

UNIVERSITA DEGLI STUDI DI BOLOGNA

Dipartimento di Ingegneria Chimica, Mineraria e delle Tecnologie Ambientali
Scuola di dottorato: Ingegneria Civile ed Architettura
Settore Scientifico Disciplinare ING-IND/28

UNIVERSITE DE TOULOUSE II-LE MIRAIL

UFR Sciences, espaces, sociétés
Département de géographie et aménagement-GEODE, UMR 5602 CNRS

TESI

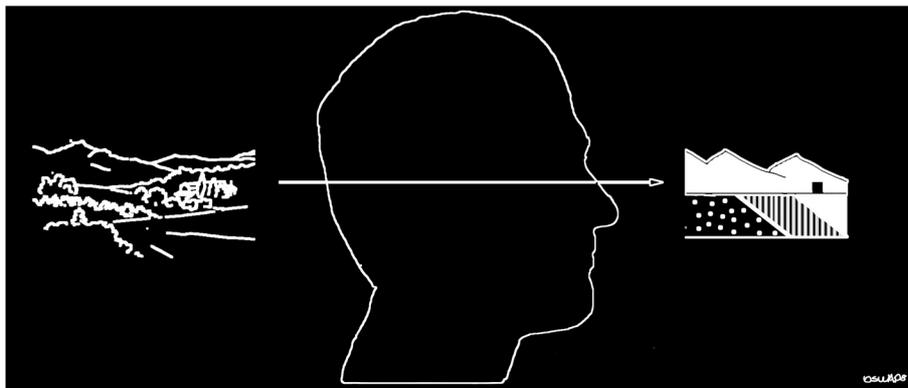
per ottenere il titolo di

Dottore di Ricerca in GEORISORSE e GEOTECNOLOGIE-XX Ciclo
e

Docteur en GEOGRAPHIE et AMENAGEMENT

Marco FOLLADOR

**MODELLIZZAZIONE SPAZIO-TEMPORALE DELLE
DINAMICHE DI USO DEL SUOLO ED ANALISI
COMPARATIVA DI DIFFERENTI APPROCCI PREDITTIVI
USO INTEGRATO DI SIG E TELERILEVAMENTO
NELLO STUDIO DEI PROCESSI DI DEFORESTAZIONE NELLA REGIONE
LA JOYANCA IN PETEN, GUATEMALA**



07 marzo 2008

Membri della commissione:

Prof. Roberto BRUNO, Università di Bologna, tutor

Prof. Martin PAEGELOW, Université de Toulouse II, co-tutor

Prof. Michel MAIGNAN, Université de Lausanne, membro effettivo

Prof. Jacques HUBSCHMAN, Université de Toulouse II, membro effettivo

Prof. Ezio CRESTAZ, SNAM Progetti, membro esperto

Prof. Nathalie VILLA, Université de Toulouse III, membro esperto

Coordinatore del dottorato:

Prof. Paolo BERRY, Università di Bologna

UNIVERSITA DEGLI STUDI DI BOLOGNA

Dipartimento di Ingegneria Chimica, Mineraria e delle Tecnologie Ambientali
Scuola di dottorato: Ingegneria Civile ed Architettura
Settore Scientifico Disciplinare ING-IND/28

UNIVERSITE DE TOULOUSE II-LE MIRAIL

UFR Sciences, espaces, sociétés
Département de géographie et aménagement-GEODE, UMR 5602 CNRS

TESI

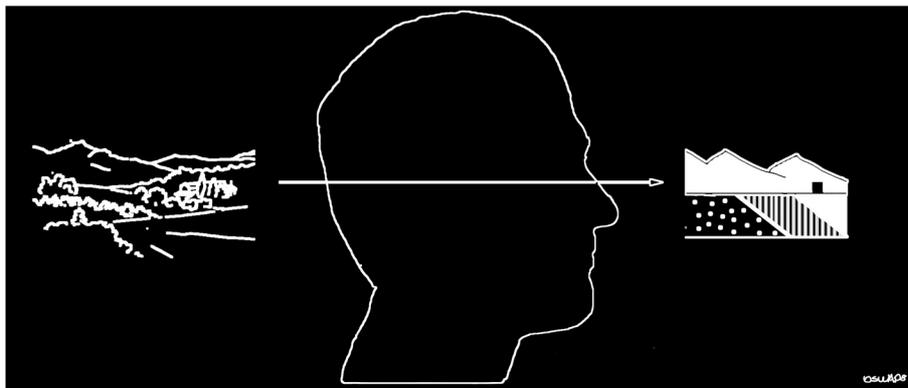
per ottenere il titolo di

Dottore di Ricerca in GEORISORSE e GEOTECNOLOGIE-XX Ciclo
e

Docteur en GEOGRAPHIE et AMENAGEMENT

Marco FOLLADOR

**MODELLIZZAZIONE SPAZIO-TEMPORALE DELLE
DINAMICHE DI USO DEL SUOLO ED ANALISI
COMPARATIVA DI DIFFERENTI APPROCCI PREDITTIVI
USO INTEGRATO DI SIG E TELERILEVAMENTO
NELLO STUDIO DEI PROCESSI DI DEFORESTAZIONE NELLA REGIONE
LA JOYANCA IN PETEN, GUATEMALA**



07 marzo 2008

Membri della commissione:

Prof. Roberto BRUNO, Università di Bologna, tutor

Prof. Martin PAEGELOW, Université de Toulouse II, co-tutor

Prof. Michel MAIGNAN, Université de Lausanne, membro effettivo

Prof. Jacques HUBSCHMAN, Université de Toulouse II, membro effettivo

Prof. Ezio CRESTAZ, SNAM Progetti, membro esperto

Prof. Nathalie VILLA, Université de Toulouse III, membro esperto

Coordinatore del dottorato:

Prof. Paolo BERRY, Università di Bologna

*Sim, passava aqui frequentemente há vinte anos...
Nada está mudado – ou, pelo menos, não dou por isto –
Nesta localidade da cidade ...Há vinte anos!...
O que eu era então! Ora, era outro...
Há vinte anos, e as casas não sabem de nada...Vinte anos inúteis (e sei lá se
o foram!
Sei eu o que é útil ou inútil?)...
Vinte anos perdidos (mas o que seria ganhá-los?)*

*Tento reconstruir na minha imaginação
Quem eu era e como era quando por aqui passava
Há vinte anos...
Não me lembro, não me posso lembrar.*

*O outro que aqui passava, então,
Se existisse hoje, talvez se lembrasse...
Há tanta personagem de romance que conheço melhor por dentro
De que esse eu-mesmo que há vinte anos passava por aqui!*

*Sim, o mistério do tempo.
Sim, o não se saber nada,
Sim, o termos todos nascido a bordo
Sim, sim, tudo isso, ou outra forma de o dizer...*

*Daquela janela do segundo andar, ainda idêntica a si mesma,
Debruçava-se então uma rapariga mais velha que eu, mais lembradamente de
azul.
Hoje, se calhar, está o quê?
Podemos imaginar tudo do que nada sabemos.
Estou parado físisca e moralmente: não quero imaginar nada...*

*Houve um dia em que subi esta rua pensando alegremente no futuro,
Pois Deus dá licença que o que não existe seja fortemente iluminado,
Hoje, descendo esta rua, nem no passado penso alegremente.
Quando muito, nem penso...
Tenho a impressão que as duas figuras se cruzaram na rua, nem então nem
agora,
Mas aqui mesmo, sem tempo a perturbar o cruzamento.*

*Olhamos indiferentemente um para o outro.
E eu o antigo lá subi a rua imaginando um futuro girassol,
E eu o moderno lá descí a rua não imaginando nada.*

*Talvez isso realmente se desse...
Verdadeiramente se desse...
Sim, carnalmente se desse...*

Sim, talvez...

[Álvaro de Campos, O engenheiro aposentado, 1932]

A mamma, papa, Enri e Fernanda.
Ai miei anni d'università

Indice

I INTRODUZIONE	i
i.I. Contesto in cui si sviluppa la ricerca	iii
i.II. Obbiettivi	v
Obbiettivi specifici	vi
i.III. Ipotesi	vii
i.IV. Indice commentato	viii
I N T R O D U C T I O N	I
i.I.Contexte de la Recherche	III
i.II.Objectifs	VI
i.III.Hypothèses	VIII
i.IV.Indice Commenté	VIII
<hr/>	
CAPITOLO 1. L'importanza dello studio delle dinamiche di uso del suolo nell'economia dei cambi global	1
1.1 Uno sguardo rapido sui cambiamenti globali	3
1.2 Dinamiche globali e cambiamenti dell'uso del suolo (LUCC)	14
1.2.1 Premessa epistemologica	14
1.2.2 Le dinamiche di uso del suolo: LUCC	15
<hr/>	
CAPITOLO No.2 La modellizzazione dei cambi di uso del suolo	23
2.1. I modelli e la modellizzazione: introduzione	25
2.2. I modelli di LUCC	38
2.3. Classificazione dei modelli di LUCC	44
2.3.1. Introduzione	44
2.3.2. I modelli statistici ed econometrici	50
2.3.3. I modelli dinamici spaziali	54

2.2.3.1. Introduzione	54
2.2.3.2. I modelli basati su Automa Cellulare (CA)	56
2.2.3.3 I modelli Ibridi Integrati	63
2.2.3.4 I modelli Multi Agente	67
2.2.3.5 I Modelli Apprendisti	71
2.2.3.6 I modelli basati su SIG	71

CAPITOLO No. 3 Descrizione dei modelli utilizzati nello studio dei LUCC in Peten	73
---	-----------

Introduzione generale	75
3.1. PNNET: una rete neurale predittiva	76
3.1.1. Introduzione	76
3.1.2 Caratteristiche generali di una rete di neuroni artificiali	78
3.1.3 L'apprendimento di una rete neurale	82
3.1.4 Il Perceptrone multi-strato (MLP)	84
3.1.5 Descrizione del modello PNNET	88
3.1.6 PNNET: descrizione dei programmi in R©	95
3.1.6.1 Il programma forme.R	96
3.1.6.2 Il programma Perceptron.R	99
3.1.6.3. Il programma Statistique.R	100
3.2 Dinamica EGO	101
3.2.1 Introduzione	101
3.2.2 L'approccio modellizzativo	103
3.2.2.1 Calcolo quantitativo dei cambiamenti futuri	103
3.2.2.2 Allocazione spaziale dei cambiamenti	106
3.2.2.3 La simulazione dell'uso del suolo	111
3.3 Il modello Geomatico combinato CA_Markov	114
3.4 Land Cover Modeler LCM	119

CAPITOLO No. 4 L'area di studio in Peten e la creazione di un database geografico **123**

4.1. La Joyanca e le problematiche ambientali in Peten	125
4.1.1. Introduzione	125
4.1.2 La Joyanca: ambiente e geomorfologia	128
4.1.3 La colonizzazione de La Joyanca	132
4.2. Uso integrato di telerilevamento e SIG per la costruzione di un database geografico	136
4.2.1 Introduzione	136
4.2.2 Pre-processamento delle immagini satellitari	140
4.2.2.1 Acquisizione delle immagini e definizione del rettangolo involvente	140
4.2.2.2 Creazione di una maschera binaria	141
4.2.2.3 Miglioramento del contrasto	143
4.2.2.4 Correzione radiometrica	146
4.2.2.5 Selezione delle bande più adeguate per le composizioni di colore	151
4.2.2.6 La dimensione del pixel	155
4.2.3 Individuazione dei cambi di uso del suolo attraverso lo studio multitemporale di immagini satellitari	158
4.2.3.1 Tecniche di analisi multitemporale	158
4.2.3.2 Gli indici vegetali e di stress idrico	160
4.2.3.3 Diagramma di dispersione dei ND	164
4.2.3.4 Analisi multitemporale RGB(NDMI)	166
4.2.4 Creazione dell'input per la simulazione	169
4.2.4.1 Classificazione delle immagini	169
4.2.4.2 Variabili esplicative	173
4.2.4.3 Analisi esplorativa e scelta delle variabili	174
4.2.4.4 Considerazioni finali	177

CAPITOLO 5. Validazione ed analisi comparativa dei modelli predittivi **179**

5.1 Evoluzione del fenomeno e dei processi-pilota	181
5.2 Applicazione dei modelli al caso della Joyanca	184

5.3 Validazione e discussione dei risultati	187
5.3.1 Analisi visuale	188
5.3.2 Analisi quantitativa	193
5.3.3 Analisi pixel-a-pixel	194
5.3.4 Land Use Cover Change Budget	195
5.3.5 Complementarietà dei modelli	198
5.3.6 Simulazione e frequenza del cambio osservato	199
5.3.7 Validazione Fuzzy	201
5.4 Conclusioni	204

CONCLUSI ONI	207
c.1. Conclusioni generali	209
c.2. Conclusioni "locali"	215
c.3. Limiti e prospettive future	217

CONCLUSI ONS	221
c.1. Conclusions Générales	223
c.2. Conclusions "locales"	229
c.3. Limites et perspectives futures	231

In Musica...	235
In Poesie...	237
Ri ngrazi amenti	239
BI BLI OGRAFIA	241

Li sta del l e Fi gure	263
Li sta del l e Tabel l e	265

Sulla tesi...

"J'aime la concision, j'abhorre la pléthore, l'obésité des thèses, ces sommes assommantes, cette adiposité que secrète, trop souvent, notre Université, délayant en mille pages ce qui pourrait se condenser en cent, pour le plus grand bénéfice du lecteur.

On ne trouvera donc pas, ici, une histoire exhaustive des (modèles) et de (la modélisation), mais une réflexion sur leur fonction, utilité et limites. On ne trouvera pas, enfin, cet étalage d'érudition, qui vise à intimider le lecteur, bien plus qu'à l'informer. Les références indispensables se concentrent dans les notes, comme autant d'incitations à poursuivre l'investigation. A chacun d'en user à sa guise".¹

[Roger, 1997, Pag. 8-9]

¹ Amo la concisione, detesto la sovrabbondanza, l'obesità delle tesi, queste moli pesanti, questa adiposità che troppo spesso, la nostra Università, spalma su mille pagine ciò che invece potrebbe essere condensato in cento, con grande beneficio per il lettore. Non troverete qui una descrizione storica esaustiva dei modelli e della modellizzazione, ma una riflessione personale sul loro funzionamento, utilità e limiti. Non troverete, qui, infine lo sfoggio d'erudizione, che vuole intimidire il lettore più che informarlo. Le referenze indispensabili si concentrano nelle note, come un'incitazione a proseguire la ricerca. Ad ognuno il piacere di usarla come gli pare.

I NTRODUZI ONE

“Pensare che una delle conseguenze della conoscenza è la diminuzione del rischio, è un’illusione. Infatti, è impossibile riuscire a comprendere tutto e alla fine, il rischio non è mai eliminato; semplicemente prende forme diverse.” (Claude Raffestin, 2005 p.14)

I NTRODUZI ONE

i.1. Contesto in cui si sviluppa la ricerca

La necessità di comprendere i meccanismi che regolano i cambiamenti globali in atto ed offrire così uno strumento di pianificazione futura, rappresenta oggi una priorità della ricerca scientifica. Quest'ultima combina osservazioni, analisi dei processi, esperimenti empirici e modellizzazione, per migliorare la nostra capacità di spiegare e prevedere la coevoluzione del sistema uomo-ambiente.

La comprensione dei cambiamenti globali passa attraverso la comprensione del ruolo fondamentale giocato dall'attività antropica nella rapida trasformazione degli ecosistemi e del paesaggio, rispetto alla lenta variabilità presente in natura.

I macro-effetti dell'impronta umana sul funzionamento del sistema terrestre, sono il risultato "sorprendente" dell'azione sinergica di micro-processi a scala locale, spesso non considerati o sufficientemente studiati (e da qui la sorpresa!). I cambiamenti globali sono dunque dei fenomeni complessi su larga scala che emergono dall'interazione simultanea di agenti indipendenti, molto meno complessi.

Numerosi progetti internazionali sono stati creati negli ultimi anni per rispondere al crescente desiderio di capire le intricate relazioni fra la società ed il suo milieu, e la maniera in cui queste possono o potranno modificare la capacità del pianeta ad accogliere la vita ed a garantire il nostro benessere. L'esatto comportamento di questo sistema è spesso imprevedibile, perché emergente piuttosto che predeterminato.

I meccanismi che portano una società a degradare il proprio intorno sono ancora poco chiari, e frequentemente descritti in maniera semplicistica. L'uso del suolo deriva dalla combinazione sinergica di più fattori, ordinati su differenti risoluzioni spazio-temporali e livelli

organizzativi che variano in funzione del contesto geografico e socio-politico, secondo la seguente espressione generale:

Uso del suolo = f (pressione, opportunità, politiche, vulnerabilità, organizzazione sociale)

dove

pressione = f (demografia, scarsità di risorse, povertà, infrastrutture)

opportunità = f (leggi di mercato, costi, tecnologia)

politiche = f (sussidi, quote, leggi, tasse, piani di sviluppo)

vulnerabilità = f (fragilità del sistema sociale e ambientale, dipendenza da capitale straniero)

organizzazione sociale = f (accesso alle risorse, interazioni urbano-rurali, cultura)

Il metabolismo industriale moderno ha incrementato i flussi di energia e di materiali risultanti dall'estrazione, la trasformazione e l'uso delle risorse naturali. Gli effetti più evidenti di queste attività sono visualizzabili attraverso i processi di cambio della copertura terrestre, dovuti principalmente, da un punto di vista economico e di estensione spaziale, all'agricoltura, alla pastorizia, alla produzione di prodotti forestali, all'espansione urbana ed all'estrazione di georisorse.

Attualmente le regioni selvagge, lievemente perturbate dall'uomo, rappresentano meno della metà delle terre emerse; in tutte le aree del mondo l'agricoltura e la pastorizia hanno rimpiazzato la vegetazione naturale per far fronte ad un aumento vertiginoso dei consumi, legato in parte alla crescita demografica ed in parte alla diffusione di uno stile di vita occidentale. Nei paesi tropicali la situazione è più grave, poiché le terre ancora disponibili per un'espansione agricola sono limitate; la foresta copre gran parte del territorio ed il suolo è povero in nutrienti. I piani di colonizzazione lungo i fronti pionieri sono il risultato di decisioni politiche che valutano i benefici a breve termine e che mirano a risolvere tensioni sociali interne o ad attirare capitali stranieri.

Le conseguenze di questo processo di distruzione della foresta primaria sono varie, e vanno dalla riduzione della biodiversità fino all'alterazione del clima globale (emissione di gas ad effetto serra e modificazione dell'albedo terrestre), dalla degradazione del suolo al comportamento anomalo dei regimi idrologici, fino all'esaurimento dei prodotti forestali ed alimentari.

In questo contesto la Riserva della Biosfera Maya (RBM) gioca un ruolo importante, poiché rappresenta un tentativo estremo di preservare la più grande area di foresta tropicale del Centro America. L'evoluzione delle dinamiche di uso del suolo e la quantificazione della loro evoluzione attraverso un'analisi ed una modellizzazione spazio-temporale, dovrebbe permettere di evidenziare le aree vulnerabili al crescente flusso migratorio dei "senza terra", e valutare l'efficacia delle misure preventive adottate dalla RBM.

i.II. Obiettivi

L'obiettivo di questa tesi è lo sviluppo di una metodologia per analizzare le dinamiche di uso del suolo in un sistema antropizzato estremamente complesso e frammentato, utilizzando informazioni facilmente accessibili, senza grande dispendio di tempo e denaro.

L'enfasi di questo lavoro non si focalizza sulla complessità strutturale od informatica dei modelli utilizzati, ma ne verifica le possibili applicazioni pedagogiche ed esploratorie, per una migliore comprensione del fenomeno studiato e per visualizzare la coevoluzione del sistema uomo-ambiente.

Cercheremo di comprendere i limiti del processo di modellizzazione di un sistema complesso, revisionando le precedenti pubblicazioni nella letteratura modellistica, ed analizzando a scala locale, i risultati di differenti approcci predittivi.

In sintesi questa tesi cerca di rispondere alle seguenti domande: *è possibile modellizzare un sistema complesso uomo-natura? È possibile studiare l'evoluzione di dinamiche di uso del suolo frammentate e caotiche, con un database limitato? Qual è l'affidabilità di un modello che lavora in condizioni non ottimali? Un modello è più utile per comprendere o per prevedere?*

Obbiettivi specifici:

- ✓ Descrivere l'evoluzione dei patterns spaziali nella regione la Joyanca, Peten, in funzione degli agenti e delle variabili che operano a scala locale. Individuare un modello concettuale della dinamica paesaggistica lungo questo fronte pioniere, evidenziando i principali criteri ambientali che influenzano i cambi di uso del suolo.
- ✓ Costruire un database omogeneo e sufficiente ad alimentare i modelli predittivi, attraverso il processamento dei dati multispettrali. Integrare, in una rappresentazione spazialmente esplicita, questa informazione numerica digitale con l'informazione socio-culturale e fitogeografica. Dimostrare l'importanza dei SIG nella fase d'integrazione dei dati.
- ✓ Identificare, attraverso tecniche di analisi multitemporale di cambio, il ruolo esatto e l'importanza di ogni variabile biofisica e socio-economica nell'evoluzione dei patterns spaziali del sistema. Comprendere l'evoluzione quantitativa e qualitativa delle dinamiche di uso del suolo passate. Identificare la tecnica di analisi multitemporale più adatta al contesto geografico ed ai dati disponibili.
- ✓ Comprendere l'utilità della modellizzazione. Verificare l'importanza pedagogica e di apprendimento del modello.

- ✓ Sviluppare un modello freeware basato su reti neurali che in passato hanno dimostrato grandi potenzialità per catturare relazioni nascoste e non lineari in sistemi complessi. Affinare la parametrizzazione del modello e migliorarne le funzioni, in base ai primi risultati simulati.
- ✓ Confrontare le performance di quattro modelli predittivi alimentati con lo stesso input. Verificarne affidabilità e precisione attraverso una validazione esaustiva. Evidenziare i limiti e le qualità di ogni modello nella rappresentazione delle differenti dinamiche di uso del suolo.
- ✓ Offrire infine uno strumento flessibile e facilmente adattabile a nuovi contesti operativi e nuove problematiche. Creare un piccolo tutoriale per i modelli freeware.
- ✓ Verificare l'efficacia delle misure adottate dalla Riserva di Biosfera Maya per preservare la foresta tropicale. Dimostrare l'utilità di un modello come "laboratorio sperimentale" per visualizzare le conseguenze di decisioni e pianificazioni economiche, amministrative ed ambientali, nel breve termine.

i.III. Ipotesi

Questa tesi di dottorato verifica le seguenti ipotesi:

- o È possibile modellizzare un sistema antropizzato, attraverso delle semplificazioni che permettano di evidenziare le dinamiche e gli agenti, in funzione dei nostri obiettivi
- o È possibile ottenere risultati soddisfacenti anche lavorando in condizioni non ottimali, cioè studiando un fenomeno complesso e caotico, con poche informazioni.

- o Il telerilevamento rappresenta una risorsa fondamentale nello studio dello stato di salute della Terra e permette di sopperire all'assenza di dati in aree di difficile accesso o con condizioni ambientali estreme. È possibile attraverso il processamento d'immagine, ottenere informazioni utili sugli agenti e sulle variabili esplicative che influenzano le dinamiche del paesaggio, senza un grande dispendio di soldi e di tempo.
- o È possibile e necessario costruire ponti fra discipline differenti, per ottenere una rappresentazione esaustiva, nei limiti del nostro sapere cognitivo, delle dinamiche del sistema uomo-ambiente.

i.IV. Indice commentato

Il primo capitolo di questa tesi offre una descrizione generale sui cambi globali, focalizzando l'attenzione sulle principali cause ed agenti responsabili della degradazione della biosfera. Abbiamo poi analizzato l'importanza delle dinamiche di uso del suolo (LUCC) nell'economia dei cambiamenti globali e rimarcato la necessità di comprendere profondamente la coevoluzione del sistema uomo-ambiente.

Il secondo capitolo si sofferma sulla teoria modellistica e propone una revisione dei principali modelli di LUCC. Una loro successiva classificazione, in base a dei criteri che riteniamo significativi, permette di inquadrare il lavoro svolto all'interno di un determinato contesto metodologico.

Il terzo capitolo descrive in maniera approfondita i modelli predittivi sviluppati o riparametrizzati. Particolare attenzione è data al metodo PNNET basato sulle reti neurali, ed ai 4 programmi costruiti in Dinamica EGO, fondati sulla teoria Bayesiana dei pesi dell'evidenza. I modelli CA_Markov e LCM sono descritti più velocemente in quanto moduli di Idrisi, già ampiamente dettagliati nel suo tutoriale. Abbiamo inoltre incluso un piccolo manuale d'uso per PNNET e Dinamica EGO.

Nel quarto capitolo si presenta l'area di studio della Joyanca in Peten, Guatemala, e le problematiche affrontate. La regione fa parte della Riserva di Biosfera Maya, la più grande foresta tropicale primaria del Centro America, che ha conosciuto negli ultimi anni un forte tasso d'immigrazione e deforestazione, parzialmente incentivata dal governo locale. La costruzione del database geografico attraverso il processamento di immagini satellitari e le tecniche di analisi multitemporale dei cambi di uso del suolo, permettono di scegliere le variabili indipendenti che influenzano le dinamiche di deforestazione nella Joyanca e preparare, con l'aiuto dei SIG, l'input per i modelli predittivi.

Il quinto capitolo si sofferma sulla fase di validazione dei risultati, evidenziando la difficoltà incontrata nel separare i limiti dei modelli dalla complessità del fenomeno e dalla qualità dei dati. La validazione ha cercato di catturare differenti aspetti nelle performance dei modelli, sottolineando come la congruenza reale-simulato diminuisca rapidamente non appena le classi di copertura passano dallo stato di persistenza allo stato dinamico.

I NTRODUCTI ON

“Pensare che una delle conseguenze della conoscenza è la diminuzione del rischio, è un’illusione. Infatti, è impossibile riuscire a comprendere tutto e alla fine, il rischio non è mai eliminato; semplicemente prende forme diverse.” (Claude Raffestin, 2005 p.14)

INTRODUCTION

i.1. Contexte de la Recherche

La nécessité de comprendre les mécanismes qui règlent les changements globaux en cours et de pouvoir ainsi offrir un moyen de planification politico-économique future, représente aujourd'hui une priorité de la recherche scientifique. Cette dernière combine des observations, des analyses des processus, des expérimentations et des modélisations, pour améliorer notre capacité d'expliquer et de prévoir l'évolution du système homme-environnement.

La compréhension des changements globaux passe par la compréhension du rôle fondamental joué par l'activité anthropique dans la rapide transformation des écosystèmes et des paysages, qui contraste avec la lente variabilité naturelle de notre Planète.

Les macro-effets de l'empreinte humaine sur le fonctionnement du système terrestre, sont le résultat «inattendu» de l'action synergique de micro-processus à l'échelle locale, souvent non considérés ou insuffisamment étudiés (d'où leur caractère inattendu.) Les changements d'utilisation du sol sont donc des phénomènes complexes macro-échelle qui résultent de l'interaction simultanée d'agents indépendants, beaucoup moins complexes.

De nombreux projets internationaux ont été lancés depuis quelques années pour répondre au besoin croissant de comprendre les relations imbriquées entre la société et son milieu, et la manière dont celles-ci peuvent ou pourront modifier la capacité de la planète à accueillir la vie et à garantir notre bien-être. Le comportement exact du système homme-environnement est souvent imprévisible, étant donné son caractère émergent plutôt que prédéterminé.

Les mécanismes qui conduisent une société à dégrader son propre environnement sont encore peu connus et souvent décrits

de manière simpliste. L'utilisation du sol dérive de la combinaison synergique de plusieurs facteurs, ordonnés suivant différentes résolutions spatio-temporelles et niveaux d'organisation qui varient en fonction du contexte géographique et sociopolitique, selon l'expression générale suivante :

Utilisation du sol = f (pression, opportunité, politique, vulnérabilité, organisation sociale)

dont

pression = f (démographie, manque de ressources, pauvreté, infrastructures)

opportunité = f (lois du marché, coûts et profits, technologie)

politique = f (subsides, quotas, lois, impôts, plans de développement)

vulnérabilité = f (fragilité du système social et environnemental, dépendance de capital étranger)

organisation sociale = f (accès aux ressources, interaction milieu urbain-rural, niveaux de scolarité)

Le métabolisme industriel moderne a incrémenté les flux d'énergie et de matière résultant de l'extraction, de la transformation et de l'utilisation des ressources naturelles. Les effets plus évidents de l'impact anthropique sur la planète peuvent être visualisés à travers les processus de changement de la couverture terrestre et de l'utilisation du sol. Ces derniers incluent principalement, d'un point de vue économique et spatial, l'agriculture, l'élevage, la colonisation des forêts, l'expansion urbaine et l'extraction de géoressources.

Actuellement les terres « vierges » ou encore faiblement perturbées par l'homme, représentent moins de la moitié des terres émergées; dans toutes les régions du monde l'agriculture et l'élevage ont remplacé la végétation naturelle pour faire face à une augmentation vertigineuse des consommations, liée en partie à l'augmentation démographique et aussi à la diffusion du modèle de vie occidental. Dans les pays tropicaux, la situation est particulièrement grave, puisque les terres encore disponibles pour une expansion agricole sont limitées ; la forêt couvre une grande partie du

territoire et le sol est pauvre en nutriments. Les programmes de colonisation et les fronts pionniers qu'ils génèrent sont le résultat de décisions politiques qui attendent des bénéfices à court terme et cherchent à résoudre des tensions sociales internes ou à attirer des capitaux étrangers.

Les conséquences de ces processus de destruction de la forêt primaire tropicale sont diverses, et vont de la réduction de la biodiversité jusqu'à l'altération du climat global (émission de gaz à effet serre et modification de l'albédo terrestre), de la dégradation du sol au comportement anormal des régimes hydrologiques, jusqu'à l'épuisement des produits forestiers et alimentaires.

Dans ce contexte, la Réserve de Biosphère Maya (RBM) joue un rôle important, puisqu'elle représente l'ultime tentative de préserver la plus grande surface de forêt tropicale de l'Amérique Centrale. L'évolution des dynamiques d'utilisation du sol et la quantification de leur évolution à travers des analyses et des modélisations spatio-temporelles devrait permettre de mettre en évidence les aires vulnérables au flux migratoire croissant de « sans terre », et d'évaluer l'efficacité des mesures préventives adoptées par la RBM.

i.II. Objectifs

L'objectif de cette thèse est celui de développer une méthodologie d'analyse des dynamiques d'utilisation du sol dans un système anthropisé extrêmement complexe et fragmenté, en utilisant des informations facilement accessibles, pour un coût et dans un délai raisonnables.

Ce travail n'est pas centré sur la complexité structurelle ou informatique des modèles utilisés, mais sur la vérification des applications possibles pour la pédagogie et l'exploration, afin de mieux comprendre le phénomène étudié et de mieux visualiser les évolutions du système société-milieu.

Nous chercherons également à cerner les limites de la démarche de modélisation d'un système complexe, à partir d'une révision de la littérature concernée et en analysant, à l'échelle locale, les résultats de différentes approches prédictives.

En synthèse, cette thèse devra répondre aux questions suivantes : *Est-il possible de modéliser un système complexe homme-nature ? Est-il possible d'étudier l'évolution des dynamiques d'utilisation du sol fragmentées et chaotiques, avec des bases de données limitées ? Quelle est la fiabilité d'un modèle mis en œuvre en conditions non-optimales ? Un modèle est-il plus utile pour comprendre ou pour prévoir ?*

Objectifs spécifiques:

- ✓ Décrire l'évolution des patterns spatiaux dans la région de la *Joyanca*, Petén, en fonction des agents et des variables qui agissent à l'échelle locale. Déterminer un modèle conceptuel de la dynamique paysagère au long de ce front pionnier, en mettant en évidence les principaux critères environnementaux qui influencent les changements de l'utilisation du sol.
- ✓ Construire une base de données homogène et suffisante pour alimenter les modèles prédictifs, à travers le traitement des données multi-spectrales. Intégrer, dans une représentation spatialement explicite, cette information numérique digitale avec les informations socioculturelles et phytogéographiques. Souligner l'importance des fonctionnalités SIG dans la phase d'intégration des données.
- ✓ Identifier, à travers des techniques d'analyse multi-temporelles des changements, le rôle exact et l'importance de chaque variable biophysique et socio-économique dans l'évolution des patterns spatiaux du

système. Comprendre l'évolution quantitative et qualitative des dynamiques d'utilisation du sol passées. Identifier la technique d'analyse multi-temporelle la mieux adaptée au contexte géographique et aux données disponibles.

- ✓ Comprendre l'utilité de la modélisation. Vérifier l'importance du rôle pédagogique et d'apprentissage du modèle.
- ✓ Développer un modèle freeware basé sur des réseaux neuronaux, qui ont déjà montré leurs grandes potentialités de mise en évidence des relations cachées et non-linéaires en systèmes complexes. Affiner les paramètres du modèle et améliorer les fonctions, sur la base des premiers résultats simulés.
- ✓ Confronter les performances des quatre modèles prédictifs alimentés avec le mêmes variables explicatives. Vérifier la fiabilité et la précision à travers une validation exhaustive. Mettre en évidence les limites et les qualités de chaque modèle dans les représentations des différentes dynamiques d'utilisation du sol.
- ✓ Offrir enfin un moyen flexible et facilement adaptable à de nouveaux contextes opérationnels et nouvelles problématiques. Créer un petit « tutorial » pour les outils freeware développés.
- ✓ Vérifier l'efficacité des mesures adoptées de la Réserve de Biosphère Maya pour préserver la forêt tropicale. Montrer l'utilité d'un modèle comme « laboratoire expérimental » pour visualiser les conséquences des décisions et de la planification économique, administrative et environnementale, à court terme.

i.III. Hypothèses

Cette thèse de doctorat vérifie les hypothèses suivantes:

- o Il est possible de modéliser un système anthropisé, à travers des simplifications qui permettent de mettre en évidence les dynamiques et les agents, en fonction de nos objectifs.
- o Il est possible d'obtenir des résultats satisfaisants même en travaillant dans de conditions non-optimales, c'est-à-dire en étudiant un phénomène complexe et chaotique, avec peu d'informations.
- o La télédétection représente une ressource fondamentale dans l'étude de l'état de santé de la Terre et permet de suppléer à l'absence de données en régions d'accès difficile ou en conditions environnementales extrêmes. Le traitement d'image délivre à bon compte des informations utiles sur les agents et sur les variables explicatives qui influencent les dynamiques de paysage.
- o Il est possible et nécessaire de construire des ponts entre différentes disciplines, afin d'obtenir une représentation exhaustive, dans les limites de notre savoir cognitif, des dynamiques du système homme-environnement.

i.IV. Indice Commenté

Le premier chapitre de cette thèse offre une description générale sur les changements globaux, en focalisant l'attention sur les principales causes et les principaux agents responsables de la dégradation de la biosphère. Nous avons ensuite analysé l'importance des dynamiques d'utilisation du sol (LUCC) dans l'économie des changements globaux et souligné la nécessité d'une compréhension profonde des co-évolutions du système homme-environnement.

Le deuxième chapitre aborde d'abord la théorie modélisatrice et propose une révision des principaux modèles de LUCC. Leur classement, ensuite, sur la base de critères que nous retenons comme significatifs, permet d'encadrer le travail effectué à l'intérieur d'un contexte méthodologique déterminé.

Le troisième chapitre décrit de manière approfondie les modèles prédictifs développés ou re-paramétrés. Une attention particulière est apportée à la méthode PNNET basée sur les réseaux des neuraux, et aux quatre programmes construits en Dynamique EGO, fondés sur la théorie Bayésienne des poids d'évidence. Les modèles CA_Markov et LCM sont décrits plus rapidement en tant que modules d'Idrisi, déjà présentés de façon exhaustive par son propre « tutorial ». Nous avons aussi proposé un petit « tutorial » pour PNNET et pour Dinamica EGO.

Dans le quatrième chapitre, le terrain d'étude de *la Joyanca* au Petén (Guatemala) est présenté et les problématiques locales abordées. Cette région fait partie de la Réserve de Biosphère Maya, la plus grande forêt tropicale primaire de l'Amérique Centrale, qui a connu depuis quelques années un fort taux d'immigration et une déforestation rapide, partiellement stimulée par le gouvernement local. La construction des bases de données géographiques à travers le traitement d'images satellitaires et les techniques d'analyse multi-temporelles des changements d'utilisation du sol, permettent de choisir les variables indépendantes qui influencent les dynamiques de déforestation dans *la Joyanca* et de préparer, avec l'aide des SIG, les inputs pour les modèles prédictifs.

Enfin, le dernier chapitre comprend la phase de validation des résultats et rend compte de la difficulté à distinguer les limites propres aux modèles de celles imputables à la complexité du phénomène elle-même ou à la qualité des données.

La validation a tenté de mettre en relief les différents aspects que recouvrent les performances des modèles, en soulignant la façon dont la congruence réel-simulé diminue rapidement lorsque les classes de couverture du sol passent de l'état de stabilité à l'état dynamique.

CAPITOLO No. 1

L'importanza dello studio delle dinamiche di uso del suolo nell'economia dei cambiamenti globali

“ Il vecchio diceva guardando lontano, immagina questo coperto di grano, immagina i frutti, immagina i fiori e pensa alle voci e pensa ai colori. E in questa pianura fin dove si perde, crescevano gli alberi e tutto era verde, cadeva la pioggia, segnavano i soli il ritmo dell'uomo e delle stagioni...”

(Il vecchio ed il bambino, Guccini, 1972)

1.1 Uno sguardo rapido sui cambiamenti globali.

Il nostro pianeta sta rapidamente cambiando per cause antropiche e naturali. La capacità della Terra di reagire ad una perturbazione, attraverso un meccanismo di autoregolazione, è stata fortemente compromessa da decenni di insensato e rapido sviluppo della nostra società [Lovelock, 2006]. La necessità di comprendere i meccanismi che regolano i cambiamenti globali in atto e di offrire così uno strumento di pianificazione politico-economica futura, rappresenta oggi una priorità della ricerca scientifica. *"Nous ne pouvons pas être optimistes à long terme qu'à la condition que le pessimisme contribue, à court terme, à changer le monde d'aujourd'hui"*¹ [Lambin, 2004, Pag.26].

Numerosi progetti sono stati avviati negli ultimi decenni per rispondere a questa necessità impellente; i primi passi mossi negli anni '80 portano alla creazione del Earth System Science Concept (ESSC, 1986), un modello integrato dell'intero sistema uomo-natura che aveva come obiettivo l'analisi e la proiezione futura dei cambiamenti climatici in atto [Steyaert, 1993]. Successivamente, grazie ad una progressiva diffusione di una coscienza ambientalista e ad un riconoscimento del mondo scientifico e politico della necessità di intervenire in tempi brevi, numerosi altri progetti di ricerca internazionali sono nati dall'integrazione di differenti discipline, per offrire una risposta completa, nei limiti delle nostre possibilità cognitive, alle pressanti problematiche globali. Fra questi ricordiamo il Global Land Project (GLP) (Fig.1-1) cofinanziato dal International Geosphere-Biosphere Programme (IGBP) e dal International Human Dimensions Programme on Global Environmental Change (IHDP), il programma DIVERSITAS (su biodiversità), il programma Land Use Cover Change (LUCC) sui

¹ Possiam guardare con ottimismo ad un futuro remoto solo se il pessimismo attuale riuscirà, in breve, a cambiare il mondo in cui viviamo.

cambi di uso del suolo ed infine il programma sui cambiamenti climatici World Climate Research Programme (WCRP).

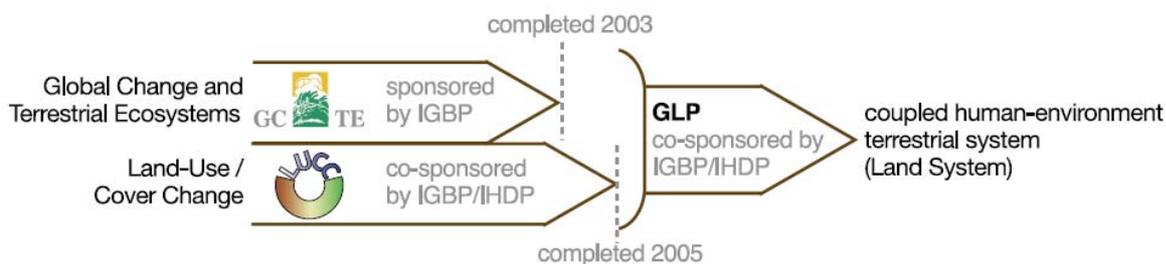


Fig 1-1: Creazione del Global Land Programme per lo studio dei cambiamenti nel sistema uomo-ambiente. Il progetto riprende ed integra il lavoro di ricerca svolto da precedenti programmi sui cambiamenti globali. [Fonte immagine: IGBP e IHDP, 2005].

Studiare il sistema uomo-ambiente significa comprendere l'influenza della natura sullo sviluppo della società (adattamento dell'uomo all'ambiente in cui vive) e l'impatto delle attività antropiche sulla natura (trasformazione dell'ambiente in cui vive).

Il dibattito ambientale attuale si focalizza su tre problematiche distinte: la dimensione dei cambiamenti in atto sul nostro pianeta, lo studio delle relazioni causa-effetto e la valutazione della nostra vulnerabilità di fronte alla possibile risposta della Terra alle perturbazioni [Lovelock, 2004]. La prima questione si basa sulla analisi dei numerosi dati oggi disponibili grazie ai nuovi sistemi di informazione, in particolare grazie alle immagini da satellite che permettono l'osservazione di un processo su differenti scale spazio-temporali. La seconda questione cerca di evidenziare le relazioni, spesso complesse e non facilmente rappresentabili attraverso un linguaggio scientifico, fra fattori antropici e fattori naturali, nell'economia di un cambiamento. I meccanismi che portano una società a degradare il proprio ambiente sono poco noti o spesso descritti in maniera semplicistica. La ultima questione è probabilmente la più dibattuta; essa si muove attraverso differenti visioni del rapporto fra uomo e

natura, da un ottimismo estremo che ripone illimitata fiducia nel "genio" umano e nella sua capacità di trovare una risposta a tutto, fino all'estremo pessimismo, che vede avvicinarsi il giorno del giudizio. I dati a supporto di una o dell'altra teoria sono spesso aneddotici, basati su osservazioni locali isolate da un contesto globale o spalmate su un intervallo temporale troppo breve. La validità di un'osservazione è sempre associata ad una rappresentatività spaziale e temporale del fenomeno [Baudelle et al., 2004]. È necessario muoversi attraverso un vasto spettro di discipline e livelli di dettaglio per dare una risposta completa e accettata dal mondo scientifico al problema dei cambiamenti globali, che metta i politici, i decisionisti e la società in generale di fronte alle proprie responsabilità. La previsione delle conseguenze ambientali del nostro attuale modello di sviluppo è fondamentale per pianificare una eventuale e ormai necessaria correzione di rotta in maniera anticipata, vista l'inerzia dei sistemi naturali e sociali e la crescente complessità dell'economia mondiale.

La nostra capacità di intervenire e modificare la natura è notevolmente aumentata con l'inizio dell'era industriale e la diffusione del lavoro meccanizzato, ma è soprattutto nel XX secolo che il nostro pianeta è stato irrimediabilmente segnato dall'impronta dell'uomo. Numerosi sono i problemi ambientali che si possono evidenziare a differenti risoluzioni spaziali (locale, regionale o globale) e temporali (eventi progressivi, ciclici o intermittenti, lenti o rapidi), che affettano differenti contesti geografici e sociali.

[Lambin, 2004] propone una classificazione delle principali problematiche ambientali in 5 gruppi: la scarsità di risorse in terra e la loro degradazione a fronte di un'espansione agricola per soddisfare la richiesta di prodotti alimentari e forestali; l'utilizzo del pianeta come discarica illimitata a fronte di un consumismo diffuso e caratteristico dello stile di vita

occidentale; l'inefficienza o incapacità nel gestire delle risorse planetarie comuni all'insieme dell'umanità, come oceani ed atmosfera, in maniera sostenibile, privilegiando gli interessi economici locali; la degradazione delle risorse biologiche e genetiche; l'utilizzo irrazionale di risorse non rinnovabili come il petrolio ed in certa maniera anche l'acqua [Shiva, 2003].

Alcuni cambiamenti ambientali possono avere ripercussioni sull'intero equilibrio mondiale, poiché interagiscono con un sistema che opera su grande scala. Ad esempio, lo scioglimento previsto dei ghiacciai della Groenlandia a fronte di un aumento di 3°C nella temperatura terrestre, libererà grandi quantitativi di gas metano intrappolato nel ghiaccio; il metano è un gas serra molto più forte del monossido di carbonio e favorirà dunque un ulteriore aumento della temperatura del pianeta. Altri cambi possono realizzarsi indipendentemente, a livello locale o regionale, ma in maniera sincronica o sinergica; è dunque necessario comprendere quali siano gli effetti cumulativi dei micro-processi sull'equilibrio dell'intero sistema naturale. Il modello MOSDEL (Model for Sustainable Development of Land use) sviluppato all'interno del programma WAVES (Water Availability and Vulnerability of Ecosystems and Society) è un esempio pratico di come integrare informazioni locali nell'analisi dei cambiamenti globali [Gaiser et al., 2003].

I dati che si riferiscono all'intero sistema terrestre, data l'ampiezza di indagine e le caratteristiche degli strumenti utilizzati (ad esempio i satelliti meteorologici con risoluzione molto grande), possono mascherare, generalizzandole, tendenze divergenti in differenti contesti geografici e storici. Nel periodo 2000-2005, ad esempio, il nostro pianeta ha registrato una perdita netta di foreste pari ad 7 milioni di ettari. Per meglio comprendere la dinamica del fenomeno dobbiamo integrare questo "inventario" globale con

un'analisi dei processi regionali su un intervallo temporale più esteso. Possiamo così evidenziare comportamenti spazio-temporali divergenti: in Europa questa tendenza si è da tempo invertita ed esiste un tasso incrementale di riforestazione legato all'abbandono delle attività agricole. L'Asia è passata da una perdita netta di circa 800.000 ettari l'anno degli anni '90 ad un recupero annuo di un milione di ettari nel periodo compreso tra il 2000 ed il 2005, principalmente come conseguenza dell'afforestazione su larga scala in Cina. L'America Latina e l'Africa hanno registrato la più ampia perdita netta annuale di foreste, rispettivamente 4,3 e 4 milioni di ettari, seguite dall'Oceania, 380000 ettari, e dal Nord e Centro America con 330000 ettari ciascuna [FAO, 2005].

Attualmente le terre selvagge, lievemente perturbate dall'uomo, rappresentano meno della metà delle terre emerse; in tutte le aree del mondo l'agricoltura ha rimpiazzato la vegetazione naturale per far fronte ad un vertiginoso aumento dei consumi legato in parte alla crescita demografica, in parte alla diffusione dello stile di vita occidentale. Nei paesi tropicali la situazione è più grave, poiché le terre ancora disponibili per un'espansione agricola sono limitate; la foresta copre gran parte del territorio ed il suolo è povero in nutrienti. I piani di sviluppo agricolo lungo il fronte pioniere di deforestazione sono il risultato di decisioni amministrative e politiche che valutano solo i benefici a breve termine, senza ponderare minimamente i costi ambientali in termini di biodiversità persa e degradazione del suolo. Nel XX secolo la superficie coltivata nel mondo è aumentata del 50%, principalmente nelle regioni tropicali; contemporaneamente si è perso un quarto delle terre fertili a causa dell'erosione, della salinizzazione e dello sfruttamento eccessivo del suolo [Berlan, 2004].

La degradazione delle georisorse è legata all'evoluzione socio-economica che, in un certo istante e luogo, spinge un

gruppo sociale verso un tentativo di appropriazione di beni ambientali. Spesso, alla base di questi processi, esistono alleanze fra chi governa e un'élite economica che ottiene un accesso privilegiato alle risorse naturali, a scapito di una moltitudine che rimane emarginata o spinta ad emigrare verso aree meno produttive. Questi cambiamenti sono il risultato di decisioni prese da amministratori interessati a ottenere benefici economici sul breve periodo. Essi ragionano solo in una dimensione locale, percepiscono cioè solamente i pericoli ed i danni che si manifestano nell'ambiente in cui operano, senza preoccuparsi degli effetti che le loro azioni possono avere, per sinergia con altri micro-processi, sull'equilibrio del sistema naturale terrestre (visione dal locale al globale). A livello mondiale le istituzioni internazionali possono proporre riforme che si scontrano con gli interessi economici di un paese (gli Stati Uniti si oppongono energicamente all'applicazione del trattato di Kyoto²), oppure stabilire regole (ad esempio regole di mercato) che finiscono inevitabilmente per avere un impatto all'interno della società, anche a piccola scala (visione dal globale al locale). Comprendere appieno un fenomeno complesso e i meccanismi che lo regolano può essere oneroso e a volte impossibile, vista la difficoltà che abbiamo a rappresentare le relazioni, spesso nascoste, fra fattori a differenti livelli organizzativi e in un contesto storico, geografico, economico e sociale spesso poco chiaro. *"... biophysical alterations and social forces generate different responses in the Northern Hemisphere than in the Southern one, in urban environments than in rural ones, and in developed countries than in developing countries"*³ [GLP, 2005, Pag.1]. Questa complessità deve dunque essere

2 "La nostra economia sta rallentando. Stiamo anche attraversando una crisi energetica e l'idea di mettere un coperchio al biossido di carbonio è economicamente insensata" [George W. Bush, 2001].

3 ...le alterazioni biofisiche e le forze sociali generano risposte differenti nell'emisfero nord rispetto all'emisfero sud, nelle città rispetto alle campagne, e nei paesi sviluppati rispetto ai paesi in via di sviluppo.

semplificata riconoscendo che, nella maggior parte dei casi, un numero limitato di relazioni note fra variabili esplicative dà origine ad un cambiamento in un sistema uomo-ambiente. Dobbiamo inoltre distinguere i fattori che agiscono in continuo, corrodendo poco a poco la capacità di reazione di un sistema ed indebolendolo progressivamente, dagli eventi improvvisi e bruschi, caratterizzati da una forte intensità temporale, un grande potere di dispersione ed una notevole magnitudine [Bandelle et al., 2004]. Queste perturbazioni, rapide e violente, possono affettare sistemi fragili e già indeboliti, e produrre cambiamenti rilevanti spesso non prevedibili.

