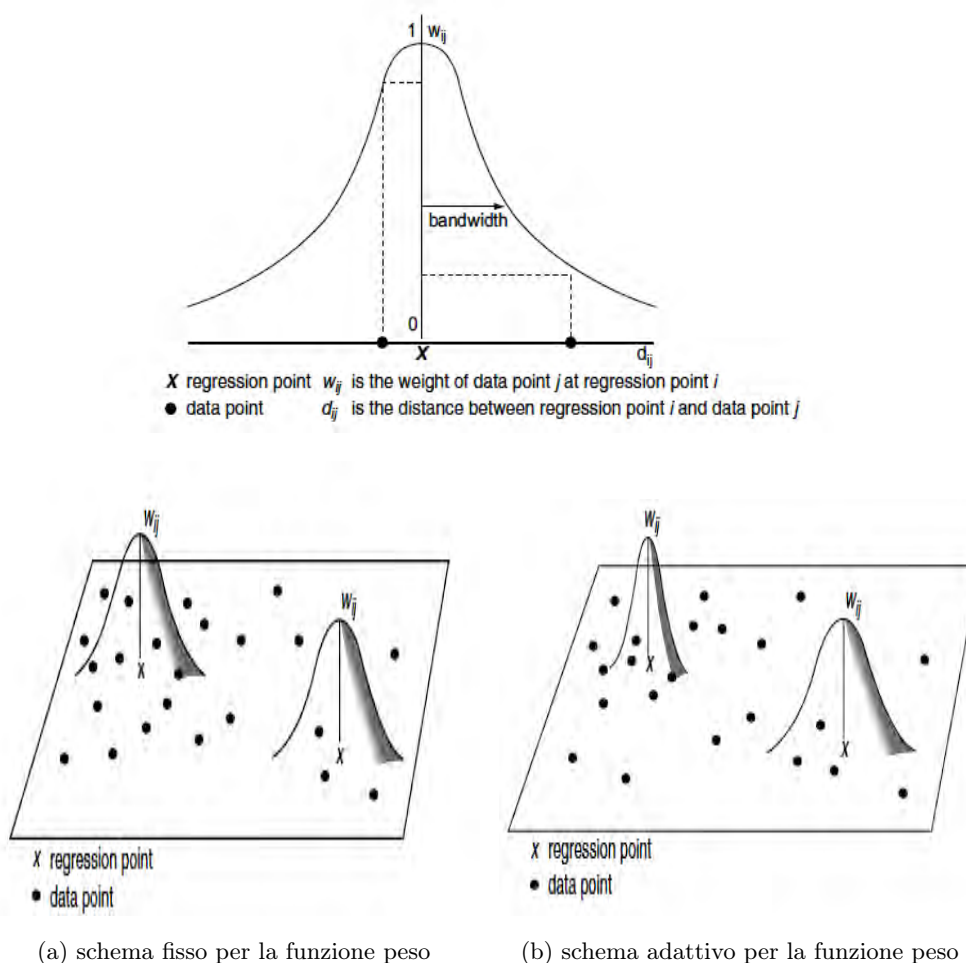


Figura 4.4: Caratteristiche della funzione peso.

Una scelta usuale per la funzione peso è una curva Gaussiana, espressa nella forma:

$$w_{ij} = \exp\left[-\frac{d_{ij}^2}{2b^2}\right] \quad (4.4)$$

con d_{ij} che è la distanza tra il punto osservato i e gli N punti vicini ($j = 1, \dots, N$) e b è il bandwidth. Questa funzione gaussiana permette a ciascuna osservazione compresa di avere un peso diverso da 0. Ciò rende possibile che ogni misura sia collegata alle altre.

Conseguentemente, la calibrazione del modello coinvolge la scelta del numero di punti vicini da includere nella stima dei parametri locali (N), o la distanza di decadimento (b) della funzione peso. Possono essere usati differenti modi per

scegliere il miglior bandwidth. Tra questi ci sono l'*Akaike Information Criterion* (AIC) (già visto nel caso dell'analisi multilevel), e la *cross validation*. Il criterio della cross-validation si basa sulla tecnica dei minimi quadrati e può essere formalizzato come segue:

$$CV = \sum_{i=1}^n [y_i - \hat{y}_{\neq i}(b)]^2 \quad (4.5)$$

dove $\hat{y}_{\neq i}(b)$ è il valore stimato di y_i con la GWR, utilizzando un bandwidth pari a b e omettendo le osservazioni relative a i . Scelta la matrice dei pesi, il valore di b che produce il minimo di CV è il bandwidth ottimale. Per dimostrare che il GWR descrive la relazione significativamente meglio di un modello globale ordinario viene implementato una ANOVA test, basata sulla distribuzione *F di Fisher*. La F-statistic si definisce come segue:

$$F = \frac{RSS_0/d}{RSS_1/d_1} \quad (4.6)$$

dove RSS_1 e RSS_0 sono la somma dei quadrati dei residui rispettivamente del modello GWR e della regressione globale (OLS), mentre d_1 e d_0 sono i gradi di libertà dei due modelli. L'ipotesi nulla è che il modello GWR non migliori il fit dei dati rispetto ad un modello globale. Più è grande la F-statistic maggiore sarà l'evidenza del miglioramento ottenuto con il fit in esame (Brunsdon et al.,1996).

La scelta se considerare un parametro globale o meno, viene fatta attraverso la definizione di un criterio statistico, *DIFF of Criterion*, sulla base di quanto riportato in Nakaya (2012), fondato sul confronto di modelli con differenti ipotesi iniziali. Per testare, infatti, la variabilità geografica del k -esimo coefficiente regressivo, si confrontano le statistiche test del fit ottenuto con una GWR considerando tutti i predittori locali (GWR_local) e un modello misto in cui solo il coefficiente globale è riferito alla variabile k -esima (GWR_kglo). Il modello GWR_local e il GWR_kglo, vengono confrontati attraverso l' AIC_c . Il migliore avrà l' AIC_c più basso e quindi, a seconda delle scelte fatte per quel fit,

il predittore sarà stazionario o meno.

L' Akaike Information Criterion corrected (AIC_c) è una statistica leggermente modificata rispetto l'AIC ed è definito come (Fotheringham et al. 2002):

$$AIC_c = 2n\ln(\sigma) + n\ln(2\pi) + \left[\frac{n + tr(S)}{n - 2 - tr(S)} \right] \quad (4.7)$$

dove n è la grandezza del campione, σ è la stima della deviazione standard del termine di errore, e $tr(S)$ è la traccia della matrice "hat" S di GWR:

$$\hat{y} = Sy$$

con y che è il vettore della variabile dipendente e \hat{y} il vettore dei valori stimati da GWR. Il test viene ripetuto per tutti i K coefficienti di regressione.

Per svolgere l'analisi, ci siamo serviti del software GWR4 di Nakaya et al. (2012), reperibile dal sito <http://gwr.nuim.ie>. Nell'output del GWR4 viene riportata la differenza tra il modello originale di GWR e quello modificato, in termini dell' indicatore AIC_c . Mantenendo fisso il coefficiente di regressione k , il $DIFF_Criterion$ mostrato in output è

$$DIFF_Criterion = AIC_c^{originale} - AIC_c^{modificato}$$

quindi se fosse positivo significherebbe che il modello GWR_kglo fitta meglio i nostri dati e sarebbe lecito considerare stazionario il coefficiente di regressione, mentre se fosse negativo potremmo concludere che il modello con tutti i coefficienti locali migliora il nostro fit di dati.

4.2 Radon indoor e Geographically Weighted Regression

Per poter applicare correttamente l'analisi spaziale ai dati standardizzati di concentrazione di radon indoor, abbiamo mediato le misure raccolte nello stesso

edificio. Infatti, si verifica sovente che tra i 757 dati disponibili dal database di ARTA, per uno stesso edificio si abbiano misure, svolte in ambienti differenti, per esempio su piani diversi. Questi valori sono utili per l'analisi multilevel, ma ovviamente sono fonte di errore per l'indagine spaziale. Per evitare questi problemi, abbiamo calcolato la media geometrica delle concentrazioni di radon misurate in uno stesso edificio, ottenendo un dataset ridotto di 710 misure. Nei boxplot di fig.(4.6) i dati della concentrazione di radon indoor sono stati raggruppati in base alla classe di permeabilità e al piano in cui è stata fatta la misura, evidenziando, come immaginato, una progressiva riduzione del valore medio salendo di piano. La concentrazione di radon indoor cresce in relazione anche all'aumento della permeabilità del suolo (tab. 4.2 e fig. 4.7).

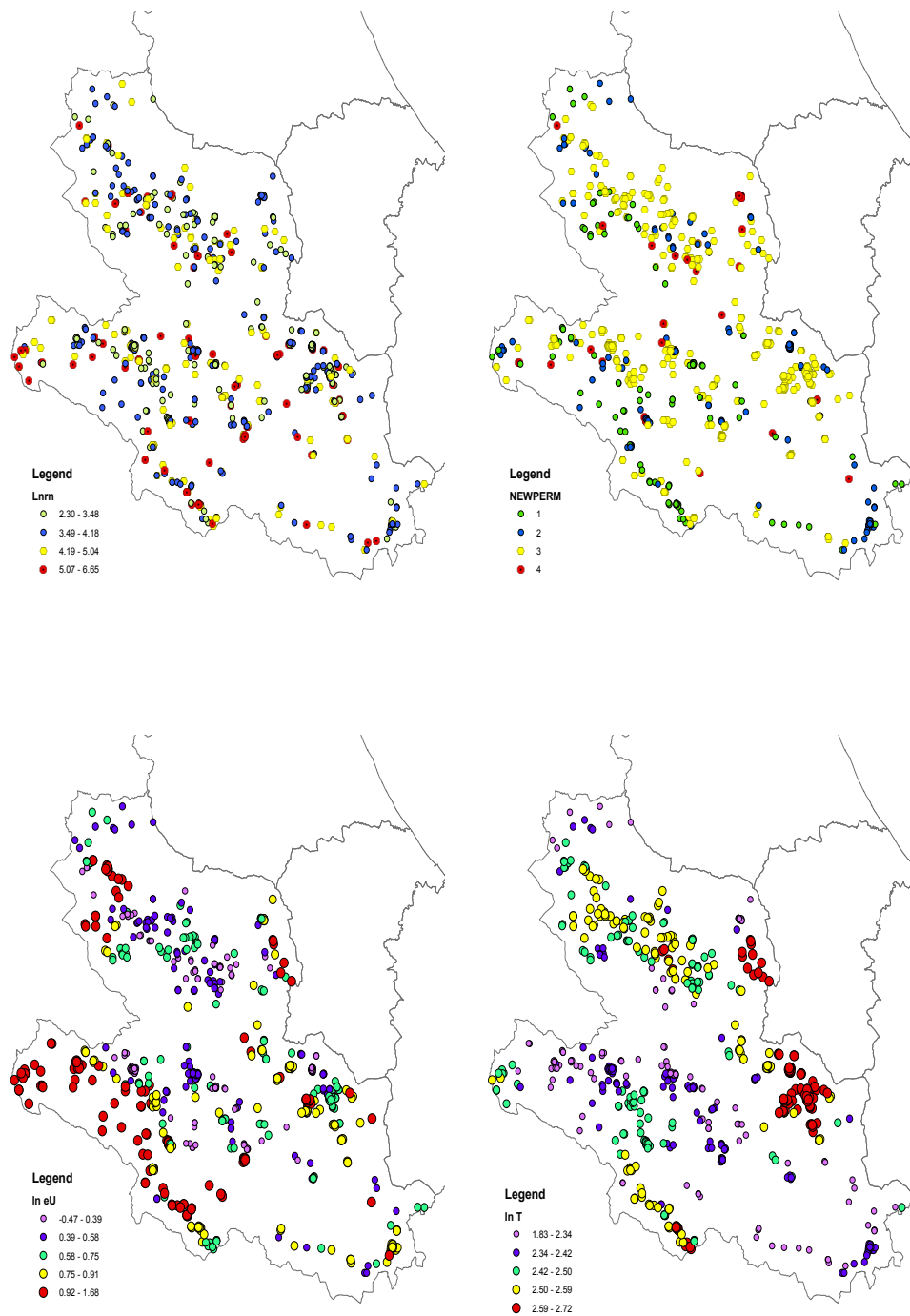


Figura 4.5: Set di dati georeferenziati nella provincia dell'Aquila; da sinistra in alto: misure radon indoor (log-trasformato) (Bq m^{-3}), classe di permeabilità (1: bassa; 2: media; 3: alta; 4: altissima); $\ln(\text{contenuto di uranio eq})$ (ppm) ; $\ln(\text{temperatura media annua})$ ($^{\circ}\text{C}$).

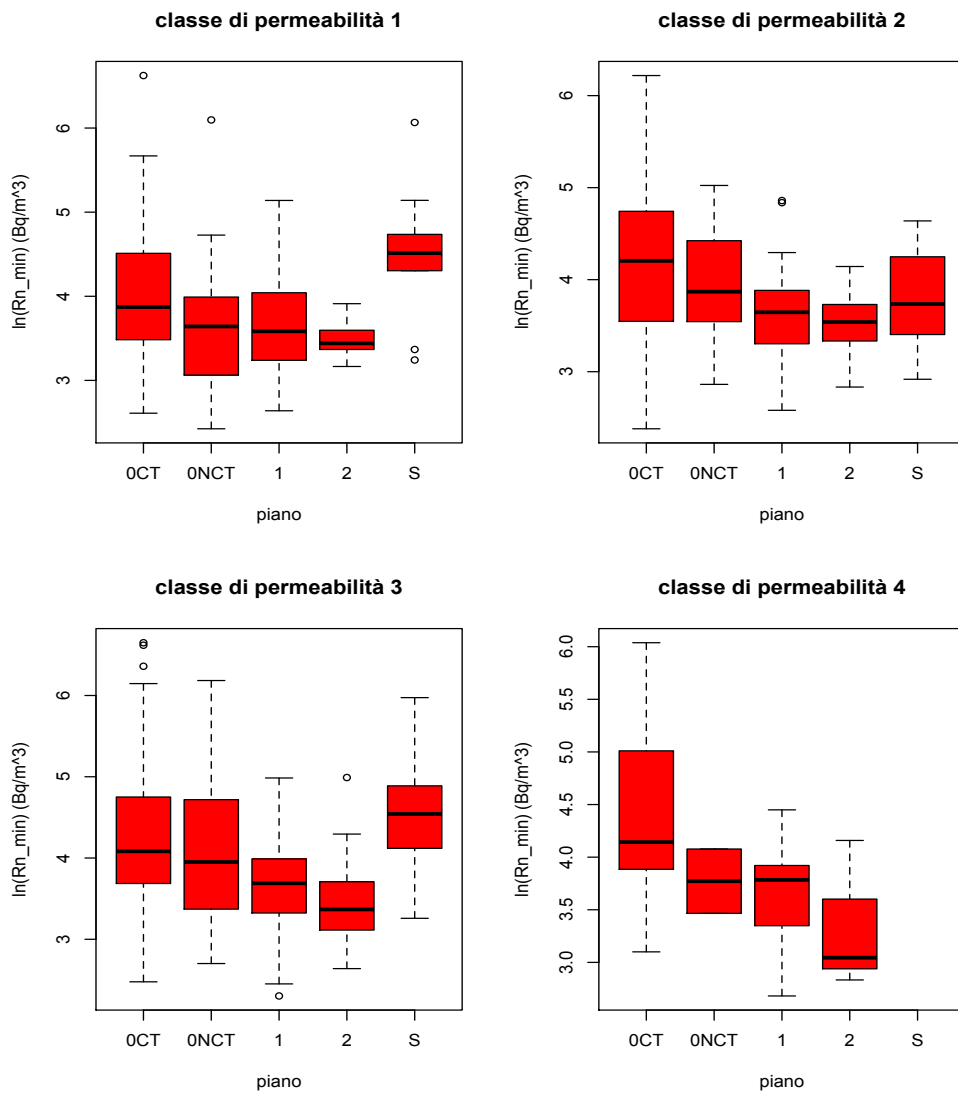


Figura 4.6: Boxplot per i dati misurati di concentrazione radon indoor raggruppati in base alla classe di permeabilità (1: bassa; 2: media; 3: alta; 4: altissima). *OCT*: piano terra a contatto con il terreno; *ONCT*: piano terra senza contatto; 1: primo piano; 2: piano > 1; *S*: sotterraneo.

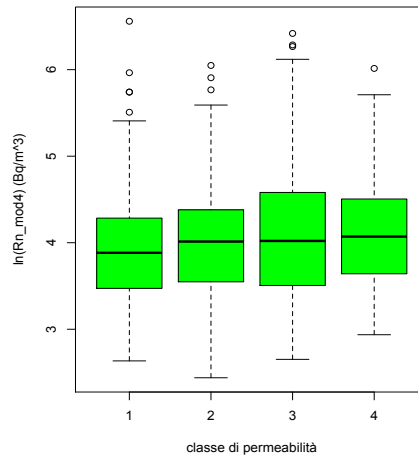


Figura 4.7: Boxplot per i dati standardizzati con il mod4 (si veda il cap [3]) raggruppati per classe di permeabilità

Per stabilire quali variabili esplicative del nostro modello siano realmente necessarie per *spiegare* i dati di radon indoor, analizziamo la relazione tra la concentrazione del gas e le variabili geologiche con una regressione multivariata globale. Il test di ipotesi t è riportato in tab.(4.1) insieme al coefficiente VIF (*variance inflation factors*) che interpreta il grado di collinearità¹ tra i predittori. Il VIF è calcolato per ciascuna variabile ed è dato da:

$$VIF_j = \frac{1}{1 - R_j^2} \quad (4.8)$$

con R_j^2 che è il coefficiente di determinazione della regressione per il parametro j - *esimo* calcolato utilizzando le restanti covarianti.

Se la j - *esima* variabile non presenta alcuna relazione lineare con le altre si ha che $R_j^2 = 0$ e $VIF_j = 1$; in presenza di quasi multicollinearità, VIF_j misura l'entità dell'aumento della varianza di β_j dovuto a tale problema e se la j - *esima* covariata dipende linearmente dalle altre si ha che $R_j^2 = 1$ e $VIF_j = \infty$.

¹*collinearità*: quando due o più variabili regressive sono linearmente dipendenti.

Capitolo 4. Geographically Weighted Regression e mappa del potenziale radon geogenico

OLS	Variabile	Stima	Stan. Error	t(Stima/SE)	Pr(> t)	VIF
	β_0 (Intercept)	4.871	0.510	9.56	0.0000	
$\ln Rn_{mod4}$	β_1 (ln perm)	0.012	0.004	2.69	0.0072	1.10
	β_2 (ln frat)	-0.034	0.049	-0.70	0.4840	10.48
	β_3 (ln eth)	-0.111	0.113	-0.98	0.3272	2.32
	β_4 (ln K)	-0.032	0.113	-0.28	0.7809	2.38
	β_5 (ln T)	-0.332	0.186	-1.78	0.0750	1.10
	β_5 (ln spess)	0.025	0.024	1.03	0.3016	1.66
	β_6 (ln por)	-0.011	0.053	-0.20	0.8428	9.45
	β_7 (ln eU)	0.254	0.096	2.64	0.0084	1.18

Tabella 4.1: Riassunto della regressione lineare multivariata (rifiutiamo l'ipotesi nulla $\beta = 0$ se $|t| \geq 1.96$)

Con la tecnica di *backward stepwise*, eliminiamo di volta in volta le variabili con t-value basso (rifiutiamo l'ipotesi nulla $\beta = 0$ se $|t| \geq 1.96$) e VIF alto, fino ad ottenere un modello con tutte le variabili significative e indipendenti. Possiamo quindi considerare solo tre variabili predittive (log-trasformate): la quantità di uranio equivalente, la temperatura media annua e la permeabilità media del suolo.

OLS	Estimate	Std. Error	t value	Pr(> t)	VIF
β_0 (Intercept)	4.6225	0.4472	10.34	0.0000	
β_3 (ln perm)	0.0107	0.0042	2.54	0.0112	1.00
β_2 (ln T)	-0.2533	0.1780	-1.42	0.1551	1.01
β_1 (ln eU)	0.1915	0.0886	2.16	0.0310	1.00

Tabella 4.2: Riassunto della regressione lineare con tre variabili esplicative, $\ln Rn_{mod4}$ variabile dipendente.

Per verificare l'esistenza di autocorrelazione nelle variabili, calcoliamo l'indice I di Moran globale. Questo coefficiente è, infatti, spesso utilizzato in geostatistica per valutare il grado di dipendenza spaziale tra i valori assunti da una variabile ed è dato da (Cliff e Ord, 1981; Upton e Fingleton, 1985):

$$I = \frac{n}{\sum_{i,j=1}^n w_{ij}} \frac{\sum_{i,j=1}^n w_{ij} (y_i - \bar{y})(y_j - \bar{y})}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (4.9)$$

dove n è il numero di unità geografiche; y_i è la variabile che descrive il fenomeno oggetto di analisi nella regione i ; \bar{y} rappresenta la media campionaria e w_{ij} è la matrice dei pesi, i quali sono inversamente proporzionali alla distanza tra il punto i e il punto j .

Se valori simili dei parametri - alti o bassi - sono localizzati frequentemente nello stesso intorno spaziale è presente una autocorrelazione spaziale positiva dei dati e $I \rightarrow 1$. Al contrario, una prossimità spaziale di valori dissimili indica una autocorrelazione spaziale negativa (o eterogeneità spaziale), con $I \rightarrow -1$. Un valore pari a 0, invece, indica un pattern spaziale non diverso da un fenomeno casuale. Il numeratore nell' eq. (4.9) si interpreta come la covarianza tra unità contigue. L'indice di I di Moran per le variabili in esame è riportato in tab. (4.3).

	I di Moran	p.value
$\ln Rn_{mod4}$	0.036	0.000
$\ln Rn_{mod3}$	0.033	0.000
$\ln Rn_{allRan}$	0.034	0.000
ln perm	0.247	0.000
ln T	0.432	0.000
ln eU	0.361	0.000

Tabella 4.3: Indice I di Moran e p-value (con l' ipotesi nulla di distribuzione casuale), calcolato in **R** con la funzione *Moran.I()* del package *ape*.

In accordo con quanto previsto, l'indice I di Moran per i parametri è positivo evidenziando la presenza di autocorrelazione spaziale, in particolar modo per le variabili esplicative di permeabilità, concentrazione di uranio equivalente e temperatura.

Occorre considerare, a questo proposito, che le misure, su cui basiamo la nostra analisi spaziale, provengono dai centri abitati sparsi non omogeneamente sul territorio in analisi. Inoltre la densità dei dati spaziali è proporzionale alla densità delle abitazioni. Questo comporta una clusterizzazione dei dati che ha delle importanti conseguenze sul processo di regressione spaziale, principalmente

dovute alla larga preponderanza nel dataset di misure proveniente dalle 3 conche intermontane più densamente popolate (L'Aquila, Avezzano, Sulmona), caratterizzate da una ridotta varietà di litotipi e di valori di permeabilità. Per ridurre tale fenomeno è preferibile attuare una semplice strategia di declusterizzazione, consistente nell'aggregare le misure in base all'appartenenza a blocchi quadrati di lato pari a 2 km. Ciò comporta, pertanto, la sostituzione del dataset originale con un dataset declusterizzato i cui dati sono ricavati dai precedenti associando al centroide di ogni cella, per ogni variabile coinvolta nella regressione, le medie geometriche dei dati ricadenti nel perimetro della cella stessa².

In tab. (4.4) vengono riportate le statistiche riassuntive dei modelli di GWR. Abbiamo utilizzato una funzione peso di tipo gaussiana con un bandwidth fisso. La miglior ampiezza di banda (b) è stimata in base al criterio AIC , cioè scegliendo il valore che permette di ottenere l' AIC più basso.

La variabile dipendente considerata è la concentrazione di radon standardizzata per le caratteristiche dell'edificio, ottenuta con l'analisi multilevel bayesiana. Quindi, come visto nel capitolo precedente, $lnRn_{mod4}$ è il dato di concentrazione media annua di radon misurato in un'abitazione, a piano terra, costruita dopo il 1960 e utilizzando laterizio e cemento. Invece, $lnRn_{mod3}$ è analoga alla precedente ma senza considerare il materiale da costruzione come potenziale predittore della misura di gas radon. Entrambi questi modelli comunque considerano il raggruppamento per comune come unico effetto random sulla distribuzione. Al contrario, il $lnRn_{allRan}$ è la variabile standardizzata simile a quella ottenuta con il $mod4$, quindi abbiamo ancora quattro regressori, ma ora questi variano per ogni raggruppamento, quindi vengono considerati tutti come effetti random (si veda il par.[3.8]).

²Aggregando i dati campionati si perde ovviamente il livello di risoluzione che abbiamo con i singoli dati. Inoltre, le possibili modalità di aggregazione sono varie e potenzialmente arbitrarie, rendendo il risultato diverso a seconda della scelta fatta. Questo problema è conosciuto in geostatistica come COSP (change of support problem) (Gotway e Young, 2002). Ulteriori approfondimenti e analisi sono comunque oltre lo scopo di questo lavoro di tesi. Ci limitiamo alla scelta di unità aggregate non molto estese, diminuendo per quanto possibile il rischio di distorsioni.

Capitolo 4. Geographically Weighted Regression e mappa del potenziale radon geogenico

variabile dip.	OLS/GWR	R^2	R_{adj}^2	AIC	AIC_c	RSS	bandwidth
$lnRn_{mod4}$	OLS	0.046	0.033	628.602	628.808	139.082	
	gwr	0.177	0.106	611.488	614.076	119.950	16686.308
	gwrMIX	0.177	0.113	609.185	611.449	119.984	14917.653
$lnRn_{mod3}$	ols	0.043	0.030	622.928	623.133	136.459	
	gwr	0.142	0.080	611.891	613.772	122.326	18872.486
	gwrMIX	0.146	0.089	609.109	610.837	121.729	16686.308
$lnRn_{allRan}$	ols	0.035	0.022	610.142	610.348	130.728	
	gwr	0.128	0.065	601.470	603.351	118.122	18872.486
	gwrMIX	0.136	0.075	598.934	600.841	117.035	16010.742

Tabella 4.4: Riassunto delle caratteristiche di fitting per alcuni dei modelli di regressioni analizzati (celle 2x2 km). Il bandwidth è espresso in metri.

I modelli di GWR mostrano un miglioramento costante del fit di dati rispetto alla regressione multivariata lineare (OLS), visibile con la riduzione delle diagnostiche test per la misura della bontà del fit: AIC , l' AIC_c e RSS (somma dei quadrati dei residui). Il coefficiente di determinazione R^2 e la versione corretta R_{adj}^2 , che tiene conto del numero di variabili indipendenti, misurano la proporzione di varianza totale spiegata dalla variabile dipendente attraverso l'uso del modello. Questi coefficienti aumentano sostanzialmente per la GWR e i valori più alti si riscontrano per i modelli con variabile dipendente $lnRn_{mod4}$. In accordo anche con quanto trovato dall'analisi multilevel (par. [3.8]), scegliamo questa variabile $lnRn_{mod4}$, come la migliore per la descrizione del fenomeno, infatti si otteneva il valore più basso di DIC (deviance information criterion). Uno sviluppo futuro sarà la validazione del modello attraverso una cross-validation, in grado di definire oggettivamente quale variabile sia più adeguata per descrivere il set di dati.

La regressione lineare multivariata per il $lnRn_{mod4}$ (tab. 4.5), con tre variabili predittive, mostra che la concentrazione di radon indoor è correlata positivamente al contenuto di uranio equivalente nel suolo e al valore della permeabilità, mentre il coefficiente di regressione è negativo per la temperatura. I risultati

confermano la relazione diretta tra concentrazione di radon e uranio nel terreno e nelle rocce e la grande influenza della permeabilità sulla variabile. La temperatura ha ovviamente un coefficiente regressivo negativo, dimostrando la diminuzione del gas indoor nelle zone con temperature medie più alte.

OLS	Variabile	Stima	Stan. Error	t(Stima/SE)	Pr(> t)
$\ln Rn_{mod4}$	β_0 (Intercept)	5.233	0.616	8.500	0.0000
	β_1 (ln eU)	0.235	0.122	1.928	0.0548
	β_2 (ln T)	-0.509	0.249	-2.048	0.0415
	β_3 (ln perm)	0.017	0.006	2.786	0.0057

Tabella 4.5: La stima dei parametri con la regressione globale (OLS) per la variabile $\ln Rn_{mod4}$ (dati aggregati in celle 2x2 km).

In tab.(4.4) il modello “gwrMIX” si riferisce ad un modello misto di GWR il cui il logaritmo della permeabilità è assunto quale regressore globale. Questa scelta è avvalorata dal risultato del *DIFF of criterion*. La tab.(4.6) mostra i risultati per i fit della variabile $\ln Rn_{mod4}$ (risultati del tutto analoghi si hanno anche per $\ln Rn_{mod3}$ e $\ln Rn_{allRan}$).

Variabile	F	DOF	DIFF of Criterion
β_0 (Intercept)	2.680	3.441	-2.757
β_1 (ln eU)	5.117	3.073	-10.108
β_2 (ln T)	5.291	3.287	-11.352
β_3 (ln perm)	1.318	3.382	2.066

Tabella 4.6: Diff of Criterion per la GWR con le tre variabile esplicative locali. La variabile dipendente è il $\ln Rn_{mod4}$.

Il risultato positivo del *DIFF of Criterion* per il coefficiente della permeabilità, porta a presupporre che la descrizione del fenomeno migliori se consideriamo il $\ln perm$ come una variabile globale, facendo questa scelta, infatti, le diagnostiche di fit migliorano (tab.4.4).

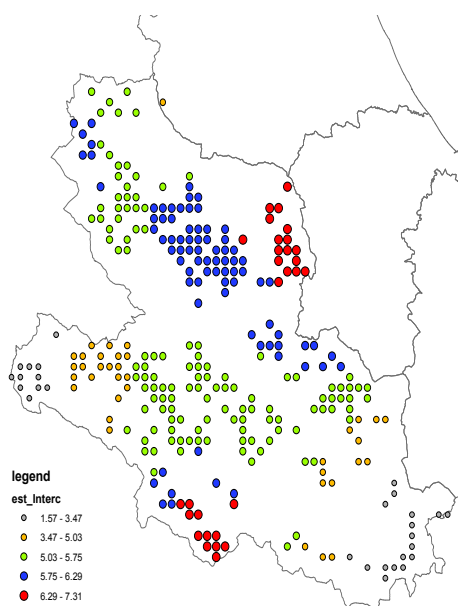
Variable	Estimate	Standard Error	t(Estimate/SE)
β_3 (ln perm)	0.020	0.006	3.083

Tabella 4.7: Stima per il coefficiente di ln perm globale.

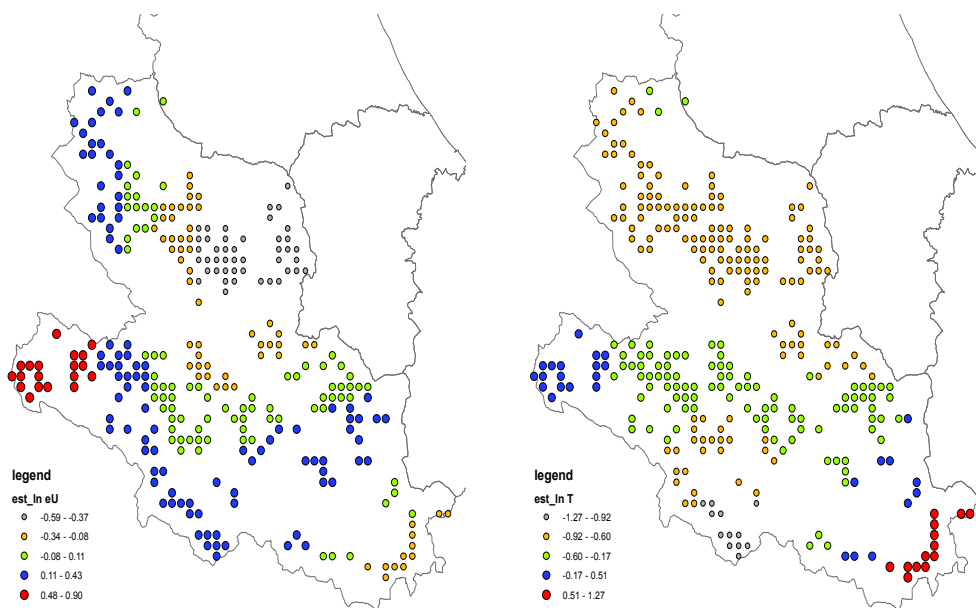
Variabile	Media	Min	Max	Mediana	LwrQuantile	UprQuantile
β_0 (Intercept)	5.338	1.565	7.308	5.612	5.184	5.937
β_1 (ln eU)	0.004	-0.592	0.901	0.072	-0.204	0.141
β_2 (ln T)	-0.500	-1.268	1.273	-0.625	-0.720	-0.481

Tabella 4.8: Stime per i coefficienti locali per il gwrMIX con il β_3 (ln perm) globale.

La distribuzione spaziale delle relazioni tra radon indoor e variabili predittive può essere mappata. Le fig.(4.8) seguenti mostrano come variano geograficamente i coefficienti locali. I valori cui fanno riferimento i diversi colori usati nelle mappe, sono definiti in base ai quantili delle corrispondenti distribuzioni.



a)

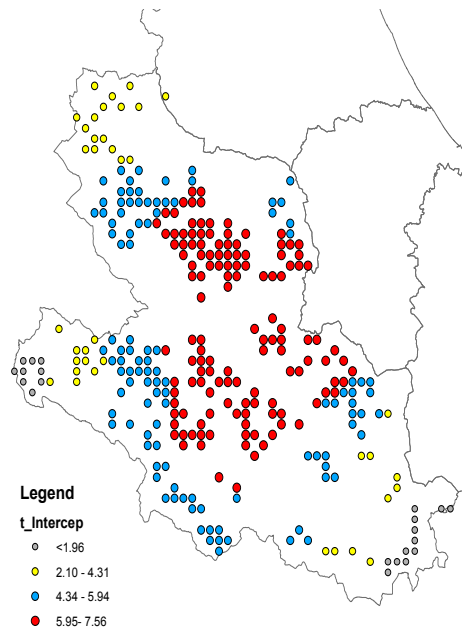


b)

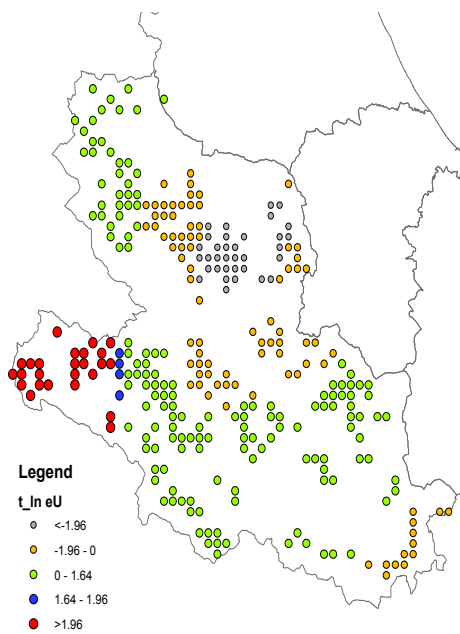
c)

Figura 4.8: Coefficienti locali per il modello gwrMIX per $\ln Rn_{mod4}$, rispettivamente a) β_0 (intercetta); b) β_1 ($\ln eU$) e c) β_2 ($\ln T$).

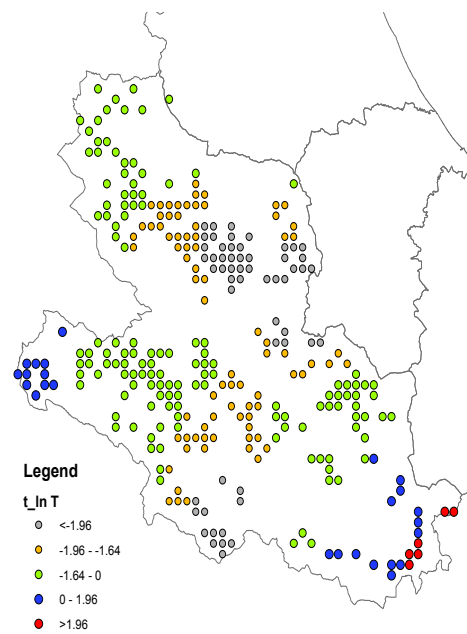
Ulteriori informazioni possono essere acquisite mappando le statistiche t , ottenute dividendo ogni stima locale del coefficiente di regressione con il corrispondente errore standard. L'indagine sulla statistica t permette di analizzare le relazioni a livello locale, indicando la significatività di ciascun parametro in ogni punto, e fornendo un'indicazione generale sulle aree geografiche in cui il modello sembra perdere efficacia. Nelle mappe vengono indicati gli intervalli per i quali il $|t|$ è maggiore di 1.96 e $1.64 < |t| < 1.96$, che corrispondono rispettivamente ad una probabilità di errore del 5% e del 10% di rifiutare l'ipotesi nulla ($\beta=0$) quando è vera.



d)



e)



f)

Figura 4.9: Statistiche t locali rispettivamente per: d) β_0 (intercetta); e) β_1 ($\ln eU$) e f) β_2 ($\ln T$).

I coefficienti di regressione relativi al contenuto di uranio equivalente (β_1) e alla temperatura media annua (β_2) mostrano un'importante variabilità spaziale. Il coefficiente β_1 è significativo ($|t| > 1.96$) solo nella parte ovest della regione (spot rossi in fig. (4.8b)) e nell'area a sud est della città di L'Aquila (spot grigi). Nella prima zona il coefficiente di regressione è positivo e varia da 0.48 a 0.90, qui il contenuto di uranio è elevato [4.5;5.5]ppm e corrisponde infatti a un'area con concentrazioni elevate di radon (fig.4.5). Al contrario, nella zona est dell'aquilano e nella parte centrale della regione (fig. 4.8b), il coefficiente regressivo negativo [-0.59; -0.33] rappresentando un' anomalia nella relazione diretta tra concentrazione di radon indoor e contenuto di uranio nel suolo. Nel fig(4.10) anche il R^2 locale in quest'area è molto basso [0.08-0.10] evidenziando l'inadeguatezza del nostro modello nella descrizione del fenomeno.

Il coefficiente β_2 per la temperatura è significativo (fig.4.9f) nella parte interna della regione con livelli di significatività del 5% e del 10%. Come accade però per l'uranio, abbiamo anche in questo caso delle irregolarità nella descrizione del fenomeno fisico da parte del modello: nella zona più ad ovest e nella parte a Sud-Ovest (fig. 4.8c) dell'area, il coefficiente regressivo è positivo.

In fine, riportiamo in fig.(4.10) la variazione spaziale del coefficiente di determinazione locale R^2 , che, come già ricordato, è una grandezza caratteristica della bontà del fit. I risultati ottenuti mostrano una descrizione della variabile dipendente abbastanza soddisfacente ($R^2 = 0.20$) per la zona al confine sud-est della provincia (Val di Sangro) e nella parte più occidentale della Marsica ($R^2 = 0.27$). La variabilità spiegata dal nostro modello, oltre ad essere indubbiamente maggiore rispetto a quella derivante dall'applicazione di una regressione multivariata globale, risulta superiore anche a quella ottenuta da Nissi et al. 2012 e Pasculli et al. 2013 (per entrambi R^2 locale medio= 0.11; mentre ora R_{new}^2 locale medio= 0.18).

