

Impatto delle caratteristiche dell'edificio sulla concentrazione del radon indoor. Un'applicazione della regressione quantile ai dati della campagna di monitoraggio in Alto Adige

Pegoretti S.¹, Borgoni R.², Verdi L.³

¹Università degli Studi di Trento - Dipartimento di Fisica - via Sommarive 14, 38050 Povo (TN)

²Dipartimento di Statistica, Università di Milano-Bicocca

³APPA Bolzano – via Amba Alagi 5, 39100 Bolzano, luca.verdi@provincia.bz.it

INTRODUZIONE

Il radon è un gas radioattivo naturalmente presente nel terreno, da cui diffonde con relativa facilità mescolandosi con l'atmosfera. Diverso invece il caso in cui questo gas si accumula in ambienti chiusi. Si parla allora di "radon indoor", le cui ripercussioni sulla salute possono risultare importanti. Basti ricordare che il radon è inquadrato al secondo posto, dopo il fumo, come causa di insorgenza di tumore polmonare.

È noto che la concentrazione del radon indoor presenta delle regolarità da un punto di vista territoriale legate ad aspetti ambientali principalmente fisici e geologici. Tuttavia esiste una larga componente della variabilità di questo fenomeno non spiegabile dalla componente territoriale. Parte di tale variabilità può essere attribuita a fattori di tipo antropico specifici degli edifici. In questo contesto risulta quindi importante disporre di modelli in grado di valutare la potenziale influenza delle caratteristiche proprie dell'edificio sede della misura, quali ad esempio il materiale da costruzione, la qualità degli infissi o il tipo di locale, sul valore misurato.

Solitamente la maggior parte dei modelli statistici proposti a tale fine si basano su un approccio regressivo mirato a valutare l'impatto sulla concentrazione media dell'inquinante (si veda tra gli altri Apte et al (1999), Lèvesque *et al.* (1997)). Questa scelta potrebbe però non essere ottimale principalmente in due situazioni: in primo luogo quando l'obiettivo è valutare il comportamento dei valori estremi della distribuzione dell'inquinante o nel caso di distribuzioni particolarmente asimmetriche. In secondo luogo appare ragionevole pensare, come confermato da molte analisi empiriche, che l'impatto di una caratteristica dell'edificio non sia costante ma dipenda dal livello di concentrazione presente.

Occorre tuttavia osservare che, anche una volta che si controlli per tali fattori, esiste una parte residua di variabilità che non appare ulteriormente spiegabile tramite fattori misurabili e che sembra opportuno quantificare per valutare l'incertezza attribuibile alle previsioni ottenibili tramite analisi di regressione.

In questo lavoro, vengono presentati i risultati ottenuti applicando l'approccio della regressione dei quantili alla campagna di monitoraggio del radon indoor in Alto Adige. Obiettivo principale del lavoro è stato stimare l'impatto di variabili secondarie caratterizzanti l'edificio sede della misura, differenziando l'effetto in funzione del valore di concentrazione. Ciò ha consentito di ottenere modelli di riferimento ("case tipo") rispetto ai quali valutare le "rischiosità" dei diversi profili edificati. L'approccio proposto consente inoltre di quantificare la componente di eterogeneità individuale presente a livello di singolo edificio non attribuibile ai fattori misurabili. La modellistica è stata sviluppata, infine, per controllare l'effetto ambientale locale tramite la codifica di una esplicativa rappresentata dal valore medio delle concentrazioni misurate in intorno spaziali dei diversi siti di misura, in modo da diversificare aree esposte a diverse concentrazioni di radon.

LA CAMPAGNA DI MONITORAGGIO IN ALTO ADIGE: DESCRIZIONE DEL DATASET IMPIEGATO

I dati cui il presente lavoro fa riferimento sono stati raccolti in successive campagne di misura condotte sull'intero territorio altoatesino, in modo da ottenere una copertura spaziale piuttosto omogenea e rappresentativa dell'area geografica in esame. I dati fanno riferimento al valore di attività di radon misurata all'interno degli edifici mediante dosimetri passivi a stato solido, esposti tipicamente per un semestre (estivo o invernale). Al fine di ottenere un valore più affidabile, i dati di concentrazione hanno subito una procedura di correzione per un fenomeno noto come "saturazione da sovrapposizione" (Verdi *et al.*, 2005), la cui influenza si manifesta in maniera sensibile sui valori più elevati, che in ambito preventivo vanno trattati con maggior attenzione.