[Lambin, 2004] propone un raggruppamento di questi fattori in 5 grandi "sindromi" o famiglie di cause che associandosi hanno un impatto rilevante sull'ambiente.

La prima sindrome è legata all'impoverimento delle georisorse dovuto ad un utilizzo intensivo ed irrazionale delle stesse (miniere, suolo agricolo, etc.). Numerose possono essere le cause che spingono in questa direzione, dall'aumento demografico all'applicazione di nuove tecnologie e pratiche produttive inadatte a ecosistemi fragili [Berlan, 2004], da una disuguaglianza sociale nella ripartizione delle risorse alla scarsità di quest'ultime (la grande ondata di migrazione verso i fronti pionieri in Amazzonia dei "senza terra" contrasta con l'abbondanza di risorse dei grandi *fazenderos*). L'impoverimento delle risorse è seguito da un aumento della pressione sulle restanti, chiudendo così un circolo vizioso.

Il secondo gruppo di variabili esplicative descrive l'aumento della vulnerabilità di una società o un sistema di fronte ad una perturbazione intensa, e la sua limitata capacità di recupero da un trauma. Queste caratteristiche sono legate ad un impoverimento economico o alla marginalizzazione sociale e politica di un gruppo di individui o di un paese. *"Saranno le popolazioni più povere del Terzo Mondo quelle più colpite dal*

cambiamento climatico, dalla siccità, dallo scioglimento dei ghiacciai, dall'esaurimento delle risorse. I contadini, i pastori, le comunità indigene diventeranno profughi ambientali in seguito alla scomparsa delle piogge, al crollo dei raccolti ed all'inaridimento dei fiumi" [Shiva, 2003, Pag.63]. La dipendenza economica e il forte debito accumulato da alcuni stati africani nei confronti di paesi terzi, portano ad una gestione catastrofica, ma redditizia nel breve periodo, delle risorse. I paesi creditori a loro volta, hanno un forte interesse nel mantenere questa condizione di dipendenza e instabilità politica per poter accedere alle grandi ricchezze dei territori africani senza doversi preoccupare dei danni ambientali che provocano in situ. Anche il nostro pianeta ha perso progressivamente la propria capacità di reagire ai cambiamenti, indebolendosi sempre più sotto il peso delle nostre azioni [Lovelock, 2006].

La terza sindrome è legata allo sviluppo economico dei paesi emergenti; in queste regioni l'impatto ambientale dovuto all'utilizzo intensivo delle terre e al progresso industriale, sono noti e ben visibili, ma considerati come secondari se rapportati ai benefici economici. *"...l'industrializzazione era vista come un simbolo di progresso, i tribunali, come dimostrano i numerosi casi giudiziari, tendevano ad appoggiare il diritto delle industrie ad inquinare l'acqua"* [Shiva, 2003, Pag.45]. La Cina rappresenta l'emblema di questa categoria: l'uso abbondante di carbone nelle centrali, la canalizzazione dei corsi d'acqua, l'assenza di regole ambientali minime, che attirano numerosi investimenti stranieri e l'esodo di massa dal mondo rurale verso i grandi centri abitati, hanno notevolmente compromesso gli equilibri interni del sistema naturale. La portata dei cambiamenti ha fatto registrare forti ripercussioni sull'ecosistema locale (scomparsa di molte specie animali autoctone, siccità dei principali fiumi spesso utilizzati come discariche, gli esodi forzati, malattie respiratorie fra la

popolazione etc.) e sull'intero sistema globale (le aree industriali cinesi sono fra i principali produttori di CO₂). Nel 2004 i costi ambientali e umani, valutati solo a livello nazionale, sono esorbitanti: 50 miliardi di euro di perdite economiche legate all'inquinamento di aria ed acqua, più di un milione di malati cronici di bronchite e 118000 decessi per cancro nelle zone rurali [Courier International, 2007].

Il quarto gruppo è legato ai cambiamenti nel rapporto fra l'uomo e la natura come conseguenza di un'evoluzione strutturale, culturale o istituzionale della società. La grande concentrazione di persone nelle metropoli asiatiche e americane, l'abbandono delle attività agricole per il lavoro industriale, sono esempi pratici di questi cambiamenti.

Infine la quinta sindrome è legata ai cambiamenti dovuti a decisioni politico-amministrative che focalizzano la loro attenzione su benefici sociali od economici, trascurando i costi ambientali della pianificazione sul lungo periodo. Il più delle volte le macro-decisioni, prese in campo internazionale, si basano su una conoscenza approssimativa delle realtà interessate e raramente amministratori o ricercatori locali sono interpellati durante la loro discussione [Park, 1993]. Il risultato è una politica di azione inefficace e distante dai problemi reali di una comunità o di un sotto-sistema naturale.

Non è facile distinguere con sicurezza i confini di ogni gruppo; il più delle volte un cambiamento è il risultato di un'azione sinergica di fattori indipendenti, appartenenti a differenti categorie. L'evoluzione economica di un paese, ad esempio, può determinare un deterioramento delle risorse o una marginalizzazione sociale; come conseguenza nuove regole istituzionali possono essere attuate per correggere lo sviluppo e la gestione dell'ambiente, etc.

Quando l'impatto ambientale di un'attività diventa significativo, delle retroazioni (*feedback*) possono condurre ad

una modificazione dei meccanismi che regolano il cambiamento. Queste retroazioni possono essere positive (favoriscono i cambi) o negative (si oppongono ai cambi), alterando la domanda di materie prime, la vulnerabilità di una società, etc. Ad esempio, i danni provocati da fenomeni naturali catastrofici (uragani, piogge torrenziali, etc.) legati alle anomalie climatiche, hanno reso la nostra società più sensibile all'idea di sviluppo sostenibile e portato alla sottoscrizione del trattato di Kyoto (feedback negativo). Parallelamente le stesse anomalie nel ciclo idrologico e climatico hanno pregiudicato la produzione agricola di vaste regioni, determinando un aumento della vulnerabilità della società che vi abita e della pressione sulle restanti risorse naturali.

Una profonda comprensione di come l'uomo stia cambiando il sistema terrestre, di quali siano le forze motrici e le dinamiche del cambiamento, e le loro conseguenze, ci potrà aiutare a quantificare e limitare i rischi che possono derivare da una perdita delle condizioni ideali di abitabilità del pianeta. L'uomo ricava una vasta gamma di prodotti dall'ambiente in cui vive e così facendo lo modifica; il nostro impatto sull'ecosistema deve essere ben compreso e reso sostenibile, poiché strettamente legato alla sicurezza alimentare della nostra società [Gregory et al., 2000]. *"Consumimos tanto que a Terra já não consegue sustentar o mundo familiar e confortável a que nos habituamos. Agora ela está mudando, de acordo com suas próprias regras internas, para um estado em que já não somos mais bem-vindos"*⁴ [Lovelock, 2006, Pag.20].

Il sistema naturale non è un sistema lineare; esiste una soglia, un limite (Fig.1-2) oltre il quale la Terra non potrà

⁴ Consumiamo così tanto che la Terra non riesce più a mantenere quel mondo familiare e confortevole a cui ci siamo abituati. Ora sta cambiando, d'accordo con le proprie regole interne, verso uno stato in cui non saremmo più i benvenuti.

più assorbire l'impatto antropico. Capire quale sia questo "sustainable limit" e quando lo raggiungeremo, è la priorità di tutti i programmi di ricerca scientifica che studiano i cambiamenti globali. È necessario ridiscutere il rapporto con la natura in cui viviamo unendo tutti i ponti che esistono fra le diverse discipline scientifiche ed gli ambienti economico-politici della società. L'International Panel of Climate Change (IPCC) individua nella concentrazione di 500 ppm di CO₂ nell'aria, la soglia oltre la quale, un aumento della temperatura media di 3°C, innescherà una serie di meccanismi di autoregolazione a catena nell'intero sistema terrestre, che finiranno inevitabilmente per avere forti ripercussioni sul nostro sistema economico, sociale e culturale.

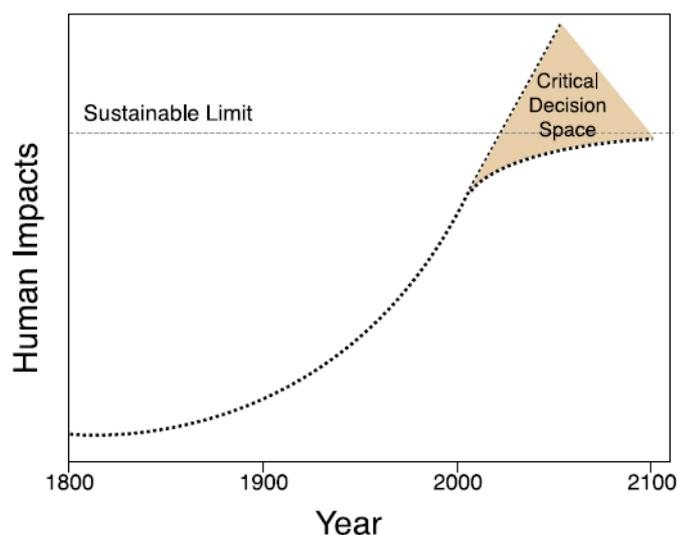


Fig 1-2: Rappresentazione concettuale della relazione fra impatto umano globale cumulativo (aumento demografico, uso delle risorse, perdita di biodiversità, etc.) e limite di sostenibilità del sistema terrestre. Oltrepassare questa soglia significherà mettere in moto una serie di meccanismi e autoregolazioni della Terra ancora poco note, che avranno pesanti ripercussioni sul nostro sistema sociale ed economico. [Fonte immagine: IGBP e IHDP, 2005].

" *Suspeitamos da existência de um limite, fixado pela temperatura ou pelo nível de CO₂ no ar. Uma vez ultrapassado, nada que as nações do mundo façam alterará o resultado... precisamos levar a sério a mudança global, e imediatamente, e*

depois fazer o possível para reduzir as pegadas dos seres humanos sobre a Terra"⁵ [Lovelock, 2006, Pag.19 e 24].

1.2 Dinamiche globali e cambiamenti dell'uso del suolo (LUCC)

1.2.1 Premessa epistemologica

Il termine LUCC (Land Use and Cover Change⁶) verrà adottato, in questa tesi, per indicare tutti i cambiamenti legati all'uso del suolo e dei sistemi terrestri. *Land Use* definisce un'attività umana, una forma di gestione di un contesto geografico, sociale ed economico; *Land Cover* definisce un cambio fisico della copertura terrestre, legato o meno ad una causa antropica diretta [Brown et al., 2004]. "*Understanding the dynamics of land system changes requires attention to land cover (biophysical conditions) and land use (human uses) as a coupled human-environment system*"⁷ [Rindfuss et al., 2004, Pag.13976]. Questa scelta epistemologica non è totalmente condivisa in ambito scientifico; in [Koomen et al., 2007] si preferisce adottare, per semplicità, *Land Use* per riferirsi sia ai cambiamenti di copertura della terra sia alla maniera in cui essa è utilizzata, pur riconoscendo l'utilità di separare questi due termini nel caso in cui una stessa area (foresta, ad es.) sia adibita a differenti funzioni (taglio selettivo, produzione agricola e parco naturale, ad es.).

In botanica la parola *copertura* indica "*la superficie di un territorio coperta da individui della stessa specie vegetale*" [Gabrielli, 2007, Pag.640]. Generalizzando questa definizione, considereremo la copertura come l'insieme degli attributi della

5 Sospettiamo che esista un limite massimo, in termini di temperatura o concentrazione di CO2 nell'aria. Una volta oltrepassato, nessuna azione dei governi altererà il risultato... dobbiamo considerare seriamente ed immediatamente i cambiamenti globali e fare il massimo possibile per ridurre l'impatto delle nostre attività sulla Terra.

6 Uso della terra e cambiamento della copertura.

7 Per capire le dinamiche dei cambiamenti del sistema terrestre bisogna prestare attenzione a *land cover* (attributi biofisici) ed a *land use* (uso umano), in un sistema unificato uomo-natura.

superficie terrestri e dell'immediato sottosuolo, cioè la parte del nostro pianeta in cui si concentra l'attività della vita.

1.2.2 Le dinamiche di uso del suolo: LUCC

*"The Earth System refers to the global-scale couplet socio-environmental system; and the terrestrial component of the Earth System - the land system - stands at the centre of understanding the relationship between human and their environment"*⁸ [GLP, 2005, Pag.3].

I LUCC rappresentano probabilmente la manifestazione più evidente dei cambiamenti globali, e sono anzitutto causati da una crescente domanda di prodotti agricoli e forestali [Gregory et al, 2000]. L'agricoltura si è espansa nei territori coperti da foresta tropicale, nelle savane, nel cerrado e nelle steppe per rispondere ai bisogni alimentari di una popolazione in crescita ed ad un consumismo occidentale ampiamente diffuso. I territori destinati alle coltivazioni sono aumentati di un 50% durante il XX secolo ed i pascoli sono più che raddoppiati negli ultimi 200 anni. In America Latina i grandi fronti di deforestazione creano nuovi spazi alle monoculture (soia in Mato Grosso, ad es.); esistono, contemporaneamente, molti esempi di piccoli agricoltori concentrati in aree marginali o lungo i grandi vettori di penetrazione [Passos, 2007] nelle regioni ancora selvagge. L'impatto ambientale di questi ultimi è poco evidente se analizzato su grande scala, ma gli effetti cumulativi e sinergici di questi micro-processi non devono essere trascurati nella valutazione dello stato di degradazione del suolo e dei cambiamenti di copertura. Il nostro lavoro di ricerca, qui descritto, parte da queste considerazioni generali, e si pone come obiettivo la rappresentazione e la

⁸ La Terra è un sistema complesso uomo-ambiente su scala globale; la componente terrestre del pianeta - il sistema terrestre - gioca un ruolo fondamentale nella comprensione delle relazioni fra l'uomo e l'ambiente in cui vive.

previsione dell'evoluzione dei patterns spaziali in una piccola regione del Peten, analizzando, su scala locale, le relazioni fra variabili esplicative, sociali, economiche e fisiche, e le dinamiche di uso del suolo osservate negli ultimi decenni.

I primi studi sui LUCC cominciarono negli anni '70 a seguito del crescente interesse verso i cambiamenti climatici e le loro relazioni con le attività umane. Solo successivamente si capì che i cambiamenti globali oltrepassavano la semplice nozione di clima, interessando molteplici aspetti dell'impatto antropico sul nostro pianeta, dalla perdita di biodiversità all'impoverimento del suolo, dalla gestione delle risorse naturali alla variazione della composizione atmosferica, e potevano avere forti ripercussioni sul piano economico, sociale ed alimentare.

La comprensione dei LUCC è passata da un approccio iniziale troppo semplicistico ad una analisi approfondita, spesso dispendiosa in termini temporali e di lavoro, della grande complessità nelle relazioni uomo-natura, basata su una ricostruzione storica che sia rappresentativa dei processi analizzati e su una multiscalarità che catturi la frammentarietà del paesaggio e delle dinamiche locali e globali.

La complessità dei LUCC deriva dalla combinazione di più fattori, che vanno dalla difficoltà di separare impatto antropico da variabilità naturale, ai complessi parametri che regolano la presa di decisioni a differenti livelli organizzativi, dalla necessità di evidenziare dei cambi su un ampio ventaglio di scale spazio-temporali, alla scarsa conoscenza e documentazione dei meccanismi di autoregolazione dell'ambiente.

La conversione di una classe di copertura (foresta, ad es.) in un'altra classe (suolo esposto, ad es.) è facilmente individuabile, poiché determina un cambio di identità e degli

attributi dell'oggetto di studio. Il monitoraggio di questo tipo di LUCC può essere facilmente realizzato attraverso uno studio multi-temporale di immagini satellitari, opportunamente scelte per catturare la frequenza spaziale e temporale dei processi (Fig.1-3)

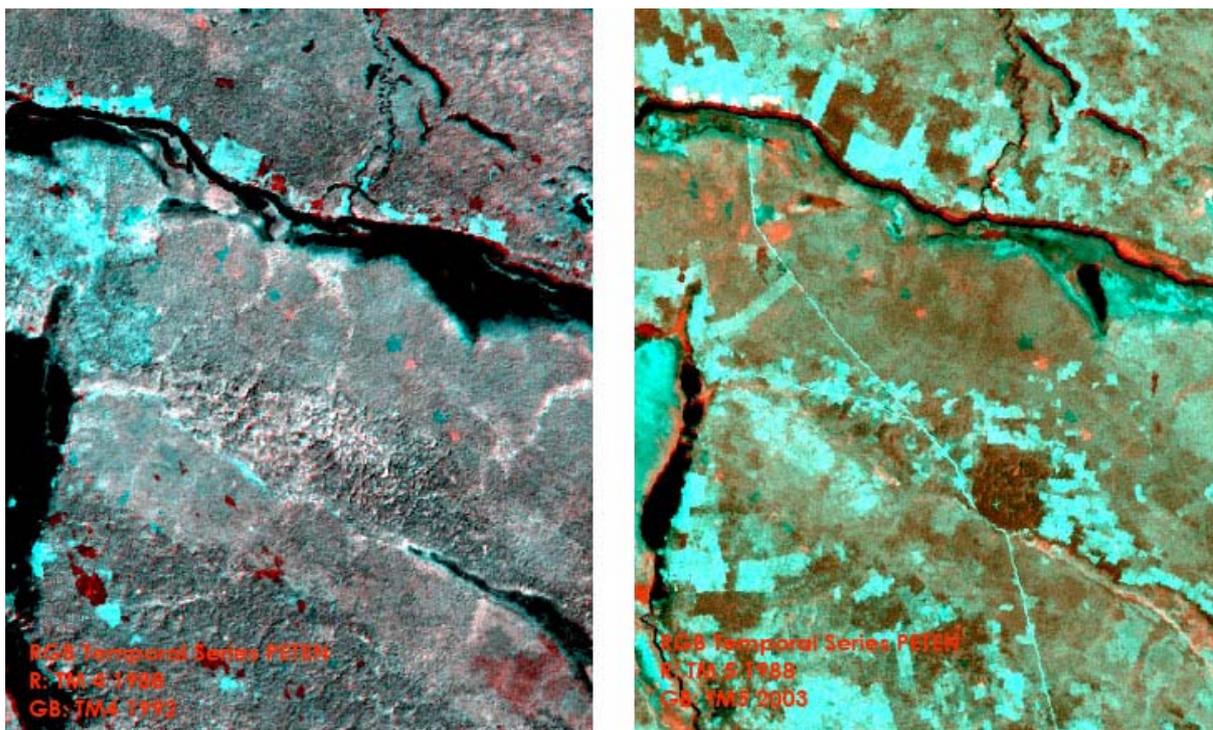


Fig 1-3: Studio multi-temporale di immagini satellitari TM Landsat (P20-R48). In azzurro i cambiamenti verificatisi nella regione La Joyanca, Peten, nell'intervallo fra le due date considerate. A sinistra: R(TM4,05/11/1988), GB(TM4,24/11/1992). A destra: R(TM5,05/11/1988), GB(ETM5,07/05/2003). Alcuni cambi evidenziati sono dovuti alle attività umane (in genere le forme geometriche caratterizzano i perimetri di deforestazione per la milpa o il pascolo) altri sono legati a fenomeni naturali (una alluvione che segue ad un periodo di pioggia intenso, ad es.). [Fonte: Geode/CNRS; Follador].

Più complessa risulta l'analisi e l'individuazione dei piccoli cambi che modificano solo alcuni attributi di una copertura terrestre; essi possono essere mal interpretati, non percepiti appieno o semplicemente ignorati. I loro effetti sul sistema naturale non sono spesso evidenziabili attraverso una discretizzazione spazio-temporale; questi micro-cambiamenti necessitano cioè di uno studio continuo per comprendere come affettino il comportamento e le caratteristiche del sistema

[Lambin et al., 2003], poiché continua è la variabilità degli attributi della superficie nello spazio e nel tempo. Mentre le conversioni della copertura terrestre sono ben documentate, poche informazioni, principalmente qualitative, esistono per descrivere le modificazioni minori degli attributi, spesso influenzate da variazioni climatiche o da caratteristiche intrinseche al sistema (stato di degradazione o di salute della vegetazione, ad es.), piuttosto che da attività antropiche dirette.

Il contesto socio-culturale ed economico di una regione può variare nel tempo e nello spazio, influenzando il processo decisionale a livello locale, regionale o globale. A piccola scala, in una comunità, la formazione culturale, di un amministratore o di un politico, la sua percezione della realtà e del rapporto uomo-ambiente, le credenze religiose ed i valori morali finiscono spesso per influire maggiormente sulla presa di decisioni rispetto agli interessi economici. Questi micro-fattori si combinano con variabili che operano a livelli organizzativi superiori (cambi nei patterns di consumo, politiche statali inefficienti e dannose sul lungo periodo, che offrono sussidi atti a favorire la migrazione dei ceti sociali più poveri in aree marginali, l'antropizzazione di terre ancora selvagge o la costruzione di strade attraverso aree isolate) che ne amplificano la portata e l'impatto sull'ecosistema locale. Presentiamo, a titolo esemplificativo, un estratto dell'intervista al sindaco di Itacarambi, Minas Gerais, Brasile, realizzata all'interno di un lavoro di ricerca sui cambi paesaggistici nel bacino del Rio Sao Francisco,

"Eu mesmo desmatei, eu desmatei um monte de trem aqui minha filha, você nem imagina o tanto que eu desmatei...Olha a incoerência, há 35 anos atrás eu peguei dinheiro emprestado do Banco do Brasil, que é do governo, juros super subsidiado, porque se não fosse isso eu não vinha pra cá, eu vim porque tinha essa facilidade. Então me deram dinheiro pra comprar o

trator, pra desmatar, eu aluguei dez tratores de esteira, cada um com um correntão, precisava de uma caminhão só pra carregar a corrente pra você ter idéia do tamanho e do peso, ai amarrava uma corrente de ca e uma de la, dois tratores puxando a corrente, aquilo ali oh, o que tivesse na frente ai saia tudo. E ta ai ainda ta ai no Brasil. Agora eles ficam querendo que a gente não tire nem um pedaço de pau, mas eles tiraram, eles tiraram, arrumaram a sua vida estruturaram sua vida. Eu acho que a amazônia tem que ser cuidada mesmo, mas acontece que se nós não tirarmos a madeira agora, vai virar deserto, tão ai os cientistas, o Al Gore ta chegando ai..."⁹
 [Renno, 2007, non pubblicato].

Le cause dirette di un cambiamento implicano un'azione fisica sulla copertura terrestre, ad opera di un individuo o di una comunità. Le cause indirette rappresentano l'insieme di regole economiche, sociali, demografiche, biofisiche, tecnologiche, che influenzano il rapporto fra l'uomo ed il proprio ambiente. Generalmente le cause indirette sfuggono al controllo di una comunità locale e possono essere viste come macro-variabili inserite in un contesto più generale. *"There are two primary scales of processes at work in LUCC, a micro-scale, in which individuals seek to achieve their objectives, and macro-scale, reflecting the context of these decisions...changes in the macro-scale are often unpredictable, given their result from complex interactions of economic, political, and transnational institutions and social*

9 Io stesso ho deforestato molto qui, lei non immagina neppure quanto...guardi che incoerenza, 35 anni fa presi in prestito soldi del governo, sussidi, per venire qui, altrimenti non sarei mai venuto senza queste agevolazioni. Mi pagarono per comprare un trattore e deforestare, io ho affittato 10 trattori e una catena enorme, per capirci avevo bisogno di un camion solo per trasportare la catena, la legavo fra due trattori, e avanzavamo e tutto quello che c'era davanti veniva sradicato. E ancora oggi succede così, qui in Brasile. Ma adesso il governo vuole che noi smettiamo di deforestare, di prendere anche un solo pezzo di legno; ma loro lo hanno fatto, loro ci hanno spinto a farlo, dato soldi, strutturato la nostra vita. Io penso che dobbiamo aver cura dell'Amazzonia, ma dobbiamo deforestare ora altrimenti poi diventerà un deserto, gli scienziati stanno spiegando questo, Al Gore sta dicendo questo..."

processes"¹⁰ [Brown et al., 2004, Pag.4]. La costruzione di scenari futuri, formulati a partire da ipotesi socio-economiche esplicite (aumento demografico, sviluppo economico, avanzamenti tecnologici, etc.), ed il loro utilizzo nella modellizzazione predittiva, permette di migliorare la nostra conoscenza delle relazioni fra macro-parametri, processi locali e cambi globali. In [Hauglustaine et al., 2004] troviamo un esempio di previsione multi-scenario delle traiettorie climatiche, come conseguenza dei cambiamenti nella composizione atmosferica (Tab.1-1); i differenti scenari (B1, B2, A2, A1FI,) includono una serie di ipotesi su crescita demografica, disparità sociali, sviluppo economico e tecnologico, elaborate di comune accordo con gli esperti di ogni settore disciplinare interessato.

Gaz à effet de serre	Temps de résidence (en années)	Période préindustrielle	2000	2100			
				B1	B2	A2	A1FI
CO ₂ (millionnières)	120	280	367	540	611	836	958
CH ₄	12	700	1760	1574	2973	3731	3413
N ₂ O	114	270	316	375	362	447	460
CF ₄	50000	0	0,082	0,208	0,384	0,397	0,341
C ₂ F ₆	10000	0	0,003	0,011	0,022	0,023	0,020
SF ₆	3200	0	0,005	0,035	0,044	0,065	0,056
HFC-23	260	0	0,015	0,032	0,034	0,033	0,034
HFC-32	5	0	0	0,008	0,017	0,020	0,020
HFC-125	29	0	0	0,058	0,102	0,107	0,134
HFC-134a	13,8	0	0,012	0,379	0,835	0,877	0,899

Tab 1-1: Esempio di modellizzazione predittiva multi-scenario. Gli scenari proposti B1, B2, A2 ed A1FI descrivono mondi possibili caratterizzati da differenti condizioni demografiche, economiche, sociali e tecnologiche, che finiscono inevitabilmente per descrivere differenti traiettorie nella composizione atmosferica e, dunque, dei cambiamenti climatici. [Fonte: Hauglustaine et al., 2004, Pag.126].

10 I processi di LUCC lavorano su due differenti scale, una micro-scala in cui gli individui cercano di raggiungere i loro obiettivi, una macro-scala che riflette il contesto in cui sono prese le decisioni... i cambiamenti nella macro-scala non sono spesso prevedibili poiché derivano da una complessa interazione fra istituzioni economiche, politiche e transnazionali, e processi sociali.

Per rappresentare un sistema complesso uomo-ambiente abbiamo bisogno di effettuare delle semplificazioni distillando alcuni processi ed elementi principali che ci permettono di descrivere ed analizzare la problematica in un linguaggio scientifico. Sulla base di una ampia casistica di LUCC, [Lambin et al., 2003] identifica alcuni patterns ricorrenti e 5 tipologie di cause principali di cambiamento di uso del suolo (Tab.1-2). I LUCC derivano dalla complessa interazione di queste variabili e dei meccanismi di retroazione (*feedback*), caratteristici dei sistemi adattabili. I cambiamenti nel sistema terrestre non sono stabili e discreti, ma si evolvono continuamente per l'azione sinergica e cumulativa di micro-processi e dinamiche globali, in un preciso contesto geografico, storico e socio-economico.

Comprendere e prevedere con esattezza il comportamento del sistema uomo-ambiente è un compito particolarmente oneroso e probabilmente impossibile; alcune difficoltà sono legate ai limiti del nostro sapere cognitivo, altre derivano dalle caratteristiche intrinseche al processo evolutivo, poiché nulla si ripete in maniera identica e "*l'évolution est créatrice et non plus seulement logique*"¹¹ [La Moigne, 1994, Pag.263].

*"Nature is too complex and depends on too many processes that are poorly understood or little monitored... We could, for example, pay more attention to nature... to accumulate information on how living things and their environments interact. That kind of data is crucial for models"*¹² [Dean, 2007, tratto dalla rubrica Books on Science].

11 L'evoluzione è creatrice e non più solamente logica.

12 La natura è troppo complessa, dicono, e dipende da troppi processi che sono compresi in maniera approssimativa o poco monitorati... Noi dobbiamo prestare più attenzione alla natura... per accumulare informazioni su come gli esseri viventi ed il loro ambiente interagiscono. Questo tipo di dati è di fondamentale importanza per i modelli.

Resource scarcity causing pressure of production on resources	Changing opportunities created by markets	Outside policy intervention	Loss of adaptive capacity and increased vulnerability	Changes in social organization, in resource access, and in attitudes
<p>Slow</p> <p>Natural population growth and division of land parcels</p> <p>Domestic life cycles that lead to changes in labor availability</p> <p>Loss of land productivity on sensitive areas following excessive or inappropriate use</p> <p>Failure to restore or to maintain protective works of environmental resources</p> <p>Heavy surplus extraction away from the land manager</p>	<p>Increase in commercialization and agro-industrialization</p> <p>Improvement in accessibility through road construction</p> <p>Changes in market prices for inputs or outputs (e.g., erosion of prices of primary production, unfavorable global or urban-rural terms of trade)</p> <p>Off-farm wages and employment opportunities</p>	<p>Economic development programs</p> <p>Perverse subsidies, policy-induced price distortions and fiscal incentives</p> <p>Frontier development (e.g., for geopolitical reasons or to promote interest groups)</p> <p>Poor governance and corruption</p> <p>Insecurity in land tenure</p>	<p>Impoverishment (e.g., creeping household debts, no access to credit, lack of alternative income sources, and weak buffering capacity)</p> <p>Breakdown of informal social security networks</p> <p>Dependence on external resources or on assistance</p> <p>Social discrimination (ethnic minorities, women, members of lower classes or castes)</p>	<p>Changes in institutions governing access to resources by different land managers (e.g., shift from communal to private rights, tenure, holdings, and titles)</p> <p>Growth of urban aspirations</p> <p>Breakdown of extended family</p> <p>Growth of individualism and materialism</p> <p>Lack of public education and poor information flow on the environment</p>
<p>Fast</p> <p>Spontaneous migration, forced population displacement, refugees</p> <p>Decrease in land availability due to encroachment by other land uses (e.g., natural reserves)</p>	<p>Capital investments</p> <p>Changes in national or global macro-economic and trade conditions that lead to changes in prices (e.g., surge in energy prices or global financial crisis)</p> <p>New technologies for intensification of resource use</p>	<p>Rapid policy changes (e.g., devaluation)</p> <p>Government instability</p> <p>War</p>	<p>Internal conflicts</p> <p>Illness (e.g., HIV)</p> <p>Risks associated with natural hazards (e.g., leading to a crop failure, loss of resource, or loss of productive capacity)</p>	<p>Loss of entitlements to environmental resources (e.g., expropriation for large-scale agriculture, large dams, forestry projects, tourism and wildlife conservation), which leads to an ecological marginalization of the poor</p>

Tab 1-2: Principali cause di LUCC individuate da [Lambin et al., 2003, Pag.224]

CAPITOLO No. 2

La modellizzazione dei cambi di uso del suolo

Critica a Protagora

Socrate: -[...] nessuna cosa è per se stessa una sola; tu non puoi, correttamente, dar nome a una cosa né a una sua qualità; se tu, per esempio, chiami alcuna cosa grande, ecco che essa potrà apparire anche piccola; se la chiami pesante, potrà apparire anche leggera; e così via per tutto il resto, perché niente è uno, né sostanza, né qualità. Dal mutar luogo, dal muoversi, dal mescolarsi delle cose fra loro, tutto diviene ciò che noi, adoprando una espressione non corretta, diciamo che è; perché niente mai è, ma sempre diviene. Su questo punto tutti i filosofi, uno dopo l'altro, a eccezione di Parmenide, si deve ammettere che sono d'accordo, [...]; e anche sono d'accordo i migliori poeti dell'uno e dell'altro genere di poesia, Epicarmo della commedia, della tragedia Omero; il quale, dicendo "Padre fu Oceano a' numi e madre Teti", intese dire che tutte le cose hanno origine dal flusso e dal movimento. O non credi che egli intese dire così?

Teeteto: - Sì, lo credo.

(Platone, Teeteto, 151 d-152)

2.1. I modelli e la modellizzazione: introduzione

*"... Modelling has become the common language of environmental science, one that is spoken in the same words by hydrologists or biologist and by natural scientists and social scientists or scientists and engineers."*¹ [Eddy, 1993, Pag. 6].

Osservare un sistema è anzitutto costruire un modello attraverso il filtro delle nostre conoscenze. Quando disegniamo uno schema mentale noi semplifichiamo la realtà basandoci più sulle relazioni qualitative che quantitative, linearizzando il comportamento degli elementi, separando componente spaziale da componente temporale ed a volte ignorando quest'ultima, trattando l'oggetto di studio isolatamente rispetto all'ambiente circostante. Quando i problemi diventano più complessi, quando le relazioni quantitative, le caratteristiche non lineari, la risoluzione spazio-temporale sono informazioni necessarie, allora la nostra capacità di studio e previsione dei cambiamenti del sistema risulta alquanto limitata. Dobbiamo perciò ricorrere a metodi numerici predittivi che sfruttano le notevoli potenzialità dei calcolatori elettronici [Costanza et al., 1998].

Modellizzare significa costruire un oggetto astratto che rappresenti la struttura e la dinamica del sistema osservato [Ramat, 2006]. Il modello è il risultato di questa attività.

Modellizzazione e **modello** sono due attributi di uno stesso processo cognitivo, ma non vanno confusi [Ferrand, 2006]. Il primo è un processo, un protocollo, un insieme di azioni per meglio comprendere il funzionamento complesso di un fenomeno, una metodologia che associa attori ed oggetti, secondo regole

¹ La modellizzazione è diventata il linguaggio comune per le discipline ambientali, parlato con le stesse parole da idrologi o biologi e scienziati naturali o ingegneri.

ed ipotesi che il modellizzatore stabilisce. Il modello è uno strumento, un filtro attraverso cui osservare e rappresentare un sistema, una sua astrazione che lo semplifica ignorando numerose caratteristiche e focalizzando l'attenzione sugli aspetti che meglio definiscono la problematica in esame. Egli è costruito in funzione di un insieme di obiettivi che determinano le ipotesi della modellizzazione ed il grado di semplificazione del modello.

I modelli hanno spesso una debole articolazione (basso numero di variabili esplicative) ma una grande precisione, restringono il campo di analisi focalizzandosi su processi via via più dettagliati ed ignorando la totalità del sistema. Il modellizzatore tende a lavorare in condizioni ottimali, su aree di piccole dimensioni oppure con dinamiche estremamente semplici descritte da una quantità e qualità di dati a volte eccessiva. Ci dicono "troppo su troppo poco" [Jorgensen, 1994]. A volte è preferibile fare la scelta opposta, aumentare la **complessità** del modello a scapito della precisione², lavorare su sistemi complessi, conoscere "un po' a riguardo di tutto". Un primo risultato della simulazione, per quanto limitato, offre un valido aiuto per meglio comprendere le relazioni nascoste, i meccanismi prima ignorati o mal interpretati. È bene ricordare però che non sempre aumentare la complessità del modello, aggiungendo nuove variabili esplicative, significa automaticamente aumentare la nostra capacità di conoscere e rappresentare un sistema (Fig.2-1).

² Questa affermazione non è vera in assoluto. La scelta è fatta sempre in funzione dei dati e degli obiettivi.

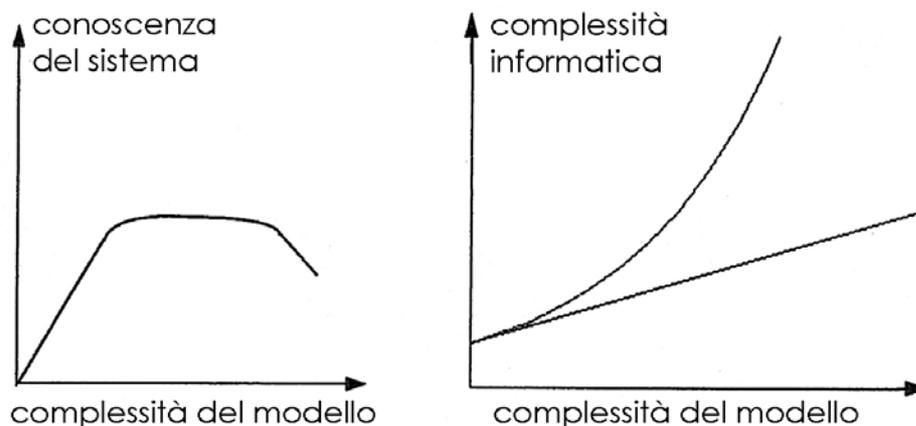


Fig 2-1: Complessità del modello e livello di conoscenza del sistema. A sinistra: l'aggiunta di variabili esplicative non porta necessariamente ad un aumento della conoscenza sul sistema; raggiunto un limite ottimale, l'aumento della complessità può determinare un aumento della confusione e dell'incertezza allontanando il comportamento del modello dal sistema reale. A destra: l'aumento del numero di variabili determina un aumento rapido della complessità informatica; durante la modellizzazione è necessario tenere a mente le capacità di elaborazione ed i tempi di calcolo del proprio computer. [Grafici adattati da Coquillard et al., 1997].

Lo scopo della modellizzazione non è offrire uno strumento complesso che si perda in inutili labirinti teorici, matematici o informatici, ed accessibile solo ad un pubblico di specialisti. *"...il ne faut pas penser que le but est d'offrir le modèle le plus beau... on ne fait pas de la modélisation pour la modélisation mais pour comprendre le fonctionnement du Monde et comprendre les impacts de certaines perturbations sur le Monde"*³ [Ramat, 2006, Pag.50]. Dobbiamo saper trovare un giusto equilibrio fra semplicità del modello e corretta rappresentazione dei fenomeni naturali o artificiali studiati. L'aggiunta di nuove variabili può aiutare il modello a meglio descrivere il sistema reale, ma allo stesso tempo può renderlo sensibile a fattori secondari che non concernono direttamente le ipotesi principali della modellizzazione creando confusione o contraddizione fra dati osservati e risultati della

3 Non dobbiamo pensare che lo scopo sia offrire il modello più "bello"... non si fa della modellizzazione per il semplice gusto di modellizzare ma per comprendere il funzionamento del Mondo e capire l'impatto che certe perturbazioni possono avere sul mondo.

simulazione. Vale la pena aumentare la complessità del modello? Cosa guadagniamo in termini di conoscenza? Serve per raggiungere il nostro obiettivo? Per rispondere a queste domande bisogna ricordare che un modello è anzitutto uno strumento ed il suo valore è definito relativamente allo scopo che ci proponiamo.

La scelta degli elementi, dei processi e degli attributi da "distillare" per rappresentare il sistema reale è influenzata dagli obiettivi e dalla **soggettività del modellizzatore**. Le scelte metodologiche, le preferenze di un formalismo o di un programma rispetto ad un altro, possono portare a risultati differenti. Due modellizzatori che lavorano sullo stesso sistema danno origine a due differenti modelli. Innanzitutto perché ognuno osserva la realtà da un punto di vista personale, partigiano, influenzato dall'appartenenza ad una disciplina specifica e dalla propria esperienza. Secondo perché le conoscenze informatiche e metodologiche differiscono ed ognuno tende ad utilizzare il formalismo o il programma che meglio conosce o con cui è abituato a lavorare. Questo ridimensiona, in parte, il valore del risultato ottenuto, poiché impregnato di una soggettività spesso non condivisa da altri rami scientifici. *"...nous possédons les éléments du problème; nous savons, d'une connaissance abstraite, comment il sera résolu, car le portrait ressemblera sûrement au modèle et sûrement aussi à l'artiste..."*⁴ [La Moigne, 1994, Pag. 263].

Per chi è stata concepita la modellizzazione? A chi servirà il modello? Come costruire uno strumento utile che soddisfi le richieste dell'**utente**? La risposta a queste domande determina la natura delle problematiche studiate e differenzia le diverse scelte durante la costruzione del modello [Ferrand, 2006]. Il modellizzatore deve interagire con gli esperti tematici del

4 Possediamo gli elementi del problema, sappiamo come sarà risolto, poiché il risultato finale assomiglierà al modello usato ed all'artista.

dominio in esame, interessati all'uso del modello e ai suoi risultati. Spesso tale cooperazione può sembrare impossibile, data la scarsa volontà dei ricercatori a condividere le proprie conoscenze od a ridiscutere le proprie competenze settoriali; altre volte si incontrano oggettive difficoltà ad integrare ipotesi e pratiche accademiche eterogenee e contraddittorie (economia ed ecologia, ad es.). *"Significant results can always emerge when bridges are erected between diverse language and mentalities, even for achieving complex goals"*⁵ [Iovine et al., 2006, Pag.859]. La **multidisciplinarietà** deve essere considerata un cardine portante, un elemento imprescindibile della modellizzazione e, più in generale, di tutta la ricerca scientifica. Troppo spesso noi saltiamo la fase di discussione e di creazione di consenso fra differenti discipline sulla natura del problema studiato e ci concentriamo direttamente sulla implementazione del modello. *"what we save in time and effort by jumping ahead is easily lost later on in attempts to forge agreement about result..."*⁶ [Costanza et al., 1998, Pag.187].

Abbiamo già detto che la prima tappa del processo di modellizzazione passa attraverso la definizione degli elementi, degli attributi e delle relazioni di un sistema reale rispetto ad un obiettivo specifico. La descrizione di un sistema si può articolare su diversi livelli, dalla semplice osservazione del fenomeno (nessun legame fra input ed output è stabilito), alla specificazione del comportamento interno di un sistema dinamico attraverso la nozione di stato e funzione di transizione, e la successiva scelta del formalismo più adatto a rappresentarla (un automa a stati finiti, ad es.). Lo sviluppo di un semplice modello concettuale offre una rappresentazione generale del comportamento del sistema (Fig.2-2); una successiva analisi

5 Risultati significativi possono essere ottenuti stabilendo ponti fra differenti linguaggi e mentalità per raggiungere obiettivi complessi.

6 ...il tempo guadagnato saltando alcune tappe iniziali è facilmente perso nel raggiungere un consenso generale sui risultati.

statistica, basata su informazioni storiche, è usata per quantificare le relazioni fra le sue diverse componenti. Il risultato, fortemente dipendente dai dati e dalla scelta delle tecniche statistiche, non garantisce però che i principali fattori responsabili dell'evoluzione del sistema siano individuati. Un database storico non sempre è in grado di "catturare" tutte le informazioni necessarie per comprendere un fenomeno. I processi spaziali registrati possono cambiare nel tempo e gli effetti delle decisioni politiche ed amministrative possono manifestarsi solo in un futuro prossimo; la capacità di previsione è dunque fortemente ridotta [Costanza et al., 1998].

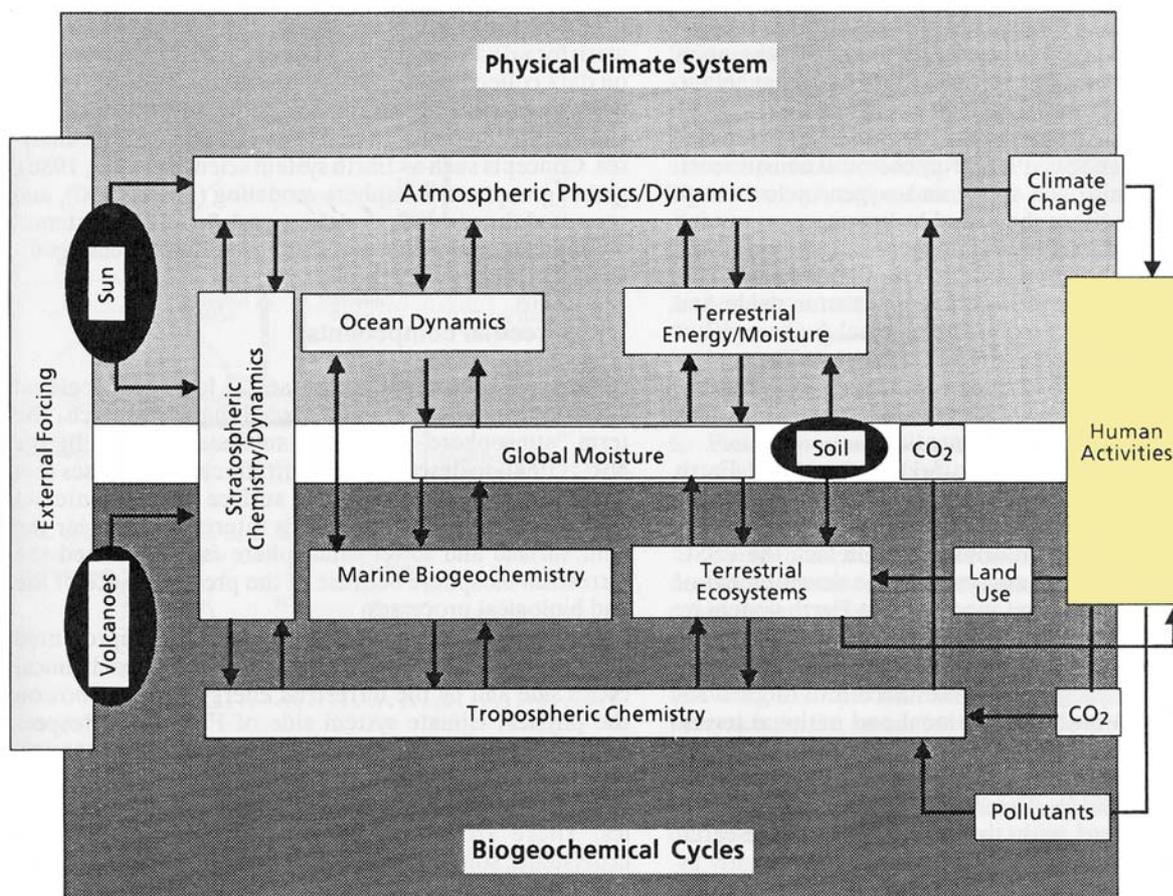


Fig 2-2: Modello concettuale dei principali processi nel nostro pianeta che operano su grande scala temporale (decenni-secoli). La costruzione di un modello concettuale richiede una buona conoscenza dei processi fisici, degli elementi e delle inter-relazioni più importanti che esistono in natura e nella società. Lo schema sopra riportato rappresenta la prima fase dello sviluppo del ESSC, Earth System Science Concept, un modello integrato del sistema Terra per comprenderne il funzionamento e prevedere i cambiamenti del clima sul lungo periodo. [Immagine adattata da ESSC, 1986].

La modellizzazione dinamica, al contrario, costruisce una rappresentazione del fenomeno includendo tutti gli aspetti attuali del sistema (leggi fisiche, fattori socio-economici, etc.), o che prevediamo influenzeranno il suo comportamento futuro (pianificazione politico-amministrativa, vincoli ambientali, etc.). Questa minor dipendenza dai dati storici garantisce una maggior flessibilità del modello e la possibilità di adattarlo a differenti problematiche (sviluppo urbano, deforestazione, scienze sociali, etc.).

A partire da una serie di **ipotesi** sul fenomeno osservato (paradigmi), cerchiamo di ampliare la nostra conoscenza del sistema. La costruzione di un primo modello schematico deve contenere tutte le informazioni necessarie per implementare il modello stesso in un qualche linguaggio di programmazione (simulazione). Questi ultimi sono appositamente disegnati per facilitare la modellizzazione dei sistemi non lineari e dinamici; in commercio esistono molti esempi di **linguaggi grafici di programmazione** fra cui ricordiamo STELLA® (prodotto dalla Isee systems), PCRaster (proposto dal Dip. di Geografia della università di Utrecht) e Dinamica EGO (sviluppato dal CSR della università di Belo Horizonte). Tutti i programmi permettono la creazione di modelli complessi partendo da operatori e funzioni elementari interconnessi fra loro e visualizzabili attraverso un'interfaccia grafica.

Il computer realizza poi, abbastanza agevolmente, una sequenza di operazioni spesso interrotta dalla necessità di ritornare sui propri passi e modificare i parametri iniziali della simulazione (Fig.2-3).

La modellizzazione ha una forte **finalità di apprendimento**; durante le prime fasi il modellizzatore rimette in gioco le proprie capacità cognitive nel selezionare le caratteristiche essenziali e nel formulare le ipotesi sui meccanismi che generano le osservazioni sul sistema in esame. "...the

explanatory models do not attempt to reproduce actual LUCC systems; instead, they concentrate on specific aspects and on modelling fundamental dynamics, in the hope that such laboratory explorations will lead to empirically relevant insights"⁷ [Parker et al., 2003, Pag.326].

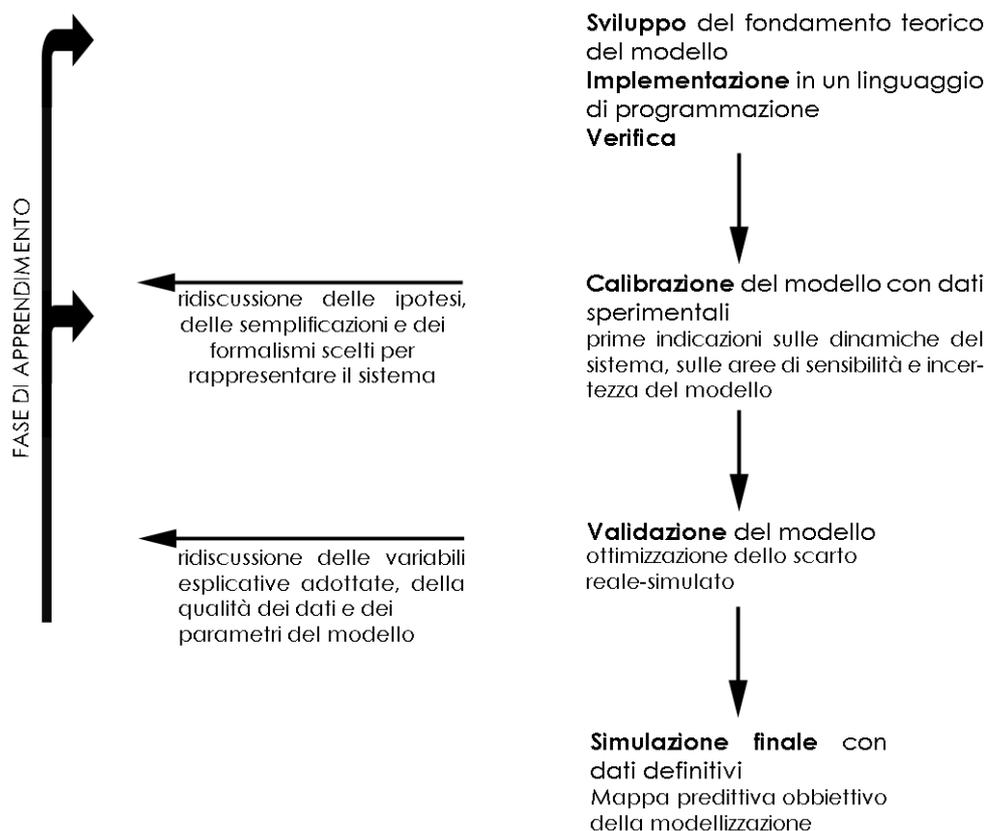


Fig 2-3: Fasi della modellizzazione. La calibrazione del modello prevede la scelta dei parametri e dei criteri ambientali che rappresentano il fenomeno studiato. La validazione consiste nel confronto fra la mappa simulata in una data nota e la realtà. Entrambi queste fasi hanno una finalità di apprendimento; l'analisi dei primi risultati sperimentali permette di evidenziare relazioni nascoste o mal interpretate fra i processi e le variabili esplicative. Il modellizzatore può ritornare sulle proprie scelte iniziali cambiando le ipotesi, i formalismi od i criteri utilizzati.

L'insieme dei formalismi è tradotto in un modello e successivamente caricato in un linguaggio di programmazione su

⁷ I modelli esplicativi non vogliono riprodurre l'attuale sistema di uso del suolo; essi vogliono, invece, concentrarsi su alcuni aspetti specifici e sulle dinamiche principali del sistema, con la speranza che questo laboratorio di esplorazione ci porti a scoprire alcune relazioni empiricamente rilevanti.

un computer, permettendoci di osservare l'evoluzione delle variabili scelte ed offrendo dei primi risultati. Se il modello riproduce fedelmente in termini quantitativi il sistema reale, ci troviamo nella situazione ideale in cui non abbiamo bisogno di ulteriori affinamenti. Più probabilmente le prime simulazioni divergeranno abbastanza dalla realtà, offrendoci la possibilità di esplorare il comportamento del modello per diversi input (variabili, parametri, etc.) ed effettuare le opportune modifiche. In ogni caso avremmo ottenuto un'informazione in più [Coquillard et al., 1997]. Durante questa fase il modello si deve considerare come uno strumento di analisi, di conoscenza di un sistema complesso che rivela relazioni e proprietà prima nascoste o mal interpretate, che ci porta ad identificare nuove variabili esplicative o che testa le nostre conoscenze e le ipotesi fatte, evidenziando le eventuali lacune teoriche nel nostro approccio. *"...other challenges may arise due to the nature of the phenomena: some parameters of the model cannot be independently determined, either for fundamental reasons (e.g. they may be empirical, not physical) or for practical reasons (e.g. direct measures are not allowed); their values could possibly be found by comparing the model outcome with a set of experimental data"*⁸ [Iovine et al., 2006, Pag.859]. Per ora i risultati della simulazione sono lontani dagli obiettivi prefissati e la finalità di studio prevale sulla finalità di predizione e pianificazione.

Una volta che i parametri, le variabili e i formalismi sono stati definitivamente scelti, coerentemente con le indicazioni emerse nella fase di validazione (minimizzo lo scarto realtà-simulato), si arriva alla formazione di un modello generalizzato, indipendente dai dati o dalla problematica

8 ...altre sfide possono presentarsi, dovute alla natura del fenomeno: alcuni parametri del modello non possono essere determinati indipendentemente, sia per ragioni di base (ad esempio sono empirici e non fisici), sia per ragioni pratiche (ad esempio non è possibile effettuare misure dirette); i loro valori possono essere determinati confrontando il risultato del modello con una serie di dati sperimentali.

affrontata. Il fondamento teorico del modello può allora essere riassunto e presentato al mondo scientifico attraverso i suoi elementi chiave per permetterne la riproduzione e l'utilizzo in contesti disciplinari differenti [Amblard et al., 2006].

I processi nel mondo reale sono tipicamente 3D, dipendenti dal tempo ed estremamente complessi; questa complessità include spesso comportamenti non lineari, componenti stocastici e cicli su multiple scale temporali e spaziali. Studiare tale soggetto, a livello individuale o come parte integrante di un sistema più grande, passa attraverso la formulazione di ipotesi e regole per spiegarne il funzionamento. Le regole possono a loro volta essere espresse attraverso formule matematiche o relazioni logiche; la complessità dei processi non permette soluzioni analitiche ed ha bisogno di una serie di semplificazioni teoriche per ridurre il numero delle equazioni utilizzate.

I modelli matematici basati su linguaggi di programmazione che realisticamente simulano l'evoluzione dei patterns spaziali nel tempo, sono sempre più utilizzati per un'analisi quantitativa, e non solo più qualitativa, delle complesse problematiche a livello locale, regionale o globale [Steyaert, 1993]. La simulazione spazio-temporale, integrata con le **informazioni socio-economiche**, offre uno strumento diagnostico e predittivo unico per la valutazione del rischio ambientale o per la gestione delle georisorse.

Le notevoli capacità di calcolo dei computer permettono oggi di sviluppare in un ambiente virtuale (*grid*), metodi numerici basati su una **discretizzazione** spazio-temporale che facilita la simulazione di fenomeni complessi. La parametrizzazione dei sub-grid permette di legare i processi che si manifestano a piccola scala, alla globalità dell'analisi. Recentemente, infatti, ci si è concentrati sull'importanza di integrare modelli che operano su differenti livelli di dettaglio per evidenziare il legame fra cambiamenti

globali ed effetti cumulativi delle attività a livello regionale [Prinz et al., 2003].

Per comprendere la complessità dei sistemi reali dobbiamo dunque muoverci attraverso differenti **scale spazio-temporali** all'interno di una disciplina, ed attraverso molteplici discipline. Lo spettro delle scale temporali può essere semplicemente diviso a seconda che lo studio si concentri sul corto (secondi-giorni), medio (giorni-mesi) o lungo (anni) periodo; analogamente, per estrapolazione, possiamo estendere i risultati delle osservazioni locali alla totalità della regione. I SIG rappresentano uno strumento estremamente utile per integrare dati provenienti da differenti sorgenti e con risoluzioni variabili, come i dati di campo (informazioni puntuali raccolti con GPS) e le scene satellitari. In [Chiles et al., 1999] si evidenzia come la frequenza, la dispersione e la correlazione spaziale di una stessa variabile possano cambiare in funzione della dimensione del supporto; questa considerazione deve essere tenuta a mente durante l'integrazione di dati con differenti risoluzioni spazio-temporali (un punto ed un istante, un valore medio su un'area durante un intervallo, etc.). Il lavoro svolto dal gruppo WAVES, descritto in [Gaiser et al., 2003], offre un ottimo esempio di modellizzazione multiscalare (MOSDEL, Model for Sustainable Development of Land use) per evidenziare il legame causale fra attività umane, benefici economici e impatto ecologico a livello regionale e globale.

Abbiamo già detto che un sistema reale può essere decomposto in sottosistemi che operano su scale spazio-temporali differenti. Analogamente un modello deriva dall'assemblaggio di sottomodelli interconnessi in una struttura gerarchica. Ogni sottomodulo è oggetto di un formalismo differente, che meglio si adatta a rappresentare il sottosistema considerato. La domanda ora potrebbe essere: "quale è il livello di dettaglio più adatto per integrare due

modelli che lavorano su scale differenti?”. Il passo temporale e la risoluzione spaziale da utilizzare deve rappresentare meglio possibile le caratteristiche e le dinamiche principali del sistema globale, relativamente agli obiettivi da raggiungere. Se vogliamo capire le conseguenze ecologiche o socio-economiche di una perturbazione nell'immediato futuro, privilegeremo una risoluzione temporale più fine rispetto ad una pianificazione strategica per raggiungere obiettivi sul lungo termine [Follador et al., 2008; Kavouras, 2001].

Sono dunque gli obiettivi che ci proponiamo, scelti coerentemente con i dati disponibili [Matheron, 1978 e 1989], ad influenzare lo sviluppo del modello ed a definire il **livello di astrazione**. Questa scelta ha importanti ripercussioni in termini di costi, tempo e sforzo lavorativo, ed è vincolata alla capacità di elaborazione del nostro computer.

Il modello che abbiamo disegnato ci serve per meglio comprendere un sistema naturale? Per concepire un qualche oggetto artificiale? O ci permetterà, per simulazione, di valutare le conseguenze delle perturbazioni sul sistema studiato? Servirà a politici e manager per raccogliere informazioni necessarie alla pianificazione futura di un'attività?

Oggi risulta chiaro che per risolvere le pressanti questioni ambientali bisogna integrare le competenze delle discipline tradizionali con i risultati offerti dai metodi predittivi. Un modello deve essere **facile da usare** ed **adattarsi** alle esigenze di diversi domini scientifici [Park, 1993]. Aggiungerei a questo un'ulteriore considerazione: un modello dovrebbe essere in grado di lavorare con **dati facilmente ottenibili**, senza grande dispendio di soldi o tempo.

L'**incertezza** dei risultati di un modello è conseguenza delle semplificazioni effettuate per rappresentare il sistema reale: solo un numero limitato di processi, infatti, viene

analizzato, e tale scelta è spesso condizionata dalle limitate conoscenze del modellizzatore; i meccanismi che regolano l'evoluzione del fenomeno sono ignorati o mal interpretati, oppure la risoluzione temporale e spaziale adottata può non essere adatta al raggiungimento degli obiettivi prefissati; non possediamo dati di buona qualità od ancora non siamo riusciti ad integrare le diverse componenti della simulazione in un tutto coerente [Christie et al., 2005].

Non è mai facile stabilire in maniera generalizzata quale sia l'approccio metodologico migliore per studiare i fenomeni complessi; ogni modello deve essere valutato nel proprio ambiente di lavoro e relativamente ai propri scopi. Possiamo dire comunque che un buon modello deve rispettare, come minimo, 3 criteri fondamentali: realismo (la simulazione deve riprodurre qualitativamente il comportamento del sistema reale); precisione (la simulazione deve riprodurre quantitativamente il comportamento del sistema con un grado di precisione adatto ai nostri obiettivi); flessibilità (possibilità di adattare lo stesso modello a differenti problematiche e discipline). Un singolo modello non può massimizzare tutti i criteri e la scelta di quale caratteristica privilegiare deve essere fatta sempre in funzione dello scopo della modellizzazione [Orzak et al., 1993] e delle informazioni disponibili.

"des modèles des phénomènes perçus ou conçus dans le temps et dans l'espace, nous disons volontiers qu'ils sont des représentations... représenter c'est aussi re-présenter, présenter à nouveau, après qu'un temps soit écoulé, en acceptant donc que la re-présentation ne soit pas nécessairement un duplicata, une copie conforme. Re-présenter, c'est proposer une représentation qui porte en elle-même sa légitimation : elle a mémoire et projet, et elle fonde sa

légitimité sur sa cohérence avec son histoire et ses devenirs"⁹

[La Moigne, 1994, Pag.266].

2.2. I modelli di LUCC

L'uso del suolo è probabilmente il risultato più visibile dell'interazione fra l'uomo e l'ambiente biofisico. I modelli di LUCC aiutano a comprendere questo intricato sistema e possono offrire un'informazione credibile sulla possibile situazione futura del nostro pianeta, relativamente agli attuali processi decisionali.

Il LUCC è un tema centrale nello studio dei cambi globali, poiché l'impatto antropico sul sistema terrestre ha forti ripercussioni sulla stabilità del clima, sulla biodiversità e sull'equilibrio di molti ecosistemi, sulla degradazione del suolo e sulla vulnerabilità del sistema uomo-ambiente.

Ricercatori ed esperti appartenenti a settori accademici e disciplinari differenti hanno iniziato a collaborare per dare la risposta più chiara possibile alle pressanti questioni ambientali. Questa multidisciplinarietà nasce dalla necessità di comprendere a fondo le complesse relazioni fra le attività umane e il sistema biofisico naturale; "*the diverse community of researchers engaged in these efforts has spawned a Land Change Science (LCS)*"¹⁰ [Rindfuss et al., 2004, Pag. 13976]. L'integrazione fra tutte queste opinioni partigiane è, in generale, realizzata attraverso un modello, il cui scopo è analizzare i dati raccolti a differenti scale spazio temporali, rappresentare le relazioni causa-effetto ed evidenziare le

9 ...diremmo che i modelli percepiti o concepiti nello spazio e nel tempo sono delle rappresentazioni...rappresentare è anche ri-presentare, presentare di nuovo dopo un certo periodo, accettando dunque che non sia esattamente un duplicato, una copia perfetta. Ri-presentare è proporre una rappresentazione che ha una legittimità in se: ha memoria ed un fondamento teorico, basa la sua legittimità sulla coerenza fra dati storici e divenire.

10 I differenti gruppi di ricerca ingaggiati in questo sforzo hanno dato origine ad una nuova scienza dei cambi di uso del suolo (LCS).

conseguenze future di un certo LUCC [Verburg et al., 2005]. *"In order to encourage cross-fertilization and comparison between models, it may be important to have a common language through which model mechanisms can be communicated"*¹¹ [Parker et al., 2003, Pag. 331].

Il programma LUCC, cofinanziato dall'IGBP e IHDP, indica nell'integrazione fra le scienze sociali, economiche, ambientali e spaziali uno degli obiettivi principali nello studio dei cambi di uso del suolo. *"As this research progresses, people and pixels need to be linked to analyze human agency and social structure in land change. The depth of understanding brought to bear from bottom-up, field-based approaches must be made spatially explicit, including efforts to integrate results with remote sensing information (pixelizing the social); remote sensing approaches that potentially increase spatial and temporal dimensions of land-use/land-cover change understanding must be pushed beyond its biophysical dimensions (socializing the pixel)"*¹² [Lambin et al., 2000, Pag.24].

La ricerca dei principali fattori che guidano i cambiamenti e la loro corretta rappresentazione in un linguaggio scientifico, è probabilmente la fase più delicata nello sviluppo di un modello realistico di LUCC [Veldkamp et al., 2001]. Un cambio è normalmente il risultato dell'azione, indipendente ma sinergica, di molteplici attori che lavorano su scale differenti e che cercano di raggiungere un determinato obiettivo o beneficio (un individuo, una comunità, un

11 Per incoraggiare un'interdisciplinarietà "fertilizzante" e un confronto fra modelli, potrebbe essere necessario utilizzare un linguaggio comune attraverso cui far comunicare le differenti metodologie adottate.

12 Man mano che la ricerca avanza, persone e pixels devono essere legati per analizzare le attività antropiche e la struttura sociale all'interno dei cambi di uso del suolo. La profondità della ricerca deve includere un'analisi dal locale al globale e viceversa; i dati di campo devono essere inseriti in un contesto spaziale ed integrati con le informazioni di telerilevamento ("pixelizzare" il sociale). Le immagini satellitari che potenzialmente aumentano la capacità di analisi temporale e spaziale dei LUCC, devono essere considerate oltre la loro semplice dimensione biofisica (socializzare i pixels).

dipartimento, uno stato). Un modello deve essere in grado di legare dinamicamente questi micro-macro-processi di LUCC al sistema naturale in cui operano, considerando eventuali risposte (retroazioni) del sistema biofisico al cambio, che possono a loro volta influenzare il rapporto futuro fra società ed ambiente. Comprendere queste dinamiche complesse e non lineari, permette al modellizzatore di evidenziare sia i processi locali sia i loro effetti emergenti a livello globale; *"definition of emergence usually concern macroscale phenomena that arise from microinteraction... however macrostructure potentially also affect units at the microscale"*¹³ [Parker et al., 2003, Pag.323].

I LUCC sono spesso rappresentati attraverso una categorizzazione dei fenomeni osservati, oscurando così una parte dei processi e delle decisioni che hanno influenzato le dinamiche di cambio [Munroe et al., 2007]. Una decisione è dunque vista come un processo discreto che si manifesta con la presenza o meno di una certa categoria di uso del suolo, fra le possibili alternative considerate; chi sono gli attori di questa decisione, gli obiettivi che si nascondono dietro una scelta, gli eventuali stati di transizione fra due categorie, la dispersione e le relazioni spaziali fra differenti processi economici e sociali, diventano dunque informazioni difficilmente misurabili e probabilmente perse. Gli effetti fisici di una certa decisione sull'ambiente sono a loro volta rappresentati da una categorizzazione della copertura terrestre, che spesso è il risultato della classificazione di immagini satellitari [Chuvieco, 2002]. Vediamo dunque come il passaggio dal mondo reale al modello sia accompagnato da una progressiva perdita di informazioni, dovuta alla sequenza di

13 La nozione di emergenza concerne fenomeni a macro-scala che derivano dall'interazione sinergica di microprocessi... comunque anche la macrostruttura affetta potenzialmente gli elementi a microscala.

semplificazioni ed approssimazioni adottate per raffigurare il sistema uomo-natura.

I LUCC sono principalmente dei fenomeni locali, inseriti però in un contesto amministrativo, economico e naturale gerarchico; il livello di analisi adatto a rappresentare le interazioni fra le variabili esplicative e gli agenti, dipende dal processo spaziale considerato. Integrare in maniera corretta queste differenti scale d'informazione, facilita la comprensione delle dinamiche di LUCC. Non esiste una metodologia standardizzata per effettuare questa operazione; ogni modellizzatore cerca di utilizzare i dati a disposizione nel modo più fruttuoso per ottenere uno scopo, sia esso aumentare la propria conoscenza di un sistema complesso o prevederne l'evoluzione sotto differenti scenari.

Includere informazioni socio-economiche nell'analisi dei LUCC può risultare complicato, sia per l'assenza frequente di una referenza spaziale in questo tipo di dati, sia per le difficoltà metodologiche che incontriamo nell'integrare la sfera sociale, politica e naturale. Un atto amministrativo può riferirsi ad un dipartimento od una regione; questa uniforme dispersione di un fattore su una macro-area contrasta con la variabilità spaziale dei micro-processi locali all'interno della stessa regione. Gli effetti di un processo economico-sociale e le forze motrici del cambio sono spesso visualizzate attraverso variabili semplicistiche, facili da misurare spazialmente, come la distanza da un oggetto rappresentativo (strada, mercato, scuola, etc.) [Serneels et al., 2001]. *"This shift from driving force to proximate causes, for data convenience, might obscure causality"*¹⁴ [Veldkamp et al., 2001, Pag.2], ignorando inoltre la variabilità e le relazioni spaziali che caratterizzano molti fenomeni economico-sociali.

14 Questo passaggio dalle forze motrici alle cause provvisorie, per convenienza, può oscurare la causalità.

I modelli di LUCC sono frequentemente utilizzati per supportare decisioni economico-politiche, pianificare investimenti ed offrire possibili soluzioni a problematiche di rischio ambientale. I risultati della simulazione ci permettono di visualizzare possibili "mondi futuri" in base alle differenti alternative adottate (sviluppo demografico ed urbanistico, attività previste, richiesta di materie prime, etc.), evidenziando eventuali rischi sociali ed ambientali legati all'attuazione di un piano o progetto specifico. *"Dynamic modeling is one such tool that helps us close spatial and temporal gaps between decision, actions and results"*¹⁵ [Costanza et al., 1998, Pag.185].

In particolare la simulazione di LUCC rappresenta un importante aiuto nella preparazione, sviluppo e valutazione di strategie e pianificazioni spaziali su larga scala [Koomen et al., 2007]. Durante la fase di preparazione (Fig.2-4) si elaborano differenti stati futuri del sistema uomo-natura analizzato, basati su una proiezione dei trend osservati nei dati storici o ipotizzando nuove condizioni socio-economiche (scenari). Il primo caso è adatto solo allo studio dell'evoluzione dei patterns spaziali nel breve periodo, poiché nulla ci garantisce che le forze motrici del cambio osservato in passato rimangano invariate; un modello, opportunamente calibrato e validato, può avere uno scarso potere predittivo se il contesto socio-economico, le variabili esplicative e le tendenze osservate, cambiano rispetto alle informazioni contenute nei dati storici utilizzati [Pontius et al., 2004]. L'uso di scenari è preferibile per studi sul lungo termine, poiché essi coprono in maniera più completa le possibili divergenze dello sviluppo economico e sociale futuro rispetto al presente [Barredo et al., 2008; Verburg et al., 2006; Hauglustaine et al., 2004].

15 La modellizzazione dinamica ci permette di colmare il vuoto che esiste fra processi decisionali, azioni e risultati.

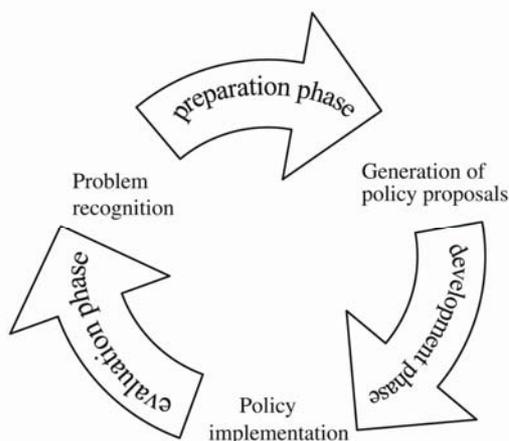


Fig 2-4: Rappresentazione delle fasi principali della pianificazione spaziale.
 1- Fase di preparazione: visualizzo future configurazioni di uso del suolo sotto differenti scenari o tendenze ricavate da dati storici. 2- Fase di sviluppo: implemento molteplici strategie politico-amministrative e visualizzo i risultati in termini di uso del suolo e impatto ambientale. 3- Fase di valutazione: evidenzio l'efficacia o le lacune dell'attuale politica amministrativa ed economica. [Immagine adattata da Koomen et al., 2007].

La seconda fase prevede l'implementazione di differenti alternative e la visualizzazione dell'impatto che ogni singolo progetto ha sul sistema terrestre. Chi prende le decisioni è dunque messo di fronte alle proprie responsabilità.

Una volta che la strategia politico-economico-amministrativa è stata scelta ed attuata, i modelli di LUCC possono essere utilizzati per valutarne l'impatto sul sistema sociale e naturale, evidenziando eventuali lacune e problematiche precedentemente non considerate.

Un modello di LUCC è uno strumento indispensabile nello studio dei sistemi complessi uomo-ambiente ed un aiuto efficace nel processo decisionale politico ed economico. Essi però non devono essere considerati delle "answer machines"¹⁶ [Agarwal et al., 2002], ma semplicemente un punto di incontro di molteplici discipline interessate allo studio ed al benessere del nostro pianeta, che condividono opinioni, metodologie e punti di vista a volte molto differenti.

¹⁶ Macchine che producono risposte

2.3. Classificazione dei modelli di LUCC

2.3.1. Introduzione

Negli ultimi vent'anni centinaia di modelli di LUCC sono stati sviluppati per rispondere alle domande di svariate discipline, dall'ecologia alla geografia, dall'economia all'informatica, dalla statistica alle scienze della Terra, ed altre ancora. Esistono in letteratura alcuni esempi di revisione sullo stato dell'arte dei modelli di LUCC, spesso caratterizzati da approcci differenti nella fase di classificazione. I criteri scelti per la categorizzazione dei modelli includono il fondamento teorico della ricerca, il livello di aggregazione spaziale adottato, il grado di rappresentazione della complessità di un sistema, il tipo di LUCC considerato, le soluzioni tecniche scelte, le metodologie di calibrazione e validazione etc.

Alcuni autori hanno preferito raggruppare i modelli in base al loro scopo finale o alla scala di lavoro; *"These models can be categorized as either whole landscape models, distributional landscape models, or spatial landscape models, depending on the amount of detail included in the models"*¹⁷ [Baker, 1989, Pag.111]. I modelli di analisi globale si concentrano sui cambi degli attributi emergenti di un sistema, ma non permettono di visualizzare i processi a microscala. I modelli distribuzionali analizzano una sub-regione e l'evoluzione di ogni classe di copertura terrestre in essa contenuta, ma possono trascurare delle importanti relazioni spaziali. I modelli spaziali permettono di descrivere la localizzazione e la configurazione di un cambio di uso del suolo. Questa prima revisione di Baker

17 Questi modelli possono essere ordinati in classi come modelli globali, distribuzionali e spaziali, dipendendo dalla quantità di dettagli contenuta nel modello.

non fa nessun riferimento su come rappresentare esplicitamente il fattore umano nel processo di modellizzazione.

Irwin e Geoghegan richiamano l'attenzione sulla necessità di sviluppare dei modelli economici di LUCC più articolati che includano anche "*the human behavioural components*"¹⁸ [Irwin et al., 2001, Pag.7], per meglio comprendere le relazioni causali fra le scelte adottate da un agente e i cambiamenti osservati nel sistema. L'articolo passa in revisione i modelli economici tradizionali di uso del suolo, evidenziando i progressi fatti in direzione di un approccio che integri una completa analisi economica in una struttura spaziale definita. I modelli vengono classificati in spazialmente espliciti e non economici (automa cellulare, modelli empirici ed ibridi), ed i modelli economici, inseriti o meno in un contesto spaziale.

[Lambin, 1997] propone una classificazione delle metodologie di monitoraggio dei LUCC in aree tropicali analizzando l'utilità dei modelli descrittivi, empirici, statistici e dinamici, per lo studio dei fenomeni di deforestazione e degradazione del suolo, e la simulazione della loro evoluzione futura sotto differenti scenari.

[Agarwal et al., 2002] seleziona 19 modelli di LUCC e li analizza in base alla loro capacità di rappresentare la complessità spaziale, temporale e antropica di un sistema (Fig.2-5).

18 Componenti descrittive del comportamento umano

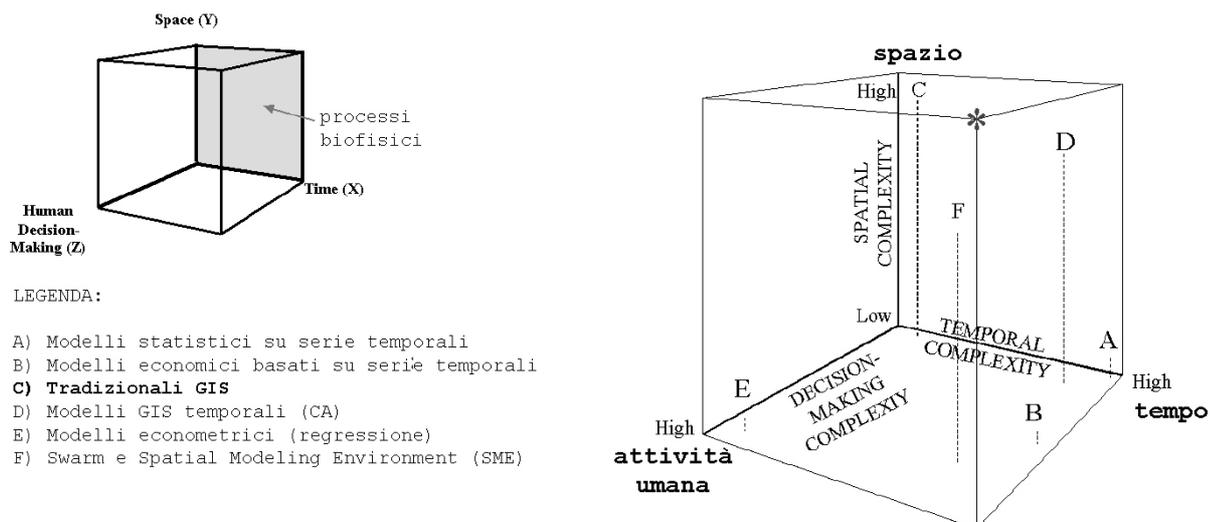


Fig 2-5: Rappresentazione 3D della complessità di un sistema. La struttura spaziale, temporale e umana è scelta come riferimento nella classificazione dei modelli di LUC. I processi naturali operano solamente sul piano (x,y). Lo studio dei sistemi complessi uomo-ambiente richiede l'integrazione della componente antropica, poiché le decisioni prese a livello economico-politico sono un fattore fondamentale nell'evoluzione dei cambi nel nostro pianeta. [Immagine adattata da Agarwal et al., 2002].

La formulazione di un indice di complessità spaziale permette di dividere i modelli in spazialmente rappresentativi e spazialmente interattivi. I primi hanno la capacità di visualizzare i dati in 2D o 3D ma non tengono conto delle relazioni topologiche fra gli elementi; i secondi definiscono esplicitamente le relazioni spaziali e le loro interazioni (in un intorno, ad es.) nel tempo.

Modelli con un basso indice di complessità temporale hanno un solo passo temporale e analizzano fenomeni di breve durata; un indice medio caratterizza modelli che possono adottare differenti passi temporali nello studio di processi di lunga durata. Infine i modelli con coefficiente elevato, hanno la capacità di rappresentare sistemi in cui compaiono azioni retroattive (*feedback*) e lags temporali, e di integrare dati raccolti su scale temporali differenti all'interno di un lungo periodo.

L'indice di complessità antropica (HDM¹⁹) descrive la capacità di un modello di catturare i processi decisionali nella rappresentazione del sistema uomo-ambiente. Un valore basso di questa grandezza indica che il modello non è stato sviluppato per tener conto dei processi economico-amministrativi che si nascondono dietro ad un cambiamento, mentre un HDM elevato caratterizza modelli che possono maneggiare agenti multipli che interagiscono attraverso domini differenti.

[Brown et al., 2004] concentrano la loro attenzione sui modelli di LUCC spazialmente espliciti su scala locale. Gli approcci modellizzativi sono divisi in due categorie: modelli empirici e modelli di simulazione dinamica. I primi sono procedimenti induttivi, poiché cercano di descrivere i processi basandosi esclusivamente sui dati raccolti in un contesto geografico e temporale preciso. Questa categoria comprende i metodi statistici di stima (come la funzione di regressione logistica), i modelli econometrici e i processi di Markov. I modelli dinamici sono costruiti per rappresentare al meglio un sistema complesso, ponderando in maniera opportuna i risultati statistici e l'opinione degli esperti sui processi di LUCC. *"They deemphasize the fitting of data and emphasize the fidelity of model elements and processes to what is known about the processes"*²⁰ [Brown et al., 2004, Pag.8]. A questa categoria appartengono, fra gli altri, i modelli Automa cellulare (CA) e Multi Agente (MA).

[Coquillard et al., 1997] riconoscono 3 tipi di processi modellizzativi: i modelli analitici, i modelli puramente stocastici ed i modelli di simulazione (Fig.2-6).

19 Human decision making: preparazione e adozione di decisioni

20 I modelli dinamici attribuiscono meno importanza all'implementazione dei dati, enfatizzando la fedeltà degli elementi del modello alle nostre conoscenze sui processi osservati.

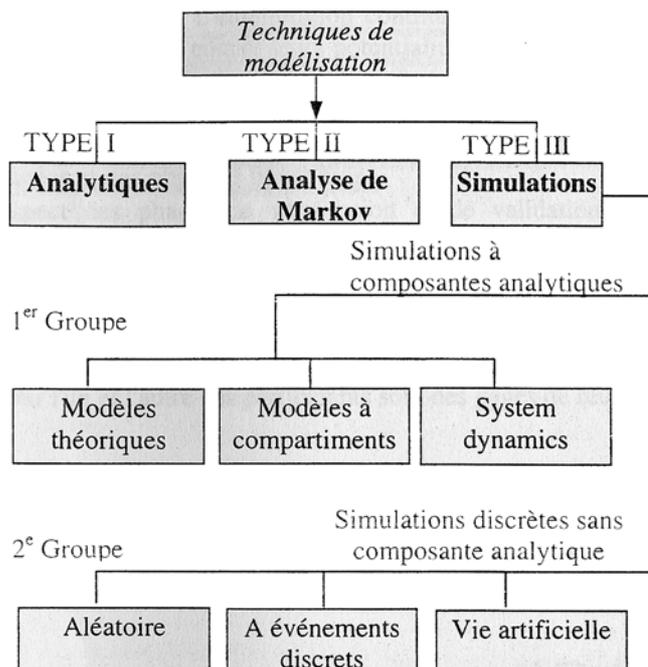


Fig 2-6: Tassonomia dei metodi di modellizzazione proposti da Coquillard e Hill. I modelli sono divisi in analitici, stocastici e di simulazione. I primi sono deterministici e continui nello spazio e nel tempo. I secondi includono una componente probabilistica nella rappresentazione del sistema. I modelli di simulazione, infine, si dividono a loro volta in metodi a componenti analitiche e metodi puramente discreti. [Immagine da Coquillard et al., 1997].

I modelli analitici o deterministici continui, utilizzano un formalismo matematico puro per rappresentare un processo. Un metodo continuo nello spazio e nel tempo si basa su delle equazioni differenziali che possono essere risolte dopo opportune semplificazioni del sistema.

Il secondo tipo di modelli introduce una componente probabilistica nella rappresentazione del sistema; l'analisi di Markov è un semplice esempio di questa categoria di metodi modellizzativi.

Infine ci sono i procedimenti di simulazione, divisi in modelli a componenti analitiche e modelli puramente discreti. Quest'ultimi prevedono una discretizzazione spazio-temporale del fenomeno osservato, che si realizza dunque in un intervallo di tempo specifico con una frequenza variabile (funzione di una legge di distribuzione) o fissata dal modellizzatore in accordo con le caratteristiche evolutive del sistema (ore, giorni,

etc.). Coquillard e Hill concentrano la loro analisi sulla modellizzazione degli ecosistemi naturali, tralasciando ogni considerazione sull'importanza di integrare la componente decisionale umana nello studio dei LUCC.

La più completa revisione dei modelli di uso del suolo è, a mio avviso, quella proposta da [Briassoulis, 2000]. L'autrice, dopo aver considerato i differenti approcci di categorizzazione presenti in letteratura, opta per una suddivisione dei modelli di LUCC basata sulla loro "modeling tradition"²¹. In base a questo criterio, che aggrega caratteristiche strutturali e tecniche risolutive, Briassoulis individua 4 classi di modelli di LUCC: i modelli statistici ed econometrici, i modelli di interazione spaziale, i modelli di ottimizzazione ed i modelli integrati. All'interno di ogni categoria i differenti approcci modellizzativi sono poi analizzati in base al loro obiettivo (descrizione, pianificazione, analisi di impatto, etc.), al livello di aggregazione spazio-temporale e settoriale, al dinamismo (statici, quasi statici o dinamici), al fondamento teorico (formalismi adottati, relazioni evidenziate, etc.), alle specifiche di costruzione del modello (tipi di variabili, parametrizzazione, soluzioni tecniche, etc.), alla qualità e tipologia di dati utilizzati. Una parte del lavoro svolto da Briassoulis sarà riproposto ed approfondito con esempi nei prossimi capitoli.

[Parker et al., 2003] rivisita velocemente i principali modelli di LUCC e li confronta con il modello multi-agente ibrido MAS²², evidenziando le qualità di quest'ultimo nella rappresentazione dei processi decisionali ed economici legati ai LUCC. Gli autori dividono i metodi di modellizzazione in 7 categorie, discutendo qualità e limiti di ogni approccio: modelli matematici, modelli di sistema, modelli statistici,

²¹ Tradizione modellizzativa

²² MAS: Multi Agent System for LUCC.

modelli basati sull'opinione di esperti, modelli evolventi (reti neurali, ad es.), modelli cellulari (Automa cellulare CA e modelli di Markov, ad es.), modelli ibridi (GEOMOD2, CLUE, simulazione spaziale dinamica, ad es.) e modelli multi-agente focalizzati sui processi antropici e il loro legame causale con i LUCC. Il metodo MAS proposto da Parker integra CA per rappresentare gli aspetti biofisici ed ecologici del sistema naturale, ed un modello multi-agente per evidenziare i fattori legati alle decisioni umane.

Vari esempi di modelli e di loro applicazioni nello studio di sistemi complessi si possono trovare nelle edizioni speciali delle riviste del settore; fra queste ricordiamo [Iovine et al., 2007; Koomen et al., 2007; Verburg et al., 2005; Veldkamp et al., 2001].

Nei prossimi capitoli descriveremo brevemente alcuni modelli di LUCC, soffermandoci su quelli che riteniamo abbiano maggiori potenzialità di sviluppo futuro. Questa revisione non vuole essere un contributo esaustivo sulle principali tecniche di modellizzazione, ma serve solamente per inquadrare il nostro lavoro in un determinato contesto metodologico.

2.3.2. I modelli statistici ed econometrici

Le tecniche statistiche sono comunemente utilizzate nello studio dei LUCC per evidenziare relazioni matematiche fra variabili dipendenti (uso del suolo, ad es.) e variabili indipendenti o predittive (distanza da una strada, altitudine, demografia, etc.).

Possiamo raggruppare questi metodi statistici in due classi: i modelli continui, che trattano l'uso del suolo come una variabile continua (superficie adibita ad un uso specifico) e modelli discreti, che considerano l'uso del suolo come una variabile discreta (differenti usi all'interno di un'area). La

regione di studio è normalmente suddivisa in parcelle, la cui forma e dimensione dipende dal livello di aggregazione scelto e dalla disponibilità di dati [Briassoulis, 2000].

In un modello continuo, per ogni zona, la distribuzione delle tipologie di uso del suolo (variabile dipendente) ed il valore delle variabili indipendenti sono implementati attraverso una regressione multipla, la cui forma generale è:

$$LUT_i = a + \beta_1 x_1 + \dots + \beta_n x_n + \varepsilon \quad (1)$$

dove "LUT_i" è l'area caratterizzata dall'i-esimo tipo di uso del suolo, "X₁...X_n" sono i valori delle variabili predittive ed "ε" rappresenta l'errore del modello statistico. Il modulo CHANGE del modello CLUE (© The CLUE Group) utilizza una regressione lineare per stimare il cambiamento dell'area occupata da un uso specifico del suolo, dovuto all'evoluzione del contesto socio-economico ed ambientale.

I modelli statistici discreti descrivono differenti tipologie di uso del suolo in funzione di un certo numero di variabili indipendenti continue o categoriche. La regressione multipla (MR) è probabilmente lo strumento più utilizzato nei modelli statistici per legare le forze motrici dei LUCC ipotizzate, ai patterns spaziali osservati; *"MR is most useful for empirically testing theoretical relationships among a set of variables, conducting rigorous hypothesis testing and ranking the relative contribution of factors"*²³ [Munroe et al., 2007, Pag.523]. I modelli basati sulla regressione lineare risultano inadatti per lo studio dei sistemi complessi uomo-natura, poiché le dinamiche ed i processi osservati tendono ad essere non lineari.

23 La regressione multipla è più utile per testare empiricamente le relazioni teoriche fra variabili, condurre un test rigoroso sulle ipotesi effettuate e classificare il contributo relativo di ogni fattore.

La natura spaziale dell'informazione può causare incertezze ed errori durante l'analisi statistica: "two problems arise when sample data has a location component: 1) spatial dependence exists between the observations and 2) spatial heterogeneity occurs in the relationships we are modeling"²⁴ [LeSage, 1999, Pag.2]. Entrambi queste grandezze devono essere valutate ed eventualmente corrette con opportune tecniche di statistica spaziale [Nelson et al., 2002; Anselin, 2002; LeSage, 1999]. Infine evidenziamo come, nel caso si utilizzi la MR per lo studio delle relazioni causali in un sistema complesso, sia necessaria un'analisi del coefficiente di multicollinearità [Serneels et al., 2001] per valutare il grado di correlazione fra le variabili indipendenti.

[Munroe et al., 2007 & 2004] utilizzano una regressione logistica multinomiale per studiare la traiettoria dei LUCC (8 classi di uso del suolo) in una regione dell'Onduras, in funzione di tre variabili indipendenti. Questo lavoro combina analisi di uso del suolo, valutazione dei patterns spaziali e tecniche econometriche per evidenziare e quantificare i fattori che hanno causato i LUCC.

Pendenza e altitudine sono considerati come i fattori geofisici più importanti per l'agricoltura mentre la distanza dai mercati regionali e locali, attraverso differenti tipi di strade, rappresenta una determinante fondamentale nell'utilizzo di una certa parcella.

Gli autori stabiliscono che un'area sarà soggetta all'utilizzo che apporterà il maggior beneficio economico al proprietario; il rendimento è espresso come funzione della distanza dai mercati e dalle caratteristiche geofisiche, secondo la formula:

24 Due problemi possono derivare dall'analisi di dati georeferenziati: 1) una dipendenza spaziale esiste fra le osservazioni, 2) un'eterogeneità attraverso lo spazio (ed il tempo) si verifica nelle relazioni che modellizziamo.

$$\ln(R_{ik}) = \alpha_{0k} + \alpha_{1k}D_i + \alpha_{2k}G_i \quad (2)$$

dove "R" rappresenta la rendita che deriva dal k-esimo uso del suolo alla locazione "i", "D" il costo del trasporto fra "i" ed il mercato, e "G" le caratteristiche geofisiche. L'equazione (2) è econometricamente stimata attraverso la formula:

$$\ln(R_{ik}) = \alpha_{0k} + \alpha_{1k}D_i + \alpha_{2k}G_i + \dots + u_{ik} \equiv A_k Z_i + u_{ik} \quad (3)$$

dove "Z" è il vettore delle variabili indipendenti, "A" il vettore dei parametri e "u" il fattore di disturbo (errore) con distribuzione aleatoria Weibull [Chomitz et al., 1996].

Per stimare la (3) supponiamo che una parcella sia adibita all'uso che apporti la maggior rendita al proprietario, cioè in un punto "i" avremmo un uso "k" se

$$R_{ik} > R_{ij}, \forall j \neq k \quad (4)$$

La (3) può essere espressa dunque da un modello logit multinomiale [Chomitz et al., 1996, Pag.493]:

$$\text{Prob}(i \text{ adibito a } k) = \frac{\exp(Z_i A_k)}{\sum_j \exp(Z_i A_k)} \quad (5)$$

dove la probabilità che la cella "i" sia adibita all'uso "k" è stimata relativamente a tutti i possibili utilizzi "j" con $j \neq k$. La (5)²⁵ può essere normalizzata ed espressa sottoforma di odds²⁶:

$$\ln[\text{Prob}(i \text{ adibito a } k) / \text{Prob}(i \text{ adibito ad altro uso})] \equiv A_k Z_i \quad (6)$$

I valori stimati "A_kZ_i" possono essere dunque interpretati come l'influenza delle variabili esplicative sulla probabilità che un uso "k" si realizzi nel punto "i" [Munroe et al., 2004].

25 L'espressione exp() indica l'esponenziale.

26 Odds: "è definito come il rapporto fra la probabilità che un evento accada sulla probabilità che esso non accada...il log(odds) è detto logit" [Bonham-Carter, 1994, Pag.306].

Le probabilità di LUCC così calcolate sono integrate in un modello dinamico (o ibrido) che consideri anche i fattori non registrati nei dati analizzati (un'evoluzione del contesto socio-economico), per proiettare l'evoluzione dei pattern di un processo nell'immediato futuro.

*"The statistic models allows us to predict where land cover conversions are most likely to take place in the near future. The models do not address the question "when are land cover conversions likely to take place?" ...they suggest some driving force to explain these changes"*²⁷ [Serneels et al., 2001, Pag.79].

Lo sviluppo di nuovi approcci basati sulla modellizzazione Bayesiana gerarchica sembrano permettere il superamento dei problemi legati alla dipendenza spazio-temporale nell'informazione, alle differenti risoluzioni dei dati ed alla propagazione dell'errore dovuta all'incertezza dalle differenti fonti d'informazione [Eckert et al., 2007; Yang et al., 2005].

2.3.3. I modelli dinamici spaziali

2.2.3.1. Introduzione

*"Models should be able to generate reliable projections into the future or to the past. This requires linking dynamically the processes (and models) of LUCC to biophysical processes (and models) to represent biophysical feedbacks to LUCC and the LUCC adaptations to biophysical changes"*²⁸ [Veldkamp et al., 2001, Pag.2].

27 I modelli statistici ci permettono di prevedere dove i cambiamenti di copertura sono più probabili nel breve termine. I modelli non rispondono alla questione "quando si realizzerà un LUCC?"... loro ci permettono di evidenziare le forze motrici che influenzano i cambi.

28 I modelli devono essere in grado di generare una previsione realistica nel futuro o nel passato. Questo richiede un legame dinamico fra processi (e modelli) di LUCC e processi biofisici (e modelli), per rappresentare le risposte biofisiche ai LUCC e l'adattamento dei LUCC ai cambi biofisici.

I modelli dinamici spaziali (o spazialmente espliciti) consentono di rappresentare le relazioni fra agenti, organismi ed ambiente, in maniera più completa rispetto ai metodi statistici. La calibrazione e la validazione del modello si basano su dati storici, utili ma spesso non sufficienti per descrivere l'evoluzione dei processi osservati, soprattutto quando i drivers ed i fattori esplicativi cambiano nel tempo [Pontius et al., 2004]. La semplice analisi statistica dell'informazione viene sostituita da una rappresentazione più fedele agli elementi ed ai processi del sistema [Brown et al., 2003], basata anche su dati aggiuntivi o sull'esperienza del modellizzatore.

Secondo [Almeida, 2003] un modello si dice dinamico se possiede una componente temporale esplicita, se i suoi input ed output variano con il tempo e se il suo stato dipende dallo stato anteriore. Il termine "spazialmente esplicito" caratterizza modelli che cercano di descrivere e prevedere l'evoluzione di attributi ambientali in sub-unità con distinta localizzazione e configurazione.

[Briassoulis] evidenzia l'importanza dei modelli spazialmente espliciti nello studio dei LUCC: *"it is generally preferable to have spatially-explicit (or georeferenced) models which produce more reliable land use changes as they reflect the actual spatial variability of various constraints on land transformations... the building of spatially explicit models which can capture the intricate interrelationships in space between the activities modeled as well as important economic-environmental interactions"*²⁹ [Briassoulis, 2000, Pag.46 e 99]. Questi modelli si focalizzano spesso su singole parcelle od

²⁹ È preferibile lavorare con modelli spazialmente espliciti (o georeferenziati) che producono una rappresentazione più realistica dei LUCC, poiché considerano la variabilità spaziale dei fattori che influenzano i cambiamenti del paesaggio... i modelli spazialmente espliciti catturano le intricate inter-relazioni nello spazio fra le attività modellizzate e le interazioni economiche ed ambientali fondamentali.

aree che corrispondono ad unità amministrative (NUTS II, ad es.)³⁰, permettendo così di legare micro-economia ed ecosistema locale nell'analisi dei Lucc. Altre volte i risultati ottenuti su differenti livelli di dettaglio spazio-temporale devono essere integrati attraverso operazioni di "downscaling" o "upscaling"³¹ [Printz et al., 2003], scegliendo opportunamente la risoluzione che evidenzia i principali processi in atto.

Proponiamo qui alcuni esempi di modelli spaziali dinamici comunemente utilizzati nello studio dei cambiamenti di uso del suolo: i modelli basati su Automa Cellulare, i Modelli Multi Agente, i modelli basati su Reti Neurali ed i Modelli Ibridi.

2.2.3.2. I modelli basati su Automa Cellulare (CA)

"Cellular automata are mathematical idealizations of physical systems in which space and time are discrete, and physical quantities take on a finite set of discrete values. A cellular automaton consists of a regular uniform lattice (or array), usually infinite in extent, with a discrete variable at each site (cell). The state of a cellular automaton is completely specified by the values of the variables at each site. A cellular automaton evolves in discrete time steps, with the value of the variable at one site being affected by the values of variables at sites in its neighbourhood on the previous time step. The neighbourhood of a site is typically taken to be the site itself and all immediately adjacent sites. The variables at each site are updated simultaneously (synchronously), based on the values of the variables in their

30 NUTS: Nomenclatura di unità territoriali per l'analisi statistica. Eurostat, direttiva europea.

31 Downscaling: diminuzione della scala, da macro a micro, aumento del dettaglio. Upscaling: aumento della scala, da micro a macro, con perdita di dettaglio.

neighbourhood at the preceding time step, and according to a definite set of local rules"³² [Wolfram, 1983, Pag.603].

Partendo da questa prima definizione di Wolfram, cerchiamo di riassumere le principali caratteristiche di un automa cellulare:

- ✓ Le cellule sono disposte in una matrice regolare a N-dimensioni.
- ✓ Ogni cellula adotta uno stato fra un numero finito di alternative.
- ✓ Il tempo è discretizzato.
- ✓ Le cellule cambiano secondo regole locali.
- ✓ Le stesse regole sono applicate a tutte le cellule.
- ✓ Ad ogni passo temporale le cellule sono aggiornate, simultaneamente o in maniera sequenziale asincrona [Hegselmann et al., 1998].

Ogni interazione in un CA si realizza all'interno di un intorno spaziale o vicinaggio, definito dal modellizzatore in base alle caratteristiche del processo studiato e alle capacità di elaborazione del proprio calcolatore elettronico (Fig.2-7).

32 Un automa cellulare (CA) è un'astrazione matematica del sistema fisico, in cui spazio e tempo sono discreti, e gli attributi assumono anch'essi un congiunto di valori discreti. Un CA è formato da un campo matriciale regolare infinito, con una variabile discreta in ogni punto (cellula). Lo stato dell'automata è completamente specificato dai valori delle variabili in ogni cellula. Un CA evolve in passi temporali discreti, ed il valore della variabile in ogni cellula è influenzato dal valore delle cellule vicine, durante i passi temporali precedenti. L'intorno (vicinaggio) di una cellula è composto dalla stessa e dalle cellule immediatamente adiacenti. Le variabili in ogni cellula sono attualizzate simultaneamente (sincronicamente) basandosi sul valore delle variabili nell'intorno scelto e nell'istante precedente, in accordo con un congiunto di regole locali predefinite.

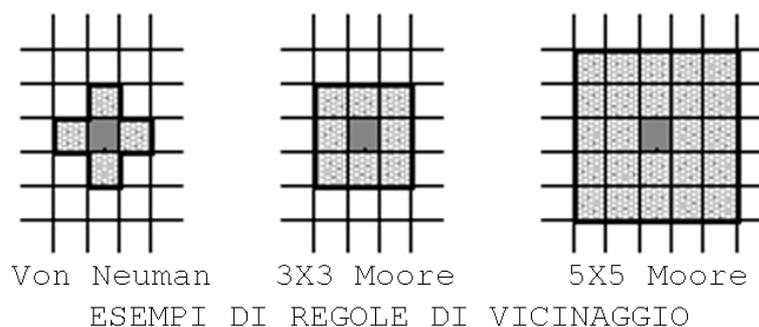


Fig 2-7: Differenti regole di vicinaggio. La scelta del vicinaggio deve tener conto dei patterns spaziali dei processi studiati e dei tempi di calcolo del proprio computer. [Immagine adattata da Hegselmann et al, 1998].

L'estensione delle relazioni spaziali catturate da un vicinaggio specifico, può non riflettere l'estensione reale dei processi in un sistema (problema di "spatial orientedness"). In questo caso si fa ricorso a regole di transizione non uniformi, che possono dinamicamente cambiare l'intensità e la configurazione delle connessioni fra le cellule; questo tipo di approccio sembra più adatto a catturare la complessità dei LUCC, caratterizzata da un'eterogeneità di agenti che agiscono su scale spaziali differenti.

In un CA le leggi fisiche che descrivono un fenomeno naturale sono tradotte in regole locali ed un macro-processo suddiviso in micro-interazioni o processi elementari [D'Ambrosio et al., 2006]. Questo permette di evidenziare le proprietà emergenti che derivano dall'interazione sinergica dei meccanismi e delle decisioni prese a livello locale; il CA può essere considerato come un sistema multi agente, basato su regole di vicinanza e su interazioni sovrapposte [Hegselmann et al., 1998].

Le attività umane sono spesso descritte in maniera semplicista, facendo coincidere un agente con una cellula della matrice e studiando l'impatto che questo ha nel proprio intorno, attraverso le regole locali di transizione. Questa approssimazione può non rispecchiare la variabilità spaziale dei processi in atto, offrendo dei risultati di utilità

limitata nella presa di decisioni politico-amministrative e nella pianificazione socio-economica sul lungo termine. [Irwin et al., 2001, Pag.10] evidenziano gli errori che si possono commettere nel considerare solamente le interazioni locali, all'interno del vicinaggio, per descrivere un LUCC: *"...A whole host of features that create extensive spatial heterogeneity across the landscape will drive actual changes in land use pattern and therefore, conclusive statements about the interaction causing the changes in actual urban form are misleading"*³³ .

È necessario utilizzare un insieme di regole complesse e gerarchiche per differenziare le decisioni e le attività antropiche su scale differenti, che interessano gruppi di cellule di dimensioni variabili.

Esistono modelli basati su CA che permettono di descrivere fenomeni dinamici e non-lineari su grande scala. Il modello SCIARA $\gamma 2$ (Simulation by Cellular Interactive Automata of the Rheology of Aetnean lava flows), ad esempio, si basa su un automa cellulare per simulare le colate di lava e detriti dall'Etna [Avolio et al., 2006]. La regione di studio è suddivisa in un congiunto di esagoni regolari; ogni cellula del vicinaggio è caratterizzata da un indice che la distingue dalle altre e permette di considerare eventuali anisotropie (Fig.2-8).

33 Un intero congiunto di caratteristiche, che definiscono l'eterogeneità spaziale estesa attraverso il paesaggio, possono influenzare i cambiamenti attuali nei patterns di uso del suolo, e dunque una valutazione conclusiva sulle interrelazioni causali nei cambi urbani può essere ingannevole (se basata solo sulle regole di transizione locale e sulle interazioni fra cellule adiacenti).

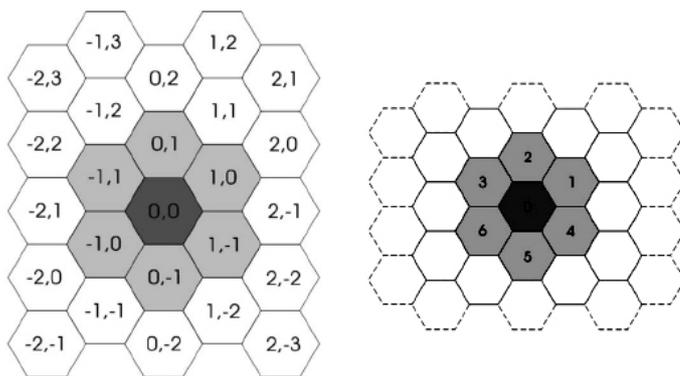


Fig 2-8: Il modello SCIARA $\gamma 2$ utilizza un CA per simulare colate di lava dall'Etna. A sinistra: il modello adotta un congiunto di esagoni regolari per discretizzare la regione in cui si evolve il fenomeno studiato. A destra: il vicinaggio esagonale adottato da SCIARA $\gamma 2$. Gli indici sono attribuiti alle celle per identificarle: 0 per la cella centrale, 1...6 per le restanti. [Immagine adattata da Avolio et al., 2006].

Alcuni formalismi devono essere adottati per definire questi fenomeni macroscopici attraverso un CA [Di Gregorio et al., 1999]. Le regole di transizione sono specificate da serie di equazioni che descrivono i processi elementari studiati e le loro interazioni (il flusso di lava, la sua progressiva solidificazione, etc.). Per un approfondimento di questi temi si rimanda a [Avolio et al., 2006; D'Ambrosio et al., 2006].

Per spiegare le caratteristiche essenziali di un approccio basato su CA presentiamo un semplice modello di "dinamiche d'opinione" [Hegselmann et al., 1998]. Supponiamo di discretizzare lo spazio attraverso un reticolato 2D e di scegliere un intorno di von Neumann (Fig.7). Ogni cellula è occupata da un individuo che possiede una propria opinione; essa rappresenta lo stato della cellula. Le opinioni appartengono ad un congiunto finito di alternative possibili, detto spazio degli stati, scelto nell'intervallo di numeri reali $[0,1]$. Ogni individuo attualizza il proprio stato considerando l'opinione dei propri vicini (ogni cellula del vicinaggio ha lo stesso peso) secondo la formula:

$$\frac{1}{\#N_i} \sum_{j \in N_i} u_j(t) \rightarrow u_i(t+1)^* \quad (7)$$

dove " u_j " rappresenta l'opinione di un individuo " j " nel periodo " t ", " $\#N_i$ " il numero di vicini dell'individuo " i " che deve prendere la decisione futura " $u_i(t+1)$ ".

Ovviamente la (7) restituisce un numero infinito di possibili opinioni; alcune correzioni devono dunque essere apportate per ottenere uno spazio finito di stati. Utilizziamo una funzione di discretizzazione che permette di trasformare l'opinione " $u_i(t+1)^*$ ", che non compare fra le possibili opzioni, in " $u_i(t+1)$ " appartenente allo spazio di stati predefiniti:

$$\frac{1}{\#N_i} \sum_{j \in N_i} u_j(t) \rightarrow \text{step function} \rightarrow u_i(t+1) \quad (8)$$

La (8) è anche chiamata regola universale di transizione [Hegselmann et al., 1998].

Il modello SpaCelle [Dubos-Paillard et al., 2003] è stato costruito per studiare l'evoluzione urbana basandosi sul paradigma della concorrenza spaziale fra diverse sottopopolazioni cellulari in interazione con il loro ambiente. Ogni individuo cellulare utilizza la propria forza vitale, che varia dalla nascita fino alla morte, per far fronte alle difficoltà che derivano dall'azione di altri individui (condizioni ambientali) che agiscono su buffer circolari via via più ampi. Lo spazio del modello è rettangolare e bidimensionale, discretizzato in cellule quadrate o esagonali. Il meccanismo di transizione "T" fra gli stati possibili (dalla nascita fino alla morte) si basa su una serie di variabili esplicative (età, forza vitale, durata della vita prevista, etc.), e permette alla cellula " i " all'istante " t ", in condizioni ambientali " $a_i(t)$ ", di passare da uno stato " $s_i(t)$ " allo stato " $s_i(t+1)$ " nell'istante successivo. L'equazione che descrive questo meccanismo è rappresentata da:

$$s_i(t+1) = T(s_i(t), a_i(t)) \quad (9)$$

Questo meccanismo seleziona la lista di regole applicabili allo stato " $s_i(t)$ "; ogni regola stabilisce la forza " $F(a_i(t))$ " di transizione, in funzione delle condizioni ambientali della cellula " i " catturate dal vicinaggio scelto. Questa è espressa dalla formula [Dubos-Paillard et al., 2003, Pag.12]:

$$F(a) = \frac{\sum_i \lambda_i (\prod_j f_j(a_j))}{\sum_i \lambda_i}, \in \mathfrak{R} [0,1] \quad (10)$$

come combinazione di funzioni elementari d'interazione " f_j " (demografiche, di prossimità, etc.) ponderate dai pesi " λ_i ".

Il principio di concorrenza adottato permette alla transizione con maggior forza di realizzarsi a scapito delle altre; se la regola di conservazione è più forte la cellula mantiene il proprio stato, altrimenti cambia oppure, se tutte le forze sono nulle, passa allo stato di morte naturale.

I modelli basati su CA hanno trovato una ampia utilizzazione nel campo dei LUCC, dallo studio dell'espansione urbana [Barredo et al., 2007; Almeida, 2003], all'analisi delle dinamiche forestali [Alonso et al., 2000], dalla valutazione del rischio ambientale [Andretta et al., 2006] alla pianificazione economica [Irwin et al., 2001]. La loro relativa facilità di applicazione, la capacità di descrivere micro-processi locali ed i loro effetti emergenti, e la qualità dei risultati ottenuti hanno contribuito a questa grande popolarità nel mondo accademico. Alcuni miglioramenti devono ancora essere fatti, soprattutto per catturare la complessità e la variabilità spaziale delle relazioni uomo-natura attraverso differenti livelli di dettaglio. Ritorneremo ulteriormente sui modelli cellulari nel corso del terzo capitolo di questa tesi che approfondirà i metodi utilizzati nello studio delle dinamiche di LUCC in Peten.

2.2.3.3 I modelli Ibridi Integrati

I modelli ibridi combinano differenti tecniche modellizzative per offrire un risultato più realistico possibile e ampiamente accettato. Ogni sub-modello descrive le caratteristiche di un dominio disciplinare specifico ed elabora delle informazioni che saranno poi utilizzate da un modello generale come input per l'analisi dei LUCC.

Il progetto CCAT³⁴ della Comunità Europea [Farmer et al., 2007; Kristensen et al., 2004] si pone come obiettivo il monitoraggio degli effetti indotti dalla Politica Agricola Comunitaria sullo sviluppo economico e sull'equilibrio ambientale delle zone rurali all'interno dei 27 Paesi Membri. L'analisi si appoggia su diversi data base (Eurostat, FAOstat, etc.) che raccolgono informazioni a livello locale (strategie agricole di un coltivatore, principali entrate, mercati locali, etc.), regionale e nazionale (attuazione delle direttive europee da parte dello stato membro, politica nazionale, prezzi dei prodotti, etc.). Questi dati sono poi elaborati attraverso una serie di strumenti di analisi e modellizzazione che permettono di calcolare degli indicatori di impatto, economico, sociale o ambientale, utilizzati poi come termini di confronto fra i singoli Paesi o fra le regioni all'interno di uno stato membro.

Gli indicatori ambientali sono calcolati a partire da un modello econometrico CAPRI³⁵ ed un modello biogeochimico DNDC³⁶.

Il modello CAPRI [Junker et al., 2003] è stato sviluppato per analizzare i cambiamenti economico-sociali ed ambientali nel settore agricolo europeo dovuti alle riforme introdotte

34 CCAT: Cross-Compliance Assessment Tool (strumento di valutazione della cross-compliance stabilita dalla politica agricola comunitaria).

35 CAPRI: Common Agricultural Policy Regionalized Impact analysis (modello per l'analisi a scala regionale dell'impatto indotto dalla politica agricola comunitaria).

36 DNDC: Denitrification and Decomposition model (modello per analisi dei processi di denitrificazione e decomposizione)

dalla CAP³⁷, a livello regionale (NUTS II) e sul mercato nazionale. CAPRI è composto da una serie di sottomodelli che catturano i differenti elementi legati alle attività agricole, dalle regole di mercato alle strategie produttive, dalla domanda di un determinato prodotto alle quote e ai sussidi stabiliti dalla Commissione Europea. CAPRI calcola una serie di indicatori per valutare lo stato di salute di un Paese Membro, il suo grado di adempimento alle norme europee, lo sviluppo economico di un determinato settore agricolo e l'impatto ambientale (ciclo di N, K, P e produzione di gas ad effetto serra; protezione del suolo; etc.). Il modello è dunque utilizzato per simulare e comparare ex-ante, nel breve e medio termine, l'impatto delle politiche agricole a livello regionale e nazionale.

Il modello DNDC è un modello biogeochimico integrato, composto da 6 sottomodelli che rappresentano i cambiamenti pedologici, la crescita della vegetazione, i processi di decomposizione, nitrificazione, denitrificazione e fermentazione [Pathak et al., 2005] nel sistema studiato. Il DNDC utilizza dati raccolti su scala locale e regionale per ottimizzare la gestione del territorio e le pratiche agricole o forestali, calcolando l'emissione di gas ad effetto serra legata ai cambiamenti nell'uso del suolo e permettendo di pianificare eventuali strategie mitigative o di sviluppo sostenibile. Il modello lega in maniera dinamica i principali fattori che intervengono nella produzione di gas serra, come la temperatura o il PH, con i drivers ecologici che li influenzano direttamente, come il clima o le attività antropiche (Fig.2-9). Un cambiamento in un driver comporta dunque un effetto a catena che determina una modificazione nei fattori esplicativi, nelle reazioni geochimiche e dunque nella produzione dei gas serra.

37 CAP: Common Agricultural Policy (politica agricola comunitaria europea).

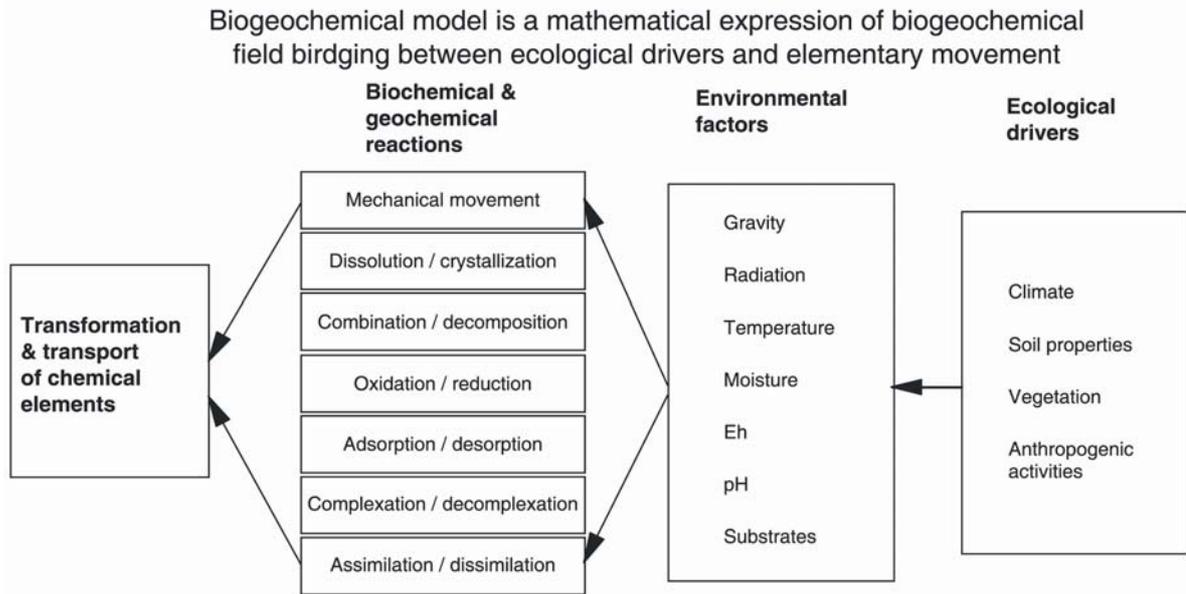


Fig 2-9: Relazioni fra fattori ambientali, drivers ecologici e reazioni biochimiche che determinano la produzione di gas serra. Il modello DNDC è strutturato a partire da questo schema concettuale. [Immagine da Li, 2007].

Il DNDC è stato ampiamente utilizzato per simulare l'emissione di gas serra nelle risaie asiatiche (Fig.2-10), sotto differenti scenari di gestione dell'acqua legati alle pratiche agricole (allagamento continuo o stagionale, parziale o totale dei campi di riso, ad es.). Queste informazioni ex-ante permettono di valutare l'impatto ambientale dei differenti usi del suolo e delle attività antropiche.

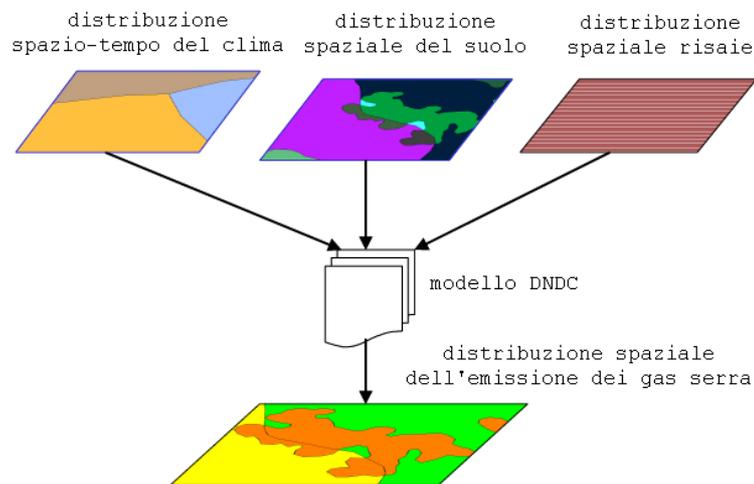


Fig 2-10: Procedimento di upscaling del DNDC nello studio dell'emissione di gas serra dalle risaie cinesi. [Immagine adattata da Pathak et al., 2005].

L'integrazione dei modelli CAPRI e DNDC permette dunque di avere una visione completa dell'impatto delle riforme proposte dalla CAP (economiche ed ambientali) sul mondo agricolo ed in natura, a livello regionale e per ogni Paese Membro della Comunità Europea.

[Rokityanskiy et al., 2007] propongono una descrizione del modello DIMA³⁸ per lo studio del peso di differenti strategie di gestione forestali sulla capacità di sequestrare CO₂ e sulla perdita di biomassa a livello globale (Fig.2-11).

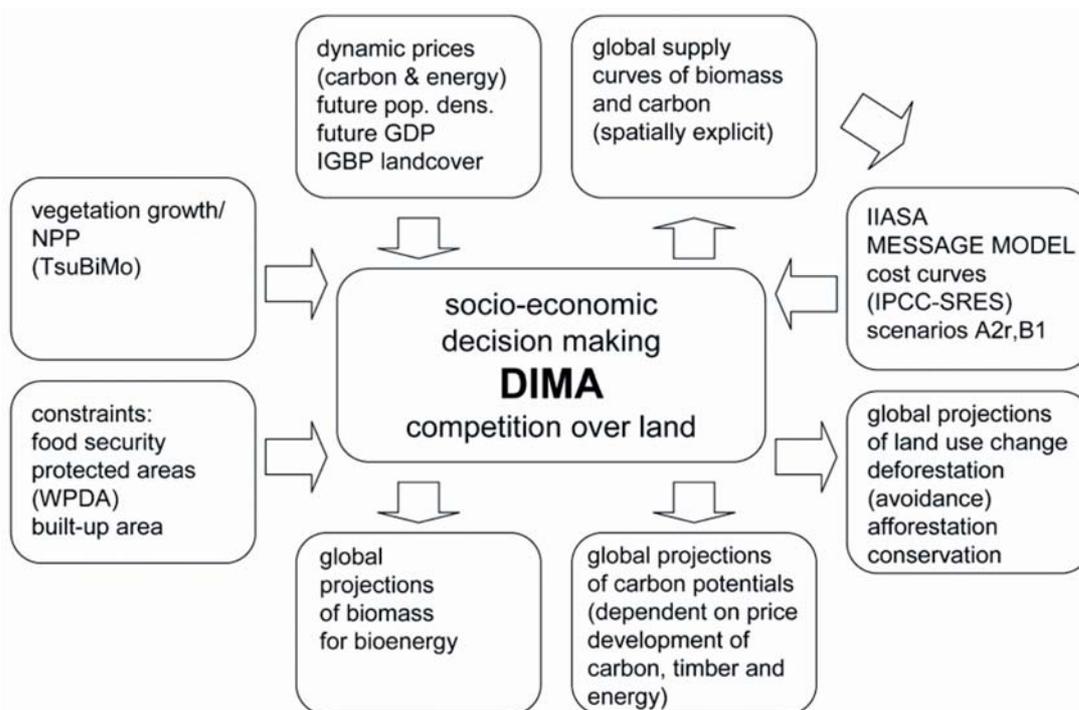


Fig 2-11: Rappresentazione dell'approccio modellizzativo integrato proposto da DIMA nello studio delle strategie di gestione forestale relativamente al sequestro di CO₂ e alla perdita di biomassa. [Immagine da Rokityanskiy et al., 2007].

Il modello specifica quale uso del suolo (afforestazione, deforestazione, preservazione del paesaggio, etc.) dovrebbe essere applicato in una regione, in base al costo della terra e della produzione forestale, alla potenzialità del terreno, alla

38 DIMA: Dynamic Integrated Model of forestry and Alternative land use (modello dinamico integrato per la gestione forestale e per l'analisi di un uso alternativo del suolo).

densità demografica ed allo sviluppo economico stimato. Il DIMA include il sottomodello MESSAGE³⁹, utilizzato per la pianificazione energetica in funzione di possibili scenari futuri di sviluppo economico-demografico. Il sottomodello TsuBiMo per lo studio della crescita di vegetazione a livello globale, permette di calcolare la produttività primaria netta (NPP) partendo dai dati sulle caratteristiche geofisiche, climatiche e di copertura globali. Il modello biofisico dunque si basa sulla scelta ottimale delle strategie di gestione forestale ed uso del suolo in base alla migliore alternativa da un punto di vista economico. Quattro scenari IPCC SRES sono utilizzati per descrivere possibili mondi futuri (differenti condizioni demografiche, sociali, etc.).

Come output il DIMA produce una previsione di 100 anni sull'utilizzo del suolo, sequestro di CO₂, impatto degli incentivi statali, politiche climatiche e richiesta energetica, dimostrando come eventuali politiche economiche basate sulle quote di CO₂ emesso-sequestrato possano essere un cardine portante di future strategie di mitigazione del clima e di sviluppo sostenibile.

2.2.3.4 I modelli Multi Agente

L'utilità e le potenzialità dei modelli multi agente sono ampiamente riconosciute all'interno dei progetti di ricerca sui cambiamenti globali e fra questi il Global Land Project (GLP).

Mentre i modelli basati su CA sono focalizzati sul paesaggio e sulle regole di transizione, i metodi multi agente danno un maggior risalto all'attività umana. "Agents are

39 MESSAGE: Model for Energy Supply Strategy Alternatives and their General Environmental impact (modello per la pianificazione energetica sotto differenti scenari rappresentativi della domanda in energia future)

crucial component in these models."⁴⁰ [Parker et al., 2003, Pag.317]. Un agente è caratterizzato da una certa autonomia d'azione, adotta una strategia che gli permette di ottenere un beneficio o di raggiungere uno scopo; egli opera su un ambiente, in relazione e comunicazione con altri attori che possono a loro volta influenzare o limitare la sua attività.

Un modello multi-agente per lo studio dei Lucc (ABM-Lucc⁴¹) è generalmente formato da due componenti [Berger et al, 2001a]. La prima parte è un modello cellulare che rappresenta lo spazio di studio e può essere basato su differenti tecniche di modellizzazione come CA o Markov. La seconda componente è un modello multi-agente (ABM) che rappresenta i processi decisionali e le attività umane.

Un ABM include una serie di regole che definiscono le relazioni fra gli agenti ed il loro ambiente, e marcano la sequenza temporale delle azioni. Ogni agente è rappresentato da un insieme di leggi che utilizza le informazioni, interne ed esterne, per influenzarne lo stato, le decisioni e le azioni.

Nello studio dei Lucc un agente può essere identificato con un coltivatore o un proprietario che sceglie come utilizzare il proprio terreno integrando valori culturali, esperienza personale, caratteristiche biofisiche dell'ambiente e contesto sociale (azioni e opinioni di altri agenti che operano in un vicinaggio). Un ABM-Lucc include, per ogni singolo agente, le regole che influenzano le sue decisioni, come la massimizzazione del profitto ad esempio, le relazioni con il proprio ambiente e le risorse disponibili. Contrariamente ai tradizionali approcci modellizzativi matematici o econometrici, un ABM non richiede la definizione di un insieme di condizioni di equilibrio, offrendo così un'elevata flessibilità che

40 Gli agenti sono gli elementi cruciali di questo approccio modellizzativo.

41 ABM-Lucc: Agent Based Model for Lucc (modello multi agente per Lucc).

permette di catturare l'eterogeneità dei processi e le complesse interdipendenze fra gli agenti e il loro ambiente.

Quando un ABM è combinato con un modello cellulare che definisce lo spazio in cui gli agenti operano, è possibile rappresentare esplicitamente i processi, la loro variabilità e distribuzione spaziale, le interazioni ed i fenomeni multi-scalari. La capacità di catturare la complessità di un sistema è probabilmente la qualità più rilevante di un ABM [Berger et al., 2001b]. Evidenziare le proprietà emergenti di un sistema, intese come il risultato sinergico di processi e componenti a micro-scala, richiede una comprensione profonda delle interazioni e dei feedback su differenti livelli di dettaglio. Un ABM può essere utilizzato come un "laboratorio sociale" per esplorare l'impatto di differenti strategie di uso del suolo sul paesaggio e sul sistema socio-economico.

[Anwar et al, 2007] analizzano l'influenza delle attività antropiche legate al commercio marino ed al turismo sulla vita marina nella regione di St.Lawrence Estuary, in Quebec. Un modello ABM è implementato su una piattaforma di modellizzazione RePast⁴² e diverse simulazioni sono prodotte per verificare l'impatto sull'ecosistema naturale di differenti strategie di turismo marino. Gli agenti scelti sono divisi in due categorie: le barche, considerate come un agente pensante in grado di percepire l'ambiente fisico in cui opera, di pianificare la propria attività e di avere memoria. Le balene sono rappresentate come una semplice entità che reagisce a stimoli esterni. Le regole di comportamento per ogni imbarcazione (Fig.2-12) sono sviluppate sulla base di osservazioni personali, relazioni scientifiche e data base SIG ed in funzione della massimizzazione del profitto (ovvero il tempo di osservazione di balene per i turisti).

42 RePast: Recursive Porous Agent Simulation Toolkit. Sviluppato dal laboratorio di Geo-informatica dell'Università di Calgary. Open source: <http://repast.sourceforge.net/download.html>.

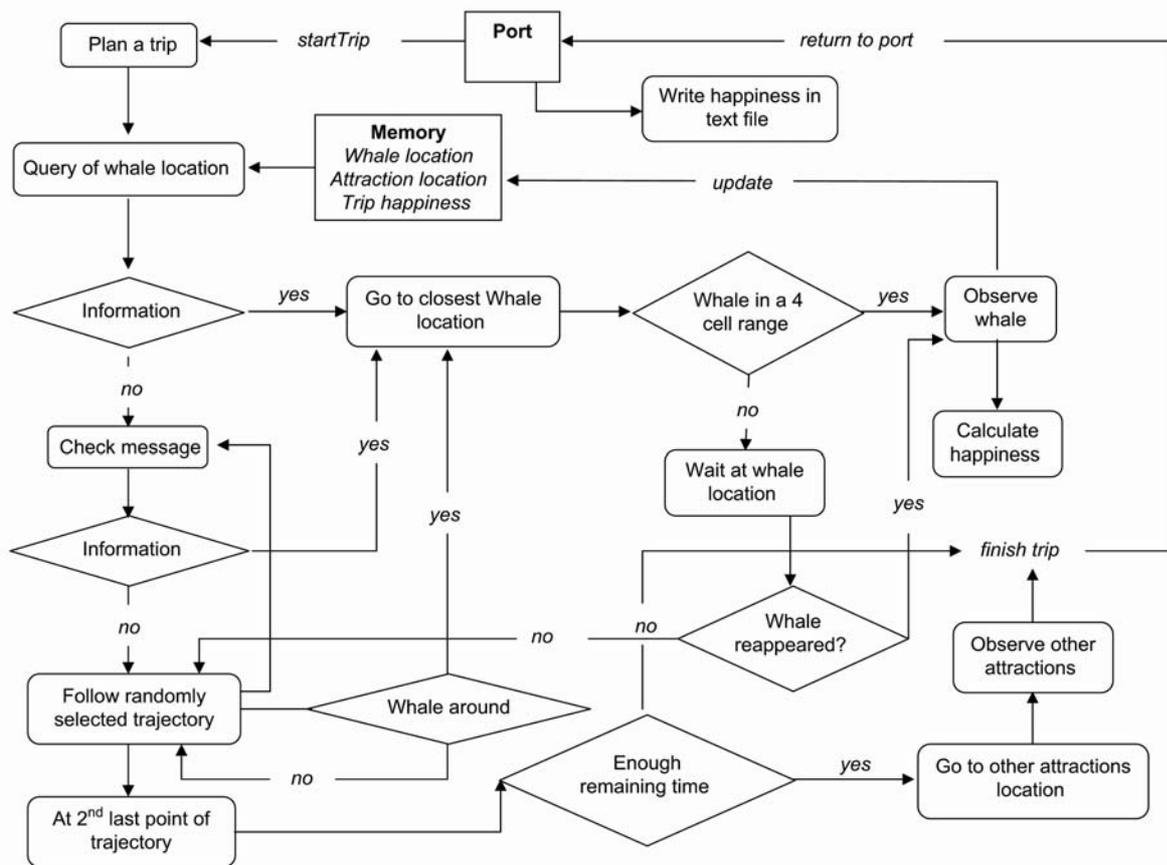


Fig 2-12: Regole di comportamento per un'imbarcazione definite in un ABM per lo studio dell'impatto del turismo marino sull'ecosistema. Una barca è considerata come un agente intelligente in grado di ottimizzare il proprio profitto, cioè il tempo di osservazione di una balena. Quest'ultima è rappresentata come un agente reattivo, che risponde a stimoli e perturbazioni esterne, dovute alle attività turistiche. [Immagine da Anwar et al., 2007].

Lo spazio fisico è rappresentato attraverso due mappe raster: la prima rappresenta i differenti usi del "suolo" (porto, acqua, campo, aree produttive, etc.); la seconda è una mappa batimetrica che registra la profondità delle acque e definisce sia i limiti di navigabilità sia l'habitat delle specie marine. Entrambi i raster hanno una risoluzione di $100 \times 100 \text{m}^2$.

Gli autori dimostrano l'utilità e le potenzialità di un ABM per valutare l'impatto di differenti strategie turistiche sull'ecosistema marino ed in generale sulle georisorse comuni.

I modelli ABM-LUCC hanno trovato un'amplia utilizzazione in vari campi applicativi, dalla gestione delle risorse agricole

[Rouchier et al., 2001] agli studi sulle dinamiche urbane [Torrens, 2002].

2.2.3.5 I Modelli Apprendisti

Per Modelli Apprendisti intendiamo qui tutti i metodi di modellizzazione che si basano su intelligenza artificiale, reti neurali e programmazione evolutiva. Questi modelli sono "allenati" con una serie di dati test per comprendere le relazioni input/output; sia le informazioni iniziali (stimolo) che il risultato finale sono noti ed introdotti nella modellizzazione. Una volta strutturato, il modello può essere utilizzato per calcolare mappe predittive nel breve-medio periodo sulla base dei nuovi dati (variabili esplicative, serie temporali, etc.).

Rimandiamo al capitolo 3 di questa tesi per una descrizione dettagliata del modello PNNET (reti neurali) sviluppato da Villa e Follador per lo studio dei LUCC in Peten.

2.2.3.6 I modelli basati su SIG

Questi metodi modellizzativi derivano dalla combinazione di strumenti e funzioni caratteristiche dei SIG (per rappresentare lo spazio ed integrare dati multiscalare) con moduli specifici sviluppati per catturare la dimensione temporale ed antropica del sistema e permettere dunque una rappresentazione dinamica dei processi in atto. GEOMOD2 [Pontius et al., 2006] e CLUE [Verburg et al., 2006; Veldkamp et al., 1996] sono due esempi molto noti di questa categoria.

I modelli CA_Markov, LCM e Dinamica EGO fanno parte di questo gruppo e verranno ampiamente descritti nel 3 capitolo di questa tesi.

CAPITULO No. 3

Descrizione dei modelli utilizzati nello studio
dei LUCC in Peten.

Moi, j'aime utiliser toutes les ressources de l'imagination, j'aime inventer la réalité et surtout la transformer. Le théâtre est absolument le contraire du réel, surtout quand il n'est pas une imitation du vrai mais son image transposée et paradoxale, ou excentrique. J'aime restructurer, réinventer la réalité, faire venir au jour des images des choses qui ne soient pas courantes, qui ne soient pas banalement inscrites dans nos cerveaux mais qui surprennent, pas en tant que jeu artificiel mais comme jeu de la raison et de l'intelligence.

(Dario Fo, 2004)

Introduzione generale

Negli ultimi anni sono stati messi in commercio decine di modelli di LUCC, caratterizzati da differenti approcci metodologici, perlopiù per rispondere alle esigenze di una determinata disciplina o problematica. Essi hanno dunque un limitato campo di applicabilità ed offrono buone performance solo all'interno del contesto operativo in cui sono stati sviluppati. Presentiamo in questo capitolo i quattro modelli predittivi utilizzati per lo studio delle dinamiche di uso del suolo in Petén, scelti per la loro economicità, per la flessibilità e la facilità d'utilizzazione, anche con un database limitato.

Il modello Predictive Neural Networks (PNNET), sviluppato da Villa e Follador in linguaggio [®]R, utilizza un approccio basato sulle reti neurali, che già in passato hanno dimostrato una grande capacità di catturare le relazioni nascoste e non lineari di un sistema complesso. PNNET è stato realizzato per offrire uno strumento gratuito, flessibile ed automatico per lo studio dell'evoluzione dei patterns spaziali di un fenomeno.

Il secondo modello è stato costruito in Dinamica EGO, un ambiente *freeware* di simulazione sviluppato dal Centro de Sensoriamento Remoto della UFMG, Brasile. Abbiamo scelto Dinamica EGO perchè appositamente progettato per lo studio delle dinamiche di uso del suolo in aree tropicali, e perché offre inoltre la possibilità di definire alcuni importanti parametri evolutivi di un processo (sojourn time, grado di isometria, etc.), caratteristica assente in altri metodi predittivi.

I modelli CA_Markov e Land Cover Modeler sono due moduli di Idrisi Andes (un SIG sviluppato da Clark-Labs, USA), e rappresentano i principali metodi predittivi utilizzati dal

gruppo di ricerca del GEODE/CNRS, per lo studio delle dinamiche di uso del suolo.

Le fasi che hanno portato alla creazione di PNNET e dei programmi in Dinamica EGO, sono qui descritti in maniera dettagliata. L'autore spera così di poter giustificare le scelte metodologiche adottate ed offrire un piccolo tutoriale per l'utilizzo di questi due modelli *freeware*.

3.1. PNNET: una rete neurale predittiva

3.1.1. Introduzione

Negli ultimi anni le reti neurali (ANNs¹) sono state frequentemente applicate allo studio di sistemi complessi, spesso con l'aiuto di dati satellitari. Le ANNs sono utilizzate per stimare informazioni tematiche ad una scala sub-pixel [Mertens et al., 2004; Foody, 1997], per classificare immagini [Joshi et al., 2006], per quantificare cambi di uso del suolo [Nemmour et al., 2006], per l'analisi di rischio [Kanungo et al., 2006], ed infine, per la previsione, sul breve periodo, delle dinamiche ambientali [Follador et al., 2008; Villa et al., 2007].

Una ANN è un insieme di unità computazionali semplici, dette neuroni, celle o nodi, legate da un sistema di connessioni, le sinapsi, che lavorano in parallelo per realizzare le trasformazioni input/output. Il numero di neuroni e connessioni definisce la topologia della rete. Un valore reale " W_{ij} ", detto peso, è attribuito ad ogni sinapsi fra due nodi " i " e " j ". Un nodo rappresenta una variabile di stato " x_i " ed è caratterizzato da un valore reale " W_i ", detto bias; " x_i " è una funzione del bias, dei pesi attribuiti alle connessioni e

1 ANNs: Artificial Neural Networks (reti neurali artificiali)

dello stato dei neuroni interconnessi [Kanewski et al., 2004], attraverso una funzione di attivazione.

In questa tesi descriveremo dettagliatamente un **Multi-Layer Perceptron**² (MLP), un modello unidirezionale ingresso-uscita, composto da 3 strati (Fig.3-1):

- Un Input: il numero di neuroni dipende dal numero dei dati in entrata, ad es. mappe tematiche.
- Un Output: il numero di neuroni dipende dai nostri obiettivi, ad es. le mappe di uso del suolo futuro.
- Uno strato intermedio nascosto: la sua dimensione e parametrizzazione è scelta in base ai risultati della fase di ottimizzazione della rete.

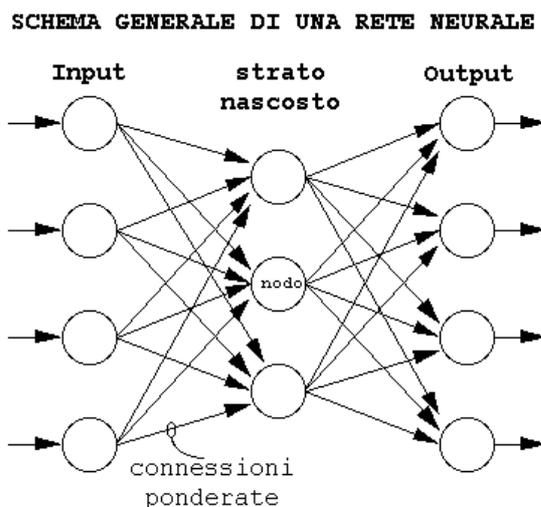


Fig 3-1: Schema di una rete neurale con un solo strato nascosto. Il modello PNNET è basato su un Perceptrone con un solo strato nascosto, sufficiente a garantire la risoluzione di quasi tutti i problemi non lineari, mantenendo semplicità strutturale e tempi di calcolo brevi.

Numerose reti neurali sono state "allenate" con le stesse informazioni in input ma diverse topologie, utilizzando un algoritmo di apprendimento assistito (error-backpropagation). La migliore ANN è stata selezionata in base a criteri

² Perceptrone multi-strato. Il primo modello di un neurone artificiale sufficientemente complesso per fornire dei comportamenti interessanti fu proposto da Rosenblatt nel 1962 e fu chiamato perceptrone.

statistici (scarto quadratico medio, ad es.) per minimizzare l'errore fra l'output simulato e l'output desiderato [Villa et al., 2007 ;Joshi et al., 2006; Lee et al., 2006]. La funzione "nnet" [Venables et al., 1999] è stata caricata nel linguaggio R[®] per la fase di apprendimento ed ottimizzazione della rete; l'implementazione del modello PNNET in R[®] ha portato allo sviluppo di tre sottoprogrammi che, applicati in ordine sequenziale, permettono di realizzare le simulazioni predittive.

3.1.2 Caratteristiche generali di una rete di neuroni artificiali

Una neurone artificiale è un modello matematico che simula un neurone biologico (Fig.3-2).

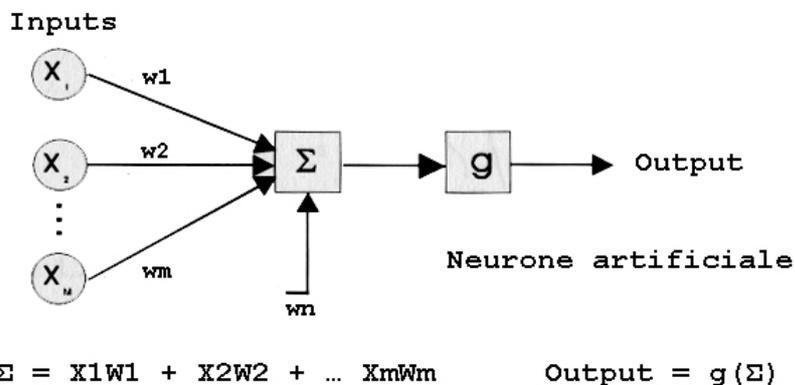
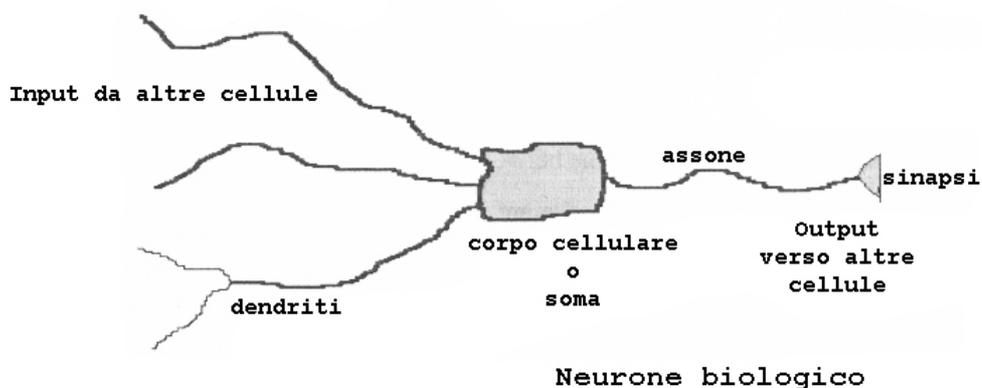


Fig 3-2: Modello di un neurone biologico e di un neurone artificiale ad una uscita. La "g" è detta funzione logistica di attivazione.

Un neurone artificiale è l'unità base di elaborazione dell'informazione in una ANN, ed è composto da tre elementi [Kanevski et al., 2004; Villa, 2002]:

- Una serie di cellule d'ingresso " $(X_i)_{i=0...m}$ " che corrispondono alle cellule biologiche che ricevono l'impulso.
- Una serie di sinapsi o connessioni, ognuna delle quali è caratterizzata da un certo peso " $(W_i)_{i=0, ...m}$ "; un valore positivo è associato ad una sinapsi eccitatrice ed un valore negativo ad una sinapsi inibitrice. Le informazioni in input verranno integrate attraverso una funzione additiva " \sum ".
- Una cellula d'uscita che rappresenta la risposta del perceptrone agli stimoli ricevuti ed è il risultato della somma ponderata degli input, limitata attraverso una funzione di trasferimento (o attivazione) " g ", di cui parleremo dettagliatamente in seguito.

Un neurone è dunque descritto matematicamente dalle seguenti formule [Mas et al., 2007]:

$$\begin{cases} v_k = \sum_{j=1}^m w_{kj} x_j + w_{0k} \\ y_k = \varphi(v_k) \end{cases} \quad (1)$$

dove " w_{0k} " è detto bias o soglia di attivazione.

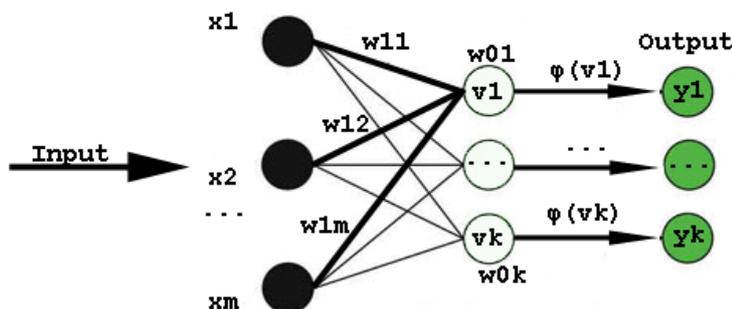


Fig 3-3: Rappresentazione grafica di una rete, in cui ogni neurone è descritto dalle equazioni (1).

Gli input ricevuti sono gli analoghi dei dendriti in un neurone biologico; essi sono combinati, spesso attraverso una semplice somma ponderata, per formare il livello di attivazione interno (Fig.3-3).

La funzione di attivazione " $\varphi(v)$ " definisce l'output di un neurone come combinazione lineare degli inputs " v ". Esistono differenti " $\varphi(v)$ " come ad esempio:

- la funzione lineare $\varphi(v)=v$, $\forall v \in \mathbb{R}$
- la funzione sigmoidea $\varphi(v)=\frac{1}{1+e^{-v}}$ (Fig.3-4)

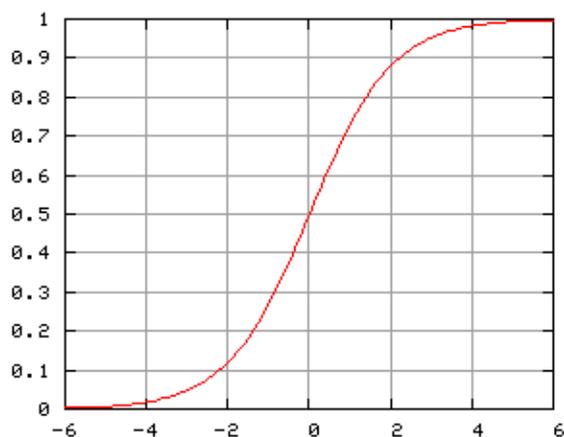


Fig 3-4: Rappresentazione grafica della funzione logistica sigmoidea.

- la funzione tangente iperbolica $\varphi(v)=\frac{e^v - e^{-v}}{e^v + e^{-v}}$

La funzione logistica di attivazione più utilizzata ha la forma:

$$\varphi(v) = \frac{1}{1 + e^{-av}} \quad , \quad \forall v \in \mathbb{R} \quad (2)$$

dove "a" è il parametro che determina la pendenza della curva; maggiore è il suo valore, più ripida sarà la curva. Le funzioni di attivazione assumono valori compresi fra [0,1] o [-1,1]; ciò permette di mantenere limitati e controllati i segnali che viaggiano all'interno della rete [Manca, 2005].

Una rete di neuroni artificiali è una collezione di semplici unità computazionali legate attraverso un sistema di interconnessioni. La disposizione dei neuroni nella rete, il numero di sinapsi e di nodi definiscono la **topologia** del ANN.

*"ANNs are parallel computational models comprised of densely interconnected adaptive processing units. These networks are fine-grained parallel implementations of nonlinear static or dynamic systems. A very important feature of these networks is their adaptive nature, where "learning by example" replace "programming" in solving problems. This feature makes such computational models very appealing in application domains where one has little or incomplete understanding of problem to be solved but where training data is readily available. Another key feature is the intrinsic parallel architecture that allows for fast computation of solution when these networks are implemented in customized hardware...The artificial neuron is the basic building block/processing unit of an ANN. It's necessary to understand the computational capabilities of this processing unit as a prerequisite for understanding the function on a network of such units"*³ [Hassoun, 1995, Pag.2].

³ Le reti di neuroni sono modelli che lavorano in parallelo e sono composte da unità di calcolo adattabili interconnesse. Le ANN sono delle implementazioni di sistemi non lineari statici o dinamici. La caratteristica più rilevante di una rete neurale è la sua natura adattabile che privilegia, per risolvere un problema, l'apprendimento con una serie di esempi piuttosto che l'attività di modellizzazione

Esistono due tipi di reti neurali:

- ✓ Le reti unidirezionali entrata→uscita (**feedforward**⁴): non esistono *feedbacks* ed il modello è considerato statico. Una volta calibrato, la sua risposta ad un input identico sarà sempre la stessa. I pesi ed i bias rimangono invariati. Il MLP è probabilmente la classe di reti unidirezionali più utilizzata.
- ✓ Le reti ricorrenti(**feedback**⁵): sono modelli dinamici. In questa classe l'output di un neurone è rimandato indietro, direttamente o passando attraverso i nodi intermedi, e riutilizzato come input. Tramite meccanismi di aggiornamento i pesi ed i bias sono modificati via via fino al livello di input. Questo ciclo si ripete fino a quando la rete non converge ad una soluzione.

3.1.3 L'apprendimento di una rete neurale

L'apprendimento di una rete neurale rappresenta la fase di messa a punto dei pesi di ogni connessione. Esistono due differenti regole di apprendimento/allenamento [Kanevski et al., 2004]: supervisionato e non supervisionato.

Apprendimento **supervisionato**: è il metodo più utilizzato. Innanzitutto bisogna costruire un insieme di dati con cui "addestrare" la rete (training set), formato da una serie di coppie (input-reale, output-reale) sperimentali. Successivamente un sottoinsieme di dati è scelto dal

vera e propria. Questa proprietà rende ANN particolarmente adatta all'applicazione su casi caratterizzati da una scarsa od incompleta conoscenza del sistema, e da una grande quantità di osservazioni e dati empirici. Un'altra caratteristica importante delle reti neurali è la loro architettura intrinseca parallela, che permette di calcolare velocemente le soluzioni, quando il modello è implementato su un computer... Il neurone artificiale è l'unità base di processamento di una ANN. È necessario comprendere le capacità di calcolo di questa unità per capire il funzionamento della rete.

4 All'avanti.

5 All'indietro.

precedente gruppo ed i suoi elementi "input[]" introdotti uno ad uno nella ANN. Ogni i -esimo risultato $output\text{-rete}(i)$ corrispondente all' i -esimo $input\text{-reale}(i)$ sarà confrontato con l' i -esimo $output\text{-reale}(i)$ desiderato. Dopo aver esaminato l'intero sottoinsieme di dati, la rete è aggiornata ed i pesi modificati; questa operazione ha lo scopo di minimizzare l'errore fra $output\text{-reale}$ ed $output\text{-rete}$ simulato (Fig.3-5). La fase che va dall'introduzione dei dati all'aggiornamento dei pesi è detta **epoca**.

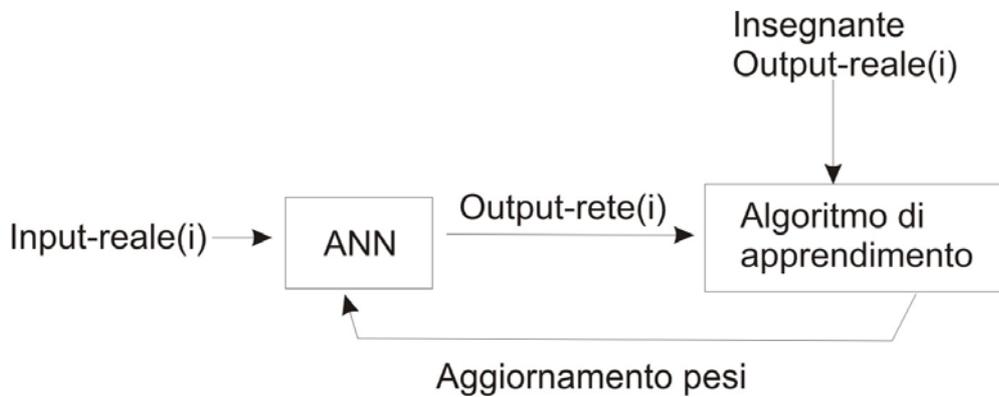


Fig 3-5: Rappresentazione grafica delle diverse tappe dell'apprendimento supervisionato.

Apprendimento **non supervisionato**: noi forniamo alla rete solo una serie di $input\text{-reale}[]$, poiché questo metodo non è accompagnato dall'insegnante ($output\text{-reale}$) e realizza una sorta di auto-apprendimento. I pesi variano in base ad una regola definita a priori e non in base all'errore $output\text{-reale}/output\text{-rete}$ (Fig.3-6).

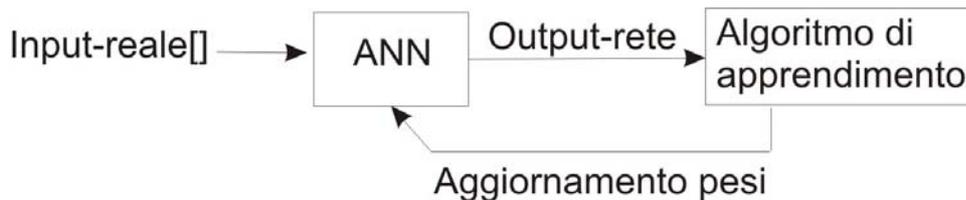


Fig 3-6: Rappresentazione grafica delle diverse tappe dell'auto-apprendimento.

Possiamo, ad esempio, utilizzare una rete formata da un input e da un livello che chiamiamo strato competitivo. Quest'ultimo consiste di neuroni che competono fra loro per meglio adattarsi (rispondere) alle caratteristiche contenute negli input-reali[]. "The network operates in accordance with a "winner-takes-all" strategy"⁶ [Kanevski et al., 2004, Pag.174]. La ANN con auto-apprendimento più famosa è la Kohonen's SOM⁷, sviluppata per identificare strutture e caratteristiche nascoste nei dati [Kohonen, 2006] e produrre una rappresentazione compressa e discretizzata, detta mappa, che preserva le relazioni topologiche dello spazio di input.

3.1.4 Il Perceptrone multi-strato (MLP)

Il perceptrone multi-strato (MLP⁸) è la classe più importante di ANN unidirezionali *feedforward* con allenamento supervisionato. Un MLP consiste di un insieme di input (strato di entrata), di nodi computazionali (strato o strati nascosti) e di nodi di output (strato di uscita). Un neurone su un livello "i" è sempre connesso al neurone sul livello successivo "i+1", senza possibilità di retroazioni.

L'analisi del comportamento dei nodi nello strato nascosto permette di meglio comprendere cosa è stato appreso dalla rete e con che grado di precisione.

Il MLP apprende attraverso un algoritmo supervisionato detto "*backpropagation algorithm*". I dati sono presentati alla ANN e propagati in avanti attraverso la rete fino alla produzione di un output; nella fase successiva di "*backpropagation*", l'output-rete è confrontato con l'output-reale desiderato ed i pesi della ANN sono corretti in accordo con una regola di minimizzazione dell'errore (Fig.5).

⁶ La rete opera secondo la strategia " il vincitore prende tutto".

⁷ SOM: Self Organising Maps

⁸ MLP: MultiLayer Perceptron

L'allenamento di un MLP è basato sulle seguenti tappe operazionali:

1. **Inizializzazione** dei pesi: dei numeri aleatori sono assegnati alle connessioni sinaptiche e ai nodi, usando una distribuzione random uniforme.
2. Elaborazione dei risultati (fase **forward**): alleniamo la ANN con un insieme di dati (Input-reale[], Output-reale[]) = T; per ogni coppia di T che indicheremo con "(x(n), d(n))", l'input è propagato attraverso la rete, strato dopo strato ("l"), come indicato dalle seguenti equazioni [Mas et al., 2007]:

$$\begin{cases} v_j^{(l)}(n) = \sum_{i=1}^{m_0} w_{ji}^{(l)} y_i^{(l-1)}(n) \\ y_j^{(l)} = \varphi_j(v_j^{(l)}(n)) \end{cases} \quad (3)$$

3. Calcolo del segnale d'errore per ogni neurone (fase **backpropagation**):

$$e_j(n) = d_j(n) - y_j(n) \quad (4)$$

dove "d_j(n)" rappresenta l'output-reale desiderato, e "y_j(n)" è l'output-rete per lo stesso neurone "j". Il gradiente locale della rete è calcolato utilizzando le seguenti equazioni:

$$\delta_j^{(l)}(n) = \begin{cases} e_j^{(L)}(n) \varphi_j(v_j^{(L)}(n)), \nabla \text{neurone } j \text{ nello strato di output } L \\ \varphi_j(v_j^{(L)}(n)) \sum_k \delta_k^{(l+1)}(n) w_{kj}^{(l+1)}(n), \nabla \text{neurone } j \text{ nello strato nascosto } L \end{cases} \quad (5)$$

dove "φ()" denota differenziazione rispetto

all'argomento.

I pesi sinaptici sono aggiornati secondo una "delta rule" ⁹ generalizzata:

$$w_{ji}^{(l)}(n+1) = w_{ji}^{(l)}(n) + \alpha[w_{ji}^{(l)}(n-1)] + \eta\delta_j^{(l)}(n)y_i^{(l-1)}(n) \quad (6)$$

dove la "**learning rate** η " è utilizzata per aggiornare i pesi fra due nodi " $i-j$ ". Il "**momentum** α " analizza la configurazione anteriore dei pesi per stabilire in che direzione ricercare il minimo errore e garantisce che l'algoritmo "backpropagation" operi velocemente quando un numero di passi consecutivi cambia i pesi nella stessa direzione.

Le fasi *forward* e *backpropagation* sono ripetute per ogni esempio " (n) " appartenente a T , fino al raggiungimento di una condizione che termina l'apprendimento. Quest'ultima è, in genere, rappresentata da un numero prefissato di epoche o da un valore dell'errore considerato come tollerabile e adatto ai nostri obiettivi. La qualità dei risultati di una ANN allenata è verificato calcolando l'errore quadratico medio fra output-rete ed output-reale desiderato.

" η " e " α " controllano la velocità e l'efficacia del processo di apprendimento. *"If the learning rate is set too high then large steps will be taken, and the system will oscillate. If it is set too low, small steps will be taken, resulting in longer training times and a greater likelihood of the search becoming trapped in a local minimum in the error surface. A careful selection of these two parameters is necessary for smooth convergence to a global minimum..."*¹⁰ [Mas et

⁹ Delta rule: regola di apprendimento basata sulla diminuzione del gradiente, per l'aggiornamento dei pesi in un MLP; il "gradient descent" è un algoritmo di ottimizzazione non lineare, per individuare il minimo locale di una funzione.

¹⁰ Se la learning rate è troppo grande, dei passi computazionali ampi saranno adottati ed il sistema oscillerà. Se il valore della learning rate è troppo basso,

al., 2007, Pag.33]. Numerosi studi sono stati realizzati per determinare i **parametri ottimali** di una ANN [Kavzoglu et al., 2003]; questa scelta dipende strettamente dal problema esaminato e dai dati a disposizione, anche se possiamo generalizzare dicendo che un " η "=[0.1, 0.2] con un " α "=[0.5, 0.6] rappresenta un ottimo compromesso per ottenere un'elevata precisione di classificazione. Alcune ANN (il modulo LCM in Idrisi Andes, ad es.) adottano delle strategie di apprendimento adattabili; questi metodi adattano la "*learning rate*" durante la fase di allenamento, in funzione delle caratteristiche della superficie e del gradiente d'errore.

Un'ultima considerazione deve essere fatta a riguardo dei **dati** scelti per l'allenamento della rete. La distribuzione dell'informazione e la dimensione del insieme di dati presentato alla ANN condiziona il risultato; per minimizzare l'errore totale, una rete deve elaborare dei dati che siano rappresentativi delle reali proporzioni delle categorie studiate. Alcune volte il volume di informazioni per ogni classe è scelto in base all'area occupata dalla stessa; in altre circostanze può essere modificato per aumentare il numero di pixels che rappresentano le classi più piccole, o, al contrario, per ridurre i casi più numerosi e ridondanti. L'ordine di introduzione dei dati può anche esso influenzare l'allenamento della rete, poiché le ultime coppie (input, output) presentate alla ANN hanno più importanza nell'aggiornamento dei pesi sinaptici rispetto alle prime. In generale questo problema è risolto adottando un ordine di introduzione random.

dei piccoli passi computazionali saranno adottati e ciò porterà a dei tempi di allenamento più lunghi e ad una più elevata possibilità che la ricerca rimanga vincolata ad un minimo locale, sulla superficie dell'errore. Una scelta accurata di questi due parametri è necessaria per garantire una convergenza della rete su un minimo globale.

3.1.5 Descrizione del modello PNNET

Per modellizzare i processi di deforestazione in Peten, consideriamo il seguente problema di regressione: per ogni pixel(i) di una mappa, la variabile target è considerata come binomiale ed il suo valore dipende dalla domanda "Rimarrà classificato come foresta alla prossima data analizzata ($t+1$)?". La variabile è codificata in forma disgiuntiva:

[1 0] in caso affermativo, [0 1] altrimenti.

Per rispondere a questa questione utilizzeremo differenti variabili predittive:

- ✓ Processi temporali: i valori dei pixels rappresentativi delle varie categorie di copertura in una certa data (t), codificati in forma disgiuntiva. Se per esempio analizziamo 5 differenti categorie di occupazione del suolo (foresta primaria, foresta secondaria, zone umide, aree coltivate e pasto), la prima sarà classificata sottoforma di un vettore a 5 coordinate: (1 0 0 0 0), la seconda: (0 1 0 0 0), etc.
- ✓ Processi spaziali: la frequenza di ogni tipo di uso del suolo, alla data (t), in una zona di influenza. Quest'ultima non è altro che una finestra quadrata di vicinaggio, centrata sul pixel analizzato e la cui dimensione può essere scelta in funzione delle caratteristiche del problema (fino a quale distanza un pixel può influenzare il valore del pixel centrale). Per rispettare le caratteristiche spaziali della mappa, pondereremo l'influenza di una cella attraverso una funzione decrescente della distanza (Fig.3-7) [Follador et al., 2008; Villa, 2005].

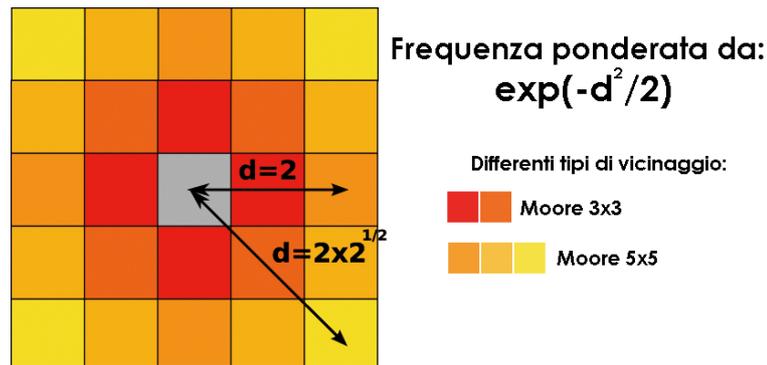


Fig 3-7: Esempi di vicinaggio e calcolo della frequenza delle categorie di uso del suolo, all'interno della zona di influenza per ogni pixel della mappa. La frequenza di una classe è calcolata mediante una funzione decrescente della distanza dal pixel centrale in esame. La finestra mobile analizza tutti i pixels dell'immagine.

- ✓ Criteri ambientali o variabili esplicative alla data (t): queste grandezze influenzano l'evoluzione dei patterns spaziali dei processi studiati. Possono essere ulteriormente suddivise in due categorie, dinamiche e statiche. Le variabili dinamiche cambiano da una data all'altra; la distanza da aree deforestate o dai vettori di penetrazione, ad esempio, sono grandezze che evolvono nel tempo, poiché le zone adibite a pascolo ed agricoltura si espandono progressivamente e nuove infrastrutture sono costruite per integrare la precedente rete viaria. I criteri ambientali statici mantengono costante la loro informazione con il passare del tempo (DEM, distanza da un corso d'acqua, ect.). Il modello PNNET può ricevere sia variabili numeriche che categoriche (sempre codificate in forma disgiuntiva).

Per modellizzare le dinamiche ambientali in Peten abbiamo utilizzato un perceptrone con un solo strato nascosto (Fig.3-8), data la sua riconosciuta capacità di risolvere la maggior parte dei problemi non lineari e di approssimare ogni funzione continua. Lo strato di input è formato da 14 neuroni: due neuroni rappresentativi delle classi tematiche in esame (foresta e area deforestata o disturbata), due neuroni per la

frequenze di tali classi in una finestra d'influenza, per ogni pixel, ed infine 10 variabili esplicative alla data (t) . Lo strato nascosto è formato da "k" nodi computazionali, scelti durante la fase di ottimizzazione della rete. Il valore di "k" è stabilito dal modellizzatore in base alla necessità di flessibilità (un numero maggiore di neuroni garantisce una maggiore flessibilità della ANN) ed all'errore reale-simulato ottenuto nella fase test di cross-validazione. Infine il livello di output è composto da due neuroni che rappresentano le probabilità di avere, per ogni pixel, foresta e non foresta alla data successiva $(t+1)$ ¹¹. Ad ogni pixel della mappa finale sarà dunque attribuita la classe che ha la probabilità di realizzazione maggiore.

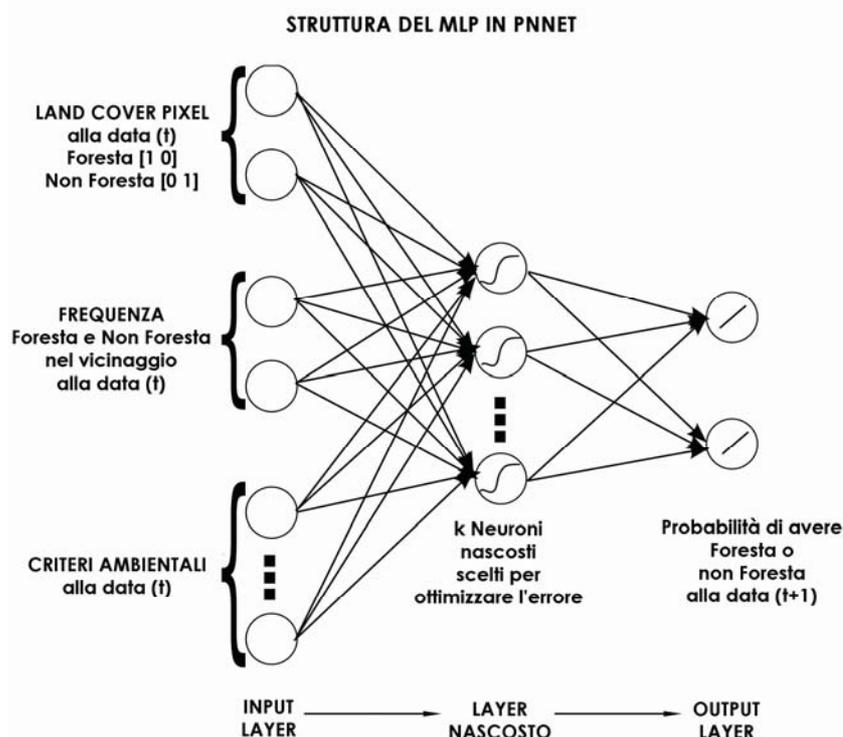


Fig 3-8: Struttura del MLP con un solo strato nascosto, utilizzato per la costruzione di PNNET. In Input vengono presentate tutte le variabili predittive raccolte alla data (t) ; in Output otteniamo le probabilità di avere Foresta o non Foresta nell'area studio del Peten, alla data successiva $(t+1)$. [Immagine adattata da Follador et al., 2008].

¹¹ $(t+1)$ rappresenta la data successiva simulata. Il valore "+1" indica un avanzamento di un passo temporale predefinito e non un aumento annuale. Ad esempio per lo studio in Peten noi abbiamo utilizzato un passo temporale di 3 anni; dunque con le informazioni della data $t=2000$ simuliamo l'evoluzione della deforestazione nella data $t+1=2003$.

Ricordiamo che il legame fra input ed output è espresso da:

$$p_w^j(x) = \sum_{i=1}^k w_{ij}^2 g(x^T w_i^1 + w_i^0) \quad (7)$$

dove "x" è il vettore contenente le variabili di ingresso, "p_w^j(x)" è il j-esimo risultato dipendente dai pesi "w", "w_i¹" è il vettore dei pesi fra lo strato di input ed i-esimo neurone nascosto, "w_i⁰" è il bias dell'i-esimo neurone nascosto e "w_{ij}²" è il peso della connessione fra i-esimo neurone nascosto ed il j-esimo output. I pesi sono scelti durante la fase di allenamento della rete su un data set rappresentativo, per ridurre l'errore fra mappa reale e simulata. Come funzione di attivazione abbiamo scelto la seguente funzione sigmoidea:

$$g(z) = \frac{1}{(1+e^{-z})} \quad (8)$$

Per lo studio delle dinamiche ambientali in Peten abbiamo utilizzato tre carte tematiche ricavate dalla classificazione di immagini satellitari acquisite nel 1998 (SPOT2, 606-315, 1998_02_24), 2000 (ETM, 20-48, 2000_03_27) e 2003 (ETM, 20-48, 2003_05_07). Ogni matrice ha (979r X 601c) pixels con una risoluzione di 20x20 m²; il volume dei dati categorici e delle variabili esplicative è troppo grande per essere caricato direttamente nei programmi che abbiamo sviluppato in R[®] e che compongono il modello PNNET. Siamo stati dunque obbligati a ridurre la dimensione del data set attraverso una semplificazione¹²: durante la fase di apprendimento e di simulazione, considereremo solamente le informazioni dei pixels frontiera. Chiamiamo di **pixel frontiera** una cella che ha, almeno , una classe di uso del suolo differente all'interno della sua finestra d'influenza. "*Après avoir repéré des larges*

12 Per una descrizione dettagliata dei metodi di pre-processamento dei dati di una ANN, si rimanda a [Villa, 2002, Capitolo 1.3.4].

*zones dans lesquelles l'occupation du sol était stable, nous avons considéré uniquement les pixels dont un voisin au moins avait une valeur d'occupation du sol différente...les autres pixels ont été considéré comme stables..."*¹³ [Villa, 2005, Pag.62]. Facciamo notare che questa campionatura, pur risultando sufficientemente adatta alle caratteristiche del nostro problema, può non rappresentare completamente la reale dispersione e variabilità dell'informazione esaminata. *"The training data must be representative of the entire distribution of values likely to be associated with a particular class. If the extent of the distribution of the data in feature space is not covered adequately the network may fail to classify new data accurately"*¹⁴ [Mas et al., 2007, Pag.26].

La metodologia utilizzata per analizzare il comportamento della rete neurale e la qualità dei suoi risultati, è composta da 3 fasi (Fig.3-9):

- ✓ Fase di allenamento: ottimizza i pesi della rete, dato un "k" numero di neuroni nascosti e un "V" dimensione dell'area di influenza;
- ✓ Fase di validazione: seleziona i valori di "k" e "V" ottimali;
- ✓ Fase di test: confronta la mappa predetta con la mappa reale nel 2003 e dona le prime statistiche.

Più precisamente, la fase di allenamento utilizza circa il 10% dei pixels frontiera delle mappe 1998/2000 per costruire

13 Dopo aver evidenziato ampie zone che permangono stabili, abbiamo considerato solamente i pixels che hanno almeno un vicino caratterizzato da un uso del suolo differente...gli altri pixels sono stati considerati come stabili (nel periodo simulato).

14 I dati scelti per allenamento della rete devono essere rappresentativi dell'intera distribuzione spazio-temporale delle caratteristiche associabili ad ogni classe. Se tale distribuzione non è adeguatamente rappresentata, la ANN può non lavorare correttamente introducendo nuovi dati (differenti da quelli usati nell'allenamento).

il training set (input, output) e permettere l'ottimizzazione dei pesi "w" descritti in (7). Per ogni coppia di pixel input/output che indichiamo con $(x_i, y_i)_{i=1..n}$, "w" è scelto per minimizzare l'errore quadratico medio fra i valori predetti dalla rete " $p_w^j(x_i)$ " in (1) a partire da un input " x_i ", ed i valori reali di output " y_i^j ":

$$E = \sum_i^n \sum_{j=1,2} (p_w^j(x_i) - y_i^j)^2 \quad (9)$$

Questa operazione di ottimizzazione è stata realizzata utilizzando un algoritmo di decremento del gradiente (gradient descent); per ovviare ai problemi legati ai minimi locali, questa fase è ripetuta 10 volte con differenti training set, scelti in maniera random, sempre rispettando le reali proporzioni foresta/non foresta nella mappa 1998. Infine selezioniamo il perceptrone con "E" minimo.

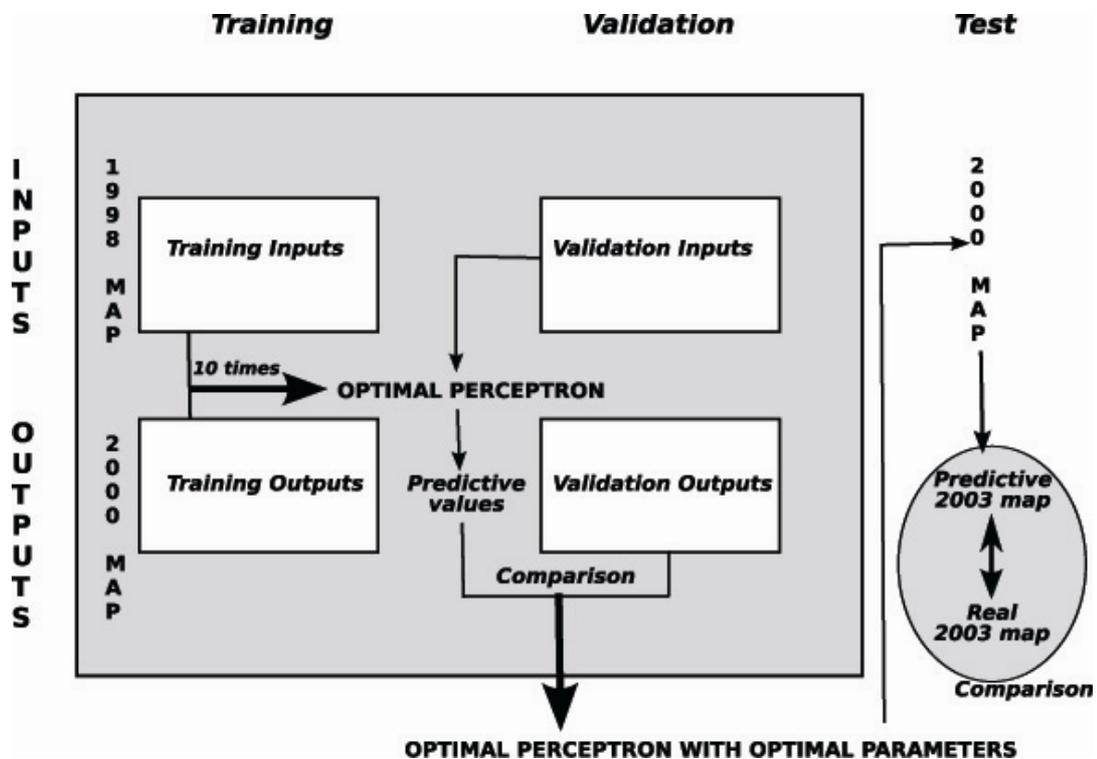


Fig 3-9: Approccio metodologico utilizzato per l'ottimizzazione di PNNET. Le mappe di uso del suolo in Peten nel 1998 e 2000 sono utilizzate nella fase di apprendimento e ottimizzazione del perceptrone. Successivamente 2000 sarà presentato a MLP ottimizzato per simulare l'evoluzione delle dinamiche di deforestazione in 2003, scelto come riferimento per la validazione finale. [Immagine da Follador et al., 2008]

Una volta ottimizzata la rete per ogni valore di "k" e "V", cerchiamo di stabilire quale sia il miglior valore per questi due parametri, attraverso la fase di validazione. Circa il 30% di tutti i pixels frontiera nelle mappe 1998 e 2000 sono selezionati, sempre rispettando le reali proporzioni fra le classi nel 1998; questo gruppo di pixels è differente dal training set precedentemente utilizzato. I pixels di input (1998) sono utilizzati in ogni perceptrone ottimizzato per produrre una mappa predittiva, che sarà poi comparata con l'output reale desiderato (2000). Ancora una volta il MLP scelto è quello che minimizza "E". Nel nostro caso la fase di validazione ha portato alla scelta di $k=5$ (Fig.3-10) e $V=3$ (vicinaggio di Moore 3x3).

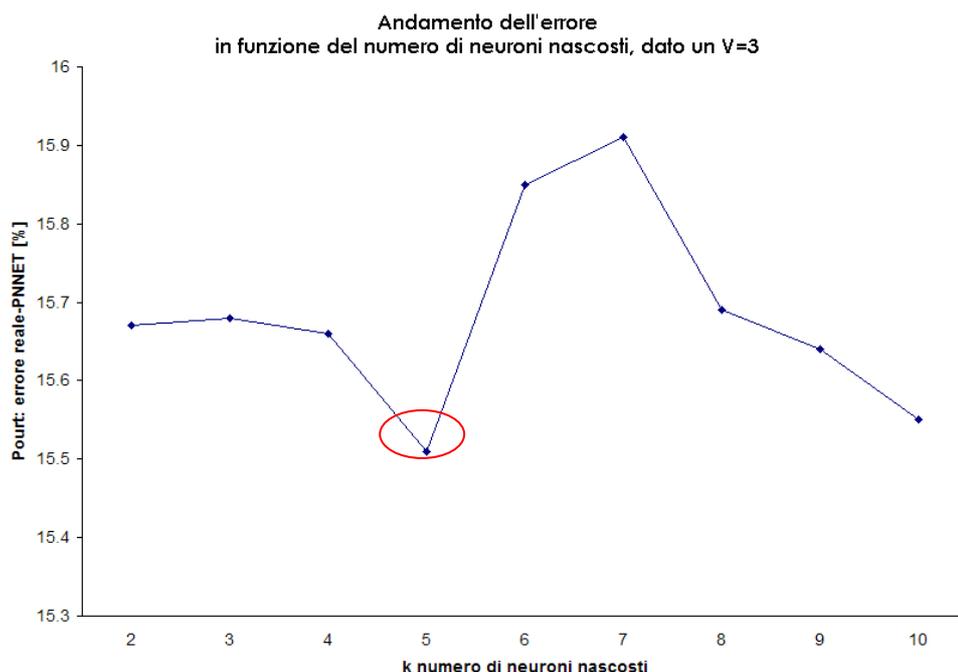


Fig 3-10: Andamento dell'errore reale-simulato "Pour" in funzione del numero di neuroni nello strato nascosto, dato un V=3. Dopo l'analisi di differenti alternative di V e k, nello studio delle dinamiche del Peten il perceptrone ottimale scelto ha $k=5$ e $V=3$.

Infine, utilizziamo tutti i pixels frontiera della mappa 2000 come input per il MLP per la costruzione della mappa predittiva 2003. La fase successiva di test offre le prime statistiche di divergenza (pixel per pixel) fra reale-simulato

in 2003, scelta come riferimento per la validazione finale. Un filtro di contiguità (automa cellulare) può essere aggiunto per aumentare l'isometria delle aree deforestate e limitare il numero dei pixel isolati, probabilmente dovuti all'approssimazione legata ai pixels frontiera.

3.1.6 PNNET: descrizione dei programmi in R[®]

*"We start with the simplest but most common form with one hidden layer. The input units just provide a "fan-out" and distribute the inputs to hidden units in the second layer. These units sum their inputs, add a constant (bias) and take a fixed function Φ_{hidden} of the result. The output units are of the same form but with output function Φ_{output} ...Software to fit feed-forward neural networks with a single hidden layer is provided in our library nnet"*¹⁵ [Venables et al., 1999, Pag.337 e 340].

Per eseguire la metodologia precedentemente descritta, abbiamo sviluppato 3 programmi in R[®], caricando in questo ambiente di programmazione la libreria "nnet"¹⁶. I nostri programmi sono freeware e liberamente scaricabili all'indirizzo internet:

http://nathalie.vialaneix.free.fr/maths/article.php3?id_article=49.

Alla stessa pagina è possibile trovare un piccolo manuale che rapidamente illustra le funzioni base di PNNET attraverso alcuni semplici esercizi. Cercheremo, in questo capitolo, di offrire una descrizione dettagliata dei programmi sviluppati e della loro applicazione nello studio dei LUCC in Peten. Gli autori (Villa e Follador) sperano così di poter integrare le nozioni base introdotte nel manuale in rete.

15 Cominciamo con la ANN più semplice e comune, formata da un solo strato nascosto. Le unità di input si limitano a distribuire le informazioni alle unità dello strato nascosto. Queste unità sommano i loro inputs, aggiungono una costante (bias) ed applicano una funzione Φ_{hidden} al risultato. Le unità di output presentano la stessa forma, ma con una funzione di uscita Φ_{output} ... gli strumenti per strutturare una rete neurale con un solo strato nascosto sono descritti nella nostra libreria nnet.

16 Per maggiori dettagli vedere <http://www.r-project.org>.

L'ossatura di PNNET è formata da 3 programmi in R[®] (in allegato nel CD):

1. *forme.R*: prepara i dati originali in una forma adatta ad essere utilizzata nel MLP descritto in (Fig.8);
2. *perceptron.R*: esegue la fase di allenamento della rete e restituisce l'errore quadratico medio del migliore perceptrone su un insieme di 10 campioni (scelti fra i dati d'allenamento);
3. *statistique.R*: restituisce l'errore reale-simulato durante la fase di apprendimento, validazione e test del MLP (Fig.9).

3.1.6.1 Il programma *forme.R*

Questa funzione ha la seguente sintassi:

```
forme(data,      taille,      variables.quant,      variables.qual,
variables.quant.temps, variables.qual.temps, V, N, iter)
```

dove:

- ✓ *data* è la matrice che contiene le carte tematiche di copertura nelle differenti date. Ogni riga rappresenta un pixel ed ogni colonna una data (quindi muovendomi lungo una riga ho l'evoluzione temporale del pixel). Le *N* categorie di copertura sono classificate da [1...*N*], mentre lo zero rappresenta eventuali pixels mascherati o non considerati nello studio;
- ✓ *taille* è semplicemente il vettore contenente il numero di righe e colonne delle mappe originali utilizzate (ovviamente tutte le mappe devono avere la stessa dimensione);
- ✓ *variables.quant* è una matrice contenente le variabili quantitative (numeriche) che rimangono costanti durante il periodo esaminato; ogni riga (stesso ordine di *data*) è un

pixel ed ogni colonna rappresenta una variabile (quindi muovendomi lungo una riga ho i valori delle differenti variabili nello stesso pixel);

- ✓ *variables.qual* è una matrice che contiene le variabili qualitative (categoriche) che rimangono stabili nel periodo esaminato; ogni riga (stesso ordine di *data*) è un pixel ed ogni colonna rappresenta una variabile. Le N variabili, rappresentate da un valore [1..N], sono codificate in forma disgiuntiva.
- ✓ *variables.quant.temps* è una matrice che contiene tutte le variabili numeriche dinamiche, che cambiano durante il periodo esaminato. Ogni riga rappresenta un pixel (stesso ordine di *data*) e le colonne rappresentano il valore delle variabili alla data *t1*, seguito dal valore in *t2* e così via (Fig.3-11);

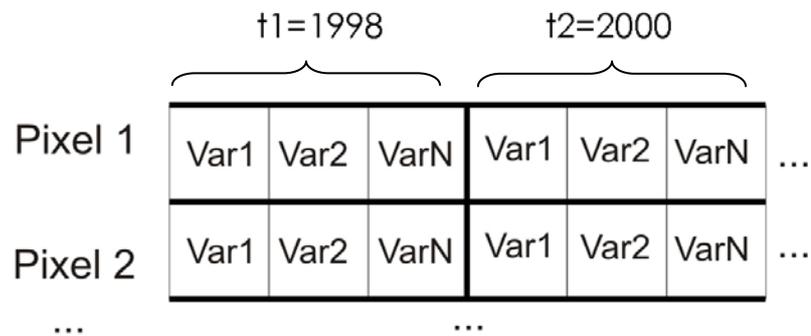


Fig 3-11: Rappresentazione grafica della disposizione dei dati nella matrice *variables.quant.temps*.

- ✓ *variables.qual.temps* è una matrice che contiene le variabili categoriche dinamiche, che cambiano durante il periodo esaminato. Ogni riga rappresenta un pixel (stesso ordine di *data*) e le colonne rappresentano il valore delle variabili alla data *t1*, seguito dal valore in *t2* e così via.
- ✓ *V* è la dimensione che definisce la zona d'influenza, descritta nel precedente capitolo. Un $V > 7$ è generalmente

sconsigliato, poiché comporta tempi di calcolo ed un'allocazione di memoria a volte eccessivi.

- ✓ N è il vettore che contiene il numero di pixels che formerà il "training data set" (dati per l'allenamento della rete) ed il "validation data set" (dati per la fase di cross-validazione). Normalmente si consiglia di non superare il valore di $N_{\text{training}} < 100.000$ e $N_{\text{test}} < 500.000$.
- ✓ $Iter$ è il numero che indica quante volte la rete ripete la fase di allenamento. Il valore scelto per default è $iter=10$.

La funzione *forme.R* crea differenti files; i più importanti sono:

- ✓ *IO.Rdata* include gli array *entrees* e *sorties* che contengono tutti i pixels scelti per la fase di allenamento e validazione della rete, e che rappresentano rispettivamente l'input e l'output (reale) del MLP. Le variabili qualitative e le mappe classificate di uso del suolo sono codificate in forma disgiuntiva; le variabili numeriche sono centrate e ridotte¹⁷;
- ✓ *selections.Rdata* contiene l'array *selections* in cui sono registrate tutte informazioni (estensione, numero) a riguardo dei pixel frontiera selezionati;
- ✓ *selections_partage.Rdata* include l'array *sapp* e la matrice *stest* che contengono le informazioni sui pixels frontiera selezionati, rispettivamente per l'allenamento ($iter=10$ training sets) e la validazione. La proporzione delle differenti categorie di uso del suolo è rispettata ed uguale a quella delle mappe originali;

¹⁷ Per un approfondimento delle tecniche utilizzate nel preprocessing dei dati, si rimanda a [Villa, 2002].

- ✓ Altri file sono creati dalla funzione *forme.R*. Essi non sono utili per comprendere il funzionamento PNNET, e dunque non verranno qui descritti; per un eventuale approfondimento si rimanda al sito web.

3.1.6.2 Il programma Perceptron.R

La sintassi di questa funzione è la seguente:

```
perceptron (entree, sorties, sapp, stest, k)
```

dove:

- ✓ *entrees*, *sorties*, *sapp* e *stest* sono matrici prodotte dalla precedente funzione *forme.R*.
- ✓ *k* è il numero di neuroni dello strato nascosto, scelto dal modellizzatore.

La funzione *perceptron.R* genera il file *perceptron.Rdata* che contiene:

- ✓ *percept_opt* che è il perceptrone ottimizzato per un certo *V* e un certo *k*, su *iter* data sets d'allenamento (training samples).
- ✓ *perf* è il vettore che contiene *iter* errori quadratici medi per ogni training sample, scelti *V* e *k*.

Numerosi perceptron, con differenti valori di *k* ($k=2\dots 10$) sono stati costruiti ed i risultati comparati per scegliere il valore ottimale di *k*; ogni risultato prodotto da *perceptron.R* deve essere rinominato e salvato a parte per evitare che il nuovo output sia sovrascritto al *percept_opt* precedente (per $k=2$ rinominiamo il file di output come *percept_opt_2*, per $k=3$ come *percept_opt_3*, etc.). Ovviamente aumentando il numero di neuroni nascosti, si aumenta parallelamente il numero delle connessioni sinaptiche e dunque dei pesi (nel nostro caso $k=2 \rightarrow 24$ pesi; $k=3 \rightarrow 35$ pesi; $k=4 \rightarrow 46$ pesi, etc.), incrementando flessibilità e tempi di calcolo della rete.

3.1.6.3. Il programma *Statistique.R*

La sintassi di questa funzione è la seguente:

```
statistique(percept_opt, entrees, sorties, selections, data)
```

Questa funzione genera due files:

- ✓ *statistiques_app.Rdata* che contiene le statistiche relative alla fase di allenamento e validazione. Dopo aver costruito le mappe predittive (registrate in *carte.app*), calcolo l'errore sui pixels frontiera selezionati (*pour*), e sull'intera carta (*pourt*), le matrici di transizione fra tutte le categorie di uso del suolo nell'ultima data analizzata (*transf.test*) ed infine le proporzioni relative di ogni categoria di uso del suolo sempre con riferimento all'ultima data (*nveg*).
- ✓ *statistiques_test.Rdata* contiene le statistiche relative alla fase di test. La mappa predittiva finale è costruita a partire dalla mappa registrata al passo temporale precedente (in *cartes.test*). Calcolo dunque l'errore sui pixels frontiera selezionati (*pour.test*) e sull'intera carta (*pourt.test*), le matrici di transizione fra tutte le categorie di uso del suolo nell'ultima data analizzata (*transf.test*) e le proporzioni relative di ogni categoria di uso del suolo sempre con riferimento all'ultima data (*nveg.test*).

La funzione *statistique.R* permette di quantificare l'errore commesso durante la simulazione dai differenti perceptroni creati (*percept_opt_2, percept_opt_3, etc.*) e di scegliere conseguentemente il valore ottimale di *k*. Come già ricordato in (Fig.3-10), nello studio dei LUCC in Peten abbiamo adottato un MLP formato da 5 neuroni nascosti (per un *V=3* predefinito).

3.2 Dinamica EGO

3.2.1 Introduzione

Dinamica EGO è la versione aggiornata del precedente modello (Dinamica) sviluppato dal CSR¹⁸ dell'Università Federale di Minas Gerais (UFMG), Brasile. Questo freeware è liberamente scaricabile al sito: <http://www.csr.ufmg.br/dinamica/EGO>.

Il modello aggrega le tradizionali funzioni dei SIG, con una serie di operatori che permettono di simulare l'evoluzione spaziale di fenomeni naturali-artificiali come la deforestazione [Follador et al., 2008; Cuevas et al., 2007] o lo sviluppo urbano [Almeida, 2003]. Le fasi della modellizzazione, dalla calibrazione alla validazione, hanno la forma di un diagramma di flusso di dati (Fig.3-12); un'interfaccia grafica di facile comprensione, permette di creare modelli semplicemente legando degli algoritmi (detti *functors*) attraverso le loro porte di input ed output. È possibile dividere l'area di studio in sub-regioni, caratterizzate da differenti contesti sociali, economici e biofisici, permettendo un'analisi specifica e "su misura" delle dinamiche locali di cambio di uso del suolo [Rodrigues et al., 2007]. Durante la calibrazione calcoliamo la matrice di transizione per ogni classe, nel periodo che intercorre fra mappa iniziale e mappa finale; la matrice di transizione può essere calcolata anche esternamente a Dinamica EGO, attraverso altri modelli (econometrici, ad es.), ed introdotta successivamente. Una mappa di probabilità d'occorrenza è prodotta per ogni transizione, attraverso il "*weights of evidence method*"¹⁹. Il numero totale di pixels che abbiamo previsto cambieranno, avanzando di un passo temporale, sono

18 CSR: Centro de Sensoriamento Remoto (centro di telerilevamento), IGC Istituto di Geoscienze. Responsabile del progetto: Prof. Britaldo Silveira Soares Filho.

19 Weights of Evidence : pesi dell'evidenza. Per maggiori dettagli sulla teoria Bayesiana vedere [Bonham-Carter, 1994].

divisi fra due differenti funzioni di transizione. La prima, *Expander*, analizza l'espansione o la contrazione di parcelle già esistenti adibite ad un certo uso del suolo (l'espansione di un'area di pascolo, ad es.). La seconda funzione, *Patcher*, genera nuove aree per una determinata classe di copertura, isolate o sparpagliate (nuovi poligoni deforestati, ad es.). La validazione è basata su una comparazione "fuzzy" fra mappa reale e simulata, all'interno di una zona di influenza per ogni pixel; questo metodo non considera solamente la congruenza pixel a pixel, ma anche la probabilità di trovare il valore corretto predetto in un intorno del pixel reale (cioè non è importante che l'area predetta come deforestata si trovi nel punto esatto o dislocata di alcuni metri).

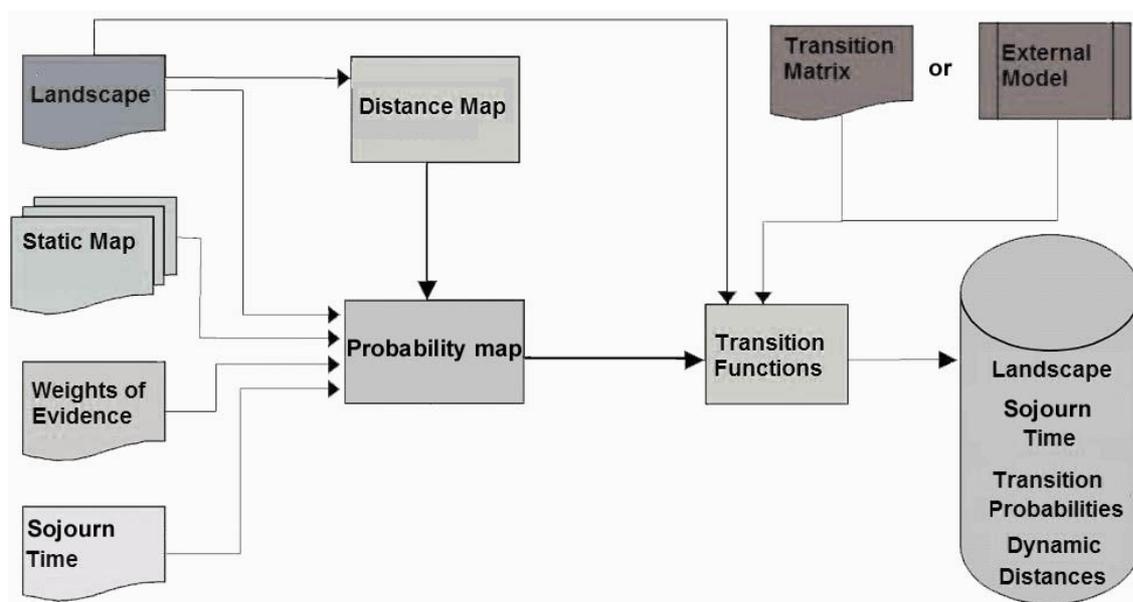


Fig 3-12:Diagramma di flusso di Dinamica. [Immagine da Godoy et al., 2008]

3.2.2 L'approccio modellizzativo

Dinamica EGO utilizza un automa cellulare per riprodurre le dinamiche del paesaggio e la maniera in cui i patterns spaziali evolvono. Il paesaggio è dunque rappresentato come un array n-dimensionale di cellule che interagiscono all'interno di un certo vicinaggio. Lo stato di ogni cellula è funzione dello stato precedente dei pixels all'interno del vicinaggio, e di

una serie di regole di transizione. Tutte le cellule sono aggiornate simultaneamente ad ogni passo temporale. Il cambiamento è ristretto ad un intorno del pixel e l'interazione fra le cellule del vicinaggio è calcolata attraverso regole locali dell'automa cellulare. Quest'ultime in Dinamica sono composte da due funzioni di transizione complementari: *Expander* e *Patcher*.

L'ambiente di simulazione integra differenti fasi che sono facilmente implementate attraverso un'interfaccia grafica o direttamente attraverso il linguaggio XLM. Riassumiamo qui i passaggi principali che portano alla produzione di una mappa predittiva ed i parametri adottati per lo studio dei LUCC in Peten.

3.2.2.1 Calcolo quantitativo dei cambiamenti futuri

Il modello *01-determine_transition_matrix.xml* (Fig.3-13 e nel CD allegato) calcola il numero di pixels per ogni transizione considerata (nel caso del Peten, foresta→deforestazione e viceversa). Il calcolo si basa su un'analisi di Markov di secondo ordine. Dato un processo stocastico $X(t)$, caratterizzato da un numero discreto di stati $S=\{1,2,...K\}$, ed una sequenza di istanti $t_1 < t_2 < \dots < t_{n-1} < t_n$, la probabilità condizionata avrà la forma [Shamshad et al., 2005]:

$$\Pr\{X(t_n)=i_n | X(t_1)=i_1, \dots, X(t_{n-1})=i_{n-1}\} = \Pr\{X(t_{n-1})=i_{n-1}\}, \in [0,1] \quad (10)$$

La probabilità condizionata:

$$\Pr\{X(t)=j | X(s)=i\} = P_{ij}(s,t) \quad (11)$$

dallo stato "i" allo stato "j", è detta di ordine $r=t-s$, $\forall 0 \leq s \leq t$ e con $1 \leq i \text{ e } j \leq k$. Una catena di Markov di ordine 2 considera dunque l'informazione dei due istanti precedenti

(t_{-1}) e (t_0) per stabilire la probabilità all'istante (t_1) studiato²⁰.

Abbiamo definito un passo temporale di un anno, per poter così prevedere i tassi futuri di cambio su base annuale. La funzione *determine transition matrix*, nel caso del Peten, è stata parametrizzata con le mappe categoriche di uso del suolo a $(t_{-1})=1998$ e $(t_0)=2000$, ed un numero di passi temporali=2.

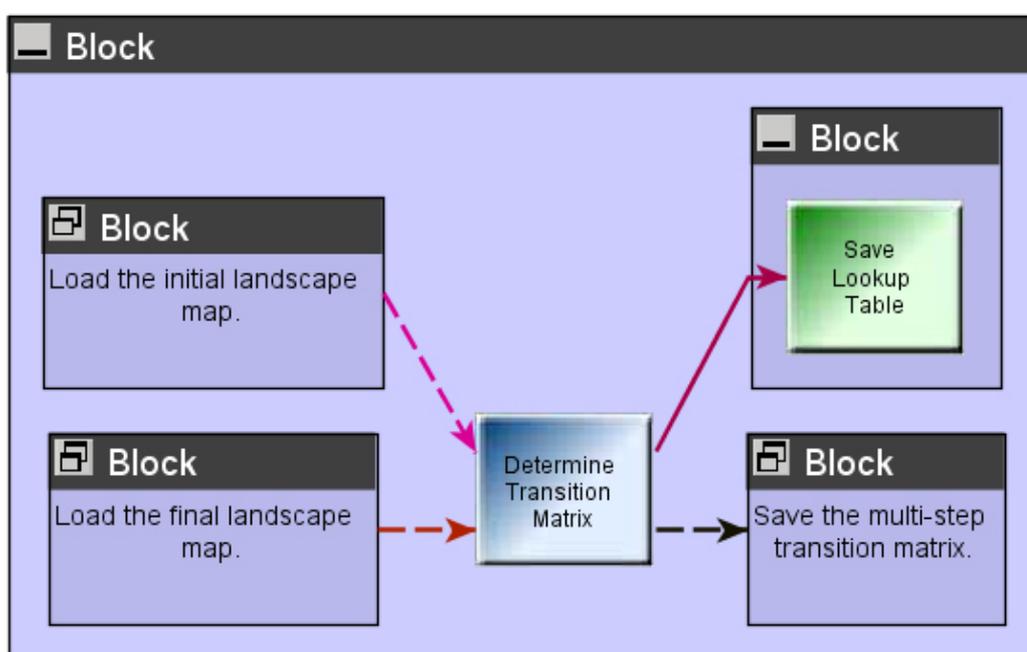


Fig 3-13: Rappresentazione grafica del modello *determine transition matrix.xml* costruito in Dinamica EGO. Questo primo modello permette il calcolo delle matrici di transizione basandosi su Markov Chains di 2° ordine (data iniziale t_{-1} - data finale $t_0 \rightarrow$ data futura t_1).

La matrice ottenuta descrive un sistema che cambia attraverso incrementi discreti di tempo²¹, in cui il valore di ogni variabile, in un dato momento, è la somma di percentuali dei valori delle variabili nell'istante precedente. La somma delle frazioni lungo una riga della matrice è uguale ad uno; la diagonale raccoglie invece le percentuali di pixels che non

20 Ordine $r=t-s \rightarrow r=1-(-1)=2$.

21 *Du point de vue du temps, les réalisations successives de la variable constituent un processus discret* [Coquillard et al., 1997, Pag.67].

subiscono cambiamenti fra la data iniziale e finale.
 “ L'ensemble du système peut se résumer à une seule matrice rassemblant les probabilités de passage de tous les états vers tous les autres”²² [Coquillard et al., 1997, Pag.68].

Una matrice di transizione di secondo ordine può essere simbolicamente rappresentata da:

$$P = \begin{bmatrix} p_{1,1,1} & p_{1,1,2} & \dots & p_{1,1,k} \\ p_{1,2,1} & p_{1,2,2} & \dots & p_{1,2,k} \\ \vdots & \vdots & \vdots & \vdots \\ p_{1,k,1} & p_{1,k,2} & \dots & p_{1,k,k} \\ p_{2,1,1} & p_{2,1,2} & \dots & p_{2,1,k} \\ p_{2,2,1} & p_{2,2,2} & \dots & p_{2,2,k} \\ \vdots & \vdots & \vdots & \vdots \\ p_{k,k,1} & p_{k,k,2} & \dots & p_{k,k,k} \end{bmatrix}$$

dove “ $p_{i,j,k}$ ” è la probabilità del prossimo stato “ k ” in base allo stato attuale “ j ” e allo stato precedente “ i ”.

Nel caso del Peten, le dinamiche studiate si riassumono attraverso lo studio binario delle transizioni fra le classi foresta e non foresta (Tab.3-1).

Matrici di Markov per t(1)=2003; dati: t(-1)=1998 & t(0)=2000

#Pixels per transizione			Prob. transizione		
	Foresta	Milpa		Foresta	Milpa
Foresta	278488	77254	Foresta	0.7828	0.2172
Milpa	22278	70552	Milpa	0.24	0.76

Tab. 3-1: La matrice di Markov adottata per lo studio dei LUCC in Peten riassume le probabilità di realizzazione delle 4 transizioni considerate ed il numero di pixels atteso. Le caselle evidenziate sulla diagonale sono indicative della persistenza delle classi.

22 L'intero sistema si può riassumere attraverso una matrice che descriva le probabilità di passaggio di ogni classe verso tutte le altre classi.

La matrice di transizione ottenuta dall'analisi delle due date precedenti ($t_{-1}=1998$ e $t_0=2000$) è successivamente utilizzata nel calcolo delle percentuali di transizione annuali, attraverso la (12):

$$P^t = HV^tH^{-1} \quad (12)$$

dove "P" rappresenta la matrice di probabilità, "H" e "V" le matrici degli autovettori e autovalori, e "t" un multiplo o una frazione del passo temporale [Soares-Filho et al., 2002].

3.2.2.2 Allocazione spaziale dei cambiamenti.

Per ogni transizione considerata (Tab.1) costruiamo una mappa di probabilità di realizzazione attraverso il metodo dei pesi dell'evidenza (WoE). Quest'ultimo è un approccio Bayesiano che permette di evidenziare le relazioni fra variabili esplicative e i Lucc osservati. In generale se una mappa, quantitativa o qualitativa, offre un'informazione utile per identificare le aree deforestate (ad es.), allora i pixels che rappresentano queste zone hanno caratteristiche proprie ben differenti dalle restanti regioni coperte da vegetazione. Per evidenziare queste diversità utilizziamo la "*likelihood ratio function*", cioè il rapporto fra due distribuzioni di frequenza [Chung et al., 2002].

Una mappa di probabilità per le due classi considerate, foresta e milpa, è calcolata considerando tutti i layers (variabili esplicative) attraverso la "*likelihood ratio function*". Dinamica EGO utilizza un metodo Bayesiano per stimare quest'ultima.

La possibilità di realizzazione di una certa transizione (la deforestazione "D", ad es.), data una mappa binaria che descrive la presenza o assenza di un pattern spaziale "C", può essere espressa dalla probabilità condizionale:

$$P\{D|C\} \equiv \frac{P\{D \cap C\}}{P\{C\}} \equiv P\{D\} \frac{P\{C|D\}}{P\{C\}} \quad (13)$$

Attraverso una manipolazione algebrica possiamo rappresentare la (13) in termini di *odds*, che sono definiti come il rapporto fra la probabilità di occorrenza di un evento e la probabilità che lo stesso non avvenga [Bonham-Carter, 1994]:

$$O\{D|C\} \equiv \frac{P\{D|C\}}{P\{\bar{D}|C\}} \equiv \frac{P\{D\}P\{C|D\}}{P\{\bar{D}|C\}P\{C\}} \equiv \frac{P\{D\}P\{C|D\}}{P\{\bar{D}\}P\{C|\bar{D}\}} \equiv O\{D\}LS \quad (14)$$

dove " \bar{D} " rappresenta l'assenza di " D " e " LS " è detto *likelihood ratio* oppure *sufficiency ratio*²³.

Esprimiamo la (14) come logaritmo naturale per definire il peso dell'evidenza positivo (positive Weights of Evidence) " W^+ " di realizzazione di " D " dato " C ":

$$Logit\{D|C\} \equiv Logit\{D\} + W^+ \quad (15)$$

Questo approccio può essere utilizzato per manipolare diverse mappe contemporaneamente e poter così studiare le relazioni fra i LUCC ed i molteplici patterns geografici considerati. Alcune ipotesi restrittive sulla dimensione dell'area studiata e sui *prior odds ratio* (14) per ogni transizione, devono essere fatte. Alla fine la probabilità condizionale per la deforestazione, dato un gruppo di " n " variabili esplicative espresse attraverso delle carte quantitative o qualitative, è data per ogni pixel da:

$$P\{D|C_1 \cap C_2 \cap \dots \cap C_n\} \equiv \frac{e^{\sum W^+}}{1 + e^{\sum W^+}} \quad (16)$$

²³ Per una descrizione più dettagliata del metodo WoE si rimanda a [Bonham-Carter, 1994].

Per realizzare questo calcolo abbiamo costruito in Dinamica EGO due modelli da caricare in sequenza:

02-WoE_calcolo_ranges_coeff.xlm e *03-WoE_correlazione.xlm*.

Nel primo modello (Fig.3-14 e CD allegato) la funzione *Determine Weights of Evidence Ranges* permette di adattare il metodo WoE, che impiega patterns binari, a mappe categoriche multiple, trattando ogni categoria "A" separatamente dalle restanti " \bar{A} ". Il metodo di categorizzazione [Agterberg et al., 1990] calcola gli intervalli in cui dividere una mappa continua (mappa di distanza, ad es.) rispettando la struttura dei dati; si stabilisce un (Dx) per una variabile continua su una scala byte (toni di grigio 0-255), utilizzato per dividere la mappa in una sequenza incrementale di buffer(Nx), compresi fra [x_{min} , $x_{min}+nDx$]. Ogni valore "n" determina un limite che separa la mappa in due parti: (Nx)=1 e ($\bar{N}x$)=0. Così, ad esempio, in una carta di distanza da una strada, tutto lo spazio compreso nel primo buffer (0-20m, ad es.) sarà considerato come uguale ad 1 mentre tutta l'area restante sarà uguale a 0 e si procederà al calcolo di " W^+ " con questa informazione. Successivamente si ripeterà il calcolo per il secondo buffer (20-40m, ad es.) e così via. La quantità " An " definisce il numero di pixel all'interno di ogni buffer e " dn " il numero di occorrenze di un certo evento "D" (deforestazione, ad es.) all'interno del buffer. Utilizzando queste informazioni si arriva alla definizione degli intervalli di categorizzazione della variabile continua²⁴.

24 Gli intervalli sono calcolati applicando l'algoritmo line-generalizing (Intergraph, 1991) al grafico $An, An \cdot \exp(W^+)$. Per una descrizione più dettagliata del metodo utilizzato si rimanda a [Intergraph, 1991] ed alle note di aiuto in Dinamica EGO.

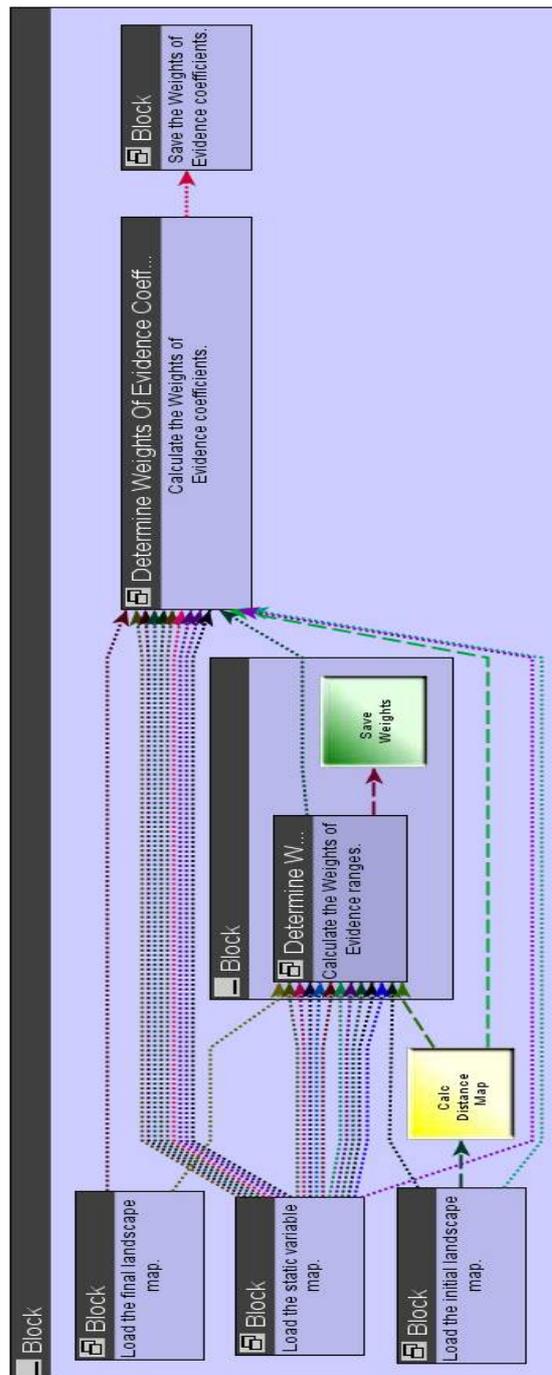


Fig 3-14: Rappresentazione grafica del modello 02-WoE_range_coeff.xlm, realizzato per il calcolo dei pesi dell'evidenza, dato un gruppo di n variabili esplicative e le mappe categoriche di uso del suolo in due istanti precedenti t_{-1} e t_0 alla data da simulare t_1 .

Il metodo WoE presenta il vantaggio, rispetto agli altri metodi statistici come la regressione lineare o logistica, di non essere soggetta alle ipotesi e restrizioni classiche degli approcci parametrici, spesso violate dai dati spaziali; inoltre l'effetto di ogni variabile spaziale può essere calcolato

separatamente. L'unica ipotesi che deve essere fatta riguarda l'indipendenza spaziale delle mappe predittive; questa è verificata dal modello *03-WoE_correlazione.xlm* (Fig.3-15 e CD allegato), attraverso dei test basati sulla comparazione di coppie di mappe (coefficiente di Cramer, coefficiente di contingenza e *Joint Information Uncertainty*²⁵). Come risultato di questi test una fra le variabili correlate può essere eliminata o combinata con la seconda formando una nuova grandezza che sostituisce le due precedenti nella modellizzazione.

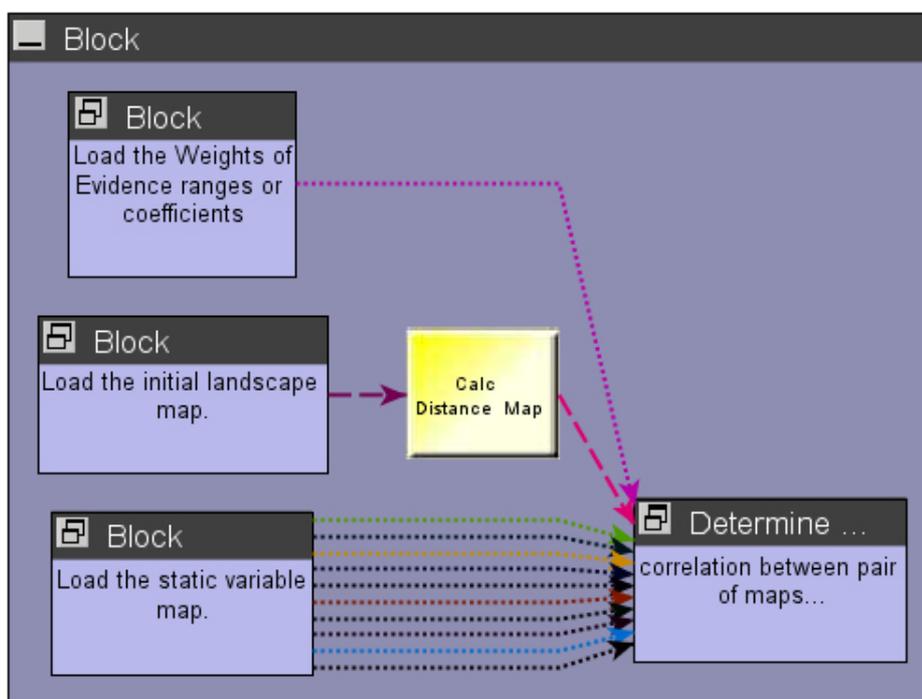


Fig 3-15: Rappresentazione grafica del modello 03-WoE_correlazione.xlm, per calcolare la correlazione fra coppie di mappe predittive.

Riassumendo l'uso integrato dei modelli *02-WoE_calcolo_ranges_coeff.xlm* e *03-WoE_correlazione.xlm* permette dunque di categorizzare variabili continue, verificare il grado di correlazione delle variabili esplicative e calcolare i pesi dell'evidenza rispetto ad ogni criterio

²⁵ Per una descrizione dettagliata di questi metodi si rimanda a [Bonham-Carter, 1994].

ambientale; l'output è registrato in un file testo (.txt) [Follador et al., 2008].

3.2.2.3 La simulazione dell'uso del suolo

Per realizzare la simulazione di LUCC in Dinamica Ego abbiamo costruito un modello *04-Run_simulazione.xlm* (Fig.3-16 e CD allegato), che permette di riprodurre l'evoluzione dei patterns spaziali prendendo in considerazione un grande numero di parametri descritti dalle funzioni utilizzate, fra cui [Follador et al., 2008]:

- ✓ *Select percent Matrix*: in Dinamica EGO le regole locali dell'automa cellulare sono determinate da due funzioni complementari: *Expander* e *Patcher*. In questo blocco noi specifichiamo la percentuale di pixels che devono essere modificati attraverso *Expander* che permette solamente l'espansione o la contrazione di parcelle appartenenti ad una certa classe, già esistenti (un'area di pascolo che si espande progressivamente a scapito della foresta, ad es.). La percentuale ($1-Expander\%$) è utilizzata in *Patcher* per creare nuove parcelle attraverso un meccanismo di "semina"²⁶ (nuovi piccoli poligoni di deforestazione, ad es.). L'operatore *Patcher* permette la transizione da uno stato $i \rightarrow j$ solamente in un intorno dei pixels che hanno uno stato differente da j [Almeida, 2003]. Per lo studio in Peten abbiamo scelto una percentuale *Expander* di 50% per la deforestazione e 70% per riforestazione (la riforestazione si manifesta come progressiva contrazione di aree defoestate ora abbandonate).

26 Seeding mechanism.

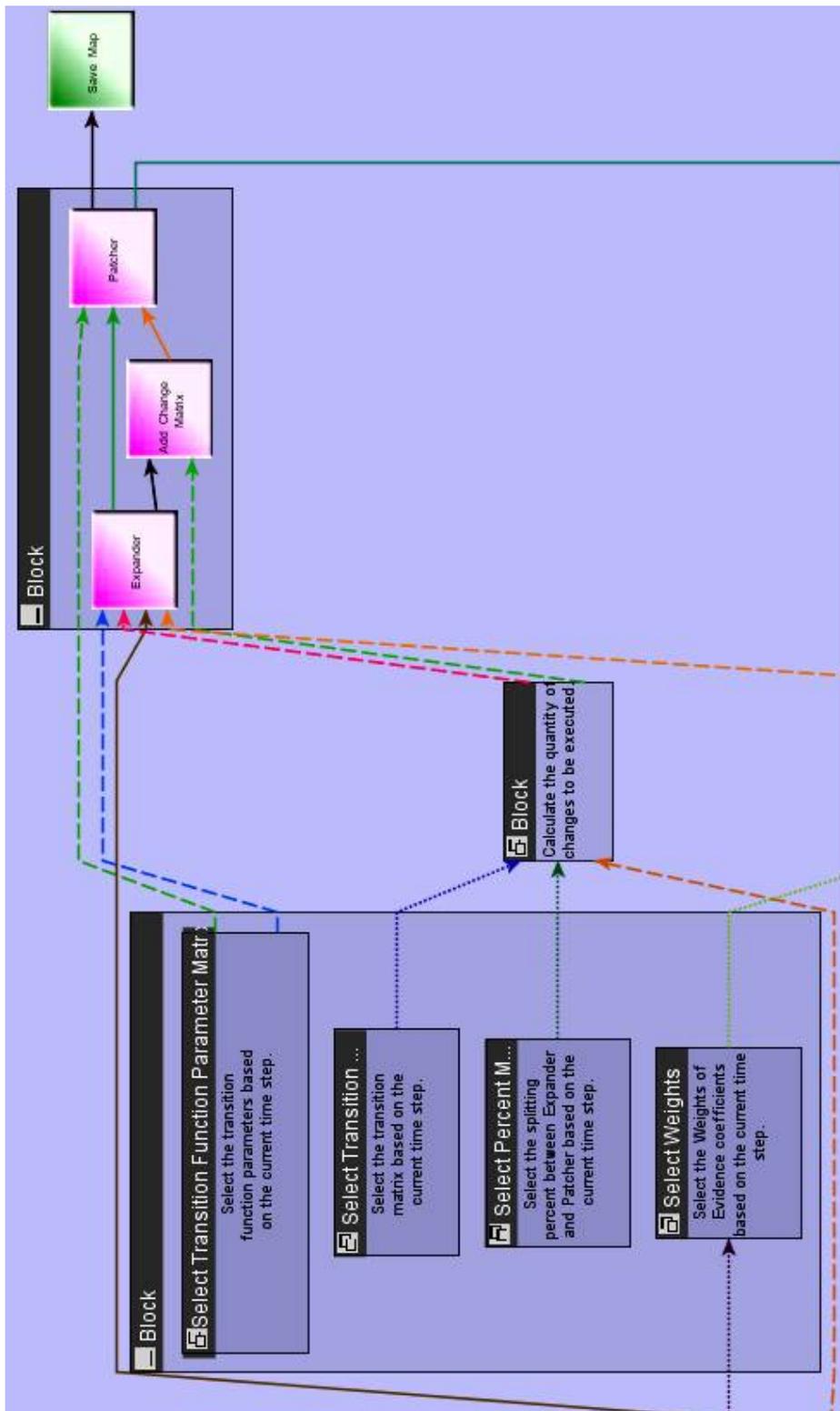


Fig 3-16: Zoom su 04-Run_simulazione.xlm. Evidenziate le funzioni che permettono di definire i più importanti parametri per una corretta rappresentazione dell'evoluzione dei patterns spaziali (isometria della parcella, dimensione media, regole di transizione dell'automa cellulare, etc.).

- ✓ *Select transition Parameter Matrix* permette di regolare i tre parametri utilizzati da *Expander* e *Patcher*: la

dimensione media (ha) dei gruppi di pixels che saranno modificati da ogni transizione, la varianza sulla dimensione media ed il grado d'isometria della parcella. Aumentando la dimensione media delle aree soggette a modificazione si ottiene un paesaggio meno frammentato e più uniforme; la varianza stabilisce di quanto può variare la dimensione di una parcella rispetto al valore medio stabilito. Infine l' isometria determina il grado di aggregazione dei pixel permettendo una rappresentazione più realistica dei patterns spaziali di ogni classe (valori >1 indicano forte coesione fra le celle e dunque un poligono compatto). Questi parametri sono importanti per visualizzare correttamente la frammentazione del paesaggio, spesso ignorata nella fase di validazione che privilegia una verifica quantitativa e di allocazione spaziale dei risultati. La frammentazione del paesaggio (influenzata dalla forma e dalla scarsa coesione delle parcelle deforestate) determina un aumento della mortalità della vegetazione, un deterioramento dello stato di salute della foresta e contribuisce ad una parte importante di emissioni di gas ad effetto serra (CO₂ e CH₄) [DeFries et al., 2007; Laurence et al., 1998]. Per lo studio delle dinamiche di LUCC in Peten noi abbiamo scelto una dimensione media di 3(ha) e 1(ha), una varianza di 2(ha) e 1(ha) ed un'isometria di 1,3 e 1,1 rispettivamente per deforestazione e riforestazione.

Dinamica EGO modella le transizioni da uno stato $i \rightarrow j$ considerando inizialmente la mappa di probabilità di cambio di uso del suolo, calcolata con WoE; i pixels con la probabilità più alta sono ordinati in un array. La selezione della cella centrale di una nuova parcella (gruppo di pixels che prevediamo cambieranno) viene effettuata leggendo l'array ordinatamente dall'inizio alla fine. Successivamente un certo insieme di pixels, attorno alla cella centrale, è scelto in base alla loro

probabilità di transizione P_{ij} [Cuevas et al., 2007]. Questo meccanismo è adottato da entrambi le funzioni *Expander* e *Patcher*.

Riassumendo potremmo dire che in Dinamica EGO le probabilità di cambio di uso del suolo, calcolate con il metodo WoE, guidano l'automa cellulare durante la simulazione (allocazione spaziale), mentre il metodo delle catene di Markov è utilizzato per definire le percentuali di cambiamento per ogni classe di uso del suolo (quantità).

3.3 Il modello Geomatico combinato CA_Markov

Questo modulo di Idrisi Kilimanjaro e Idrisi Andes (Clark Labs, USA) sarà qui brevemente presentato, poiché il lavoro svolto si è limitato alla riparametrizzazione di un modello già esistente al fine di adattarlo ai dati in esame. Per un approfondimento sul modello CA_Markov e sulle funzioni-operatori che adotta, si rimanda a [Paegelow et al., 2005] oppure al manuale di Idrisi [Eastman, 2003 & 2006].

A differenza dei metodi puramente matematici, il modello geomatico combinato permette di integrare l'analisi statistica dei dati con l'opinione e l'esperienza del modellizzatore o del geografo, al fine di evidenziare le relazioni fra le dinamiche ambientali osservate e le variabili esplicative. CA_Markov combina tre strumenti modellizzativi: una valutazione multi-criterio (MCE²⁷) per calcolare le mappe di probabilità per ogni categoria della variabile studiata (uso del suolo); un'analisi di Markov di secondo ordine per il calcolo delle percentuali di transizione; l'ultima tappa prevede l'integrazione delle precedenti informazioni attraverso una valutazione multi-

27 MCE: Multi-criteria evaluation.

obbiettivo (MOLA²⁸) e l'utilizzo di un automa cellulare (una sorta di filtro di contiguità) per garantire una rappresentazione realistica dei patterns spaziali.

Durante la fase di calibrazione del modello utilizziamo le mappe classificate di uso del suolo degli ultimi due istanti precedenti alla data di riferimento per la validazione; i dati di "allenamento" sono completati dalla lista di variabili esplicative (fattori ambientali, sociali, etc.). Il metodo adottato può essere considerato come un modello supervisionato, che oppone ad un approccio automatico come quello proposto da PNNET, una simulazione guidata dall'opinione e dalle conoscenze del modellizzatore o di un esperto del settore. CA_Markov deriva dall'integrazione di 3 strumenti disponibili in Idrisi Kilimanjaro e Idrisi Andes [Follador et al., 2008]:

✓ *MCE - Allocazione spaziale delle percentuali di cambio.*

Questo metodo crea delle mappe di probabilità di realizzazione per ogni classe di uso del suolo, basandosi su regole che legano le variabili indipendenti al fenomeno studiato. Queste regole possono essere stabilite sia attraverso tecniche statistiche come la regressione logistica o l'analisi delle componenti principali, sia in base all'esperienza ed all'opinione del modellizzatore. Le mappe di probabilità sono utilizzate poi per l'allocazione spaziale delle percentuali di transizione calcolate con Markov.

I criteri ambientali comprendono sia le maschere binarie, che escludono dalla simulazione le aree che non interessano, sia i fattori, variabili nello spazio, che influenzano il grado di favorevolezza per ogni classe di uso del suolo (Tab.3-2).

28 MOLA: Multi-objective Land Allocation.

Explanatory Variables (Factors)	Type
Distance from roads in 2000	Dynamic
Distance from roads in 2002	Dynamic
Distance from cleared areas 1998	Dynamic
Distance from cleared areas 2000	Dynamic
Distance from Pipeline	Static
Distance from Rio San Pedro	Static
Distance from Southern lakes	Static
Distance from southern river	Static
Distance from villages	Static
Digital Elevation Model - high/low lands	Static

Mask	Reason
Archeological sites	Protected areas
Water and wet lands	Excluded

Tab 3-2: Criteri ambientali adottati nello studio delle dinamiche di deforestazione in Peten. I criteri ambientali includono i fattori, cioè mappe continue che descrivono la variabilità spaziale delle grandezze considerate, e le maschere binarie, che oscurano le aree che non consideriamo nello studio.

Ogni grandezza è rappresentata attraverso le unità di misura più adatte (distanza in metri, pendenza in gradi, etc.); i fattori devono essere dunque standardizzati per poterli successivamente comparare e analizzare assieme. Standardizzare significa riportare tutte le informazioni su una scala di misura comune, in questo caso una scala byte da 0 a 255 (scala di toni di grigio). Questa operazione può essere fatta manualmente o attraverso la funzione *fuzzy*. Successivamente i fattori sono ponderati a coppie attraverso la matrice di Saaty [Saaty, 1977]; questo procedimento è realizzato dalla funzione *Weight*, ricordando che $\sum p_i = 1$. I fattori sono poi ordinati per peso crescente. All'interno della successiva funzione *MCE-OWA* scelgo il grado di rischio e di compensazione che vogliamo adottare durante la produzione delle mappe di idoneità (Fig.3-17); questa operazione porta alla definizione di un secondo gruppo di pesi-contestuali, ricordando che $\sum pc_i = 1$



Fig 3-17: Strategia utilizzata in MCE-OWA. Abbiamo scelto di lavorare con una compensazione fra fattori intermedia e un rischio basso. Il vertice superiore del triangolo rappresenta la situazione di combinazione lineare dei fattori, caratterizzata da pesi-contestuali uguali.

Scegliamo di restare fra la condizione di rischio minimo (AND) e la semplice combinazione lineare (vertice del triangolo che attribuisce a tutti i fattori un peso contestuale uguale). In altre parole abbiamo deciso di attribuire al fattore con "p" minore un "pc" maggiore; quest'ultimo diminuisce progressivamente fino al fattore con "p" maggiore, a cui attribuisco un "pc" minimo. Per la costruzione della mappa di idoneità alla deforestazione in Peten abbiamo ottenuto i seguenti risultati (Tab.3-3):

Fattori - Deforestazione	Weight	OWA
DEM- modello di elevazione digitale	0.04	0.26
Distanza pipeline	0.05	0.25
Distanza deforest.in 1998	0.06	0.17
Distanza rio San Pedro	0.07	0.15
Distanza deforest.in 2000	0.10	0.08
Distanza infrastrutture in 2000	0.12	0.02
Distanza fiumi sud	0.13	0.02
Distanza laghi sud	0.13	0.02
Distanza villaggi	0.13	0.02
Distanza infrastrutture in 2002	0.17	0.01

Tab 3-3: Pesì e pesi-contestuali utilizzati in CA-Markov per la creazione della mappa di idoneità alla deforestazione nell'area test in Peten. Questo risultato deriva dalla combinazione delle funzioni Weight e MCE-OWA di Idrisi.

✓ *Catene di Markov - analisi quantitativa del cambio*

Come in Dinamica EGO, utilizziamo l'analisi di Markov di secondo ordine²⁹ per calcolare il valore della variabile studiata all'istante t_{+1} (2003) in funzione del valore all'istante t_0 (2000) e t_{-1} (1998). Le matrici di transizione registrano la probabilità di ogni classe di copertura di cambiare verso un'altra classe, ed il numero di pixels che saranno modificati (Tab.1). L'algoritmo genera anche un gruppo di mappe di probabilità condizionale per ogni categoria di uso del suolo, evidenziandone la probabilità di realizzazione, per ogni pixel, dopo un determinato numero di passi temporali. Questo calcolo è effettuato partendo dalla ultima data disponibile t_0 (2000).

✓ *MOLA - Allocazione dei pixels multi-obiettivo*

L'allocazione spaziale delle percentuali di transizione di ogni classe di copertura viene realizzata attraverso un automa cellulare utilizzando le mappe di idoneità calcolate con MCE. L'algoritmo MOLA integra tutte le informazioni precedentemente calcolate e, bilanciando i possibili stati di uso del suolo per ogni pixel, realizza la mappa predittiva finale. Le due classi milpa e foresta sono introdotte nella procedura MOLA come obiettivi, con peso uguale. Ogni obiettivo è definito utilizzando la corrispondente mappa di idoneità calcolata con MCE, ed il numero di pixel per ogni categoria, definito con MCA. MOLA è un processo iterativo che assegna ogni pixel con elevata probabilità di appartenenza, alle rispettive classi di uso del suolo, fino ad esaurire tutti i pixel previsti per ogni categoria. Quando un pixel presenta un'elevata probabilità di realizzazione per più classi, sarà assegnato alla categoria che presenta il valore più elevato [Berger et al., 2005].

29 Per maggiori dettagli sulle Catene di Markov si rimanda al Capitolo 3.2.2.1 di questa tesi.

Infine applichiamo un filtro spaziale di contiguità; l'automa cellulare utilizzato iterativamente 10 volte diminuisce la probabilità di realizzazione dei pixels isolati favorendo la creazione di parcelle più compatte.

3.4 Land Cover Modeler LCM

Land Cover Modeler è il nuovo modulo integrato di Idrisi Andes (Clark Labs, USA), per lo studio delle traiettorie paesaggistiche e delle dinamiche di uso del suolo. LCM aggrega differenti strumenti per analizzare i LUCC passati, calcolare la probabilità di cambiamenti futuri e l'evoluzione del fenomeno studiato, il suo impatto sulla biodiversità e sull'equilibrio ecologico, integrando anche eventuali strategie e pianificazioni politiche future nella simulazione.

Analogamente a quanto fatto per CA_Markov, presentiamo qui una rapida descrizione di LCM, poiché il lavoro si è limitato ad una riparametrizzazione di un modello già esistente per adattarlo ai nostri dati ed obiettivi. Inoltre gran parte della teoria di modellizzazione utilizzata da LCM è stata ampiamente trattata nei capitoli precedenti (reti neurali, analisi di markov). Per un ulteriore approfondimento del metodo e delle funzioni che adotta, si rimanda al manuale di Idrisi Andes [Eastman, 2006].

La funzione *Land Cover Change analysis*, primo passo del processo modellizzativo con LCM, analizza i LUCC fra 1998 ed il 2000, scelte come date per l'"apprendimento". È possibile visualizzare facilmente la quantità guadagnata e persa da ogni classe di uso del suolo nel periodo di studio ed i trend per ogni transizione.

La seguente funzione *Land Cover Change modeling* permette di verificare le relazioni e l'utilità informativa di ogni variabile esplicativa rispetto alle due dinamiche studiate

(deforestazione e riforestazione) attraverso un test di Cramer (vedi capitolo 4.2). Le variabili quantitative utilizzate possono essere statiche o dinamiche (Tab.2). Le prime sono costanti ed esprimono la favorevolezza per ogni transizione. Le variabili dinamiche cambiano nel tempo (distanza dalle aree deforestate, in continua espansione, ad es.) e devono essere dunque ricalcolate ad ogni iterazione durante il periodo simulato. Una volta terminata questa fase, passiamo al calcolo delle mappe di probabilità di transizione utilizzando un Perceptrone Multi-strato³⁰ (MLP). Otteniamo dunque, per ogni pixel, la probabilità che in esso si abbia riforestazione e deforestazione. Per la costruzione del data base di "allenamento" della rete realizziamo un campionamento aleatorio sia sulle celle che cambiano nell'intervallo studiato (1998-2000), sia su quelle che rimangono invariate. Metà dei pixels così selezionati sono successivamente utilizzati per il calcolo dei pesi sinaptici e l'ottimizzazione della rete; la seconda metà invece è utilizzata per la validazione e la verifica dei risultati. Abbiamo costruito una rete neurale formata da 10 neuroni sullo strato di Input, uno strato nascosto con 4 neuroni e un output con due nodi che restituiscono la probabilità di deforestazione e riforestazione per ogni pixel. Il MLP è stato parametrizzato con un *learning rate* di 0.000121 ed un *momentum factor* di 0.5. Dopo 5000 interazioni abbiamo raggiunto una precisione di 87%, valore considerato accettabile come suggerito in [Eastman, 2006]. La successiva fase di classificazione permette di creare le mappe di probabilità di deforestazione e riforestazione.

La funzione *Land Cover Change prediction* permette di integrare l'analisi di Markov di secondo ordine, che dona le quantità di pixels soggetti a cambiamento (vedi capitoli

30 Il Perceptrone Multi-strato (Multi-layer perceptron, MLP) è stato ampiamente descritto nel capitolo 3.1 di questa tesi, riguardante le reti neurali ed il modello PNNET.

precedenti), con le mappe di idoneità calcolate con MLP, per l'allocazione spaziale degli stessi. Abbiamo prodotto una carta definitiva di uso del suolo (*hard*) in cui ogni pixel appartiene ad una classe specifica, e una mappa *soft* che registra la probabilità di appartenenza di ogni pixel alle possibili classi di uso del suolo, evidenziando così le zone più vulnerabili e più colpite dalla deforestazione.

CAPITULO No. 4

L'area di studio in Peten e la creazione di un data base geografico

Allora questa volta possiamo iniziare davvero. Arrivano da tutte le parti, i barbari. E un po' questo ci confonde, perché non riusciamo a tenere in pugno l'unità della faccenda, un'immagine coerente dell'invasione nella sua globalità. Ci si mette a discutere delle grandi librerie, dei fast-food, dei reality show, della politica in televisione, dei ragazzini che non leggono, e di un sacco di cose del genere, ma quello che non riusciamo a fare è guardare dall'alto, e scorgere la figura che gli innumerevoli villaggi saccheggianti disegnano sulla superficie del mondo. Vediamo i saccheggi, ma non riusciamo a vedere l'invasione. E quindi a comprenderla.

Credetemi: è dall'alto, che bisognerebbe guardare. E' dall'alto che forse si può riconoscere la mutazione genetica, cioè le mosse profonde che poi creano, in superficie, i guasti che conosciamo.

(Alessandro Baricco. I barbari)

4.1. La Joyanca e le problematiche ambientali in Peten

4.1.1. Introduzione

L'area test su cui si sono concentrati i nostri sforzi modellizzativi, è una piccola regione al Nord-Ovest del Peten chiamata la Joyanca (Fig.4-4).

Dopo decenni di regime autoritario sfociato in una lunga guerra civile, il Guatemala ha intrapreso verso la fine degli anni '80 un processo di democratizzazione conclusosi con la firma degli Accordi di Pace nel 1996 [Effantin, 2006]. Questa transizione è caratterizzata da una nuova ingerenza internazionale giustificata da una crescente preoccupazione ecologista a livello mondiale; dopo 40 anni di politica coloniale agraria anarchica, nel 1989 il governo decide di adottare una prima legge di conservazione delle foreste tropicali e di creare il Consiglio Nazionale delle Aree Protette (CONAP).

A partire dagli anni '60 il Peten, uno dei 22 dipartimenti del Guatemala, che rappresenta da solo un terzo della superficie del paese, è diventato la principale destinazione della migrazione rurale nazionale [Effantin et al., 2004]. Questo movimento di colonizzazione di un'area marginale è stato incentivato dal governo, che voleva da un lato risolvere i conflitti sociali legati ai senza terra e ad un accesso ineguale alle risorse naturali, dall'altro garantire una presenza umana ed un controllo di un'area frontiera (Fig.1).

Il Peten, e più in particolare la regione sud del dipartimento, ha perso negli ultimi 40 anni quasi il 45% della sua superficie forestale; *"Los bosques son ahora raros en el sur del Peten y así mismo el desarrollo"*¹ [Nittler et al., 2005, Pag.5]. Dalla maggior parte dei boschi si ricavavano

1 I boschi nel sud del Peten sono oggi rari e analogamente possiamo dire dello sviluppo.

specie pregiate come il Caoba e il Cedro Spagnolo; una volta esaurita la propria attività estrattiva, l'industria del legno abbandonava l'area incurante di prevenire una eventuale conversione della foresta in pascolo o campo ad opera degli immigrati.

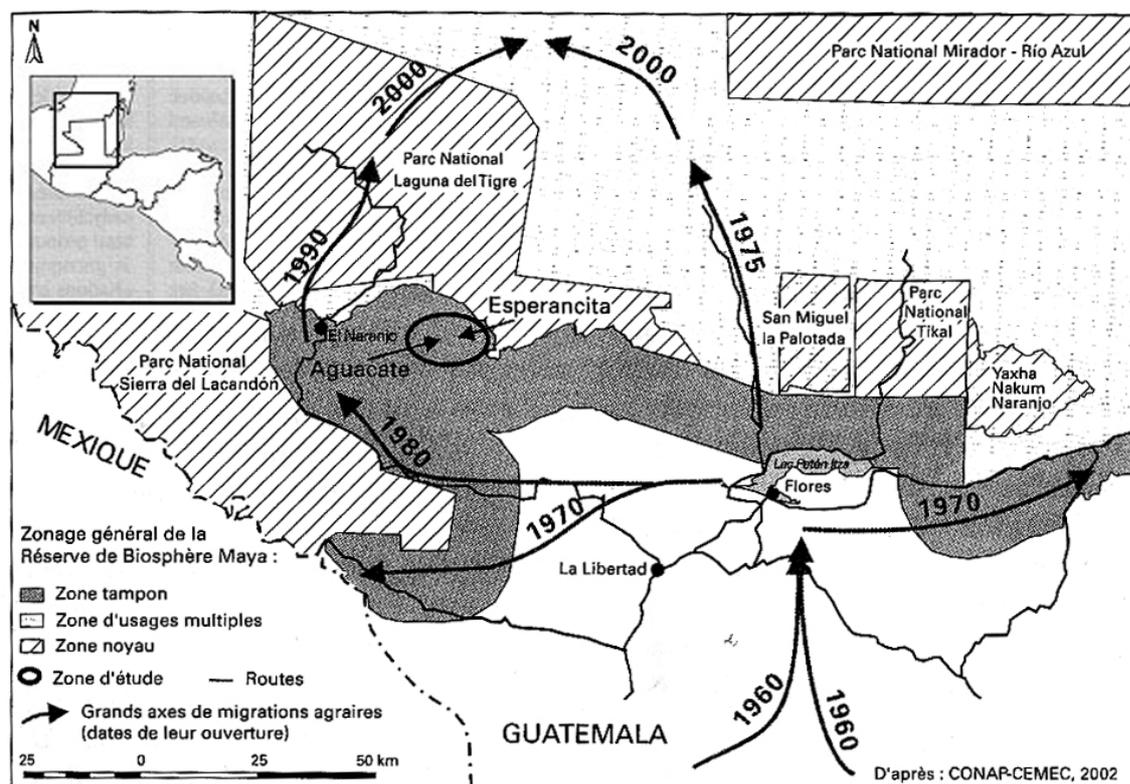


Fig 4-1: Migrazioni rurali, Riserva della Biosfera Maya e localizzazione dell'area di studio (evidenziata dall'elisse) in Peten. [Immagine da Effantin et al., 2004].

Per salvare il nord del Peten dallo stesso destino, nel 1990 il governo del Guatemala decide di creare numerose aree protette sul 70% della regione. La Riserva della Biosfera Maya (RBM) occupa, con i suoi 2,1 milioni di ettari (767.000 ha di area protetta, 848.440 zona di uso multiplo e 497.500 di zona di transizione) [Nittler et al., 2005], la metà del dipartimento e rappresenta la più grande superficie continua di foresta tropicale del Centro America [FIPA/AID-CONAP, 2002].

Come già detto, queste riserve furono create per rispondere alla crescente pressione di paesi stranieri, particolarmente di USA e Germania; la pianificazione fu minima e la popolazione locale fu scarsamente consultata. L'obbiettivo era prevenire, nel più breve tempo possibile, una nuova invasione incontrollata dei contadini che scappavano dalla povertà del sud sovrappopolato, dai grandi proprietari terrieri interessati a monopolizzare le risorse naturali e da fenomeni di violenza legati alla guerriglia².

La RBM è strutturata secondo il modello universale delle riserve di biosfera dell'Unesco, che lega obiettivi di gestione ambientale ai differenti livelli di protezione e restrizione d'uso del bosco. Essa è formata da (Fig.1):

1. Zona centrale (*zona nucleo*) messa sotto protezione totale e soggetta alle sole attività di ricerca, conservazione e turismo sostenibile.
2. Zona di transizione (*zona de amortiguamiento*) che deve proteggere la zona centrale dalle eventuali ondate migratorie di senza terra o di grandi proprietari a caccia di nuove terre. L'agricoltura è permessa con restrizioni per stabilizzare la frontiera agraria.
3. Zona di uso multiplo (*zona de usos multiples*) che protegge la zona centrale grazie a dei contratti fra la CONAP ed i gruppi di utilizzatori per garantire una gestione sostenibile della foresta.

Più di 80.000 persone vivono in questa regione [Effantin et al., 2004]; l'introduzione di nuovi regolamenti e procedure di regolarizzazione dei possedimenti ha provocato, fra le

² Con il termine immigrati indichiamo in questa tesi l'intera popolazione formatasi in Peten a partire dagli anni '60 dall'integrazione di gruppi nativi "puros peteneros" con delle comunità arrivate successivamente, i "ladinos", che parlano solamente lo spagnolo. Per una descrizione più dettagliata si rimanda a [Effantin, 2006, Pag.19].

differenti comunità installate in Peten, numerosi conflitti e tensioni. L'importante per il CONAP era di far sentire la presenza dello Stato in un'area di illegalità diffusa, limitando l'invasione agricola e la pastorizia nella foresta tropicale, il traffico di droga e di specie protette, gli incendi forestali ed il saccheggio dei siti archeologici. La soluzione fu incontrata dando alle comunità locali ed alle industrie delle concessioni a lungo termine (25 anni) nella zona di uso multiplo e dunque la responsabilità legale del bosco; questa scelta ha portato in breve ad una diminuzione importante dei fenomeni illegali di deforestazione, di incendio ed appropriazione della terra, garantendo una gestione più sostenibile dell'ecosistema tropicale.

Ovviamente questa analisi su grande scala può nascondere ed oscurare gli agenti e le variabili che giocano un ruolo importante nell'evoluzione dei processi locali e che, per sinergia, danno origine alle proprietà emergenti precedentemente descritte. Il nostro studio si concentrerà ora su una piccola regione del Peten che si trova a cavallo fra la zona di transizione e la zona centrale; osserveremo le dinamiche di LUCC che si svilupparono dal 1987 al 2003, cercando di comprendere le relazioni fra i fattori chiave e cambiamenti di uso del suolo, facilitando così la costruzione di carte predittive per verificare l'evoluzione dei patterns spaziali in relazione agli obiettivi della RBM.

4.1.2 La Joyanca: ambiente e geomorfologia

La nostra area test, definita attorno al sito archeologico della Joyanca, si trova nella parte a nord-ovest del Peten, che si caratterizza per un ambiente semplice e ricorrente in cui l'acqua occupa una posizione preponderante [Arnauld et al., 2004]. Questa zona si trova all'interno della regione di *bosque humedo subtropical* [FAO, 2002; Cruz, 1976], con una

pluviometria media annuale compresa fra 1500 e 2500 mm ed una temperatura media di 25°C, che permette lo sviluppo di una fitta vegetazione semidecidua. Il Peten è caratterizzato da un paesaggio lievemente ondulato (Fig.4-3) compreso fra 100 e 300 m sul livello del mare; il suolo è formato da calcari evaporitici e da dolomite micro-granulare [Métailié et al., 2000], con elevata presenza di argille nelle aree inondabili. Il manto freatico alimenta numerose lagune e contribuisce al drenaggio superficiale attraverso importanti corsi d'acqua come il Rio San Pedro (Fig.4-2).



Fig 4-2: La Joyanca è anche chiamata la regione delle lagune. L'acqua è un elemento preponderante di questo paesaggio. [Foto di Métailié J.P., GEODE-CNRS].

“Los ambientes de la pequena region de la Joyanca se encuentran organizados, de norte a sur, de acuerdo con una alternancia de mesetas inclinadas hacia el norte, de 100 a 300 m de altitud, y de bajos pantanosos o lacustres, hacia los 50-70m. Se hallan generalmente limitados por fallas este-oeste, las cuales han originado un relieve disimétrico, con largas vertientes de escara inclinacion hacia el norte y frentes de cuesta cortos y empinados que miran hacia el sur y dominan los pantanos y las lagunas; el sitio de la Joyanca se encuentra en una de estas vertientes con muy poco desnivel, a orillas del

frente sur que domina un amplio cibal"³ [Arnauld et al., 2004, Pag. 37].

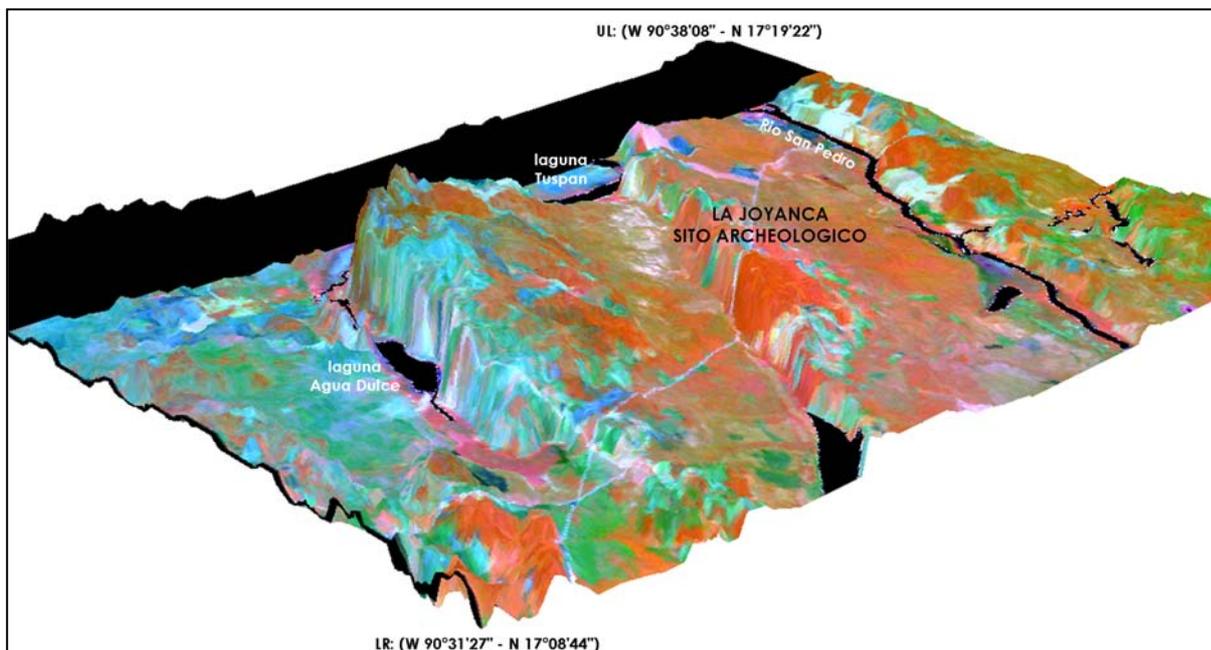


Fig 4-3: Immagine RGB 4-5-2 (ETM, 20-48, 07/05/03) stesa su un modello digitale del terreno ricavato da una immagine Shuttle Radar (il DEM ha un fattore di esagerazione verticale=2, per evidenziare i versanti e la morfologia del paesaggio lievemente ondulato).

Il paesaggio risulta particolarmente frammentato ed a volte è difficile definire i limiti esatti dei vari elementi che lo compongono: foresta tropicale densa, foresta secondaria e campi abbandonati (detti *guamiles*, con differenti gradi di rigenerazione), foresta umida, orti famigliari, pascolo, suolo esposto ed ambienti lacustri.

Sui versanti possiamo distinguere due diverse entità (Fig.4-4): un'area elevata, caratterizzata da suoli calcarei fratturati, ben drenati e fertili, dove si sviluppa un bosco di grandi alberi, detto *selva alta*. In queste zone si trovano gli

³ I paesaggi della piccola regione della Joyanca sono organizzati, da nord a sud, secondo un'alternanza di altipiani inclinati verso nord, fra i 100 ed i 300 m, e da terre inondabili e lacustri, fra i 50-70m. Sono generalmente limitati da faglie est-ovest, che hanno dato origine ad un rilievo asimmetrico, con lunghi versanti poco inclinati verso nord e piccole scarpate ripide che guardano a sud e dominano le aree innondate e le lagune; il sito della Joyanca si trova su uno di questi versanti con piccolo dislivello, sul bordo di una scarpata che domina un ampio sibal.

antichi insediamenti Maya, che ben conoscevano il grande potenziale agricolo delle zone alte. A quote minori si riscontra la presenza di antichi letti fluviali, poco incisi ma di grande larghezza, che contribuiscono a drenare la *meseta* durante la stagione delle piogge, formando delle vere e proprie paludi. Queste regioni sono chiamate *bajos* e sono caratterizzate da suoli molto pesanti con più del 60% di argille che possono superare i 2 m di spessore, spesso inondati durante la stagione delle piogge. Il loro potenziale agricolo è debole e vi cresce un bosco più basso, meno denso ed un sottobosco di piante erbacee. In prossimità delle lagune dopo i lunghi periodi di pioggia si formano delle paludi più o meno temporanee, dette *cibales* dove prospera una vegetazione erbacea. Questi due ultimi ambienti, durante la stagione secca, ardono con molta facilità e sono sede di grandi incendi che spesso sfuggono al controllo degli immigrati [Arnauld et al., 2004; Métaillié et al., 2000].

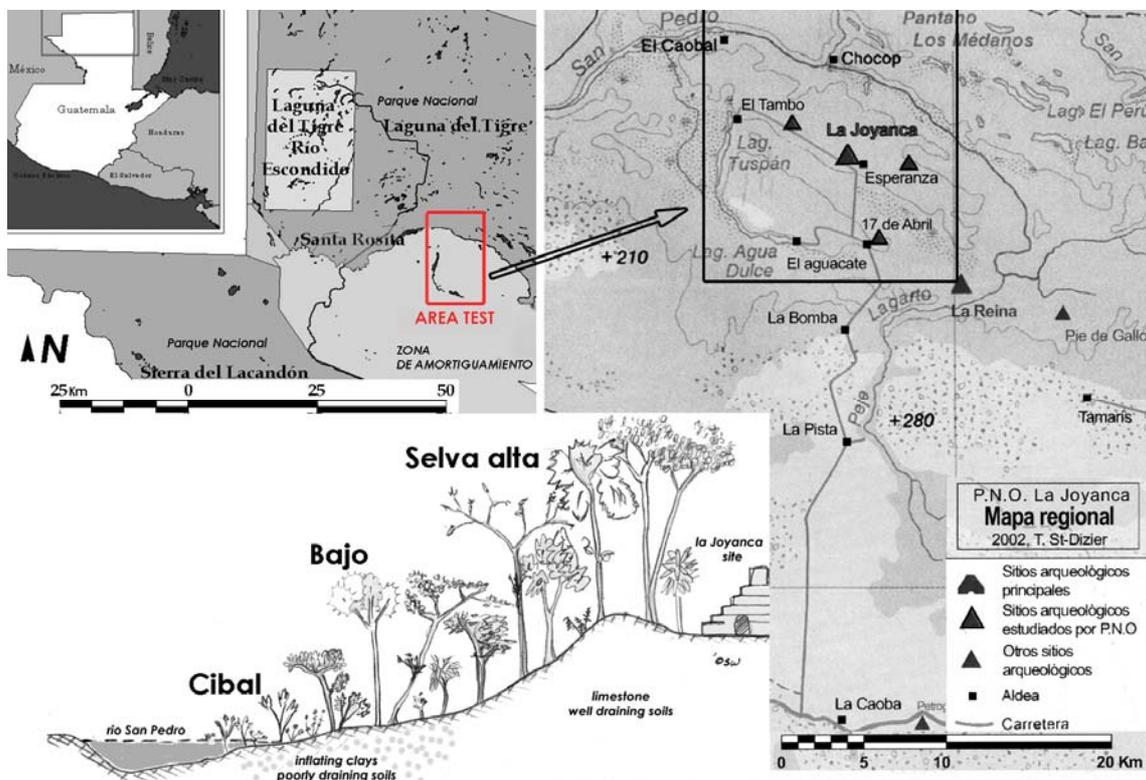


Fig 4-4: Contesto geografico e schema morfologico-vegetale, importante per capire le diverse fasi di appropriazione della foresta da parte degli abitanti. [Immagine adattata da Follador et al., 2008].

4.1.3 La colonizzazione de La Joyanca

La prima occupazione della regione e la sua quasi totale deforestazione inizia nel periodo Classico Maya (100-900 d.C) e termina con il collasso di questa civilizzazione, il conseguente abbandono delle terre ed un lento ritorno della densa foresta tropicale originale [Follador et al., 2006; Geoghegan et al, 2001].



Fig 4-5: L'agricoltura taglia-brucia, detta Milpa, è caratteristica della colonizzazione delle aree tropicali. La dimensione della milpa è in genere di un ettaro e garantisce principalmente la sussistenza della famiglia o comunità. [Foto di Métaillié J.P., GEODE/CNRS].

A partire dagli anni '80, periodo in cui si osservano i primi poligoni di deforestazione al nord del rio San Pedro, la colonizzazione è stata alquanto rapida. I coloni hanno rimpiazzato la copertura boschiva densa e continua con un nuovo sistema "agrosilvopastorale".

L'agricoltura *roza-quema*⁴ (Fig.4-5), detta ***milpa***, interessa generalmente aree di un ettaro ricavate dalla foresta primaria o secondaria.

*"...Lancadon continue to practice swidden farming (Fig.4-6). Milpa farming is centered around the Mesoamerican cultigens of corn, beans and squash, as well as supplemental crops of chiles, manioc and tomatoes...most Lancadon actively cultivate their Milpa for approximately three years, after which fields are allowed to return to jungle growth...Overgrown milpas are never completely abandoned. Instead they continue to be sources of food and resources...Finally Lancadon agriculture incorporates forest regeneration and re-growth as essential elements of production. After several years (from 3 to 7 years) jungle vegetation is allowed to engulf fields and new milpas are established elsewhere. All together, this agricultural system produces a diverse landscape..."*⁵ [Kashanipour et al., 2004, Pag.50].

4 L'agricoltura taglia e brucia (slash and burn agriculture) è una fonte importante di emissioni di gas ad effetto serra. Vedi rapporto ASB Program.

5 ... I Lancadon (popolazione Maya indigena del sud del Chiapas) continuano a praticare una forma di agricoltura mobile. La milpa è fondata sulle culture mesoamericane del mais, fagioli e zucche, ma anche su campi supplementari di peperoncino, manioca e pomodoro...la maggior parte delle attività dei Lacandon coltiva la milpa per 3 anni circa, dopo di che permette il ritorno della foresta. La milpa però non è mai completamente abbandonata e continua ad essere una fonte di cibo e risorse...Infine l'agricoltura dei Lancadon incorpora la rigenerazione della copertura forestale e la ricrescita del bosco come elemento essenziale nella produzione. Dopo alcuni anni di sfruttamento (3-7 anni) la giungla fagocita nuovamente i campi e la milpa è ricreata altrove. Tutte assieme queste attività agricole creano un paesaggio diversificato...".



Fig 4-6: Esempio di agricoltura-mobile. Dopo alcuni anni la milpa è abbandonata e si permette la ricrescita della foresta secondaria. Una nuova milpa è creata in un'area vicina. [Foto di Métaillié J.P, GEODE-CNRS].

Le terre migliori per la *milpa* sono quelle situate sulle parti più elevate delle *mesetas*, ben drenate e relativamente povere in argilla. Queste aree sono state divise, durante la fase di legalizzazione delle proprietà, all'inizio degli anni '90, fra le comunità indigene che praticano un'agricoltura sostenibile di sussistenza.

Nelle aree di *bajos* lo spesso strato argilloso passa rapidamente da un'umidità eccessiva alla siccità durante la stagione secca; il loro rendimento agricolo non è molto buono e per questo, oggi, sono adibite per lo più ad attività pastorizie (Fig.4-7). La dimensione delle aree deforestate è maggiore rispetto alla milpa (dipende dalla capacità della famiglia di acquistare nuovo bestiame ed espandere l'area di pascolo), e non vi sono in genere segni di riforestazione secondaria. L'attività pastorizia determina generalmente un forte impoverimento delle terre adibite a questo uso.



Fig 4-7: Attività pastorizie nelle aree di bajo, ricche in argilla e dunque poco adatte alla Milpa. La dimensione delle parcelle deforestate è maggiore e tende ad espandersi con gli anni, con limitati casi di riforestazione secondaria. [Foto di Métaillié J.P., GEODE-CNRS].

Iniziato verso la fine degli anni '80 il processo di colonizzazione ha portato, in meno di 20 anni, ad una totale trasformazione del paesaggio ed ad una notevole degradazione della foresta tropicale. In particolare i grandi incendi del 1998 e quelli del 2000 hanno notevolmente danneggiato la maggior parte degli ambienti fino ad allora preservati. Il fuoco rimane il mezzo privilegiato dai nuovi immigrati per liberare uno spazio nel denso manto vegetale (Fig.4-8).

Mapa 4
Incendios forestales en el Petén 2005

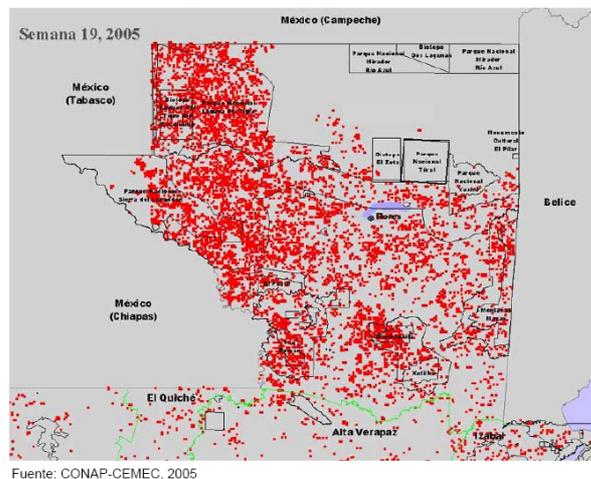


Fig 4-8: Mappa degli incendi registrati nel Peten nel 2005. La maggior parte di questi incendi sono dolosi e legati alla milpa; i contadini perdono spesso il controllo del fuoco che loro stessi hanno appiccato e l'incendio può propagarsi su grandi superfici portando ad una rapida degradazione dell'ambiente originario. [Immagine da Gomez et al., 2005]

Oggi anche le zone meno adatte alle attività agricole o pastorizie come il *cibal*, sono oggetto di interesse da parte degli abitanti o degli ultimi arrivati, desiderosi di trovare nuove terre da poter utilizzare; si registrano dunque fenomeni di deforestazione caotici, sparpagliati, spesso guidati dall'arbitrarietà o dalla disperazione del contadino, piuttosto che da una logica di potenzialità produttiva.

Infine apriamo una piccola parentesi sulle terre a nord del Rio San Pedro. In questa zona si sono stabiliti i primi insediamenti verso il 1988; l'utilizzazione del suolo per agricoltura di sussistenza e pastorizia si è poi progressivamente espansa verso nord e lungo il letto del fiume. Nei primi anni del 2000 si è registrata una brusca frenata in questa tendenza, spiegata non da motivi biofisici od ambientali, ma dalle recenti attività legate al narcotraffico in queste aree.

La regione La Joyanca presenta dunque una complessità paesaggistica, di processi e di agenti, estremamente elevata. Modellizzare questa realtà risulta molto difficile ed a volte utopico, poiché non sempre le semplificazioni ed i formalismi adottati nella rappresentazione dei LUCC, riescono a catturare il caos e l'aleatorietà dei processi antropici.

4.2. Uso integrato di telerilevamento e SIG per la costruzione di un database geografico

4.2.1 Introduzione

Le immagini satellitari rappresentano un'importante fonte di dati periodica ed economica, in termini di tempo e denaro . L'utilità di queste informazioni è particolarmente evidente nello studio di aree di difficile accesso (foresta tropicale o montagne, ad es.) oppure con condizioni ambientali estreme (zone polari e deserti, ad es.). Il telerilevamento permette di

raccogliere dati in Paesi in cui le condizioni di instabilità politica e militare, attuale o passata, non hanno permesso di sviluppare un database sufficientemente dettagliato o affidabile⁶, e dove un eventuale lavoro sul campo risulterebbe troppo pericoloso o interdetto.

In Guatemala, l'Istituto Geologico Nazionale (IGN) ha cominciato solo recentemente a produrre le prime carte tematiche (vegetazione, geologia, etc.) su scala nazionale; la risoluzione adottata è troppo grossolana per poter essere utile in uno studio a livello locale. Analogamente, i primi dati numerici sulla regione del Peten, frutto del lavoro di ONG, organizzazioni statali⁷ ed internazionali, offrono una visione semplicistica e troppo generalizzata dei fenomeni di deforestazione e di uso del suolo nella regione La Joyanca.

Data l'assenza di informazioni dettagliate sull'area test, quasi tutti i dati che useremo come input per la modellizzazione, sono ricavati dall'analisi ed il trattamento di una serie di 11 immagini Landsat (ETM e TM, Path 20/Row 48) e SPOT (Path 606-607/Row 315)⁸ dal 1988 al 2003, e dai rilievi aerofotogrammetrici.

Il nostro progetto ha come obiettivo quello di evidenziare le variabili e le dinamiche strutturali che influenzano l'utilizzo del suolo, attraverso l'analisi delle informazioni satellitari e l'utilizzo dei dati raccolti *in situ* da [Effantin-Touyer, 2006]. Modellizzare questi processi antropici e legare gli effetti locali con i dati raccolti su una scala più ampia, attraverso un'operazione di upscaling in un SIG, può aiutare a meglio comprendere quali siano i fattori che

6 Vedi ad esempio il progetto Niger-HYCOS del WMO.

7 Fra queste ricordiamo il Ministerio de Agricultura, Ganaderia Y Alimentacion (MAGA), l'Unidad de Politicas e Informacion Estrategica (UPIE), l'Instituto Nacional de Bosques (INAB), la già citata CONAP e l'Asociacion de Comunidades Forestales de Peten (ACOFOP).

8 Immagini SPOT ottenute grazie all'accordo ISIS, Salleron/©CNES 2003.

influenzano realmente la degradazione di questo ambiente tropicale [Li, 2007; Geoghegan et al., 2001].

Le immagini sono state acquisite con periodicità irregolare in funzione della loro disponibilità sul mercato e privilegiando le scene con scarsa nuvolosità (Tab.4-1).

DATE	SATELLITE	CANALI	N° CANALI	PATH ROW	Risoluzione
(1°P) 07/05/2003	Ldt-ETM	TM	7	20-48	30m - p15m
(S) 30/03/2002	Spot4	XI	4	606-315	20m
(S) 30/03/2002	Spot4	M	1	606-315	p10m
(S) 27/03/2000	Ldt-ETM	TM	7	20-48	30m-p15m
(S) 27/03/2000	Spot4	XI	4	607-315	20m
(S) 10/02/2000	Spot4	M	1	607-315	p10m
(S) 24/02/1998	Spot2	XS	3	606-315	20m
(S) 30/03/1994	Spot3	P	1	607-315	p10m
(P) 24/11/1992	Ldt-TM5	TM	6	20-48	30m-p15m
(P) 05/11/1998	Ldt-TM4	TM	6	20-48	30m-p15m
(S) 15/02/1987	Foto aerea		1	1:30000	5m

(1°P): inizio stagione delle piogge

(P): stagione delle piogge

(S): stagione secca

Tab 4-1: Lista delle immagini satellitari ed aeree utilizzate.

Questa selezione delle immagini satellitari influenza indirettamente il periodo adottato per la calibrazione e la validazione dei metodi predittivi; le mappe scelte come informazione iniziale e finale della modellizzazione sono, infatti, il risultato del processamento e della classificazione dei dati telerilevati a disposizione. In condizioni ottimali (abbondanza di dati), i periodi analizzati dovrebbero far riferimento a condizioni socio-economiche simili, che garantiscano una certa stabilità delle variabili esplicative adottate [Almeida, 2003]. Se quest'ultime dovessero cambiare fra il periodo analizzato nella calibrazione del modello, ed il periodo simulato, i risultati che dovremmo attenderci non saranno certo rappresentativi della reale evoluzione dei patterns spaziali [Pontius et al., 2006; Pontius et al., 2004].

Le foto aree raccolte nel 1987 sono state digitalizzate con un dpi⁹=150 e assemblate in un mosaico con il programma ER Mapper. I punti adottati per il loro georeferenzamento, sono gli stessi utilizzati per la correzione geometrica locale delle immagini satellitari. Il passaggio dal valore numerico della scala 1:30000, alla risoluzione digitale in metri è stata realizzata attraverso la formula:

$$R[m] = \frac{0,0254}{dpi} * scala \quad (1)$$

dove 0,0254 rappresenta la costante di trasformazione fra pollici e metri (1 pollice = 2,54cm).

Non esistono per la regione della Joyanca indicatori socio-economici sottoforma di attributi numerici o di carte; le informazioni utilizzate per migliorare la simulazione fanno riferimento alle opinioni ed all'esperienza di sociologi e geografi che hanno in passato lavorato in quest'area.

Ciò ha permesso di creare una zona di bassa probabilità di deforestazione a nord del Rio San Pedro, dove si è registrata dal 2000 un'attività legata al narcotraffico, che scoraggia ed intimorisce i contadini (informazioni ottenute da Effantin-Touyer). I dati catastali sulla divisione delle terre attorno alla Joyanca fra le popolazioni indigene, definiscono un poligono, corrispondente all'area di concessione comunitaria, in cui si registra un ritmo di deforestazione minore ed un'attività agricola più sostenibile. La differenziazione fra suoli produttivi nella parte alta delle *mesetas* e suoli argillosi dei *bajos* fatta in [Arnauld et al., 2004; Métaillé et al, 2000], ci ha spinto a creare a una mappa di potenzialità agricola in funzione dell'altitudine: le terre più elevate sono interessate da attività basate sulla milpa, più sostenibili e

9 Dpi: *dot per inch*, punti per pollice. È una unità di misura della risoluzione ed esprime la quantità di punti stampati lungo una linea di 2,54cm.

che garantiscono una periodica riforestazione. Le terre umide in basso sono adibite alla pastorizia; la deforestazione interessa superfici maggiori e procede come espansione di aree di pascolo preesistenti.

4.2.2 Pre-processamento delle immagini satellitari

4.2.2.1 Acquisizione delle immagini e definizione del rettangolo involvente

Le informazioni satellitari sono state acquisite in mesi e stagioni a volte differenti; non è dunque possibile eliminare completamente la variabilità "scena-a-scena" dovuta alle condizioni atmosferiche ed alla fenologia. Le immagini del 05/11/88 e del 24/11/92 fanno riferimento alla stagione delle piogge in cui si registra un vigore vegetale ed un grado di umidità elevato. Confrontare questi dati con altri registrati durante la stagione secca può fornire false indicazioni sui cambi e sulle dinamiche ambientali realmente avvenute nel periodo in esame; l'analisi multitemporale deve, in questo caso, essere assistita dall'esperienza del geografo che integra i risultati ottenuti dal processamento delle immagini con le proprie opinioni per separare variabilità naturale ed impatto antropico.

*"Las características orbitales de un satélite permiten que las imágenes sean adquiridas periódicamente y en condiciones similares de observación, facilitándose así cualquier estudio que requiera una dimensión temporal"*¹⁰ [Chuvieco, 2002, Pag.195]. Nella realtà è difficile lavorare in queste condizioni ottimali; la possibilità per un ricercatore di accedere alle immagini satellitari che meglio si adattano ai propri obiettivi, è spesso limitata dai costi eccessivi o dalla disponibilità nel periodo richiesto. Il GEODE/CNRS ha

¹⁰ Le caratteristiche orbitali dei satelliti permettono un'acquisizione delle immagini periodica ed in condizioni simili di osservazione, facilitando così ogni studio che includa la dimensione temporale.

selezionato le scene per il Peten tenendo in considerazione questi due ultimi criteri; in particolare le condizioni meteo e la presenza frequente di nuvole, hanno rappresentato i principali ostacoli alla costruzione di un database più completo ed omogeneo.

Dalle immagini originali abbiamo ritagliato un poligono rettangolare (Fig.4-9) che include la nostra area test della Joyanca (circa 12 X 19.5 km²). I vertici del rettangolo sono definiti dalle seguenti coordinate:

$$\left\{ \begin{array}{l} \text{UL: (W } 90^{\circ}38'08'' \text{ - N } 17^{\circ}19'22'') \\ \text{LR: (W } 90^{\circ}31'27'' \text{ - N } 17^{\circ}08'44'') \end{array} \right.$$

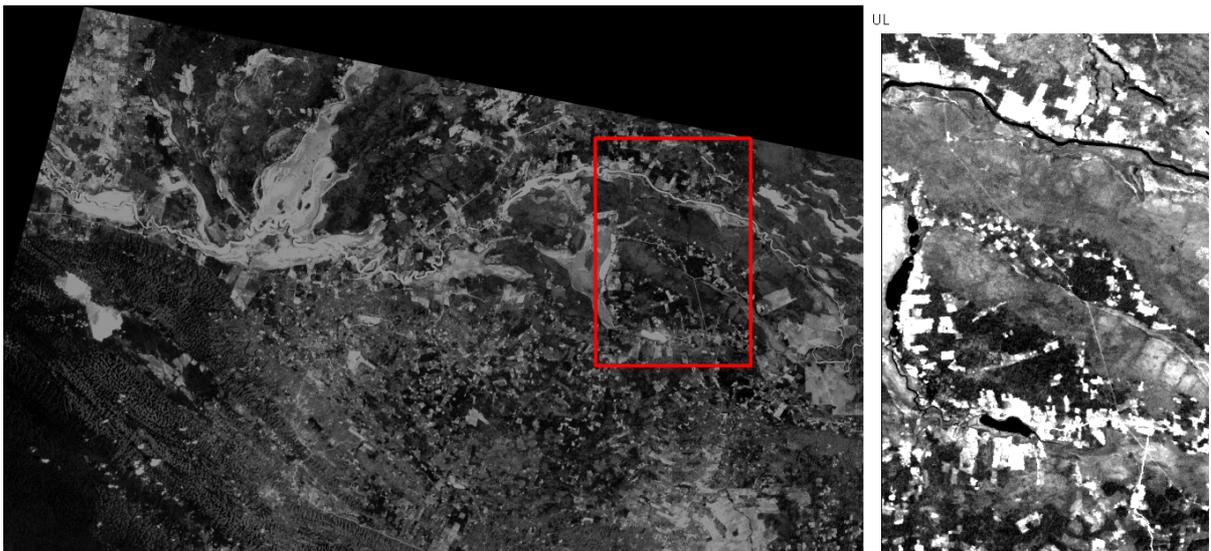


Fig 4-9: L'area di studio è definita da un rettangolo involvente, i cui vertici sono definiti dalle coordinate sopra indicate UL e LR. Ogni poligono è dunque ritagliato dal dataset originale e successivamente registrato. A sinistra: immagine pan SPOT del 10/02/2000, 606-315. A destra uno zoom sull'area selezionata: immagine PCA1 (Componente Principale n1).

4.2.2.2 Creazione di una maschera binaria

Successivamente abbiamo creato una maschera binaria (Fig.4-10) per isolare le zone non interessate dalle dinamiche di deforestazione, come i bacini idrici e le aree paludose, ed i pixels che non apportano informazioni utili o rappresentano un

disturbo nell'analisi dei LUCC, come le nuvole e la loro ombra [Wilson et al., 2002]. A questa prima maschera abbiamo aggiunto poi un poligono che delimita il sito archeologico della Joyanca, area protetta da guardie armate ed in cui la *selva alta* rimane preservata. Infine un'ulteriore zona a sud-est della regione è stata esclusa, poiché interessata da processi di deforestazione che rappresentano l'arrivo di un nuovo fronte pioniere¹¹ dal sud, caratterizzato da dinamiche strutturali ed agenti differenti da quelli considerati nel nostro studio.

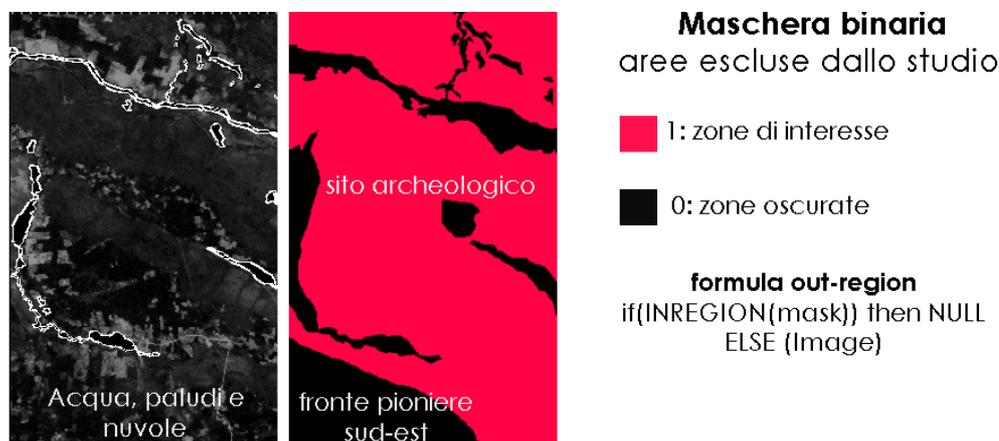


Fig 4-10: Costruzione della maschera binaria in ER Mapper. A Sinistra: i bacini idrici e le nuvole sono vettorizzati (linea bianca). A questa prima maschera aggiungiamo un poligono involvente il sito archeologico protetto ed il fronte pioniere sud-est. A destra: maschera finale. La maschera è applicata alle immagini satellitari originali attraverso un algoritmo di fusione che ho chiamato Out-region.

Una maschera è dunque un gruppo di poligoni irregolari che hanno lo scopo di isolare le zone prive di importanza nell'economia della ricerca o che possono causare confusione durante la classificazione delle immagini e nell'analisi dei cambi [Hayes et al., 2003]. L'operazione è stata realizzata utilizzando ER Mapper [Earth Resource Mapping, 1998]: i poligoni sono stati disegnati utilizzando un digitalizzatore e

¹¹ Processo di appropriazione spaziale di territori vergini. Si manifesta principalmente attraverso la deforestazione e l'incendio del bosco o delle savane [Hypergeo].

successivamente applicati alle scene originali. L'algoritmo di fusione si basa sulla formula *Out-Region* (2):

if(INREGION(mask)) then NULL, ELSE (Image) (2)

che porta il valore digitale¹² dei pixels all'interno dei poligoni a zero (ND=0).

4.2.2.3 Miglioramento del contrasto

Il miglioramento del contrasto cerca di adattare la risoluzione radiometrica dell'immagine alla capacità di visualizzazione del monitor. Il ND di un pixel, infatti, può non corrispondere al livello visuale NV, ossia alla intensità di grigio o di colore con cui si visualizza il pixel sullo schermo; dobbiamo dunque modificare parzialmente entrambi i parametri. *"A tecnica de realce de contraste tem por objetivo melhorar a qualidade das imagens sob os critérios subjetivos do olho humano...O contraste entre dois objetos pode ser definido como a razão entre os seus níveis de cinza médios. A manipulação do contraste consiste numa transferencia radiometrica em cada pixel, com o objetivo de aumentar a discriminação visual entre os objetos na imagem. Realiza-se a operação ponto a ponto, independentemente da vizinhança"*¹³ [INPE, 2004, Aula4, Pag.1].

Generalmente le tecniche di miglioramento del contrasto sono applicate a due casi opposti: 1) l'ampiezza dei ND è minore di quella degli NV; 2) il numero di ND della immagine

12 L'unità minima in cui si divide il territorio osservato rappresenta anche l'unità minima di informazione nell'immagine, detta pixel. Il tono di grigio (0-255) con cui ogni pixel appare sul monitor, si definisce attraverso un valore numerico che corrisponde alla codifica della radianza realizzata dal sensore al momento dell'acquisizione dell'immagine. Tale valore si chiama numero digitale ND. Ogni pixel registra differenti ND, uno per ogni banda spettrale. [Chuvieco, 2002; Lillesand et al., 1979].

13 La tecnica di miglioramento del contrasto ha l'obbiettivo di migliorare la qualità delle immagini, in base al criterio soggettivo dell'occhio umano. La manipolazione del contrasto modifica il valore radiometrico di ogni pixel, per aumentare la capacità di discriminazione visuale fra gli oggetti dell'immagine. L'operazione si realizza pixel a pixel, senza considerare il vicinaggio.

supera il numero degli NV possibili. Nel primo caso dovremmo applicare un'espansione del contrasto originale, nel secondo una riduzione dello stesso [Chuvieco, 2002].

Un istogramma dell'immagine (Fig.4-11) descrive la distribuzione statistica dei toni di grigio, in termini di numero di pixels per ogni tono. Possiamo definire il contrasto digitale in funzione dei valori ND_{max} e ND_{min} dell'immagine. Esistono diverse misure di dispersione comunemente utilizzate in statistica:

$$C_1 = ND_{max} / ND_{min}$$

$$C_2 = ND_{max} - ND_{min} \tag{3}$$

$$C_3 = \sigma(ND)$$

che offrono anche tre definizioni possibili del contrasto digitale.

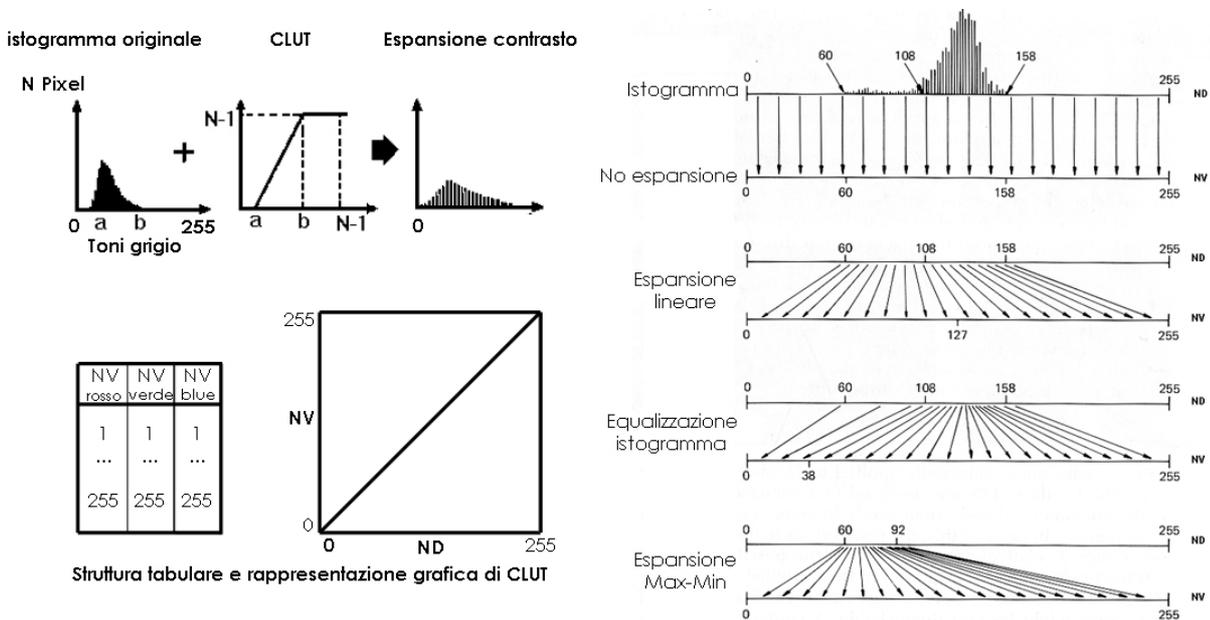


Fig 4-11: Il miglioramento del contrasto digitale permette di migliorare la capacità dell'occhio umano di distinguere gli elementi dell'immagine. Questa operazione prevede in generale una espansione dell'istogramma dei ND su tutta la scala byte 0-255. Questo passaggio fra ND e NV è realizzato mediante la CLUT. Noi abbiamo scelto un'espansione speciale che prevede la specificazione di un Max e un Min dei ND di interesse.

La CLUT¹⁴ è una matrice numerica (Fig.11) che indica il NV con cui si rappresenta sullo schermo ciascun ND dell'immagine. Essa funge da filtro fra i dati originali memorizzati sul disco e l'intensità di colore con cui li visualizzo. Grazie alla CLUT possiamo trasformare le relazioni fra ND e NV con il criterio che riteniamo più opportuno (trasformazione lineare, quadratica, equalizzazione istogramma, etc.) [Lillesand, 1979]. Se non applichiamo nessun miglioramento del contrasto la CLUT sarà rappresentata graficamente dalla diagonale che passa per (0,0) e (255,255); un'eventuale espansione o compressione dei ND modifica la pendenza della retta [Chuvieco, 2002]. Come regola generale possiamo dire che quanto maggiore sarà l'inclinazione della curva applicata all'istogramma, maggiore sarà il contrasto [INPE, 2004]. Bisogna comunque prestar attenzione a non creare saturazione (Fig.4-12); questo problema determina la compressione di un intervallo di toni di grigio (esterno al punto di intersezione) in un unico valore, perdendo dunque delle informazioni utili. La saturazione si realizza quando la pendenza della retta di CLUT è eccessiva.

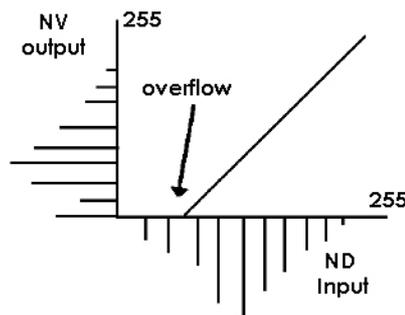


Fig 4-12: Problema di saturazione (overflow) nel miglioramento del contrasto digitale. Questo problema causa la perdita di informazioni dettagliate, poiché più valori di grigio vengono compressi in un'unica colonna dell'istogramma di uscita.

Tutte le immagini satellitari ed aeree acquisite sono state sottoposte al miglioramento del contrasto attraverso una trasformazione lineare o speciale (in cui specifico

14 CLUT: Color Look Up Table. Tabella dei colori.

nell'istogramma, i valori max e min dei toni di grigio delle classi studiate).

Queste operazioni sono state realizzate con l'aiuto della funzione *Imagem-Contraste* di SPRING 4.0 (INPE, São Paulo). L'espansione dei ND dell'immagine originale è stata effettuata manualmente su ogni banda monocromatica, per un totale di 41 layers. In alcuni casi la procedura ha notevolmente incrementato la capacità di distinguere i differenti elementi presenti nell'immagine (Fig.4-13).

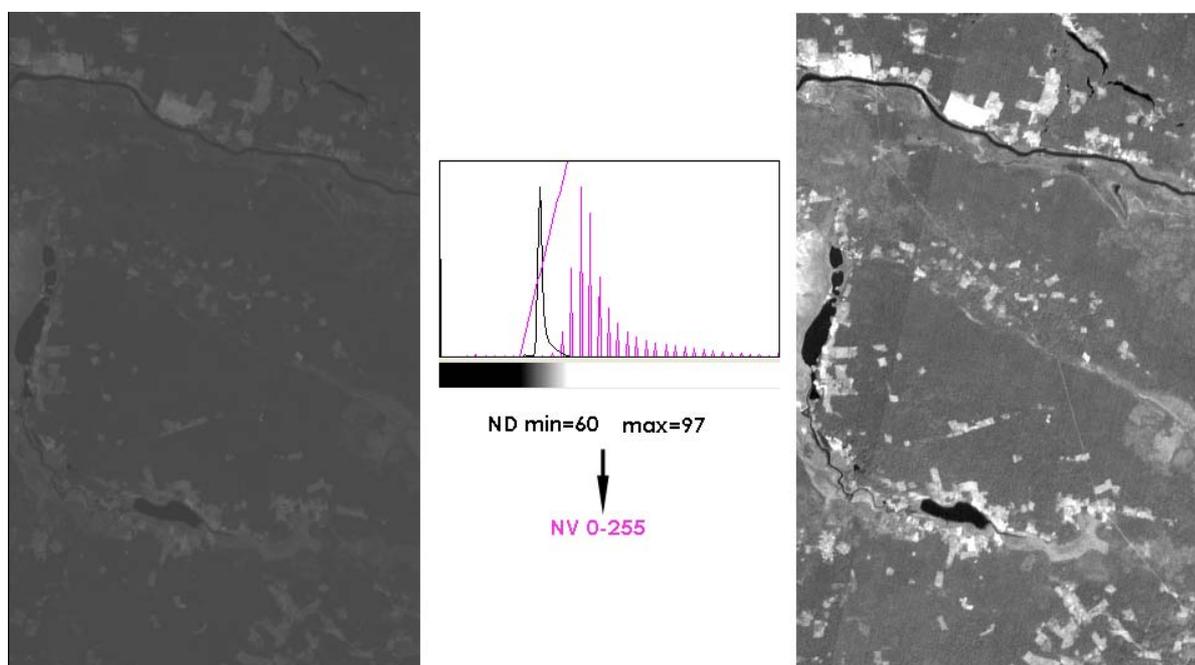


Fig 4-13: Espansione dei toni di grigio nell'istogramma dell'immagine originale. Operazione realizzata in SPRING 4.0. A sinistra: immagine XS1 (SPOT2, 606-315, 24/02/98). A destra: immagine contrastata.

4.2.2.4 Correzione radiometrica

La consistenza radiometrica fra elementi rappresentati in una serie multitemporale di immagini satellitari, è difficile da mantenere per i cambiamenti che si possono registrare, durante il periodo esaminato, nelle caratteristiche del sensore, nelle condizioni atmosferiche, nell'angolo di insolazione e nell'angolo di veduta del sensore. Per correggere queste grandezze e garantire una maggiore omogeneità dei dati

si applica una correzione radiometrica sulla serie di immagini. Questa scelta non è sempre necessaria, soprattutto quando lavoriamo con un'unica data; molti autori preferiscono analizzare i dati originali, considerando che ogni ulteriore trasformazione può determinare una parziale perdita di informazioni [Chica-Olmo et al., 2000].

*"When and how to correct the atmospheric effects depend on the remote sensing and atmospheric data available, the information desired and the analytical methods to extract the information"*¹⁵ [Song et al., 2001, Pag.230].

Il termine correzione radiometrica designa tutte quelle tecniche che modificano i ND originali per avvicinarli al valore che essi avrebbero nell'immagine, in caso di ricezione ideale. In questa categoria troviamo dunque le operazioni di ripristino di pixels e linee perse, la correzione dello "striping", il calcolo della energia riflessa da un oggetto al suolo, ed altre ancora.

Esistono due tipi di correzione radiometrica comunemente utilizzate per normalizzare una serie multitemporale di immagini satellitari: la correzione assoluta e la correzione relativa [Chen et al., 2005]. La prima permette di ottenere il valore della riflettanza di un oggetto, misurato al suolo; l'operazione richiede molte informazioni di difficile accesso, specialmente per le serie storiche, come certe caratteristiche dei sensori registrate simultaneamente alle condizioni atmosferiche. Il primo passo prevede la trasformazione dei valori di ND dell'immagine nella radianza misurata dal satellite, mediante la relazione (4):

$$L_{sat} = DN * Gain + Offset \quad (4)$$

¹⁵ Quando e come correggere gli effetti atmosferici dipende dal telerilevamento effettuato, dai dati sulle condizioni atmosferiche disponibili, dal tipo di informazione che si desidera e dal metodo utilizzato per estrarla.

dove "Gain" e "Offset" sono parametri che dipendono dalla calibrazione del sensore.

Successivamente questo valore della radianza è trasformato nella riflettanza misurata al suolo, mediante la relazione (5):

$$\varphi_{surface} = \frac{(L_{sat} - L_{path})\pi}{E\tau} \quad (5)$$

dove " L_{path} " rappresenta la radianza catturata dal sensore come risultato della riflessione delle particelle e molecole nell'atmosfera [Kaufman, 1993], " E " è l'irradianza¹⁶ sulla parcella al suolo e " τ " è la trasmissione dell'atmosfera [Lillesand et al., 1979]. Per il calcolo della riflettanza al suolo la (5) si basa su misure *in situ* di tutti i parametri citati; la difficoltà di ottenere informazioni affidabili rende questo metodo poco pratico.

La correzione relativa cerca di adattare le caratteristiche radiometriche delle immagini satellitari ad una scena o a dei valori, scelti come sistema di riferimento. "*Relative radiometric normalization is an image based correction method achieved by setting the multitemporale images into a common scale without extra parameters from other measurements*"¹⁷ [Chen et al., 2005, Pag.64]. A questa categoria appartengono [Hong et al., 2008] il metodo *Dark Objective Subtraction* (DOS) [Chuvieco, 2002], il *Pseudoinvariant Feature methods* (PIF) [Schott et al., 1988] e il *Ridge method* [Song et al., 2001], fra gli altri.

Per la correzione delle immagini dal 1988 al 2003 del Peten, abbiamo scelto il metodo DOS, in funzione dei dati disponibili e visti buoni risultati offerti da questa tecnica nell'analisi multitemporale di cambi [Cheng et al., 2005]. Il

16 Irradianza: flusso radiante incidente su una superficie per unità di area [Watt/m²].

17 La normalizzazione radiometrica relativa è un metodo basato sulla correzione dell'immagine ottenuta portando tutta la serie multitemporale di scene su una scala comune, senza bisogno di parametri supplementari da altre misure.

punto di partenza di questo metodo consiste nell'assumere che le aree con forte assorbenza (acqua ed ombra) dovrebbero presentare una radianza spettrale molto prossima allo 0. In realtà l'istogramma di un'immagine sempre presenta un ND_{min} superiore a questo valore, effetto dovuto alla dispersione atmosferica. Un'approssimazione semplicistica del metodo può essere rappresentata dalla traslazione dell'istogramma:

$$ND_{i(pixel),k(banda)}^c = ND_{i(pixel),k(banda)} - ND_{min,k(banda)} \quad (6)$$

dove " $ND_{min,k}$ " rappresenta il valore minimo registrato nella banda "k". La (6) permette dunque di sottrarre dal valore originale ND del i-esimo pixel nella k-esima banda, il valore minimo " $ND_{min,k}$ ", situando l'origine dell'istogramma a zero. La scelta degli "oggetti neri" di riferimento (ND_{min}) per la sottrazione richiede un po' di attenzione; precedenti ricerche consigliano di utilizzare corpi idrici profondi e costanti, con acqua limpida, oppure zone scure di vegetazione coperte dall'ombra delle nuvole. Noi abbiamo optato per i due grandi laghi Laguna Tuspan e Laguna Agua Dulce, il cui livello si mantiene costante nel periodo studiato. Fra le informazioni utilizzate nella simulazione dal 1998 al 2003, solamente l'immagine SPOT2 del 24/02/98 ha richiesto una normalizzazione radiometrica DOS (Fig.4-14); le scene ETM del 2000 e 2003 erano già state precedentemente corrette.

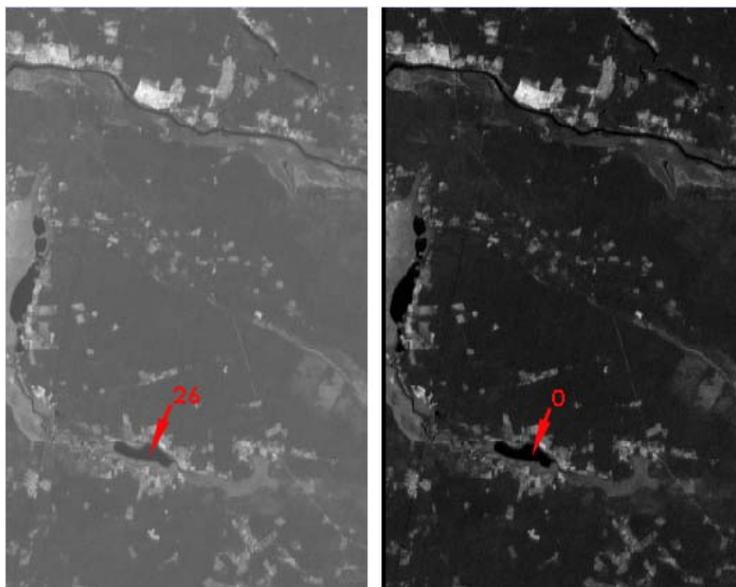


Fig 4-14: Correzione radiometrica DOS. La Laguna Tuspan e la Laguna Agua Dulce sono state prese come riferimento per la normalizzazione dei ND. Immagine XS2 (SPOT2, 606-315, 24/02/1998).

Per realizzare la correzione atmosferica abbiamo utilizzato la funzione *ATMOSC* di Idrisi Andes [Eastman, 2006]. I parametri richiesti sono stati messi a disposizione dalla SPOT IMAGE, come allegato (.PDF) alle immagini acquistate (Fig.4-15).

Work Order	: 0131432-6 (72716101)			
Volume	: CDROM 2040001365			
Scene ID	: 4 608-315 01-03-31 16:56:58 2 I			
Product Code	: SCSG1C00N			
Date	: 2003-01-16 10:27:34			

Scene Parameters

Scene ID	4 608-315 01-03-31 16:56:58 2 I			
K-J identification	608-315			
Date	2001-03-31 16:57:01			
Instrument	HRVIR 2			
Shift Along Track	4 1200			
Preprocessing level	2A			
Spectral mode	XI			
Number of spectral bands	4			
Spectral band indicator	XS1	XS2	XS3	XS4
Gain number	5	5	3	3
Absolute calibration gains (W/m²/sr/μm)	2.11525	2.68607	1.29300	8.19106
Orientation angle	10.3 degree			
Incidence angle	L29.6 degree			
Sun angles (degree)	Azimut: 127.2 Elevation: 68.5			
Number of lines	3662			
Number of pixels per line	4580			

Scene Center Location

Latitude	N017° 18' 40"
Longitude	W090° 08' 10"
Pixels number	2248
Line number	1823

Corners Location

Corner	Latitude	Longitude	Pixel n°	Line n°
1	N017° 38'09"	W090° 25'04"	780	1
2	N017° 30'37"	W089° 42'02"	4580	756
3	N017° 06'30"	W090° 33'17"	1	2908
4	N016° 59'01"	W089° 50'21"	3800	3662

Files Parameters

	Record Length	Number of Records	Offset
VOLD_01.DAT	360	5	0
LEAD_01.DAT	3960	27	0
IMAG_01.DAT	5400	14649	0
TRAI_01.DAT	1080	3	0
NULL_01.DAT	360	1	0

Fig 4-15: Esempio dei parametri di un'immagine SPOT; una parte di queste informazioni sono necessarie per la correzione radiometrica realizzata.

4.2.2.5 Selezione delle bande più adeguate per le composizioni di colore

A partire dai dati multi-spettrali raccolti possiamo generare differenti composizioni di colore (CC), applicando ciascuno dei tre colori primari (rosso, verde, blue) ad una banda distinta dell'immagine, secondo il criterio e l'ordine

che riteniamo più opportuno. Questo processo permette di visualizzare simultaneamente informazioni di distinte regioni dello spettro, facilitando la delimitazione visuale di alcune classi di copertura. La selezione delle bande ed il loro ordine di applicazione nei canali RGB, dipende dal tipo di sensore e dalla finalità dello studio: per la caratterizzazione geologica si fa spesso affidamento alle composizioni basate sulle bande SWIR, MIR, TIR [Follador, 2005], mentre la composizione NIR,R,G facilita la cartografia di masse vegetali, bacini idrici e città [Follador, 2006; Chuvieco, 2002].

In letteratura sono stati proposti differenti indici statistici e cromatici per offrire un criterio obbiettivo di selezione delle 3 bande più adeguate per un'immagine RGB, all'interno del database multi-spettrale. *"Besides trial-and-error visual analysis, is there no way quantitatively to identify which three bands should produce the optimum color composite?"*¹⁸ [Jensen, 1986, Pag.90].

[Chavez et al., 1984] hanno sviluppato l'*Optimum Index Factor* (OIF) per analizzare tutte le 20 possibili combinazioni che si possono ottenere con 6 bande TM Landsat, concludendo che generalmente, maggiore è l'OIF, maggiore è l'apporto di informazioni. Il metodo è applicabile ad ogni gruppo di dati multispettrali, raccolti con altri sensori.

L'algoritmo utilizzato per il calcolo del OIF, per ogni sottogruppo di 3 bande è:

$$OIF = \frac{\sum_{k=1}^3 s_k}{\sum_{j=1}^3 abs(r_j)} \quad (7)$$

18 Oltre alla semplice analisi visuale, esiste una maniera di stabilire quantitativamente quali 3 bande produrranno la miglior composizione RGB?

dove "s_k" è la deviazione standard per k-esima banda, "r_j" il valore assoluto del coefficiente di correlazione fra le bande prese in esame.

L'OIF è stato ampiamente utilizzato per studi geologici e geomorfologici [Follador, 2005 & 2004; Al-juaidi et al., 2003; Mirzavand G., 2003, Paradella et al., 2001], per la fusione di dati multispettrali [Genderen et al., 1998] e per valutazioni di impatto ambientale [Haack et al., 1995].

I risultati di questa tecnica mostrano un comportamento simile; il valore dell'OIF, calcolato sui database TM Landsat (Tab.4-2), privilegia la combinazione di una banda del visibile (B1, B2 e B3), una dell'infrarosso prossimo (B4) ed una dell'infrarosso medio (B5 e B7), identificando così i tre assi principali di variazione in una immagine TM.

Bande	Tot DevSt	Tot Corr.	OIF
145	116.26	1.08	107.6481
245	113.79	0.94	121.0532
345	120.43	1.12	107.5268
157	124.37	2.43	51.1811
147	122.35	0.99	123.5859
457	120.26	1.16	103.6724
135	124.54	2.41	51.6763
125	117.9	2.3	51.2609
134	122.52	1.23	99.6098
257	121.9	2.48	49.1532
235	122.07	2.47	49.4211
357	128.54	2.66	48.3233
124	115.88	1.22	94.9836
247	119.88	0.82	146.1951
347	126.52	1	126.5200
234	120.05	1.07	112.1963
137	130.63	2.6	50.2423
127	123.99	2.49	49.7952
237	128.16	2.63	48.7300
123	124.16	2.76	44.9855

**Calcolo del OIF fra 6 bande TM
(Landsat-ETM, 27/03/00, P20-R48)**

r	B1	B2	B3	B4	B5	B7	DevSt.TM
B1	1	0.9	0.92	-0.2	0.68	0.79	40.82
B2	0.9	1	0.94	-0.02	0.72	0.8	38.35
B3	0.92	0.94	1	-0.11	0.81	0.89	44.99
B4	-0.2	-0.02	-0.11	1	0.2	0	36.71
B5	0.68	0.72	0.81	0.2	1	0.96	38.73
B7	0.79	0.8	0.89	0	0.96	1	44.82

matrice di correlazione (r)

(S)

Tab 4-2: Il calcolo dell'OIF per la serie multispettrale Landsat-ETM del 27/03/00, esclusa la banda termica, ha evidenziato come la CC TM(2-4-7), sia la combinazione ottimale.

Per tutti i database multispettrali inoltre, la combinazione in colore meno adeguata (ridondanza dell'informazione) è risultata la CC(3-2-1), anche detta True

color composite, poiché essa rappresenta l'immagine percepita dall'occhio umano. Il calcolo dell'OIF per le serie TM, ha portato alla selezione delle seguenti composizioni: CC(1-4-5) per il 1988, CC(1-4-7) per il 1992, CC(2-4-7) per il 2000 (Tab.2) ed infine CC(2-4-5) per il 2003.

La scelta delle CC ottimali permette di evidenziare le differenti classi di uso del suolo, diminuendo il volume di informazioni trattate. In particolare il contrasto fra la banda NIR (vigore vegetale) e la banda SWIR-MIR¹⁹ (suolo esposto) sottolinea il livello di degrado della foresta, l'espansione delle zone prive di copertura vegetale e le aree che invece presentano una progressiva riforestazione (Fig.4-16).

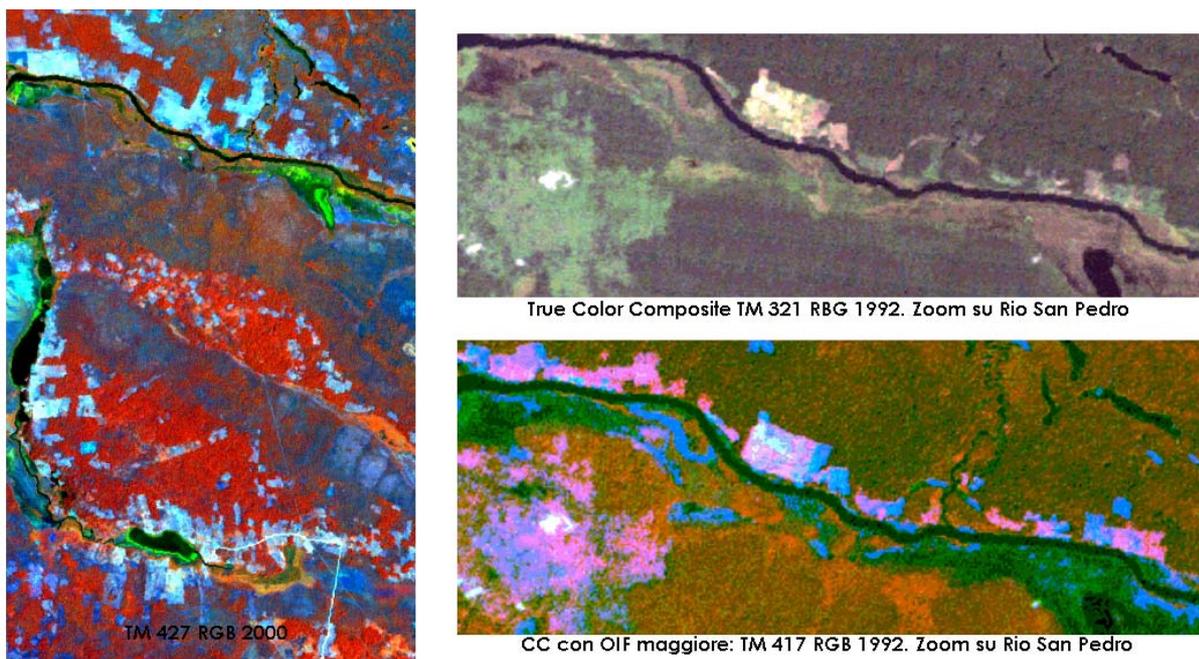


Fig 4-16: A sinistra: CC TM 247 ottimale con un OIF=146,19 calcolato dai dati multispettrali del 27/03/00. A destra: zoom su CC TM 321 e TM 417. Confronto fra immagine reale (OIF= 3,38) e composizione in colore ottimale (OIF= 21,74). Dati multispettrali del 24/11/92.

19 NIR: Near Infrared, infrarosso prossimo [0,7-1,3 μ m]. SWIR: Short Wave Infrared, infrarosso medio prossimo [1,3-2,5 μ m]. MIR: Medium Infrared, infrarosso medio [2,5-3,7 μ m]. VIS: visibile, Blue[0,4-0,5 μ m], Verde[0,5-0,6 μ m] e Rosso [0,6-0,7 μ m].

4.2.2.6 La dimensione del pixel

*"The choice of grid resolution was related with the cartographic and statistical concepts: scale, computer processing power, positional accuracy, size of delineations, inspection density, spatial autocorrelation structure and complexity of terrain...it was concluded that no ideal grid resolution exist, but rather a range of suitable resolutions"*²⁰
[Hengl, 2006, Pag.1283].

La risoluzione spaziale definisce l'oggetto più piccolo che può essere evidenziato in un'immagine; questa dimensione corrisponde alla minima unità informativa di un raster, detta pixel. Le principali proprietà di quest'ultimo, come la dimensione, la forma e la localizzazione, sono definite principalmente dai sensori in commercio e non direttamente dalle proprietà del suolo.

Spesso nel processamento dei dati satellitari si dà per scontato che ogni pixel rappresenti una sola classe di copertura terrestre (Fig.4-17); in realtà questa situazione non è molto frequente e può condurre ad evidenti errori di classificazione dell'immagine [Foody, 2006].

La quantità di pixels misti in un'immagine telerilevata, è spesso considerevole e dipende sia dalle caratteristiche del sensore (risoluzione spaziale) sia dal mosaico di classi di copertura al suolo, e dunque dal contesto geografico. Per questo, molte volte, dei metodi di classificazione che offrono eccellenti risultati in un certo ambiente (condizioni ottimali: piccole aree con grandi parcelle omogenee e una variabilità topografica insignificante), possono fallire se applicati in altre regioni.

²⁰ La scelta della risoluzione del grid è influenzata da concetti cartografici e statistici: scala, capacità di elaborazione del computer, precisione nella localizzazione, dimensione delle delineazioni, densità, autocorrelazione spaziale e complessità del paesaggio...concludiamo che non esiste una risoluzione ideale, ma una serie di risoluzioni adeguate ai nostri obiettivi.

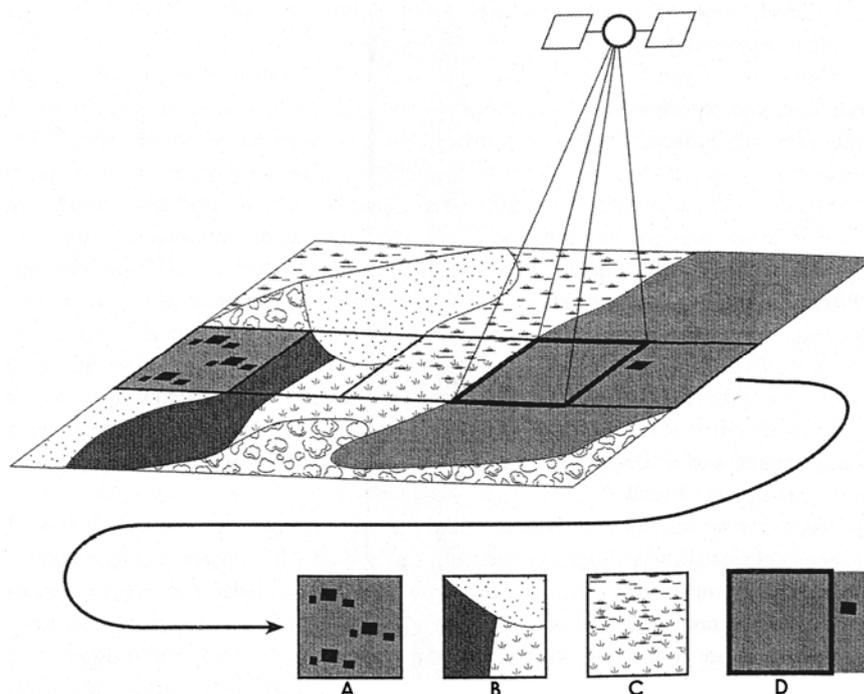


Fig 4-17: Cause principali dei pixels misti. A) Il pixel cattura un'area di foresta abitata da un piccolo villaggio. B) Il pixel cattura un'area con 3 classi tematiche discrete, divise da confini nitidi. C) Si tratta di un'area di transizione fra due classi di uso del suolo, ad esempio fra milpa abbandonata e foresta. D) Il pixel cattura una parcella omogenea, il cui valore spettrale risulta comunque influenzato dalla presenza di oggetti (casa) molto vicini. [Immagine da Foody, 2006].

Generalmente, aumentando la risoluzione spaziale, il segnale raccolto dal sensore risulta essere più vicino alle reali condizioni della copertura [Chuvieco, 2002]. La porzione di pixels misti aumenta dunque quando la risoluzione spaziale diventa più grossolana (Fig.4-18).

Adottare una risoluzione più fine comporta parallelamente un aumento delle righe e delle colonne che compongono l'immagine, e dunque un aumento del volume dei dati. Nel nostro lavoro abbiamo sperimentato differenti alternative di pixel, considerando sia la frammentarietà del paesaggio, sia la capacità di elaborazione del nostro computer.

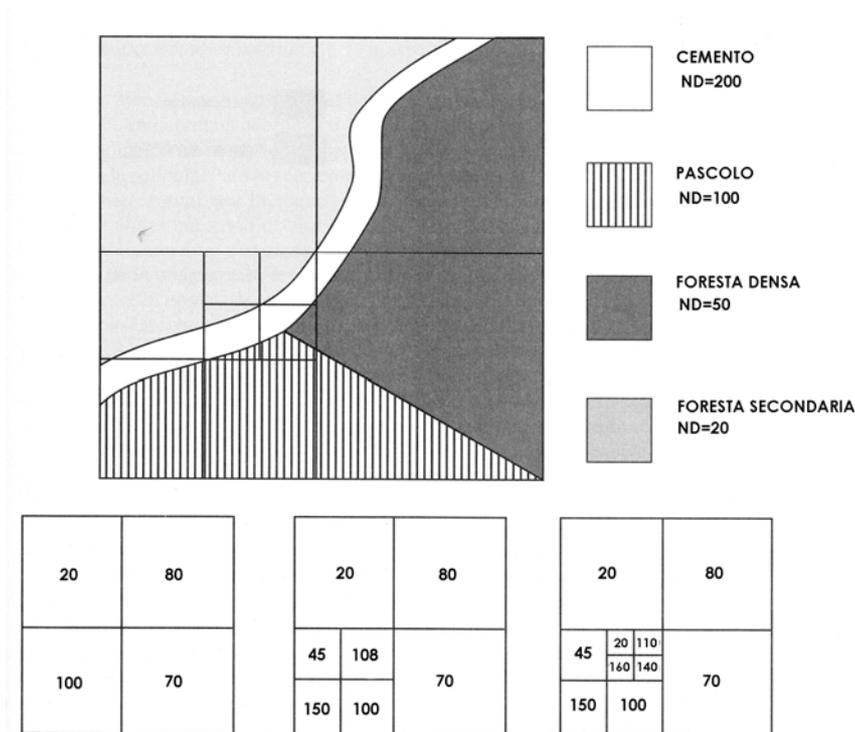


Fig 4-18: Relazione fra dimensione del pixel e capacità di catturare la frammentazione della copertura terrestre. [Immagine corretta da Chuvieco, 2002].

La scelta finale è stata di adottare una risoluzione di 20m, che bene si adatta alle caratteristiche dei nostri modelli predittivi e agli obiettivi prefissati. Molti studi su deforestazione tropicale si affidano in genere a supporti maggiori, adeguati a descrivere le dinamiche di uso del suolo a macro-scala ma inadatti ad evidenziare processi e fattori a livello locale. Nel nostro caso, una risoluzione troppo dettagliata (5m, 10m, 15m), oltre a creare un inutile aumento di volume dell'informazione, provoca un'allocazione di memoria eccessiva sul server in fase di simulazione con PNNET. Abbiamo cercato di risolvere il problema dividendo le carte originali in rettangoli di uguale dimensione, utilizzati separatamente in simulazione ed infine riassemblati. Gli effetti frontiera, lungo i margini di ogni singola parte, rendono però i risultati poco realistici. La scelta di aumentare la risoluzione a 20m è stata dunque fortemente influenzata dal comportamento e dai

risultati dei modelli predittivi, come suggerito da [Hengl, 2006].

4.2.3 Individuazione dei cambi di uso del suolo attraverso lo studio multitemporale di immagini satellitari

4.2.3.1 Tecniche di analisi multitemporale

*"All digital change detection is affected by spatial, spectral, temporal and thematic constraints. The type of method implemented can profoundly affect the qualitative and quantitative estimates of the disturbance. Even in the same environment, different approaches may yield different change maps."*²¹ [Coppin et al., 1996, Pag.214].

Le risorse naturali, fra queste anche le foreste tropicali, sono continuamente soggette a dei cambiamenti, intendendo per cambiamento un'alterazione nell'identità delle componenti della copertura terrestre [Pelekis et al., 2004] oppure un loro movimento spaziale-spettrale nel tempo. La variazione delle caratteristiche spettrali, facilmente rilevabile durante il processamento delle immagini, è dovuta alle modificazioni del segnale riflesso o emesso da una parcella, in seguito ad un cambio di copertura oppure all'alterazione della densità o dello stato fenologico. I cambi spaziali implicano trasformazioni della forma e della dimensione delle parcelle, che possono mantenere la stessa copertura [Chuvieco, 2002].

I metodi di analisi multitemporale sono caratterizzati da due differenti tipi di errore: l'errore di omissione e l'errore di commissione. Il primo si riferisce ad un cambio che è stato ignorato o parzialmente evidenziato (carenza di informazione); il secondo, invece, è il risultato di un cambio rilevato ma non

²¹ L'analisi digitale dei cambi è influenzata dalle caratteristiche spaziali, spettrali, temporali e tematiche in esame. Il metodo utilizzato può profondamente modificare la stima qualitativa e quantitativa dei cambi. Anche se riferiti allo stesso ambiente, due approcci differenti daranno origine a differenti carte di cambio.

realmente avvenuto (falso allarme). La scelta del metodo più adatto al contesto geografico e storico in esame, minimizza uno o l'altro errore in funzione degli obiettivi e dei dati.

[Coppin et al., 1996] presentano una revisione dei principali metodi di analisi dei cambi di uso del suolo, concludendo che la differenza di immagini (in cui il valore 0 indica persistenza ed i valori $\neq 0$ un cambio), e la trasformazione lineare multitemporale dei dati (basata su PCA e Tasseled Cup²²), sono gli approcci che offrono maggior precisione nello studio dei LUCC in aree forestali. L'analisi basata sulle Componenti Principali può comunque presentare dei limiti legati al fatto che, spesso, i valori di intensità che descrivono i cambi in esame, rappresentano una piccola frazione dell'informazione satellitare e possono essere registrati nelle componenti minori (PC4, PC5, etc.), mescolati con il disturbo nei dati [Hame, 1991]. Queste ultime componenti sono in genere trascurate durante la classificazione e dunque l'informazione è persa [Sabins, 1987].

Una descrizione più dettagliata di questi metodi è presentata in [Chuvieco, 2002]; l'autore divide le tecniche di analisi multitemporale in due gruppi, a seconda che esse utilizzino immagini continue o categorizzare. Nel primo caso si impiegano delle tecniche quantitative: differenze, regressioni, PCA, etc., mentre nel secondo si confrontano immagini precedentemente classificate (Delta classification); la precisione di quest'ultimo approccio è totalmente dipendente dall'accuratezza della classificazione iniziale.

[Sader et al., 2003; Wilson et al., 2002 ; Hayes et al., 2001] sottolineano come il metodo RGB-NDVI, basato sulla visualizzazione delle mappe degli indici vegetali nei tre

²² Per una descrizione dettagliata dell'Analisi delle Componenti Principali (PCA) e della tasseled cup, si rimanda a [Tso et al., 2001].

canali di colore (Red, Green, Blue), sia la tecnica più accurata ed efficace per l'analisi di una serie temporale di immagini Landsat TM e l'individuazione dei cambi in area forestale. Gli autori evidenziano come questa tecnica permetta di trattare contemporaneamente 3 date successive, e dunque risulti più vantaggiosa con datasets di grandi dimensioni.

Per queste ragioni e per la sua semplicità rappresentativa abbiamo scelto il metodo RGB-NDVI e RGB-NDII (vedi prossimi capitoli), per l'analisi e la visualizzazione dei cambi registrati nella regione della Joyanca dal 1998 al 2003 [Follador et al., 2006].

4.2.3.2 Gli indici vegetali e di stress idrico

L'utilizzo del rapporto (*ratio*) fra i pixels delle diverse bande per discriminare masse vegetali ed aree deforestate, deriva dal peculiare comportamento radiometrico della vegetazione. Numerose pubblicazioni in letteratura [Bruno et al., 2006; Chen et al., 2005; Turner et al., 2004; Sader et al., 2003; Chuvieco, 2002] evidenziano come la firma spettrale (Fig.4-19) di una pianta sana mostri un chiaro contrasto fra la regione del VIS (in particolare nella banda Rossa [0,6-0,7 μm]) e la regione del NIR ([0,7-1,3 μm]).

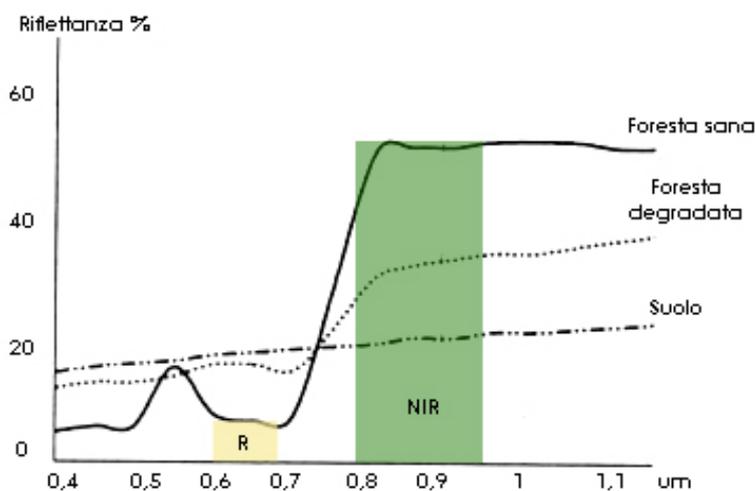


Fig 4-19: Contrasto spettrale della vegetazione sana fra la banda Rosso e la banda infrarosso prossimo NIR.

Questo contrasto spettrale permette di separare con relativa facilità la vegetazione sana dalle altre classi di copertura. Quando la foresta è degradata o stressata (frammentazione paesaggio, siccità, ad es.) la differenza fra la banda R e la banda NIR diminuisce (Fig.18). Diciamo dunque che quanto maggiore sarà la differenza fra la riflettanza nel NIR e nel R, maggiore sarà il vigore vegetale della regione osservata e la sua biomassa [Chowdhury et al., 2004]. Analogamente si può evidenziare come un suolo esposto abbia praticamente gli stessi valori nel NIR e R, con un contrasto molto basso o tendente allo zero.

Basandosi su questo principio, sono stati elaborati numerosi indici di vegetazione, che indichiamo con NDVI²³ e che sono generalmente espressi, per ogni i-esimo pixel, da:

$$NDVI_i = \frac{\varphi_i(NIR) - \varphi_i(R)}{\varphi_i(NIR) + \varphi_i(R)} \quad (8)$$

dove " $\varphi_i(NIR)$ " e " $\varphi_i(R)$ " indicano rispettivamente la riflettanza del i-esimo pixel nella banda dell'Infrarosso Prossimo e nella banda del Rosso. Questi intervalli spettrali corrispondono alla banda B4 (NIR) e B3 (R), per i sensori Landsat TM, ed alle bande XS2 (R) e XS3 (NIR) per lo SPOT.

La (8) è definita in termini di riflettanza al suolo, richiedendo dunque tutte le correzioni radiometriche assolute viste precedentemente. Noi utilizzeremo le informazioni registrate nei ND delle immagini; il valore dei NDVI, anche se calcolato relativamente alle condizioni atmosferiche ed al sensore adoperato, ci permette di pervenire alle stesse conclusioni.

[Bruno et al., 2006] riprendono la formula sviluppata da [Guyot et al., 1994] per correggere radiometricamente gli

23 Normalized Difference Vegetation Index.

errori sistematici che dipendono dal sensore considerato e gli effetti atmosferici, evidenziando come sia necessaria una conversione fra riflettanza apparente e ND dell'immagine. Le formule sono state normalizzate (scala Byte) e adattate alla presenza della maschera binaria, precedentemente applicata ai dati Landsat (9) e SPOT (10):

$$NDVI_{TM+Mask} = \frac{(ND_{NIR} - 0.801 * ND_R)}{(ND_{NIR} + 0.801 * ND_R) + 1} + 127 \quad (9)$$

$$NDVI_{Spot+Mask} = \frac{(ND_{NIR} - 0.663 * ND_R)}{(ND_{NIR} + 0.663 * ND_R) + 1} + 127 \quad (10)$$

L'analisi visuale delle mappe multitemporali NDVI, permette di seguire chiaramente l'evoluzione delle dinamiche di deforestazione ed uso del suolo nella Joyanca (Fig.4-20). Per rendere più evidente il contrasto fra le zone degradate e la foresta densa, mostriamo in seguito le immagini non mascherate del 1988 e del 2003, in cui risalta il vigore vegetale della selva alta preservata attorno al sito archeologico.

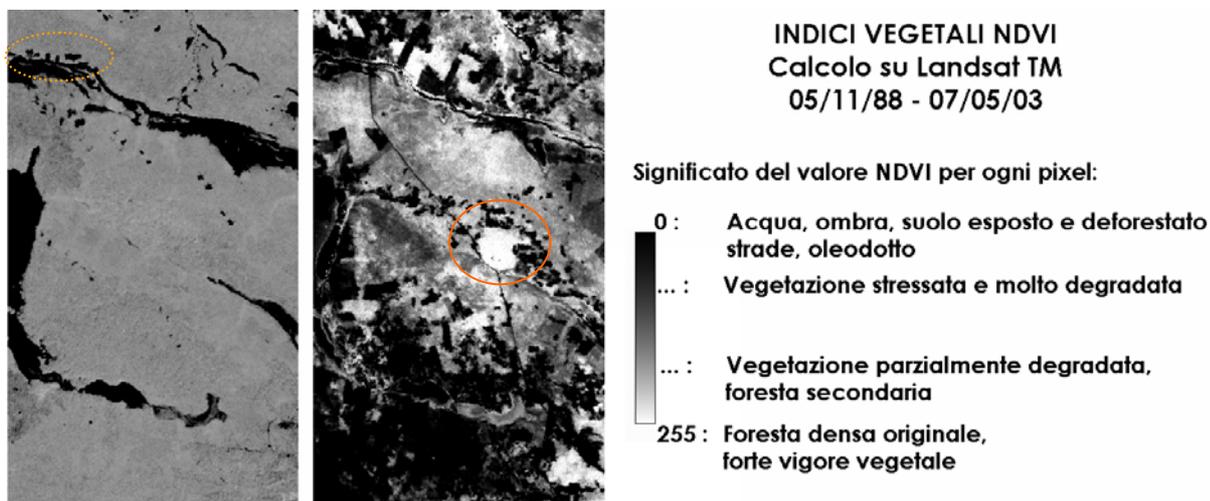


Fig 4-20: Mappe degli indici vegetali NDVI calcolate sul database multispettrale Landsat del 1988 e 2003. A sinistra: TM 05/11/88, primi insediamenti al nord del Rio San Pedro, evidenziati in rosso. In questa immagine possiamo notare una uniformità del manto vegetale, denso e in buona salute (grigio chiaro). Le zone nere rappresentano i bacini idrici, alcune ombre di nuvole ed i primi poligoni di deforestazione (forme geometriche rettangolari evidenziate). A destra: ETM 07/05/03, la foresta densa originale è stata fortemente degradata e sostituita da attività agricole e pastorizie. La selva alta attorno al sito archeologico, area preservata, mostra un forte vigore vegetale (poligono di colore bianco, evidenziato), che contrasta con il nero delle parcelle deforestate. Le zone di bajo centrali presentano toni che vanno dal grigio scuro al grigio chiaro, indicativi di una fase di transizione e ricrescita dopo i grandi incendi del 1998 e 2000.

La carta NDVI del 1988 mostra una foresta uniforme, densa e preservata, rappresentata da una tonalità omogenea di grigio chiaro (ND elevato); le grandi aree nere rappresentano i bacini idrici e le paludi. Le piccole parcelle nere (ND basso) a nord del Rio San Pedro presentano forme geometriche quasi regolari ed indicano i primi insediamenti umani nella regione; le restanti zone scure sparpagliate nell'immagine hanno forme irregolari e rappresentano le ombre delle nuvole sulla vegetazione. La carta del 2003 evidenzia una situazione totalmente differente ed un paesaggio molto frammentato. Le numerose parcelle regolari (e non) di colore nero, indicano le aree di suolo esposto, deforestato o incendiato, nonché i soliti bacini idrici. Le poche zone di foresta densa originale, con forte vigore vegetale, sono caratterizzate da un colore quasi bianco che contrasta con i toni scuri dell'immagine. Le aree di *bajo* mostrano differenti graduazioni di grigio ed indicano una fase di transizione e ricrescita dopo i grandi incendi del 1998 e 2000.

Questi indici rappresentano un ottimo aiuto per discriminare la vegetazione ed evidenziare il suo stato di salute. Essi sono definiti a partire dal comportamento spettrale di una pianta nel NIR e R, e possono non essere idonei allo studio di altre classi di copertura. Nonostante la nostra ricerca sia focalizzata sullo studio delle dinamiche di deforestazione, abbiamo deciso di affiancare ai NDVI un altro indicatore per meglio evidenziare le zone erose ed incendiate. [Chuvieco, 2002] dimostra come il contenuto di acqua nella vegetazione e nel suolo sia ben rappresentato dal contrasto fra le bande SWIR e NIR. Quando l'umidità del suolo o della vegetazione aumenta, parallelamente il valore di riflettanza diminuisce nel SWIR. Come fatto in precedenza per il NDVI,

definiamo ora l'indice di stress idrico NDII²⁴ attraverso il contrasto fra le due bande dell'infrarosso prossimo e medio:

$$NDII_{Mask} = \frac{(ND_{NIR} - ND_{SWIR})}{(ND_{NIR} + ND_{SWIR}) + 1} + 127 \quad (11)$$

I valori " ND_{SWIR} " sono raccolti nella banda B5 di Landsat TM e nella banda SWIR di SPOT.

[Wilson et al., 2002] dimostrano come la classificazione basata su NDII, riesca ad individuare i cambi in una area forestale del Maine in maniera più accurata rispetto al tradizionale NDVI. [Chuvieco, 2002] evidenzia come le carte NDII siano un efficace strumento per cartografare zone incendiate, erose o soggette a siccità. Abbiamo dunque deciso di utilizzare l'informazione apportata da entrambi gli NDVI e NDII, creando un nuovo indice NDMI²⁵ come media dei valori precedentemente calcolati:

$$NDMI = \frac{ND_{NDVI} + ND_{NDII}}{2} \quad (12)$$

Questo indice sarà successivamente utilizzato per l'analisi multitemporale dei cambi avvenuti fra il 1998 ed il 2003, periodo scelto per la successiva calibrazione e validazione dei modelli predittivi.

4.2.3.3 Diagramma di dispersione dei ND

Per verificare il grado di informazione apportato da ogni banda, si fa spesso uso dei diagrammi di dispersione bivariati che localizzano la posizione spettrale dei pixel lungo le due bande scelte come assi. Quando la nuvola di punti, che rappresenta la descrizione multidimensionale dello spettro di

24 NDII: Normalized Difference Infrared Index. Questo indice è stato chiamato in differenti modi: indice di stress idrico, indice di umidità o NDII, a seconda degli autori.

25 NDMI: Normalized Difference Medium Index.

ogni singola categoria di uso del suolo [Jensen, 1986], forma una linea o un'elisse molto allungata, allora le bande sono molto correlate per cui è sufficiente considerarne una sola. Al contrario, se la nuvola di punti presenta un'alta dispersione, la correlazione fra le bande sarà piccola ed entrambi devono essere considerate poiché offrono un'informazione contrastata [Chuvieco, 2002]. La costruzione dei diagrammi di dispersione ha evidenziato ancora l'utilità delle bande R, NIR e SWIR (Fig.4-21) per l'individuazione e la classificazione delle famiglie di pixels che rappresentano le singole categorie di uso del suolo.

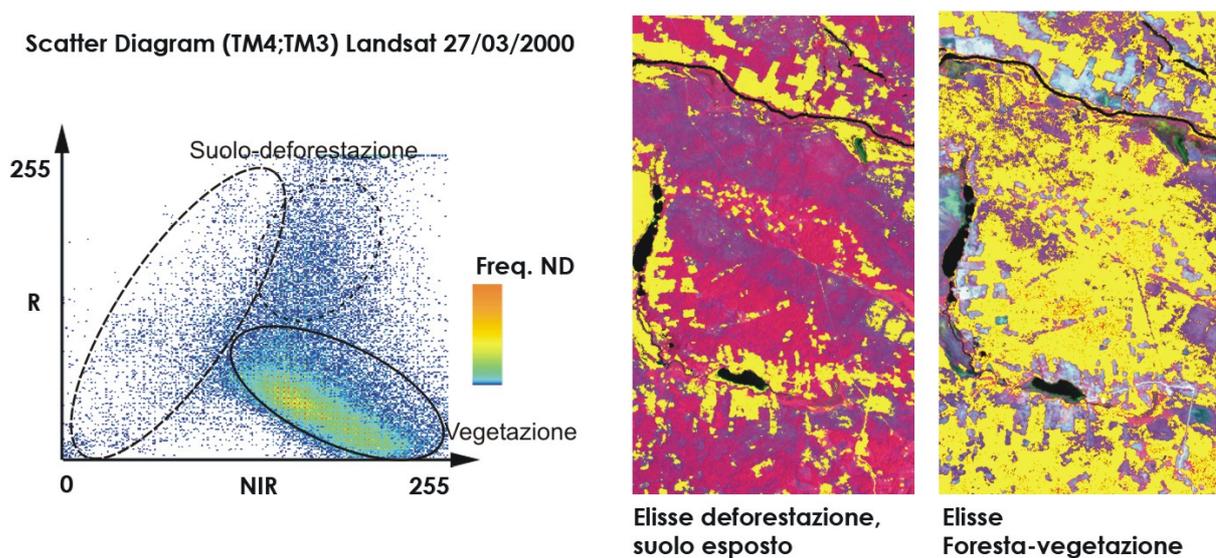


Fig 4-21: Nuvola di correlazione con bande NIR e R. L'analisi visuale delle famiglie di pixel, ha evidenziato ancora l'utilità delle bande R, NIR e SWIR per classificare e separare spettralmente le classi di copertura studiate.

Possiamo individuare facilmente le elissi o famiglie di pixels che classificano ogni categoria di uso del suolo; la verifica di questo *clustering* è stato realizzato attraverso la funzione *scattergram* di ER Mapper [Earth Resource Mapping, 1998]. I punti caratterizzati da un alto valore nel NIR e un basso ND nel R, si trovano all'interno dell'elisse che rappresenta la vegetazione; viceversa i punti che ricadono lungo la diagonale ($ND_{NIR}=ND_R$) o che hanno valori nel R maggiori

del NIR, identificano le elissi correlate alla deforestazione e suolo esposto. Analogo comportamento si può osservare nel diagramma che utilizza come assi SWIR e NIR.

La scelta dei NDVI e NDII, basati sul contrasto spettrale che registriamo fra R, NIR e SWIR, fra le differenti categorie di copertura, è dunque apparsa adatta al nostro studio multitemporale dei cambi di uso del suolo nella Joyanca.

4.2.3.4 Analisi multitemporale RGB(NDMI)

Per l'analisi multitemporale dei cambi avvenuti nella regione La Joyanca fra il 1998 ed il 2003, abbiamo dunque utilizzato il metodo basato sulla visualizzazione simultanea dei NDMI in tre date successive, attraverso i canali Rosso Verde e Blue (RGB) del monitor. Ogni pixel ha un colore che deriva dalla combinazione dei tre valori di NDMI nei rispettivi canali. Questa tonalità è interpretata con l'aiuto dell'analisi visuale delle immagini originali e ponderando il valore di ogni NDMI nella sequenza temporale. Abbiamo concentrato la nostra attenzione sulle date 1998-2000-2003, che hanno un passo temporale simile e dunque soggetto a minor incertezza. La scelta di intervalli maggiori, infatti, può produrre dei risultati troppo generalizzati ed non corretti; l'analisi multitemporale utilizzata si basa su eventi discreti, mentre nessuna informazione è fornita su ciò che succede fra gli istanti selezionati. È quindi più facile dedurre l'evoluzione di un processo attraverso una sequenza di immagini sufficientemente ravvicinate nel tempo, senza cadere in evidenti errori di valutazione. La classificazione della carta finale è stata realizzata grazie alle opinioni di ricercatori che hanno lavorato nell'area durante il periodo studiato e consultando le precedenti pubblicazioni su argomenti e metodologie simili [Sader et al., 2003; Wilson et al., 2002 ; Hayes et al., 2001].

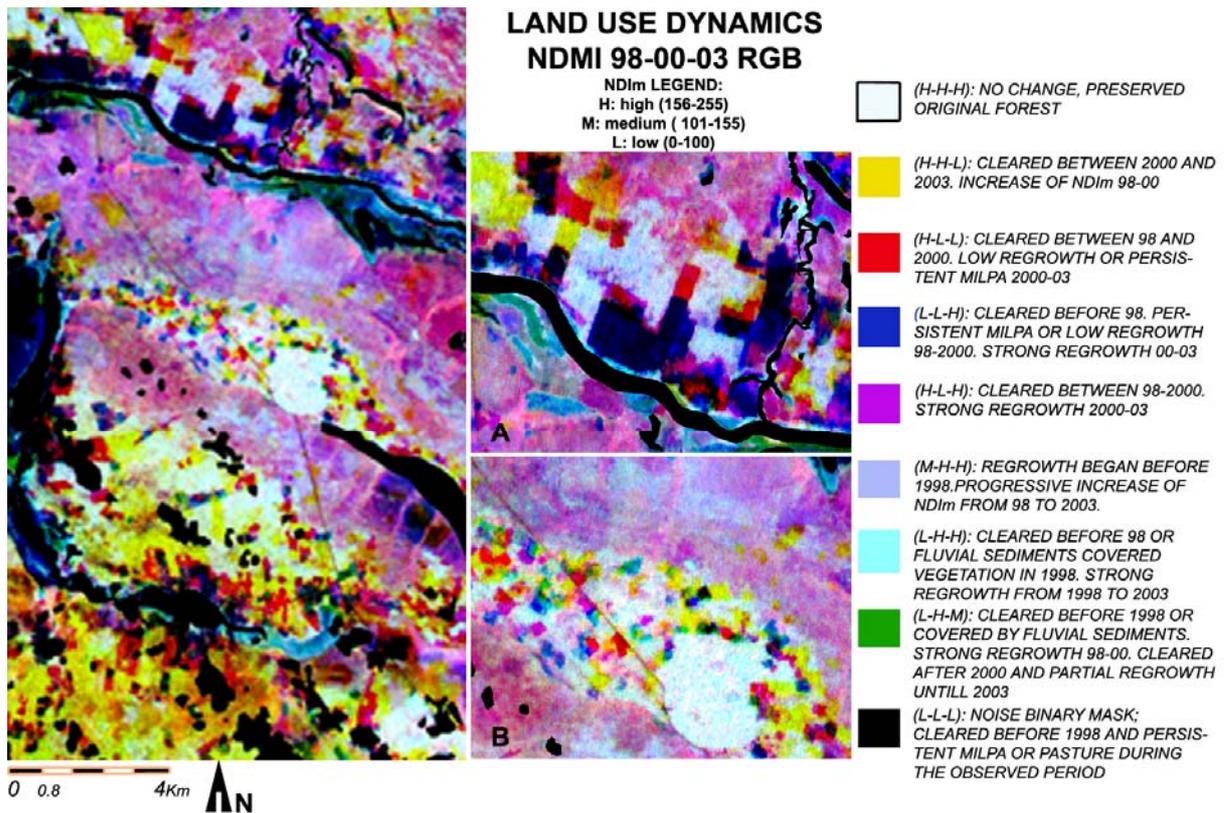


Fig 4-22: Analisi multitemporale delle dinamiche di uso del suolo nella Joyanca. Metodo basato su RGB (NDMI): attraverso il Rosso visualizzo i NDMI nella prima data 24/02/98; attraverso il Verde gli NDMI del 27/03/00 ed infine nel Blue gli NDMI del 07/05/03. I colori risultanti derivano dalla composizione dei ND nei tre canali e permettono dunque di seguire l'evoluzione degli indici nel tempo. A) Zoom a nord del rio San Pedro, evidenzia la persistenza delle prime aree deforestate. B) Zoom sull'area archeologica preservata La Joyanca. [Immagine da Bruno et al., 2006].

Il metodo RGB(NDMI) evidenzia la frammentazione e la dinamicità del paesaggio dal 1998 al 2003. I differenti colori sottolineano l'evoluzione dell'uso del suolo con un passo temporale di due o tre anni. Il colore bianco deriva dalla composizione di ND_{NDMI} elevati nelle tre date considerate ed indica dunque una foresta densa e preservata durante tutto l'arco temporale studiato. Come possiamo vedere queste parcelle sono molto ridotte e destinate probabilmente a scomparire; solamente la *selva alta* attorno al sito archeologico della Joyanca (Fig.4-22, B) dovrebbe mantenere le proprie caratteristiche originali, essendo parte di una zona protetta.

La carta RGB(NDMI) permette di sottolineare la rapidità e l'importanza dei cambiamenti intercorsi nella regione in soli cinque anni. L'utilizzo del fuoco come strumento di deforestazione causa un veloce e spesso incontrollato degrado della foresta originale. I toni di colore violetta rappresentano una vegetazione di *bajo* che è stata incendiata nel 1998 o nel 2000, e che ora sta progressivamente ricrescendo. Il fronte di deforestazione che arriva da sud-est, mascherato in fase di simulazione, è rappresentato dal colore giallo. Gli stessi toni rappresentano una parte di foresta decidua che ha perso le foglie nel 2003. Questa confusione nella classificazione è dovuta al fatto che il database del 07/05/2003 fa riferimento alla fine della stagione secca (massima siccità), che anticipa le piogge, mentre le altre due date rappresentano l'inizio della stessa stagione. L'assenza di una regolarità temporale nell'acquisizione delle immagini satellitari non ha permesso di eliminare questi effetti stagionali e fenologici. Questa classificazione è stata poi corretta attraverso un'analisi visuale e consultando le mappe fitogeografiche della regione, realizzate da [Métaillié et al., 2000]. Il metodo RGB(NDMI) offre la possibilità di interpretare e classificare i cambi nella foresta, seguendo l'evoluzione dei patterns spaziali in tre date successive. Possiamo così rispondere alle domande "da dove?" e "verso cosa?", per un'interpretazione temporale dei cambi nel periodo studiato [Hayes et al., 2001].

L'informazione offerta da RGB(NDMI) ci aiuta a meglio comprendere i processi di uso del suolo nella Joyanca e sarà utile nella fase di simulazione per la scelta dei criteri ambientali, per parametrizzare correttamente i modelli e valutare i risultati ottenuti.

4.2.4 Creazione dell'input per la simulazione

Una volta terminata la fase di preprocessamento delle immagini e di analisi dei cambi, abbiamo iniziato la preparazione dei dati con cui alimentare i modelli predittivi. Il periodo scelto per la calibrazione e validazione dei metodi di simulazione si basa sulle informazioni raccolte nel database multispettrale del 1998, del 2000 e del 2003.

4.2.4.1 Classificazione delle immagini

Per creare le mappe di uso del suolo in queste date abbiamo realizzato una classificazione supervisionata delle combinazioni in colore (scelte in accordo con il loro OIF). Nonostante la grande frammentarietà del paesaggio, abbiamo individuato quattro classi principali: foresta alta (foresta originale densa), foresta secondaria (vegetazione del *bajo* + stadio avanzato della ricrescita), zone umide paludose (aree di *cibal*) e deforestazione²⁶ (milpa + pascolo + suolo esposto). Inizialmente la classificazione è stata realizzata attraverso un algoritmo di massima probabilità che rappresenta il metodo più robusto e più utilizzato nel trattamento di immagini satellitari. Si considera che i ND in seno ad ogni categoria siano disposti secondo una distribuzione normale; questa ipotesi, teoricamente non sostenibile, ci permette però di descrivere praticamente la categoria attraverso una funzione di probabilità a partire dal valore della media e della deviazione standard. Se analizziamo una sola banda, la probabilità che un pixel "x", definito da un certo " ND_x " appartenga alla classe A,

²⁶ Deforestazione: chiameremo questa classe di Milpa, nonostante il termine identifichi generalmente una forma di agricoltura (vedi capitolo 4.1.3). Chiameremo dunque Milpa tutte le forme di disturbo e cambio profondo della foresta densa originale.

può essere calcolata attraverso la media " $\overline{ND_A}$ " e la varianza " σ_A^2 " della classe, secondo:

$$p\{x|A\} = \frac{1}{\sqrt{2\pi\sigma_A^2}} \exp\left[-\frac{(-ND_x - \overline{ND_A})^2}{2\sigma_A^2}\right] \quad (13)$$

Questo calcolo è effettuato per ogni categoria definita in classificazione; il pixel "x" sarà assegnato alla classe A solo se:

$$p\{x|A\} \geq p\{x|B\}, \quad \forall B \neq A \quad (14)$$

La media e la deviazione standard per ogni categoria, vengono calcolate sui pixels considerati durante la fase di "addestramento" (definisco sull'immagine piccoli poligoni rappresentativi di ogni classe). È inoltre possibile stabilire degli intervalli di probabilità per ogni categoria, facilitando così l'esclusione di quei pixels con una probabilità di appartenenza molto bassa (le code delle distribuzioni normali). Noi abbiamo scelto un limite di accettazione di 99%, il che significa che solo 1% dei pixels saranno ignorati, compensando la possibilità che qualche pixel sia stato incluso nei poligoni di addestramento di una classe per errore, o che si trovi sul confine fra due classi. Per ridurre la confusione fra le categorie, ossia la sovrapposizione fra le loro distribuzioni di probabilità, si consiglia di acquisire, per ogni classe, un numero significativo di campioni di piccole dimensioni. L'analisi della matrice di classificazione permette di evidenziare la percentuale di pixel classificata correttamente e la confusione media fra le categorie; nel nostro caso abbiamo rimarcato la difficoltà a separare *selva alta* da *bajo*, specie nelle immagini del 1988 e 1992 che fanno riferimento alla stagione delle piogge in cui la foresta secondaria ha un forte

vigore vegetale che la rende spettralmente simile alla foresta densa (Fig.4-23).

L'algoritmo di massima verosimiglianza lavora pixel a pixel, considerando esclusivamente il valore ND. Per migliorare l'accuratezza della nostra classificazione abbiamo utilizzato un algoritmo MAXVER-ICM di Spring 4.0 (INPE, Sao Paulo) che combina un classificatore di massima probabilità con un algoritmo Interacted Conditional Modes, che include la dipendenza spaziale fra ND. Nella prima fase l'immagine è categorizzata per l'algoritmo di massima probabilità considerando solo il valore digitale dei pixels. Nella fase seguente l'ICM considera l'informazione contestuale dell'immagine: la classe attribuita ad un pixel dipenderà tanto dal ND di quest'ultimo, quanto dalle classi attribuite ai suoi vicini.

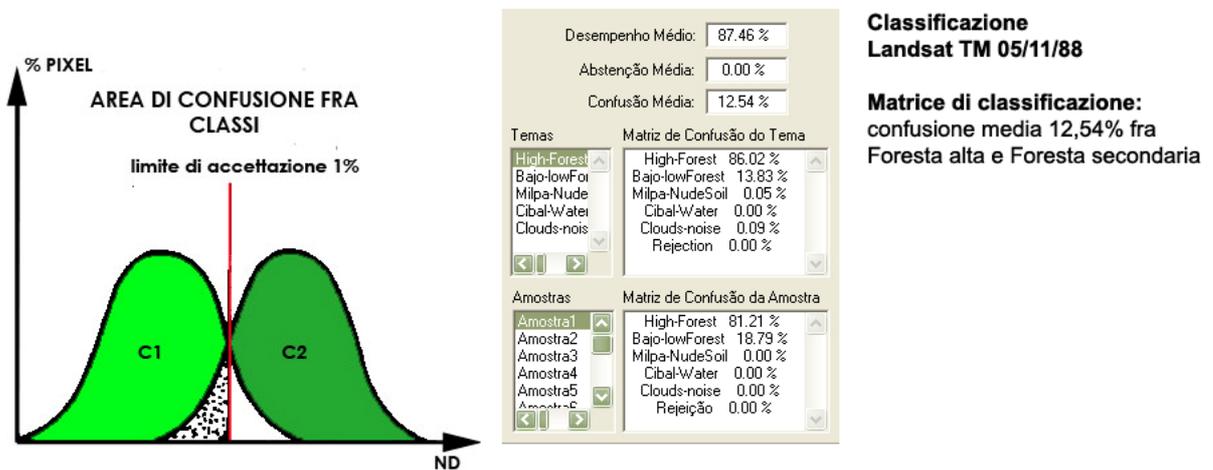


Fig 4-23: Classificazione con algoritmo di massima probabilità, delle immagini TM del 05/11/88. Lo spettro della Foresta alta e della foresta secondaria sono parzialmente sovrapposti e questo crea una confusione nella classificazione di queste due categorie.

L'utilizzo dell'algoritmo MAXVER-ICM aumenta di qualche punto percentuale la precisione di classificazione fra foresta alta e secondaria, ma il grado di confusione permane elevato

(10%). La questione non è di importanza rilevante nel nostro studio, poiché la carta classificata finale che alimenta i modelli predittivi, sarà binaria e rappresenterà solamente due classi: Foresta e Deforestazione (Fig.4-24).

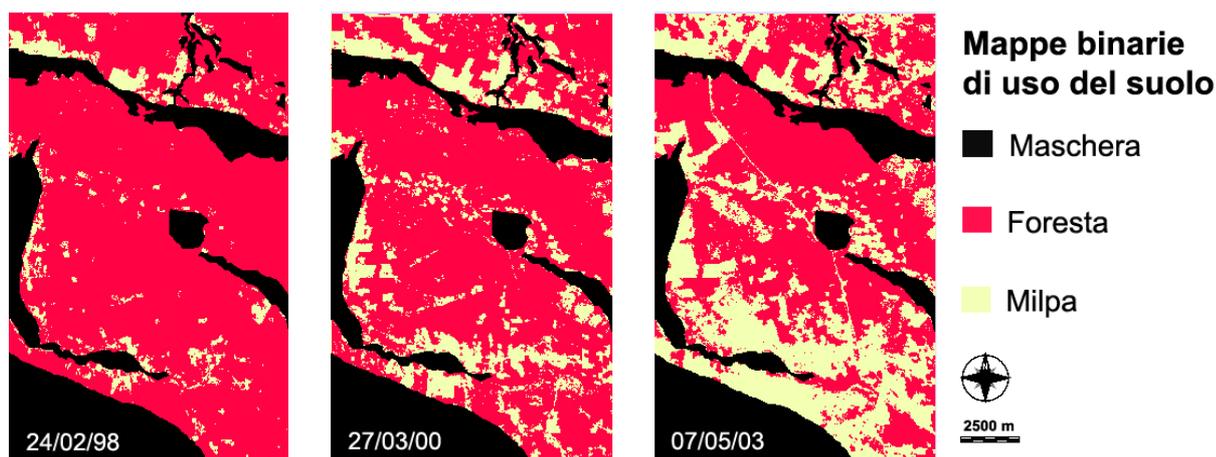


Fig 4-24: Immagini classificate del 1998-2000-2003. Le classi di uso del suolo sono ridotte a due: Foresta e Milpa (deforestazione). Queste carte alimenteranno i modelli predittivi e sono necessarie per l'analisi quantitativa dei cambi avvenuti. Le prime due immagini sono utilizzate in fase di calibrazione del modello per simulare 2003, scelta come referenza per la validazione dei risultati.

In [Chowdhury et al., 2004] si evidenzia come l'analisi della tessitura²⁷ in un'immagine, aumenti la separabilità statistica delle classi di copertura. Gli autori utilizzano, in ERDAS [Leica Geosystems], una finestra mobile 3x3, calcolando la varianza, e dunque l'eterogeneità spettrale della riflettanza, attorno al pixel centrale. [Chica-Olmo et al., 2006 & 2000] spiegano come un pixel sia spazialmente correlato con i suoi vicini, e che tale dipendenza può essere quantificata ed introdotta nel classificatore. Gli autori dimostrano le potenzialità del variogramma per l'analisi dell'autocorrelazione spaziale dei dati radiometrici; ogni classe di copertura ha un certo pattern di variabilità spaziale e dunque un differente variogramma. Questa informazione sulla

²⁷ Tessitura: variazione dei toni di grigio nei pixels dell'immagine, in funzione della scala

tessitura può essere considerata come un ulteriore attributo delle categorie di uso del suolo per migliorare la fase di classificazione.

Analogamente possiamo rimarcare una differente tessitura fra *selva alta* (alternanza di luci ed ombre che crea una variabilità locale maggiore ed un'immagine più "ruvida") e foresta secondaria (che presenta caratteristiche più omogenee). Queste informazioni devono essere integrate all'informazione radiometrica per migliorare un'eventuale classificazione delle classi di foresta (per il calcolo della biomassa, ad es.).

4.2.4.2 Variabili esplicative

L'analisi visuale delle immagini rialzate e migliorate, ha permesso di individuare i fattori che influenzano l'evoluzione della deforestazione nella Joyanca. Grazie ad un digitalizzatore abbiamo vettorizzato le strade, i villaggi ed i bacini idrici; operazioni logiche ed algebriche fra raster hanno permesso di quantificare le aree deforestate nel periodo studiato e le differenze fra le singole date. Successivamente abbiamo generato le matrici delle variabili indipendenti considerate, come le mappe di distanza. Queste ultime concernono tanto elementi lineari (strade, fiumi), quanto poligoni (aree deforestate, villaggi). Alcune mappe di distanza sono state calcolate ad ogni passo temporale perché riferite a variabili dinamiche che si evolvono nel tempo (distanza dalle strade e dalle aree deforestate). