

SCUOLA DI INGEGNERIA E ARCHITETTURA

Dipartimento di Ingegneria Civile, Chimica, Ambientale e dei Materiali - DICAM

*CORSO DI LAUREA MAGISTRALE IN INGEGNERIA PER L'AMBIENTE E IL
TERRITORIO*

TESI DI LAUREA

in
Gestione delle Risorse Idriche M

STUDIO DEGLI EFFETTI DI VARIABILI IDROLOGICHE E GEOMORFICHE
SULL'ESPANSIONE URBANA NEL NORD-EST DELLA CINA

CANDIDATO
Grandi Giulia

RELATORE:
Chiar.ma Prof. Ceola Serena

CORRELATORE:
Chiar.mo Prof. Bertuzzo Enrico

Anno Accademico 2017/2018

Sessione III

SOMMARIO

Introduzione.....	6
1 Perché la Cina: contesto sociale e culturale, inquadramento geografico e climatico ...	9
1.1 La crescita.....	11
1.2 La geografia e il clima della Cina	12
1.2.1 Morfologia e idrografia	12
1.2.2 Clima	15
1.3 L'area di studio.....	18
2 I dati a disposizione e quelli di interesse per lo studio.....	20
2.1 La scelta della variabile da modellizzare	20
2.1.1 Dati di distribuzione e densità di popolazione.....	20
2.1.2 I dati di luminosità notturna.....	23
2.2 La scelta dei parametri di influenza.....	26
2.2.1 Digital Elevation Model (DEM).....	27
2.2.2 Distanza dalla costa.....	29
2.2.3 Distanza dai corsi d'acqua.....	32
2.2.4 Temperatura	34
2.2.5 Piovosità	37
2.2.6 Land Cover	39
3 Elaborazione preliminare dei dati per l'individuazione delle caratteristiche del modello di espansione urbana	41
3.1 La definizione di classi di parametri per lo studio delle luci notturne	41
3.1.1 Elevazione: suddivisione in classi.....	42
3.1.2 Distanza dalla costa: suddivisione in classi	46
3.1.3 Distanza dai corsi d'acqua: suddivisione in classi	49
3.1.4 Temperatura: suddivisione in classi.....	52
3.1.5 Piovosità: suddivisione in classi.....	56
3.1.6 Copertura del terreno: suddivisione in classi	59
3.2 Ricerca di eventuali correlazioni tra i fattori ambientali considerati	60
4 Realizzazione del modello matematico di espansione urbana	63
4.1 Il modello teorico	63
4.2 Lo script per la calibrazione numerica e la simulazione di anni futuri.....	72
4.2.1 La creazione di una maschera territoriale per il calcolo della probabilità di urbanizzazione	73

4.2.2	La ricerca di un valore di soglia per la rappresentazione degli stati “urbanizzato” e “non urbanizzato”.....	77
4.2.3	L’algoritmo	78
4.3	Risultati e considerazioni	88
	Conclusioni e prospettive future.....	95
	Indice delle Figure e delle Tabelle.....	100
	Ringraziamenti.....	106
	Riferimenti e Bibliografia	107

***“Everything is related to everything else,
but near things are more related than distant things”***

- Waldo Tobler, first law of geography¹.

Introduzione

La popolazione mondiale oggi supera i 7 miliardi e mezzo di persone, ed è destinata a crescere nei prossimi anni fino a raggiungere i 9,8 miliardi nel 2050, secondo le più recenti stime ONU (2017)². Il pattern spaziale della popolazione non appare uniformemente distribuito sulla superficie terrestre ma risulta essere concentrato in specifiche aree in cui la densità è molto elevata e l'urbanizzazione estrema, a contrasto con altre zone che appaiono quasi disabitate. Uno studio del 2003 ad opera di Conservation International³ evidenzia infatti che il 97,6% della popolazione mondiale è concentrata in poco più della metà della superficie disponibile sul pianeta; l'indagine sulle cause di questa distribuzione non omogenea e la modellizzazione del fenomeno potrebbero portare alla definizione della struttura degli insediamenti umani futuri e alla previsione degli effetti antropici sull'ambiente.

L'uomo, infatti, interagisce da sempre con il territorio in cui si insedia, modificandone pesantemente l'aspetto e le caratteristiche; la disponibilità di risorse naturali e le criticità ambientali conseguenti da inquinamento e deforestazione sono solo alcune ragioni per cui oggi è necessario provare a prevedere dove la popolazione si distribuirà negli anni a venire, soprattutto in un'ottica di cambiamento climatico globale. Evidentemente, il processo di urbanizzazione provoca localmente una trasformazione non solo dell'aspetto della superficie terrestre ma anche del clima, della vegetazione, delle riserve idriche e delle componenti dell'atmosfera. Le previsioni di espansione urbana servirebbero quindi come supporto alla definizione dell'impatto umano sull'ambiente; per esempio, al fine di valutare la Human Footprint dei prossimi anni non è necessario solo conoscere quanto la popolazione aumenterà, ma anche dove e con quale modalità. Lo sviluppo di una grande città porterebbe con sé la costruzione di infrastrutture con un peso maggiore sull'ambiente rispetto alla crescita di insediamenti in aree poco sviluppate o, ancora, una crescita localizzata graverebbe di più sulle risorse rispetto a una crescita più uniformemente distribuita.

Così come l'uomo influenza il clima dell'area in cui si stabilisce fin dalla preistoria⁴, arrivando a provocare i cambiamenti che osserviamo attualmente, il clima ha effetti sull'uomo e ne influenza evidentemente la distribuzione, non a caso le grandi aree disabitate del pianeta sono zone desertiche o ricoperte di ghiaccio. È importante chiedersi come reagirà la popolazione al verificarsi dei grandi cambiamenti ambientali dovuti all'effetto del riscaldamento globale: aumento generale delle temperature e dei fenomeni estremi, variazioni del livello dei mari e della vegetazione. Identificare fattori ambientali che influenzano l'urbanizzazione, e in che misura, può portare anche alla previsione di come cambieranno gli insediamenti se uno di tali fattori dovesse subire una variazione. Un modello analitico in grado di simulare e prevedere il processo di urbanizzazione potrebbe essere utilizzato

per molteplici applicazioni multidisciplinari: previsione dell'impatto antropico sull'ambiente, analisi di migrazioni derivate dal peggioramento delle condizioni ambientali in una determinata area, valutazione del rischio idrogeologico, stime di benessere e crescita economica.

In questo elaborato di tesi, si è cercato quindi di individuare se esistano delle forzanti ambientali che governano il processo di urbanizzazione e in che misura esse potrebbero influenzare la distribuzione di popolazione costruendo un modello stocastico spazialmente esplicito in grado di descrivere la crescita urbana in una regione nel nord-est della Cina. Il primo step è stato quindi quello di definire un'area di studio limitata, per rendere i calcoli meno onerosi, ma significativa. Per via della grande variabilità geografica e climatica e della immensa crescita di popolazione verificatasi negli ultimi 30 anni, l'area adeguata allo studio è stata appunto individuata nel nord-est della Cina, concentrandosi su un quadrato di circa 1000 x 1000 km² nei dintorni della città di Pechino.

Per ricostruire la crescita urbana dell'area di studio nel recente passato e successivamente simulare uno scenario futuro, sono state utilizzate immagini satellitari di luminosità artificiale notturna, quali indicatori dell'area urbanizzata, disponibili dall'anno 1992 al 2013. Come forzanti ambientali sono stati considerati fattori idrologici, geomorfici e climatici nella forma di elevazione, distanza dalla costa e dai corsi d'acqua, valori medi annuali di piovosità e temperatura e tipologia di copertura del suolo. Si è ipotizzato che tali forzanti ambientali potessero condizionare il fenomeno della crescita urbana, e che quindi potessero influenzare la distribuzione di popolazione, soprattutto come già evidenziato in studi precedenti (i.e. C. Small, 2003⁵). Si può ritenere, infatti, che il processo di urbanizzazione tenda a svilupparsi principalmente lungo i corsi d'acqua o la linea di costa, dove sono favorevoli le vie di scambio o il reperimento di materie prime, o in zone abbastanza pianeggianti, dove è più facile realizzare infrastrutture, o ancora dove il clima favorisca la vita e le attività umane.

Consultando diverse banche dati, come quella del JRC⁶, dell'USGS HydroSHEDS⁷, del SEDAC⁸ e di Google Earth Engine⁹, sono stati reperiti i dati sotto forma di mappe GIS, successivamente elaborate in ambiente numerico per analizzare eventuali relazioni con i dati di luminosità notturna.

Lo studio ha rilevato una forte dipendenza delle luci artificiali con le forzanti identificate. Tale relazione è stata impiegata nella creazione di un modello stocastico spazialmente esplicito in grado di definire la probabilità di urbanizzazione, ossia la probabilità che un pixel non urbanizzato si trasformi in urbanizzato in un definito intervallo di tempo. Un ruolo fondamentale nel definire questa probabilità sarà da imputare anche alla vicinanza con punti dell'area che risultano già urbanizzati, secondo il principio della geografia di W. Tobler per cui oggetti vicini si influenzano maggiormente tra loro rispetto ai punti più lontani¹⁰.

Il presente elaborato di tesi è strutturato nel seguente modo: nel primo capitolo, si analizzeranno le peculiarità ambientali e sociali della Cina e in particolare dell'area di studio selezionata mentre nel

secondo capitolo saranno esposte per esteso le scelte legate ai dati utilizzati; nel terzo capitolo si esamineranno poi le elaborazioni preliminari per la scelta delle funzioni inserite nel modello, spiegato infine nel quarto capitolo.

Per riassumere, si è cercato di rispondere a domande quali: *perché la popolazione si distribuisce preferenzialmente in alcune aree piuttosto che in altre? I processi di cambiamento di popolazione e di densità sono influenzati da fattori ambientali, oltre che sociali? E, infine, Si può modellizzare questo processo in modo da prevedere anche andamenti futuri?*

1 Perché la Cina: contesto sociale e culturale, inquadramento geografico e climatico

Per poter definire un modello che descriva la distribuzione della popolazione e il suo comportamento spazio-temporale, è stato necessario individuare un'area che avesse assistito, negli anni, a una forte crescita demografica. Dato che si utilizzeranno, sia come variabile da modellizzare sia come parametri del modello, immagini acquisite con telerilevamento, gli anni da esaminare sono quelli per cui sono disponibili fotografie satellitari. Dunque, il problema preliminare è quello di individuare un'area i cui si sia verificata una forte crescita di popolazione negli ultimi anni del XX secolo.

Osservando le mappe della Global Population Density Grid, prodotta dal SEDAC¹¹, in cui è riportata la densità di popolazione a livello globale, si notano diverse aree in cui essa è aumentata tra l'anno 1970 e il 2000 (Figura 1.1). In particolare, le variazioni maggiori si verificano nei paesi del BRICS (Brasile, Russia, India, Cina, Sudafrica) che, come è noto, sono stati (e sono tuttora) protagonisti di una forte crescita economica e di popolazione.

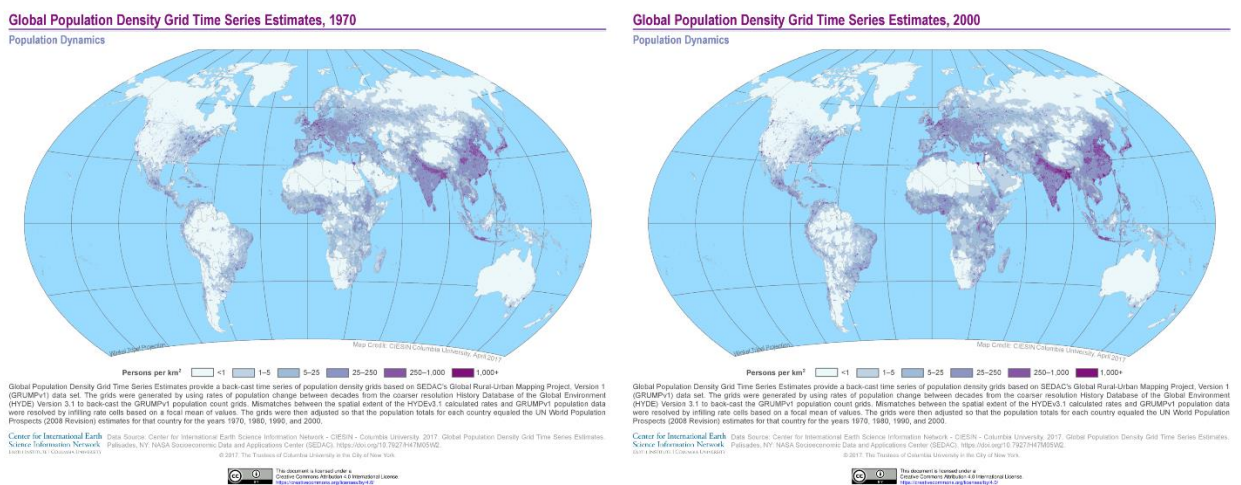


Figura 1.1: Variazione della densità di popolazione mondiale dall'anno 1970 al 2000. GPW prodotta dal SEDAC, NASA.

Ancora più esplicative sono le immagini satellitari di luminosità artificiale notturna del Defense Meteorological Program (DMSP) Operational Line-Scan System (OLS)¹², ovvero le fotografie satellitari che mostrano come appare la superficie terrestre durante la notte (Figura 1.2). Tramite le tali immagini, è possibile individuare, da un anno al successivo, le nuove aree antropizzate o l'aumento di luminosità di quelle già presenti. L'aumento di luminosità può essere collegato sia alla

crescita di urbanizzazione, e quindi anche alla crescita economica del paese, sia all'aumento di popolazione o di densità di popolazione.

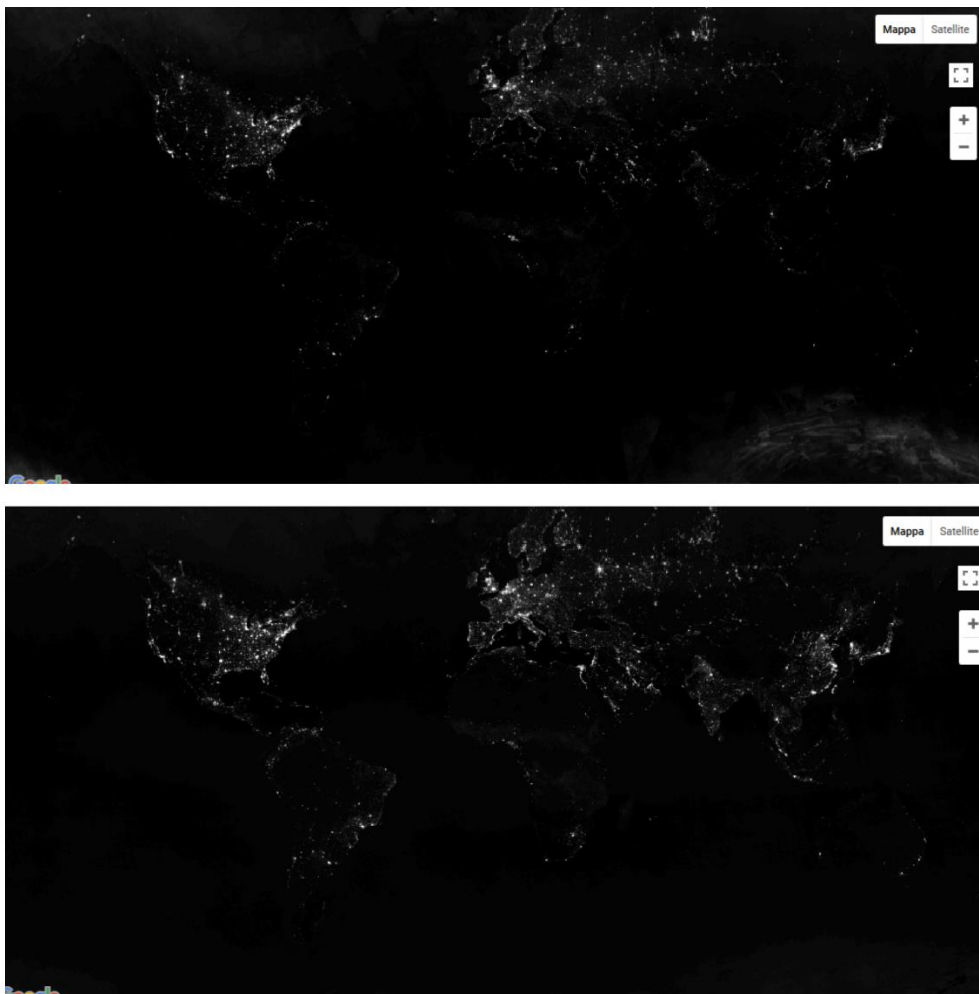


Figura 1.2: confronto tra le fotografie di luci notturne nel 1992 e nel 2014. Fonte: database di Google Earth Engine.

Tra il 1992 e il 2014 (estremi di disponibilità della serie) si nota la nascita di nuove luci nel Sud America e in Medio Oriente, si osserva uno spiccato incremento di luminosità in Europa e America del Nord ma, al di sopra di tutto, risulta evidente il salto compiuto dall'Asia. India, Cina e Sud Est Asiatico sono protagonisti di un passaggio dalla quasi assenza di luci a una luminosità altissima in poco più di vent'anni.

Per questo motivo, data anche la disponibilità di altri dati di interesse e la vastità del continente, che comporta una morfologia variegata e la compresenza di climi molto diversi tra loro, lo studio si è focalizzato sulla Cina, in particolare nell'area del Nord-Est nei dintorni di Pechino.

I motivi che rendono la Cina il territorio più adeguato ad affrontare questa analisi, sono esposti per esteso nei paragrafi seguenti.

1.1 La crescita

Dal 2011, la Cina ha raggiunto il secondo posto nell'economia mondiale (dietro gli Stati Uniti), con tassi annuali di crescita del PIL di oltre il 10%, e continua a crescere, sebbene a velocità più ridotta¹³. Per rendere meglio l'idea dell'impresa cinese, questi obiettivi sono stati raggiunti a partire dalla fine del 1970, rendendo il Paese una superpotenza emergente in pochi decenni¹⁴; prima di allora, la Cina si presentava come un territorio arretrato economicamente e tecnologicamente, chiuso verso gli altri paesi e basato su una società rurale.

A partire dal 1978, il Governo propone diverse riforme che aprono il Paese all'occidentalizzazione e ha inizio la crescita. Lo sviluppo economico porta la percentuale di poveri a diminuire dal 53% nel 1981 all'8% nel 2001¹⁵, mentre la popolazione passa da poco meno di un miliardo di abitanti del 1978 agli oltre 1,3 miliardi attuali¹⁶.

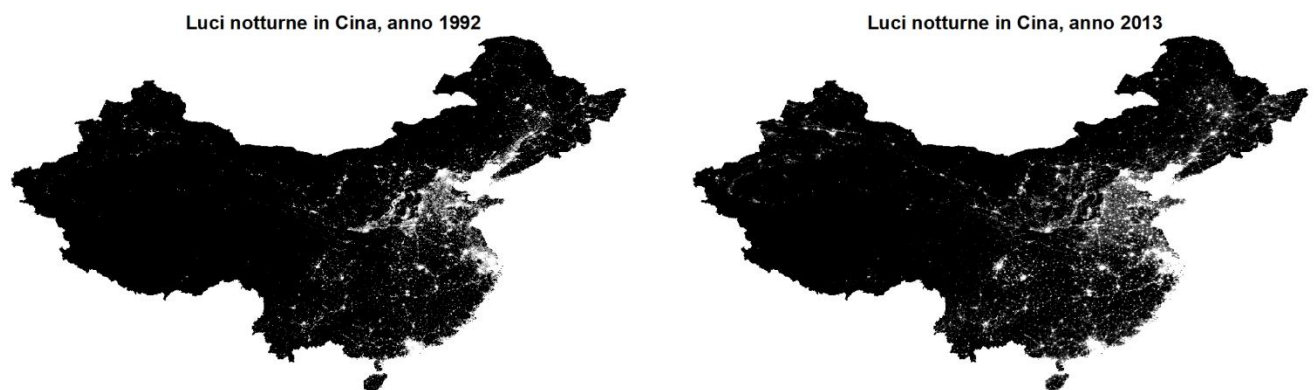


Figura 1.3: Variazione delle luci notturne in Cina tra il 1992 e il 2013, dati DMSP-OLS forniti dal SEDAC (NASA) ed elaborati in ambiente GIS.

La crescita di popolazione ed economia implica un aumento delle infrastrutture, dei trasporti (soprattutto autostrade e ferrovie veloci), del numero di abitazioni e, dunque, l'intensità complessiva delle luci notturne cresce notevolmente (Figura 1.3). Dalle immagini satellitari, tuttavia, si osserva una distribuzione della popolazione e delle luci non omogenea; le città più importanti, le luci e le vie di comunicazione principali sono concentrate ad est lungo la costa, mentre ad ovest la luminosità è quasi assente. Secondo la tesi qui esposta, i motivi sono da ricercare soprattutto nella geografia e nel clima del Paese, come si esaminerà nei paragrafi successivi.

È bene notare che anche questo aspetto di variazione spaziale può essere utile per la costruzione del modello, poiché un territorio quanto più vario morfologicamente, climaticamente e anche per

densità di popolazione permetterà di osservare meglio come i parametri influenzino la variabile da modellizzare.

1.2 La geografia e il clima della Cina

Quello della Cina è, infatti, uno dei territori che presenta maggior variabilità geografica e climatica, data anche l'ampiezza della sua superficie, che raggiunge quasi i 10 milioni di km².

Il Paese si estende per circa 5.250 km in direzione Est-Ovest e 5.500 km in direzione Nord-Sud, coprendo una grande distanza in entrambe le direzioni. Questa notevole estensione lungo entrambi gli assi definisce un territorio molto vario per morfologia, idrografia, vegetazione e, soprattutto, clima. Si passa dalle alte montagne al confine con il Nepal, a ovest, fino ai quasi 14.000 km di costa, a est, che danno origine a zone desertiche con climi semiaridi nell'interno o zone boschive molto umide, ricche di grandi laghi, nelle pianure a contatto col mare¹⁷.

Di seguito, si esamineranno le peculiarità geografiche del Paese (morfologia, idrografia, vicinanza alla costa) e si analizzeranno le principali aree climatiche e i fattori che le influenzano (copertura vegetale, precipitazione, temperatura e umidità).

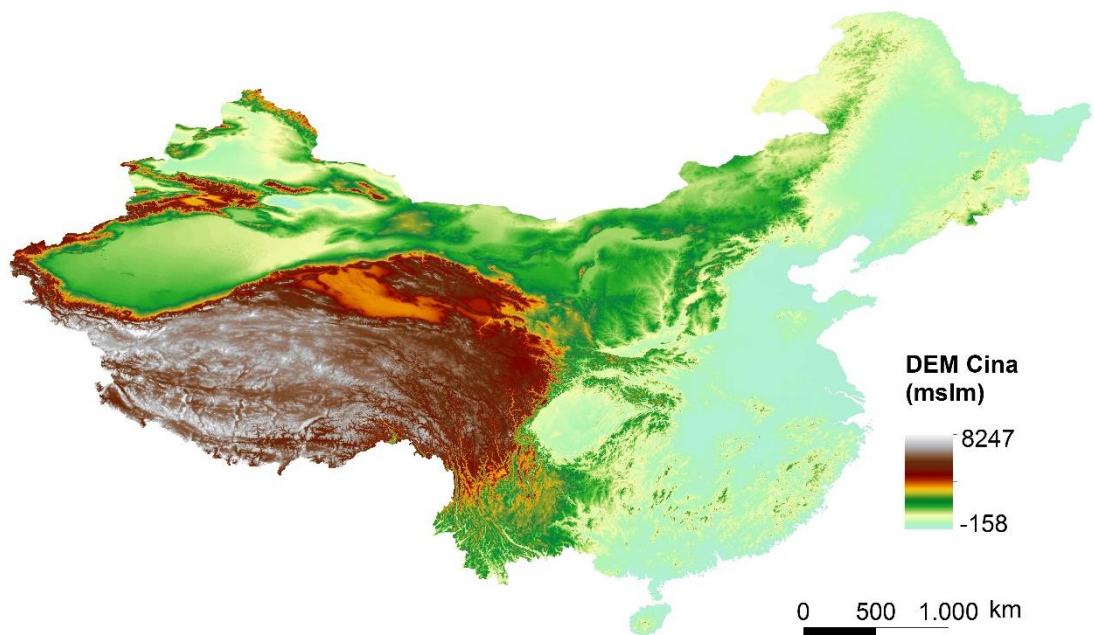
1.2.1 Morfologia e idrografia

Come anticipato, il territorio cinese presenta diverse zone ad altitudini differenti; l'85% dell'area totale, tuttavia, supera i 500 m.s.l.m. mentre un terzo del territorio è occupato da montagne, possiamo dunque intenderlo come un Paese prevalentemente collinare-montuoso¹⁸.

Sul suolo cinese (al confine con Nepal e Pakistan) troviamo le cime più alte del mondo appartenenti alla catena montuosa dell'Himalaya: Everest, K2, Lhotse, Makalu sono tutti rilievi che superano un'altitudine di 8.500 m.s.l.m.; complessivamente, l'altopiano del Tibet risulta l'area più elevata del Paese e del mondo, con un'altitudine media di 4.000 m.s.l.m. Le vette più elevate si trovano dunque a Sud-Ovest, mentre le quote decrescono spostandosi a Est, fino a raggiungere il livello del mare. Questa variazione di altitudine determina forme morfologiche altrettanto varie: montagne, altopiani, colline, pianure, bacini e deserti fino ai numerosi chilometri di costa.

La varietà di quote può essere osservata in Figura 1.4, si passa dagli 8.850 m di altitudine del monte Everest a Ovest, fino ai -155 m del lago Ayding a Nord.

La lunga costa cinese, le grandi catene montuose, la presenza di laghi e foreste, ne influenzano il clima che risulterà a sua volta piuttosto variegato.



**Figura 1.4: Andamento digitale delle quote del terreno con risoluzione di 30" di angolo (cioè 1 km lineare).
Fonte: USGS HydroSHEDS.**

Tra i fattori in gioco, la presenza della costa è fondamentale per definire il clima cinese; il mare, infatti, influenza l'escursione termica, riducendo le differenze di temperature tra giorno e notte e tra estate e inverno, rendendo le temperature mediamente più miti e più favorevoli all'insediamento umano. Anche le vette elevate dell'Himalaya, influenzando i movimenti di venti e masse d'aria, hanno effetti sulla temperatura; inoltre, bloccando le precipitazioni, danno luogo ad aree desertiche alle loro pendici.

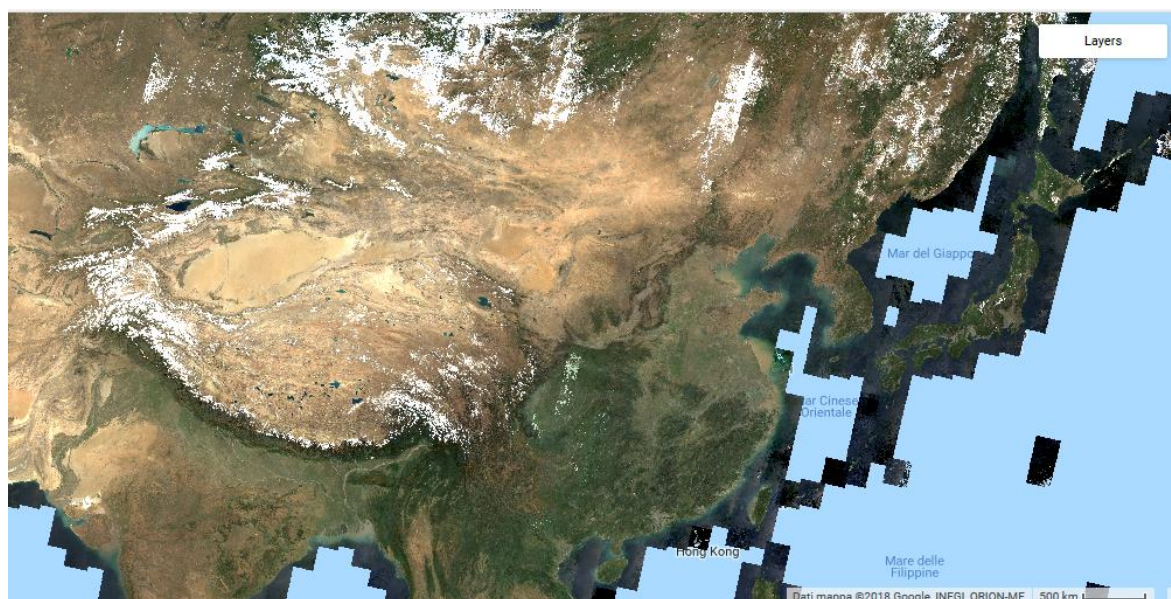


Figura 1.5: Immagine satellitare del territorio cinese ottenuta dal sistema di Landsat 8 tramite letture di riflettanza della superficie. Fonte: dataset di Google Earth Engine.

Osservando le immagini satellitari della superficie terrestre (in Figura 1.5 è riportata quella di Landsat 8¹⁹) è possibile osservare la diversificazione con cui si presenta il suolo cinese: ampie zone desertiche a nord-ovest si contrappongono alle aree verdi del sud-est, le alte catene montuose a sud-ovest declinano fino alle coste, che risultano piuttosto frastagliate a nord e più lineari a sud.

Un altro parametro importante per capire la morfologia del Paese è la distribuzione della vegetazione; in Figura 1.6 è possibile osservare in dettaglio la copertura del terreno, ottenuta dall'elaborazione delle immagini di MODIS²⁰.

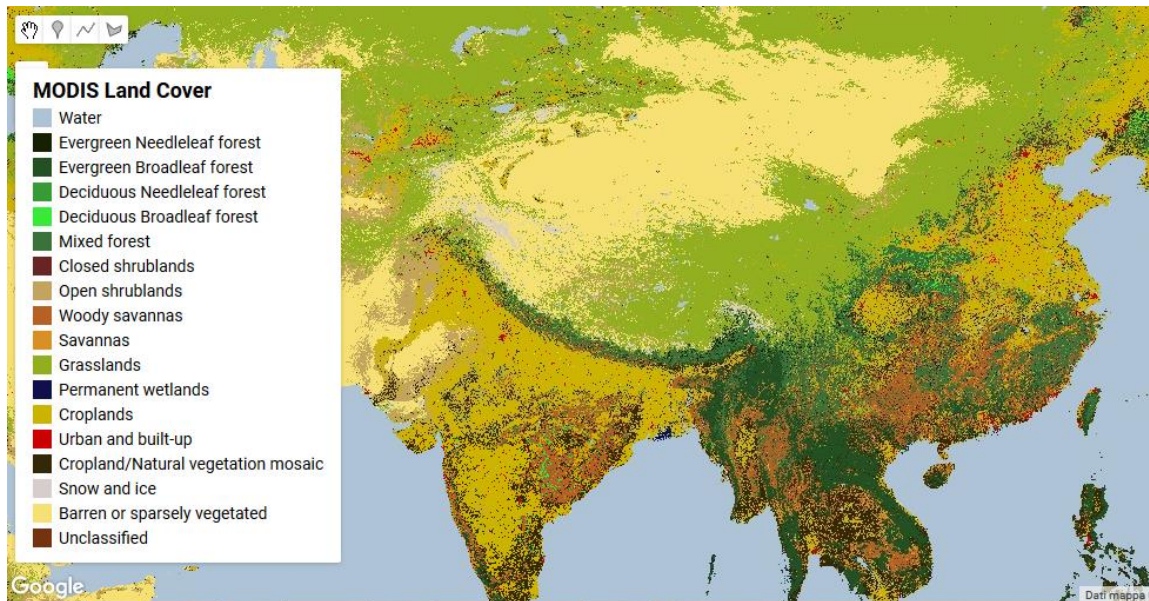


Figura 1.6: MODIS Land Cover Map, mappa di copertura vegetale per il territorio cinese prodotta a partire dalle fotografie di MODIS. Fonte: dataset di Google Earth Engine.

Si nota ancora una volta una distribuzione di vegetazione eterogenea e composta da specie molto diverse: mentre lungo la catena montuosa dell'Himalaya prevalgono foreste perenni di latifoglie o conifere, man mano che l'altitudine diminuisce verso est compaiono foreste decidue tipiche delle zone temperate. Il comportamento è diverso se si procede verso nord, qui abbiamo il rapido passaggio da foreste sempreverdi a praterie e, in seguito, a zone semi-desertiche sterili in cui si presenta solo vegetazione sparsa. La fertilità del suolo aumenta, invece, verso la costa, e non è un caso che proprio in vicinanza del mare si trovino le più vaste aree coltivate del Paese.

Se è vero che la copertura vegetale è influenzata dai fattori climatici, si possono immaginare le diverse aree climatiche presenti nel territorio già osservando questi primi dati; si può presupporre che l'umidità aumenti dove compare maggiore copertura vegetale di tipo forestale o agricola, mentre dove la vegetazione risulta quasi assente si verificheranno zone più aride e meno piovose.

Anche l'idrografia può essere un fattore di influenza per l'insediamento umano. La vicinanza a corsi d'acqua può, infatti, determinare lo sviluppo di un'area agricola o industriale, ma, soprattutto se

navigabile, può anche rappresentare una importante via di collegamento con la costa o altri centri abitati. In Figura 1.7 sono evidenziati i numerosissimi corsi d'acqua della Cina ottenuti dal database di HydroSHEDS²¹, i quali si presentano come una fitta rete di canali, perenni o effimeri. I tre fiumi principali, il Fiume Giallo, il Fiume Azzurro e lo Xi, hanno tutti direzione ovest-est e sfociano quindi in mare, comportamento comune al 50% delle acque totali, che segue infatti l'andamento delle quote del terreno²². Un cospicuo numero di corsi d'acqua ha, invece, un percorso endoreico e non raggiunge la costa descrivendo invece traiettorie chiuse nell'entroterra che portano le acque a disperdersi per evaporazione o infiltrazione, motivo per cui si formano ampie aree desertiche. In questi territori, i fiumi non riescono a drenare adeguatamente la superficie per via dell'aridità del terreno e delle scarse precipitazioni, le loro acque formano quindi estese riserve sotterranee o evaporano prima di raggiungere la costa.

Sono presenti anche numerosi laghi, sia nelle regioni aride e calde del nord-ovest, dove l'affioramento di acque sotterranee forma numerosi bacini con forte salinità²³, sia nelle zone montuose e umide del sud-est ma anche lungo i rilievi più acclivi dell'ovest²⁴.

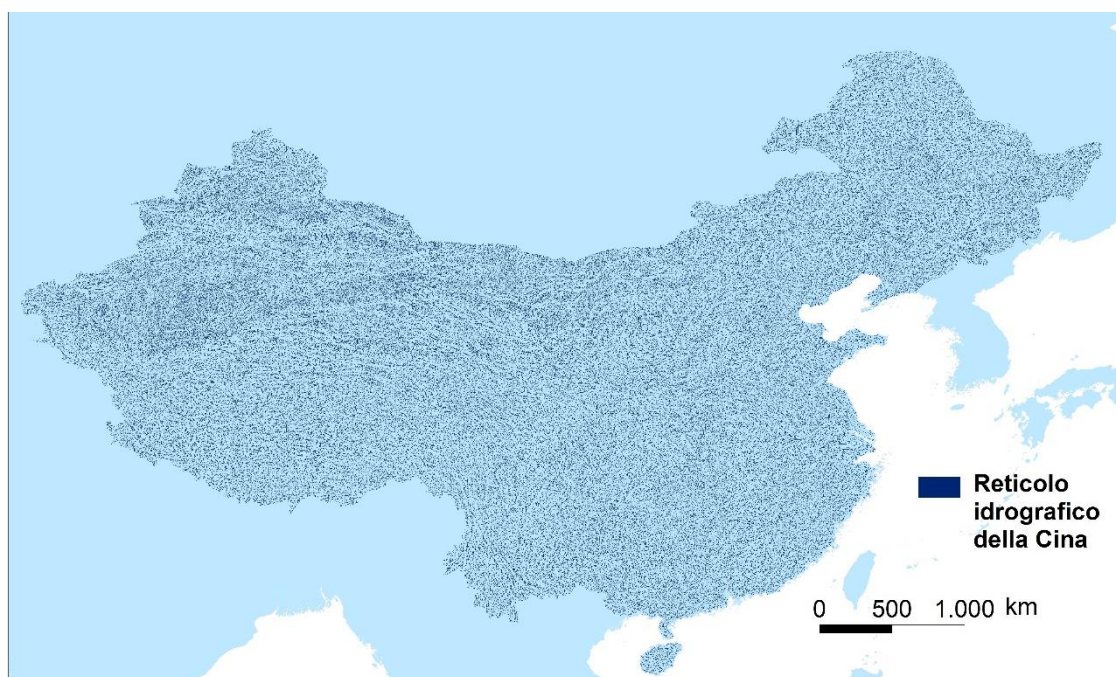


Figura 1.7: Reticolo idrografico della Cina. Fonte: database di HydroSHEDS.

1.2.2 Clima

Come esposto finora, la Cina ha un vastissimo territorio che presenta forme morfologiche e coperture del suolo molto variegata attraversate da corsi d'acqua perenni a portate ingenti e torrenti effimeri o

sotterranei. Un riassunto delle diverse zone morfologiche che si possono osservare è riportato in Figura 1.8.

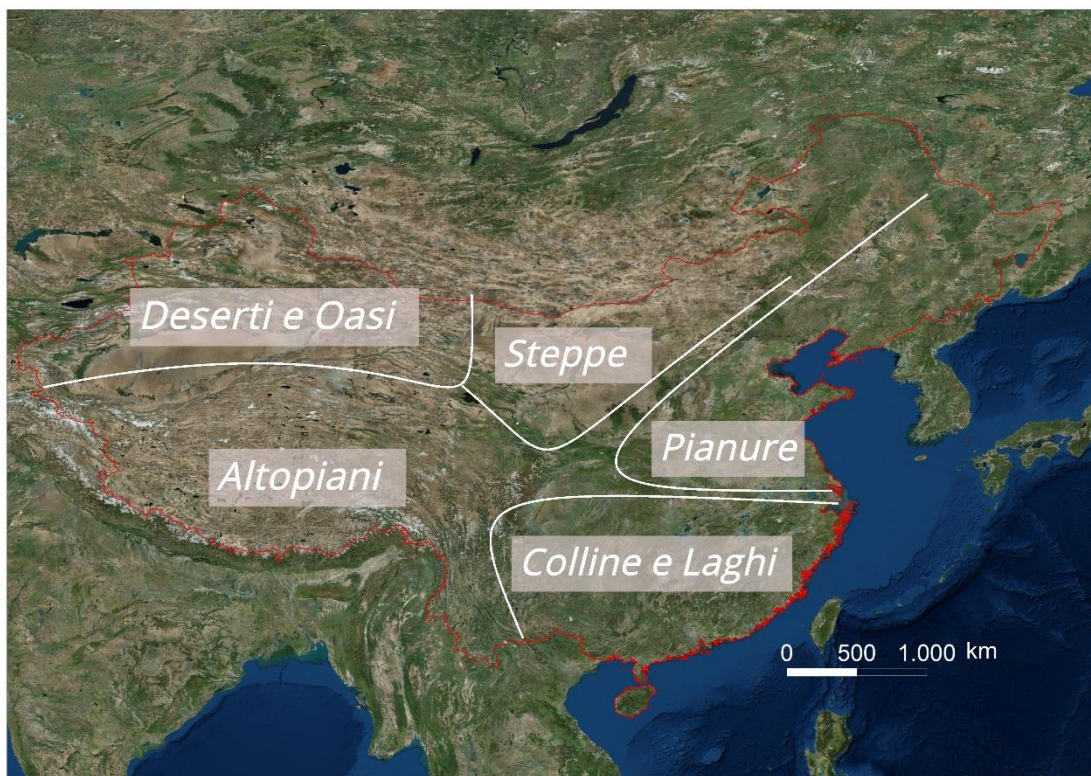


Figura 1.8: Macro-aree morfologiche della Cina.

Per ottenere più informazioni sul clima, è possibile reperire i dati di temperatura media nei mesi di gennaio e luglio (°C) e piovosità media annuale (mm) (Figura 1.9).

Mediamente, le zone più calde sono quelle del sud-est, che presentano anche una minor escursione termica ma un'elevata piovosità, accentuata dai monsoni nei mesi invernali; confrontando con la Figura 1.6 si nota che in questo territorio sono presenti le foreste più ampie del Paese. Il clima qui è Subtropicale Umido, con estati molto lunghe e piogge intense per tutto l'anno. Per contro, a sud-ovest si incontra la catena Himalayana caratterizzata dal clima degli altipiani, le temperature registrate sono le minime in entrambe le stagioni, le precipitazioni sono soprattutto nevose e la piovosità risulta scarsa. Salendo a nord-ovest, si incentrano dapprima un clima desertico, caratterizzato da lunghe estati molto calde, inverni corti ma rigidi e piovosità scarsa, e poi un clima semiarido delle steppe con piovosità scarsissima e grandi escursioni termiche tra estati e inverni. A nord-est, infine, il clima si presenta con caratteristiche continentali temperate; estati lunghe e miti si contrappongono a inverni abbastanza freddi, la piovosità è abbastanza elevata soprattutto nei mesi più caldi. All'estremo nord est, la zona della Manciuria vede un clima subartico con estati corte e temperature molto rigide in inverno.

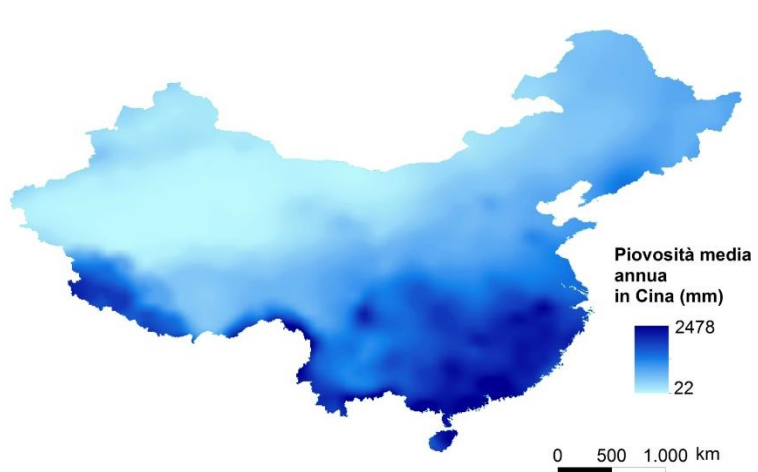
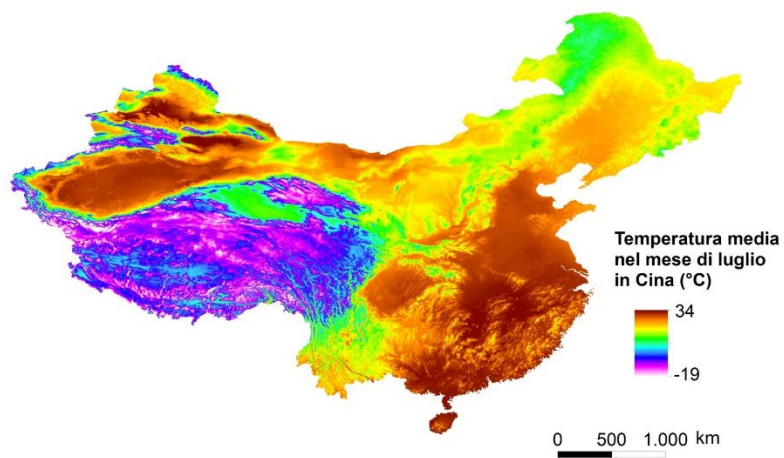
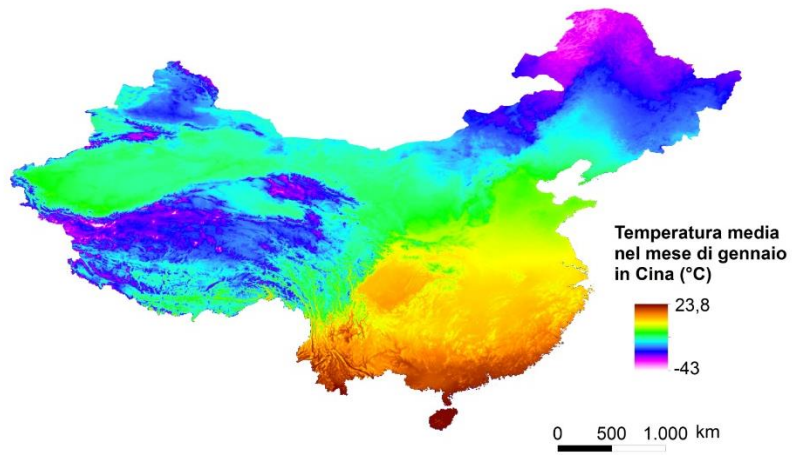


Figura 1.9: Temperature medie nel mese più freddo e più caldo e piovosità media in Cina. Fonte: NOAA, FAO database, poi elaborati in ambiente GIS.

1.3 L'area di studio

Appare chiaro che analizzare ed eseguire i calcoli necessari alla costruzione del modello sull'intero territorio cinese risulterebbe troppo oneroso in termini di tempo ma anche dal punto di vista computazionale.

Più è vasto il territorio analizzato per determinare il legame tra intensità di luce e parametri geomorfici e idrologici che ne determinano l'evoluzione nel corso degli anni, maggiore è la probabilità di osservare situazioni estreme; per esempio, se si studiasse il fenomeno a scala globale, si incontrerebbero punti in cui l'intensità luminosa è totalmente assente anche se vicino alla costa e in territorio pianeggiante, è l'esempio dei poli, oppure si potrebbero osservare luci intense ad elevate altitudini, caso dell'Altopiano del Messico²⁵. In questo caso, il modello sarebbe più difficile da determinare e presenterebbe andamenti complessi. D'altro canto, un modello a scala minore esaminerebbe un territorio poco vario e potrebbe funzionare solo per quell'area limitata, senza avere valenza generale.

Cercando di bilanciare queste osservazioni, si è deciso di focalizzare le analisi su un'area quadrata di circa 1000 km di lato, in modo da osservare una buona variazione sia della luminosità notturna che dei parametri e da rendere i calcoli più snelli. L'area individuata è situata a nord-est, nei dintorni di Pechino (Figura 1.10).

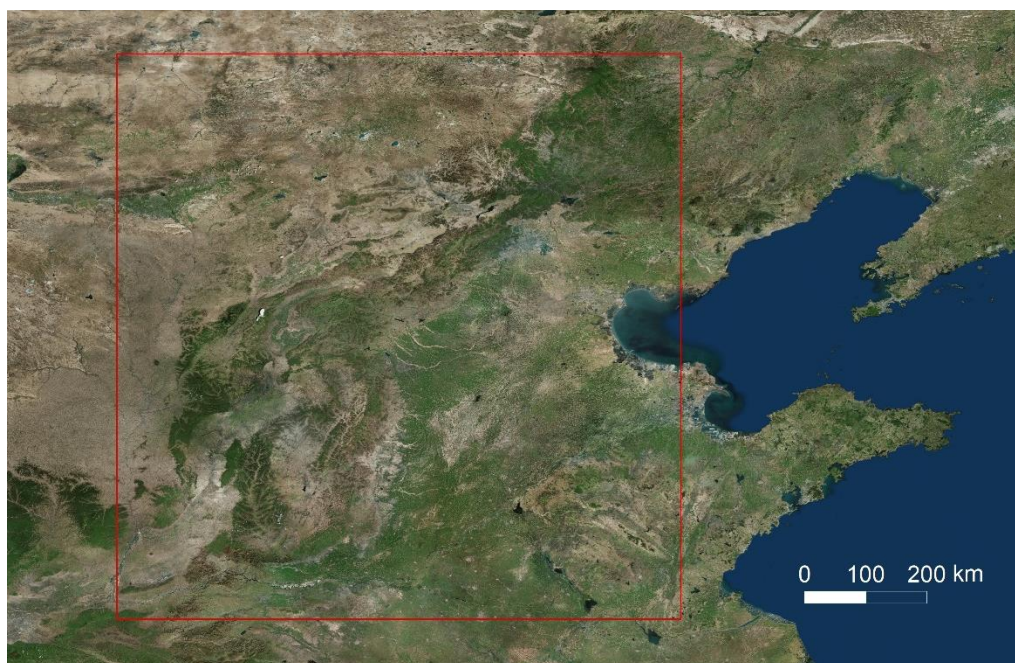


Figura 1.10: Immagine satellitare dell'area di studio (Nord-Est della Cina, nei dintorni di Pechino). Fonte: Basemap "World Imagery" di ArcGIS.

In questo territorio sono localizzate importanti città e si è assistito a una forte crescita di luminosità e a un marcato infittimento del tessuto urbanizzato negli anni di osservazione satellitare; inoltre, il territorio presenta diverse forme morfologiche ospitando ampie pianure coltivate, montagne ricche di foreste e, a nord-ovest, steppe più aride. Il clima passa da continentale caldo, lungo le coste a est, a semiarido, oltre le catene montuose a ovest, a continentale umido, all'estremo nord.

Le variabilità di popolazione e ambiente appaiono buone e l'area individuata ha quindi tutte le caratteristiche necessarie per l'analisi.

2 I dati a disposizione e quelli di interesse per lo studio

2.1 La scelta della variabile da modellizzare

Lo scopo dell'elaborato era, dunque, quello di definire un modello per la descrizione, e, magari, la previsione, della distribuzione di popolazione nel tempo e nello spazio a scala locale, con un raggio di studio comunque abbastanza esteso.

Dopo aver definito l'area di studio, è naturalmente necessario definire la variabile da modellizzare. Per quanto riguarda la popolazione, le possibilità di esame sono sostanzialmente due: ricercare i dati di popolazione, o di densità di popolazione, oppure utilizzare le variazioni di luminosità notturna (Nighttime Lights), come accennato al Capitolo 1.

Vediamo di seguito le differenze nell'utilizzare un dato piuttosto che l'altro.

2.1.1 Dati di distribuzione e densità di popolazione

La prima tipologia di dato che si può ricercare per mappare la popolazione in una determinata area è, intuitivamente, quella che descrive la densità di popolazione.

Questi dati si possono ritrovare alla sezione EU Science Hub della Commissione Europea, alla tematica Global Human Settlement Layer, un database²⁶ realizzato dal Joint Research Centre e dal DG REGIO. Lo scopo di questo database è quello di produrre informazioni spaziali che descrivano la presenza umana a livello globale, utilizzando le informazioni satellitari in diversi istanti temporali (tramite classificazione di immagini ottenute da Landsat) integrate con dati di censimento o altre informazioni geografiche. La mission degli sviluppatori vuole che queste informazioni siano *“Riproducibili, scientificamente difendibili, a piccola scala, sinottiche, complete, a livello globale”*²⁷.

I layer a disposizione in questo database sono sostanzialmente di tre tipologie: mappe di superficie edificata, di insediamenti umani (basati sul grado di urbanizzazione) o di densità di popolazione; per questo studio si è scelto di focalizzarsi su queste ultime informazioni, utilizzando i raster della Global Human Settlement Population Grid (GHS-POP)²⁸.

I diversi file raster a disposizione riportano, a diverse risoluzioni (250 m o 1 km), la densità di popolazione come numero di persone per cella (pixel).

Come si può osservare in Figura 2.1, queste informazioni si ottengono dall'integrazione tra dati di popolazione, ricavati dai censimenti, e immagini satellitari poi elaborate per ottenere mappe tematiche di superficie edificata, ovvero mappe di uso del suolo antropizzato.

La base per la costruzione dei layer GHS è la Gridded Population of the World (GPW v.4)²⁹ fornita dal SEDAC (Socioeconomic Data and Applications Center), il data center della NASA ospitato dal CIESIN Earth Institute (Center for International Earth Science Information Network) della Columbia University.

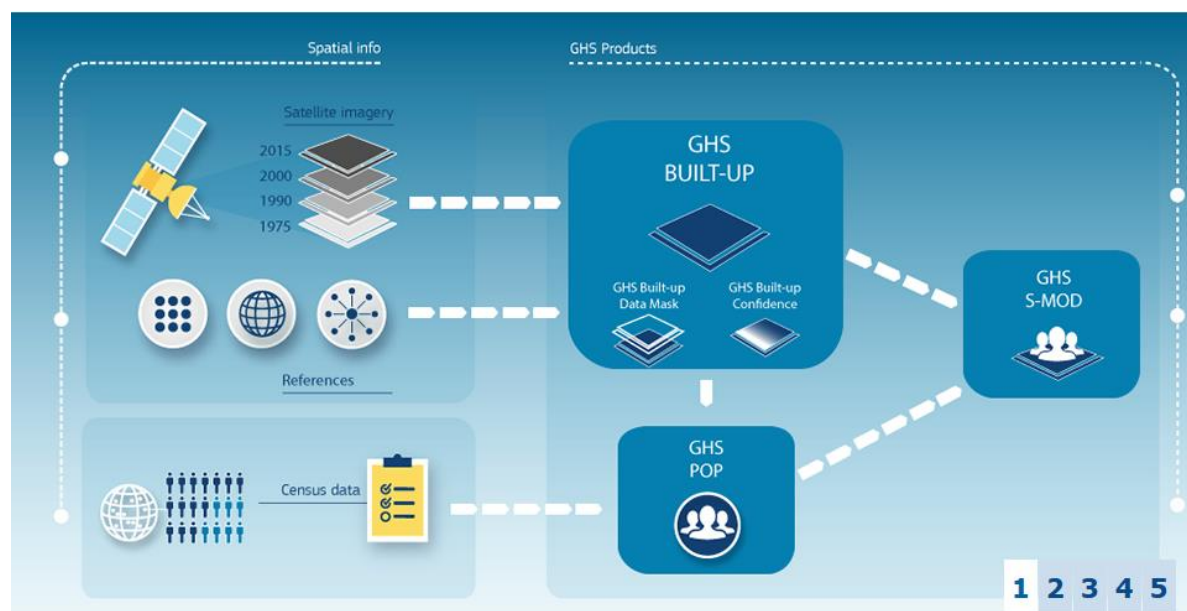


Figura 2.1: I prodotti scaricabili dal database GHS, fonte: <https://ghsl.jrc.ec.europa.eu/data.php>.

Rispetto alla GHS-POP, la GPWv4 non fornisce il numero di persone per cella ma per confini amministrativi o di censimento. Per questo motivo, risulta vantaggioso utilizzare dati disaggregati da questi confini e forniti pixel per pixel, processo con cui si costruisce la GHS-POP.

Uno svantaggio nell'uso dei layer forniti dalla Commissione Europea è che essi sono disponibili solo per gli anni 1975, 1990, 2000 e 2015; non si può, quindi, pensare di ottenere un andamento anno per anno della densità di popolazione. Ovviamente, si potrà comunque ricercare una correlazione tra questo dato e i parametri dello studio ma sarà supportata da un numero di osservazioni limitate.

Per questo motivo, ed altri esposti di seguito, si è esplorata la possibilità di studiare, piuttosto che la densità di popolazione, la luminosità notturna nell'area di interesse.

L'ultima caratteristica da riportare per questi file raster è il sistema di riferimento. Trattandosi di dati a scala globale per cui sono riportati i dati di popolazione per cella, i raster sono forniti in sistema World Mollweide, con la peculiarità di mantenere inalterate le aree nel passaggio dalla realtà alla rappresentazione³⁰. Per questo tipo di proiezioni, dette equivalenti, le aree non vengono infatti deformate e si mantengono proporzionate a quelle reali, a discapito di angoli e forme che vengono modificati³¹.

Questa proprietà rende il sistema World Mollweide piuttosto utile per rappresentazioni in cui l'area sia più significativa della forma spaziale degli oggetti, come nel caso della densità di popolazione, ottenuta dal rapporto tra numero di persone occupanti una cella e l'area della cella stessa:

$$Densità = \frac{Popolazione_{cella}}{Area_{cella}}$$

Come si può intuire dalla Figura 2.2, in cui è riportata la GHS-POP per l'area della Cina nel 2015, la proiezione deforma gli oggetti soprattutto verticalmente, allungando le forme più vicine all'equatore e schiacciandole man mano che si avvicinano ai poli. Questo fenomeno si deve alle caratteristiche della rappresentazione che riporta i paralleli come linee orizzontali a distanze variabili (si infittiscono allontanandosi dall'equatore) e i meridiani come linee a curvatura variabile (la curvatura aumenta allontanandosi dal meridiano che incontra l'equatore nel suo punto medio, rappresentato da una retta verticale) a distanze costanti (Figura 2.3).

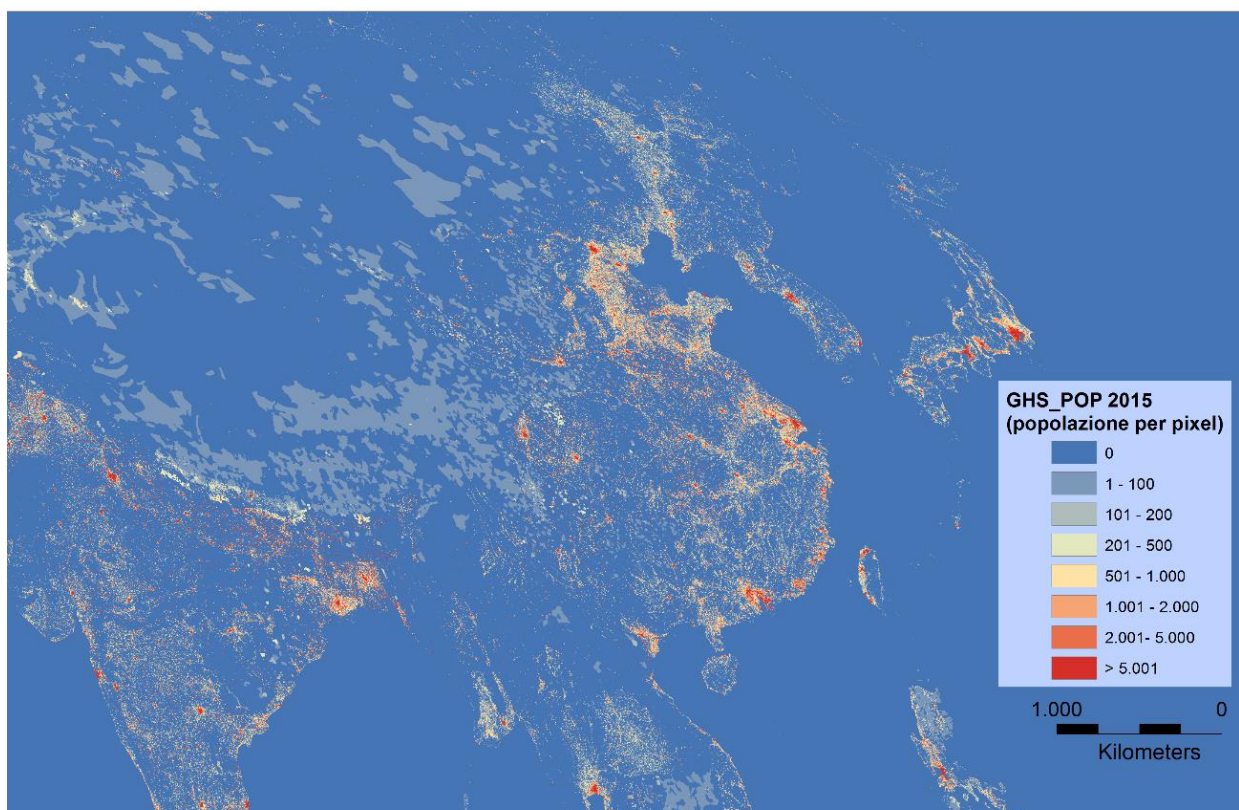


Figura 2.2: Rappresentazione della GHS-POP per l'area dell'Asia Orientale nell'anno 2015. Sono visibili notevoli deformazioni delle forme rispetto alle rappresentazioni cartografiche conformi.

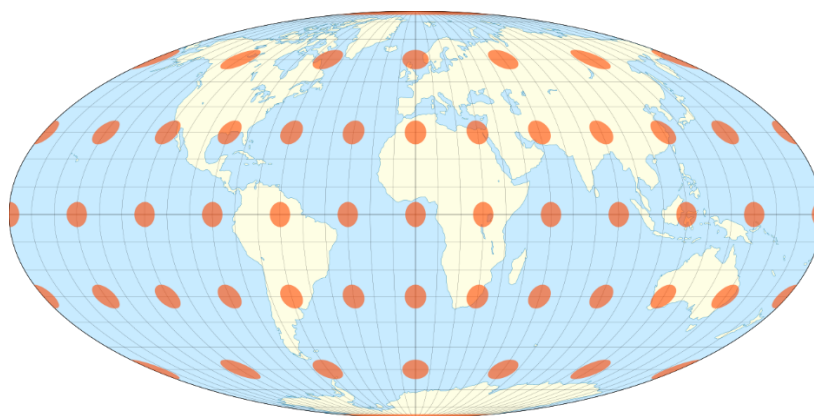


Figura 2.3: Forme di paralleli e meridiani nella rappresentazione World Mollweide, fonte: Justin Kunimune - Own work, CC BY-SA 4.0, <https://commons.wikimedia.org/w/index.php?curid=66467569>

In Tabella 2.1 sono riportate brevemente le caratteristiche riassuntive dei dati di densità di popolazione riportati nel database della GHS-POP.

Tematismo	Densità di popolazione per cella
Tipologia di dato	Raster
Sistema di Riferimento	World Mollweide
Risoluzione disponibile	250 m, 1 km
Anni per cui sono disponibili le osservazioni	1975, 1990, 2000, 2015

Tabella 2.1: Proprietà dei dati della Global Human Settlement Population Grid.

2.1.2 I dati di luminosità notturna

Per avere maggior disponibilità di osservazione rispetto ai dati di densità di popolazione, ma anche per ottenere ulteriori informazioni, si sono quindi utilizzati anche i dati di luminosità notturna nell'area di studio.

Come accennato al Capitolo 1, si tratta di mappe ottenute registrando le radiazioni luminose visibili e vicine all'infrarosso che provengono dalla superficie terrestre durante la notte. I satelliti deputati a questo compito sono quelli del Defense Meteorological Satellite Program's Operational Linescan System (DMSP/OLS), un programma gestito dal NOAA.

Rispetto ai dati di densità di popolazione, è chiaro che i dati di luce notturna, fornendo informazioni sull'attività oltre che alla presenza umana, contengono anche informazioni sull'economia, le aree urbanizzate o lo sviluppo delle infrastrutture del Paese³².

Le applicazioni di questi dati sono rapidamente aumentate negli ultimi anni, spaziando in diversi campi: dallo studio dell'urbanizzazione³³ e della distribuzione della popolazione allo studio dell'inquinamento, dalla diffusione di malattie fino all'economia.

Per quanto riguarda l'economia e la società, uno studio³⁴ di Bennett M. e L. C. Smith del 2017 ha dimostrato come questi dati rappresentino con una buona precisione la realtà. Inoltre, la correlazione tra luci notturne e variazione di popolazione sembra essere migliore per Paesi in rapida crescita come, appunto, la Cina.

Una peculiarità di questi dati è che essi non rappresentano solo la distribuzione delle luci artificiali, ma riportano anche informazioni sull'intensità luminosa che può assumere un valore da 0, per pixel in cui le luci sono assenti, a 63, pixel a massima luminosità. In questo modo si possiede un'espressione quantitativa con cui è possibile creare una classificazione delle luci o definirne un comportamento spazio-temporale.

Il valore è ottenuto tramite elaborazioni delle immagini fotografate dai satelliti che devono essere filtrate per rimuovere gli effetti di disturbo: nuvole, luci provenienti da vulcani o incendi e riverbero della luce solare e lunare³⁵.

Un ulteriore vantaggio è la numerosità della serie. Se per la densità di popolazione avevamo a disposizione solo 4 anni di osservazioni, per le luci notturne possiamo disporre di un file raster per ogni anno dal 1992 al 2013 (gli ultimi anni devono ancora essere corretti e revisionati).

Tematismo	Luminosità notturna di una cella
Tipologia di dato	Raster
Sistema di Riferimento	GCS-WGS84
Risoluzione disponibile	1 km
Anni per cui sono disponibili le osservazioni	Dal 1992 al 2013

Tabella 2.2: Proprietà dei Nighttime Light Data.

Come si può notare dalla Tabella 2.2, che riassume le caratteristiche dei dati di luminosità, si tratta ancora di file raster con risoluzione a 1 km, questa volta basati sul sistema di riferimento GCS-WGS84. In questo tipo di rappresentazione vengono modificate le aree degli oggetti, mentre se ne preservano gli angoli e, quindi, le forme.

La risoluzione disponibile risulta perciò minore di quella della GHS-POP (vedi Tabella 2.1), sicuramente ciò implica un minore dettaglio nelle osservazioni; tuttavia, una rappresentazione meno dettagliata può anche risultare un vantaggio poiché rende il calcolo meno oneroso avendo, a parità di area, un numero minore di pixel da analizzare.

Per analisi su piccole aree sarebbe opportuno utilizzare un dettaglio maggiore per poter osservare andamenti locali più precisi. Per questo lavoro, che si concentra un'area abbastanza vasta di 1000 km per 1000 km (come esposto al Capitolo 1), un numero troppo elevato di pixel da elaborare

porterebbe a tempi di calcolo molto ampi e si può allora accettare una perdita di dettaglio per ottenere risultati a una scala maggiore, la risoluzione di 1 km risulta allora più che sufficiente.

In conclusione, per via della numerosità di dati e della completezza dell'informazione che fornisce, riportando non solo dati sulla densità di popolazione ma anche sullo sviluppo della società e dell'economia, avendo una risoluzione accettabile per gli scopi e i mezzi di questo studio, si è deciso di assumere come variabile da modellizzare la luminosità notturna nell'area di interesse.

Dunque, tutti i dati utilizzati saranno riportati a una risoluzione di 1 km (30 secondi di grado) e in sistema WGS84, perché queste sono le caratteristiche dei dati di luce e poiché risulta più semplice mantenere le forme degli oggetti nella rappresentazione simili a quelle reali.

In Figura 2.4 sono riportate le immagini di luci notturne nell'area di studio negli anni 1992 (a sinistra) e 2013 (a destra).

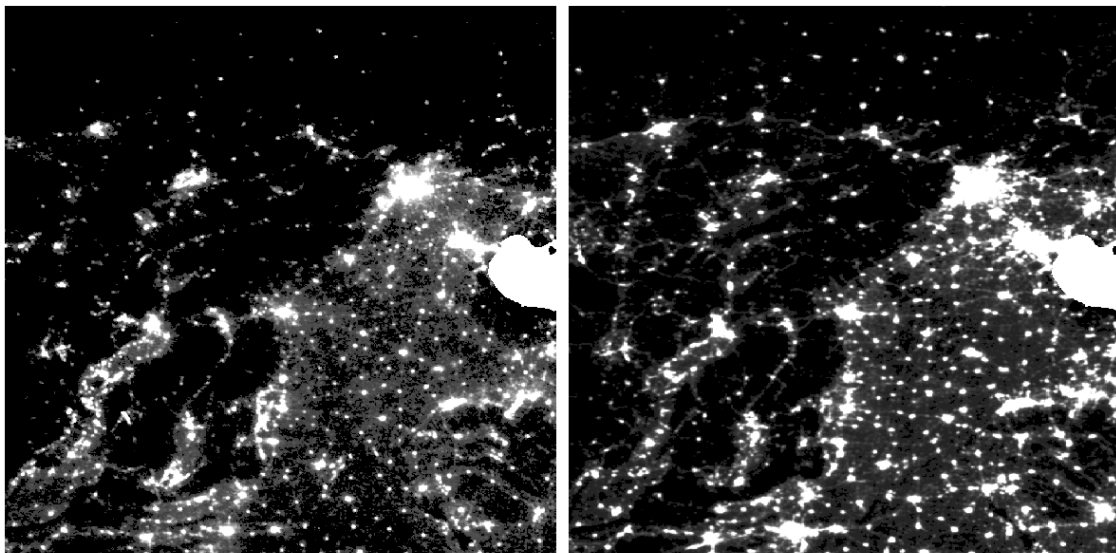


Figura 2.4: Rappresentazione delle luci notturne all'interno dell'area di studio nell'anno 1992 (sinistra) e 2013 (destra), fonte: raster da DMSP/OLS, NOAA, poi elaborate in ArcGIS.

Si possono da subito fare due osservazioni: innanzitutto le luci sono maggiori a est, zona limitrofa alla costa, mentre diminuiscono velocemente verso ovest, con ampie zone buie, compatibilmente alla distribuzione non omogenea della popolazione cinese esposta al Capitolo 1. Inoltre, si può notare un infittimento delle luci tra i due anni e un complessivo aumento di luminosità, più marcato comunque nelle zone orientali.

Come esposto al Capitolo 1, era necessario individuare una zona che presentasse una buona variazione di popolazione nel periodo di studio. La variabile scelta, le luci notturne, vedono un netto incremento negli anni disponibili, ulteriore conferma della loro adeguatezza per questo studio.

La densità di popolazione per il momento è stata scartata come variabile per il modello; essa riporta comunque informazioni utili e sarà infatti utilizzata nell'analisi dei dati riportata nei capitoli successivi.

2.2 La scelta dei parametri di influenza

Il primo passo verso la creazione di un modello per definire l'andamento della luminosità notturna è la scelta dei parametri da cui essa può dipendere.

Innanzitutto, è bene ricordare che esistono diverse possibilità per indagare la variazione di distribuzione dell'uomo e il suo sviluppo a scala globale e locale, ognuna delle quali può utilizzare approcci diversi. Si può, per esempio, pensare di descrivere il fenomeno con parametri sociali come la sicurezza e il benessere³⁶, oppure è possibile ipotizzare una dipendenza da fattori economici: tipologia di settori economici presenti³⁷ o sviluppo tecnologico in una determinata area³⁸.

In questo studio, i parametri a cui sarà correlata la distribuzione della popolazione nel tempo e nello spazio, sono tutti di tipo ambientale.

In particolare, possono essere riassumibili come parametri:

- Di posizione;
- Climatici;
- Idrologici;
- Morfologici.

I fattori ipotizzati per la ricerca di correlazioni con la luminosità, e quindi la presenza umana, sono derivati da esperienze e da osservazioni precedenti, discusse brevemente nei paragrafi seguenti.

È intuitivo ritenere che la popolazione sia più portata a stanziarsi nei pressi di corsi d'acqua o in vicinanza alla costa, per avere maggiori vie di commercio o scambio, ma anche per il facile accesso alla risorsa idrica. Inoltre, l'esperienza suggerisce che gli insediamenti di dimensioni maggiori si verifichino in pianura, dove lo spazio per costruire è maggiore, più facile da modellare e, spesso, ricco di riserve d'acqua. I parametri di posizione, quindi, sono legati alla vicinanza ai fiumi e alla costa, mentre i parametri morfologici sono caratterizzati da altitudine e tipologia di terreno. Si è ricercata, infatti, la tipologia di copertura del terreno, per verificare se esistono tipologie di suolo preferenziali nello sviluppo di insediamenti umani. Essendo però difficile tradurre questo parametro in modo quantitativo, sarà utilizzato solo per le analisi preliminari al Capitolo 3, e non per la costruzione del modello vero e proprio. Inoltre, si può ritenere che la popolazione preferisca insediarsi soprattutto dove il clima è più favorevole sia alla vita che alle attività umane (agricoltura, allevamento, trasporti). Osservando il clima che caratterizza l'area (Capitolo 1) possiamo immaginare che le luci

notturne siano più intense dove le temperature saranno più miti ma non troppo elevate, ovvero in zone dove il clima non si presenti in forme estreme.

Parlando di clima, si è ritenuto opportuno includere un altro parametro, idrologico: la piovosità. In aree in cui si verifica scarsità di pioggia si avrà anche una popolazione limitata (il caso più esemplificativo, e più estremo, è quello delle aree desertiche di tutto il mondo). Per contro, anche una precipitazione eccessiva potrebbe scoraggiare gli insediamenti umani; si può pensare, per esempio, alle zone in cui si verificano uragani o monsoni. Questi ultimi sono presenti anche nel territorio analizzato, motivo per cui risulta più difficile fare previsioni sul legame tra la luminosità e la precipitazione nell'area di studio.

Vediamo quindi in che tipologie di dato si traducono i parametri esposti sopra, analizzando la disponibilità di informazioni e la loro significatività.

Per osservare chiaramente le relazioni tra questi parametri e la distribuzione delle luci notturne, si rimanda al capitolo successivo in cui saranno affrontate diverse elaborazioni matematiche per stabilire quali sono i fattori che sanciscono una luminosità più o meno intensa.

2.2.1 Digital Elevation Model (DEM)

Esiste un tipo di informazione che, in qualche modo, contiene in sé tutti i parametri di posizione e morfologia illustrati sopra: il Digital Elevation Model, ovvero il modello digitale delle quote del terreno. Si tratta tipicamente di file raster che riportano, per ogni pixel, l'elevazione sul livello del mare.

Utilizzando questi dati, è facile ottenere informazioni sulla morfologia dell'area di studio: in base alla quota possiamo definire se stiamo osservando un'area montuosa o pianeggiante.

La quota potrebbe anche contenere in sé informazioni sulle caratteristiche climatiche, solitamente a quote più basse corrispondono climi più miti e, magari, influenzati dalla presenza del mare. A quote maggiori, invece, si realizzano climi più rigidi, caratterizzati anche da precipitazioni nevose che potrebbero sfavorire gli insediamenti umani (queste considerazioni saranno poi confermate nel Capitolo 3).

Poiché contiene informazioni di diversa natura, il DEM è, dunque, un parametro rilevante per lo studio. I raster sono disponibili, a diverse risoluzioni, sul portale dell'USGS HydroSHEDS e sono prodotti dal WWF³⁹. In Tabella 2.3 possiamo vedere riassunte le proprietà di questi file.

Rispetto ai dati di luce e di popolazione si utilizzerà in questo caso, come per tutti gli altri parametri, un solo raster. Si può effettivamente ritenere che le quote siano rimaste invariate per tutti gli anni di osservazione poiché non si cercano valori con precisioni inferiori al metro e i cambiamenti di quota sono dell'ordine dei millimetri all'anno per fenomeni di subsidenza (sebbene la Cina, e in particolare

Pechino, siano oggetto di un fenomeno di subsidenza estremo che raggiunge anche i 20 mm all'anno⁴⁰), o al più del centimetro, per eventi sismici.

Tematismo	DEM (m.s.l.m.)
Tipologia di dato	Raster
Sistema di Riferimento	GCS-WGS84
Risoluzione disponibile	100m, 500 m, 1 km

Tabella 2.3: Caratteristiche del raster contenente i dati di DEM

Dalla tabella, si può notare che il sistema in cui sono riportati i dati è già il WGS84 ed è disponibile la risoluzione a 1 km, per cui non sono necessarie elaborazioni del file, se non il ritaglio del raster alla sola area di influenza. Il DEM riportato in Figura 2.5 presenta un'ampia area pianeggiante nei pressi della costa (a est) e quote che aumentano rapidamente verso ovest fino a un massimo di circa 3000 m sul livello del mare.

Da questa immagine si può intuire una certa correlazione con la distribuzione delle luci notturne, visibili in Figura 2.4. Le luci più intense, infatti, sembrano disporsi nelle aree a minor quota, con i massimi di luminosità che si verificano in presenza dell'ampia zona pianeggiante a est. Le correlazioni numeriche saranno studiate analiticamente nel dettaglio nel Capitolo 3.

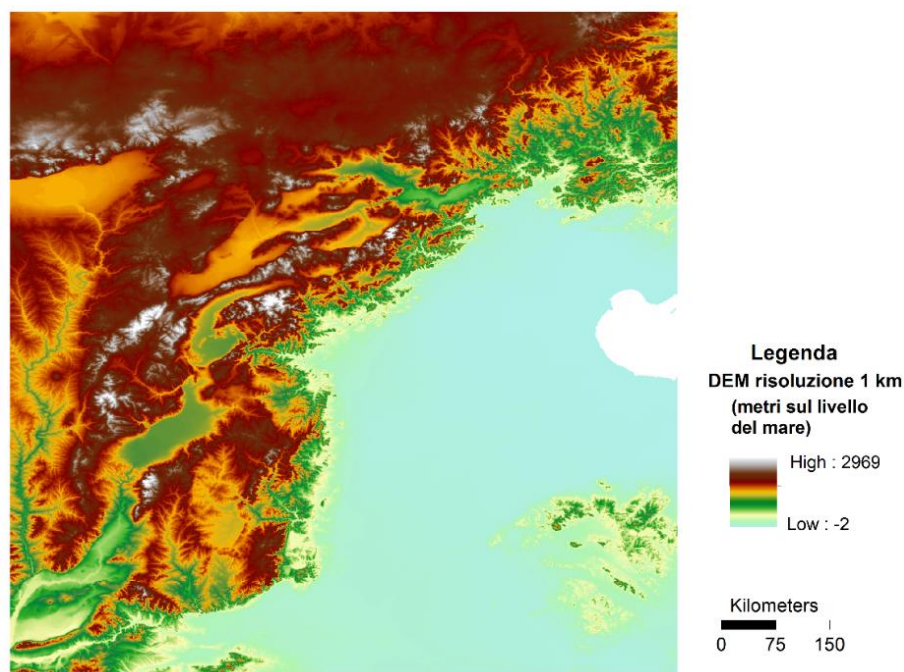


Figura 2.5: Rappresentazione del DEM per l'area di studio, fonte: USGS HydroSHEDS poi elaborata in ArcGIS

2.2.2 Distanza dalla costa

Si è deciso di ricercare un file che riportasse, punto per punto, la distanza dalla linea di costa ma, sebbene si abbiano consultati diversi database online, non sono stati trovati dei raster già contenenti questa informazione. Il dato di cui si necessita, è dunque stato creato in ambiente ArcGIS a partire da un file vettoriale che ricalca la linea di costa a livello mondiale (Figura 2.6).



Figura 2.6: Linea raster che ricalca la linea di costa a livello globale, fonte: <http://openstreetmapdata.com>.

Le grafiche vettoriali, a differenza dei raster, rappresentano le informazioni tramite l'uso di primitive matematiche⁴¹ con il vantaggio di ridurre le dimensioni del file. Il dato non ha risoluzione se non quella massima consentita dal monitor, poiché ogni linea è rappresentata appunto da funzioni e non da pixel.

È possibile utilizzare uno degli strumenti di calcolo a disposizione su ArcMap per ottenere, punto per punto, le distanze da questa linea. Per procedere in questo modo, si devono però fare alcune osservazioni.

Innanzitutto, essendo il file della linea di costa un vettore, mentre si vuole ottenere un file raster, si possono verificare problemi nella sovrapposizione tra il raster creato e quelli disponibili come le luci e il DEM. Ciò è dovuto al fatto che i raster hanno una risoluzione limitata mentre il file vettoriale permette un livello di dettaglio molto maggiore; dalla Figura 2.7, in cui è mostrato il confronto tra la linea di costa del DEM e quella del vettoriale, si notano infatti diverse incongruenze. Se si volesse ottenere un raster di distanze dalla costa a partire da questo file vettoriale, quindi, si otterrebbe un'area diversa da quella degli altri raster disponibili e le informazioni non sarebbero valide.

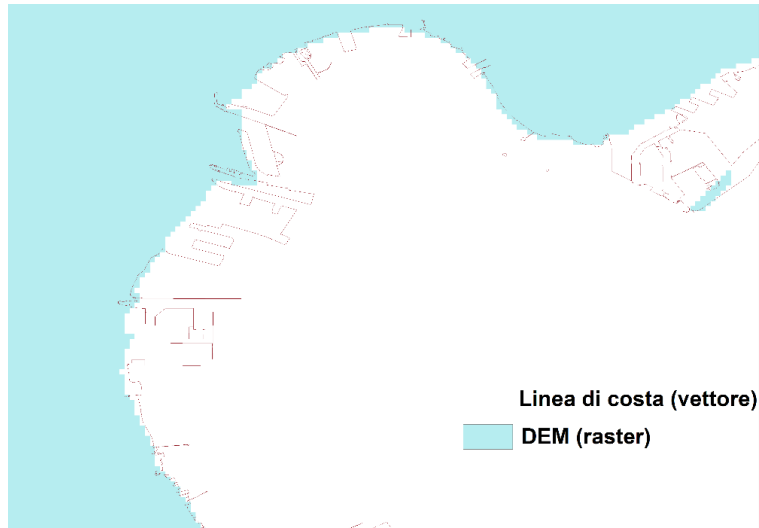


Figura 2.7: Confronto tra la linea di costa del file raster del DEM e del file vettoriale

Per ovviare a questo problema, si può cercare di trasformare il file di costa vettoriale in uno raster alla stessa risoluzione dei precedenti e osservare se le differenze tra le linee di costa risultano allora minori; ciò si può eseguire con il comando *Feature to Raster* della sezione *Conversion Tools* di ArcToolbox. È sufficiente inserire il vettore input e specificare la dimensione dei pixel che si vogliono ottenere, si può richiedere all'operazione di costruire celle della stessa misura di quelle del DEM o delle luci notturne, ovvero 1 km.

Il risultato di questo processo è riportato in Figura 2.8, e si nota come la nuova linea di costa si avvicini notevolmente a quella del raster DEM. In questo modo, si è ottenuto un file da cui è possibile calcolare la distanza per ogni pixel.

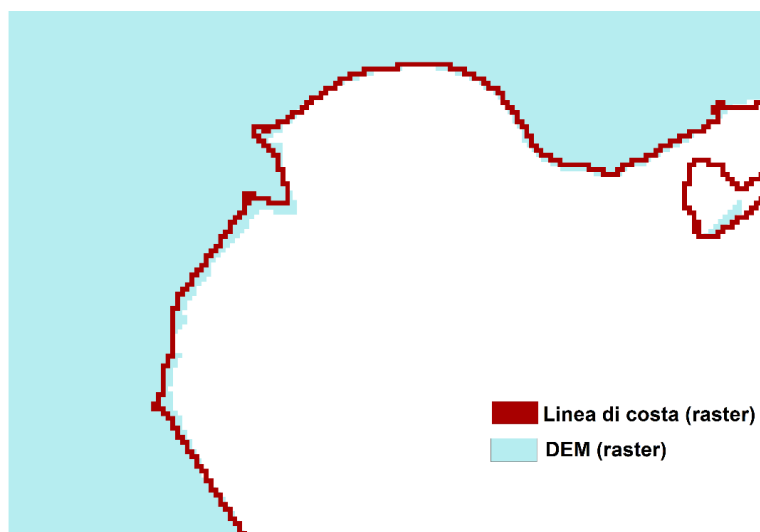


Figura 2.8: Confronto tra le linee di costa del file raster del DEM e di quello ottenuto dalla rasterizzazione del vettoriale di partenza

È adesso necessario definire quale distanza si vuole considerare; si può pensare di calcolare una distanza inclinata in tre dimensioni, cioè il segmento che tenga conto anche dell'altitudine dei punti rispetto alla costa, oppure una semplice distanza euclidea in pianta, ovvero la retta più breve che colleghi i punti in due dimensioni.

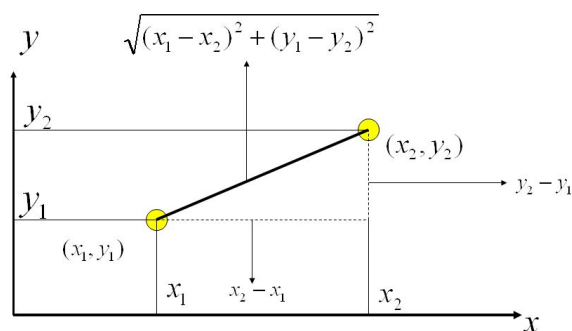


Figura 2.9: Rappresentazione della distanza euclidea in tre dimensioni: mentre con la distanza inclinata si osserva l'effetto della quota, con la distanza in pianta esso sparisce e si osserva la sola distanza x_2-x_1 .

Senza dubbi la prima è la più corretta da usare, infatti punti vicini alla costa in linea d'aria ma ad altitudine elevata, risultano comunque più lontani di punti a quota ridotta e a distanze maggiori.

È lecito semplificare il problema e calcolare la distanza euclidea in pianta ricordando che tra i parametri del modello è già stata inserita la quota nelle informazioni del DEM. Si è quindi utilizzata la distanza euclidea in pianta, assumendo che eventuali effetti della quota saranno contenuti nel parametro DEM e le due informazioni si completeranno nel modello integrato. È possibile ottenere un raster che abbia i pixel della dimensione voluta (in questo caso 1 km) e che associ a ogni punto questa informazione con lo strumento *Euclidean Distance* degli *Spatial Analyst Tools*.

Il risultato, espresso in gradi e ritagliato per l'area di interesse, è riportato in Figura 2.10. I valori limite osservati sono di 0° per punti sulla linea di costa, fino a un massimo di circa 8,3°, corrispondenti a circa 1000 km. La Tabella 2.4 riporta riassunte le caratteristiche di questo raster.

Tematismo	Distanza 2D dalla costa (gradi)
Tipologia di dato	Raster
Sistema di Riferimento	GCS-WGS84
Risoluzione disponibile	1 km

Tabella 2.4: Caratteristiche del raster che associa a ogni punto la sua distanza euclidea (in pianta) dalla costa.

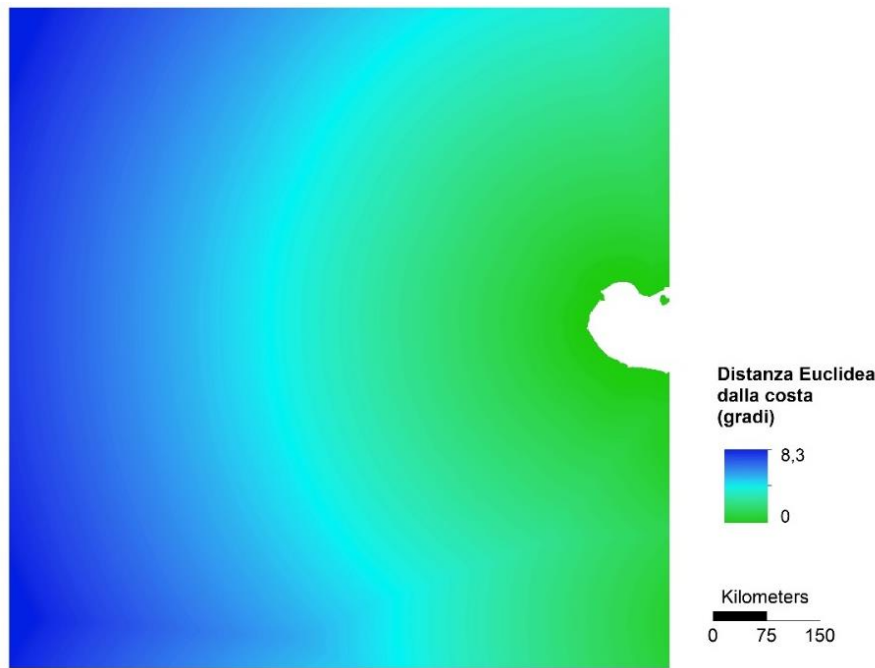


Figura 2.10: File raster ottenuto calcolando la distanza euclidea di ogni pixel dalla costa.

2.2.3 Distanza dai corsi d'acqua

Sotto le stesse ipotesi valide per il calcolo della distanza dalla costa, si è ottenuto un raster che associa a ogni pixel la distanza euclidea dai corsi d'acqua in due dimensioni.

Mentre è abbastanza facile definire quale sia la linea di costa, in questo caso è stato più complicato definire quali fossero i corsi d'acqua da cui calcolare la distanza. Se si osservano le linee di drenaggio riportate in Figura 1.7, si nota la abbondanza dei corsi d'acqua della Cina, che ricoprono quasi l'intera superficie del Paese. In un primo momento sono state usate per il calcolo proprio le linee di drenaggio, che rappresentano sia corsi d'acqua importanti, sia torrenti o ruscelli effimeri. Solitamente, queste linee non coincidono con quelle dei fiumi riportati nella cartografia del Paese, proprio perché in essa sono rappresentati solo quelli più significativi, con acque perenni e portate ingenti.

Si può, inoltre, ritenere che la popolazione non si insedi vicino ai corsi d'acqua indistintamente, ma che scelga preferibilmente quelli a portata maggiore che consentano una grande disponibilità di risorsa (sia per l'uso idropotabile che per la produzione di energia, l'utilizzo nell'industria e la navigabilità). Dunque, in seconda istanza, si è preferito migliorare l'individuazione dei corsi d'acqua di interesse e filtrare quelli delle linee di drenaggio per osservare solo i più significativi. La condizione utilizzata per il filtro è stata quella di evidenziare solo le celle con una Flow Accumulation maggiore di 5000 km². La Flow Accumulation è un raster che associa ad ogni pixel il numero di celle che confluiscono in quel pixel seguendo le linee di drenaggio. Una volta reperito questo dato dal database di HydroSHEDS, è stato possibile definire i corsi d'acqua di maggior portata come quelli serviti da

almeno 5000 celle, ovvero 5000 km², essendo la dimensione di una cella 1 km². Il valore di soglia per la definizione del numero di celle è stato suggerito da studi precedenti (i.e. Mard et al. nel 2018⁴²). In Figura 2.11 sono riportate le differenze tra le linee di drenaggio e i corsi d'acqua con Flow Accumulation maggiore di 5000 km². Questi ultimi sono in numero molto minore e possono aiutare a definire un parametro di distanza dai corsi d'acqua più significativo.

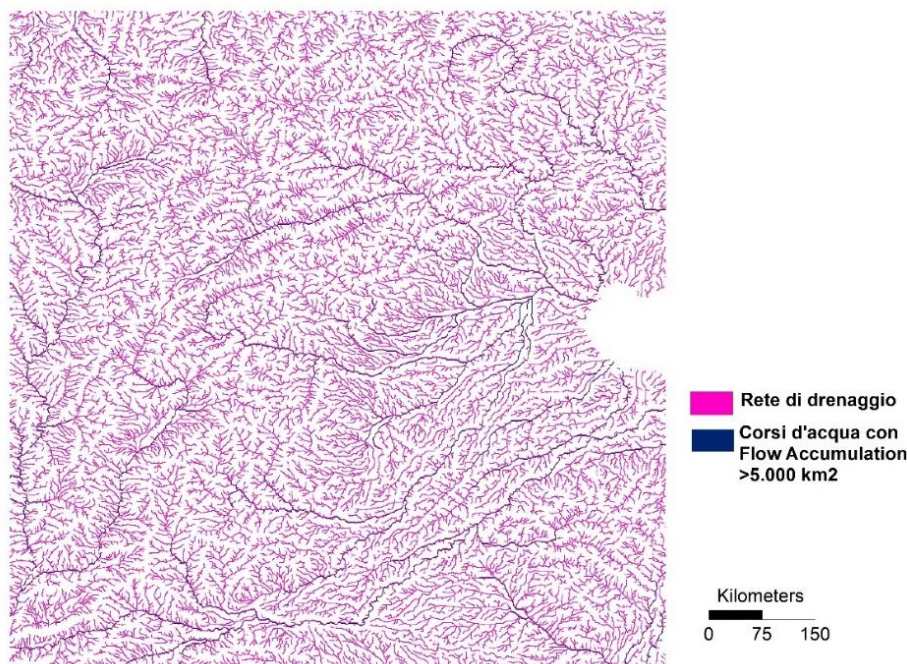


Figura 2.11: Differenza tra il raster delle linee di drenaggio e quello ottenuto imponendo una Flow Accumulation > 5000 km².

Per ottenere il raster di distanza da questi corsi d'acqua, sono stati usati ipotesi e procedimenti identici a quelli usati per calcolare la distanza dalla costa. Il raster definitivo (Figura 2.12) rappresenta quindi la distanza euclidea in pianta (espressa in gradi) dai corsi d'acqua con Flow Accumulation maggiore di 5000 km², in Tabella 2.5 sono riportate le sue caratteristiche definitive.

.Tematismo	Distanza 2D dai corsi d'acqua (gradi)
Tipologia di dato	Raster
Sistema di Riferimento	GCS-WGS84
Risoluzione disponibile	1 km

Tabella 2.5. Caratteristiche del raster contenente le distanze dai corsi d'acqua.

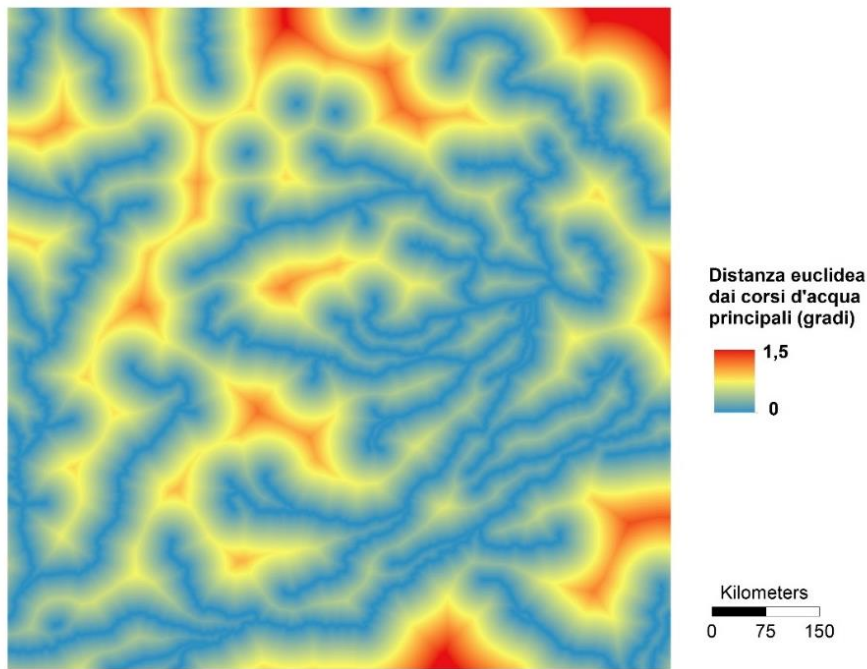


Figura 2.12: File raster ottenuto calcolando la distanza euclidea di ogni pixel dai corsi d'acqua principali.

2.2.4 Temperatura

La distribuzione di popolazione a livello globale segue fortemente quella delle temperature. Le zone più popolate sono quelle delle medie latitudini e basse altitudini, dove le temperature sono per lo più miti e l'escursione termica è controllata dal mare. Per questo motivo, si è considerato anche un parametro di temperatura all'interno del modello.

Ma quale temperatura è più rappresentativa per l'area di interesse? Possiamo considerare la temperatura media annua, che dà informazioni sul comportamento medio di un pixel senza considerare una stagionalità; oppure si potrebbe utilizzare la temperatura massima, ipotizzando che la popolazione si concentri nelle zone in cui questa è abbastanza elevata, o, all'opposto, si potrebbe considerare la temperatura minima raggiunta, verificando che le zone meno abitate sono quelle dove si raggiungono temperature più rigide.

Per queste ragioni, sono stati selezionati due diversi parametri di temperatura: la temperatura media annua e l'escursione termica massima annua per ogni pixel, seguendo il suggerimento di studi precedenti di C. Small e J. E. Cohen⁴³. Mentre la prima contiene l'informazione sul comportamento climatico medio, la seconda dà informazioni sulle massime e le minime e, quindi, delle variazioni che si verificano rispetto alla media. Le ipotesi preliminari sono che le luci notturne risulteranno più intense nelle zone con temperatura media più mite, a discapito di zone con medie troppo rigide o troppo torride, e con escursioni termiche minori, che caratterizzano climi più ospitali e meno estremi.

I dati di temperatura sono stati reperiti dal sito WorldClim⁴⁴ nella versione 2.0; sono disponibili le temperature massime, minime e medie mensili calcolate come la media dall'anno 1970 al 2000. Si hanno, dunque, a disposizione 3 pacchetti di raster contenenti 12 file ciascuno, uno per ogni mese dell'anno.

In Figura 2.13 sono riportate le temperature minime, massime e medie per i mesi di gennaio (mese più rigido) e luglio (mese più mite) per l'area di studio.

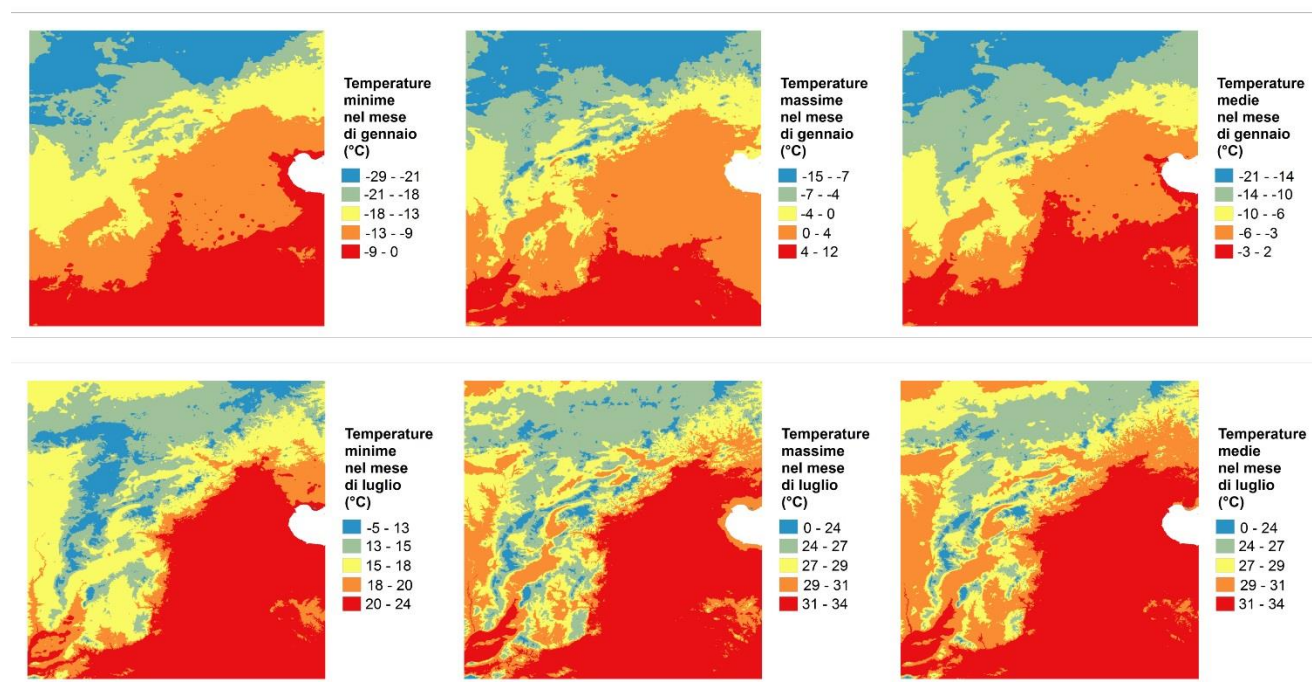


Figura 2.13: Confronto tra i raster di temperature minime (a sinistra), massime (al centro) e medie (a destra) nei mesi di gennaio (in alto) e luglio (in basso).

Dalle immagini si nota subito una grande variabilità delle temperature nel tempo e nello spazio, con il sud-est che raggiunge sempre le temperature più elevate e il nord ovest che rimane sempre più freddo.

La variabilità, come spiegato approfonditamente nel Capitolo 1, è molto importante per la costruzione di un modello poiché il parametro diventa più significativo se possiamo osservarne grandi variazioni. Per ottenere un unico dato in grado di rappresentare in modo sintetico e a livello annuale la temperatura media e l'escursione termica a partire dai numerosi file disponibili si è compiuta un'elaborazione in MATLAB. Per ricavare la temperatura media annuale per ogni pixel si è semplicemente calcolata la media tra ognuno dei 12 raster riportanti le medie mensili:

$$T_i = \frac{\sum_1^{12} T_{m,i}}{12}$$

Con

- T_i : temperatura media annua del pixel i-esimo;
- $T_{m,i}$: temperatura media del mese m nel pixel i-esimo.

Mentre per calcolare l'escursione termica punto per punto si è calcolata la differenza tra il massimo mai raggiunto nelle 12 temperature massime e il minimo delle 12 temperature minime in ogni pixel:

$$\Delta T_i = T_{max_i} - T_{min_i}$$

Con:

- ΔT_i : escursione termica massima annuale nel pixel i-esimo;
- T_{max_i} : temperatura massima raggiunta nel pixel i-esimo considerando tutti i mesi;
- T_{min_i} : temperatura minima registrata nel pixel i-esimo considerando tutti i mesi.

A partire dai file di partenza, dunque, si è ottenuta una matrice in formato geotiff in MATLAB (Figura 2.14) che può anche essere caricato in ArcGIS; le caratteristiche di questo file saranno le stesse dei raster di partenza, riportate in Tabella 2.6.

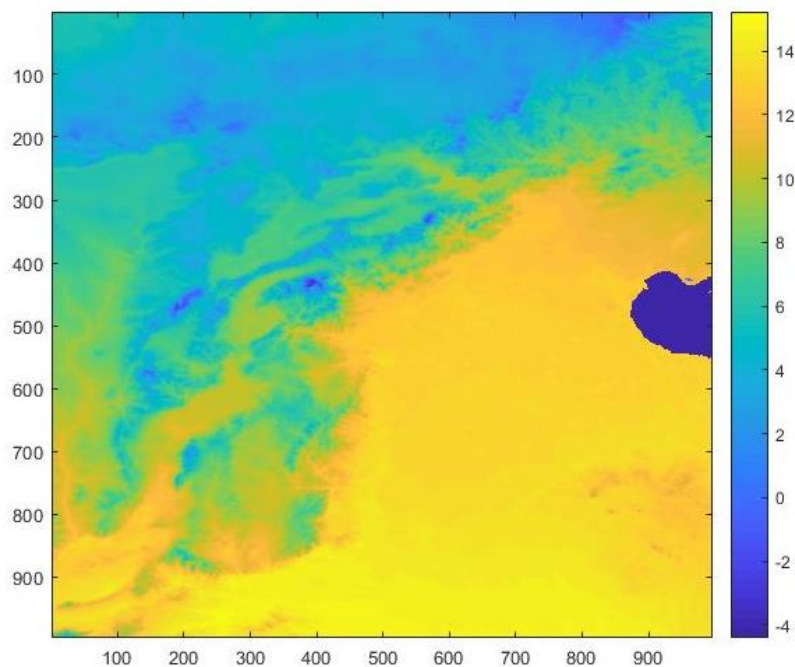


Figura 2.14: Aspetto del file geotiff di temperatura media ottenuto in MATLAB.

Tematismo	Temperatura media, minima e massima (°C)
Tipologia di dato	Raster
Sistema di Riferimento	GCS-WGS84
Risoluzione disponibile	1 km

Tabella 2.6: Caratteristiche dei raster di temperatura media, minima e massima disponibili per i 12 mesi.

2.2.5 Piovosità

L'ultimo parametro quantitativo selezionato, poiché associato a dati numerici e non a caratteristiche qualitative dell'area di studio, è la piovosità. Il clima di una determinata zona non è, infatti, determinato dalla sola temperatura, ma anche da venti, altitudine, vicinanza alla costa e dagli eventi atmosferici.

La piovosità abbondante può favorire l'agricoltura o l'allevamento, ricarica le falde e aumenta la disponibilità idrica ma, influenzando la portata dei corsi d'acqua, può fungere anche da deterrente per lo sviluppo urbano qualora si verificano frequenti allagamenti delle zone limitrofe agli argini. Una piovosità troppo scarsa, d'altro canto, determina sicuramente una disponibilità idrica minore e uno svantaggio per la popolazione.

All'interno del database GAEZ della FAO, è disponibile un raster che riporta per ogni pixel la piovosità media annuale (in millimetri) e, dunque, il dato di interesse è facilmente reperibile⁴⁵.

Come riassunto in Tabella 2.7, la risoluzione disponibile, pari a 10 km, è però inferiore a quella selezionata per lo studio, pari a 1 km.

Tematismo	Piovosità (mm)
Tipologia di dato	Raster
Sistema di Riferimento	GCS-WGS84
Risoluzione disponibile	10 km

Tabella 2.7: Caratteristiche del raster di piovosità.

Utilizzando ArcGIS, si è creato dunque un nuovo raster fatto di pixel di dimensioni compatibili con i precedenti, in modo da avere lo stesso numero di pixel da studiare. Il tool di ArcGIS che permette di eseguire questo passaggio è *Resample* all'interno del *Raster Processing* nella sezione *Data Management Tools*, grazie al quale è possibile suddividere l'area di 100 km² di un pixel iniziale (10 km x 10 km) in 100 pixel da 1 km² assegnando a ogni nuovo pixel il valore di quello di partenza (Figura 2.15).

In questo modo, il raster creato ha i pixel della dimensione voluta, ma ciò non aumenta la sua precisione, poiché non si sono assegnati nuovi valori ai pixel finali, ma si è mantenuto quello di partenza. Il vantaggio è, quindi, di avere lo stesso numero di dati da poter confrontare con gli altri raster, anche se i dati di pioggia saranno “fittizi” perché ottenuti copiando il valore riportato a una risoluzione minore su un numero maggiore di pixel.

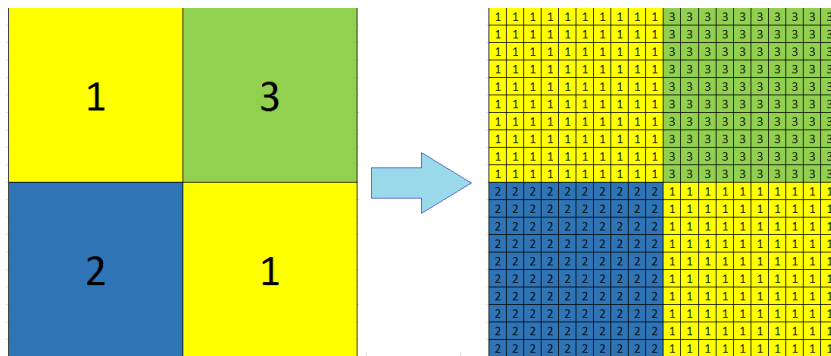


Figura 2.15: Esempio di funzionamento dello strumento Resample. Ogni pixel è diviso in celle di 1 km² assegnando a ogni cella il valore iniziale del pixel di 100 km².

In Figura 2.16 è rappresentato il raster finale, che mostra come nell’area di studio siano presenti diverse zone di piovosità, con una buona variabilità di valori. Confrontandolo con la Figura 2.13, osserviamo che le piovosità maggiori si raggiungono a sud, dove le temperature sono mediamente più miti, mentre a nord, dove le temperature sono abbastanza rigide, si verifica una piovosità molto minore.

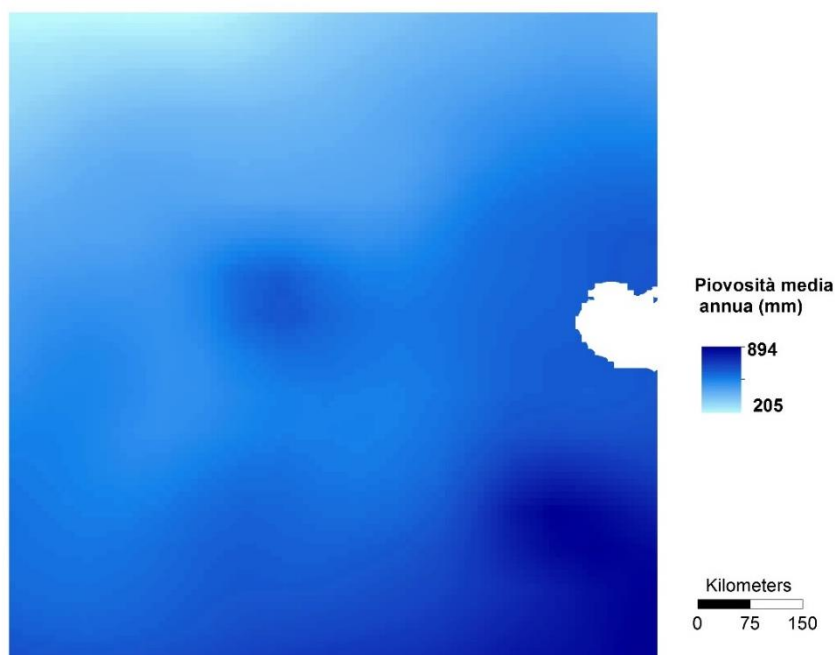


Figura 2.16: File raster finale di piovosità.

2.2.6 Land Cover

C. Small e J. E. Cohen, studiando i parametri che influenzano la distribuzione di popolazione a livello globale (2004)⁴⁶, avevano suggerito di includere in eventuali studi successivi la tipologia di copertura di suolo. L'intuizione di base è che la morfologia del terreno e la sua costituzione possano favorire l'insediamento umano, è il caso di pianure erbose o colline argillose facili da modellare, o che possano renderlo impossibile, ad esempio ampie zone calanchive spoglie e franabili oppure zone rocciose e desertiche.

Sono disponibili diversi raster per la copertura del suolo, derivati dalle foto satellitari a 36 bande spettrali di MODIS, satellite di proprietà della NASA. Un esempio, quello utilizzato per lo studio, è il Global Land Cover by National Mapping Organizations (GLCNMO)⁴⁷, che deriva dalle fotografie di MODIS per l'anno 2013 e prevede 20 campi diversi (Tabella 2.8).

Valore	Tipo di dato	Valore	Tipo di dato
1	Broadleaf Evergreen Forest	11	Cropland
2	Broadleaf Deciduous Forest	12	Paddy field
3	Needleleaf Evergreen Forest	13	Cropland / Other Vegetation Mosaic
4	Needleleaf Deciduous Forest	14	Mangrove
5	Mixed Forest	15	Wetland
6	Tree Open	16	Bare area,consolidated(gravel,rock)
7	Shrub	17	Bare area,unconsolidated (sand)
8	Herbaceous	18	Urban
9	Herbaceous with Sparse Tree/Shrub	19	Snow / Ice
10	Sparse vegetation	20	Water bodies

Tabella 2.8: Tipologie di copertura del suolo disponibili per il raster utilizzato nello studio

Si tratta di una rappresentazione molto dettagliata, che non interessa più di tanto per questo studio, tanto più che si tratta di un dato qualitativo che non potrà essere utilizzato per elaborazioni matematiche, a meno che non si decida di assegnare dei valori numerici ad ogni campo secondo una convenzione da ricercare.

In questo elaborato, verrà utilizzato solo per considerazioni generali sull'area di studio ma non sarà inserito come parametro del modello. Per questo motivo, si è deciso di accorpare tipologie di copertura del suolo simili tra loro per creare una mappa più semplice e immediata, che consenta rapide considerazioni sull'area di studio:

- I valori da 1 a 6 sono stati classificati sotto un'unica voce: "foreste";

- I valori da 7 a 10 sono stati raggruppati alla voce “vegetazione sparsa”;
- I valori da 11 a 13 sono stati classificati come “terreni agricoli”;
- I valori 14 e 15 sono stati raggruppati come “aree umide”;
- I valori 16 e 17 sono stati classificati come “aree spoglie”;
- I valori 18, 19 e 20 si sono tenuti separati e conformi alla classificazione originaria.

Con queste semplificazioni, il raster di copertura del suolo (le cui caratteristiche sono riportate in Tabella 2.9) assume l'aspetto mostrato in Figura 2.17.

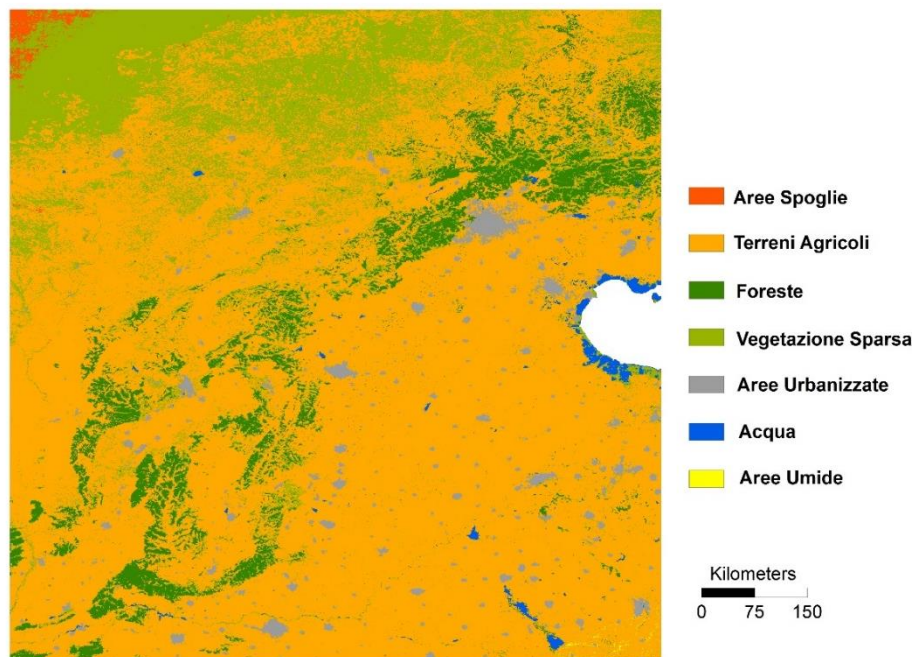


Figura 2.17: File raster di copertura del suolo nell'area di studio.

Si nota innanzitutto la grande presenza di terreni agricoli o campi coltivati che coprono la quasi totalità dell'area di studio. Foreste e aree a vegetazione sparsa si presentano solo in corrispondenza delle quote più elevate (si noti confrontando con il DEM in Figura 2.5), mentre a nord-ovest, dove la precipitazione risulta minore si incontrano aree spoglie. Il territorio si presenta quindi notevolmente antropizzato, con larga presenza di campi ad uso agricolo e ampie e frequenti aree urbanizzate.

Tematismo	Land Cover
Tipologia di dato	Raster
Sistema di Riferimento	GCS-WGS84
Risoluzione disponibile	1 km

Tabella 2.9: Caratteristiche del raster rappresentante la copertura di suolo

3 Elaborazione preliminare dei dati per l'individuazione delle caratteristiche del modello di espansione urbana

3.1 La definizione di classi di parametri per lo studio delle luci notturne

Per ottenere un modello di espansione urbana che descriva l'evoluzione dell'espansione urbana utilizzando dati satellitari di luminosità artificiale notturna, è necessario dapprima osservare come tali valori di luminosità siano legati alle singole variabili ambientali, di natura idrologica e geomorfologica. Al Capitolo 2 sono state avanzate diverse ipotesi basate sulle osservazioni delle mappe ottenute dalle immagini satellitari; si deve però procedere a una verifica numerica per osservare come varia esattamente il comportamento delle luci artificiali notturne al variare dei fattori ambientali individuati. Quel che si vuole ottenere, quindi, è un legame tra il valore della luminosità in funzione dell'elevazione, della distanza dai fiumi e dalla costa, della temperatura, della piovosità e della copertura del suolo. Prima di procedere, sono necessarie alcune considerazioni preliminari.

L'area di studio è individuata da un quadrato di lato 1000 km (a meno di qualche errore di approssimazione nel ritaglio dovuta alla conversione da gradi a chilometri lineari) con pixel di risoluzione 1 km; dunque i pixel da studiare in totale sono circa 1 milione. L'identificazione di una relazione tra la luminosità notturna e i vari parametri utilizzando i pixel singolarmente sarebbe molto dispendioso in termini di tempo di calcolo e i risultati potrebbero essere di difficile interpretazione per l'eccessiva quantità di punti da mostrare in grafico. Dunque, è necessario raggruppare in qualche modo i dati per osservarne il comportamento generale, rendendo così i calcoli meno onerosi e i risultati più chiari.

Si è deciso, quindi, di creare classi in cui suddividere i dati delle variabili ambientali e di calcolare, per ognuna di esse, la luminosità media. Il procedimento consiste nel calcolare la media di intensità luminosa dei soli pixel appartenenti a ogni classe, in modo da produrre un andamento medio della luminosità all'aumentare o diminuire dei valori dei parametri.

Logicamente, questo dato non riprodurrà perfettamente il comportamento delle luci e inoltre i risultati che si otterranno saranno fortemente influenzati dalla scelta delle classi poiché, cambiandone le estensioni, cambia il numero di pixel considerato e, dunque, la media di luminosità. Data la numerosità dei dati da studiare, questo procedimento si prospetta però necessario e può comunque essere indicativo del comportamento generale che la variabile assume al variare delle condizioni ambientali.

Nei prossimi paragrafi verrà spiegato come sono state costruite le classi per ogni campo di studio e come si presenta la luminosità media al variare di esse. Questi comportamenti saranno indici del legame tra le luci notturne e i fattori ambientali, saranno quindi utili per definire le funzioni del modello integrato al Capitolo 4.

3.1.1 Elevazione: suddivisione in classi

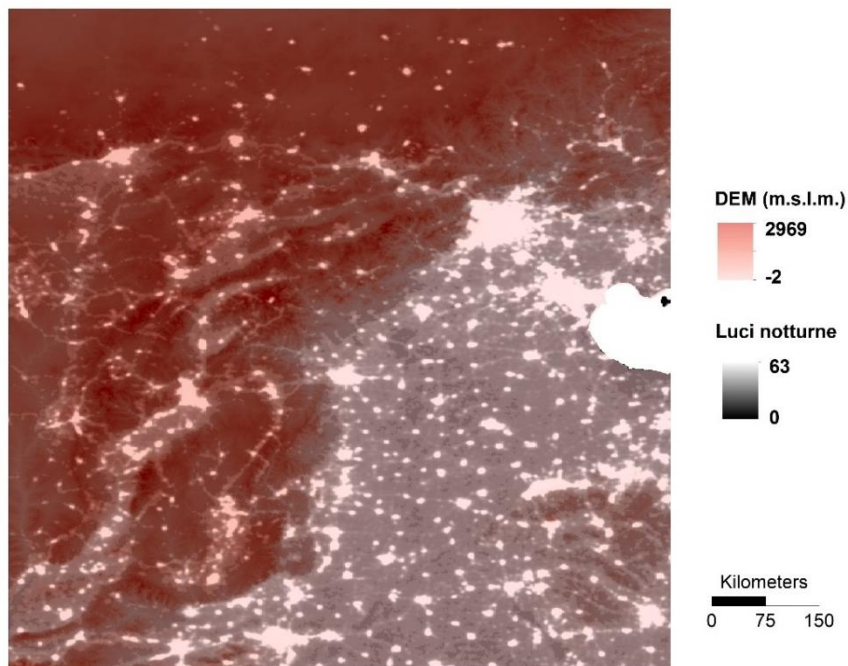


Figura 3.1: Disposizione delle luci notturne confrontata con l'andamento delle quote del terreno.

In Figura 3.1 sono stati sovrapposti i raster di luminosità artificiale notturna e di DEM per individuare se esistono quote preferenziali per la distribuzione di popolazione. In effetti, le luci più intense (ovvero le città più estese e sviluppate) si dispongono in corrispondenza delle quote più basse mentre in corrispondenza delle catene montuose le luci si diradano e diventano quasi assenti in corrispondenza dei picchi più elevati.

Da queste considerazioni abbastanza qualitative possiamo immaginare che le classi di elevazione dovranno essere più fitte sulle quote basse per vedere eventuali variazioni della luminosità media, mentre si potranno costruire classi più estese per le quote maggiori, dove comunque la luminosità appare molto bassa.

Le classi ipotizzate per l'elevazione sono riportate in Tabella 3.1, mentre la Figura 3.2 rappresenta il confronto tra il raster DEM così classificato e la luminosità notturna. Come già osservato in Figura 3.1, l'intensità luminosa maggiore appartiene senza dubbi alla prima classe, mentre oltre i 1500 metri sul livello del mare le luci sono quasi assenti.

Range di variabilità	Da -2 a 2969 m.s.l.m.
Classe 1	Da -2 a 20 m.s.l.m.
Classe 2	Da 20 a 50 m.s.l.m.
Classe 3	Da 50 a 500 m.s.l.m.
Classe 4	Da 500 a 1500 m.s.l.m.
Classe 5	Oltre i 1500 m.s.l.m.

Tabella 3.1: Classi per la suddivisione dei dati di DEM.

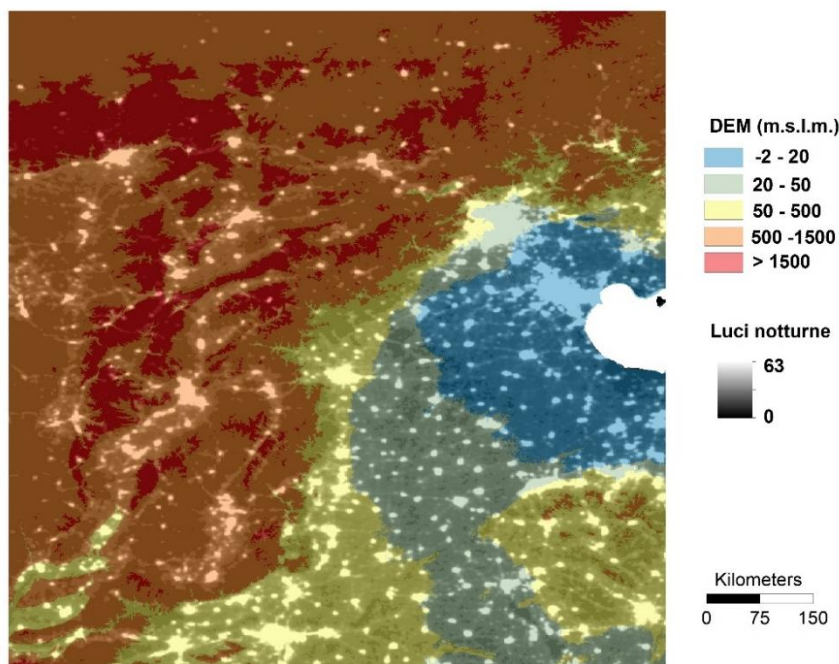


Figura 3.2: Confronto tra la luminosità notturna e le classi create per il DEM.

Per ottenere la luminosità media in ogni classe, si è costruito un codice di elaborazione numerica in MATLAB per isolare i soli punti appartenenti a ogni classe e poi calcolarne l'intensità luminosa media. Il procedimento prevede di caricare in MATLAB i due file raster in formato *geotiff*, in questo modo il programma legge i valori di ogni pixel come elementi di matrici 1000 per 1000 (nella pratica 996 per 996 per via delle approssimazioni) e può compiere su di esse le elaborazioni numeriche necessarie, oltre che rappresentarle come immagini.

Si lavora quindi sulla matrice di DEM creando delle maschere che rappresentino singolarmente le classi riportate in Tabella 3.1, su cui poi verrà ritagliata la matrice di intensità luminosa, in modo da poter calcolare la media dei soli elementi di luce appartenenti a una data classe di elevazione. Per ogni classe in cui è stato suddiviso il DEM, la maschera è definita come segue: le quote vengono

sostituite dai valori 1, rappresentativi dei dati che ricadono all'interno della classe considerata, e NaN, se i dati non rispettano le condizioni. Basterà poi eseguire una moltiplicazione elemento per elemento tra le maschere così composte e le matrici di luminosità per ottenere nuove matrici di luminosità, una per ogni anno, che riportano i valori di intensità luminosa solo nei pixel che appartengono a una certa classe di elevazione e NaN altrove.

A questo punto, è possibile calcolare la media della luminosità di ogni classe, senza considerare i NaN (utilizzando la funzione *nanmean*). Automatizzando il processo con l'utilizzo di strutture (*struct*), è possibile caricare tutti i raster di luminosità disponibili per i 22 anni e tutte e cinque le classi di elevazione per ottenere una matrice finale che riporta sulle colonne i valori medi di luminosità per ogni classe in un determinato anno e sulle righe i valori medi di luminosità di una singola classe per tutti gli anni di studio (Tabella 3.2).

Luminosità media per ogni anno per classi di DEM					
Classi (mslm)	-2 - 20	20 - 50	50 - 500	500 - 1500	>1500
Anni					
1992	8.01	6.92	6.61	1.78	0.19
1993	8.53	7.23	7.05	2.09	0.26
1994	8.89	7.84	7.35	2.18	0.31
1995	9.85	8.86	8.46	2.44	0.35
1996	10.21	8.87	8.01	2.3	0.29
1997	9.7	8.91	8.36	2.25	0.29
1998	9.9	9.09	8.35	2.22	0.31
1999	9.05	8.49	7.63	1.92	0.27
2000	9.53	8.76	7.82	2.16	0.35
2001	10.3	9.1	8.08	2.14	0.33
2002	10.38	9.74	8.85	2.25	0.34
2003	10.42	9.39	8.79	2.28	0.38
2004	11.67	10.48	10.05	2.91	0.58
2005	11.54	9.6	9.85	2.8	0.49
2006	12	10.73	10.74	2.89	0.5
2007	12.45	11.19	10.88	2.81	0.48
2008	12.26	11.18	10.91	2.94	0.49
2009	14.27	13.34	12.06	3.06	0.39
2010	14.24	12.92	12.13	3.24	0.54
2011	14.12	13.11	12.39	3.64	0.84
2012	13.82	12.83	11.81	3.64	0.81
2013	16.19	14.48	13.19	3.71	0.77

Tabella 3.2: Risultati della luminosità media per ogni classe di DEM e per ogni anno di osservazione.

I risultati numerici confermano le ipotesi preliminari, la luminosità media diminuisce man mano che l'altitudine aumenta. Complessivamente, si osserva un aumento della luminosità media in ogni classe

dal 1992 al 2013, sebbene nelle prime classi si noti un incremento molto maggiore rispetto alle ultime. In Figura 3.3 sono riportati per brevità i grafici della luminosità media per le classi di elevazione nel primo e ultimo anno di osservazione (gli anni intermedi mostrano andamenti simili).

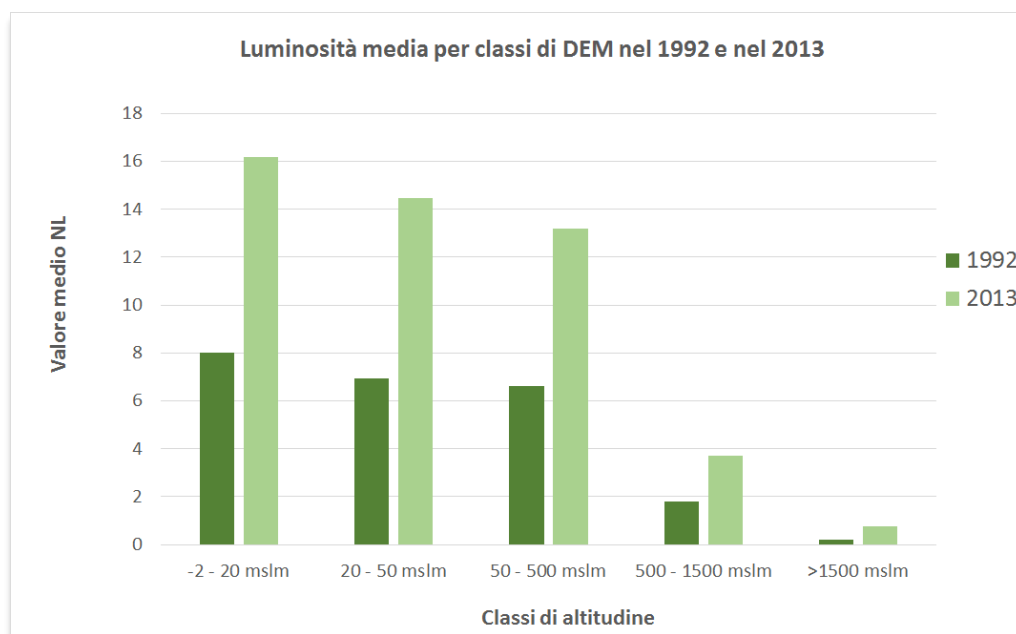


Figura 3.3: Andamento della luminosità media per classi di DEM nel 1992 e nel 2013.

Inoltre, in Figura 3.4, possiamo osservare l'andamento della luminosità media di ogni classe di DEM nel tempo per osservare le variazioni che hanno coinvolto le diverse classi di elevazione. Dal grafico, si nota chiaramente che le classi maggiormente interessate dalla crescita di luminosità (e cioè di urbanizzazione) sono le prime tre, mentre sopra i 500 metri le luci non subiscono grandi variazioni e risultano quasi costanti negli anni.

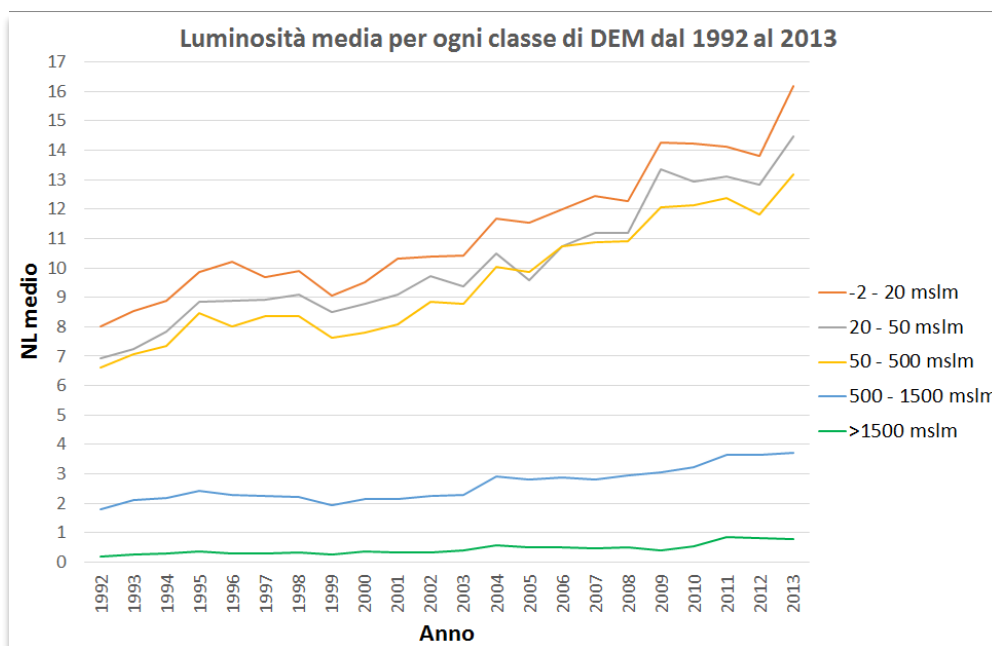


Figura 3.4: Andamento della luminosità media per ogni classe di DEM nel tempo.

In generale, quindi, la variazione di luminosità notturna diminuisce all'aumentare dell'altitudine e nel modello si dovrà tenere quindi conto di questo comportamento per costruire una previsione spaziale e temporale dell'aumento di urbanizzazione negli anni futuri. Come osservato dalle immagini satellitari, nell'area in esame la luminosità è cresciuta notevolmente nell'arco degli anni osservati; l'aumento ha interessato però soprattutto le aree ad altitudine minore, portando quindi a un aumento della densità di popolazione nelle aree pianeggianti o collinari.

3.1.2 Distanza dalla costa: suddivisione in classi

Seguendo le considerazioni accennate al Capitolo 2, ci si aspetta che la luminosità notturna di un'area sia maggiore se questa si trova in prossimità della costa, mentre potrebbe diminuire man mano che la distanza aumenta.

Queste ipotesi sono confermate da uno studio analogo a questo effettuato da C. Small nel 2004⁴⁸, in cui si conclude che *"the resultant average population densities show strong localization within 100 km of seacost"*. Quindi la distanza dalla costa sembra influenzare la presenza umana fino a un raggio di 100 km. Nel caso in esame la distanza dalla costa varia da 0 a 1000 km; seguendo quindi l'esempio in letteratura, le classi di distanza sono state scelte abbastanza ampie, dell'ordine delle centinaia di chilometri (Tabella 3.3).

Range di variabilità	Da 0 a 1000 km
Classe 1	Da 0 a 100 km
Classe 2	Da 100 a 300 km
Classe 3	Da 300 a 500 km
Classe 4	Oltre i 500 km

Tabella 3.3: Classi di distanza dalla costa.

Il procedimento per il calcolo della luminosità media di ogni classe è analogo a quello descritto per l'elevazione; in ambiente MATLAB si sono create le maschere che seguissero le condizioni di ogni classe e poi si è creata la nuova matrice di luminosità moltiplicando le maschere per le matrici di luminosità di partenza.

I risultati sono riportati in Tabella 3.4, mentre la Figura 3.5 riporta il confronto tra l'andamento della luminosità media per classe nel 1992 e nel 2013.

Luminosità media per ogni anno per classi di Distanza dalla Costa				
Classi (km)	<100	100 - 300	300 – 500	> 500
Anni				
1992	9.04	5.91	2.69	2.65
1993	9.76	6.19	3.09	2.92
1994	9.68	6.6	3.26	3.09
1995	10.96	7.44	3.75	3.47
1996	11.41	7.24	3.68	3.27
1997	10.93	7.19	3.8	3.29
1998	10.93	7.33	3.89	3.23
1999	10.04	6.71	3.56	2.87
2000	10.48	7.03	3.68	3.11
2001	11.43	7.49	3.7	3.09
2002	11.45	7.88	3.99	3.35
2003	12.11	7.81	3.66	3.44
2004	13.32	8.96	4.26	4.17
2005	13.38	8.62	3.91	4.04
2006	14.05	9.15	4.37	4.3
2007	14.66	9.39	4.67	4.14
2008	14.54	9.19	4.73	4.32
2009	16.59	10.65	5.65	4.4
2010	17.19	10.58	5.49	4.59
2011	16.91	10.49	5.84	5.1
2012	16.56	10.04	5.59	5.11
2013	19.19	11.47	6.04	5.41

Tabella 3.4: Risultati della luminosità media per ogni classe di distanza dalla costa e per ogni anno di osservazione.

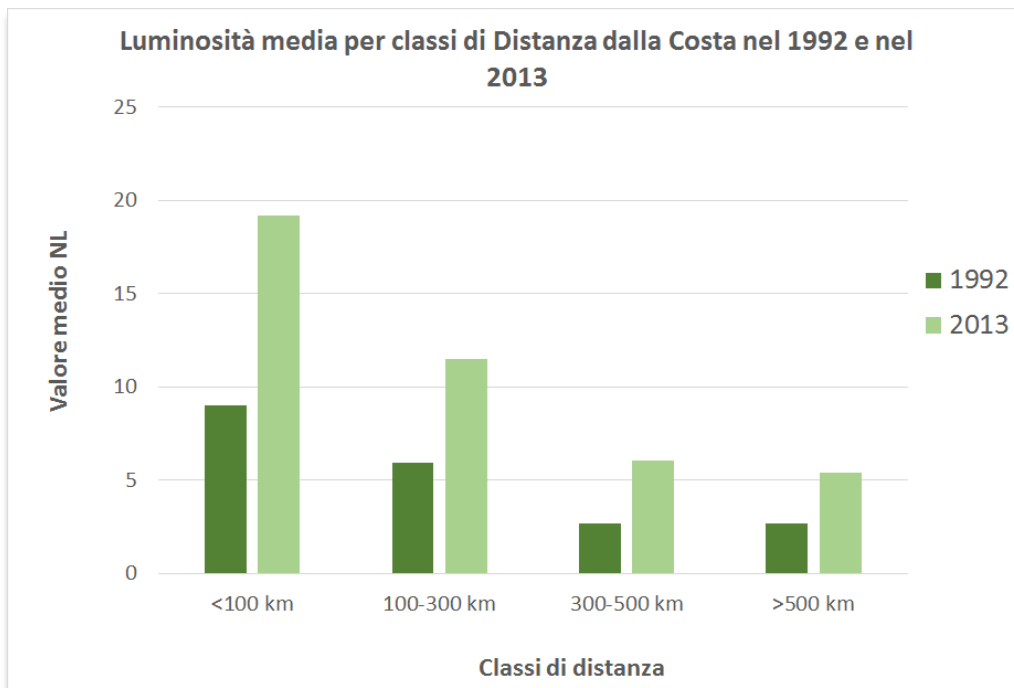


Figura 3.5: Andamento della luminosità media per classi di distanza dalla costa nel 1992 e nel 2013.

Dai risultati si evince che la luminosità media diminuisce all'aumentare della distanza dal mare, come da ipotesi iniziale; il mare si conferma un parametro importante per l'uomo e ne influenza l'insediamento. Si può pensare che questo fosse vero soprattutto in passato, quando la navigazione rappresentava il collegamento più veloce tra i porti più importanti ma nel presente, con lo sviluppo dei mezzi di trasporto su gomma, rotaia e aerei, non è detto che questo comportamento sia confermato.

Per verificare se ancora oggi i punti in cui la densità di popolazione è in aumento sono quelli limitrofi alla costa, si può produrre un grafico dell'andamento della luminosità media per ogni classe nel tempo (Figura 3.6). Effettivamente, la classe che vede la luminosità aumentare in maniera più rapida è sempre la più vicina alla linea di costa. Anche le altre classi hanno un aumento evidente; a differenza dell'elevazione, che vedeva la luminosità nelle classi a quote maggiori rimanere quasi costanti, le classi più lontane dalla costa vedono un aumento di luminosità più marcato.

Anche in questo caso, la funzione che lega i valori di luminosità alla distanza dalla costa da inserire all'interno del modello sarà comunque decrescente all'aumentare di tale distanza.

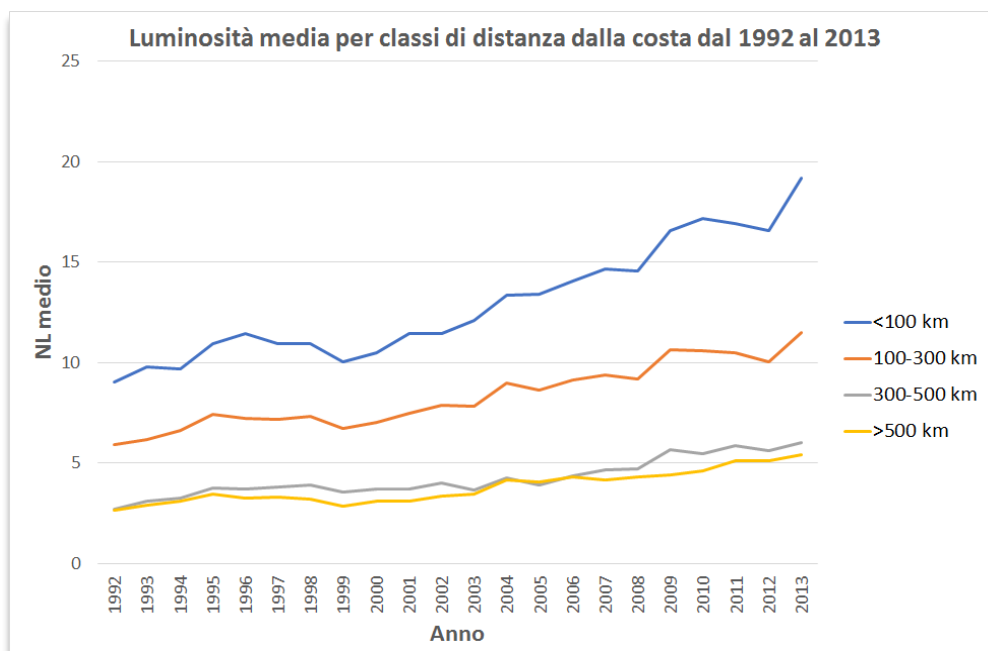


Figura 3.6: Andamento della luminosità media per ogni classe di distanza dalla costa nel tempo.

3.1.3 Distanza dai corsi d'acqua: suddivisione in classi

Seguendo gli stessi procedimenti esposti per la quota e l'elevazione e la distanza dalla costa, si esamina ora la variazione di luminosità media al variare della distanza dai corsi d'acqua principali. L'ipotesi era anche in questo caso di osservare un andamento decrescente della luminosità notturna media all'allontanarsi dai fiumi.

Inizialmente, si era ipotizzato che la distanza dai corsi d'acqua oltre la quale le luci risultassero meno intense fosse abbastanza piccola, entro 10 km. Gli andamenti che si ottengono con una classificazione chilometro per chilometro fino a una distanza di 10 km non sono, però, soddisfacenti poiché mostrano una diminuzione poco marcata per i primi chilometri e non danno informazioni sui punti più lontani, che si trovano a quasi 200 km di distanza dal corso d'acqua più vicino (Figura 3.7). L'informazione sulla disposizione urbana in prossimità dei corsi d'acqua può risultare comunque utile per analizzare i comportamenti a piccola scala, ma per un'area di studio di 1000 km di lato è più utile osservare classi di distanza più ampie. Seguendo lo studio di C. Small 2004 sopra citato⁴⁹, in cui si sostiene che la distribuzione di popolazione non dipenda più dalla posizione rispetto ai corsi d'acqua oltre i 100 km, si sono quindi selezionate classi di estensione maggiore (Tabella 3.5).

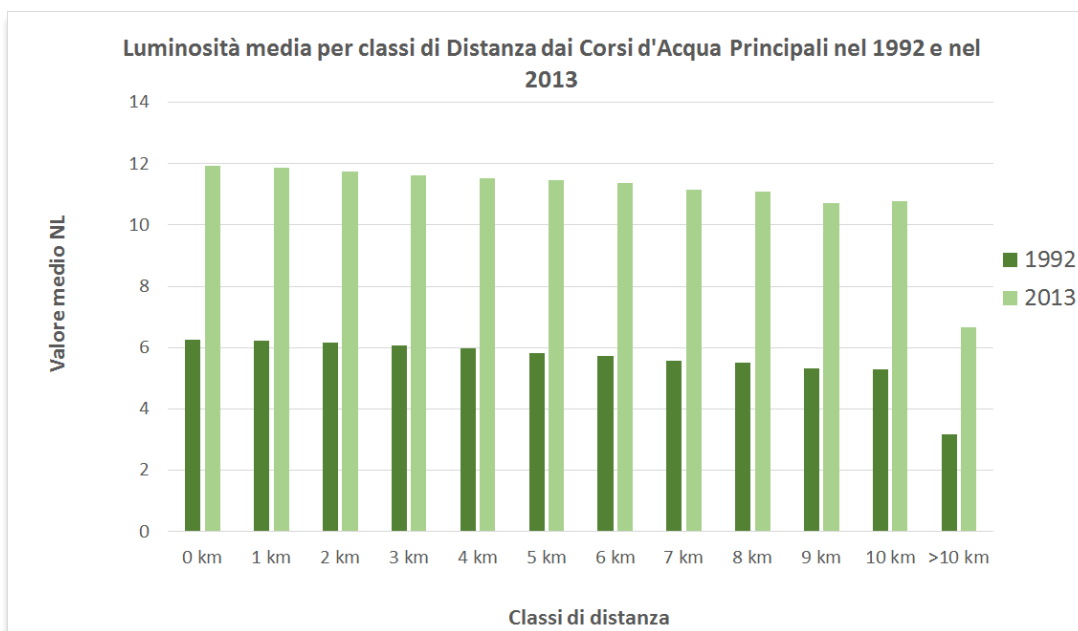


Figura 3.7: Andamento della luminosità media nel 1992 e nel 2013 per ogni chilometro di distanza dai corsi d'acqua.

Range di variabilità	Da 0 a 180 km
Classe 1	Da 0 a 10 km
Classe 2	Da 10 a 50 km
Classe 3	Da 50 a 100 km
Classe 4	Oltre i 100 km

Tabella 3.5: Classi di distanza dai corsi d'acqua.

Utilizzando queste nuove classi, si ottengono andamenti che decrescono velocemente tra la prima classe e le successive e mostrano l'effetto della distanza dai fiumi anche per i punti più lontani (Figura 3.8). I risultati numerici di luminosità media per ogni classe e per ogni anno sono riportati in Tabella 3.6.

Luminosità media per ogni anno per classi di Distanza dai Corsi d'Acqua				
Classi (km)	<10	10 - 50	50 – 100	> 100
Anni				
1992	5.77	3.81	1.45	0.93
1993	6.21	4.13	1.69	1.2
1994	6.56	4.36	1.72	1.19
1995	7.35	4.89	2.09	1.59
1996	7.19	4.82	1.83	1.47
1997	7.07	4.81	1.97	1.72
1998	7.17	4.8	2	1.82
1999	6.58	4.36	1.74	1.6
2000	6.89	4.6	1.89	1.73
2001	7.2	4.74	1.9	1.74
2002	7.57	5.03	2.11	1.99
2003	7.51	5	2.11	2.08
2004	8.6	5.84	2.61	2.24
2005	8.34	5.63	2.4	2.23
2006	8.89	6	2.68	2.55
2007	8.94	6.09	2.79	2.64
2008	9.07	6.15	2.75	2.43
2009	10.06	6.85	3.21	2.61
2010	10.13	6.95	3.2	2.56
2011	10.39	7.26	3.47	2.69
2012	10.24	7.06	3.31	2.67
2013	11.34	7.78	3.6	3.24

Tabella 3.6: Risultati della luminosità media per ogni classe di distanza dai corsi d'acqua e per ogni anno di osservazione.

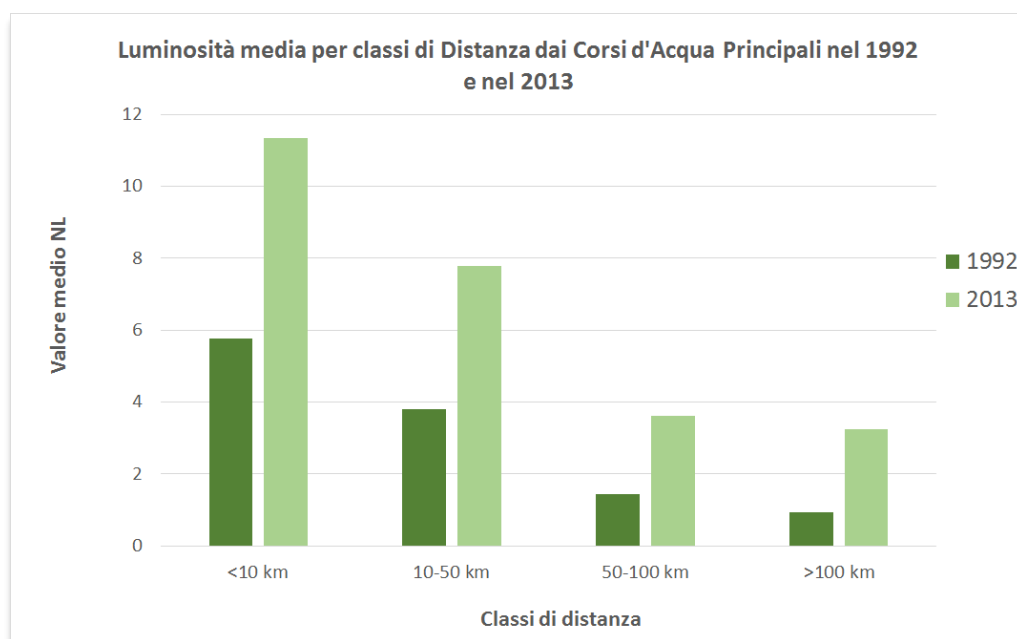


Figura 3.8: Andamento della luminosità media per classi di distanza dai corsi d'acqua nel 1992 e nel 2013.

L'ipotesi iniziale è comunque confermata da entrambe le scelte di classificazione; la luminosità media tende a diminuire man mano che la distanza dai corsi d'acqua aumenta e quindi si dovrà inserire nel modello una funzione decrescente anche per questo parametro.

In Figura 3.9 è invece rappresentato l'andamento delle classi nel tempo.

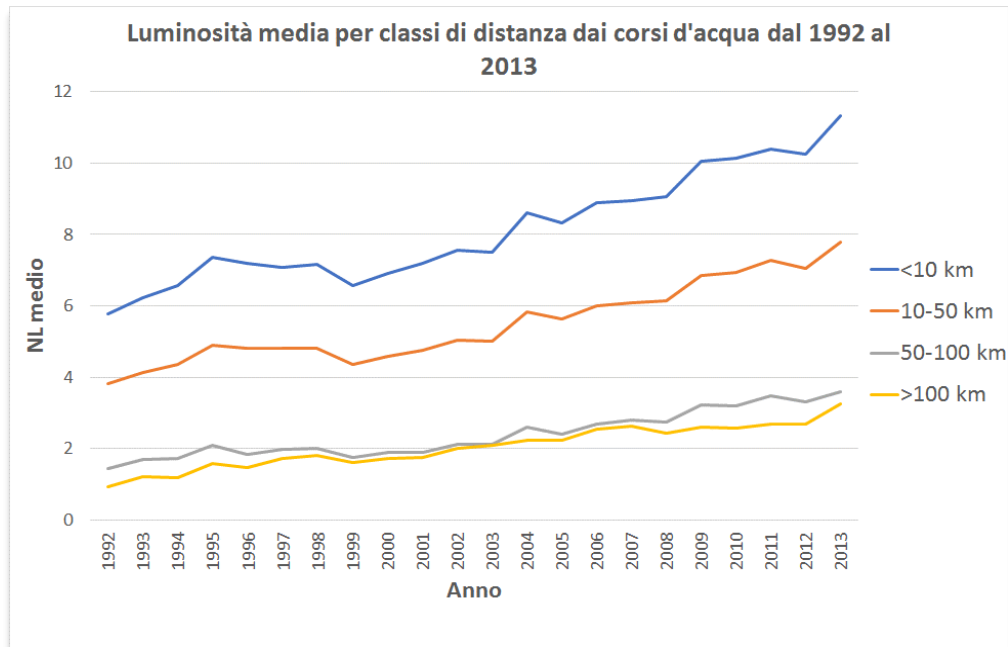


Figura 3.9: Andamento della luminosità media per ogni classe di distanza dai corsi d'acqua nel tempo.

Anche in questo caso, la crescita di luminosità ha interessato soprattutto le classi più vicine ai corsi d'acqua ma anche le ultime classi mostrano una crescita evidente, lievemente meno marcata rispetto all'andamento di quelle della distanza dalla costa.

Forse la luminosità cresce anche ad elevate distanze dai fiumi grazie alla presenza di abbondante risorsa idrica sotterranea; in generale, comunque, il trend conferma un aumento della luminosità notturna negli anni, che diventa meno marcato per le ultime classi.

3.1.4 Temperatura: suddivisione in classi

Per quanto riguarda la temperatura, si erano individuati 2 possibili parametri che potessero influenzare la luminosità: la temperatura media annuale in ogni pixel e l'escursione termica massima annuale in ogni pixel. Il processo di calcolo della luminosità media per classi illustrato nei casi precedenti è stato ripetuto per entrambi i parametri in modo da valutarne l'effettiva utilità.

Il campo di variazione della temperatura media va da -5°C a +14°C, per cui le classi avranno un range abbastanza piccolo (dell'ordine di qualche grado) e sono riportate in Tabella 3.7.

Range di variabilità	Da -5 a +14 °C
Classe 1	Meno di 5 °C
Classe 2	Da 5 a 8 °C
Classe 3	Da 8 a 10 °C
Classe 4	Oltre i 10 °C

Tabella 3.7: Classi di temperatura media annuale.

Contrariamente alle forzanti precedenti, ci si aspetta questa volta che la luminosità aumenti all'aumentare della temperatura, tanto più che si stanno esaminando zone fredde o temperate in cui la temperatura media annuale va da bassa a mite, e non raggiunge mai valori tanto elevati da scoraggiare l'insediamento umano.

Questa ipotesi è confermata dai risultati numerici, in Figura 3.10 è riportato l'andamento della luminosità media di ogni classe.

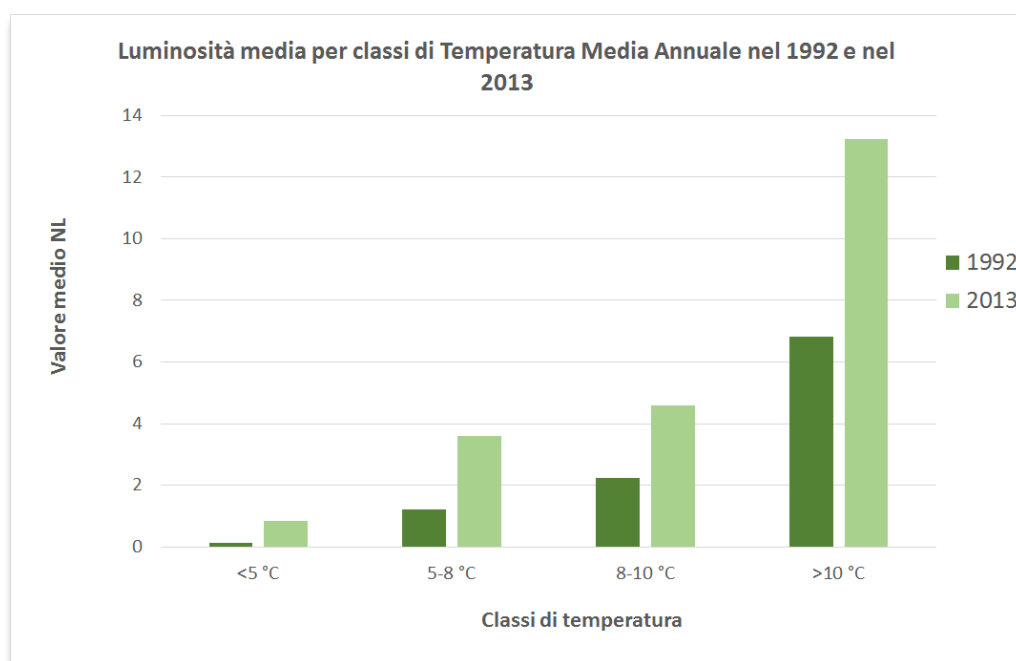


Figura 3.10: Andamento della luminosità media per classi di temperatura media annuale nel 1992 e nel 2013.

L'escursione termica massima annuale presenta un range di variabilità compreso tra 0.4 °C e 54 °C e le classi scelte sono riportate in Tabella 3.8.

Range di variabilità	Da 0.4 a 54 °C
Classe 1	Meno di 30 °C
Classe 2	Da 30 a 40 °C
Classe 3	Da 40 a 50 °C
Classe 4	Oltre i 50 °C

Tabella 3.8: Classi di escursione termica massima annuale.

Dalla Figura 3.11, si nota invece che l'andamento della luminosità media in funzione dell'escursione termica non segue un trend preciso come per la temperatura media; se nel 1992 notiamo una funzione decrescente, indice che la popolazione tenderebbe a disporsi preferibilmente nelle zone in cui la variazione climatica è minore, nel 2013 (ed anche in tutti gli altri anni di osservazione) essa mostra un andamento a campana. Per spiegare il comportamento variabile dettato dall'escursione termica, si può ipotizzare esistano, all'interno dell'area di studio, zone con escursione termica moderata ma a temperature miti e altresì zone con escursione termica ridotta ma con temperature più rigide. La popolazione pare preferisca insediarsi dove le temperature sono per lo più miti e non necessariamente dove l'escursione termica è contenuta, significa che è portata a privilegiare sia zone ad escursioni termiche medio-basse sia zone dove queste risultano più elevate, se le temperature sono complessivamente più elevate. È il caso dei climi temperati, dove estati ed inverni vedono temperature variabili ma mai estreme, contrapposti ai climi rigidi in cui la temperatura varia poco durante l'anno ma rimane sempre piuttosto bassa.