Ogni singola misura che compone il dataset di riferimento è stata accuratamente georeferenziata ed è accompagnata da una ricca serie di informazioni secondarie codificate in variabili sia di tipo qualitativo che quantitativo, volte alla descrizione del contesto geografico ed edilizio che caratterizza ogni singola misura. Risultano così disponibili informazioni relative alle proprietà del terreno e dell'edificio sede della misura, come piano di esposizione, qualità degli infissi, tipo di materiale da costruzione, tipologia dell'edificio, ecc.

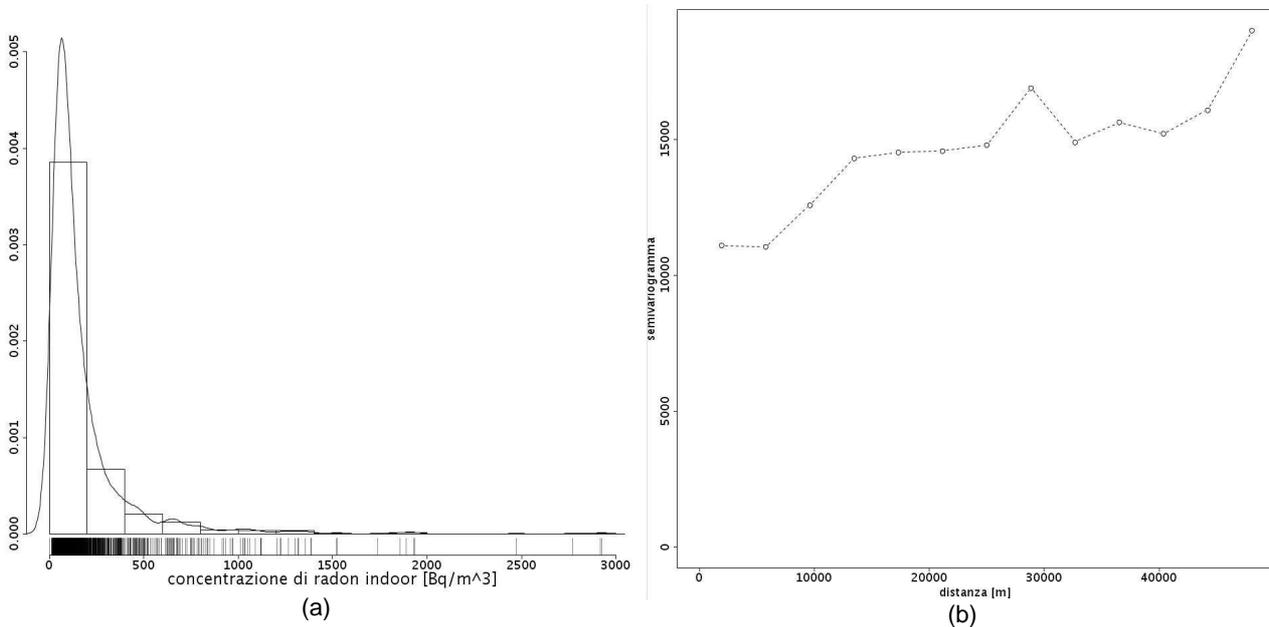
Nello specifico, i dati utilizzati nel presente lavoro fanno riferimento a un valore di concentrazione convertito a media annuale mediante opportuni coefficienti, ricavati sperimentalmente da misure successive condotte per entrambi i semestri. Inoltre, per ragioni di uniformità, si è preventivamente deciso di limitare l'analisi alle sole misure condotte al *piano zero* e nelle *abitazioni*; questa pre-selezione ha portato a una riduzione della numerosità del dataset operativo a 1523 valori di concentrazione. Infine, ogni singola misura è georeferenziata e porta informazioni relative al *contatto* dell'abitazione con il terreno e al *tipo di materiale* da costruzione. La tabella 1 riporta alcune statistiche di sintesi relative alla concentrazione di attività di radon indoor per il dataset totale e ripartite per le classi delle due covariate utilizzate nella costruzione dei modelli descritti in seguito.

Tabella 1: statistiche di sintesi relative alla concentrazione di attività di radon indoor

	<i>min</i>	15%	25%	50%	75%	85%	<i>max</i>	<i>media</i>	<i>s.q.m.</i>	<i>n (%)</i>
Totale	2	42	54	92	183	281	2924	177	262	1523
contatto = sì	12	58	79	146	289	461	2924	263	341	582 (38%)
contatto = no	2	37	46	72	133	193	2918	123	176	941 (62%)
materiale = sassi	12	52	73	131	250	430	2918	235	227	481 (32%)
materiale = altro	2	39	49	78	156	234	2924	150	316	1042 (68%)

La figura 1-(a) mostra la distribuzione della concentrazione del radon indoor mentre, in figura 1-(b) è riportato il variogramma empirico per la variabile di studio. Si nota, in particolare, la presenza di un'elevata componente spaziale non strutturata che suggerisce come i fattori territoriali spieghino solo una parte della variabilità del fenomeno.

Figura 1: (a) Istogramma sperimentale e stima a kernel della p.d.f. corrispondente. (b) Variogramma sperimentale per il valore di concentrazione di radon indoor.



LA REGRESSIONE QUANTILE

La regressione quantile è stata introdotta da Koenker e Bassett (1978) in contesti di tipo economico. Sia X un insieme di variabili secondarie e Y la variabile di risposta rappresentante la concentrazione del radon indoor. L'obiettivo dell'analisi statistica è quello di valutare la relazione fra le variabili X e la variabile di risposta e quantificarla numericamente.

Il modello di regressione standard è definito dalla relazione $Y = \mu(x, \beta) + \varepsilon$ dove β è un vettore di parametri incogniti e ε è un errore stocastico a media 0 e varianza costante. In altri termini il modello descrive come la media condizionata di Y , $E(Y | X = x)$, si modifica in funzione del set di variabili. Spesso per

la media condizionata viene assunta un'espressione lineare nei parametri incogniti, $\mu(x,\beta) = x'\beta$, stimati, in contesti standard, tramite minimi quadrati.

Nel caso di una variabile continua Y , il quantile di ordine τ , $0 < \tau < 1$, è definito come quel valore ξ_τ , che soddisfa $F(\xi_\tau) = P(Y \leq \xi_\tau) = \tau$, dove $F(y)$ rappresenta la funzione di ripartizione di Y . La regressione quantile è una metodologia statistica che specifica la relazione fra i quantili della distribuzione condizionata della variabile dipendente, $Q_{Y|X}(\tau | x)$, e l'insieme delle variabili secondarie X tramite una relazione $\xi(x,\beta)$ con β vettore di parametri incognito. Considerato un campione di n osservazioni, i parametri possono essere stimati, risolvendo un problema di ottimo analogo a quello dei minimi quadrati, come

$$\sum_{i=1}^n \rho_\tau(y_i - \xi(x_i, \beta))$$

dove $\rho_\tau(u)$, nota come *check function*, è data da $\rho_\tau(u) = u \cdot (\tau - I(u < 0))$ e $I(A)$ è la funzione indicatrice tale che $I(A) = 1$ se A è verificata e 0 altrimenti. In particolare in questo lavoro si considera il caso in cui $\xi(x,\beta)$ è una funzione lineare dei parametri, $\xi(x,\beta) = x'\beta$. In questo caso il problema di minimo precedente può essere convenientemente risolto tramite metodi di programmazione lineare (Koenker, 2005).

I parametri della regressione quantile possono essere interpretati in modo analogo a quelli dei più usuali modelli di regressione lineare. Il singolo parametro rappresenta la variazione nel quantile specificato dovuta ad una variazione unitaria nella corrispondente variabile esplicativa. Nel caso di variabili binarie, ampiamente utilizzate nel presente lavoro, l'interpretazione è analoga: il parametro rappresenta la variazione nel quantile condizionato ottenuta passando da una delle due categorie, assunta come riferimento, all'altra. Si può osservare che l'insieme dei parametri è specifico del quantile condizionato di volta in volta considerato ovvero $Q_{Y|X}(\tau | x) = x'\beta(\tau)$. In questo modo la regressione permette di valutare un effetto differenziato per la medesima variabile a seconda del livello (quantile) a cui viene considerate la variabile di risposta.

RISULTATI

In questo paragrafo riportiamo i risultati dell'analisi eseguita. I parametri di edificio considerati sono stati *materiale* e *contatto* nella forma binaria descritta nella tabella 1. Al fine di poter valutare effetti differenziati nei quattro possibili profili abitativi è stato inserito un termine d'interazione fra i due fattori. La componente ambientale è stata controllata nei modelli stimati definendo per ogni edificio i la variabile

$$\tilde{y}_i = \frac{1}{m_i} \sum_{j \in \partial_i} y_j \text{ dove } \partial_i \text{ rappresenta il vicinaggio dell'edificio } i\text{-esimo e } m_i \text{ il numero di siti presenti nell'intorno}$$

∂_i . Il vicinato è definito come l'insieme dei siti che distano da i non più di 7 km. Tale variabile esplicativa rappresenta indirettamente potenziali fattori ambientali locali che possono influenzare la concentrazione del radon indoor. Tale approccio è quindi, nel principio, vicino all'autoregressione quantile descritta nel caso di dati longitudinali da Koenker (2005) estesa al contesto di serie spaziali di dati.

La figura 2 mostra come l'effetto delle covariate si modifichi con il livello di concentrazione. Si nota come l'effetto del contatto con il suolo e dell'uso del sasso come materiale di costruzione sia positivo, cioè tali caratteristiche favoriscono una concentrazione più alta, e tenda a essere progressivamente più marcato a concentrazioni progressivamente più alte (grafico A e B). Il termine d'interazione, che rappresenta un effetto differenziato per gli edifici in sasso a contatto con il terreno, ha lo stesso tipo di andamento pur non essendo significativo su tutto il range dei quantili considerati (figura 2-C). Appare, infine, un moderato, sebbene significativo, effetto della media del vicinato, inclusa nel modello in forma lineare (figura 2-D), ad indicare il fatto che esiste una componente locale ambientale influente sulla concentrazione del radon indoor.

La metodologia utilizzata consente inoltre di quantificare l'eterogeneità non osservata (cioè non spiegabile tramite i fattori misurabili considerati nel modello) propria dell'edificio per ciascun profilo costruttivo. Tale eterogeneità può essere quantificata tramite un intervallo di probabilità prefissata ottenuto per differenza dai quantili previsti dal modello per ognuno dei quattro profili costruttivi specificati in tabella 1. In particolare si è considerato l'intervallo di probabilità identificato dal primo e quarto quintile (probabilità associata all'intervallo pari a 0.6). La figura 3 riporta tali ampiezze per i quattro profili abitativi per due diversi livelli di concentrazione media del vicinaggio: 75 Bq/m^3 (valore della media nazionale) e 170 Bq/m^3 (valore medio dell'Alto Adige). Si nota come tale componente di eterogeneità aumenta per i profili abitativi caratterizzati da concentrazioni più elevate e all'aumentare della "rischiosità" dell'area stimata tramite la media del vicinaggio.

La capacità predittiva della modellistica statistica diventa, in queste situazioni, ovviamente inferiore. Lo stesso grafico riporta anche gli intervalli quantile al 60% per altri due modelli considerati: il primo esclude l'effetto dei vicini e il secondo esclude, oltre al vicinaggio, anche il termine d'interazione. Si evidenzia come,

controllando per l'effetto ambientale, la capacità predittiva migliori sensibilmente (cioè l'intervallo è più stretto) suggerendo che tale fattore risulta essere importante nello spiegare la complessa variabilità del fenomeno. Come già suggerito dal grafico di figura 2, invece, differenziare i profili costruttivi tramite un termine d'interazione, appare poco rilevante e con impatto differenziato per i diversi profili abitativi.

Figura 2: Coefficienti di regressioni per le variabili secondarie considerate al variare dell'ordine del quantile. La fascia ombreggiata rappresenta l'intervallo di confidenza al 95% del parametro. La linee orizzontali rappresentano la stima tramite minimi quadrati del parametro e il relativo intervallo di confidenza.

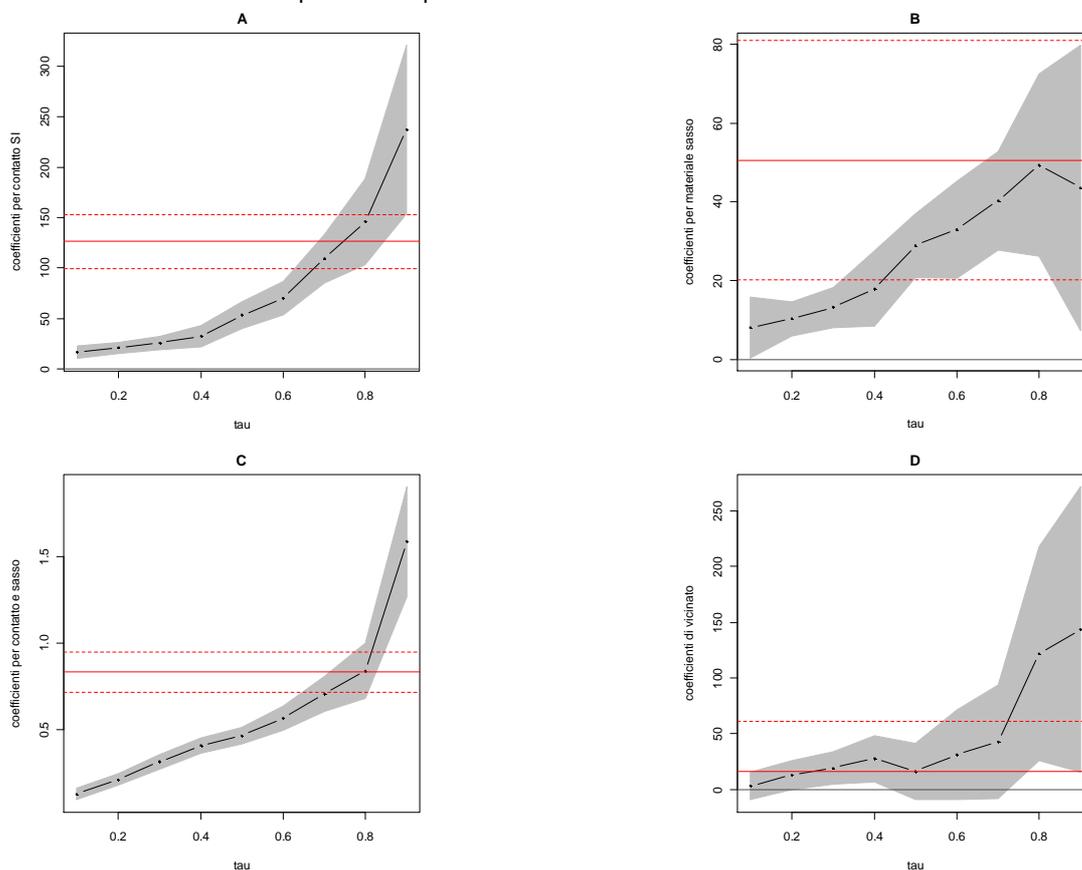
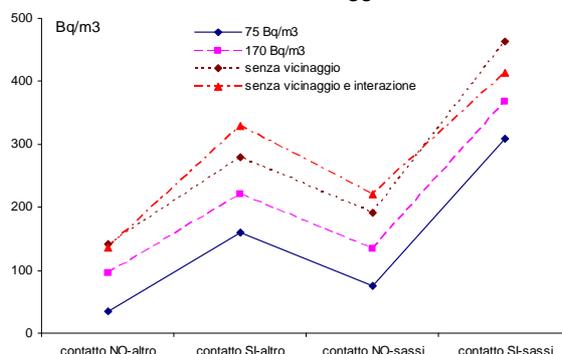


Figura 3: intervalli percentili al 60% per diversi profili abitativi. 75 Bq/m³ e 170 Bq/m³ rappresentano valori prefissati della media del vicinaggio



BIBLIOGRAFIA

- Apte MG, Price PN, Nero AV, Revzan KL (1999) Predicting New Hampshire indoor radon concentrations from geologic information and other covariates *Environmental Geology* 37, 181-194
- Koenker R, Bassett GS, (1978) Regression quantiles. *Econometrica* 46, 33-50
- Koenker R, (2005): *Quantile Regression*. Cambridge University Press.
- Lèvesque B, Gauvin D, McGregor RG, Martel R, Gingras S, Dontigny A, Walker W, Lajoie P, Létourneau E. 1997. Radon in residences: influences of geological and housing characteristics. *Health Physics* 72: 907-914.

Verdi L., Pegoretti S. e Ferrarini M., *Saturation on LR-115 Radon Detectors: Potential and Real Effects on Radon Mapping*, in International Conference, 4th Dresden Symposium: Survey of Geo-Hazards, 26–30 settembre 2005, Dresda, Germania.