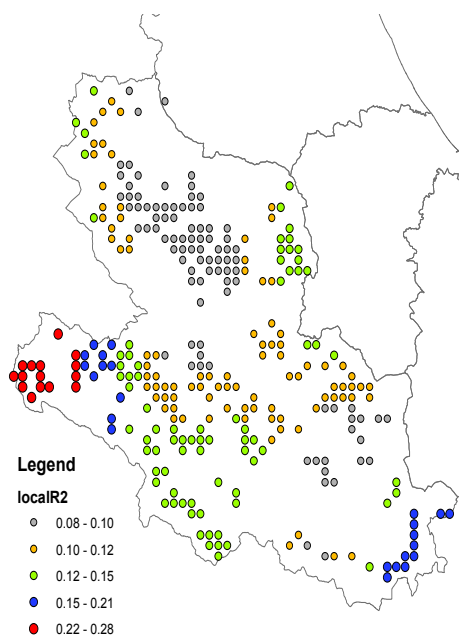


Figura 4.10: Coefficiente di determinazione locale R^2 .

4.3 Il potenziale radon geogenico

Il potenziale radon geogenico (PRG) è una quantità rappresentativa del rischio radon, connessa con le caratteristiche ambientali dell'area considerata. Al contrario della semplice mappatura delle concentrazioni del gas indoor rilevate nell'ambito di campagne di misura, il PRG fornisce una stima più vicina a “ciò che la terra emette”, dipendente unicamente dall'ambiente geologico. La concentrazione di radon indoor, infatti, è fortemente influenzata, come abbiamo evidenziato con l'analisi multilevel, da fattori naturali e antropici come il tipo di costruzione, i materiali utilizzati, le abitudini di vita, il clima, ed è temporalmente variabile e caratteristica di un particolare edificio.

L'idea di creare una mappa del potenziale radon geogenico è intimamente connessa con il concetto di radon-prone area. Una corretta definizione del potenziale radon potrebbe fornire una ragionevole guida per l'identificazione di queste zone, in particolare quando il numero e/o la qualità delle misure speri-

mentali di radon indoor è inadeguato.

In accordo con quanto proposto dalla nuova raccomandazione Euratom Basic Safety Standards Directive (BSS; EC 2011), ma già espresso, per l'Italia, nel Dec.Leg. n.241 del 25/05/2000, tutti i Paesi membri devono individuare eventuali zone caratterizzate da elevate concentrazioni di radon negli edifici, stabilire dei limiti di riferimento nazionali e programmare un piano di azione per la salvaguardia della salute pubblica (Art. 103, EC, 2011).

In particolare, gli Stati devono stabilire dei livelli di riferimento nazionali che non superino (come media annua) i 200 Bq m^{-3} per le nuove abitazioni e le nuove costruzioni con accesso pubblico, 300 Bq m^{-3} per le abitazioni esistenti, e 300 Bq m^{-3} per edifici pubblici con accesso pubblico, permettendo dei livelli di azioni massimi di 1000 Bq m^{-3} solo per ambienti caratterizzati da basso tempo di occupazione.

In attesa che il BSS diventi una legge Europea, i vari Paesi membri hanno avviato discussioni di metodo per la mappatura di una grandezza rappresentativa del rischio radon per la popolazione.

Tra i progetti intrapresi, molti si basano sulla misura sistematica della concentrazione di radon indoor su un campione di abitazioni la cui distribuzione territoriale sia la più omogenea possibile. I dati raccolti vengono pre-processati in modo da definire una nuova variabile da mappare, il potenziale radon geogenico, riferita ad una situazione standard e caratteristica del rischio radon dovuto a sorgenti naturali (Friedmann 2005, Hunter et al 2009). Altri tipi di studi invece hanno basato l'individuazione di tali aree utilizzando le sole informazioni geologiche, analizzando la capacità di produzione e di trasporto di radon delle unità geologiche che costituiscono il territorio (Gundersen and Schumann, 1996; Orlando et al., 2000; Kemski et al., 2001; Ielsch et al., 2010).

Il nostro approccio prevede di combinare le misure di concentrazione di radon indoor con i dati geologici e, in generale ambientali, disponibili. Una volta infatti stabilita l'influenza di questi sulla concentrazione misurata, possiamo stimare il potenziale radon geogenico in ogni nuovo punto, purché si conoscano i valori delle

variabili predittive. La forza di questa nuova metodologia è nel fatto di riuscire a combinare più informazioni utili per la definizione del potenziale, evitando la necessità di avere misure di concentrazione per ogni centro abitato. Inoltre si tiene fortemente in considerazione l'aspetto spaziale del fenomeno e la non stazionarietà della relazione che lega le variabili, per un'analisi quanto più precisa e realistica.

4.3.1 Mappatura per la Provincia dell'Aquila

Sfruttando le potenzialità offerte da un GIS nel quale è stata implementata la Carta Tecnica Regionale³, abbiamo individuato i centri abitati localizzati all'interno di celle 2x2 km (appartenenti alla stessa griglia utilizzata in precedenza per l'analisi GWR) prive di misure di radon. Abbiamo associato un valore di temperatura media annua (ricavato con la tecnica di interpolazione descritta al par [A.5]), un valore di uranio equivalente (ricavato con l'applicazione del block kriging) ed un valore medio di permeabilità stimato, ove disponibile, sulla base del lavoro di Palermi e Desiderio ovvero della mappa elaborata in Pasculli et al. 2013 (vedi fig.4.3).

Il modello GWR calibrato con il nostro dataset, sarà in grado di predire il valore della variabile dipendente (concentrazione di radon indoor in un'abitazione standard) in nuovo punto nell'area di studio, una volta definiti i valori delle variabili indipendenti.

Il software GWR4.0 permette di determinare i coefficienti locali della regressione locali per ognuno dei nuovi punti. Consideriamo il set di osservazioni per le variabili indipendenti nel nuovo punto p_0 , $x_0 = (x_{01}, x_{02}, \dots, x_{0p})$. Il valore stimato per y_0 (che è il valore vero di concentrazione standardizzata di radon in p_0) è:

$$\hat{y}_0 = x_0^T \hat{\beta}(p_0) \quad (4.10)$$

³link: <http://www.regione.abruzzo.it/xcartografia/>

dove

$$\hat{\beta}(p_0) = [X^T W(p_0) X]^{-1} X^T W(p_0) y \quad (4.11)$$

e $W(p_0)$ è noto perché i parametri che costituiscono la matrice sono stati determinati nel processo di calibrazione del modello.

In output troviamo anche il t-value e l'errore standard associato a ciascun coefficiente stimato, quindi siamo in grado di stabilire quali parametri siano significativi in quella cella e quale sia la precisione associata ad essi.

Abbiamo 219 nuove celle, quindi per un totale di 517 punti sommando le 298 celle usate per la calibrazione del modello. Il modello utilizzato è il gwrMiX con la variabile dipendente $\ln Rn_{mod4}$, descritto nella sezione precedente. Il riassunto delle statistiche viene riportato in tab.(4.9).

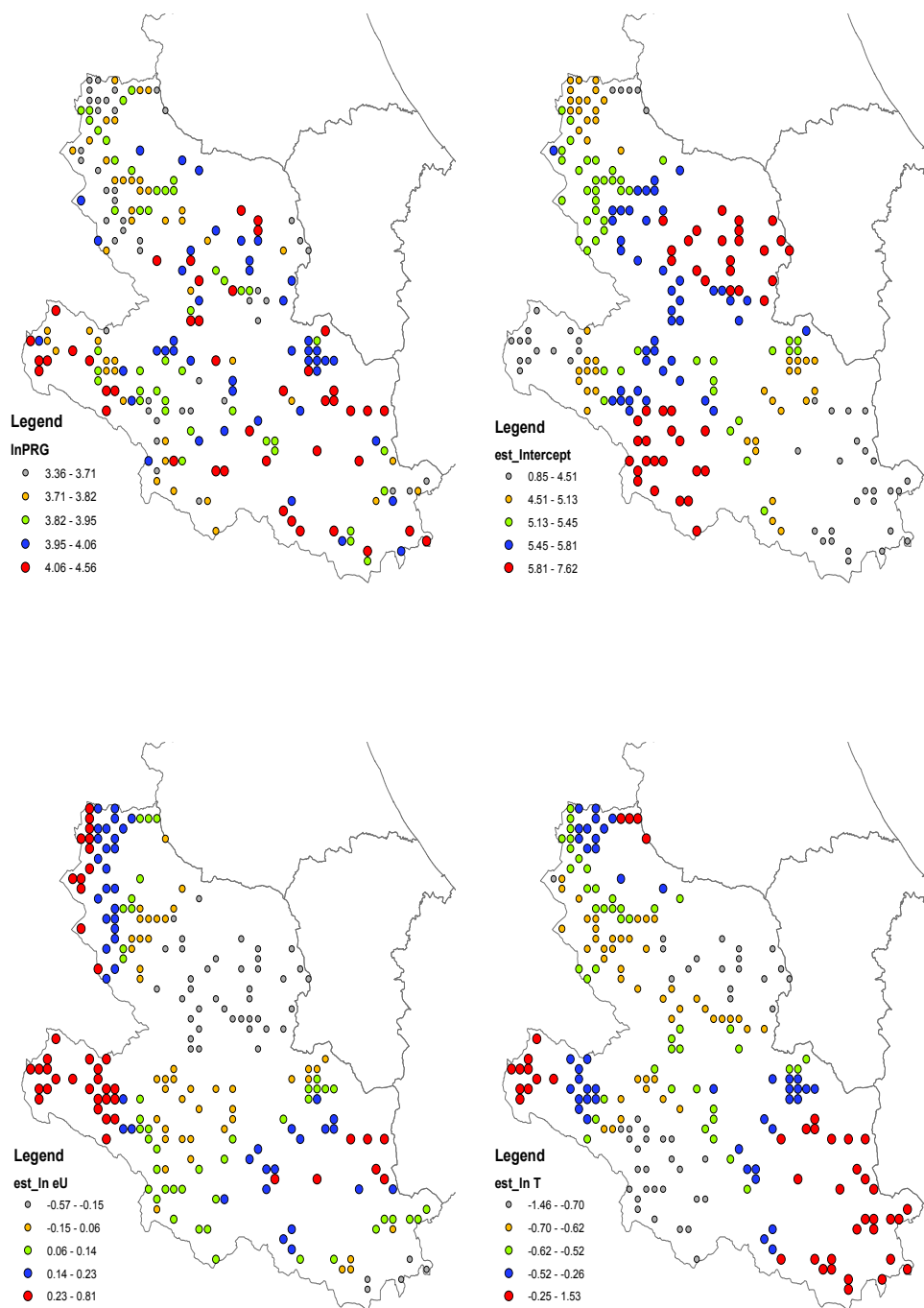


Figura 4.11: Stime per i 219 nuovi punti rispettivamente per: il potenziale radon geogenico (log-trasformato) in Bq m^{-3} ; coefficienti di regressione per l'intercetta, contenuto di uranio e temperatura. Intervalli stabiliti con i quantili delle varie distribuzioni.

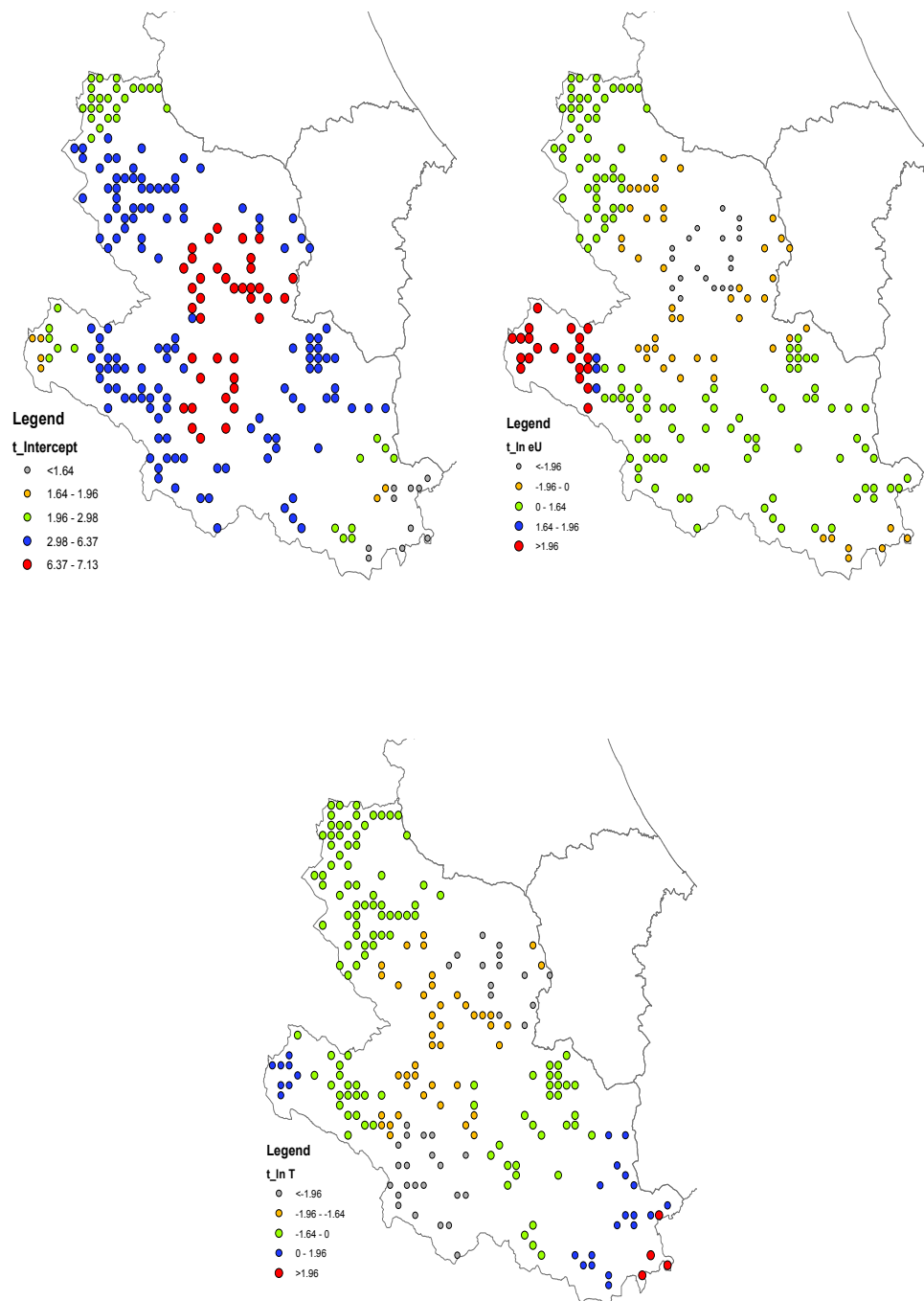


Figura 4.12: t – value dei coefficienti di regressioni stimati nei 219 nuovi punti di griglia.

	n	AM	SD	GM	GSD	MAX	MIN
\hat{y}_{gwrMIX}	298	58.40	14.68	56.86	1.25	128.48	31.26
$\hat{y}_{newPoint}$	219	50.39	11.64	49.18	1.24	95.22	28.93
tot	517	55.01	14.03	53.47	1.26	128.48	28.93

Tabella 4.9: Riassunto statistiche per il potenziale radon geogenico stimato. \hat{y}_{gwrMIX} è la concentrazione di radon Rn_{mod4} stimata con modello gwrMIX (si veda [4.2]); $\hat{y}_{newPoint}$ è il potenziale radon stimato nei nuovi punti. Valori in Bq m⁻³.

Anche in questo caso ritroviamo delle anomalie nella descrizione con il nostro modello. I coefficienti di regressione sia per l' uranio che per la temperatura assumono, in alcune zone, un segno (negativo per l'uranio equivalente e positivo per le temperature) opposto a quanto suggerito dalla realtà dei fenomeni fisici cui si riferiscono. Il modello evidentemente è inadeguato in queste aree ed è necessario un approfondimento ulteriore specialmente riguardante la complessa struttura geologica che le caratterizza.

Mettendo insieme quanto ottenuto fino ad ora, otteniamo la prima mappa del potenziale radon geogenico per la provincia dell' Aquila.

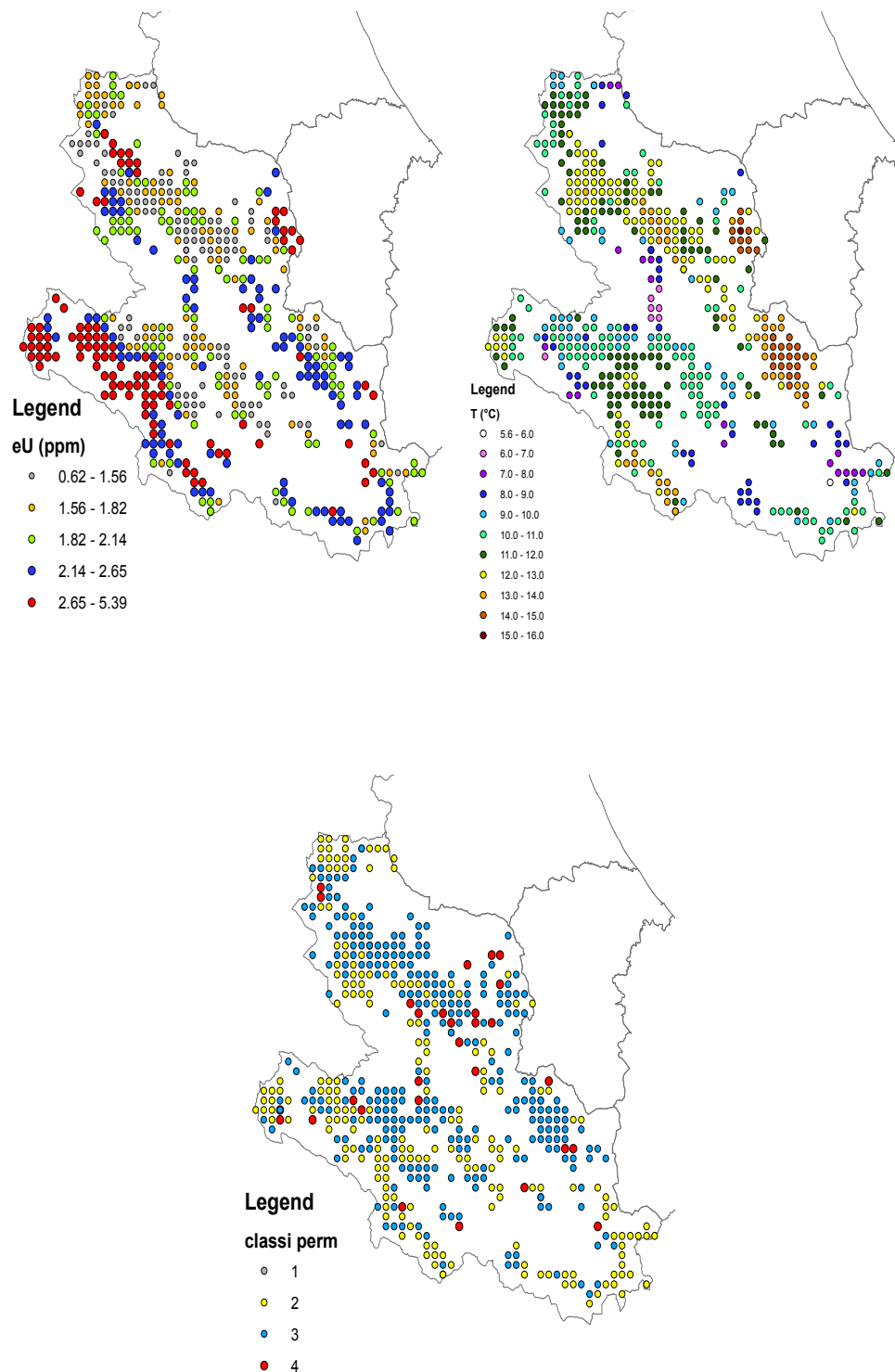


Figura 4.13: Valori delle variabili predittive georeferenziate. Classi di permeabilità: 1: bassa; 2: media; 3: alta; 4: altissima.

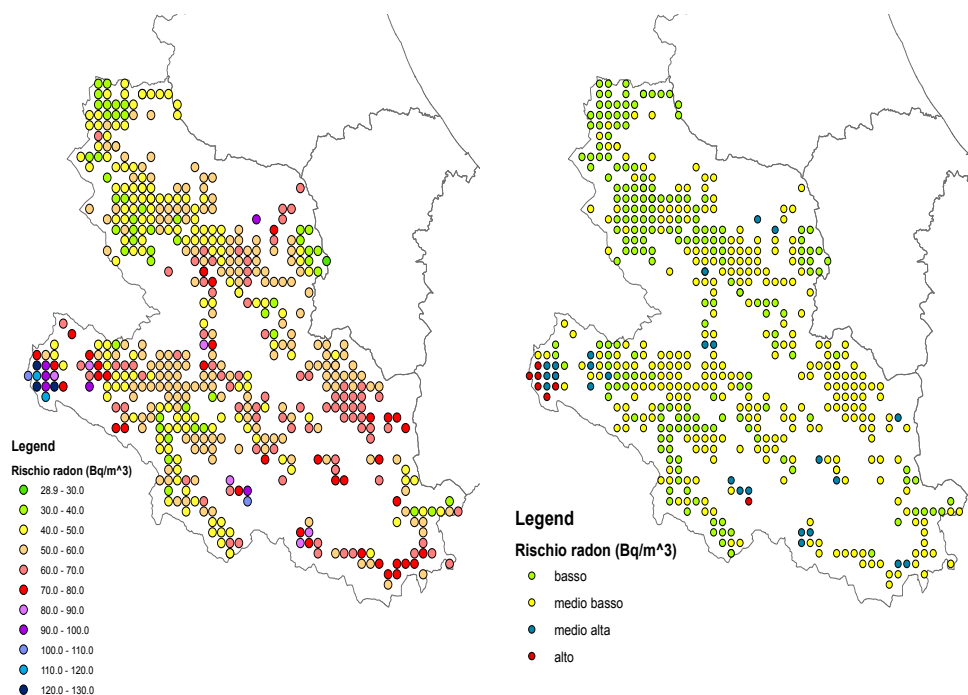


Figura 4.14: Mappa del rischio radon per la provincia dell' Aquila.

I valori stimati di concentrazione di radon indoor (Rn) per un edificio standard (potenziale radon) sono stati raccolti in quattro classi differenti di rischio radon:

- basso: $\leq 50 \text{ Bq m}^{-3}$;
- medio basso: $50 < Rn \leq 75 \text{ Bq m}^{-3}$;
- medio alto: $75 < Rn \leq 100 \text{ Bq m}^{-3}$;
- alto: $Rn > 100 \text{ Bq m}^{-3}$.

Nella mappa (fig. 4.14) sono prevalenti le zone a bassa e medio-bassa concentrazione. Solo l'area più ad ovest delle Marsica e una piccola zona tra i comuni di Collelongo e Villavallelonga presentano concentrazioni elevate superiori ai 100 Bq m^{-3} .

Nella parte centrale e a sud della provincia, il potenziale radon aumenta, fino a raggiungere valori massimi nella Marsica. Questa infatti è l'area a maggiore

concentrazione di radon a cui corrisponde anche un elevato contenuto di uranio nel suolo (fino a 5.5 ppm per eU in prossimità di Oricola (AQ)). E' da sottolineare che il R^2 locale per il modello gwrMIX raggiunge i valori massimi proprio in quest'area.

Nella zona centrale, nei pressi di Villavallelonga, ritroviamo una possibile area a rischio radon, infatti oltre ad esserci una concentrazione di uranio elevata fig.(4.13), la temperatura media è bassa e la permeabilità è elevata, tutte condizioni considerate favorevoli.

Per quanto ottenuto in questa analisi, la concentrazione media annua di radon indoor in un edificio standard nell'area aquilana è 55 Bq m^{-3} , con deviazione standard di 14 Bq m^{-3} , ben al di sotto dei limiti di azione indicati nelle normative nazionali e internazionali.

Data l'estrema variabilità dei livelli di radon, non si può comunque escludere la possibilità di osservare, attraverso una misurazione diretta della concentrazione di radon, livelli alti in aree caratterizzate da un potenziale basso e viceversa. Questa mappa fornisce solo la tendenza generale della variazione del potenziale radon geogenico e non è adatta per l'estrapolazione diretta della stima di concentrazione di radon indoor per uno specifico edificio.

4.3.2 Considerazioni finali

Il lavoro di tesi ha permesso di individuare le aree a elevata concentrazione di radon per il territorio della Provincia dell'Aquila. L'estensione della classificazione delle radon-prone areas a tutta la regione, servendosi di questo nuovo approccio statistico, può essere utilizzata dall'ARTA e dalla Regione Abruzzo per ottemperare agli obblighi di legge (D. Lgs 241/2000).

L'incertezza da associare alla previsione locale del potenziale radon è di difficile definizione. L'analisi comporterebbe il calcolo dell'intervallo di confidenza locale da associare alla stima, sulla base degli errori standard di ogni variabile indipendente. Sarebbe comunque opportuno applicare un approccio Monte Carlo o tecniche di campionamento bootstrap che consentano di ricavare una distribu-

zione di stime della regressione GWR, da cui ricavare l'intervallo di confidenza, tuttavia questo va oltre gli scopi del lavoro di tesi e per il momento tralasciamo questo aspetto.

La valutazione della correttezza o meno della nostra analisi statistica è stata fatta utilizzando diagnostiche test, essenzialmente derivate dai residui, cioè dalla differenza tra i valori stimati (calcolati con il modello) e quelli misurati. Tuttavia, queste diagnostiche possono essere fortemente influenzate da problemi di over-fitting, rendendo spesso la valutazione ottimistica (l'incertezza associata alla predizione è inferiore a quella reale). Un modo per affrontare queste possibili distorsioni è utilizzare una tecnica di *cross validation*, che permetterebbe di conoscere effettivamente la capacità del modello di predire il valore vero. Purtroppo, al momento non sono disponibili librerie, sia per l'analisi multilevel che per la GWR, che permettono di implementare rapidamente questo tipo di strumenti. Uno sviluppo futuro del lavoro prevede questo controllo.

Conclusioni

La Regione Abruzzo, per ottemperare agli obblighi di legge (Dec. Leg. 241/2000), ha affidato all'ARTA il compito di individuare eventuali zone ad elevata concentrazione di radon (radon-prone areas). Anche la Comunità Europea dell' Energia Atomica (Euratom) sta per pubblicare una nuova raccomandazione, che diventerà legge europea, che richiama l'obbligo per gli Stati membri di individuare tali aree per la salvaguardia della salute pubblica (EC, 2011). Questo lavoro di tesi si basa sulla convinzione che un'adeguata definizione di una quantità nota come potenziale radon geogenico, che ne assicuri in particolare un uso operativo, possa condurre ad una mappatura del territorio che si configuri quale utile strumento per l'attuazione delle politiche radon sia a livello nazionale che locale. Le autorità possono trarre vantaggio per la progettazione di campagne di misura del radon e per la stesura di piani di intervento sul patrimonio edilizio, sia a livello preventivo che correttivo.

Il potenziale radon geogenico è una grandezza fisica direttamente collegata alle caratteristiche geologiche e geofisiche del territorio, utilizzabile quale indicatore quantitativo del rischio radon per la popolazione. Nella definizione operativa da noi adottata, è una stima della concentrazione del gas presente in una specifica regione, dovuta alle sole proprietà geologiche dell'area in esame, indipendente da fattori antropici e temporalmente costante su tempi di scala geologici.

Per stimare questa nuova variabile abbiamo utilizzato un nuovo approccio modellistico che analizza la relazione tra le misure di concentrazione di radon indoor e i fattori geologici, in generale ambientali, dell'area in esame, tenendo conto, soprattutto, della variabilità spaziale del fenomeno.

La concentrazione di radon indoor è un fenomeno complesso ed è fortemente influenzata da fattori geologici (contenuto di uranio nel suolo, permeabilità, porosità, fratturazione, etc.) e da variabili più propriamente legate alle abitudini di chi vive all'interno dell'edificio e alle caratteristiche dell'edificio stesso (ventilazione, gradiente termico interno-esterno, tipologia di edificio, piano in cui viene fatta la misura etc.). Per individuare le aree in cui la concentrazione di radon può essere elevata, si è scelto di analizzare prima l'influenza delle caratteristiche dell'edificio, ottenendo così una variabile quanto più possibile filtrata delle condizioni specifiche in cui viene fatta la misura, e successivamente la relazione tra questa e le caratteristiche geologiche.

I dati di concentrazione radon indoor, pertanto, sono stati pre-processati con un'analisi multilevel bayesiana. Associando delle distribuzioni di probabilità ad ogni parametro dell'inferenza statistica, l'analisi bayesiana multilevel consente di stimare i parametri statistici di queste distribuzioni minimizzando l'effetto della ridotta dimensione del campione per ciascuna delle unità in cui raggruppiamo il nostro set di dati. Il modello multilevel a effetti misti tiene in considerazione opportunamente il numero diverso di misure tra un'unità geografica (i comuni) e l'altra. Il metodo di stima dei parametri (i coefficienti di regressione e le relative varianze) con simulazioni Monte Carlo basate sulle proprietà delle catene di Markov (MCMC), ha il vantaggio di considerare l'incertezza associata con la stima dei parametri random, fornendone una stima attendibile.

Le diagnostiche test, misurando la bontà del fit, hanno permesso di scegliere il modello migliore: un modello multilevel a effetti misti con quattro coefficienti di regressione fissi (il piano, la tipologia dell'edificio, il materiale da costruzione e l'anno di edificazione) ed un effetto random, dato dal comune di appartenenza dell'edificio al quale si riferisce la misura di radon.

Grazie agli indici di autocorrelazione spaziale e all'analisi GWR abbiamo mostrato, per il fenomeno in esame, la presenza di effetti localizzati e le differenze di associazione tra le variabili nello spazio. Si sono evidenziati chiari segni di correlazione spaziale nelle distribuzioni delle variabili geologiche e dei dati di concentrazione di radon indoor in tutta la zona della provincia dell'Aquila.

Le uniche variabili geologiche/ambientali significative per la regressione locale si sono rivelate essere il contenuto di uranio equivalente nel terreno, la permeabilità del suolo e la temperatura media annua dell'aria esterna. I risultati locali della regressione identificano aree dove i coefficienti sono negativi e altre dove sono positivi. Questo è un vantaggio rispetto ai risultati della regressione globale, che maschera l'eterogeneità geografica e le associazioni complesse che potrebbero esistere tra le variabili sul territorio, permettendoci di avere una descrizione accurata e puntuale del fenomeno in studio.

La GWR consente di stimare anche i coefficienti di regressione locali dove non abbiamo misure di concentrazione del gas indoor, così da poter predire il potenziale radon atteso, una volta noti i valori delle variabili predittive.

La mappa finale del potenziale radon geogenico, presentata in questo lavoro, fornisce una guida per l'individuazione delle radon-prone areas del territorio della Provincia dell' Aquila. In particolare è possibile valutare il rischio radon per le zone dove non si hanno o sono scarse le misure di concentrazione di radon negli edifici.

Questa mappatura, basandosi su tecniche avanzate di indagine statistica, è molto più accurata rispetto a quanto ottenuto in studi precedenti. Siamo in grado, infatti, di "esplorare" la non stazionarietà spaziale del fenomeno radon e della relazione tra questo e alcune variabili ambientali e geologiche.

Questo nuovo approccio modellistico vuole fornire anche un contributo al dibattito scientifico che si è sviluppato dal 2010 e tuttora in corso, per iniziativa del Joint Research Center della Commissione Europea, intorno alla definizione stessa di potenziale radon geogenico e la relativa mappatura in tutto il territorio europeo (Dubois et al 2010; Gruber et al, 2013).

La concentrazione media annua di radon indoor in un edificio standard stimata, in accordo con il potenziale geogenico medio del territorio della provincia dell'Aquila, è 55 Bq m^{-3} , valore che, rispetto ai limiti della nuova raccomandazione Euratom (200 Bq m^{-3} per gli edifici di nuova costruzione e 300 Bq m^{-3} per quelli già esistenti), può ritenersi bassa. Ma ovviamente, data la marcata variabilità del fenomeno, non è possibile escludere che in una certa percentuale di edifici la concentrazione di radon possa effettivamente superare i limiti di riferimento.

Gli sviluppi futuri di questo lavoro prevedono un'accurata validazione del modello utilizzato per la descrizione del fenomeno attraverso una cross validation leave-one-out sia per i risultati dell'analisi multilevel che per la stima del GWR. Attraverso tecniche di simulazione Monte Carlo potremo definire, inoltre, con maggior accuratezza anche l'incertezza di stima.

Analisi spaziale per la mappatura delle temperature medie

Avendo come riferimento il lavoro svolto da Hudson e Wackernagel (Hudson e Wackernagel, 1996), la temperatura media nella provincia dell'Aquila è stata mappata utilizzando l'approccio geostatistico del kriging con external drift. Nel capitolo inizialmente si descrive la tecnica geostatistica, i principi cardine e gli scopi, nella seconda parte si riportano i risultati della stima delle temperature medie annue.

La geostatistica si compone di strumenti statistici sia deterministici che probabilistici atti alla comprensione e alla modellizzazione di dati che manifestino un qualche tipo di correlazione spaziale (*variabili regionalizzate*).

La sua particolarità, rispetto alle altre discipline “classiche”, risiede nella possibilità di prendere in considerazione la variabilità spaziale sia a *corto* che a *lungo* raggio, ovvero di considerare allo stesso tempo rispettivamente la correlazione spaziale e l'eventuale presenza di un *trend* — andamento sistematico (aumento o diminuzione) delle misure della variabile sotto studio nello spazio, su distanze medio-lunghe in relazione alle dimensioni dell'intera area geografica esaminata (Cressie, 1991).

La geostatistica si configura come un insieme di tecniche matematiche e statistiche adatto a tutti quei settori in cui l'elemento determinante sia la *continuità spaziale* del fenomeno sotto esame; rispetto alla statistica classica, infatti, il maggior contributo è stato quello di implementare delle tecniche opportune per affrontare situazioni nelle quali risulti centrale la possibilità di considerare le due caratteristiche fondamentali dei dati geostatistici, che sono:

- la *non-ripetitività*: per ogni localizzazione campionata esiste infatti solitamente una sola osservazione;
- la *dipendenza*: in generale, valori osservati in localizzazioni differenti manifestano una qualche correlazione.

La geostatistica è quindi in grado di:

- costruire un modello (tipicamente di natura stocastica) che riproduce e descrive la variabilità spaziale riscontrata sperimentalmente dai campionamenti disponibili;
- utilizzare questo modello per stimare i valori della variabile di interesse in localizzazioni non ancora campionate, stabilendo anche l'entità dell'incertezza che accompagna tali stime.

La maggior parte delle proprietà spaziali che caratterizzano un processo, variano in modo così complesso che la descrizione non può essere definita in modo deterministico. Per far fronte a questa incertezza spaziale, è richiesto un approccio differente rispetto ai metodi dell'analisi spaziale deterministica tradizionale. La geostatistica moderna affronta questo problema mediante un approccio di tipo probabilistico (*processo stocastico*), che è in grado di stimare sia i valori incogniti della variabile di interesse sia l'accuratezza di tale stima.

Le variabili di tipo spaziale sono definite variabili regionalizzate (VR), per enfatizzare due aspetti fondamentali che le caratterizzano (Matheron, 1963):

- un aspetto casuale o aleatorio, che tiene conto delle variabilità locali del fenomeno;

- un aspetto strutturale o regionalizzato, che riflette le caratteristiche globali del fenomeno (*trend*).

Il modello fondamentale per una variabile regionalizzata $Z(x)$ osservata nella localizzazione x , dove x rappresenta il vettore, bi-dimensionale, di componenti spaziali $(u; v)$, viene scritto nel modo seguente:

$$Z(x) = m(x) + R(x)$$

questo modello probabilistico assume che la componente casuale sia descritta da $R(x)$ e quella strutturale da $m(x)$.

Questo implica quindi che in ogni localizzazione x_i , il valore osservato $z_k(x_i)$ della variabile campionata sia la realizzazione di una variabile aleatoria (VA) $Z_k(x_i)$, il cui valore di aspettazione (o valore atteso), indicato con $E[\cdot]$, vale proprio

$$E[Z_k(x_i)] = m(x_i)$$

In un generico punto x in cui la variabile regionalizzata non è stata campionata, il valore $z(x)$ non è noto, tuttavia risulta ben definito; può infatti venir interpretato come una realizzazione della VA $Z(x)$. L'intera famiglia delle variabili aleatorie $Z(x)$, con $u \in D$, D dominio/area di studio, prende il nome di funzione aleatoria (FA), o *processo stocastico*; tipicamente, la definizione di FA è limitata a VA collegate a uno stesso attributo, sia z , così che un'altra FA $W(x)$, $u \in D$ sarà definita per trattare un differente attributo w del fenomeno.

Punto fondamentale nel trattamento geostatistico di un set di dati georeferenziati è la definizione di uno strumento atto a descriverne e modellarne analiticamente la variabilità spaziale; lo strumento tradizionalmente utilizzato è il variogramma $\gamma(h)$, una funzione che descrive, al variare della distanza h tra due punti, quanto i valori della variabile di interesse, misurati nei due punti, siano scorrelati. In linea di principio, è senz'altro lecito supporre che due valori sperimentali della variabile siano tanto più simili tra loro, quanto più sono spa-

zialmente vicini i punti di misura. In altri termini, la correlazione spaziale tra due punti tende ad indebolirsi all'aumentare della distanza tra essi.

Determinato il modello di variogramma che meglio descrive la correlazione spaziale dei dati campionari, si può implementare un secondo strumento, che consentirà di stimare i valori assunti dalla variabile anche nei punti in corrispondenza dei quali non sono disponibili misure. Lo strumento in questione è il cosiddetto kriging, uno stimatore lineare che minimizza la varianza di stima senza introdurre errori sistematici; esso consente, in sostanza, di ricostruire i valori della variabile sui punti di una griglia regolare estesa su tutto il dominio di interesse, attraverso una media dei valori sperimentali pesati in ragione inversa della distanza, secondo il modello di variogramma individuato nell'analisi preliminare.

Autocorrelazione L'autocorrelazione definisce il grado di dipendenza spaziale tra i valori assunti da una variabile campionata. Una caratteristica intuitiva dell'ambiente è che le sue proprietà sono in relazione fra di loro in una qualche scala, grande o piccola che sia. Questa situazione è definita autocorrelazione, quindi valori campionati in luoghi vicini tra di loro, tendono ad avere comportamenti simili, mentre valori di una stessa variabile misurati in campioni raccolti in luoghi lontani tra di loro tendono ad avere comportamenti differenti, o almeno tendono a differire dai valori medi che si riscontrano nei due luoghi stessi. In tal senso, la correlazione fra i valori della variabile tende a diminuire con l'aumentare della distanza.

Modello stazionario e non-stazionario In statistica è comune assumere che un processo è *stazionario* o *stazionario al secondo ordine*, quando la sua distribuzione è invariante sotto traslazione. Allo stesso modo, una funzione random stazionaria è omogenea e ripetibile nello spazio. Un processo stocastico è detto *stazionario in senso stretto* se tutti i momenti sono invariati per traslazione. Usualmente però si richiede per la definizione di stazionarietà che solo i primi

2 momenti, cioè la media e la varianza, siano invariati sotto traslazione. Nel contesto spaziale è richiesto cioè che:

- il valore di aspettazione (o la media) della funzione $Z(x)$ sia costante in tutti i punti x . Questo significa che

$$E[Z(x)] = m(x) = m \tag{A.1}$$

è indipendente da x e h ;

- la covarianza¹ $Cov(x, x + h)$ tra due punti x e $x + h$ sia indipendente da x e funzione solo di h (lag), così che:

$$E[Z(x)Z(x + h)] - m^2 = C(h) \tag{A.2}$$

$C(h)$ è la funzione covarianza e esprime la correlazione tra le variabili $Z(x)$ e $Z(x + h)$ in funzione della distanza h .

In particolare quando $h = 0$, la covarianza torna ad essere la varianza di $Z(x)$ costante, detta anche *varianza a priori* $C(0)$. La correlazione tra le variabili $Z(x)$ e $Z(x + h)$ tende invece ad indebolirsi con l'aumento della distanza tra i due punti di appoggio e si annulla se due variabili diventano indipendenti. La distanza a cui ciò accade è detta *range*.

Le funzioni random *non stazionarie* sono quelle che non rispettano le condizioni (A.1) e (A.2). Più esplicitamente, sono quelle che soddisfano anche ad una delle due condizioni: · la media $E[Z(x)] = m(x)$ non è costante nel campo; · la funzione covarianza o non esiste (perchè non esiste la varianza) o non è invariante per traslazione.

¹In teoria della probabilità la covarianza di due variabili aleatorie è un numero $Cov(X, Y)$ che fornisce una misura di quanto le due varino assieme, ovvero della loro dipendenza. La covarianza di due variabili aleatorie X e Y è il valore atteso dei prodotti delle loro distanze dalla media:

$$Cov(X, Y) = E[(X - E[X])(Y - E[Y])]$$

A.1 Variogramma

Il variogramma è la funzione più comune della geostatistica, usata nelle applicazioni principalmente per caratterizzare la variabilità spaziale di un fenomeno regionalizzato. Sia S il dominio in cui è definita la funzione random $Z(x)$ e siano x_0 e $x_0 + h$ una coppia di punti appartenenti ad S e distanti h . La differenza tra $Z(x_0)$ e $Z(x_0 + h)$ definisce una nuova VA detta accrescimento o *incremento*:

$$[Z(x + h) - Z(x)]$$

La sua semi-varianza, è per definizione il *variogramma* (Matheron, 1965):

$$\gamma(x, h) = \frac{1}{2} \text{Var}[Z(x + h) - Z(x)] \quad (\text{A.3})$$

in generale funzione di h e del punto di appoggio x_0 .

Esaminiamo il comportamento della funzione (A.3) nel caso di modello stazionario e, scrivendo esplicitamente la varianza, otteniamo:

$$\gamma(x, h) = \frac{1}{2} \text{Var}[Z(x + h) - Z(x)] \quad (\text{A.4})$$

$$= \frac{1}{2} E\{[Z(x + h) - Z(x)]^2\} - E[Z(x + h) - Z(x)]^2 \quad (\text{A.5})$$

poiché per ipotesi la media m è invariata per traslazione si ha

$$E[Z(x + h) - Z(x)] = 0$$

e l'eq. (A.3) diventa:

$$\gamma(x, h) = \frac{1}{2} E\{[Z(x + h) - Z(x)]^2\} \quad (\text{A.6})$$

Ora, sviluppata la relazione si trova:

$$\gamma(x, h) = \frac{1}{2} \{Var[Z(x+h)] + Var[Z(x)] - 2Cov[Z(x+h), Z(x)]\} \quad (A.7)$$

$$\text{per la stazionarietà} \rightarrow \gamma(h) = \frac{1}{2} \{2Var[Z(x)] - 2C(h)\} \quad (A.8)$$

$$\rightarrow \gamma(h) = C(0) - C(h) \quad (A.9)$$

quindi dalla (A.9) la funzione variogramma è invariante per traslazione ed è strettamente legata alla funzione covarianza. Si noti che il limite superiore della $\gamma(h)$ è la varianza a priori $C(0)$, chiamata *sill*, infatti quando non c'è correlazione tra le sue misure $Z(x+h)$ e $Z(x)$, si ha:

$$\gamma(h) = \frac{1}{2} Var[Z(x+h) - Z(x)] \quad (A.10)$$

$$= \frac{1}{2} \{Var[Z(x+h)] + Var[Z(x)]\} = \sigma^2 \quad (A.11)$$

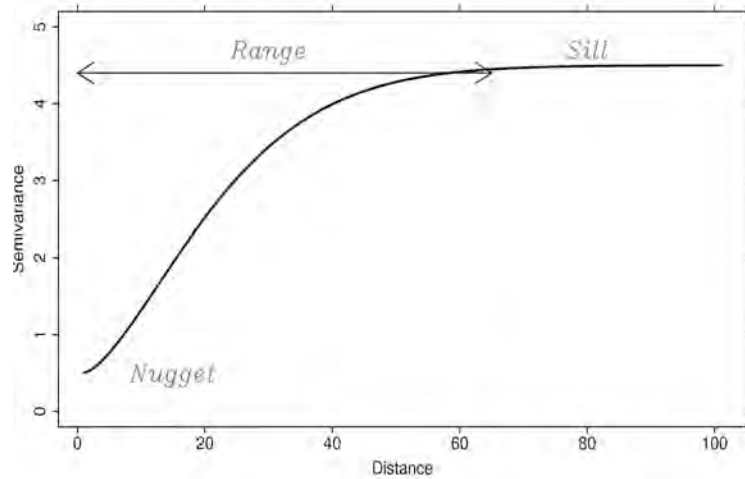


Figura A.1: Variogramma tipico

A.1.1 Variogramma sperimentale

Il variogramma sperimentale si stima dai dati, $z(x_1), z(x_2), \dots$ che sono la realizzazione della VA $Z(x)$, per mezzo della relazione per valori discreti:

$$\gamma(h) = \frac{1}{2m(h)} \sum_{i=1}^{m(h)} \{z(x_i) - z(x_i + h)\}^2 \quad (\text{A.12})$$

dove $z(x_i)$ e $z(x_i + h)$ sono i valori veri di Z nei punti spaziali (x_i) e $(x_i + h)$, e $m(h)$ è il numero di coppie a distanza h . Modificando h , si ottiene un altro set di semivarianze così da costruire il variogramma sperimentale. Generalmente si sceglie un lag e si calcolano le semivarianze per distanze multiple intere di questo intervallo.

Per un grigliato regolare, le semivarianze possono essere calcolate sulle righe e sulle colonne della griglia e l'incremento di lag è l'intervallo della stessa. Se le misure invece sono raccolte su una maglia irregolare, è allora necessario introdurre oltre alla distanza h un direzione angolare θ con una tolleranza $\Delta\theta$ ed una tolleranza Δh sulla distanza; così facendo tutte le coppie di campioni aventi distanza compresa tra $h - \Delta h$ e $h + \Delta h$ e allineate secondo una direzione compresa tra $\theta - \Delta\theta$ e $\theta + \Delta\theta$ contribuiscono al calcolo del variogramma.

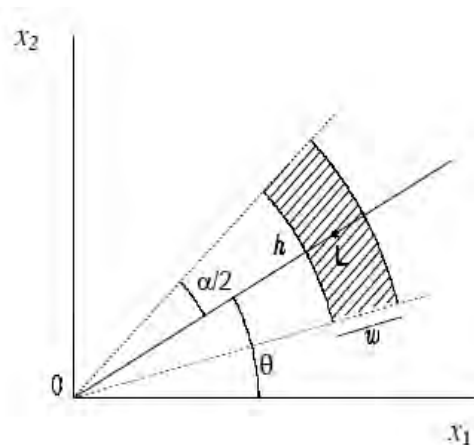


Figura A.2: Discretizzazione della distanza in bins per una maglia irregolare

In **R** la funzione variogramma è contenuta del pacchetto *gstat*, come vedremo

approfonditamente in seguito.

A.1.2 Caratteristiche del variogramma

- *Continuità.* La maggior parte delle variabili ambientali sono continue, quindi ci aspettiamo che la $\gamma(h)$ sia nulla per $h = 0$. In pratica però spesso a lag nulla corrisponde un'ordinata positiva, che rende la funzione discontinua. Questa discrepanza è detta *nugget effect* e rappresenta la necessità di modellare variazioni su piccola scala, o errori di misurazione, che possono influire sull'accuratezza della previsione.

La differenza tra il sill del modello ed il nugget rappresenta la proporzione della varianza totale che può essere attribuita alla dipendenza spaziale dei dati dalla griglia di campionamento utilizzata. Un'altra importante informazione fornita dal nugget si riferisce all'adeguatezza della griglia utilizzata per descrivere il fenomeno oggetto di studio.

- *Incremento monotono.* La semivarianza cresce con l'incremento della distanza (*lag*). Questo indica che a distanze piccole i valori di $Z(x)$ sono simili, ma a grande lag le differenze sulla media crescono. L'incremento monotono indica che il processo è spazialmente dipendente.
- *Sill e range.* Come già anticipato, il *sill* o sella è il limite superiore del variogramma quando la funzione è limitata (sempre nel caso stazionario di secondo ordine). La distanza alla quale il variogramma raggiunge la sella è il *range*. Per distanze maggiori, i punti sono spazialmente indipendenti.
- *Anisotropia.* La variazione spaziale potrebbe non essere la stessa in tutte le direzioni. Per esplorare l'anisotropia dei dati, il variogramma dovrebbe essere calcolato in almeno tre direzioni. Su una griglia regolare, di solito si calcola il variogramma sulle righe, colonne e sulle diagonali principali. Se il range dei variogrammi cambia con la direzione e una semplice trasformazione di coordinate la rimuove, allora si tratta di anisotropia geometri-

ca. Se cambia invece il sill in funzione della direzione, potrebbero essere presenti zone orientate con medie diverse ed è indicata come anisotropia zonale.

A.1.3 Fit del variogramma

Una volta che il variogramma sperimentale è stato calcolato, è necessario eseguire un fit con il metodo dei minimi quadrati dei valori sperimentali della semivarianza per ogni lag (distanza). Questo fit può essere fatto con un modello singolo o con una serie di modelli (variogrammi annidati). Facendo il fit dei dati sperimentali con una funzione matematica continua, si costruisce un modello della struttura di variabilità dell'intero dominio spaziale, non solo dei punti in cui abbiamo le misure. Tale informazione può essere utilizzata per il processo di stima. Le caratteristiche principali che la funzione di fit deve essere in grado di rappresentare sono:

- l'incremento monotono con l'aumento della distanza h ;
- il limite superiore (sill) o l'asintoto;
- l'intercetta positiva sull'ordinata (nugget);
- l'anisotropia.

Ci sono diverse funzioni utilizzate per il fit delle variabili ambientali, qui trattiamo solo le più comuni, cioè la funzione sferica e esponenziale. Le equazioni sono date per la forma isotropica con $h = |h|$. L'effetto nugget è indicato con c_0 e il partial sill c è la differenza tra sill (varianza a priori) e nugget.

Funzione sferica. E' la funzione comunemente usata per i modelli di fit. L'equazione è:

$$\gamma(h) = \begin{cases} c_0 + c \left\{ \frac{3h}{2a} + \frac{1}{2} \left(\frac{h}{a} \right)^3 \right\} & \text{per } h \leq a \\ c_0 + c & \text{per } h > a \\ 0 & \text{per } h = 0 \end{cases} \quad (\text{A.13})$$

Il parametro della distanza a indica il range, cioè la distanza massima oltre la quale non c'è correlazione tra i punti spaziali.

Funzione esponenziale.

$$\gamma(h) = c_0 + c \left\{ 1 - \exp\left(-\frac{h}{r}\right) \right\} \quad (\text{A.14})$$

Ora nel modello il parametro della distanza è r . Il sill è raggiunto più dolcemente del precedente modello e asintoticamente, pertanto il range è infinito. Per avere una misura della distanza entro cui si manifesta la correlazione è stato introdotto un range pratico a' , definito come la distanza alla quale viene raggiunto il 95% del sill e risulta $a' = 3a$.

A.2 La stima spaziale

Uno degli obiettivi principali dell'analisi spaziale è la produzione di stime spaziali, che permettono di passare dall'informazione discreta dei valori misurati alla descrizione continua del fenomeno nello spazio. Lo scopo è quello di ottenere una matrice regolare di informazione basata su una griglia regolare con un certo passo abbastanza fitto sulla quale costruire la rappresentazione della distribuzione spaziale stimata della variabile, quindi attribuire un valore alla variabile in un punto in cui essa non è nota. Lo strumento principale nel processo di stima è lo stimatore lineare. Il valore che si attribuisce cioè ad un punto del campo è calcolato mediante una combinazione lineare dei valori misurati nelle vicinanze del punto da stimare. Sia n il numero di punti noti che fanno parte del dominio che raccoglie tutti i valori utili alla stima in x_0 e $z(x_0)$ è la variabile incognita in x_0 , $z'(x_0)$ lo stimatore lineare avrà la forma:

$$z'(x_0) = \sum_{i=1}^n \lambda_i z(x_i) \quad (\text{A.15})$$

quindi il valore stimato è dato dalla media pesata di un certo numero di dati, infatti nella stima vengono utilizzati solo le misure fatte in punti che si trovano

in un intorno di x_0 , perché generalmente le misure più lontane influiscono difficilmente sul valore della stima.

Ponendoci in ambito geostatistico, definiamo $Z(x)$ come la funzione di distribuzione random assunta per interpretare in senso probabilistico il fenomeno in studio e siano $C(h)$ e $\gamma(h)$ le funzioni covarianza e variogramma. Indicando con $Z(x_0)$ e $Z(x_\alpha)$ le variabili aleatorie dei punti x_0 e x_α si ha per lo stimatore lineare:

$$Z^*(x_0) = \sum_{\alpha=1}^n \lambda_\alpha Z(x_\alpha) \quad (\text{A.16})$$

A questa è associato un *errore di stima* dato dalla differenza tra il valore vero e il valore stimato:

$$Z(x_0) - \sum_{i=1}^n \lambda_i Z(x_i) \quad (\text{A.17})$$

Lo stimatore deve inoltre avere la proprietà di essere *corretto*, cioè essere di media nulla, che equivale, in condizioni di stazionarietà, alla condizione sui predittori:

$$\sum_{i=1}^n \lambda_i = 1 \quad (\text{A.18})$$

La qualità della stima dipende dall'ampiezza degli errori di stima, essi sono caratterizzati da un distribuzione di probabilità a media nulla, e la varianza, detta *varianza di stima* σ_s^2 , è assunta quale grandezza per quantificare l'accuratezza della stima.

In funzione del variogramma $\gamma(h)$ si ricava:

$$\sigma_s^2 = \text{Var}[Z^*(x_0)] = E[(Z^*(x_0) - Z(x_0))^2] = \quad (\text{A.19})$$

$$= 2 \sum_i \lambda_i \gamma(x_i - x_0) - \sum_i \sum_j \lambda_i \lambda_j \gamma(x_i - x_j) \quad (\text{A.20})$$

con $\gamma(x_i - x_j)$ che è la semivarianza di Z tra i punti x_i e x_j e $\gamma(x_i, x_0)$ è la semivarianza tra l' i -esimo punto campione e il target x_0 . Le semivarianze sono derivate dal modello del variogramma che fitta i valori discreti a distanze limitate del variogramma sperimentale.

A.3 Il kriging

Il kriging è il *miglior* metodo di stima nell'analisi spaziale geostatistica, cioè è uno stimatore non polarizzato e la varianza dell'errore è minima. Spesso viene indicato come BLUP (Best Linear Unbiased Predictor), nel senso di Best (minimizzazione della varianza di stima), Linear (in quanto è un modello lineare) e Unbiased (la media di stima non è affetta da errori sistematici) Predictor.

A seconda della finalità della previsione e delle caratteristiche dei dati campionati, il kriging può essere fatto puntualmente o su aree di varia grandezza (block kriging). Esistono diversi tipi di kriging, la formulazione originaria è conosciuta come ordinary kriging (Journel and Huijbregts 1978) ed è il metodo più largamente utilizzato per i processi stazionari; per lo scopo di questa tesi utilizzeremo il kriging with external drift, cioè una variante dell'ordinary kriging per i dati che invece presentino una dipendenza da una variabile esterna.

A.3.1 Ordinary Kriging

L' *ordinary kriging* è il metodo più impiegato, è basato sull'assunzione di media sconosciuta e processo localmente stazionario. Per determinare i pesi che danno luogo alla stima migliore, cioè quelli che garantiscono una maggiore precisione, si minimizza la varianza di stima (A.20) con il metodo dei moltiplicatori di Lagrange:

$$\frac{\partial \sigma_s^2}{\partial \lambda_i} = 0 \text{ per } i = 1, \dots, n \quad (\text{A.21})$$

$$\text{con il vincolo: } \sum_{i=1}^n \lambda_i = 1 \quad (\text{A.22})$$

si ottiene:

$$\frac{\partial \sigma_s^2}{\partial \lambda_i} = \frac{\partial}{\partial \lambda_i} \left[2 \sum_i \lambda_i \gamma(x_i - x_0) - \sum_i \sum_j \gamma(x_i - x_j) + 2\mu(1 - \sum_i \lambda_i) \right] = 0 \quad (\text{A.23})$$

Svolgendo le derivate e utilizzando la condizione di correttezza (A.22) per determinare la $(n+1)$ -esima variabile μ (parametro di Lagrange), abbiamo:

$$\begin{cases} \sum_{j=1}^n \lambda_j \gamma(x_i - x_j) + \mu = \gamma(x_i - x_0) & \text{per ogni } i \\ \sum_{j=1}^n \lambda_j = 1 \end{cases} \quad (\text{A.24})$$

questo è un sistema lineare di *kriging* con $n + 1$ equazioni in $n + 1$ incognite e lo stimatore associato è detto *kriging*. Le incognite sono i pesi λ_i e il parametro μ . In forma matriciale, con $\gamma(x_i - x_0) = \gamma_{i0}$:

$$\begin{bmatrix} \gamma_{11} & \gamma_{12} & \dots & \gamma_{1i} & \dots & \gamma_{1n} & 1 \\ \gamma_{21} & \gamma_{22} & \dots & \gamma_{2i} & \dots & \gamma_{2n} & 1 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \gamma_{i1} & \gamma_{i2} & \dots & \gamma_{ii} & \dots & \gamma_{in} & 1 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \gamma_{n1} & \gamma_{n2} & \dots & \gamma_{ni} & \dots & \gamma_{nn} & 1 \\ 1 & 1 & \dots & 1 & \dots & 1 & 0 \end{bmatrix} \times \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \dots \\ \lambda_i \\ \dots \\ \lambda_n \\ \mu \end{bmatrix} = \begin{bmatrix} \gamma_{10} \\ \gamma_{20} \\ \dots \\ \gamma_{i0} \\ \dots \\ \gamma_{n0} \\ 1 \end{bmatrix} \quad (\text{A.25})$$

quindi

$$A\lambda = b$$

con A la matrice delle semivarianze tra due punti x_i e x_j , b il vettore delle semivarianze tra i punti delle variabili note e il target x_0 , λ è il vettore dei pesi e del moltiplicatore di Lagrange. I pesi del kriging e il moltiplicatore di Lagrange sono ottenuti invertendo la matrice A :

$$\lambda = A^{-1}b$$

I pesi λ_j inseriti nell'eq (A.15) danno la stima di Z nel punto x_0 . Si dimostra che se le posizione x_i sono distinte, il sistema è sempre regolare e ammette quindi sempre soluzione unica. Tale sistema non dipende dall'entità da stimare ma esclusivamente dalla posizione reciproca dei punti di misura, cioè dalle informa-

zioni che si utilizzano per effettuare la stima e dalla funzione variogramma. La varianza di stima nell' eq. (A.20), che ora possiamo chiamare *varianza di kriging*, può essere riscritta tenendo conto del (A.24):

$$\sigma_k^2 = \sum_i \lambda_i \gamma(x_i - x_0) + \mu \quad (\text{A.26})$$

Block Ordinary Kriging

Se ciò che si vuole stimare, con la stessa base informativa, non è una entità puntuale, ma il valore medio della variabile Z_B su un supporto areale B allora:

$$Z_B = \frac{1}{B} \int_B Z(x) dx$$

I pesi del kriging, relativi alla stima $[Z_B]' = \sum_{j=1}^n \lambda_j(B) Z(x_j)$ si ottengono dalla risoluzione di un sistema strutturalmente analogo al precedente, ma che presenta, essendo cambiata l'entità da stimare, una differenza nei termini noti:

$$\begin{cases} \sum_{j=1}^n \lambda_j \gamma(x_i - x_j) + \mu(B) = \frac{1}{B} \int_B \gamma(x_i - x_0) dx & \text{per } i = 1, \dots, n \\ \sum_{j=1}^n \lambda_j(B) = 1 \end{cases} \quad (\text{A.27})$$

La varianza di stima associata è data da:

$$\sigma_{B_k}^2 = E[(Z^*(B) - Z(B))^2] = \quad (\text{A.28})$$

$$= \sum_i^n \lambda_i \bar{\gamma}(x_i - B) - \gamma(B, B) + \mu(B) \quad (\text{A.29})$$

Tale stimatore è più comunemente noto con il nome di *varianza di block kriging*. Nel sistema precedente il termine noto esprime il valore medio che la funzione $\gamma(h)$ assume quando un estremo di h è fisso in x_i e l'altro estremo x varia in B :

$$\bar{\gamma}(x_i, B) = \frac{1}{|B|} \int_B \gamma(x_i, x) dx \quad (\text{A.30})$$

$\gamma(B, B)$ è la media della semivarianza all'interno di B:

$$\bar{\gamma}(B, B) = \frac{1}{|B|^2} \int_B \int_B \gamma(x, x') dx dx' \quad (\text{A.31})$$

La stima del valore medio della variabile in B, ottenuta con il block kriging (A.27), coincide con la media in B delle stime puntuali ottenute con il kriging puntuale (A.24), infatti confrontando i sistemi si verifica che:

$$[Z_B]' = \sum_{j=1}^n \lambda_j(B) z(x_j) = \frac{1}{B} \int_B Z'(x) dx \quad (\text{A.32})$$

A.3.2 Kriging con drift esterno

Come abbiamo visto, il kriging è un metodo geostatistico per trovare il miglior stimatore lineare sotto l'assunzione di stazionarietà al secondo ordine. In realtà i processi potrebbero essere non stazionari, come appunto la temperatura dell'aria in un dato territorio, cioè mostrano una componente di variabilità su scala regionale o una variabilità che ingloba un trend, vale a dire una variazione sistematica della variabile più o meno accentuata. Quando il processo non è stazionario, la media non è più costante ma varia nello spazio, possiamo scrivere:

$$E[Z(x)] = m(x) = \sum_p A_p G^p \text{ per } p = 1, \dots, k \quad (\text{A.33})$$

I coefficienti A_p sono sconosciuti e non è necessario stimarli, è sufficiente infatti introdurli nel sistema del kriging. In questi casi si utilizzano funzioni e metodi particolari per la descrizioni di questi fenomeni. Uno dei criteri è conosciuto come external kriging (o universal kriging se l'external drift non è una funzione lineare di una variabile esterna, ma $m(x)$ è modellata solo come una funzione lineare o quadratica delle coordinate). La variabile di interesse è modellata da un funzione random $Z(x)$, con x il vettore di coordinate, data dalla sovrapposizione di due componenti: un termine deterministico di drift non noto $m(x)$ e una funzione random stazionaria con media nulla $Y_R(x)$ che verrà modellizzata con

il variogramma:

$$Z(x) = m(x) + Y_R(x)$$

Applichiamo così l'ordinary kriging alla componente stazionaria dei residui $Y_R(x)$, che è una funzione random a cui abbiamo sottratto il trend $m(x)$. Per formare il variogramma dei residui, si stima il drift con una regressione lineare ordinary least square (OLS):

$$\gamma_R = \frac{1}{2} \text{Var}[(r(x) - r(x+h))]^2$$

con

$$r(x) = Z_{\text{misurata}}(x) - \widehat{Z^*[1]}_{\text{stimata con OLS}}$$

$\widehat{Z^*}$ è il valore medio $m(x) = \sum_p A_p G^p$ che abbiamo supposto essere $m(x) = T(h) = a_0 + a_h h$

Richiamiamo quanto fatto precedentemente per la stima puntuale con l'ordinary kriging, la trattazione areale sarà omessa ma ottenibile in maniera analoga. Indichiamo con $Y_0 = Z(x_0)$, la stima Y_0^* di Y_0 implica (Webster e Oliver, 2001):

$$E(Y_0^* - Y_0) = \sum_{i=1}^N \lambda_i m(x_i) - m(x_0) = 0 \quad (\text{A.34})$$

poiché $m(x_i) = \sum_{p=1}^k A_p G^p(x_i)$ e sostituendo questo in (A.34) si ha:

$$\sum_{p=1}^k A_p \left[\sum_{i=1}^N \lambda_i G^p(x_i) - G^p(x_0) \right] = 0 \quad (\text{A.35})$$

quindi si ottiene la condizione

$$\sum_{I=1}^N \lambda_i G^p(x_i) = G^p(x_0) \text{ per } p = 1, 2, \dots, k$$

I pesi λ_i , come per il caso stazionario, sono le soluzione del sistema lineare che ora ha k parametri di Lagrange $\mu_1, \mu_2, \dots, \mu_k$. Questo sistema lineare ha

una soluzione unica se i k vettori $G^p(x_i)$, con $i = 1, 2, \dots, N, p = 1, 2, \dots, k$, sono linearmente indipendenti. Il sistema kriging nel caso di non stazionarietà può essere scritto ora come:

$$\begin{cases} \sum_{j=1}^n \lambda_j \gamma_R(x_i - x_j) + \mu_0 + \sum_{p=1}^k \mu_p G^p(x_i) = \gamma_R(x_i - x_0) & \text{per ogni } i \\ \sum_{j=1}^n \lambda_j(x_0) = 1 \\ \sum_{j=1}^n \lambda_j G^p(x_j) = G^p(x_0) & \text{per } p = 1, 2, \dots, k \end{cases} \quad (\text{A.36})$$

con γ_R che è il variogramma dei residui $Y_R(x) = Z(x) - m(x)$. Nel sistema (A.36) non sono presenti i coefficiente A_p , quindi non è necessario che si conosca il loro valore, mentre la nuova variabile esterna $G^p(x)$ deve essere fornita in ogni punto x_0 . Utilizzando la notazione matriciale come per l'OK:

$$A\lambda = b$$

ma ora la matrice A e i vettori λ e b contengono anche le componenti della funzione di trend:

$$A = \begin{bmatrix} \gamma_{11} & \gamma_{12} & \dots & \gamma_{1n} & 1 & G^1(x_1) & G^2(x_1) & \dots & G^K(x_1) \\ \gamma_{21} & \gamma_{22} & \dots & \gamma_{2n} & 1 & G^1(x_2) & G^2(x_2) & \dots & G^K(x_2) \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \gamma_{n1} & \gamma_{n2} & \dots & \gamma_{nn} & 1 & G^1(x_n) & G^2(x_n) & \dots & G^K(x_n) \\ 1 & 1 & \dots & 1 & 0 & 0 & 0 & \dots & 0 \\ G^1(x_1) & G^1(x_2) & \dots & G^1(x_n) & 0 & 0 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ G^K(x_1) & G^K(x_2) & \dots & G^K(x_n) & 0 & 0 & 0 & \dots & 0 \end{bmatrix} \quad (\text{A.37})$$

$$\lambda = \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \dots \\ \lambda_n \\ \mu_0 \\ \mu_1 \\ \mu_2 \\ \dots \\ \mu_K \end{bmatrix} \text{ e } \mathbf{b} = \begin{bmatrix} \gamma_{10} \\ \gamma_{20} \\ \dots \\ \gamma_{n0} \\ 1 \\ G^1(x_0) \\ G^2(x_0) \\ \dots \\ G^K(x_0) \end{bmatrix} \quad (\text{A.38})$$

Per trovare i pesi e i moltiplicatori di Lagrange si inverte la matrice \mathbf{A} :

$$\lambda = \mathbf{A}^{-1}\mathbf{b}$$

inserendo i pesi nella $Z^*(x_0) = \sum_{i=1}^N \lambda_i G^p(x_i)$ si ottiene la stima.

La varianza del kriging $\sigma_{ked}^2 = \mathbf{b}^T \lambda$ si riscrive come:

$$\sigma_{ked}^2 = Var[Z^*(x_0) - Z(x_0)] = \sum_i^N \lambda_i \gamma_R(x_i - x_0) + \sum_{p=1}^k \mu_p G^p(x_0) \quad (\text{A.39})$$

A.4 Cross validation leave-one-out

In ogni studio di tipo geostatistico il numero di decisioni soggettive e interdipendenti è spesso così elevato che, prima di intraprendere la fase finale di stima, è buona norma validare l'intero modello implementato e la procedura di kriging stessa. Lo scopo è quindi quello di verificare e confrontare l'influenza dei vari parametri implementati (modello di (co)-variogramma, raggio di ricerca per il vicinaggio, numero di punti da impiegare nella stima, tipo di kriging, ecc.) al fine di ottenere l'insieme di quelli 'ottimali'; in quest'ottica, la cross validation si configura come uno strumento sia qualitativo che quantitativo. L'idea di base è la seguente: il campionamento i -esimo $z(x_i)$ viene temporaneamente escluso dal dataset, e il suo valore $\widehat{z(x_i)}$ viene stimato mediante gli altri dati sperimentali

che ne costituiscono il vicinaggio; si determina quindi il residuo i -esimo come $r(x_i) = z(x_i) - \widehat{z(x_i)}$, e il campionamento $z(x_i)$ viene così reintrodotta nel dataset. L'intera procedura viene ripetuta per tutti gli N valori disponibili. Si opterà quindi per il modello/metodo che manifesta il comportamento migliore, una volta che si sia definita una opportuna misura per identificare il risultato migliore. Indicando con l'indice k una delle K procedure di stima o varianti di implementazione sotto esame, tra i criteri cui far riferimento per l'analisi dei residui si è scelto di considerare il RMSE (Root Mean Squared Error dei residui). Questo è l'errore minimo quadrato che indica la precisione, idealmente piccolo, ragione per cui il modello migliore minimizzerà la media dei residui :

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (z(x_i) - \widehat{z(x_i)})^2} \quad (\text{A.40})$$

A.5 Temperature medie nella provincia aquilana.

Ora vediamo come il kriging con external drift sia utile per predire i valori medi di temperatura nella provincia aquilana.

Le indagini radon condotte nella regione Abruzzo iniziarono, come già visto, nei primi anni novanta e sono a tutt'oggi in corso. Esse prevedono previsioni di misurazioni di radon indoor di durata annuale, in edifici scelti in modo da rappresentare un campione il più possibile omogeneo sul territorio abruzzese e rappresentativo del patrimonio edilizio regionale. Al fine di studiare l'influenza della temperatura media esterna sulla misura di concentrazione di radon, si è provveduto ad associare a questa variabile una stima della temperatura media annua del sito cui si riferisce la misura.

Allo scopo sono stati presi in considerazione dati medi di temperatura giornalieri, relativi a 54 stazioni termometriche del Servizio Idrografico e Mareografico Nazionale (SIMN), registrati dal 1991 al 2011.

I siti del sistema di rilevamento del servizio idrografico considerati per la maggior parte si trovano nel territorio della provincia aquilana e i restanti sono

posizionati in località limitrofe (fig. A.3). Sono considerati influenti per l'interpolazione le misure provenienti da luoghi troppo distanti da quelli di nostro interesse.

I dati giornalieri registrati dalle stazioni termometriche sono stati raccolti e analizzati, uno schema riassuntivo è riportato nella tabella (tab. A.1). Capita spesso che, con riferimento al ventennio 1991-2011, manchino le misure di periodi più o meno lunghi, non permettendo una corretta analisi. Abbiamo indicato nella tabella con *NA* il numero di temperature medie annuali che non è stato possibile calcolare con i dati a disposizione.

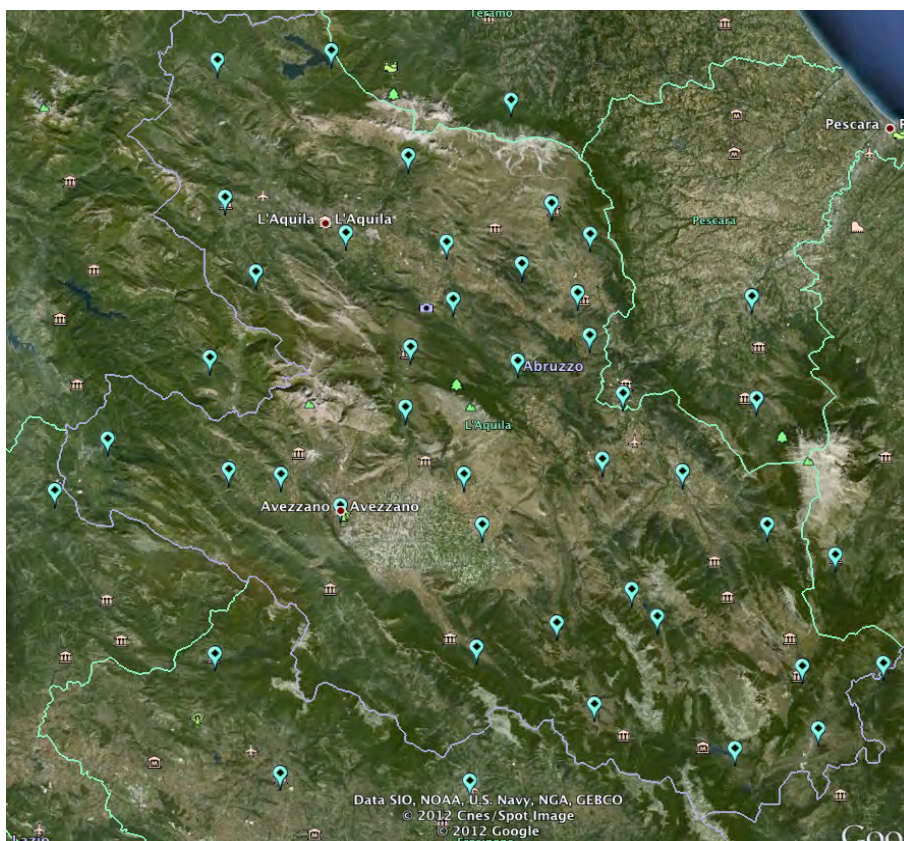


Figura A.3: Siti del servizio idrografico utilizzato per l'interpolazione delle temperature.

Appendice A. Analisi spaziale per la mappatura delle temperature medie

Comune	h	AM	SD	med	min	max	q _{1/4}	q _{3/4}	NA
tornimparte	873	11.42	1.43	11.98	8.82	13.10	10.08	12.53	1
sulmona	372	14.71	0.72	14.77	13.32	15.98	14.34	15.04	0
scanno	1039	11.34	0.52	11.49	10.35	12.05	11.19	11.73	1
roccaraso	1231	7.03	0.58	7.13	5.96	7.72	6.78	7.57	4
pescocostanzo	1461	8.10	1.26	8.63	4.86	9.19	8.14	8.79	1
montereale	913	10.98	1.10	11.42	8.36	12.22	10.81	11.68	0
l'aquila	595	12.53	0.51	12.69	11.32	13.44	12.43	12.82	0
goriano sicoli	958	9.85	2.20	10.41	6.07	12.43	8.30	11.89	0
assergi	902	11.52	0.98	11.89	9.73	12.65	10.37	12.28	4
caramanico	700	14.14	0.85	14.27	12.30	15.40	13.71	14.80	1
isola del gransasso	553	12.98	0.69	13.24	11.70	13.88	12.70	13.50	3
...
gioia vecchio	1375	6.44	1.47	6.60	3.80	8.20	5.80	7.45	14
san benedetto dei marsi	687	9.73	2.07	10.00	6.30	11.90	8.80	11.15	14
cerchio	834	9.85	1.97	10.20	6.20	11.60	9.58	11.20	15
ovindoli	1363	6.67	1.40	6.70	4.80	8.30	5.68	7.80	15
avezzano	708	11.88	0.48	11.80	11.40	12.80	11.65	11.88	15
villavallelonga	945	10.17	1.01	10.35	8.30	11.20	10.05	10.73	15
carsoli	640	11.28	2.56	12.30	7.20	13.70	10.50	12.70	16

Tabella A.1: Tabella riassuntiva delle temperature medie annuali (1991-2011) in °C . *h*= altitudine della stazione termometrica; *AM* = media aritmetica delle temperature; *SM*= deviazione standard; *med* = la mediana; *MIN* e *MAX* temperature minime e massime registrate; q_{1/4} e q_{3/4} primo e terzo quartile; NA = numero di valori annuali mancanti.

Le stazioni termometriche prese in considerazione fanno parte di tre compartimenti del Servizio Idrografico:

- il compartimento di Pescara, che raccoglie la maggior parte dei siti utilizzati e da cui abbiamo a disposizione la quasi totalità delle informazioni dal 1991 al 2011;

- il compartimento di Roma, con le stazioni di Scurlola Marsicana, Borgorose, Trevi nel Lazio, Arsoli, Sora, Alatri, Carsoli e Tagliacozzo. I dati sono stati presi dal sito <http://www.idrografico.roma.it/> che permette di ottenere gli annali

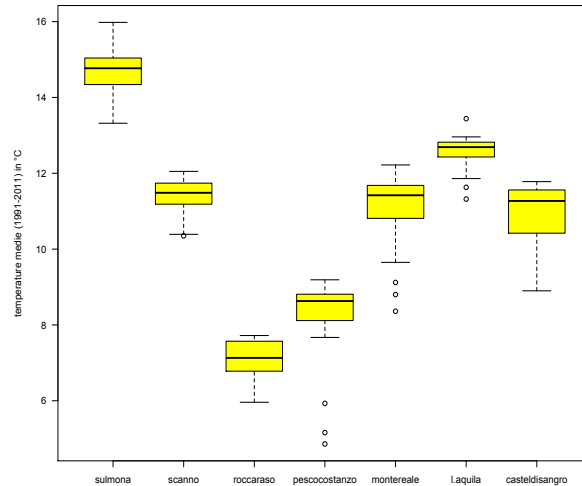


Figura A.4: Boxplot per le temperature medie in riferimento al periodo 1991-2011 di alcuni siti.

idrologici fino al 2002 e le misure giornaliere dal 2002 ad oggi.

- il compartimento di Napoli con Gioia Vecchio, Cerchio, San Benedetto dei Marsi, Ovindoli, Avezzano e Villavallelonga. I dati sono scarsi, si fermano al 1997 e sono reperibili on line su <http://193.206.192.243/annali/>. A fronte di questa difficoltà, per calcolare la temperatura media annua abbiamo preso in considerazione le misure del decennio 1987-1997.

Tuttavia è stata necessaria una scrematura del numero delle stazioni esaminate, i siti cioè con pochi dati o misure evidentemente errate sono stati eliminati, così come una delle stazioni tra due che distano meno di 5 Km. Infatti il limite di questa trattazione semplificata è l'incapacità di spiegare variazioni su piccola scala. Non sono stati considerati i siti di Bagnaturo, distante da Sulmona 2.62 km, di Sant'Eufemia a Majella distante dalla stazione termometrica di Caramanico Terme di 3 km, di Villalogo, preferendo Scanno, di Pescocostanzo e di Bussi.

Per rendere il set di dati il più possibile omogeneo e poter prevedere la temperatura media annua dei siti in cui i dati mancano per un periodo breve, a causa

ad esempio di un guasto dei termometri o di problemi nella registrazione, si è provveduto a stimare il valore atteso di temperatura con un approccio empirico: - considerando le sole stazioni a una distanza minore o uguale di 15 km del sito in esame, si calcola la variazione media tra le temperature delle località e la si applica alla temperatura non nota.

Ad esempio, per il comune di Campo di Giove le osservazioni termometriche sono iniziate nel 1999; per stimare la temperatura T_{CdG}^{91-00} del decennio 1991-2000, abbiamo dapprima calcolato le variazioni medie ΔT_i della temperatura misurata a Campo di Giove e le stazioni dei comuni entro un raggio di 15 Km (Sulmona e Palena) nel decennio 2001-2010, ne abbiamo fatto la media e sommato il risultato alle temperature medie di Sulmona e Palena per il periodo 1991-2000, in modo da avere il più possibile un set di dati completo:

$$\begin{aligned} \Delta T_{Sulmona} &= T_{CdG}^{01-11} - T_{Sulmona}^{01-11} \\ \Delta T_{Palena} &= T_{CdG}^{01-11} - T_{Palena}^{01-11} \\ \rightarrow \Delta T &= \frac{\Delta T_{Sulmona} + \Delta T_{Palena}}{2} \\ T_{CdG_{Sulmona}}^{91-00} &= T_{Sulmona}^{91-00} + \Delta T_{Sulmona} \\ T_{CdG_{Palena}}^{91-00} &= T_{Palena}^{91-00} + \Delta T_{Palena} \\ \rightarrow T_{CdG}^{91-00} &= \frac{T_{CdG_{Sulmona}}^{91-00} + T_{CdG_{Palena}}^{91-00}}{2} \end{aligned}$$

	n	AM	SD	GM	GSD	MAX	MIN
$102 \leq h \leq 550$	8.00	14.52	1.02	14.49	1.07	16.25	13.23
$551 \leq h \leq 785$	13.00	11.92	1.14	11.87	1.10	14.14	10.18
$786 \leq h \leq 1020$	13.00	11.10	0.99	11.06	1.10	12.40	8.90
$1021 \leq h \leq 1480$	9.00	8.46	1.68	8.32	1.21	11.34	6.15

Tabella A.2: Statistiche riassunte per i valori di temperatura media annua calcolati grazie ai dati giornalieri di 43 stazioni termometriche del S.I.M.N. L'altitudine h è espressa in metri e le temperature in °C.

A.5.1 Influenza dell' altitudine.

La variazione spaziale della temperatura è controllata da una moltitudine di fattori, tra i quali la topografia, la latitudine e la distanza dalla costa. In letteratura esistono indagini su vasta scala sulla variabilità spaziale di grandezze climatiche, tra cui anche la temperatura, in relazione ai fattori che le influenzano. In riferimento alla temperatura, i modelli prevalentemente utilizzati si basano sull'impiego di soli parametri di quota e latitudine, e la relazione analitica maggiormente utilizzata è una regressione lineare. Aumentando la latitudine, appunto, l'altezza del sole sull'orizzonte decresce e quindi si ha una variazione della durata e dell'intensità dell'insolazione. All'aumentare della quota, invece, l'aria diviene sempre più rarefatta e povera di vapore acqueo, quindi sempre meno capace di assorbire e trattenere energia. Il gradiente termico verticale (*lapse rate* $\Gamma = -\frac{dT}{dz}$) in funzione della altitudine nella troposfera (fino a 10-15 km dal suolo) è in media di circa - 0.6 °C ogni 100 metri di altezza (Barry e Chorley, 1987).

La ragione della diminuzione della temperatura con l'aumentare dell'altitudine si spiega con il fatto che gli strati più bassi dell'atmosfera sono i più caldi, in quanto l'effetto di riscaldamento esercitato dalla radiazione aumenta con la densità del mezzo attraversato, e per conseguenza gli strati inferiori essendo i più densi si riscaldano maggiormente. D'altra parte gli strati inferiori d'aria acquistano calore anche a causa del contatto con la superficie terrestre. La riduzione della densità conseguente al riscaldamento fa sì che le masse d'aria calda tendano a sollevarsi e, incontrando pressioni minori, in questo movimento ascensionale si dilatano e perdono calore (Visconti, 2001).

Con una semplice regressione lineare multivariata, analizziamo i nostri dati medi di temperatura in funzione dell'altitudine e della latitudine:

$$T(h, Lat, Lon) = a_{interc.} + a_h h + a_{Lat} Lat$$

Appendice A. Analisi spaziale per la mappatura delle temperature medie

	Estimate	Std. Error	t value	Pr(> t)	
$a_{interc.}$	-41.0192	30.2274	-1.36	0.1824	
a_h	-0.0065	0.0005	-12.71	1.27E-15	***
a_{Lat}	1.238E-05	6.489E-06	1.91	0.0636	.

Tabella A.3: Coefficienti della regressione lineare. La verifica delle ipotesi che i coefficienti di regressione sono significativamente diversi da zero, viene fatta con il valore del test di Student $t = \frac{\hat{\beta}_i}{se(\hat{\beta}_i)}$. Il grado di significatività è espresso dal numero di asterischi.

$$R^2 = 0.8022; R_{adj}^2 = 0.7923$$

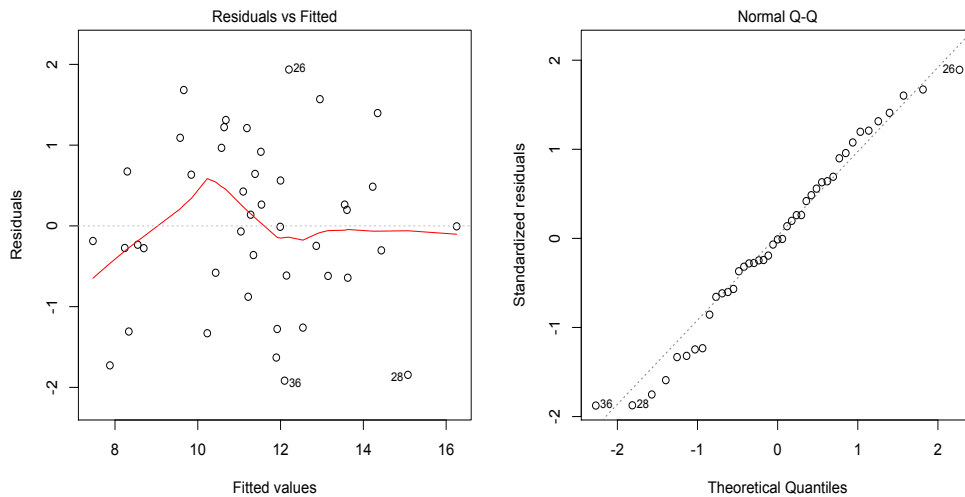


Figura A.5: Verifica della linearità : grafico dei residui verso i valori previsti; normalità dei residui standardizzati che seguono una distribuzione gaussiana $N(0,1)$

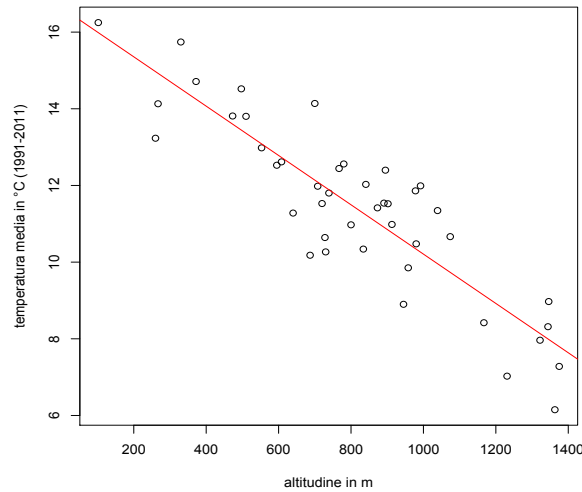


Figura A.6: Andamento della temperatura media annua in funzione dell'altitudine (in m s.l.m). L'indice di determinazione per la regressione lineare è $R^2 = 0.8022$.

Come evidenziato nella tab. (A.3) il coefficiente di regressione è significativo ($p - value < 0.05$) per la sola altitudine. Utilizzando il kriging con l'altitudine come drift esterno, siamo in grado di predire il valori di temperatura media annuale nei punti dove è stata misurata la concentrazione di radon indoor tenendo conto della variabilità della morfologia del territorio.

I risultati ottenuti con la funzione *krige* del pacchetto *gstat* in **R**, sono rappresentati nella fig (A.8). In tab. (A.4) sono riportate le caratteristiche del variogramma sperimentale dei residui $Y_R = Z(x) - m(x)$ con $m(x) = a_{intercept} + a_h h$ (si veda quanto riportano nella sezione A.3.2) e i valori caratteristici del modello di fit esponenziale.

in tab (a), np è il numero di coppie, $dist$ è la separazione media in metri tra queste e γ è la semivarianza media dei residui;

Appendice A. Analisi spaziale per la mappatura delle temperature medie

	np	dist	γ		
1	6	6491.053	0.442044		
2	32	10325.982	0.6518355	cross validation	
3	50	15331.222	0.8501252	mean_error	0.03
4	93	21888.275	1.1931128	MSE	0.84
5	97	28791.466	1.0681907	cor_obspred	0.91
6	117	36358.254	1.3751008	cor_predres	-0.03
7	144	44887.07	1.3311131	RMSE	0.92
	model	psill	range	RMSE_sd	0.40
1	Nug	0	0	(b) cross validation	
2	Exp	1.490187	17340.17		

(a) variogramma dei residui

Tabella A.4: a) proprietà del variogramma sperimentale dei residui e funzione di fitting; b) diagnostiche ottenute con la cross validation leave-one-out.

In tab.(A.4) (b) vengono riportate le diagnostiche test della cross-validation leave-one-out dei dati di temperatura utilizzando il modello di variogramma scelto (fig.A.8).

. Le diagnostiche test sono:

- mean_error: la media dei residui della cross validation, idealmente piccolo.
- MSE: mean squared error;
- cor_obspred: correlazione tra i valori osservati e predetti, idealmente;
- cor_predres: correlazione tra i valori predetti e i residui, idealmente piccola;
- RMSE: Root Mean Squared Error dei residui, idealmente piccolo;
- RMSE_sd: RMSE diviso per la deviazione standard dei valori osservati. Fornisce una misura della variazione dei residui contro la variabilità dei valori osservati.

Come riportato in fig.(A.8) il miglior fit è raggiunto con un modello esponenziale con *range pratico* = 17340 m, oltre il quale due punti possono essere considerati indipendenti, nugget nullo e sill=1.45.

Appendice A. Analisi spaziale per la mappatura delle temperature medie

h	n	AM	SD	GM	GSD	MAX	MIN
$320 \leq h \leq 550$ (a)	196	13.85	0.68	13.83	1.05	15.11	12.20
$550 < h \leq 785$ (b)	262	11.72	0.93	11.68	1.08	13.82	10.16
$785 \leq h \leq 1020$ (c)	183	10.82	0.89	10.78	1.09	12.76	8.61
$1020 \leq h \leq 1480$ (d)	74	9.07	1.09	9.00	1.13	11.15	6.21

Tabella A.5: Statistiche riassunte per i valori di temperatura media annua predetti con KED. L' altitudine h è espressa in metri e le temperature in °C.

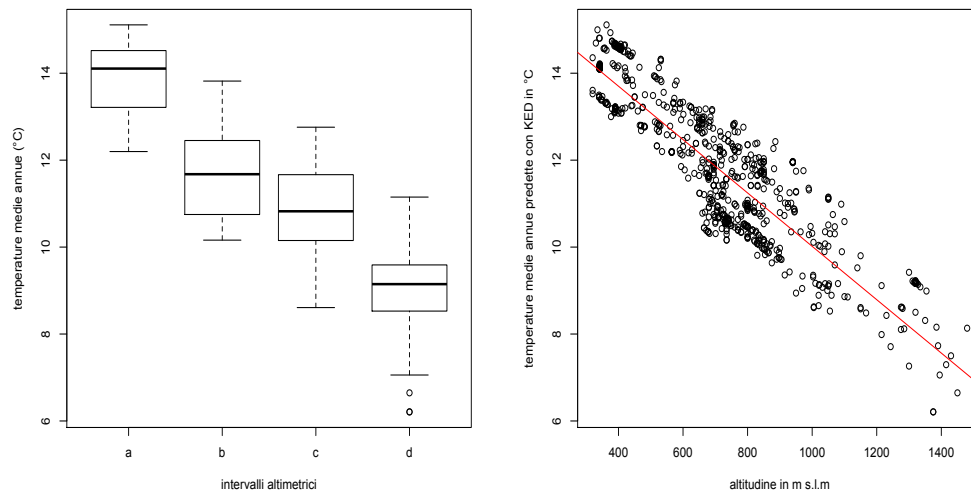
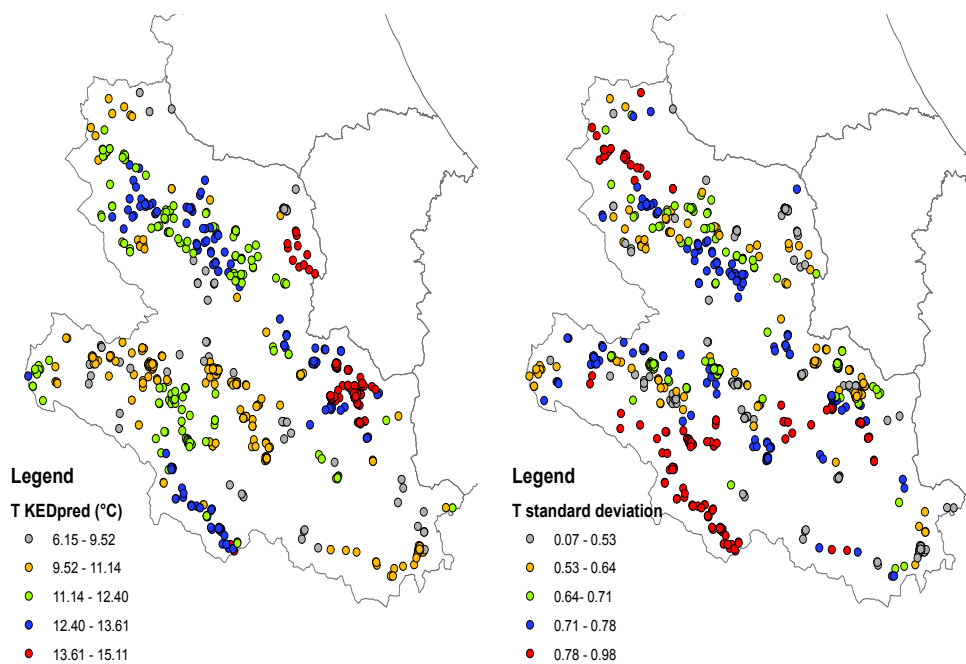


Figura A.7: Boxplot e variabilità delle temperature stimate in funzione dell'altitudine. Le lettere a, b, \dots fanno riferimento agli intervalli altimetrici riportati in tab.(A.5).



Experimental variogram and fitted variogram model

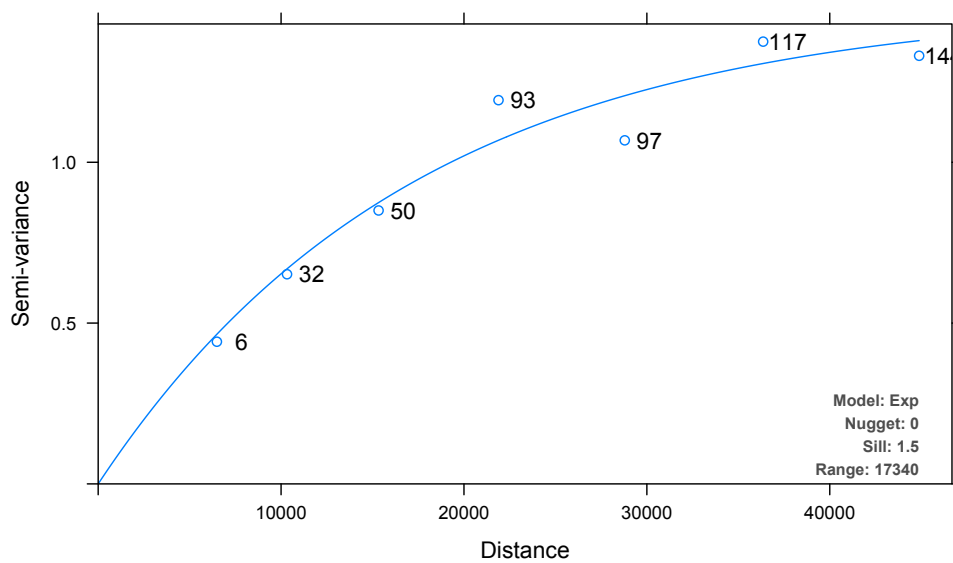


Figura A.8: Output del kriging con drift estremo per le temperature medie annuali. In alto gli intervalli per le mappe sono scelti in base ai quantili.

Appendice **B**

Codici in R

B.1 Test di normalità per i dati log-trasformati di radon

```
library(nortest)
library(fBasics)
rOUT<-8;
j<-1
p<-0;
A<-0;
B<-0;
while (rOUT<= 12) {
  D<-0;
  while (D<=5){
for(i in 1:n){
#sottrazione del fondo e correzione per D
radon$lnr0[i]<-log((radon$Rn_fa[i] -rOUT)/2 +
  sqrt((radon$Rn_fa[i] -rOUT)^2/4 + D^2))
}
#D' Agostino test
a<- dagoTest(radon$lnr0);
  if(a@test$p.value[1] > p) {p<-a@test$p.value[1];
    A<-D;
    B<-rOUT};
  D<- D +0.1 }
  rOUT<-rOUT +0.1}
#Shapito Wilk test
s<-shapiroTest(radon$lnr0);
if(s$p.value > p) {p<-s$p.value;
  A<-D;
  B<-rOUT}
```

B.2 Analisi multilevel bayesina con R2jags

B.2.1 Modello per l'intercetta random e coefficienti fissi

```

library(R2jags)
ipianoct<- factor(radon$Ipianoct)
cated<- factor(radon$Cated)
ind.anno <- factor(radon$Indanno)
ind.mat<- factor(radon$Indmat)
Comune.name <- as.vector(radon$Comune)
uni.name <- unique(Comune.name)
J <- length(uni.name)
com.id <- rep(NA,J)
for (i in 1:J){
  com.id[Comune.name==uni.name[i]] <- i }
n<-nrow(radon)
X<-cbind(piano1,piano2,pianoS,nct,c_scuola,
         c_pubblico,c_luogodilavoro,pietra,
         z_altro,y_primadel1919, y_19191960)
K<-ncol(X)
data<-list("n"=n,"J"=J, "K"=K, "y"=y, "com.id"=com.id,
          "X"=X)
radon.inits <- function (){ list
  (at=rnorm(J),Br=rnorm(K),mu.a=rnorm(1),sigma.y=runif(1),
  sigma.a=runif(1))}
parameters <- c ("at","Br", "sigma.y","mu.a", "sigma.a",
  "e.y")
jags<-jags(data,radon.inits,parameters,
  model.file="mod4.txt", n.chain=3, n.iter=10000)
r.squared.y <-
  1-mean(apply(jags$BUGSoutput$sims.list$e.y,1,var))/var(y)

```

Codice jags mod4.txt

```

model {
  for (i in 1:n){
    y[i] ~ dnorm (y.hat[i], tau.y)
    y.hat[i] <- at[com.id[i]] + inprod(Br [], X[i,])
    e.y[i] <- y[i]-y.hat[i]
  }
  tau.y <- pow(sigma.y, -2)
  sigma.y ~ dunif (0, 10)
  for (j in 1:J){
    at[j] ~ dnorm (mu.a, tau.a)
    e.a[j] <- at[j]-mu.a
  }
}

```

```

mu.a ~ dnorm (0, .00001)
tau.a <- pow(sigma.a, -2)
sigma.a ~ dunif (0, 1000)
for (k in 1:K){
Br[k] ~ dnorm (0, .0001)}}

```

Codice Jags del raggruppamento per comune e per classe litologica

```

model {
for (i in 1:n){
y[i] ~ dnorm (y.hat[i], tau.y)
y.hat[i] <- at[com.id[i]] + bt[lito.id[i]]
}
for (j in 1:J){
at[j] ~ dnorm (mu.a, tau.a)
e.a[j] <- at[j]-mu.a
}
for (k in 1:Jl){
bt[k] ~ dnorm (mu.b, tau.b)
e.b[k] <- bt[k]-mu.b}
mu.a ~ dnorm (0, .0001)
tau.a <- pow(sigma.a, -2)
sigma.a ~ dunif (0, 100)
mu.b ~ dnorm (0, .0001)
tau.b <- pow(sigma.b, -2)
sigma.b ~ dunif (0, 100)
tau.y <- pow(sigma.y, -2)
sigma.y ~ dunif (0, 100)
}

```

B.2.2 Modello a effetti misti

```

#esempio con effetti fissi piano e Cated; effetti random:
Indanno e Indmat
X<-cbind(costpietra, z_altro,y_primadel1919, y_19191960)
Z<-cbind(piano1,piano2,pianoS,nct,c_scuola,
c_pubblico,c_luogodilavoro)
K<-ncol(X)
T<-ncol(Z)
W<-diag(K)
data<-list("n"=n,"J"=J, "K"=K, "T"=T, "Z"=Z, "y"=y,
"com.id"=com.id, "X"=X, "W"=W)
inits <- function (){
list (B.raw=array(rnorm(J*K),c(J,K)), EF= rnorm(T),
mu.raw=rnorm(K),

```

```

sigma.y=runif(1), Tau.B.raw=rwish(K+1,diag(K)),
  xi=runif(K))}
parameters <- c ("B", "mu", "sigma.y", "sigma.B", "rho.B",
  "EF", "e.y")
jags.matanno<-jags(data,inits,parameters,
  model.file="~/Desktop/winbugs/mixedmodel.txt",
  n.chain=3, n.iter=10000)

```

Codice jags mixedmodel.txt

```

model {
for (t in 1:T){
EF[t] ~ dnorm (0, .00001)
}
sigma.y ~ dunif (0, 10)
tau.y <- pow(sigma.y, -2)
# K è il num di coeff random
# T è il num di coeff fissi
for (j in 1:J){
B.raw[j,1:K] ~ dnorm (mu.raw[], Tau.B.raw[,])
for (k in 1:K){
B[j,k] <- xi[k]*B.raw[j,k]
}}
for (k in 1:K){
mu.raw[k] ~ dnorm (0, .00001)
xi[k] ~ dunif (0, 100)
mu[k] <- xi[k]*mu.raw[k]
}
Tau.B.raw[1:K,1:K] ~ dwish (W[,], df)
df <- K+1
Sigma.B.raw[1:K,1:K] <- inverse(Tau.B.raw[,])
for (k in 1:K){
for (k.prime in 1:K){
rho.B[k,k.prime] <- Sigma.B.raw[k,k.prime]/
sqrt(Sigma.B.raw[k,k]*Sigma.B.raw[k.prime,k.prime])
}
sigma.B[k] <- abs(xi[k])*sqrt(Sigma.B.raw[k,k])
}
for (i in 1:n){
y.hat[i] <- inprod(EF[], Z[i,]) + inprod(B[com.id[i],],
X[i,])
y[i] ~ dnorm (y.hat[i], tau.y)
e.y[i] <- y[i]-y.hat[i]}}

```

B.3 Kriging con drift esterno

```

#Regressione lineare per T(h)
t.lm<-lm(t_12~ h+ Lat, temp)
plot(t.lm)
#Variogramma
library(gstat)
library(sp)
temp <- read.table("temperatureIdrografico.csv", header=T,
  sep=";", dec=".")
coordinates(temp) <- c("Lon", "Lat")
t.var <- variogram(t~h,temp)
temp.fit = fit.variogram(t.var, model = vgm(1, "Exp",
  15000))
plot(temp.fit)
#Cross validation leave-one-out
kr.cv = krige.cv(t~h, temp, mm)
rmse= sqrt(mean(kr.cv$residual^2))
#Krige con drift esterno
t.uk <- krige(t~h, temp, radon_place, fit.var)
plot(t.uk)

```

B.4 Indice I di Moran

```

library(ape)
#matrice dei pesi inversamente proporzionali alla distanza
moran.dists <- as.matrix(dist(cbind(radon_gwr$Lon,
  radon_gwr$Lat)))
moran.dists.inv <- 1/moran.dists
diag(moran.dists.inv)<-0
Moran.I(radon_gwr$ln_t,moran.dists.inv)
Moran.I(radon_gwr$ln_perm,moran.dists.inv)
Moran.I(radon_gwr$lnu_blk,moran.dists.inv)

```

Bibliografia

- Andersen A. (1992). Entry of soil gas and radon into houses. *Riso-R-623(EN)*.
- Andersen A.; Ulbak K. e altri (2001). Mapping indoor radon-222 in Denmark: design and test of the statistical model used in the second nationwide survey. *Sci. Tot. Env.*, **272**.
- Andersen A.; Ulbak K. e altri (2007). Prediction of ²²²Rn in Danish dwellings using geology and house construction information from central databases. *Radiat Prot Dosimetry*, **123**.
- ANPA (2000). *Sistema informativo territoriale per la valutazione del potenziale di esalazione di radon dal suolo*. Serie stato dell'ambiente 9/2000.
- Appleton J.; Miles J. e altri (99). Pilot study of the application of Tellus airborne radiometric and soil geochemical data for radon mapping. *J.Env.Rad.*, **1687–1697**.
- Apte M.; Price P.; Nero A.; Revzan R. (1999). Predicting New Hampshire indoor radon concentrations from geological information and other covariates. *Env. Geol.*, **37**, 181– 194.
- ARTA (2008). *Il problema radon in Abruzzo*. Quaderni ARTA.
- Barry R.; Chorley R. (1987). Atmosphere, weather and climate, methuen & co. Ltd., London, UK, pp. 274–328.

- Bellotti E.; Di Carlo G.; Di Sabatino D.; Ferrari N.; Laubenstein M.; Pandola L.; Tomei C. (2007). γ -ray spectrometry of soil samples from the Provincia dell'Aquila (Central Italy). *Appl.Rad. Isot.*, **65**, 858–865.
- Bochicchio F. (2008). The radon issue: Considerations on regulatory approaches and exposure evaluations on the basis of recent epidemiological results. *Applied Radiation and Isotopes*, **66**, 1561– 1566.
- Bochicchio F.; Campos Venuti G. e altri (1996). Results of the Representative Italian National Survey on Radon Indoors. *Health Physics*, **71**(5), 741–748.
- Bochicchio F.; Campos, G. Venuti S. e altri (2005). Annual average and seasonal variations of residential radon concentration for all the Italian Regions. *Radiation Measurements*, **40**, 686–694.
- Borgoni R.; Tritto V.; Bigliotto C.; de Bartolo D. (2011). A Geostatistical Approach to Assess the Spatial Association between Indoor Radon Concentration, Geological Features and Building Characteristics: The Case of Lombardy, Northern Italy. *Int. J. Environ. Res. Public Health*, **8**.
- Bossew P.; Lettner H. (2007). Investigations on indoor Radon in Austria, part 1: Seasonality of indoor radon concentrations. *J.Env.Rad.*, **98**(3), 329–345.
- Bossew P.; Dubois G.; Tollefsen T. (2008). Investigations on indoor Radon in Austria, part 2: Geo-logical classes as categorical external drift for spatial modelling of the Radon potential. *J.Env.Rad.*, **99**(81-97).
- Brunsdon C.; Fotheringham A. S.; Charlton M. E. (1996). Geographically weighted regression: a method for exploring spatial nonstationarity. *Geographical analysis*, **28**(4), 281–298.
- Ciotoli G.; Lombardi S.; Annunziatellis A. (2007). Geostatistical analysis of soil gas data in a high seismic intermontane basin: Fucino Plain, central Italy. *Journal Of Geophysical Research*, **112**.

- Cliff A. D.; Ord J. K. (1981). *Spatial processes: models & applications*. Pion London.
- Commissione Europea (EC) (2011). Proposal for a Council Directive laying down basic safety standards for protection against the dangers arising from exposure to ionizing radiation. *Draft presentato per Article 31 Euratom Treaty for the opinion of the European Economic and Social Committee (Brussels, 29.9.2011)*.
- Cressie N. A. C. (1991). *Statistics for Spatial Data*. John Wiley & Sons.
- Darby S.; Hill D. e altri (2005). Radon in homes and risk of lung cancer: collaborative analysis of individual data from 13 European case-control studies. *Brit. Med. J.*, **330**.
- Darby S.; Hill D.; Deo H. e altri (2006). Residential radon and lung cancer: detailed results of a collaborative analysis of individual data on 7148 subjects with lung cancer and 14208 subjects without lung cancer from 13 epidemiologic studies in Europe. *Scand J Work Environ Health*, **32**.
- de Leeuw J.; Meijer E. (2008). *Introduction to Multilevel Analysis, in Handbook of Multilevel Analysis*. Springer.
- della Salute M. (2002). Piano Nazionale Radon.
- Di Carlo P.; Pitari G.; De Luca N.; Battisti D. (2009). Observations of surface radon in Central Italy. *Environ Geol*.
- Dubois G. (2005). An overview of radon surveys in Europe. *Office for Official Publications of the European Communities*.
- Dubois G.; Bossew P. (2006). A European Atlas of Natural Radiations including harmonized radon maps of the European Union. What do we have, what do we know, quo vadimus? *European Commission-DG Joint Research Center. Institute for Environment and Sustainability*.

- Dubois G.; Bossew P.; Friedmann H. (2007). A Geostatistical autopsy of the Austrian indoor radon survey (1992-2002). *The Science of the Total Environment*, **377**.
- Dubois G.; Bossew P.; Tollefsen T.; De Cort M. (2010). First steps towards a European atlas of natural radiation: status of the European indoor radon map. *Journal of environmental radioactivity*, **101**(10), 786–798.
- Fattori S. (2006). *Misure di Spettroscopia Gamma di Campioni di Suolo della Provincia dell'Aquila*. Tesi non pubblicata, Università degli Studi Dell'Aquila.
- Font, Ll and Baixeras, C (2003). The RAGENA dynamic model of radon generation, entry and accumulation indoors. *Science of the total environment*, **307**(1), 55–69.
- Fotheringham A.; Charlton M.; Brunson C. (1998). Geographically weighted regression: a natural evolution of the expansion method for spatial data analysis. *Environment and Planning A*.
- Fotheringham A.; Brunson C.; Charlton M. (2002). *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*. Chichester.
- Friedmann H. (2005). Final results of the Austrian radon project. *Health physics*, **89**(4), 339–348.
- Gelman A.; Hill J. (2007). *Data Analysis Using Regression and Multilevel/Hierarchical Models*.
- Gelman A.; Carlin J.; Stern H.; Rubin D. (2003). *Bayesian Data Analysis*. Chapman & Hall.
- Ghisetti F.; Vezzani L. (1998). *Carta geologica della Regione Abruzzo in scala 1:100.000*. Regione Abruzzo.
- Goldstein H. (2002). *Multilevel Statistical Models*. John Wiley & Sons.

- Gotway C. A.; Young L. J. (2002). Combining incompatible spatial data. *Journal of the American Statistical Association*, **97**(458), 632–648.
- Gruber V.; Bossew P.; De Cort M.; Tollefsen T. (2013). The European map of the geogenic radon potential. *J. Radiol. Prot.*
- Gunby J.; Darby S.; Miles J. e altri (1993). Factors affecting indoor radon concentrations in the United Kingdom. *Health Phys.*
- GWR 4 Development Team (2012). *GWR4 Windows Application for Geographically Weighted Regression Modelling*. url:<http://gwr.nuim.ie>.
- Hox J. (2010). *Multilevel analysis: Techniques and applications*. Routledge Academic.
- Hudson G. (1992). *Kriging Temperature in Scotland using the External Drift Method*. Geostatistics Tróia.
- Hudson G.; Wackernagel H. (1994). Mapping temperature using kriging with external drift: Theory and an example from scotland. *International Journal Climatology*.
- Hunter N.; Muirhead C. R. e altri (2009). Uncertainties in radon related to house-specific factors and proximity to geological boundaries in England. *Radiation protection dosimetry*, **136**(1), 17–22.
- Ielsch G.; Cushing M.; Combes P.; Cuney M. (2010). Mapping of the geogenic radon potential in France to improve radon risk management: methodology and first application to region Bourgogne. *J Environ Radioactivity*.
- International Commission on Radiological Protection (2008). *ICRP Publication 103: Recommendations of the ICRP, Annals of the ICRP*.
- Journel A. G.; Huijbregts C. J. (1978). *Mining geostatistics*, volume 600. Academic press London.

- Kemski J.; Siehl A.; Stegemann R.; Valdivia-Manchego M. (2001). Mapping the geogenic radon potential in Germany. *Science of the total environment*, **272**(1), 217–230.
- Leung Y.; Mei C.; Zhang W. (2000a). Statistical tests for spatial non-stationarity based on the geo-geographically weighted regression model. *Env. Plan.*
- Leung Y.; Mei C.; Zhang W. (2000b). Statistical tests for spatial nonstationarity based on the geographically weighted regression model. *Environment and Planning A*.
- Matheron G. (1963). Principles of geostatistics. *Economic geology*, **58**(8), 1246–1266.
- Miles J. (1994). Mapping the proportion of the housing stock exceeding a radon reference level. *Radiation Protection Dosimetry*, **56**(1–4), 207–210.
- Miles J. (1998). Mapping radon-prone areas by lognormal modeling of house radon data. *Health Phys.*, **74**, 370–8.
- Miles J. (2001). Temporal variation of radon levels in houses and implications for radon measurement strategies. *Radiat. Protect. Dosimetry.*, **93**, 369–376.
- Miles J.; Algar R. (1997). Measurements of radon decay product concentrations under power lines. *Radiation protection dosimetry*, **74**(3), 193–194.
- Miles J.; Appleton J. (2005). Mapping variation in radon potential both between and within geological units. *J Radiol Prot*, (25), 257–76.
- Miles J.; Ball K. (1996). Mapping radon-prone areas using house radon data and geological boundaries. *Environment International*.
- Miles J.; Appleton J. e altri (2007). Indicative atlas of radon in england and wales. *Health Protection Agency*.
- Nakaya T. (2012). *GWR4 User Manual*. GWR 4 Development Team.

- Nazaroff W. (1992). Radon transport from soil to air. *Rev. Geophys.*, (30), 137–160.
- Nazaroff W.; Moed B.; Sextro R. (1988). Soil as source of indoor radon: generation, migration and entry. *Radon and its Decay Products in Indoor Air*.
- Nero A.; Schwehr M.; Nazaroff W.; Revzan K. (1986). Distribution of airborne radon- 222 concentrations in U.S. homes. *Science*.
- Nissi E.; Sarra A.; Palermi S. (2012). Radon level in dwellings and uranium content in soil in the Abruzzo region: a preliminary investigation by GWR. *Advanced Statistical Methods for the Analysis of large data-sets*.
- Olea R. (1999). *Geostatistics for Engineers and Earth Scientists*. Kluwer Academic Publishers.
- Orlando P. e altri (2000). Il Sistema Informativo Territoriale per la valutazione del Potenziale di Esalazione di Radon dal Suolo. *Rapporto ANPA, serie Stato dell'Ambiente*, **9**, 2000.
- Palermi S. (2010). *A measurement survey aimed to assess the exposure to radon of the population of Abruzzo region*. Tesi non pubblicata per la Scuola di Specializzazione in Fisica Medica. Università La Sapienza, Roma.
- Palermi S. e altri (2012). Indagine per la mappatura del radon in abruzzo. In *V Convegno Nazionale - Novara*.
- Palermi S.; Pasculli A. (2008). Radon mapping in Abruzzo, Italy. In *4th Canadian Conference on Geohazards Québec City Canada, May*.
- Pasculli A.; Palermi S.; Sarra A.; Piacentini T.; Miccadei E. (2013). A novel methodology for the analysis of radon potential by environmental geology and local spatial statistical approach. *sottoposto a Environmental Modelling & Software*.

-
- Plummer M. (2003). *JAGS: A Program for Analysis of Bayesian Graphical Models Using Gibbs Sampling*. Proceedings of the 3rd International Workshop on Distributed Statistical Computing (DSC 2003), Vienna, Austria.
- Plummer M. (2012). *JAGS is Just Another Gibbs Sampler*. url: <http://mcmc-jags.sourceforge.net>.
- Price P. (1997). Predictions and maps of county mean indoor radon concentrations in the mid- atlantic states. *Health Physics*.
- Price P.; Nero A.; Gelman A. (Health Physics). Bayesian prediction of mean indoor radon concentrations for minnesota counties. 1996.
- R Development Core Team (2009). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. url: <http://www.R-project.org>.
- Raspa G.; Salvi F.; Torri G. (2010). Probability mapping of indoor radon-prone areas using disjunctive kriging. *Radiat Prot Dosimetry*.
- Sarra A.; Nissi E.; Palermi S. (2012). Residential radon concentration in the Abruzzo region (Italy): a different perspective for identifying radon prone areas. *Environmental and Ecological Statistics*, **10**(2), 219–247.
- Schumann R. R.; Gundersen L. (1996). Geologic and climatic controls on the radon emanation coefficient. *Environment International*, **22**, 439–446.
- Servizio Geologico d'Italia e Progetto CARG (2012). *Cartografia Geologica d'Italia alla scala 1:50.000*. Servizio Geologico d'Italia e Progetto CARG. ISPRA.
- Spiegelhalter D. J. e altri (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **64**(4), 583–639.

- Tanner A. (1980). Radon migration in the ground: a supplementary review. *Natural Radiation Environment III*.
- Tanner A. (1994). Measurement and determination of radon source potential, a literature review. *National Institute of Standards and Technology*.
- UNSCEAR, (A cura di) (2000). *Sources and Effects of Ionizing Radiation*. UNSCEAR 2000 Report to the General Assembly, with Scientific Annexes. Vol I: Sources.
- Upton G.; Fingleton B. e altri (1985). *Spatial data analysis by example. Volume 1: Point pattern and quantitative data*. John Wiley & Sons.
- Verdi L.; Weber A.; Stoppa G. (2004). Indoor radon concentration forecasting in south tyrol. *Radiation Protection Dosimetry*.
- Visconti G. (2001). *Fondamenti di fisica e chimica dell'atmosfera*. CUEN.
- Wackernagel H. (1998). *Multivariate Geostatistics*. Springer.
- Webster R.; Oliver M. A. (2007). *Geostatistics for Environmental Scientists*. John Wiley & Sons.
- Zeeb H.; Shannoun F., (A cura di) (2009). *WHO Handbook on indoor radon: a public health perspective*. World Health Organization.

Ringraziamenti

Desidero ringraziare Sergio Palermi per l'infinita disponibilità, la pazienza con cui mi ha seguito e il supporto costante durante tutto il lavoro di tesi. Gli sono profondamente grata anche per aver rinunciato a gran parte del suo tempo libero permettendomi di portare avanti questo lavoro.

Un grazie particolare anche a Piero Di Carlo, la sua positività, l'entusiasmo, la cortesia, sono stati per me elementi essenziali.

Ringrazio Annalina Sarra (Università Gabriele D'Annunzio) per i suggerimenti in materia di GWR e Giovanni Desiderio (ARTA Abruzzo) per il lavoro, faticoso, sulla codifica delle permeabilità del suolo.

Grazie al mio caro amico Sebastiano per le risate e le chiacchierate. E' stato un piacere affrontare tutto questo con te.

Dedico quanto fatto in questi anni a mia madre, a mio padre, convinta di averlo reso orgoglioso di me se solo fosse qui, e a Fabio, l'uomo speciale che ho la fortuna di avere accanto.