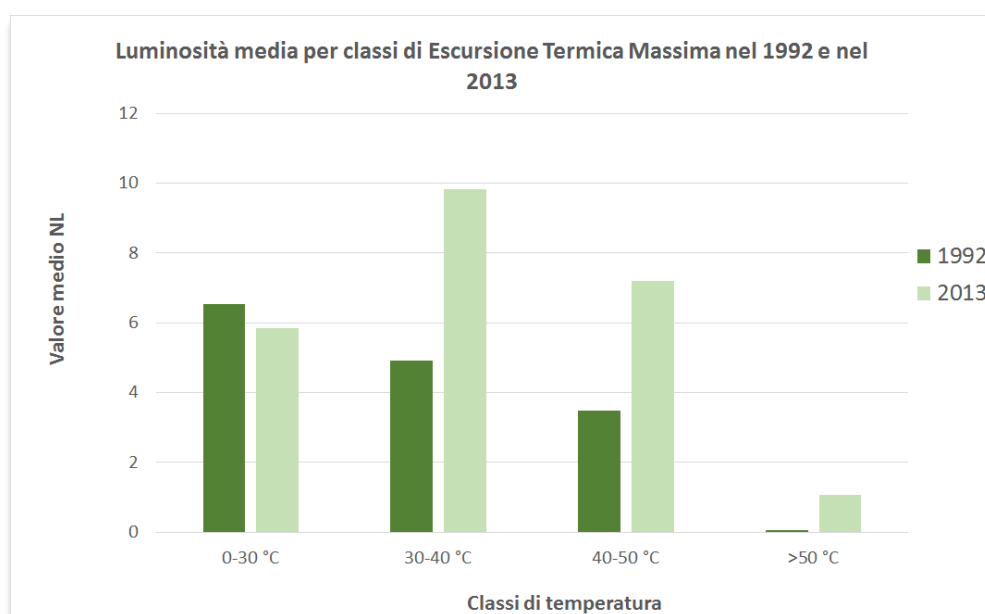


Figura 3.11: Andamento della luminosità media per classi di escursione termica massima annuale nel 1992 e nel 2013.

Visti gli andamenti dei valori di luminosità media in funzione della temperatura media annuale e dell'escursione termica massima annuale, nel presente elaborato di tesi, ai fini della stesura del modello integrato, Per avere un parametro dall'andamento più chiaro e immediato, in modo da poter scegliere senza dubbi una funzione della luminosità notturna, si è selezionata solamente la temperatura media annuale, scartando quindi l'escursione termica massima annuale. I risultati di luminosità media per le classi di temperatura media annuale, sono riportati per esteso in Tabella 3.9.

Luminosità media per ogni anno per classi di Temperatura Media Annuale				
Classi (°C)	<5	05 - 08	08 - 10	>10
Anni				
1992	0.12	1.2	2.25	6.81
1993	0.19	1.66	2.62	7.19
1994	0.18	1.66	2.73	7.62
1995	0.21	1.91	3.02	8.6
1996	0.19	1.77	2.93	8.42
1997	0.17	1.77	2.8	8.46
1998	0.2	1.85	2.73	8.48
1999	0.17	1.54	2.35	7.79
2000	0.24	1.85	2.64	8.06
2001	0.23	1.79	2.66	8.41
2002	0.24	1.8	2.91	8.94
2003	0.3	1.86	2.88	8.84
2004	0.45	2.48	3.71	10.05
2005	0.43	2.34	3.52	9.69
2006	0.42	2.42	3.64	10.47
2007	0.46	2.47	3.55	10.63
2008	0.49	2.63	3.67	10.63
2009	0.48	2.6	3.89	12.1
2010	0.55	3	4.29	11.95
2011	0.79	3.64	4.58	12.12
2012	0.9	3.65	4.32	11.74
2013	0.84	3.59	4.6	13.24

Tabella 3.9: Risultati della luminosità media per ogni classe di temperatura media annuale e per ogni anno di osservazione.

In Figura 3.12 sono riportati gli andamenti della luminosità media di ogni classe per gli anni di osservazione. Si nota chiaramente come l'unica classe ad aver visto un incremento notevole è quella relativa a temperature superiori ai 10 °C; le altre classi hanno incrementi molto meno marcati, ad esempio, la classe con valori di temperatura media annua inferiori ai 5 °C sembra restare quasi costante, sintomo che la crescita vera e propria è da attribuire alle zone con temperature medie più elevate. L'andamento è simile a quello presentato dall'elevazione; i punti che vedono la luminosità aumentare più rapidamente sono quelli in cui la quota è minore e la temperatura più mite.

Concludendo, la relazione tra luminosità e temperatura media annuale da inserire all'interno del modello sarà quindi crescente all'aumentare della temperatura.

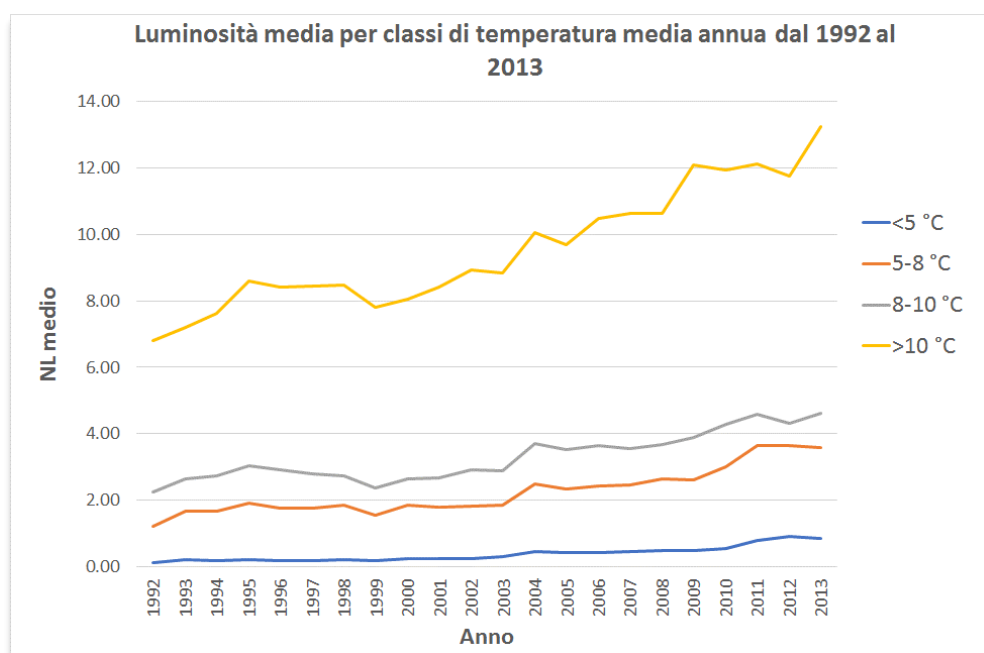


Figura 3.12: Andamento della luminosità media per ogni classe di temperatura media annua nel tempo.

3.1.5 Piovosità: suddivisione in classi

La piovosità rappresenta il parametro per cui è più difficile prevedere l'influenza sulle luci notturne. Abbiamo già evidenziato nel Capitolo 2 i problemi associati a una scarsità o a un eccesso di precipitazione, tanto più che l'area di studio può essere influenzata dai monsoni estivi che potrebbero dare luogo a una piovosità anche troppo violenta.

Suddividendo la piovosità media, che si presenta nell'area di studio con valori da 200 a quasi 900 mm, nelle classi riportate in Tabella 3.10 si ottengono gli andamenti riportati in Figura 3.13

Range di variabilità	Da 205 a 894 mm
Classe 1	Meno di 350 mm
Classe 2	Da 350 a 450 mm
Classe 3	Da 450 a 550 mm
Classe 4	Oltre i 550 mm

Tabella 3.10: Classi di piovosità media.

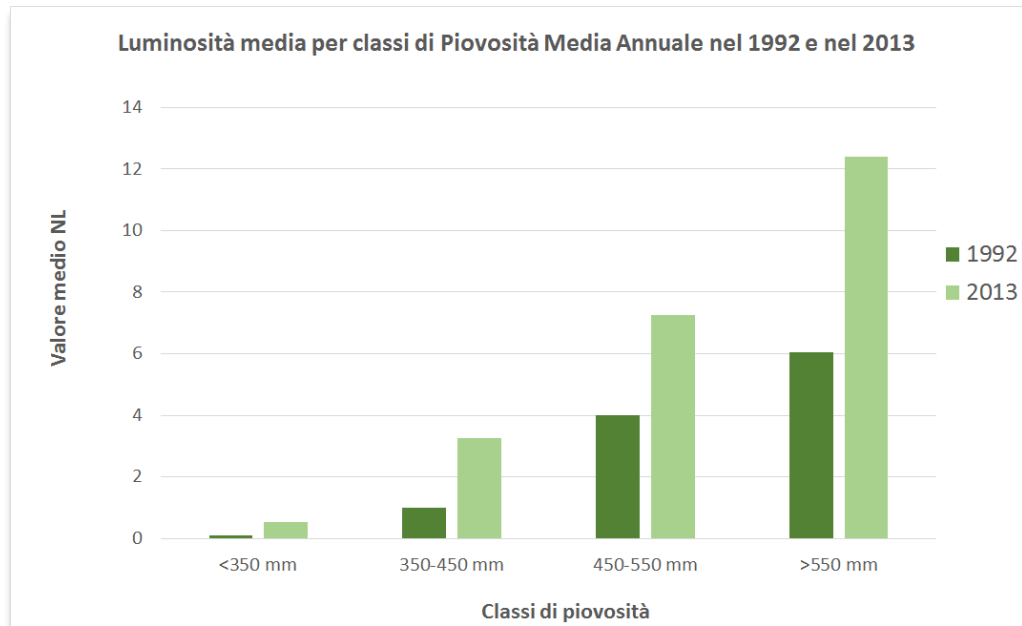


Figura 3.13: Andamento della luminosità media per classi di piovosità media nel 1992 e nel 2013.

Similmente alla dipendenza con la temperatura media, la luminosità notturna tende ad aumentare notevolmente dove le precipitazioni risultano più abbondanti. I dati completi per ogni anno di osservazione sono riportati in Tabella 3.11. Tale andamento dovrà quindi essere opportunamente descritto nel modello integrato mediante una funzione crescente tra luminosità e piovosità.

È interessante notare dalla Figura 3.14 come la classe con valori di precipitazione media annua superiori a 550 mm abbia assistito nel tempo a un forte aumento della luminosità media. Le classi rimanenti presentano invece un aumento poco accennato, tanto che la luminosità della classe con precipitazioni medie annue inferiori a 350 mm risulta quasi costante negli anni. La piovosità, come l'elevazione e la temperatura, sembra rappresentare un fattore di influenza della distribuzione di popolazione più forte rispetto alla distanza dai fiumi e dalla costa. I dati sembrano suggerire che mentre l'uomo può adattarsi a vivere in luoghi più distanti dal mare o dai fiumi, grazie alle nuove vie di comunicazione o alle tecnologie che permettono di estrarre e trasportare risorsa idrica, le condizioni ottimali di temperatura, piovosità e morfologia del terreno sembrano essere fattori imprescindibili per l'insediamento umano.

Le considerazioni fatte finora saranno poi migliorate da un'analisi di correlazione tra i parametri ambientali, per osservare se alcuni di questi sono dipendenti da altri e, quindi, possono essere eventualmente omessi o modificati (Capitolo 4).

Luminosità media per ogni anno per classi di Piovosità Media Annuale				
Classi (mm)	<350	350-450	450-550	>550
Anni				
1992	0.1	1	3.99	6.04
1993	0.14	1.43	4.43	6.3
1994	0.13	1.35	4.84	6.58
1995	0.13	1.57	5.32	7.53
1996	0.14	1.48	5.16	7.36
1997	0.13	1.43	5.06	7.47
1998	0.14	1.57	5.03	7.49
1999	0.12	1.31	4.6	6.83
2000	0.16	1.62	4.8	7.11
2001	0.15	1.53	5.05	7.37
2002	0.16	1.55	5.29	7.91
2003	0.22	1.6	5.08	7.96
2004	0.32	2.13	6.09	8.98
2005	0.3	2.07	5.88	8.6
2006	0.29	2.1	5.96	9.56
2007	0.32	2.24	5.85	9.79
2008	0.37	2.32	6	9.74
2009	0.38	2.42	6.49	11.15
2010	0.36	2.68	6.79	10.96
2011	0.5	3.27	6.92	11.26
2012	0.55	3.46	6.74	10.82
2013	0.53	3.26	7.26	12.39

Tabella 3.11: Risultati della luminosità media per ogni classe di distanza dai corsi d'acqua e per ogni anno di osservazione.

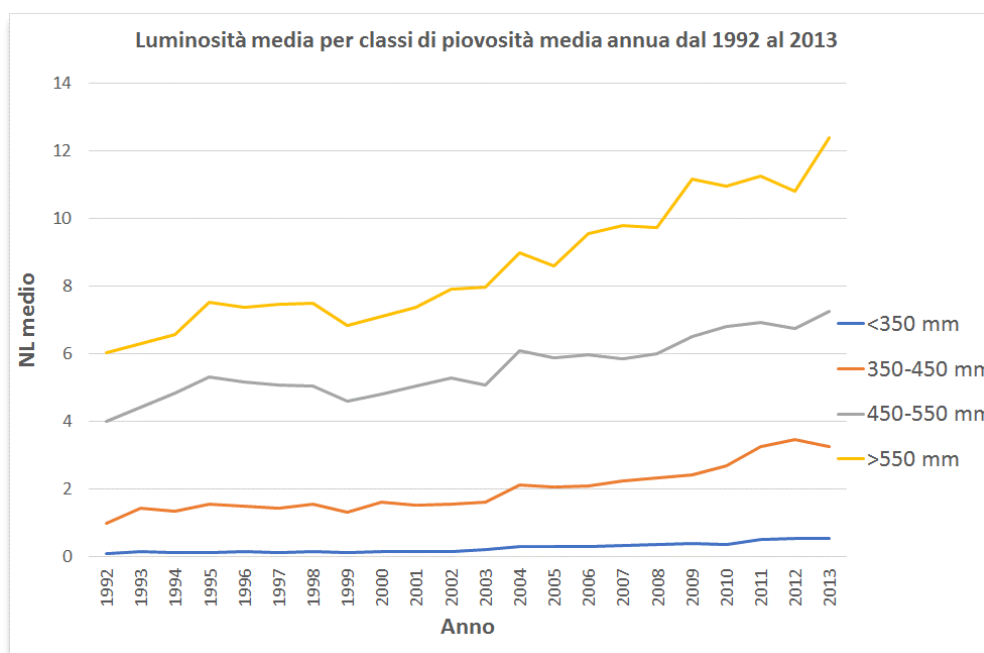


Figura 3.14: Andamento della luminosità media nel tempo per ogni classe di piovosità.

3.1.6 Copertura del terreno: suddivisione in classi

Per il parametro di copertura del suolo sono necessarie considerazioni differenti da quelle utilizzate finora. Innanzitutto, non si tratta di una classificazione su base quantitativa e quindi non si può realizzare un trend della luminosità a meno di non creare una classificazione. Per ottenere informazioni sull'area di studio, si può comunque calcolare in quali tipologie di copertura del suolo si concentrino prevalentemente le luci artificiali notturne. Il problema che può nascere attuando questo procedimento è intuibile osservando le tipologie di copertura del terreno (Tabella 2.8); essendo presente la voce "area urbanizzata", è chiaro che le luci artificiali raggiungeranno la massima intensità proprio in questa classe.

La seconda osservazione necessaria è che, a differenza degli altri parametri, non si può ritenere che la copertura del suolo sia rimasta invariata durante il periodo di osservazione; l'aspetto della superficie terrestre è spesso modificato, anche da un anno all'altro, soprattutto dall'uomo. Per analizzare correttamente tale fenomeno, si dovrebbe disporre di una serie temporale di copertura del suolo da confrontare anno per anno con la luminosità notturna. Data la complessità dell'inserimento di un parametro variabile nel tempo all'interno del modello, considerando anche che il DEM contiene indirettamente delle informazioni sul tipo di suolo, si è deciso di non utilizzare il Land Cover come variabile ambientale all'interno del modello. Di seguito sono comunque riportate alcune elaborazioni numeriche.

Per migliorare la stima del contributo della tipologia di copertura di suolo sulla luminosità, si potrebbe sostituire l'informazione di DEM con un coefficiente di *Roughness*, ovvero coefficienti che descrivono l'asperità del suolo e racchiudono quindi informazioni combinate sull'altitudine e sulla copertura del suolo (ad altitudini maggiori corrispondono suoli più ruvidi ed aspri, mentre colline e pianure hanno un coefficiente minore; suoli spogli avranno aspetto più aspro mentre la copertura vegetale individua terreni più dolci)⁵⁰.

In Figura 3.15, sono riportati i risultati della luminosità media nelle classi in cui sono state raggruppate le tipologie di copertura del suolo illustrate al Capitolo 2. Si è rappresentato il solo anno 2013 poiché è l'anno per cui è stato realizzato il raster di copertura del terreno reperito per lo studio.

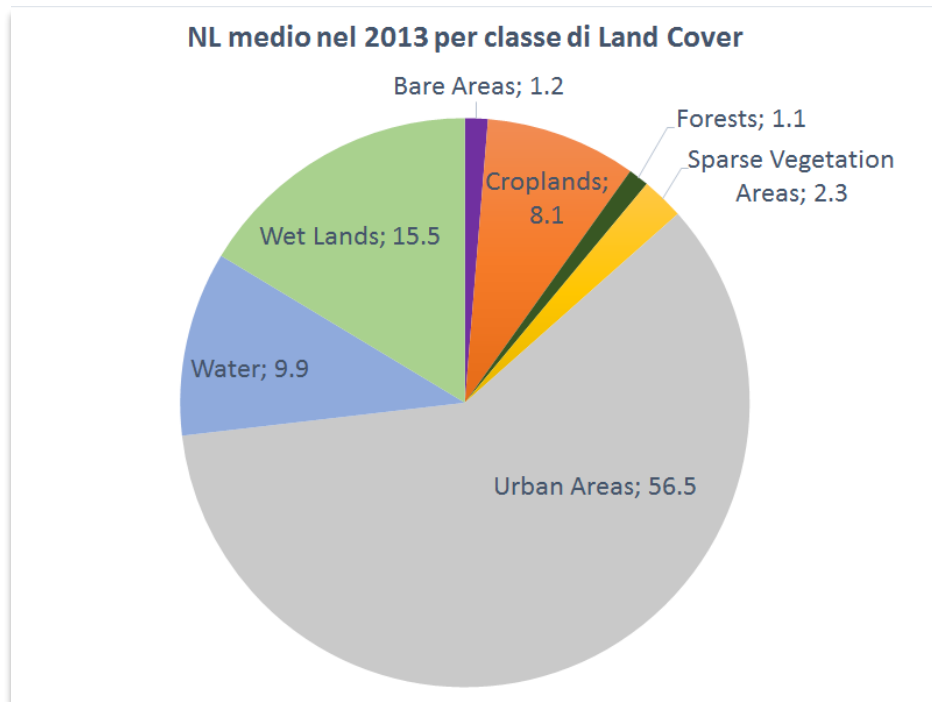


Figura 3.15: Valori di luminosità media per classi di copertura del suolo nel 2013.

Il grafico dà la conferma che le luci artificiali sono più intense nelle aree urbanizzate. Per ottenere informazioni su quale tipologia di copertura del suolo stata sia la più soggetta ad urbanizzazione, si dovrebbe verificare da un anno all'altro quali classi hanno cambiato aspetto passando da naturali ad antropizzate. In questo studio, non si è esplorata questa possibilità, concentrandosi su forzanti ambientali di tipo quantitativo.

3.2 Ricerca di eventuali correlazioni tra i fattori ambientali considerati

Prima di procedere con la costruzione del modello vero e proprio, è bene individuare eventuali correlazioni tra i fattori fin qui selezionati, in modo da poterne eventualmente ridurre il numero, ottenendo calcoli più snelli ed evitando ridondanza di dati. L'analisi è stata condotta in modo visivo, costruendo tutti i grafici che riportassero i valori di un parametro in funzione di un altro (Figura 3.16). Osservando come si dispongono i dati per tutte le coppie di variabili ambientali, si osserva che l'unica correlazione ben visibile sembra essere quella tra l'altitudine e la temperatura media annuale. La correlazione è inversa: al crescere dell'altitudine la temperatura decresce quasi linearmente.

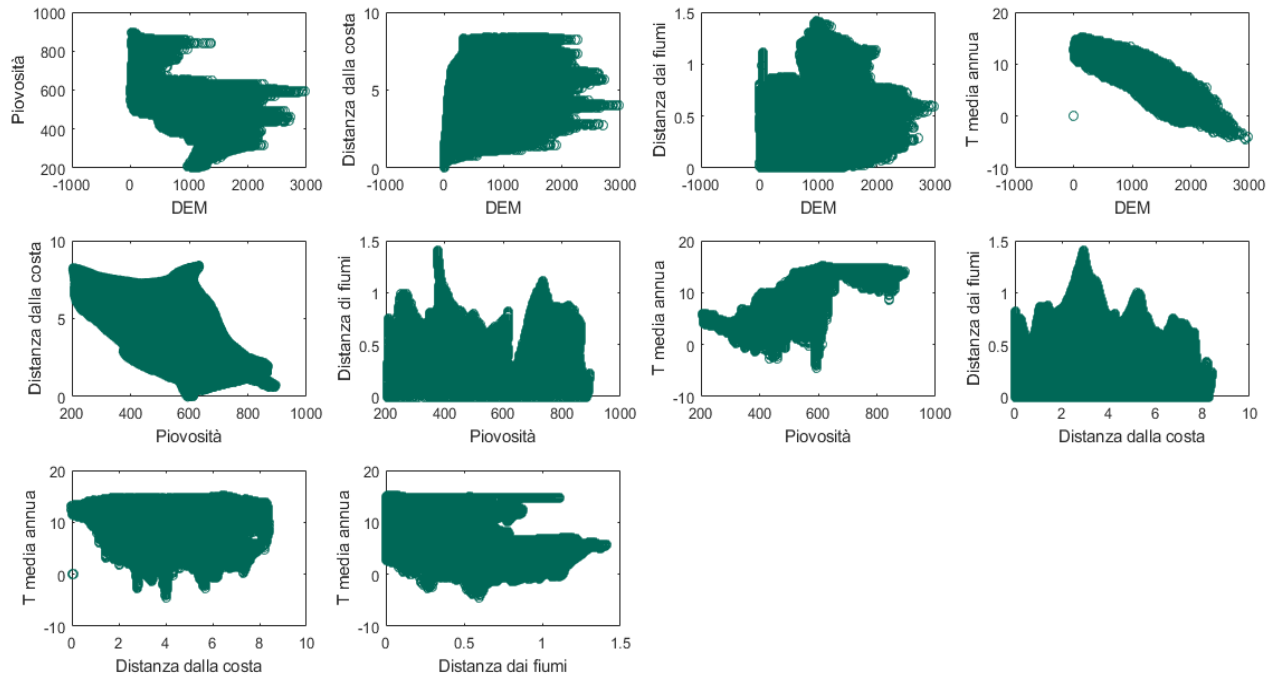


Figura 3.16: Risultati grafici per la ricerca di correlazione tra i parametri ambientali.

Tali risultati si sono poi verificati calcolando la matrice di correlazione di Pearson grazie alla funzione *coeffcorr* di MATLAB, come mostrato in Tabella 3.12. Il coefficiente di correlazione lineare è definito come il rapporto tra la covarianza delle variabili e il prodotto delle due varianze:

$$\rho_{xy} = \frac{\sigma_{xy}}{\sigma_x \sigma_y} \quad \text{con la proprietà: } -1 \leq \rho_{xy} \leq +1$$

Il coefficiente sarà quindi compreso tra i valori di -1, per cui le variabili sono perfettamente inversamente correlate, e +1, per cui le variabili sono perfettamente direttamente correlate, se il valore del coefficiente è prossimo a 0, la due variabili non risultano correlate.

La matrice che si ottiene è sempre simmetrica poiché riporta in ogni cella il coefficiente di Pearson ottenuto incrociando le variabili; i risultati sulla diagonale principale saranno quindi sempre pari a 1 poiché ottenuti dal calcolo della covarianza tra una variabile e se stessa, mentre nelle altre posizioni sono riportati i risultati del rapporto tra grandezze diverse.

	Elevazione	T media annua	Piovosità	Distanza dalla costa	Distanza dai fiumi
Elevazione	1.00	-0.93	-0.72	0.58	0.36
T media annua	-0.93	1.00	0.82	-0.38	-0.36
Piovosità	-0.72	0.82	1.00	-0.49	-0.13

Distanza dalla costa	0.58	-0.38	-0.49	1.00	0.04
Distanza dai fiumi	0.36	-0.36	-0.13	0.04	1.00

Tabella 3.12: Matrice di correlazione di Pearson per Altitudine e Temperatura media

Il coefficiente di correlazione tra altitudine e temperatura media annuale risulta molto prossimo a -1 e quindi è presente una forte correlazione inversa tra i due parametri. Per questo motivo, nelle due simulazioni del modello che verranno sviluppate al Capitolo 4, le due variabili non saranno utilizzate contemporaneamente nel set di forzanti ambientali. Si nota anche una, meno marcata, correlazione tra la piovosità media annuale con DEM e temperatura media annuale, essendo comunque meno evidente, la piovosità è stata considerata insieme alle altre nei set che si esamineranno, verrà poi valutata a sua significatività in base ai risultati di un test sulla complessità del modello.

4 Realizzazione del modello matematico di espansione urbana

4.1 Il modello teorico

Il modello ipotizzato per descrivere il processo di urbanizzazione nell'area di studio del presente elaborato di tesi è un modello stocastico spazialmente esplicito (i.e F. Finger et al. 2014⁵¹, J. B. Dunning, 1995⁵²) che permette di associare a ogni pixel la probabilità che esso passi da una condizione *non urbanizzata* a una *urbanizzata* definita in un istante di tempo. Il fenomeno dell'urbanizzazione è quindi rappresentato da due possibili stati del sistema:

- Pixel urbanizzati;
- Pixel non urbanizzati.

Tale distinzione è definita mediante l'identificazione di un opportuno valore di soglia di luminosità oltre il quale un pixel può essere considerato urbanizzato. L'area di studio viene quindi rappresentata dall'insieme X contenente tutti i pixel (di numerosità pari a N) il cui stato è definito come:

$$x_i(t) = \begin{cases} 1 & \text{per pixel urbanizzati (i.e. } NL_i > \text{valore di soglia)} \\ 0 & \text{per pixel non urbanizzati (i.e. } NL_i \leq \text{valore di soglia)} \end{cases}$$

Dove i rappresenta il generico pixel $\in X$ e t identifica un anno compreso nel periodo di studio, dal 1992 al 2013.

Pertanto, in un generico istante di tempo t , l'insieme X può essere suddiviso in due sottoinsiemi $U(t)$ e $W(t)$ che rappresentano rispettivamente i pixel urbanizzati e non urbanizzati (dall'inglese *wild*). L'insieme dei pixel urbanizzati, U, avrà numerosità:

$$N_u(t) = \sum_{i \in X} x_i(t) \quad (1)$$

Mentre l'insieme dei pixel non urbanizzati W avrà numerosità:

$$N_w(t) = N - N_u(t) \quad (2)$$

Lo stato di un generico pixel appartenente all'insieme X può modificarsi tra due istanti di tempo successivi, ovvero:

$$x_i(t) = k \rightarrow x_i(t + 1) = l \quad (3)$$

Questa transizione di stato da k a l , dove k ed l assumono valori pari a 0 o 1, si osserva per ogni istante di tempo t con una probabilità $p_{k,l}$. quindi, è possibile definire la seguente matrice delle probabilità di transizione:

$$\underline{P} = \begin{bmatrix} p_{0,0} & p_{0,1} \\ p_{1,0} & p_{1,1} \end{bmatrix} \quad (4)$$

Dove:

- $p_{0,0}$ rappresenta la probabilità che un pixel non urbanizzato all'istante t rimanga tale all'istante successivo (i.e. $x_i(t) = 0 \rightarrow x_i(t + 1) = 0$);
- $p_{0,1}$ rappresenta la probabilità che un pixel non urbanizzato all'istante t diventi urbanizzato in $t+1$ (i.e. $x_i(t) = 0 \rightarrow x_i(t + 1) = 1$);
- $p_{1,0}$ rappresenta la probabilità che un pixel urbanizzato all'istante t risulti non urbanizzato in $t+1$ (i.e. $x_i(t) = 1 \rightarrow x_i(t + 1) = 0$);
- $p_{1,1}$ rappresenta, infine, la probabilità che un pixel urbanizzato all'istante t risulti tale anche all'istante $t+1$ (i.e. $x_i(t) = 1 \rightarrow x_i(t + 1) = 1$).

Essendo \underline{P} una matrice stocastica, valgono le seguenti proprietà:

$$\begin{cases} p_{0,0} + p_{0,1} = 1 \\ p_{1,0} + p_{1,1} = 1 \end{cases} \quad (5)$$

In particolare, nel caso in esame, $p_{1,0}$ risulta nulla, ossia non si prende in considerazione il processo di abbandono delle aree urbane poiché è un processo molto più lento rispetto a quello di urbanizzazione, e dunque si può trascurare, di conseguenza risulta $p_{1,1} = 1$.

$p_{0,1}$ rappresenta invece la probabilità di urbanizzazione, oggetto di studio del seguente elaborato di tesi, e sarà, d'ora in avanti, indicata con $p_i(t)$, di conseguenza sarà: $p_{0,0} = 1 - p_i(t)$.

Essendo disponibili i dati di luminosità annuali, gli intervalli di tempo considerati risultano pari a un anno. Il processo di urbanizzazione in un definito intervallo di tempo prevede quindi la transizione dello stato di un certo numero di pixel da quello non urbanizzato a quello urbanizzato. Tali pixel, definiti *in espansione* sono rappresentati, nel passaggio da due anni consecutivi, dall'insieme E di numerosità:

$$N_E(t) = N_U(t + 1) - N_U(t) \quad (6)$$

Il processo di urbanizzazione viene quindi rappresentato da una probabilità di transizione per definire quali dei pixel non urbanizzati parteciperanno all'espansione in un determinato anno. Questa probabilità è costituita da due componenti, una che esprime l'effetto di un pixel urbanizzato su quelli non urbanizzati, l'altra dipendente dalle forzanti ambientali identificate ai capitoli precedenti, vediamole in dettaglio.

- *Kernel*⁵³: identifica l'effetto della distanza di un pixel non urbanizzato da altri urbanizzati, in altre parole rappresenta la probabilità che pixel non urbanizzato in un istante di tempo t , $x_{j \neq i}(t) = 0$, divengano urbanizzati all'istante di tempo $t+1$, $x_{j \neq i}(t + 1) = 1$ in funzione della distanza tra due pixel i e j , con $x_i(t) = 1$. È lecito infatti aspettarsi che un punto non urbanizzato che si trova vicino ad aree già urbanizzate risulti più facilmente urbanizzabile in futuro rispetto a un punto che si trova più lontano, anche in accordo con il primo principio della geografia di W. Tobler⁵⁴.
- *Forzanti ambientali*: rappresentano invece l'effetto di variabili idrologiche o geomorfiche sullo stato del sistema. Si considera quindi che l'altitudine, la temperatura e la piovosità medie annue, la distanza dal mare e dai corsi d'acqua principali possano influenzare la probabilità di un pixel di compiere la transizione, secondo gli andamenti evidenziati dall'analisi dei dati riportata al capitolo precedente di questo elaborato.

Da un punto di vista matematico, la probabilità di urbanizzazione $p_i(t)$ è definita, per ogni step temporale, come:

$$p_i(t) = C \cdot f_k(kernel_i) \cdot f_{DEM}(DEM_i) \cdot f_{Tm}(Tm_i) \cdot f_{SD}(SD_i) \cdot f_{RD}(RD_i) \cdot f_{Rain}(Rain_i) \quad (7)$$

In cui:

- C è la costante di normalizzazione che rende la probabilità adimensionale e minore o uguale a 1;
- $f_k(kernel_i)$ è l'effetto della distanza di un pixel non urbanizzato da altri pixel urbanizzati;
- $f_{DEM}(DEM_i)$ è l'effetto dell'altitudine sull'urbanizzazione;
- $f_{Tm}(Tm_i)$ è l'effetto della temperatura media sull'urbanizzazione;
- $f_{SD}(SD_i)$ è l'effetto della distanza dal mare sull'urbanizzazione;
- $f_{RD}(RD_i)$ è l'effetto della distanza dai corsi d'acqua principali sull'urbanizzazione;

- $f_{Rain}(Rain_i)$ è l'effetto della piovosità sull'urbanizzazione.

Si impone quindi che la probabilità che un pixel diventi urbanizzato in uno step temporale dipenda dalla distanza con i pixel già urbanizzati e dall'influenza delle forzanti ambientali descritte nei capitoli precedenti. Vediamo di seguito come valutare ciascun elemento dell'espressione 8.

La costante C si definisce normalizzando la probabilità rispetto alla somma delle probabilità dei pixel non urbanizzati all'istante t:

$$C = \frac{f_k(kernel_i) \cdot f_{DEM}(DEM_i) \cdot f_{Tm}(Tm_i) \cdot f_{SD}(SD_i) \cdot f_{RD}(RD_i) \cdot f_{Rain}(Rain_i)}{\sum_{i \in W(t)} [f_k(kernel_i) \cdot f_{DEM}(DEM_i) \cdot f_{Tm}(Tm_i) \cdot f_{SD}(SD_i) \cdot f_{RD}(RD_i) \cdot f_{Rain}(Rain_i)]} \quad (8)$$

Definiamo ora l'effetto della distanza dai pixel urbanizzati. Si vuole imporre una probabilità decrescente con l'aumentare della distanza, e si hanno quindi a disposizione due funzioni ad andamento decrescente, una legge di potenza:

$$K(dist_{ij}) = dist_{ij}^{-\gamma} + k_0 \quad (9)$$

O una legge esponenziale:

$$K(dist_{ij}) = \exp\left(-\frac{dist_{ij}}{\gamma}\right) + k_0 \quad (10)$$

In entrambe le espressioni compaiono due parametri:

- γ – *Parametro di Influenza (shape factor)*: parametro che definisce quanto l'influenza dei pixel urbanizzati decade con la distanza.
 $\gamma > 0$ sempre, all'aumentare del parametro, si impone all'urbanizzazione una struttura più compatta poiché la probabilità di urbanizzazione lontano dai pixel già urbanizzati decresce rapidamente;
- k_0 : aggiunge alla probabilità calcolata in funzione della distanza una quantità di base, in modo che ogni pixel sia caratterizzato da una probabilità di espansione minima, anche i più distanti da quelli urbanizzati. Il significato è di tenere conto della componente casuale che determina l'espansione anche di punti lontani da quelli già urbanizzati.

Nell'ambito dello studio sono state valutate entrambe, per le differenze nel risultato del modello si rimanda al Paragrafo 3. Nel calcolo della probabilità, il kernel è moltiplicato per lo stato del pixel i -esimo; si tratta di pesare lo stato di ogni pixel rispetto alla posizione relativamente agli altri pixel urbanizzati, bisogna quindi realizzare una convoluzione spaziale bidimensionale:

$$\sum_{j \in X(t)} x_j(t) \cdot K(\text{dist}_{i,j}) \quad \forall i \in W(t) \quad (11)$$

La matrice di partenza di input, quella contenente gli stati iniziali dei pixel, viene moltiplicata elemento per elemento per una matrice dei pesi, il *kernel*; tutti gli elementi così ottenuti vengono poi sommati per avere un unico valore, il risultato viene trascritto nella matrice finale in corrispondenza della posizione del pixel di partenza (Figura 4.1).

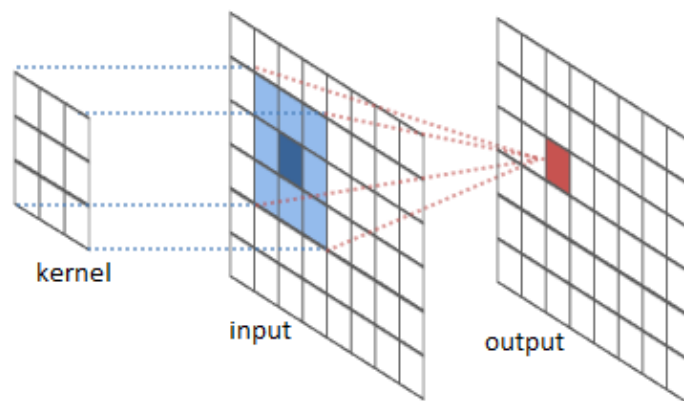


Figura 4.1: funzionamento della convoluzione spaziale⁵⁵.

Nel caso in esame, il kernel è costituito dalle funzioni della distanza dai pixel urbanizzati, dunque per ogni pixel si tratta di calcolare:

$$G = A * K \quad (12)$$

Dove A è la matrice input contenente gli stati di ciascun pixel e K è la matrice del kernel ottenuta calcolando le funzioni esponenziale o di potenza della matrice di distanza (espressioni 9 e 10) e imponendo che $K(1,1)=0$ poiché si considera nullo l'effetto di un pixel su se stesso.

Per il calcolo veloce della quantità di convoluzione, si può utilizzare il Teorema di convoluzione e passare alle trasformate di Fourier⁵⁶:

$$F(G) = F(A) \cdot F(K) \quad (13)$$

Essendo la trasformata di Fourier una funzione invertibile, l'espressione per calcolare la matrice G diventa:

$$G = \text{inv}F(F(A) \cdot F(K)) \quad (14)$$

Le funzioni che caratterizzano l'effetto delle forzanti ambientali sono invece state ipotizzate esponenziali, decrescenti o crescenti sulla base delle relazioni con la luminosità artificiale notturna individuate al Capitolo 3.

- Per l'elevazione è stata scelta una funzione esponenziale decrescente nella forma:

$$f_{DEM}(DEM_i) = \exp\left(-\frac{DEM_i}{k_D}\right) \quad (15)$$

Con $k_D > 0$: parametro che definisce l'influenza dell'altitudine sulla probabilità, all'aumentare del suo valore l'esponente diminuisce e la diminuzione della probabilità è meno marcata.

- Per la temperatura media è stata scelta una funzione esponenziale decrescente quanto più la temperatura media del pixel i -esimo si allontana dalla temperatura media massima (T_{max}) registrata su tutta l'area di studio:

$$f_{Tm}(Tm_i) = \exp\left(-\frac{T_{max} - Tm_i}{k_T}\right) \quad (16)$$

Con $k_T > 0$: parametro che definisce l'influenza della differenza tra la temperatura massima annuale registrata nell'area e la temperatura media annuale del pixel, all'aumentare del parametro l'entità dell'esponente diventa minore.

- Per la distanza dalla costa si è scelta una funzione esponenziale decrescente:

$$f_{SD}(SD_i) = \exp\left(-\frac{SD_i}{k_{SD}}\right) \quad (17)$$

Con $k_{SD} > 0$: parametro che definisce l'entità dell'effetto della distanza dal mare sulla probabilità. Se k_{SD} aumenta, l'effetto diventa meno deciso.

- Anche per la distanza dai corsi d'acqua principali risulta:

$$f_{RD}f(RD_i) = \exp\left(-\frac{RD_i}{k_{RD}}\right) \quad (18)$$

Con $k_{RD} > 0$: parametro che definisce l'entità dell'effetto della distanza dai corsi d'acqua principali sulla probabilità. Se k_{RD} aumenta, l'effetto diventa meno deciso.

- Per la piovosità l'esponenziale risulta invece crescente, data la propensione delle luci ad aumentare in prossimità dei punti con piovosità media annuale maggiore:

$$f_{Rain}(Rain_i) = \exp\left(\frac{Rain_i}{k_R}\right) \quad (19)$$

Con $k_R > 0$: parametro che definisce l'entità dell'influenza della piovosità sulla probabilità di urbanizzazione. All'aumentare di k_R l'esponente diminuisce e l'aumento di probabilità all'aumentare della piovosità è attenuato.

Si sono quindi individuati tutti i parametri del modello: γ , k_0 , k_D , k_T , k_{SD} , k_{RD} , k_R , il cui valore (sempre positivo) verrà stimato applicando il metodo della massima verosimiglianza. La funzione di verosimiglianza è definita come la probabilità congiunta di ottenere l'urbanizzazione allo step temporale successivo a quello di partenza. Ipotizzando che gli eventi siano indipendenti tra loro, poiché il passaggio all'urbanizzazione di un pixel sarà funzione solo delle forzanti ambientali e della distanza dai pixel urbanizzati, si potrà calcolare la probabilità congiunta dell'espansione dei pixel come la produttoria delle singole probabilità:

$$P\left(\bigcap_{i \in W} E_i\right) = \prod_{i \in W} P(E_i) \quad (20)$$

E quindi la funzione di verosimiglianza risulta la produttoria delle probabilità che le celle diventino urbanizzate o meno, il risultato risulta però essere un numero molto piccolo, che può essere approssimato a zero. Quindi, come usuale in questi casi, si valuterà invece la log-verosimiglianza passando quindi dalla produttoria delle probabilità alla sommatoria dei logaritmi delle probabilità che le celle divengano urbanizzate.

Per tutti i pixel che compiono la transizione in un singolo step temporale, quindi, sarà:

in $t \rightarrow t + 1$

$$LK_t = \prod_{i \in E} p_i(t) \cdot \prod_{i \in W-E} (1 - p_i(t)) \quad (21)$$

Passando quindi ai logaritmi si ottiene:

$$\log LK_t = \sum_{i \in E} \log(p_i(t)) + N_E \sum_{i \in W} \log(1 - p_i(t)) - \sum_{i \in E} \log(1 - p_i(t)) \quad (22)$$

Rispetto alla (21) nella (22) compare un ultimo termine aggiuntivo che viene sottratto agli altri. Il significato è che nel calcolo non si deve tenere conto dei pixel che hanno già compiuto la transizione. Vediamo un esempio per chiarire:

P1	P2	P3	P4	P5	P6	P7
0	1	0	1	0	0	0

➔

P1	P2	P3	P4	P5	P6	P7
0	1	1	1	1	0	1

Figura 4.2: Semplificazione del processo di espansione in un intervallo di tempo.

Consideriamo la transizione semplificata mostrata in figura 4.2, in essa si hanno:

- P2 e P4 pixel urbanizzati all'istante t che rimangono tali;
- P1 e P6 pixel non urbanizzati che rimangono tali;
- P3, P5 e P7 pixel non urbanizzati che cambiano stato.

I pixel P2 e P4 sono quindi caratterizzati da una probabilità di transizione, poiché risultano già nello stato 1, tutti gli altri pixel sono caratterizzati da una probabilità come quella descritta sopra.

La probabilità che i tre pixel divengano urbani è definita come:

$$\log(L_3) = p_3(1 - p_1)(1 - p_5)(1 - p_6)(1 - p_7)$$

$$\log(L_5) = p_5(1 - p_1)(1 - p_6)(1 - p_7) \cdot (1 - p_3)$$

$$\log(L_7) = p_7(1 - p_1)(1 - p_6)(1 - p_5) \cdot (1 - p_3)$$

E la probabilità totale sarebbe il prodotto tra queste tre. In realtà, appare chiaro che, essendo P3 già urbanizzato, non si deve reinserire il termine $(1 - p_3)$ nelle espressioni successive, e lo stesso

ragionamento vale per P5; il significato dell'ultimo termine della (22) è proprio quello di escludere le probabilità di un pixel che ha già cambiato stato in precedenza.

Osservando questo esempio, si nota che l'espressione (22) non risulta ancora completa poiché esprime la probabilità che i pixel si accendano seguendo un determinato percorso: il primo a cambiare stato è P3, seguito da P5 e infine P7. Si deve quindi inserire un termine che tenga conto del fatto che ogni percorso di espansione ha la stessa probabilità degli altri di verificarsi, la (22) risulta quindi, raccogliendo sotto la stessa sommatoria i termini dell'insieme E:

$$\log(LK_T(t)) = \sum_{i \in E} [\log(p_i(t)) - \log(1 - p_i(t))] + N_E(t) \cdot \sum_{i \in W} (1 - p_i(t)) + \log N_E(t)! \quad (23)$$

L'ultimo termine esprime il numero di permutazioni possibili all'interno dell'insieme E in un certo istante di tempo, che può essere definito mediante l'approssimazione di Stirling⁵⁷ che permette il calcolo approssimato di fattoriali per numeri piuttosto elevati:

$$\begin{aligned} \log(LK_T(t)) = & \sum_{i \in E} [\log(p_i(t)) - \log(1 - p_i(t))] + N_E(t) \cdot \sum_{i \in W} (1 - p_i(t)) \\ & + (N_E(t) \log N_E(t) - N_E(t)) \end{aligned} \quad (24)$$

La 24 esprime la formula della probabilità condizionata dell'espansione totale all'interno di un istante di tempo, ovvero la log-verosimiglianza totale per ogni step temporale.

Considerando tutte le transizioni da t_0 a $t_{end}-1$, avremo la log-verosimiglianza totale:

$$\log(LK) = \sum_{t=t_0}^{t=t_{end}-1} \log(LK_T(t)) \quad (25)$$

Ovvero la funzione da massimizzare è la somma di tutte le log-verosimiglianze che descrivono la probabilità in tutti gli step temporali, poiché le transizioni sono tra loro indipendenti.

Nel paragrafo seguente, si tradurranno questi passaggi in termini numerici, osservando gli algoritmi necessari alla definizione della probabilità qui descritta, della calibrazione e della simulazione in avanti del modello.

4.2 Lo script per la calibrazione numerica e la simulazione di anni futuri

Di seguito sono esposte le caratteristiche del codice numerico necessario alla stima dei parametri del modello, utili anche per simulare possibili scenari di urbanizzazione futura. Come premessa, sono esposte le tipologie di modelli prese in considerazione nell'ambito di questo studio.

Innanzitutto, data la stretta dipendenza tra il DEM e la temperatura media annuale (si veda l'analisi di correlazione al Capitolo 3), si è preferito costruire due set di modelli in cui tali variabili sono prese in considerazione separatamente, evitando così dati ridondanti (Tabella 4.1). In totale, ognuno dei due set di modelli prevede al massimo sei parametri da stimare.

Set 1		Set 2	
Variabili	Parametri	Variabili	Parametri
Distanza da pixel urbanizzati	γ, k_0	Distanza da pixel urbanizzati	γ, k_0
DEM	k_D	Temperatura media annuale	k_T
Piovosità	k_R	Piovosità	k_R
Distanza dalla costa	k_{SD}	Distanza dalla costa	k_{SD}
Distanza dai corsi d'acqua	k_{RD}	Distanza dai corsi d'acqua	k_{RD}

Tabella 4.1: I diversi set di variabili ambientali scelti per il modello.

Inoltre, entrambi i modelli sono stati valutati sia con un kernel esponenziale, sia con un kernel definito da legge di potenza per poter osservare le diverse performance. In totale, quindi, sono state lanciate quattro simulazioni.

All'interno di ognuna di queste quattro prove, si sono in poi valutate tutte le combinazioni di parametri possibili, considerando il kernel sempre attivo, poiché la distanza da altri pixel urbanizzati è la condizione minima per il calcolo della probabilità di urbanizzazione. In totale si sono analizzate 16 combinazioni di modelli per entrambi i set di studio ed entrambi i kernel. Le combinazioni possibili sono riportate in Tabella 4.2 dove con \checkmark sono indicati i parametri attivi mentre - rappresenta i parametri non considerati nella combinazione.

In questo modo sono stati valutati tutti i casi possibili, poiché non è scontato che il modello migliore sia necessariamente quello in cui tutte le variabili sono presenti. I risultati per ogni modello e per ogni combinazione di parametri sono poi riportati e commentati al Paragrafo 3.

Modello	Set 1 (kernel con legge esponenziale e di potenza)						Set 2 (kernel con legge esponenziale e di potenza)						Note
	γ	k_0	k_D	k_R	k_{SD}	k_{RD}	γ	k_0	k_T	k_R	k_{SD}	k_{RD}	
1	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	Tutte le componenti sono attive
2	✓	✓	-	-	-	-	✓	✓	-	-	-	-	Solo la componente kernel è attiva
3	✓	✓	✓	-	-	-	✓	✓	✓	-	-	-	Sono attive la componente kernel più una componente ambientale aggiuntiva
4	✓	✓	-	✓	-	-	✓	✓	-	✓	-	-	
5	✓	✓	-	-	✓	-	✓	✓	-	-	✓	-	
6	✓	✓	-	-	-	✓	✓	✓	-	-	-	✓	
7	✓	✓	✓	✓	-	-	✓	✓	✓	✓	-	-	Sono attive la componente kernel più due componenti ambientali aggiuntive
8	✓	✓	✓	-	✓	-	✓	✓	✓	-	✓	-	
9	✓	✓	✓	-	-	✓	✓	✓	✓	-	-	✓	
10	✓	✓	-	✓	✓	-	✓	✓	-	✓	✓	-	
11	✓	✓	-	✓	-	✓	✓	✓	-	✓	-	✓	
12	✓	✓	-	-	✓	✓	✓	✓	-	-	✓	✓	
13	✓	✓	✓	✓	✓	-	✓	✓	✓	✓	✓	-	Sono attive la componente kernel più tre componenti ambientali aggiuntive
14	✓	✓	✓	-	✓	✓	✓	✓	✓	-	✓	✓	
15	✓	✓	-	✓	✓	✓	✓	✓	-	✓	✓	✓	
16	✓	✓	✓	✓	-	✓	✓	✓	✓	✓	-	✓	

Tabella 4.2: Tutte le possibili combinazioni di parametri prese in esame per ciascuno dei set e kernel simulati.

Vediamo adesso alcune considerazioni preliminari per la costruzione dello script e le funzioni utilizzate per ottenere la calibrazione dei parametri e, successivamente, la simulazione futura.

4.2.1 La creazione di una maschera territoriale per il calcolo della probabilità di urbanizzazione

Osservando l'area di studio (Figura 1.10), si notano chiaramente aree in cui sono presenti laghi (molto evidenti soprattutto a sud-est) o una porzione di mare. È chiaro che in queste zone la probabilità di aumento delle luci, e quindi di urbanizzazione, risulterà pari a 0, essendo aree in cui non è possibile l'edificazione. Dunque, per prima cosa si è deciso di creare una maschera territoriale per rendere nullo il calcolo della probabilità di urbanizzazione in corrispondenza di laghi e mare.

Per quanto riguarda il mare, esso era già escluso dal raster di luminosità poiché i dati si fermano alla linea di costa.

I laghi, invece, non erano esclusi dal territorio del raster, ed è stato necessario reperire un file in cui fossero indicati per poi sovrapporli alla luminosità e creare la maschera definitiva. Online si possono trovare diversi shape vettoriali che rappresentano i laghi della Cina⁵⁸, essi però non rappresentano gli specchi d'acqua dell'area di studio con sufficiente precisione ma riportano solo quelli più estesi, probabilmente perché realizzati ricalcando immagini satellitari a una scala maggiore (Figura 4.3).

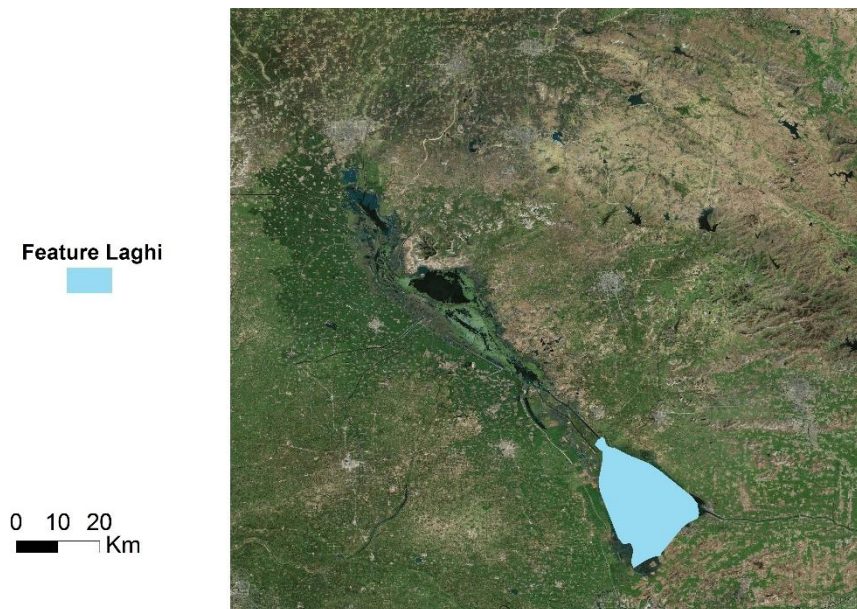


Figura 4.3: Dettaglio della feature dei laghi reperita online a confronto con la Basemap Imagery di ArcGIS.
Fonte: <https://www.naturalearthdata.com/>.

Per ottenere quindi una rappresentazione dei laghi più fedele alla realtà, è stata creata una feature poligonale partendo dalla basemap *Imagery* disponibile in ArcGIS. Si sono, cioè, creati poligoni vettoriali ricalcando le immagini satellitari degli specchi d'acqua, con il vantaggio di poter zoomare sull'area di studio in modo da rappresentare con una buona precisione tutti i laghi. Per poter confrontare la posizione dei laghi con il raster delle luci artificiali notturne, è stato necessario convertire la feature in un dataset raster, grazie al *conversion tool* di ArcGIS, il cui risultato è mostrato in Figura 4.4.

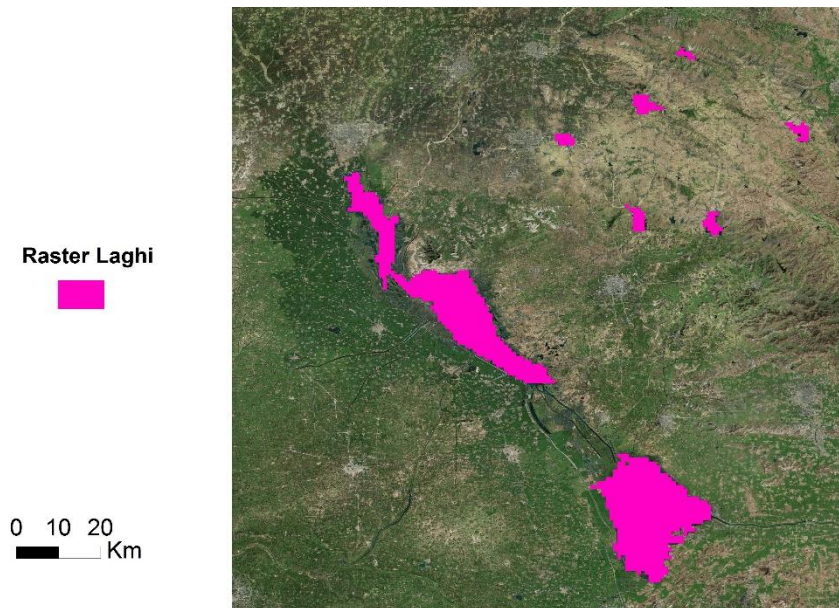


Figura 4.4: Confronto tra il raster laghi creato e la Basemap di ArcGIS in un dettaglio dell'area di studio.

Confrontando le due immagini, si nota come il raster creato ad hoc per questo studio riporti un numero di laghi molto maggiore e con forme più precise, rispetto allo shape utilizzato inizialmente. La maschera vera e propria da applicare ai raster di luminosità è stata ottenuta in ambiente MATLAB creando una matrice che riporta 0 in corrispondenza di laghi e mare e 1 in tutti i punti in cui invece l'urbanizzazione è possibile (Figura 4.5). Per ottenere questa matrice è stato sufficiente sostituire a una matrice di luminosità qualsiasi il valore 0 dove fossero presenti i NaN (in modo da individuare la zona di mare) e 1 in tutti gli altri punti; caricando poi il raster laghi si è inserito il valore 0 dove fosse presente uno specchio d'acqua e 1 in tutti gli altri punti. Le due matrici così ottenute sono state poi sommate ottenendo una matrice con valori:

- 1 in corrispondenza di mare e laghi;
- 2 in tutti gli altri punti.

Per ottenere la maschera finale, è stato imposto un valore pari a 0 per i punti a valore iniziale 1 e 1 per i punti a valore inizialmente 2. Per tutti i punti a valore 0 della matrice maschera, quindi, la probabilità di urbanizzazione si imporrà nulla.

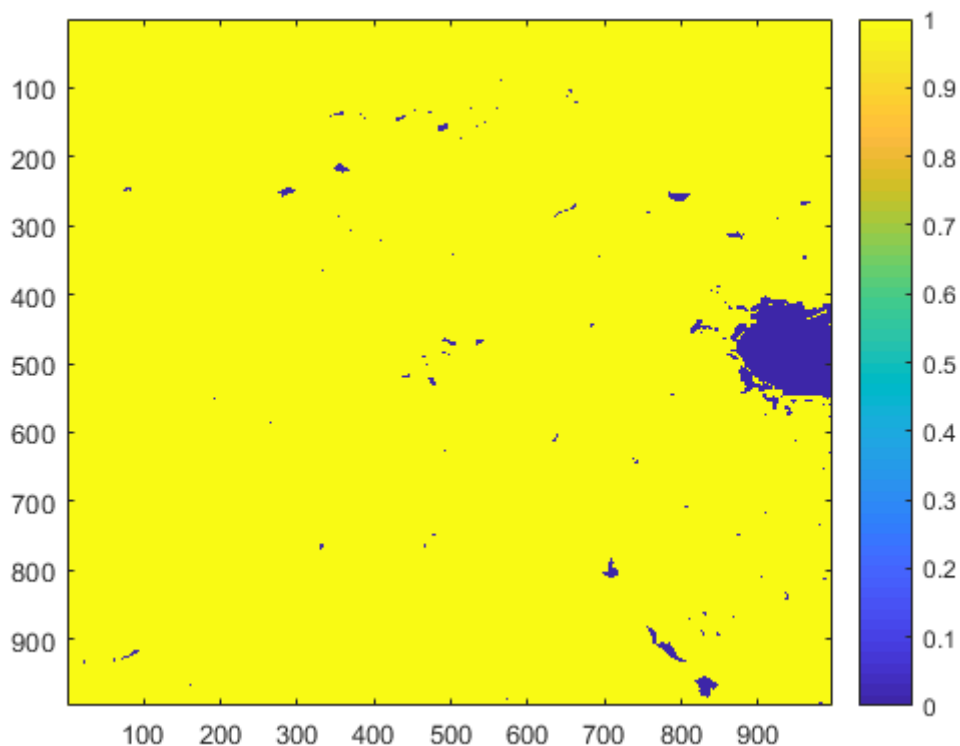


Figura 4.5: Matrice maschera che esclude le aree con presenza di laghi o mare dal calcolo della probabilità di urbanizzazione.

Si potrebbe pensare di inserire in questa maschera anche le infrastrutture viarie maggiori o i corsi d'acqua, essendo anche questi punti destinati a rimanere non urbanizzati; si può considerare infatti che le infrastrutture principali restino tali anche per gli anni successivi, e dunque l'intensità luminosa resterà costante negli anni.

Tale possibilità è stata esplorata all'interno di questo studio e a questo scopo era stata creata una maschera che contenesse anche le autostrade presenti nell'area di studio (essendo quella per il corso dei fiumi già disponibile dal raster della flow accumulation o del reticolo idrografico). Questi dati non sono poi stati inseriti nella versione definitiva poiché la risoluzione dei file raster (1 km) supera di molto la dimensione reale degli elementi che rappresentano (al più un centinaio di metri) e dunque avrebbe coperto anche aree limitrofe ai fiumi o alle strade, escludendo dal calcolo diversi punti utili. Questo effetto è rappresentato in Figura 4.6 in cui si confronta il raster di posizione dei corsi d'acqua con le immagini satellitari della basemap *Imagery* di ArcGIS; dalla figura si nota chiaramente come i corsi d'acqua riportati dal raster comprendano anche porzioni di territorio contigue.

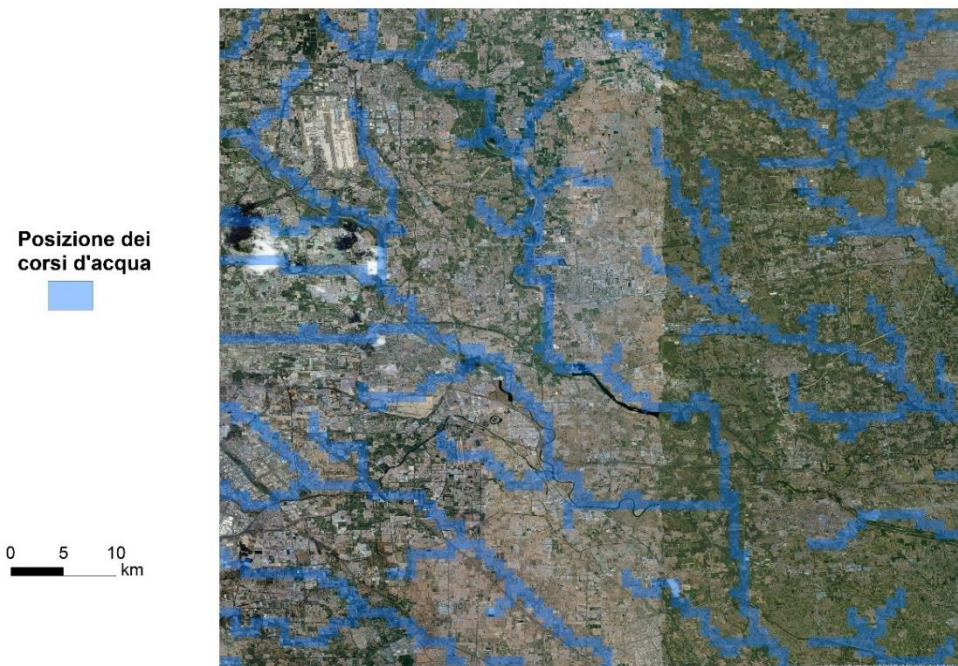


Figura 4.6: Confronto tra il raster di posizione dei corsi d'acqua fornito da HydroSHEDS e a basemap Imagery di ArcGIS.

4.2.2 La ricerca di un valore di soglia per la rappresentazione degli stati “urbanizzato” e “non urbanizzato”

Seguendo la teoria del modello utilizzato (esposta al Paragrafo 1 di questo capitolo) i pixel che appartengono alle aree urbanizzabili della matrice maschera devono essere suddivisi in due insiemi: urbanizzati o non urbanizzati. È quindi necessario definire un valore di soglia che sancisca la divisione tra uno stato e l'altro per identificare, all'interno di ogni anno, quanti pixel sono già urbanizzati (accesi) e quanti invece non lo sono (spenti) e creare matrici di luminosità fatte di soli due valori, 1 o 0.

Confrontando le immagini satellitari disponibili nella basemap *Imagery* di ArcGIS con le immagini di luminosità artificiale (Figura 4.6) si sono individuati i valori che caratterizzano i pixel ai confini dei centri abitati, che possono essere considerare i limiti delle aree urbanizzate.

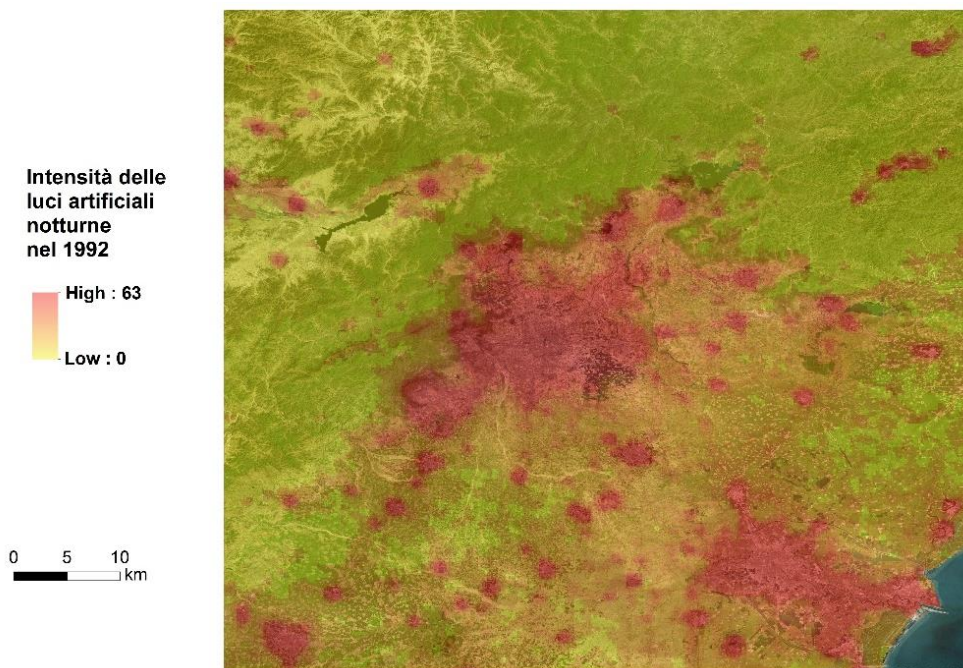


Figura 4.7: Confronto tra la basemap Imagery di ArcGIS e la luminosità artificiale nell'anno 1992 per la città di Pechino.

Analizzando l'area di Pechino, per le periferie e i piccoli centri abitati si riscontrano valori di intensità luminosa compresi tra 28 e 40 circa; per non sottovalutare il fenomeno, si è selezionato un valore di soglia pari a 30. Tale scelta è ovviamente soggettiva e sarebbe opportuno trovare delle linee generali per definire l'urbanizzazione sulla base del valore di luminosità. Dunque, un pixel risulterà urbanizzato se l'intensità luminosa registrata è strettamente maggiore di 30, in caso contrario sarà considerato non urbanizzato e se ne calcolerà la probabilità di urbanizzazione, $p_i(t)$.

4.2.3 L'algoritmo

Di seguito sono analizzati i passaggi salienti dello script che permette di calcolare la probabilità di urbanizzazione definita al Paragrafo 1; saranno analizzate le funzioni di supporto, le eventuali approssimazioni numeriche adottate e gli algoritmi di calcolo utilizzati.

4.2.3.1 Definizione della matrice degli stati

Innanzitutto, si sono compilate le matrici degli stati di ogni pixel per ogni anno osservato, cioè dal 1992 al 2013. Come spiegato in precedenza, lo stato del pixel viene definito semplicemente in base al rapporto con il valore di soglia stabilito (Paragrafo 2), dunque la matrice degli stati sarà composta da:

- 0 dove il valore di luminosità del pixel non supera quello di soglia, ossia:

$$x_i(t) = 0 \text{ per } NL \leq \text{Soglia}$$

- 1 per tutti i pixel con luminosità maggiore del valore di soglia, ossia:

$$x_i(t) = 1 \text{ per } NL > \text{Soglia}$$

In realtà, è stata fatta una considerazione aggiuntiva poiché non è detto che se un pixel supera il valore di luminosità di soglia in un certo anno esso risulti urbanizzato definitivamente. Si è quindi deciso di considerare urbanizzati solo i pixel che superassero il valore di soglia per almeno due anni consecutivi, in modo da non considerare quei punti che possono essersi accesi occasionalmente. I punti che rispettano questa condizione, sono poi considerati urbanizzati per tutto il periodo successivo, il che porta a sovrastimare il numero di punti urbanizzati, ma si ottiene un andamento più regolare del fenomeno di urbanizzazione nel tempo (Figura 4.9). L'anno 2013, non essendo disponibili i dati per l'anno successivo, è ottenuto dalla semplice divisione in base al valore di soglia. A meno dell'ultimo anno, quindi, le matrici degli stati non coincidono con le matrici di luminosità artificiale divise sulla base del valore di soglia. L'aspetto delle matrici degli stati definite per tutti gli anni di osservazione è riportato in Figura 4.8.

Matrici degli stati dei pixel dal 1992 al 2013

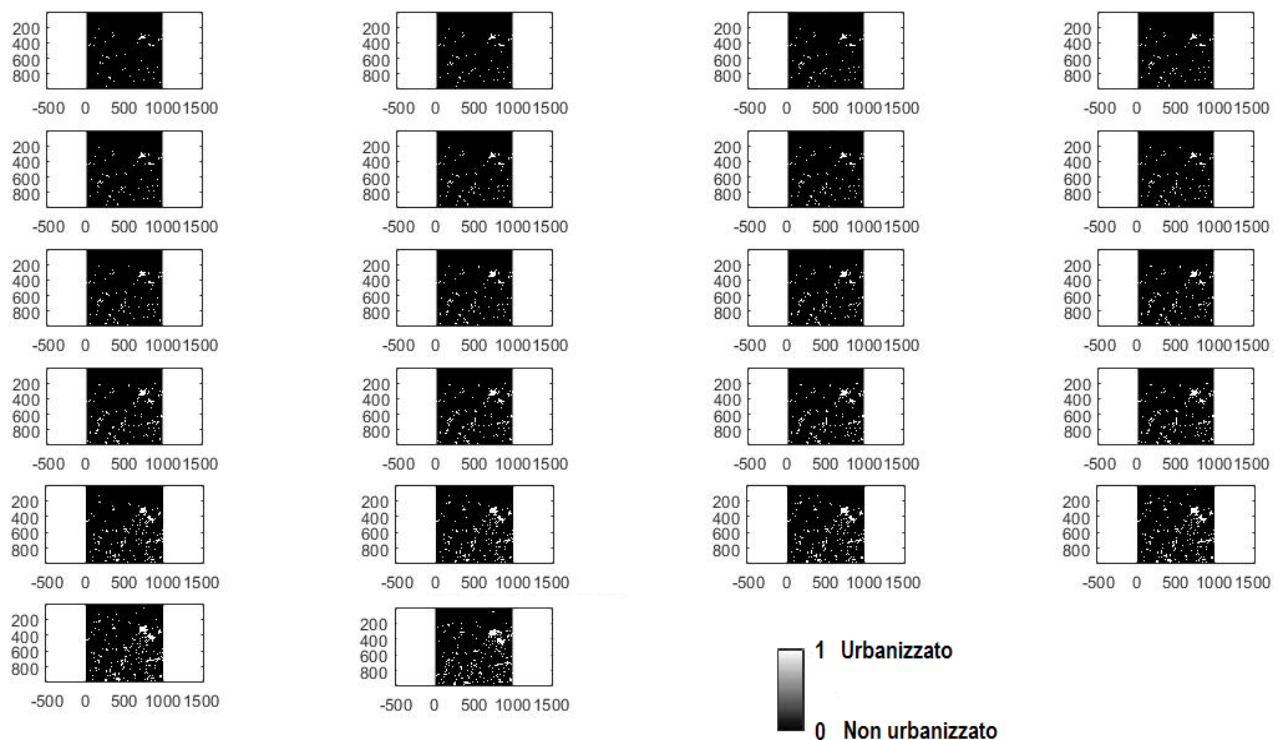


Figura 4.8: Matrici degli stati dall'anno 1992 al 2013.

Nel grafico illustrato in Figura 4.9 è invece riportato il numero di pixel urbanizzati per ogni anno sia considerando il semplice numero di pixel che supera il valore di soglia per l'anno di osservazione (linea blu), sia considerando solo i pixel che superano il valore per almeno due anni consecutivi (linea rossa). La differenza è chiara; l'urbanizzazione ottenuta inserendo la condizione di superamento per due anni consecutivi mostra un andamento più regolare ed evidenzia la possibilità non trascurabile per cui un pixel si urbanizza per un singolo anno non risulta tale negli anni successivi.

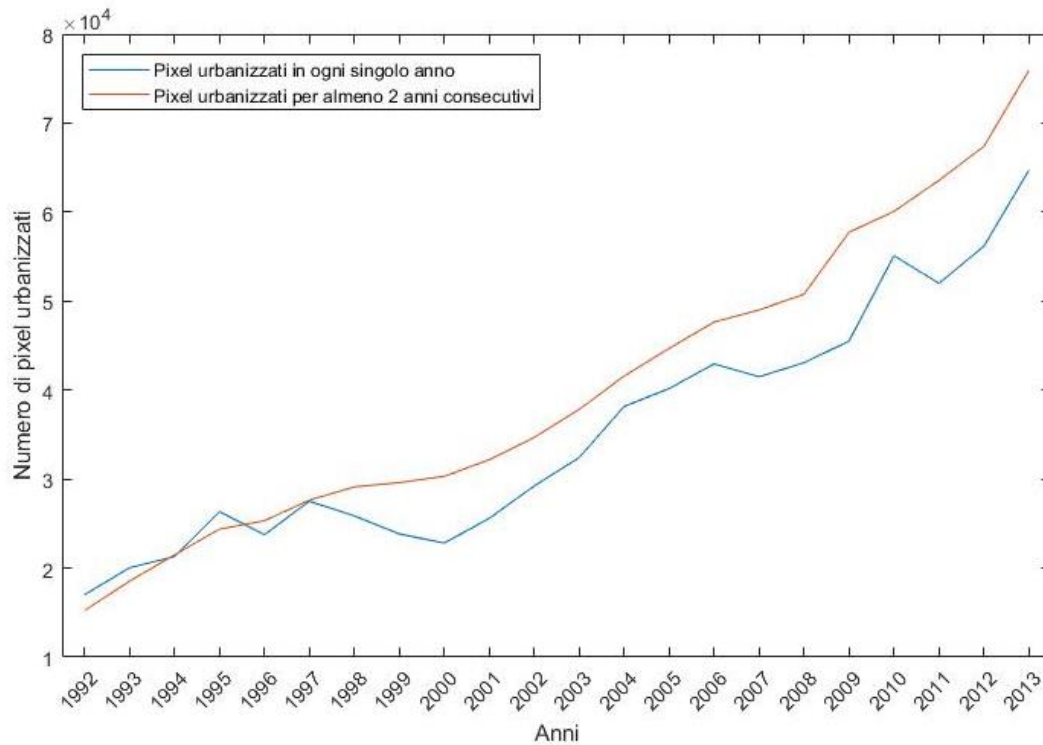


Figura 4.9: Confronto tra il numero di pixel urbanizzato applicando semplicemente il valore di soglia a tutti gli anni (in blu) e il numero di pixel urbanizzati ottenuto considerando solo i punti che hanno superato il valore di soglia per almeno due anni consecutivi (in rosso).

Le matrici così suddivise, quindi, saranno caratterizzate punto per punto da una probabilità di transizione che risulterà:

- $p_{1,1}(t) = 1$ per tutti i punti urbanizzati all'istante t che rimangono tali negli istanti successivi;
- $p_{1,0}(t) = 0$ per tutti i punti urbanizzati all'istante t e non urbanizzati all'istante $t+1$;
- $p_{0,1}(t) = p_i(t)$ per tutti i punti non urbanizzati all'istante t che divengono urbanizzati all'istante $t+1$;
- $p_{0,0}(t) = 1 - p_i(t)$ per tutti i punti non urbanizzati all'istante t che rimangono tali all'istante $t+1$.

Vediamo adesso altri passaggi necessari alla definizione di questa probabilità.

4.2.3.2 Calcolo della convoluzione spaziale

Al Paragrafo 1 è stato definito il processo di convoluzione spaziale. Per definire la matrice risultante dalla convoluzione del kernel sulla matrice degli stati, è necessario osservare l'aspetto della matrice di distanza (Figura 4.10). Il modello richiede il calcolo della distanza di ogni singolo pixel i rispetto a tutti gli altri pixel appartenenti all'area di studio. Tale operazione dovrebbe essere ripetuta poi per ogni pixel (i.e. 996×996 volte), dilatando notevolmente i tempi di calcolo computazionale. Per velocizzare la procedura si definisce una matrice di distanza in cui la distanza tra due pixel è definita in funzione delle differenze tra gli indici di posizione (riga e colonna) dei pixel stessi.

Se si si calcola la distanza tra un pixel e gli altri che compongono la matrice degli stati di partenza, la matrice di distanza risulterà simmetrica e dall'andamento periodico; la distanza tra i pixel aumenta verso il centro della matrice fino a un punto di massimo, dopodiché i valori tornano a decrescere in modo simmetrico rispetto agli assi verticale e orizzontale. Un comportamento analogo si verifica per il kernel, poiché ottenuto da una legge esponenziale o di potenza applicata alla matrice di distanza. L'andamento del kernel risulterebbe quindi toroidale, ovvero i punti più distanti dal pixel di partenza avrebbero lo stesso valore di quelli più vicini trovandosi nella posizione opposta, ovvero simmetrica.

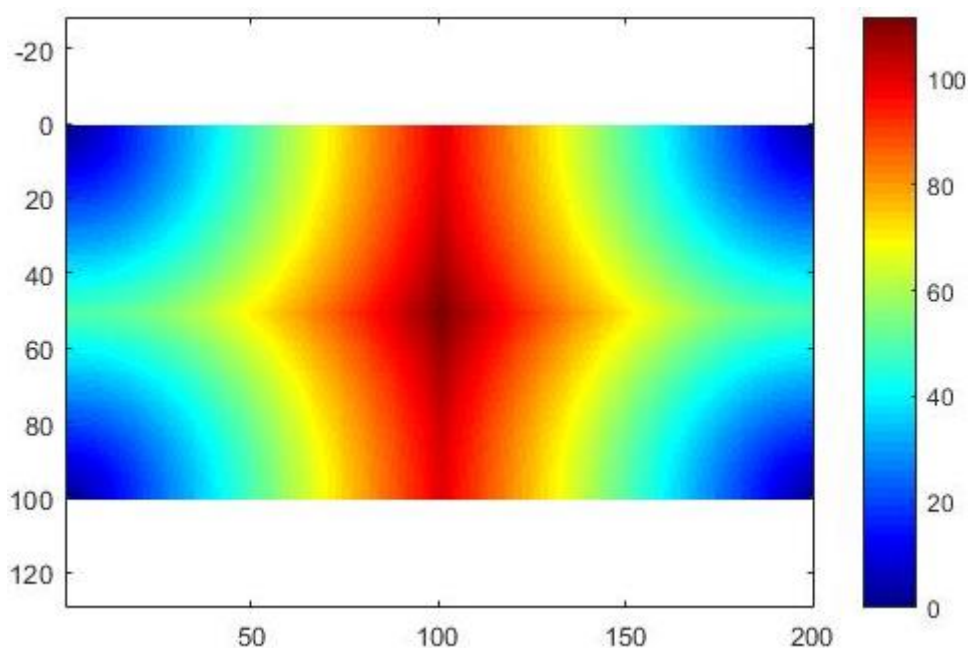


Figura 4.10: Matrice di distanza con dominio periodico.

In questo modo, però, la distanza ottenuta risulterebbe fittizia e ai pixel più lontani da quelli urbanizzati verrebbe assegnata una probabilità più alta di quella reale. Per ovviare a questo inconveniente, si può procedere raddoppiando il dominio della distanza e del kernel, ottenendo quindi una matrice doppia rispetto a quella di partenza; in questo modo la matrice di distanza corretta per i

punti del dominio di studio sarà riportata nel quadrante superiore sinistro della nuova matrice di distanza (Figura 4.11). All'interno del dominio di studio non si avrà più periodicità e quindi i punti a maggiore distanza dal pixel in esame, saranno effettivamente quelli con valori maggiori di distanza.

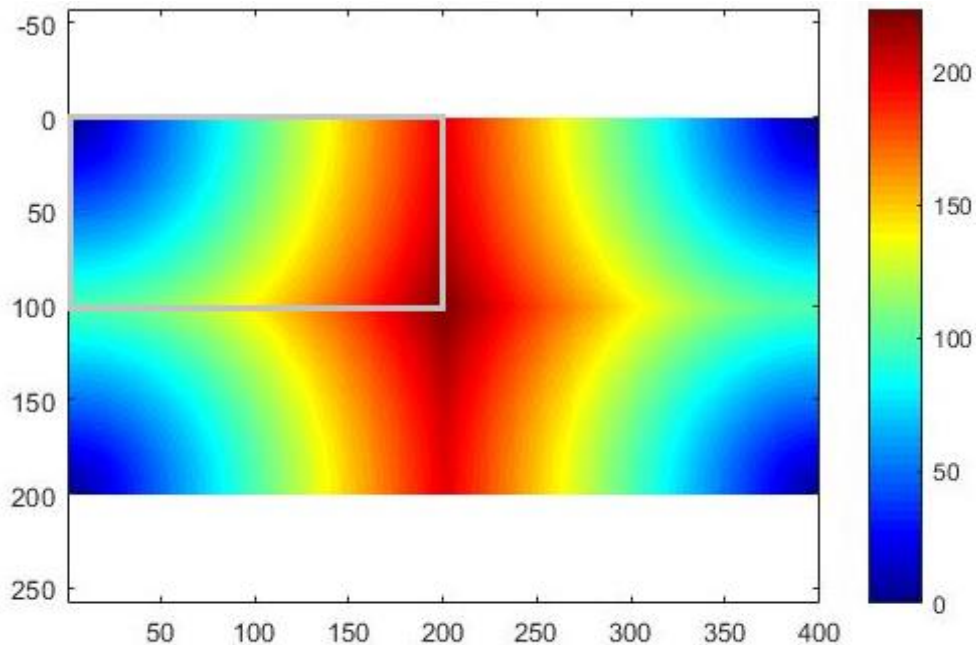


Figura 4.11: Aspetto della matrice di distanza con dominio raddoppiato, il rettangolo in grigio rappresenta il dominio di studio.

Un esempio della matrice del kernel ottenuta applicando alla matrice di distanza una legge di potenza, è riportato in Figura 4.12. Il pixel in esame (in alto a sinistra) ha un valore di kernel nullo poiché non ha effetto su sé stesso; valori della matrice decrescono poi rapidamente allontanandosi da questo punto poiché si vuole attribuire un peso minore ai pixel più lontani da quelli già urbanizzati.

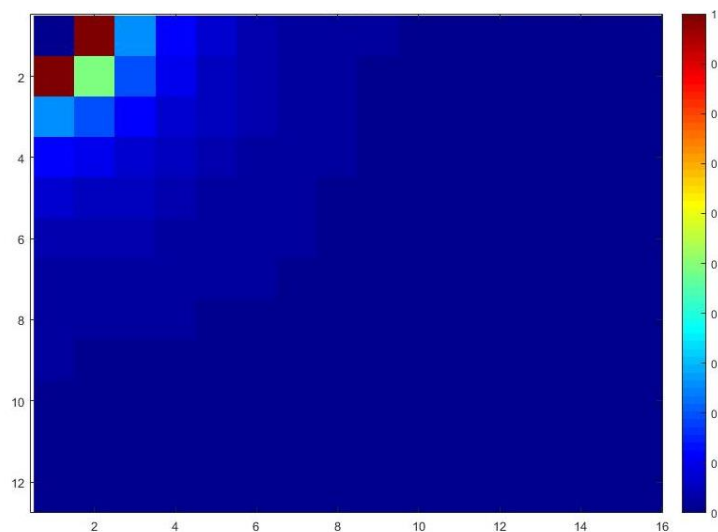


Figura 4.12: Immagine zoomata del kernel doppio ottenuto con legge di potenza della distanza.

Questi valori saranno poi sottoposti a un prodotto di convoluzione con la matrice degli stati. Per calcolare la convoluzione spaziale, si può applicare la trasformata di Fourier con il vantaggio di scambiare il prodotto di convoluzione con quello algebrico⁵⁹ (Paragrafo 1). Nel paragrafo successivo vediamo quindi come è possibile ottenere questa operazione in ambiente MATLAB.

4.2.3.3 La Fast Fourier Transform

La trasformata di Fourier permette di analizzare una funzione non periodica suddividendola, nel suo spettro, in somma di funzioni periodiche semplici. Sia $f(x, y)$ una funzione definita in \mathbb{R}^2 , assolutamente integrabile da $-\infty$ a $+\infty$, sia C definito a pezzi su ogni intervallo, si definisce la Trasformata di Fourier la funzione:

$$\hat{f}(v_x, v_y) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y) e^{-i2\pi(v_x x + v_y y)} dx dy \quad \mathbb{R}^2 \rightarrow C \quad (26)$$

La funzione è invertibile:

$$F^{-1}[\hat{f}](x, y) = f(x, y) \quad (27)$$

In un caso discreto, come quello di studio, la Trasformata di Fourier viene definita, per un insieme di numeri $[x, y]$, dalla sommatoria⁶⁰:

$$F(v_x, v_y) = \sum_{-\infty}^{+\infty} \sum_{-\infty}^{+\infty} f[x, y] e^{-i(v_x x + v_y y)} \quad (28)$$

La risoluzione diretta di questa espressione può tuttavia risultare molto onerosa, richiedendo una complessità di calcolo N^2 per ottenere N output.

Per questo motivo, si è utilizzato un algoritmo di calcolo della Fast Fourier Transform (FFT)⁶¹ in ambiente MATLAB (funzione `fft`⁶²) per calcolare la trasformata di Fourier Discreta e la sua inversa (`ifft2`) scomponendole in fattori, passando a una complessità di calcolo molto inferiore, pari a $N \log N$. Il risultato di questo calcolo è rappresentato in Figura 4.13 in cui è riportata la convoluzione calcolata tramite la FFT tra la matrice del kernel e quella degli stati per tre ipotetici pixel urbanizzati.

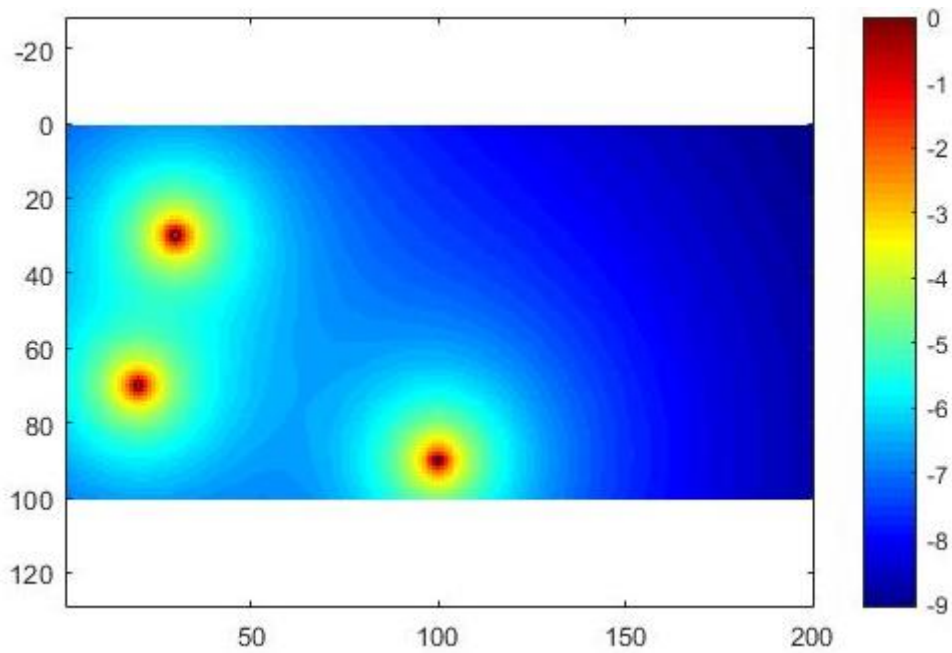


Figura 4.13: Rappresentazione della convoluzione spaziale calcolata tramite la funzione `fft` di MATLAB per tre punti.

Il risultato è analogo a quello che si otterrebbe applicando un metodo di calcolo esplicito, il cui risultato è riportato in Figura 4.14, le differenze tra i due metodi corrispondono a valori dell'ordine 10^{-16} . Dunque, si può accettare l'utilizzo della FFT per ottenere il calcolo della convoluzione ed ottenere la componente di distanza dai pixel urbanizzati nel calcolo della probabilità di urbanizzazione.

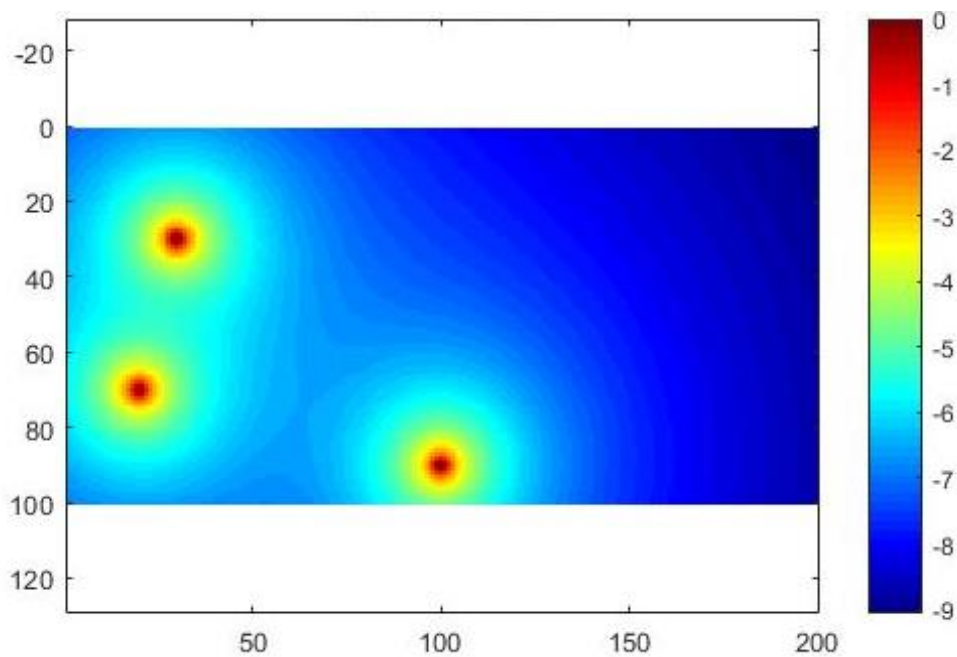


Figura 4.14: Risultato della convoluzione calcolata con metodo esplicito.

4.2.3.4 Calibrazione e selezione dei modelli

I due set di modelli considerati (applicando sia il kernel con legge esponenziale che con legge di potenza) prevedono un numero di parametri che va da 2 a 7, a seconda della combinazione di forzanti ambientali che si intende utilizzare, riportato in Tabella 4.2. Il valore di questi parametri deve essere stimato utilizzando un processo di calibrazione, in modo da ottenere un modello che approssimi al meglio possibile i dati osservati a disposizione, cioè l'urbanizzazione definita dalle matrici degli stati dall'anno 1992 al 2013.

La calibrazione automatica consiste in generale nell'impostazione di una funzione obiettivo che si andrà poi a massimizzare o minimizzare tramite un algoritmo di calcolo; in questo caso di studio si è scelto di cercare quei parametri che massimizzassero la funzione di log-Verosimiglianza (log-Likelihood Function, log-LK) data dall'espressione (25). Per ottenere questi valori, è stato utilizzato l'algoritmo di MATLAB *fmincon* che permette di cercare, seguendo diverse iterazioni partendo da punti casuali nel dominio dei parametri, il minimo di una funzione, fino a che non si raggiunge un criterio di stop. Poiché l'algoritmo è costruito per minimizzare la funzione obiettivo, mentre la log(LK) deve essere massimizzata, si prenderà come migliore set di parametri quello che minimizza $-\log(LK)$. Per avere la certezza che il punto di minimo individuato per la funzione obiettivo sia un minimo assoluto e non locale, l'algoritmo ripete la ricerca a partire da tre punti diversi presi casualmente nel dominio parametri, in modo da esplorare il più possibile lo spazio di variabilità della funzione.

Il dominio dei parametri delle forzanti ambientali è stato definito in base alla variabilità dei dati osservati, in Tabella 4.3 sono riportati i valori limite per ogni parametro.

Parametro	Limiti di variabilità
γ	0.1 – 10
k_0	0.000001 – 0.005
k_T	5 – 100
k_{RD}	2 – 80
k_{SD}	0.02 – 4
k_R	200 - 10000
k_{DEM}	0.1 - 5000

Tabella 4.3: Intervalli di variabilità dei parametri del modello.

Una volta identificati i parametri che massimizzano la funzione di Verosimiglianza per ogni set di forzanti ambientali è necessario confrontare le performance delle diverse combinazioni di modelli per identificare il modello migliore. Per farlo, è stato utilizzato il test di verifica delle informazioni di Akaike

(Akaike Information Criterion) che permette di valutare il modello migliore sulla base della performance del modello (i.e. la funzione di verosimiglianza) e della complessità (i.e. numero di parametri). Il concetto è che si può pensare di utilizzare un numero di parametri minore per avere un modello più semplice e con meno ridondanza, se esso porta a risultati accettabili.

Il criterio di Akaike è definito dalla formula:

$$AIC = -2 \log(LK) + 2n_p$$

Con LK funzione di verosimiglianza e n_p numero di parametri.

Evidentemente, si deve preferire il modello con AIC minore poiché si deve ottenere una massima $\log(LK)$ e, di conseguenza, il primo termine deve risultare minimo.

4.2.3.5 Simulazione dei dati osservati e previsione di scenari futuri

Si è deciso, dopo aver effettuato la calibrazione del modello sugli anni di osservazione, di simulare l'andamento osservato (1992-2013) e quello futuro dell'urbanizzazione per i vent'anni successivi all'ultimo per cui sono disponibili i dati (2013-2033).

Per effettuare previsioni future, è necessario valutare il tasso di espansione urbana negli anni futuri (ovvero l'espansione urbana in uno step temporale, ovvero in un singolo anno). Ciò risulta in realtà semplice, poiché l'espansione è una forzante esterna nota; si può infatti ritenere che la crescita di urbanizzazione e di popolazione siano direttamente correlate e quindi il loro rapporto, ovvero la densità di popolazione urbana, risulterà circa lo stesso in ogni step temporale. Per questo motivo, si considererà nella simulazione in avanti un numero di pixel annuale che passa dallo stato 0 allo stato 1 come una quantità costante, pari a quella che si verifica mediamente negli anni osservati.

Si considera allora un numero di nuovi pixel urbanizzati ogni anno pari a 3000, derivanti da osservazioni precedenti (Figura 4.9) che prevedono un numero di pixel urbanizzati nell'anno 2013 pari a 75916 e 15228 nel 1992.

$$E_{media}(1992 - 2013) = \frac{75916 - 15228}{22} = 2759$$

In Figura 4.15 si può osservare l'andamento del numero di pixel urbanizzati negli anni osservati e in quelli stimati. Negli anni di stima la crescita è presa lineare con una variazione di 3000 pixel per anno, ma anche l'andamento negli anni osservati si avvicina a quello ipotizzato, a meno di qualche fluttuazione.

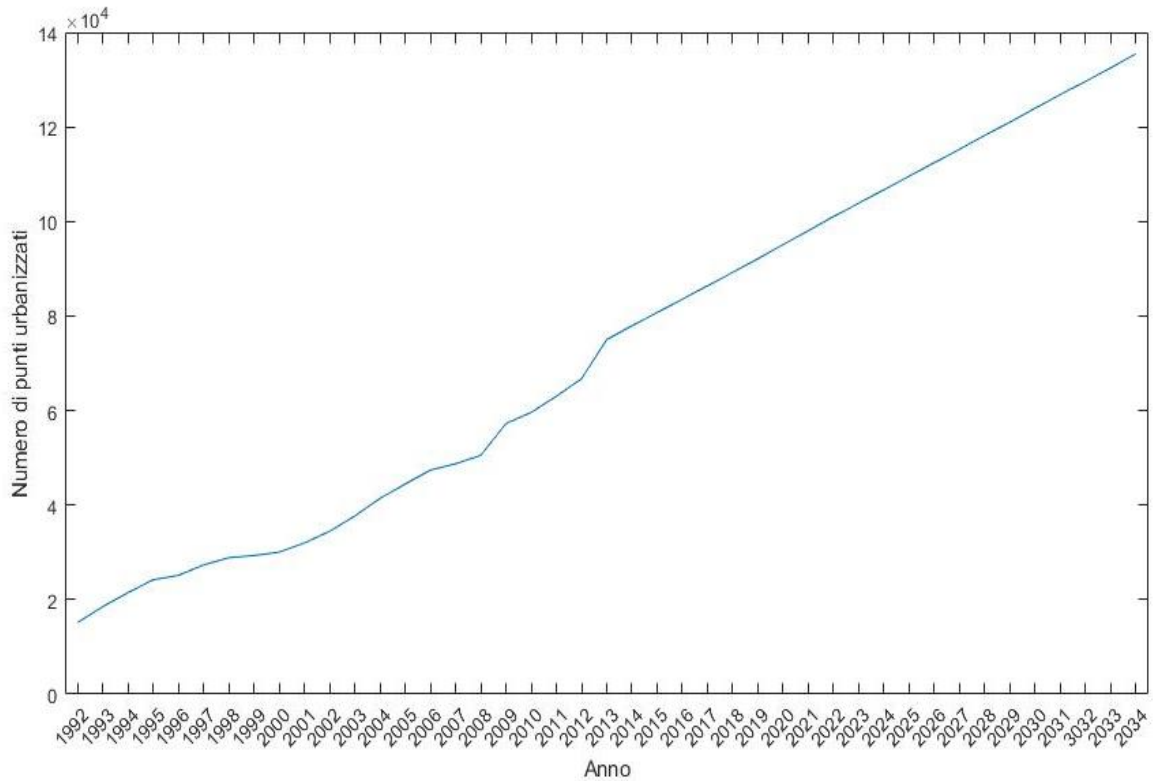


Figura 4.15: Andamento dell'urbanizzazione negli anni osservati (1992-2013) e in quelli stimati (2014-2033).

La previsione di urbanizzazione è quindi ottenuta conoscendo il numero di pixel che compiranno la transizione dallo stato 0 allo stato 1 e individuando quali siano questi punti nell'area di studio sulla base della probabilità descritta al Paragrafo 1. Il processo è così composto:

- Sono noti i pixel urbanizzati e non urbanizzati nell'anno 1992:

$$U(0) = U(1992)$$

$$W(0) = W(1992)$$

- Per ogni step temporale di un anno, il numero di pixel urbanizzati aumenta di 3000 unità disposte nei punti a maggiore probabilità:

$$U(i + 1) = U(i) \mid W(i) < p$$

$$W(i + 1) = W(i) \ \& \ \text{not}(U(i + 1))$$

Le immagini delle matrici ottenute applicando questo calcolo a tutti gli anni simulati (in totale 43 anni) sono riportate per il best model al Paragrafo 3.

4.3 Risultati e considerazioni

Una volta effettuata la calibrazione dei parametri dei vari set di modelli, è possibile confrontare le performance di ogni combinazione per verificare quale sia il migliore sulla base del risultato del criterio di Akaike. Il programma numerico è stato impostato per restituire una tabella di sintesi del test AIC per tutte le combinazioni di parametri ambientali. I risultati sono riportati, alle Tabelle da 4.4 a 4.7, rispettivamente per il set 1 con kernel esponenziale, set 1 con kernel di potenza, set 2 con kernel esponenziale e set 2 con kernel di potenza. Nelle tabelle le combinazioni di modelli sono ordinate in funzione di ΔAIC , in ordine crescente.

Set 1- kernel con legge esponenziale

Modello	Combinazione di parametri						Numero di parametri attivi	-log(LK)	AIC	ΔAIC
	γ	k_0	k_D	k_{SD}	k_{RD}	k_R				
13	✓	✓	✓	✓	✓	-	5	-276673	553356	0
1	✓	✓	✓	✓	✓	✓	6	-276680	553371	15
8	✓	✓	✓	-	✓	-	4	-276692	553391	35
14	✓	✓	✓	-	✓	✓	5	-276700	553410	54
7	✓	✓	✓	✓	-	-	4	-276711	553430	74
3	✓	✓	✓	-	-	-	3	-276721	553449	93
16	✓	✓	✓	✓	-	✓	5	-276737	553484	128
9	✓	✓	✓	-	-	✓	4	-276773	553553	197
15	✓	✓	-	✓	✓	✓	5	-277230	554470	1114
10	✓	✓	-	✓	✓	-	4	-277275	554559	1203
11	✓	✓	-	✓	-	✓	4	-277319	554646	1290
4	✓	✓	-	✓	-	-	3	-277403	554812	1456
12	✓	✓	-	-	✓	✓	4	-277575	555157	1801
6	✓	✓	-	-	-	✓	3	-277920	555846	2490
5	✓	✓	-	-	✓	-	3	-278060	556125	2769
2	✓	✓	-	-	-	-	2	-278159	556321	2965

Tabella 4.4: Risultati della calibrazione per il set 1 con kernel esponenziale per tutte le combinazioni di parametri.

Set 1- kernel con legge di potenza

Modello	Combinazione di parametri						Numero di parametri attivi	-log(LK)	AIC	Δ AIC
	γ	k_0	k_D	k_{SD}	k_{RD}	k_R				
1	✓	✓	✓	✓	✓	-	5	-275229	550467	0
2	✓	✓	✓	✓	✓	✓	6	-275236	550485	18
3	✓	✓	✓	✓	-	-	4	-275245	550497	30
4	✓	✓	✓	✓	-	✓	5	-275258	550527	60
5	✓	✓	✓	-	-	-	3	-275269	550544	77
6	✓	✓	✓	-	✓	✓	5	-275272	550553	86
7	✓	✓	✓	-	✓	-	4	-275331	550669	202
8	✓	✓	✓	-	-	✓	4	-275374	550756	289
9	✓	✓	-	✓	✓	✓	5	-275551	551113	646
10	✓	✓	-	✓	✓	-	4	-275611	551229	762
11	✓	✓	-	✓	-	✓	4	-275642	551292	825
12	✓	✓	-	✓	-	-	3	-275680	551365	898
13	✓	✓	-	-	✓	✓	4	-275951	551909	1442
14	✓	✓	-	-	-	✓	3	-276084	552174	1707
15	✓	✓	-	-	✓	-	3	-276271	552547	2080
16	✓	✓	-	-	-	-	2	-276318	552640	2173

Tabella 4.5: Risultati della calibrazione per il set 1 con kernel di potenza per tutte le combinazioni di parametri.

Set 2 - kernel con legge esponenziale

Modello	Combinazione di parametri						Numero di parametri attivi	-log(LK)	AIC	Δ AIC
	γ	k_0	k_D	k_{SD}	k_{RD}	k_R				
1	✓	✓	✓	✓	✓	-	5	-277047	554104	0
2	✓	✓	✓	✓	✓	✓	6	-277047	554106	2
3	✓	✓	✓	✓	-	-	4	-277136	554280	176
4	✓	✓	✓	✓	-	✓	5	-277146	554302	198
5	✓	✓	-	✓	✓	✓	5	-277230	554470	366
6	✓	✓	-	✓	✓	-	4	-277275	554558	454
7	✓	✓	-	✓	-	✓	4	-277319	554647	543
8	✓	✓	-	✓	-	-	3	-277403	554811	707
9	✓	✓	✓	-	✓	✓	5	-277476	554962	858
10	✓	✓	✓	-	✓	-	4	-277533	555074	970
11	✓	✓	✓	-	-	✓	4	-277569	555146	1042
12	✓	✓	-	-	✓	✓	4	-277574	555157	1053
13	✓	✓	✓	-	-	-	3	-277590	555186	1082
14	✓	✓	-	-	-	✓	3	-277800	555607	1503
15	✓	✓	-	-	✓	-	3	-278059	556125	2021
16	✓	✓	-	-	-	-	2	-278158	556321	2217

Tabella 4.6: Risultati della calibrazione per il set 2 con kernel esponenziale per tutte le combinazioni di parametri.

Set 2 - kernel con legge esponenziale

Modello	Combinazione di parametri						Numero di parametri attivi	-log(LK)	AIC	Δ AIC
	γ	k_0	k_D	k_{SD}	k_{RD}	k_R				
1	✓	✓	✓	✓	✓	-	5	-275520	551051	0
2	✓	✓	✓	✓	✓	✓	6	-275548	551108	57
3	✓	✓	-	✓	✓	✓	5	-275556	551122	71
4	✓	✓	✓	✓	-	✓	5	-275580	551171	120
5	✓	✓	-	✓	✓	-	4	-275611	551229	178
6	✓	✓	-	✓	-	✓	4	-275650	551308	257
7	✓	✓	✓	✓	-	-	4	-275656	551320	269
8	✓	✓	-	✓	-	-	3	-275679	551365	314
9	✓	✓	-	-	✓	✓	4	-275956	551920	869
10	✓	✓	✓	-	✓	✓	5	-275957	551925	874
11	✓	✓	✓	-	-	✓	4	-275980	551968	917
12	✓	✓	✓	-	✓	-	4	-275985	551978	927
13	✓	✓	✓	-	-	-	3	-276013	552032	981
14	✓	✓	-	-	-	✓	3	-276072	552151	1100
15	✓	✓	-	-	✓	-	3	-276271	552547	1496
16	✓	✓	-	-	-	-	2	-276318	552640	1589

Tabella 4.7: Risultati della calibrazione per il set 2 con kernel di potenza per tutte le combinazioni di parametri.

Da questi risultati si evince che, per entrambi i set esaminati ed entrambi i kernel, il modello che raggiunge il miglior compromesso tra la massimizzazione del logaritmo della funzione di log-verosimiglianza e il numero di parametri del modello è quello che tiene in considerazione tutte le forzanti ambientali, ad eccezione della piovosità. Il motivo può essere ricercato nella correlazione tra la forzante piovosità e le variabili temperatura media ed elevazione (Capitolo 3). Tale correlazione appare meno elevata di quella esistente tra elevazione e temperatura media ma può spingere in favore di un modello più semplice, caratterizzato da un numero di parametri minore.

Tra tutte le combinazioni ottimali ottenute con differenti set di parametri e differenti funzioni di kernel, il modello migliore in assoluto è quello che utilizza una legge potenziale del kernel e che prevede le forzanti ambientali del set 1, eccetto la piovosità. Le caratteristiche del modello migliore tra quelli indagati sono riportate in Tabella 4.8.

Forzanti	-Distanza da altri pixel urbanizzati; -Elevazione; -Distanza dalla costa; -Distanza dai corsi d'acqua principali.
Numero di parametri	5
Parametri ottimali	- $\gamma=3.9171$; - $k_0=0.0014$; - $k_D=2690.7$; - $k_{SD}=45.9245$; - $k_{RD}=3.9999$
Tipologia di kernel	Legge di potenza
AIC	550467

Tabella 4.8: Caratteristiche riassuntive del modello ottimale tra quelli esaminati.

Vediamo quindi i risultati ottenuti mediante l'applicazione di questo modello ottimale, sia in termini di riproduzione del processo di urbanizzazione ricavato dai dati di luminosità nel periodo 1992-2013, sia in termini di proiezione di scenari futuri, fino al 2033. Nelle Figure da 4.16 a 4.18 sono riportate le immagini del confronto tra il modello e i dati osservati; si riportano in grigio le aree urbanizzate calcolate in precedenza e in rosso i punti in espansione nell'anno successivo.

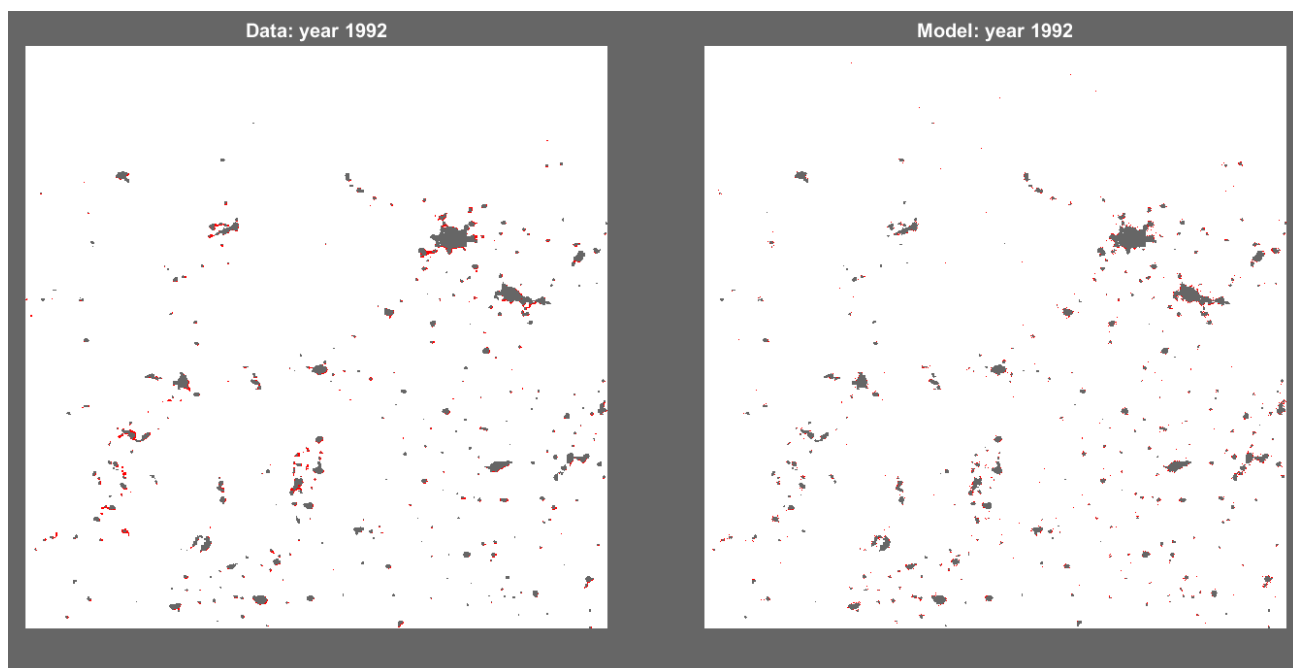


Figura 4.16: Simulazione del processo di urbanizzazione nell'anno 1992: confronto tra i dati osservati (sinistra) e il modello (destra). I pixel grigi rappresentano le aree urbanizzate nel 1992, i pixel rossi identificano le aree in espansione nel 1993.

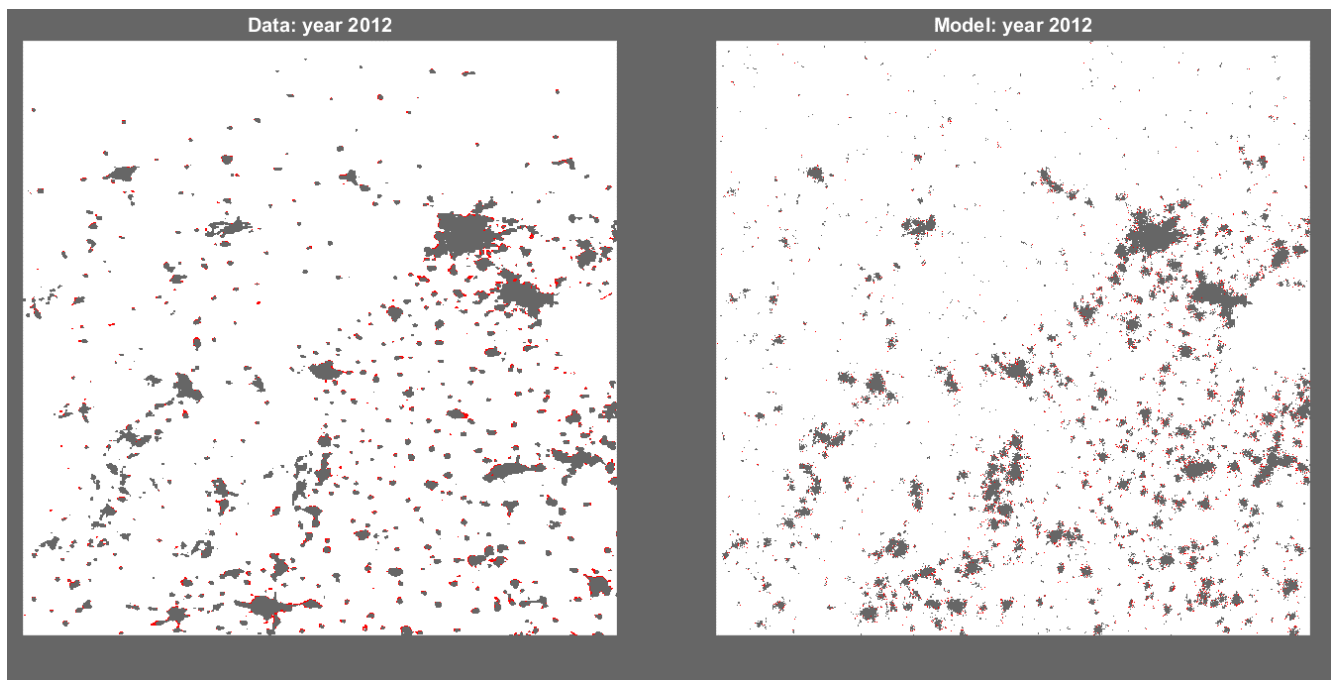


Figura 4.17: Simulazione del processo di urbanizzazione nell'anno 2012 (penultimo anno di osservazioni): confronto tra i dati osservati (sinistra) e il modello (destra). I pixel grigi rappresentano le aree urbanizzate nel 2012, i pixel rossi identificano le aree in espansione nel 2013.

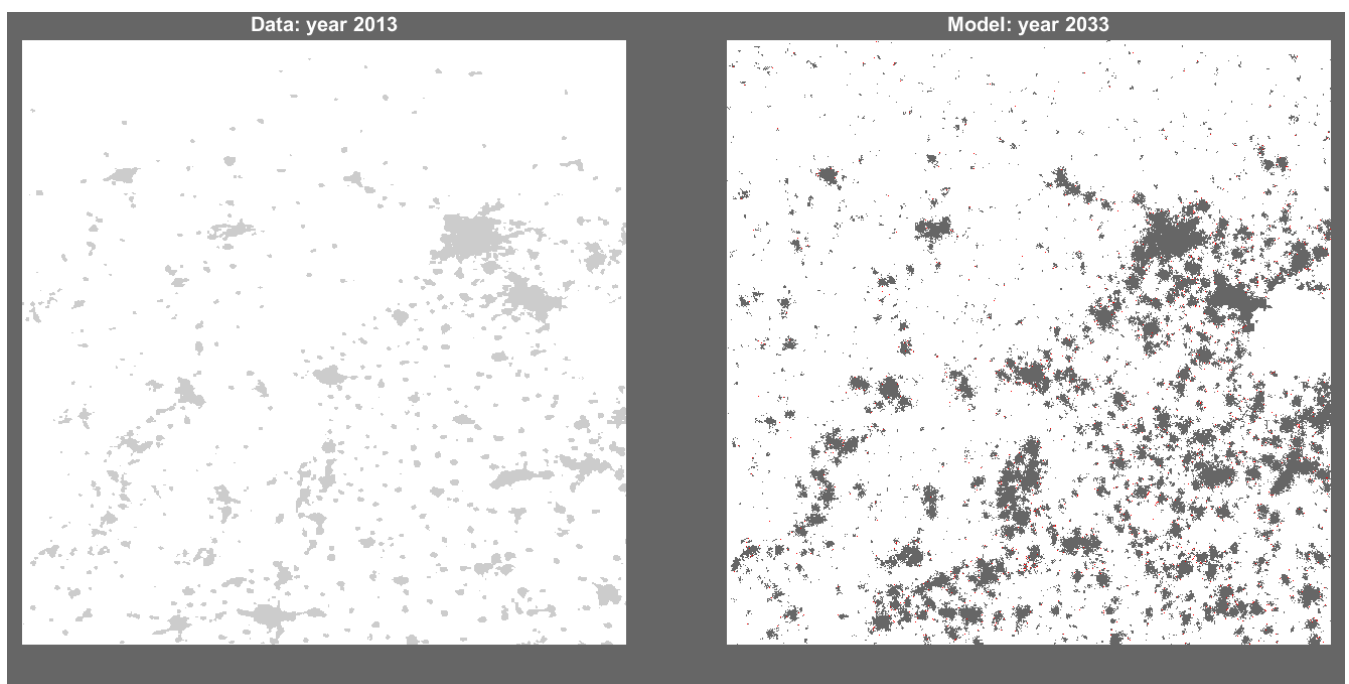


Figura 4.18: Simulazione del processo di urbanizzazione: scenario futuro nell'anno 2033, proiezione dei punti urbanizzati (in grigio).

Osservando la proiezione dei punti in espansione rispetto a quelli che sono risultati effettivamente urbanizzati, si nota che i primi appaiono più caotici e dispersi nell'area di studio rispetto ai secondi;

l'urbanizzazione reale, infatti, risulta molto più concentrata attorno ai punti già urbanizzati, rispetto a quella simulata. Complessivamente, comunque, le forme e la posizione dei cluster simulati risultano simili a quelle reali e il modello può essere tenuto in considerazione per lo scopo previsto.

Per valutare le differenze tra la stima e la realtà, è possibile utilizzare una analisi ROC⁶³ (Receiver Operating Characteristic), applicata per l'ultimo anno di osservazione, il 2013. L'analisi consiste innanzitutto nel confronto delle immagini reali e di quelle simulate, per ottenere il numero di volte in cui si verifica uno dei quattro casi riportati in Figura 4.19.

		valore vero		totale
		p	n	
predizione risultato	p'	Vero Positivo	Falso Positivo	P'
	n'	Falso Negativo	Vero Negativo	N'
totale		P	N	

Figura 4.19: Analisi ROC: Possibili casi che si verificano nel confronto delle immagini. Fonte: https://it.wikipedia.org/wiki/Receiver_operating_characteristic.

In pratica, si possono verificare quattro situazioni differenti:

- Il modello predice un punto urbanizzato ed esso appare urbanizzato anche nella realtà, Vero Positivo (TP);
- Il modello predice un punto urbanizzato ma esso risulta non urbanizzato nella realtà, Falso Positivo (FP);
- Il modello predice un punto non urbanizzato ed esso appare non urbanizzato, Vero Negativo (TN);
- Il modello predice un punto non urbanizzato ma esso appare urbanizzato, Falso Negativo (FN).

Una volta nota la frequenza di questi avvenimenti, è possibile ottenere alcuni indici per verificare la bontà del modello⁶⁴:

- *Recupero*, fornisce il tasso di veri positivi rispetto alla somma di veri positivi e falsi negativi (per questo è anche chiamato *True Positive Rate*), affinché la stima risulti corretta deve essere il più vicino possibile a 1:

$$Recall = TP_{rate} = \frac{TP}{TP + FN} = 0.69 \quad 0 \leq Recall \leq 1$$

- *False positive rate*, si definisce come il numero di falsi positivi rispetto alla somma dei falsi positivi e dei veri negativi, si tratta della probabilità di rigettare l'ipotesi nulla quando in realtà sarebbe corretta, per questo deve risultare il più possibile prossimo allo zero:

$$FP_{rate} = \frac{FP}{FP + TN} = 0.09 \quad 0 \leq FP_{rate} \leq 1$$

- *Precisione*⁶⁵, identifica il tasso di veri positivi rispetto al totale delle previsioni positive. Il significato è quello di verificare il numero di ipotesi effettivamente positive rispetto a quelle prese come tali dal modello:

$$Precision = \frac{TP}{TP + FP} = 0.38 \quad 0 \leq Precision \leq 1$$

- Si può invece definire l'*Accuratezza* come il totale dei casi corretti (TP e TN) rispetto ai casi totali:

$$Accuracy = \frac{TP + TN}{N_{tot}} = 0.89 \quad 0 \leq Accuracy \leq 1$$

Osservando gli ultimi due indici, il risultato ottenuto è piuttosto basso in termini di precisione e conferma il comportamento osservato dalle immagini alle Figure 4.16-4.18, il modello considera come urbanizzati diversi punti che nella realtà non lo sono. Tale risultato è anche legato alla scelta di classificare come urbanizzati tutti quei pixel aventi per almeno due anni consecutivi $x_i(t) = x_i(t + 1) = 1$. In termini di accuratezza, il modello raggiunge invece un valore piuttosto elevato, pari all'89%; in altre parole, predice il valore corretto di urbanizzazione o non urbanizzazione per l'89% dei pixel dell'area di studio e si può quindi ritenere sufficientemente affidabile.

Il risultato, dunque, risulta molto accurato, in grado cioè di prevedere il valore corretto per un numero rilevante di pixel, ma poco preciso, tendendo infatti a prevedere l'urbanizzazione in diversi punti non corretti.

Conclusioni e prospettive future

In questo elaborato di tesi, si è cercato di simulare e prevedere l'urbanizzazione in una regione nel nord-est della Cina tramite un modello stocastico spazialmente esplicito. Lo scopo di questo modello è quello di riprodurre il fenomeno di urbanizzazione osservato nel periodo 1992-2013 per poi rappresentare il pattern futuro degli insediamenti umani descrivendo non solo l'entità della crescita urbana, ma anche dove essa si collocherà nello spazio. La peculiarità di questo studio è quella di cercare di descrivere l'urbanizzazione osservata nell'area di interesse utilizzando solo variabili ambientali e territoriali dipendenti dalla posizione, dal clima, dalla morfologia.

Per determinare quali fossero le forzanti ambientali che governano il fenomeno e quale fosse la loro relazione con la disposizione nello spazio della crescita urbana, si è innanzitutto definita un'area di studio che avesse assistito, negli ultimi anni, a uno sviluppo urbano evidente. È stato quindi valutato l'andamento dell'urbanizzazione a livello globale tramite l'osservazione delle immagini satellitari di luci artificiali notturne allo scopo di definire un'area di studio rappresentativa. Le immagini hanno mostrato, per gli anni in cui sono disponibili, un'evidente crescita luminosa soprattutto nei paesi appartenenti al cosiddetto BRICS (Brasile, Russia, India, Cina, Sudafrica) che, non a caso, è costituito dai Paesi in cui si è verificata la maggiore crescita economica e di popolazione dalla seconda metà del Novecento ad oggi. L'area adeguata allo sviluppo dello studio, inoltre, doveva presentare un territorio il più possibile eterogeneo dal punto di vista geomorfico e climatico, in modo da osservare diverse caratteristiche ambientali e con un ampio spettro di variabilità. Per questi motivi, l'area di studio è stata individuata nel nord-est della Cina, nei dintorni di Pechino, in un'area di circa 1000 x 1000 km².

All'interno di quest'area, si è poi ipotizzato quali potessero essere le condizioni ambientali favorevoli o sfavorevoli all'insediamento umano sulla base dell'esperienza o di studi precedenti (C. Small 2004⁶⁶, C. Small e J. E. Cohen 2004⁶⁷). Sulla base di questi ragionamenti sono state selezionate le forzanti ambientali da inserire nel modello. I risultati delle analisi sulle correlazioni tra l'intensità luminosa e le forzanti ambientali selezionate, hanno poi confermato queste ipotesi iniziali. È risultato che:

- L'uomo tende a insediarsi prevalentemente in zone ad altitudine minore. L'intensità luminosa decresce infatti all'aumentare della quota, probabilmente perché risulta più facile costruire infrastrutture o abitazioni in aree pianeggianti, poco acclivi e impervie;

- L'intensità luminosa decresce all'aumentare della distanza dai corsi d'acqua e dal mare, elementi che da sempre costituiscono importanti vie di comunicazione e di scambio ma anche punti di accesso alla risorsa idrica;
- All'aumentare della temperatura media annuale la luminosità aumenta, confermando la propensione dell'uomo a preferire temperature più miti ai climi rigidi. Non si sono invece trovati riscontri precisi nella correlazione tra la luminosità e l'escursione termica massima annuale, indice del fatto che la variabilità di temperatura non rappresenta un fattore rilevante quanto la temperatura media nella scelta dell'area di insediamento;
- All'aumentare della piovosità media annuale, anche la luminosità notturna aumenta. La pioggia si conferma un parametro significativo per lo sviluppo di popolazione per via dell'effetto sul terreno, sui corsi d'acqua e sulle falde acquifere, garantendo lo sviluppo delle attività umane in diversi settori economici (agricoltura, allevamento ma anche industria).
- Risulta più difficile collegare l'intensità luminosa alla copertura del suolo, l'esperienza suggerirebbe un nesso tra l'urbanizzazione e, per esempio, la copertura vegetale o la tipologia di terreno. Suoli più teneri e ricchi di vegetazione sembrerebbero essere preferiti a territori rocciosi, aridi o spogli, ma per indagare queste ipotesi non è sufficiente l'informazione qualitativa sulla copertura del suolo.

Si è quindi confermata la dipendenza dell'urbanizzazione con le seguenti forzanti ambientali: altitudine, distanza dai corsi d'acqua principali, distanza dal mare, temperatura media annuale e piovosità media annuale. Nel modello finale è stato anche aggiunto il contributo all'espansione urbana dato dalla vicinanza tra un pixel non urbanizzato e altri già urbanizzati allo stesso istante di tempo.

Il modello sviluppato considera l'urbanizzazione come un processo stocastico che si realizza secondo una catena di processi markoviani in cui ogni transizione è indipendente dalle altre e priva di memoria. Ad ogni pixel non urbanizzato i al tempo t ($x_i(t) = 0$) è associata una probabilità $p_i(t)$ di risultare urbanizzato all'istante successivo $t+1$ ($x_i(t + 1) = 1$):

$$t \rightarrow t + 1$$

$$x_i(t) = 0 \xrightarrow{p_i(t)} x_i(t + 1) = 1$$

Tale probabilità $p_i(t)$ dipende dalla funzione matematica scelta per valutare l'effetto della distanza del pixel non urbanizzato da altri urbanizzati (kernel) e da funzioni delle forzanti ambientali elaborate

sulla base delle relazioni individuate tra queste e la luminosità artificiale. La definizione di questa probabilità per ogni pixel non urbanizzato in ogni istante di tempo permette quindi di costruire il pattern futuro dell'urbanizzazione. Per definire i parametri del modello necessari al calcolo della probabilità, è stata effettuata una calibrazione sui dati osservati (dall'anno 1992 al 2013) con lo scopo di massimizzare la funzione di log-verosimiglianza.

Data la forte correlazione inversa individuata tra il DEM e la temperatura media annuale si sono studiati due set di parametri che comprendessero una sola delle due forzanti, in modo da valutare le performance di entrambe ma evitare dati ridondanti. I due set sono poi stati valutati sia con kernel di legge esponenziale che di potenza, per poter osservare quale dei due casi ottenesse risultati migliori. Al fine di considerare tutte le possibilità e verificare se esistessero altri parametri sovrabbondanti, si sono valutate per tutti i set e per entrambi i kernel tutte le possibili combinazioni di forzanti ambientali possibili, mantenendo però il kernel sempre attivo come condizione minima, e si è selezionato il modello migliore sulla base del test di verifica delle informazioni di Akaike.

I risultati ottenuti mostrano che:

- A parità di set e di combinazione, il modello migliore si ottiene attribuendo al kernel una legge di potenza piuttosto che una legge esponenziale. Questo risultato indica che l'effetto dei pixel urbanizzati cala molto velocemente all'aumentare della distanza con i pixel non urbanizzati;
- A parità di kernel e di combinazione, i risultati migliori si ottengono includendo nel set di forzanti ambientali l'altitudine piuttosto che la temperatura (set 1). La crescita urbana è quindi influenzata maggiormente dall'elevazione del terreno che dalla temperatura; mentre l'uomo può adattarsi a vivere in climi meno miti, appare più scoraggiato a insediarsi in territori più elevati e, probabilmente, più aspri o spogli;
- Di tutte le combinazioni possibili, quella da preferire sulla base del test di Akaike è, a parità di set e di kernel, quella che include tutte le forzanti ambientali meno la piovosità media annuale. Probabilmente perché questo parametro era risultato leggermente linearmente correlato a DEM (coefficiente di Pearson pari a -0.72) e temperatura media annuale (coefficiente di Pearson pari a 0.82).

Il risultato migliore, per concludere, si ottiene considerando un modello con kernel di legge di potenza, con un set di forzanti ambientali che include: altitudine, distanza dai corsi d'acqua principali e distanza dal mare.

Si nota immediatamente una caratteristica importante del modello ottimale; le forzanti ambientali che realizzano il risultato migliore sono tutte di tipo geomorfico, restano escluse le forzanti climatiche e

idrologiche. Questo risultato implica che l'urbanizzazione è governata soprattutto dalle caratteristiche del territorio studiato, mentre risultano meno rilevanti quelle climatiche. L'insediamento umano risulta quindi legato strettamente alle risorse idriche disponibili, alla disponibilità di vie di comunicazione e commercio, allo spazio disponibile per l'edificazione e la costruzione di infrastrutture, mentre risulta meno ostacolato da condizioni climatiche avverse, probabilmente per la capacità di adattamento climatica che contraddistingue l'uomo.

Il modello così sviluppato è risultato, tramite un confronto ROC sui risultati del modello per l'ultimo anno di osservazione (2013) e l'urbanizzazione reale osservata per quello stesso anno, piuttosto accurato, assegnando lo stato corretto all'89% dei pixel dell'area di studio. Risulta invece minore la precisione, pari al 38%, poiché il modello tende a prevedere i punti urbanizzati in modo molto più disperso rispetto ai dati osservati nella realtà, anche per via delle scelte iniziali per la definizione dell'urbanizzazione. Analisi future dovranno valutare le performance del modello da un punto di vista statistico, considerando la distribuzione di probabilità delle dimensioni dei cluster urbanizzati.

Il risultato è comunque frutto anche di scelte soggettive, per esempio la modalità di definizione dello stato di urbanizzazione. Il valore di luminosità che è stato selezionato come soglia per il passaggio da uno stato di non urbanizzazione a uno di urbanizzazione, per esempio, può essere modificato in base alle applicazioni o anche sulla base di eventuali valori determinati da studi precedenti sulla luminosità o sulla densità di popolazione. Maggiore sarà il valore di soglia selezionato e minore risulterà allora la superficie urbanizzata. Inoltre, i punti considerati urbanizzati nel modello elaborato in questo studio sono solo quelli che appaiono tali per almeno due anni consecutivi. Tale scelta è stata effettuata in modo da avere un andamento dell'urbanizzazione il più possibile regolare, trascurando pertanto la probabilità di abbandono urbano. Si potrebbero valutare i risultati del modello anche prendendo come urbanizzati tutti i punti che superano il valore di soglia nell'anno di interesse; probabilmente si osserverebbe, in questo caso, un aumento della precisione del modello. A questo scopo, si potrebbero anche ripetere le analisi sostituendo alle immagini di luminosità artificiale mappe del suolo più precise, quali mappe di uso del suolo o di copertura del terreno.

Nello studio, come in altri precedenti, si era suggerita la possibilità di analizzare la relazione tra la copertura del suolo e la luminosità; questi dati non sono poi stati inclusi nella costruzione del modello per via della loro natura qualitativa e variabile nel tempo. Si potrebbe però utilizzare un parametro numerico che indichi l'asperità del suolo, tenendo conto della morfologia, del tipo di terreno e della copertura vegetale. Un esempio di questi parametri, è il *Coefficiente di Asperità (Roughness⁶⁸)* che contiene le informazioni del DEM e della copertura del suolo nella forma di un coefficiente numerico da valutare punto per punto. Il suo utilizzo è raccomandato per eventuali prove future in sostituzione all'elevazione per aggiungere al modello le informazioni sull'aspetto del suolo. Infine, i risultati sono

influenzati dalla scala di applicazione, analizzando con gli stessi metodi riportati in questo elaborato un'area di estensione diversa, si potrebbero ottenere correlazioni differenti tra la luminosità e le forzanti ambientali, ma anche un modello ottimale con set di parametri o funzione di kernel diversi. A una scala maggiore, anche globale, il clima potrebbe infatti influire maggiormente sull'urbanizzazione poiché si osserverebbero variazioni molto più estreme (per esempio, si potrebbero trovare climi artici fatti di temperature troppo rigide, inospitali, o climi tropicali in cui la piovosità elevata permette uno sviluppo agricolo importante e favorisce l'insediamento umano). Il modello, dunque, può essere applicato a qualsiasi scala, con aree di studio molto diverse tra loro. Il risultato migliore potrebbe però essere diverso da quello individuato in questo studio, così come le funzioni che descrivono l'effetto delle forzanti ambientali che andrebbero indagate tramite un'analisi ad hoc per il territorio di interesse.

Per rispondere alle domande preliminari, quindi, il modello spazialmente esplicito sviluppato in questo studio dimostra che le condizioni ambientali (soprattutto quelle geomorfiche) influenzano fortemente lo sviluppo degli insediamenti umani e che si può descrivere con buona accuratezza il processo di urbanizzazione attraverso un numero limitato di parametri legati ad esse. Utilizzando le immagini di luci artificiali notturne come indice dell'urbanizzazione, è quindi possibile riuscire a prevedere quali saranno le aree che vedranno un aumento di popolazione, e quindi di urbanizzazione, negli anni futuri. Per una data area di studio, sarà sufficiente definire l'influenza dei parametri ambientali sulla luminosità notturna in base alle osservazioni satellitari disponibili per poter poi definire la probabilità di urbanizzazione di tutti quei pixel che risultano nel presente non urbanizzati.

È stato ampiamente spiegato quanto sia importante conoscere la futura disposizione della popolazione per valutare i futuri effetti legati all'interazione umana con l'ambiente circostante e viceversa. In questo elaborato di tesi si è cercato di rispondere a queste esigenze in modo matematico, fornendo anche un possibile supporto alla pianificazione delle future azioni umane che dovranno essere messe in atto per limitare i problemi dell'effetto antropico sul clima, sul paesaggio e sulle risorse. Gli anni a venire, infatti, saranno i più importanti nella storia delle interazioni tra l'uomo e l'ambiente; gli effetti dell'urbanizzazione, della crescita di popolazione e dei bisogni ad esse collegati hanno modificato, soprattutto nell'ultimo secolo, le caratteristiche del pianeta Terra. Cercando di prevedere lo sviluppo di questi fenomeni si fornirebbe un punto di partenza per l'elaborazione di soluzioni per porre un argine a questi mutamenti.

Indice delle Figure e delle Tabelle

Figura 1.1: Variazione della densità di popolazione mondiale dall'anno 1970 al 2000. GPW prodotta dal SEDAC, NASA.....	9
Figura 1.2: Confronto tra le fotografie di luci notturne nel 1992 e nel 2014. Fonte: database di Google Earth Engine.....	10
Figura 1.3: Variazione delle luci notturne in Cina tra il 1992 e il 2013, dati DMSP-OLS forniti dal SEDAC (NASA) ed elaborati in ambiente GIS.....	11
Figura 1.4: Andamento digitale delle quote del terreno con risoluzione di 30" di angolo (cioè 1 km lineare). Fonte: USGS HydroSHEDS.....	13
Figura 1.5: Immagine satellitare del territorio cinese ottenuta dal sistema di Landsat 8 tramite letture di riflettanza della superficie. Fonte: dataset di Google Earth Engine.....	13
Figura 1.6: MODIS Land Cover Map, mappa di copertura vegetale per il territorio cinese prodotta a partire dalle fotografie di MODIS. Fonte: dataset di Google Earth Engine.....	14
Figura 1.7: Reticolo idrografico della Cina. Fonte: database di HydroSHEDS.....	15
Figura 1.8: Macro-aree morfologiche della Cina.....	16
Figura 1.9: Temperature medie nel mese più freddo e più caldo e piovosità media in Cina. Fonte: NOAA, FAO database, poi elaborati in ambiente GIS.....	17
Figura 1.10: Immagine satellitare dell'area di studio (Nord-Est della Cina, nei dintorni di Pechino). Fonte: Basemap "World Imagery" di ArcGIS.....	18
Figura 2.1: I prodotti scaricabili dal database GHS, fonte: https://ghsl.jrc.ec.europa.eu/data.php .	21
Figura 2.2: Rappresentazione della GHS-POP per l'area dell'Asia Orientale nell'anno 2015. Sono visibili notevoli deformazioni delle forme rispetto alle rappresentazioni cartografiche conformi.....	22
Figura 2.3: Forme di paralleli e meridiani nella rappresentazione World Mollweide, fonte: Justin Kunimune - Own work, CC BY-SA 4.0, https://commons.wikimedia.org/w/index.php?curid=66467569	23
Figura 2.4: Rappresentazione delle luci notturne all'interno dell'area di studio nell'anno 1992 (sinistra) e 2013 (destra), fonte: raster da DMSP/OLS, NOAA, poi elaborate in ArcGIS.....	25
Figura 2.5: Rappresentazione del DEM per l'area di studio, fonte: USGS HydroSHEDS poi elaborata in ArcGIS.....	28
Figura 2.6: Linea raster che ricalca la linea di costa a livello globale, fonte: http://openstreetmapdata.com	29
Figura 2.7: Confronto tra la line di costa del file raster del DEM e del file vettoriale.....	30

Figura 2.8: Confronto tra le linee di costa del file raster del DEM e di quello ottenuto dalla rasterizzazione del vettoriale di partenza	30
Figura 2.9: Rappresentazione della distanza euclidea in tre dimensioni: mentre con la distanza inclinata si osserva l'effetto della quota, con la distanza in pianta esso sparisce e si osserva la sola distanza x_2-x_1	31
Figura 2.10: File raster ottenuto calcolando la distanza euclidea di ogni pixel dalla costa.	32
Figura 2.11: Differenza tra il raster delle linee di drenaggio e quello ottenuto imponendo una Flow Accumulation > 5000 km ²	33
Figura 2.12: File raster ottenuto calcolando la distanza euclidea di ogni pixel dai corsi d'acqua principali.....	34
Figura 2.13: Confronto tra i raster di temperature minime (a sinistra), massime (al centro) e medie (a destra) nei mesi di gennaio (in alto) e luglio (in basso).	35
Figura 2.14: Aspetto del file geotiff di temperatura media ottenuto in MATLAB.	36
Figura 2.15: Esempio di funzionamento dello strumento Resample. Ogni pixel è diviso in celle di 1 km ² assegnando a ogni cella il valore iniziale del pixel di 100 km ²	38
Figura 2.16: File raster finale di piovosità.	38
Figura 2.17: File raster di copertura del suolo nell'area di studio.	40
Figura 3.1: Disposizione delle luci notturne confrontata con l'andamento delle quote del terreno.	42
Figura 3.2: Confronto tra la luminosità notturna e le classi create per il DEM.	43
Figura 3.3: Andamento della luminosità media per classi di DEM nel 1992 e nel 2013.....	45
Figura 3.4: Andamento della luminosità media per ogni classe di DEM nel tempo.	46
Figura 3.5: Andamento della luminosità media per classi di distanza dalla costa nel 1992 e nel 2013.	48
Figura 3.6: Andamento della luminosità media per ogni classe di distanza dalla costa nel tempo.	49
Figura 3.7: Andamento della luminosità media nel 1992 e nel 2013 per ogni chilometro di distanza dai corsi d'acqua.	50
Figura 3.8: Andamento della luminosità media per classi di distanza dai corsi d'acqua nel 1992 e nel 2013.	51
Figura 3.9: Andamento della luminosità media per ogni classe di distanza dai corsi d'acqua nel tempo.	52
Figura 3.10: Andamento della luminosità media per classi di temperatura media annuale nel 1992 e nel 2013.	53
Figura 3.11: Andamento della luminosità media per classi di escursione termica massima annuale nel 1992 e nel 2013.....	54

Figura 3.12: Andamento della luminosità media per ogni classe di temperatura media annua nel tempo.....	56
Figura 3.13: Andamento della luminosità media per classi di piovosità media nel 1992 e nel 2013.	57
Figura 3.14: Andamento della luminosità media nel tempo per ogni classe di piovosità.....	58
Figura 3.15: Valori di luminosità media per classi di copertura del suolo nel 2013.	60
Figura 3.16: Risultati grafici per la ricerca di correlazione tra i parametri ambientali.	61
Figura 4.1: funzionamento della convoluzione spaziale.....	67
Figura 4.2: Semplificazione del processo di espansione in un intervallo di tempo.....	70
Figura 4.3: Dettaglio della feature dei laghi reperita online a confronto con la Basemap Imagery di ArcGIS. Fonte: https://www.natureearthdata.com/	74
Figura 4.4: Confronto tra il raster laghi creato e la Basemap di ArcGIS in un dettaglio dell'area di studio.....	75
Figura 4.5: Matrice maschera che esclude le aree con presenza di laghi o mare dal calcolo della probabilità di urbanizzazione.	76
Figura 4.6: Confronto tra il raster di posizione dei corsi d'acqua fornito da HydroSHEDS e a basemap Imagery di ArcGIS.	77
Figura 4.7: Confronto tra la basemap Imagery di ArcGIS e la luminosità artificiale nell'anno 1992 per la città di Pechino.	78
Figura 4.8: Matrici degli stati dall'anno 1992 al 2013.	79
Figura 4.9: Confronto tra il numero di pixel urbanizzato applicando semplicemente il valore di soglia a tutti gli anni (in blu) e il numero di pixel urbanizzati ottenuto considerando solo i punti che hanno superato il valore di soglia per almeno due anni consecutivi (in rosso).	80
Figura 4.10: Matrice di distanza con dominio periodico.	81
Figura 4.11: Aspetto della matrice di distanza con dominio raddoppiato, il rettangolo in grigio rappresenta il dominio di studio.	82
Figura 4.12: Immagine zoomata del kernel doppio ottenuto con legge di potenza della distanza..	82
Figura 4.13: Rappresentazione della convoluzione spaziale calcolata tramite la funzione fft di MATLAB per tre punti.	84
Figura 4.14: Risultato della convoluzione calcolata con metodo esplicito.....	84
Figura 4.15: Andamento dell'urbanizzazione negli anni osservati (1992-2013) e in quelli stimati (2014-2033).....	87

Figura 4.16: Simulazione del processo di urbanizzazione nell'anno 1992: confronto tra i dati osservati (sinistra) e il modello (destra). I pixel grigi rappresentano le aree urbanizzate nel 1992, i pixel rossi identificano le aree in espansione nel 1993.	91
Figura 4.17: Simulazione del processo di urbanizzazione nell'anno 2012 (penultimo anno di osservazioni): confronto tra i dati osservati (sinistra) e il modello (destra). I pixel grigi rappresentano le aree urbanizzate nel 2012, i pixel rossi identificano le aree in espansione nel 2013.	92
Figura 4.18: Simulazione del processo di urbanizzazione: scenario futuro nell'anno 2033, proiezione dei punti urbanizzati (in grigio).	92
Figura 4.19: Analisi ROC: Possibili casi che si verificano nel confronto delle immagini. Fonte: https://it.wikipedia.org/wiki/Receiver_operating_characteristic	93
Tabella 2.1: Proprietà dei dati della Global Human Settlement Population Grid.	23
Tabella 2.2: Proprietà dei Nighttime Light Data.	24
Tabella 2.3: Caratteristiche del raster contenente i dati di DEM.	28
Tabella 2.4: Caratteristiche del raster che associa a ogni punto la sua distanza euclidea (in pianta) dalla costa.	31
Tabella 2.5: Caratteristiche del raster contenente le distanze dai corsi d'acqua.	33
Tabella 2.6: Caratteristiche dei raster di temperatura media, minima e massima disponibili per i 12 mesi.	37
Tabella 2.7: Caratteristiche del raster di piovosità.	37
Tabella 2.8: Tipologie di copertura del suolo disponibili per il raster utilizzato nello studio.	39
Tabella 2.9: Caratteristiche del raster rappresentante la copertura di suolo.	40
Tabella 3.1: Classi per la suddivisione dei dati di DEM.	43
Tabella 3.2: Risultati della luminosità media per ogni classe di DEM e per ogni anno di osservazione.	44
Tabella 3.3: Classi di distanza dalla costa.	47
Tabella 3.4: Risultati della luminosità media per ogni classe di distanza dalla costa e per ogni anno di osservazione.	47
Tabella 3.5: Classi di distanza dai corsi d'acqua.	50
Tabella 3.6: Risultati della luminosità media per ogni classe di distanza dai corsi d'acqua e per ogni anno di osservazione.	51
Tabella 3.7: Classi di temperatura media annuale.	53
Tabella 3.8: Classi di escursione termica massima annuale.	54

Tabella 3.9: Risultati della luminosità media per ogni classe di temperatura media annuale e per ogni anno di osservazione.....	55
Tabella 3.10: Classi di piovosità media.....	56
Tabella 3.11: Risultati della luminosità media per ogni classe di distanza dai corsi d'acqua e per ogni anno di osservazione.....	58
Tabella 3.12: Matrice di correlazione di Pearson per Altitudine e Temperatura media	62
Tabella 4.1: I diversi set di variabili ambientali scelti per il modello.....	72
Tabella 4.2: Tutte le possibili combinazioni di parametri prese in esame per ciascuno dei set e kernel simulati.....	73
Tabella 4.3: Intervalli di variabilità dei parametri del modello.....	85
Tabella 4.4: Risultati della calibrazione per il set 1 con kernel esponenziale per tutte le combinazioni di parametri.....	88
Tabella 4.5: Risultati della calibrazione per il set 1 con kernel di potenza per tutte le combinazioni di parametri.....	89
Tabella 4.6: Risultati della calibrazione per il set 2 con kernel esponenziale per tutte le combinazioni di parametri.....	89
Tabella 4.7: Risultati della calibrazione per il set 2 con kernel di potenza per tutte le combinazioni di parametri.....	90
Tabella 4.8: Caratteristiche riassuntive del modello ottimale tra quelli esaminati.....	91

Ringraziamenti

Ringrazio tutta la mia famiglia, per primi i miei genitori per avermi sempre sostenuto e per avermi insegnato l'importanza della conoscenza e della curiosità, fin da bambina. Ringrazio i miei fratelli per aver ascoltato i miei discorsi prima degli esami e per aver saputo sdrammatizzare anche le situazioni più buie.

Ringrazio Stefano, innanzitutto per la pazienza e poi per aver fornito l'attrezzatura tecnologica necessaria allo svolgimento di questo lavoro, senza te questo (e molte altre cose) non sarebbe stato possibile.

Un grazie a tutti i parenti; a tutti quelli che mi hanno incentivato a studiare sempre e a quelli che mi hanno insegnato il valore della praticità.

Grazie ai compagni e alle compagne di corso, che con me quest'anno iniziano un nuovo percorso verso il futuro (e chissà dove ci porterà?). Un grazie agli amici, ai colleghi e a tutti quelli che hanno speso una parola di incoraggiamento per aiutarmi a tagliare finalmente questo traguardo.

Grazie, infine, alla professoressa Serena Ceola che mi ha sempre stimolata e grazie a questo lavoro mi ha insegnato un nuovo metodo per osservare il mondo.

Riferimenti e Bibliografia

- [1] W. Tobler, (1970) "A computer movie simulating urban growth in the Detroit region". *Economic Geography*, 46(Supplement): 234-240.
- [2] *World Population Prospects*, ONU, revisione del 2017.
- [3] R. Mittermeier et al. (2003), "Wilderness: Earth's Last Wild Places", *Conservation Intl.*
- [4] P. Acot (2004), "Storia del clima", editore Donzelli.
- [5] C. Small (2003), *Global population Distribution and Urban Land Use in Geophysical Parameter Space*, Lamont-Doherty earth Observatory, Columbia University, New York.
- [6] <https://ec.europa.eu/jrc/en>.
- [7] <https://hydrosheds.cr.usgs.gov/>.
- [8] <http://sedac.ciesin.columbia.edu/>.
- [9] <https://earthengine.google.com/>.
- [10] W. Tobler (1970), "A computer movie simulating urban growth in the Detroit region". *Economic Geography*, 46(Supplement): 234-240.
- [11] Center for International Earth Science Information Network - CIESIN - Columbia University. 2017. *Global Population Density Grid Time Series Estimates*. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC). <https://doi.org/10.7927/H47M05W2>.
- [12] *Image and data processing by NOAA's National Geophysical Data Center. DMSP data collected by US Air Force Weather Agency.*
- [13] <https://www.bbc.com/news/world-asia-pacific-13017882>.
- [14] <http://www.worldbank.org/en/country/china/overview#3>.
- [15] The World Bank, <http://www.worldbank.org/>.
- [16] The World Bank, <http://www.worldbank.org/>.
- [17] <https://en.wikipedia.org/wiki/China>.
- [18] CIA World Factbook-China geography, <https://www.cia.gov/library/publications/the-world-factbook/geos/ch.html>.
- [19] USGS Landsat 8 Collection 1 Tier 1 Raw Scenes, https://developers.google.com/earth-engine/datasets/catalog/LANDSAT_LC08_C01_T1.

-
- [20] MCD12Q1.006 MODIS Land Cover Type Yearly Global 500m, NASA LP DAAC at the USGS EROS Center, https://developers.google.com/earth-engine/datasets/catalog/MODIS_006_MCD12Q1.
- [21] USGS HydroSHEDS, <https://hydrosheds.cr.usgs.gov/>.
- [22] https://en.wikipedia.org/wiki/List_of_saltwater_lakes_of_China.
- [23] <http://www.cina-turismo.com/idrografia.html>.
- [24] <http://blog.zonageografia.deascuola.it/articoli/i-laghi-della-cina-stanno-sparendo>.
- [25] C. Small (2003), "Global population Distribution and Urban Land Use in Geophysical Parameter Space", Lamont-Doherty earth Observatory, Columbia University, New York.
- [26] <https://ghsl.jrc.ec.europa.eu/>.
- [27] GHSL - Global Human Settlement Layer, A new open and free tool for assessing the human presence on the planet.
- [28] European Commission, Joint Research Centre (JRC); Columbia University, Center for International Earth Science Information Network - CIESIN (2015): GHS population grid, derived from GPW4, multitemporal (1975, 1990, 2000, 2015). European Commission, Joint Research Centre (JRC) [Dataset] PID: http://data.europa.eu/89h/jrc-ghsl-ghs_pop_gpw4_globe_r2015a.
- [29] <http://sedac.ciesin.columbia.edu/data/collection/gpw-v4>.
- [30] https://en.wikipedia.org/wiki/Mollweide_projection.
- [31] Professor S. Gandolfi, "Appunti di Cartografia", insegnamento di Topografia per il corso di laurea in ingegneria per l'ambiente e il territorio, Unibo, A.A. 2008-2009.
- [32] X. Li, C. Elvidge, Y. Zhou, C. Cao & T. Warner (2017) "Remote sensing of night-time light", *International Journal of Remote Sensing*, 38: 5855-5859.
- [33] Y. Ju, I. Dronova, Q. Ma & X. Zhang. (2017). "Analysis of Urbanization Dynamics in Mainland China Using Pixel-Based Night-Time Light Trajectories from 1992 to 2013." *International Journal of Remote Sensing* 38: 6047–6072. doi:10.1080/01431161.2017.1302114.
- [34] M. Bennett, and L. C. Smith. (2017), "Using Multitemporal Night-Time Lights Data to Compare Regional Development in Russia and China, 1992–2012." *International Journal of Remote Sensing* 38: 5962–5991. doi:10.1080/01431161.2017.1312035.
- [35] C. D. Elvidge, K. Baugh, M. Zhizhin, F. C. Hsu & T. Ghosh (2017) "VIIRS night-time lights, *International Journal of Remote Sensing*", 38:21, 5860-5879, doi: 10.1080/01431161.2017.1342050.
- [36] J. Bongaarts (2009), "Human population growth and the demographic transition", *Philos Trans R Soc Lond B Biol Sci.* 2009 Oct 27; 364(1532): 2985–2990. doi: 10.1098/rstb.2009.0137.
- [37] M. Sidrauski (1967), "Rational choice and patterns of growth in a monetary economy", Massachusetts Institute of Technology.

[38] M. Kremer (1993), "Population Growth and Technological Change: One Million B.C. to 1990", *The Quarterly Journal of Economics*, Volume 108, Issue 3, 1 August 1993, Pages 681–716.

[39] <https://hydrosheds.cr.usgs.gov/datadownload.php?reqdata=30demg>.

[40] H. Gong, Y. Pan, L. Zheng et al. (2018), "Long-term groundwater storage changes and land subsidence development in the North China Plain (1971–2015)", *Hydrogeol J* 26: 1417. <https://doi.org/10.1007/s10040-018-1768-4>.

[41] Professor M. Barbarella, "Appunti di cartografia numerica e sistemi informativi territoriali", insegnamento di Cartografia Numerica M per il corso di laurea in ingegneria per l'ambiente e il territorio, Unibo, A.A. 2015/2016.

[42] J. Mård1, G. Di Baldassarre, M. Mazzoleni (2018), "Nighttime light data reveal how flood protection shapes human proximity to rivers", *Science Advances* 22 Aug 2018, Vol. 4, no. 8, eaar5779. DOI: 10.1126/sciadv.aar5779.

[43] C. Small & J. E. Cohen (2004), "Continental Physiography, Climate, and the Global Distribution of Human Population," *Current Anthropology* 45, no. 2 (April 2004): 269-277. <https://doi.org/10.1086/382255>.

[44] <http://worldclim.org>, dati creati da Steve Fick e Robert Hijmans.

[45] FAO/IIASA, 2011. *Global Agro-ecological Zones (GAEZ v3.0)*. FAO Rome, Italy and IIASA, Laxenburg, Austria.

[46] C. Small & J. E. Cohen (2004), "Continental Physiography, Climate, and the Global Distribution of Human Population," *Current Anthropology* 45, no. 2 (April 2004): 269-277.

[47] *Global Land Cover by National Mapping Organizations: GLCNMO Version 3* © Geospatial Information Authority of Japan, Chiba University and Collaborating Organizations.

[48] C. Small (2004): "Global Population Distribution and Urban Land Use in Geophysical Parameter Space", Lamont-Doherty Earth Observatory, Columbia University, New York, New York, [https://doi.org/10.1175/1087-3562\(2004\)008<0001:GPDAUL>2.0.CO;2](https://doi.org/10.1175/1087-3562(2004)008<0001:GPDAUL>2.0.CO;2).

[49] C. Small (2004): "Global Population Distribution and Urban Land Use in Geophysical Parameter Space", Lamont-Doherty Earth Observatory, Columbia University, New York, New York, [https://doi.org/10.1175/1087-3562\(2004\)008<0001:GPDAUL>2.0.CO;2](https://doi.org/10.1175/1087-3562(2004)008<0001:GPDAUL>2.0.CO;2).

[50] M. Meybeck, P. Green and C. Vörösmarty (2001), "A New Typology for Mountains and Other Relief Classes: An Application to Global Continental Water Resources and Population Distribution", *Mountain Research and Development*, Vol. 21, No. 1 (Feb., 2001), pp. 34-45, Published by: International Mountain Society. https://www.jstor.org/stable/3674130?seq=1&cid=pdf-reference#references_tab_contents.

[51] F. Finger, A. Knox, E. Bertuzzo, L. Mari, D. Bompangue, M. Gatto, I. Rodriguez-Iturbe, A. Rinaldo (2014), "Cholera in the Lake Kivu region (DRC): Integrating remote sensing and spatially explicit epidemiological modeling", *Water Resour. Res.*, 50, 5624-5637, doi: <https://doi.org/10.1002/2014WR015521>.

[52] J. B. Dunning, JR., D. J. Stewart, B. J. Danielson, B. R. Noon, T. L. Root, R. H. Lamberson, E. E. Stevens (1995), "Spatially explicit population models: current forms and future uses", *Ecological Applications*, 5(1), 1995, pp. 3-11, 1995 by the Ecological Society of America.

[53] D. Rybski, A. G. C. Ros, J. P. Kropp (2013), "Distance-weighted city growth", *Physical Review E* 87, 042114 (2013). DOI: 10.1103/PhysRevE.87.042114.

[54] W. Tobler, (1970) "A computer movie simulating urban growth in the Detroit region". *Economic Geography*, 46(Supplement): 234-240.

[55] <https://www.apprendimentoautomatico.it/apprendimentoautomatico-wpblog/identificazione-di-oggetti-in-immagini-con-convolutional-neural-network-python-e-mxnet/>.

[56] M. Bramanti, C. D. Pagani, S. Salsa (2009), "Analisi matematica 2", Zanichelli Editore.

[57] https://it.wikipedia.org/wiki/Approssimazione_di_Stirling.

[58] <https://www.natureearthdata.com/>.

[59] M. Bramanti, C. D. Pagani, S. Salsa, 2009, "Analisi matematica 2", Zanichelli Editore.

[60] https://en.wikipedia.org/wiki/Discrete_Fourier_transform.

[61] https://en.wikipedia.org/wiki/Fast_Fourier_transform.

[62] <https://it.mathworks.com/help/matlab/ref/fft.html>.

[63] https://it.wikipedia.org/wiki/Receiver_operating_characteristic.

[64] Tesi di laurea di Andrea Ruggeri, "Analisi di dati satellitari a scala globale per la stima dell'impatto antropico sull'inquinamento dei corsi d'acqua".

[65] https://en.wikipedia.org/wiki/Accuracy_and_precision.

[66] C. Small (2004): "Global Population Distribution and Urban Land Use in Geophysical Parameter Space", Lamont-Doherty Earth Observatory, Columbia University, New York, New York, [https://doi.org/10.1175/1087-3562\(2004\)008<0001:GPDAUL>2.0.CO;2](https://doi.org/10.1175/1087-3562(2004)008<0001:GPDAUL>2.0.CO;2).

[67] C. Small & J. E. Cohen (2004), "Continental Physiography, Climate, and the Global Distribution of Human Population," *Current Anthropology* 45, no. 2 (April 2004): 269-277. <https://doi.org/10.1086/382255>.

[68] M. Meybeck, P. Green & C. Vörösmarty (2001), "A New Typology for Mountains and Other Relief Classes: An Application to Global Continental Water Resources and Population Distribution", *Mountain Research and Development*, Vol. 21, No. 1 (Feb., 2001), pp. 34-45, Published by: International Mountain Society. https://www.jstor.org/stable/3674130?seq=1&cid=pdf-reference#references_tab_contents.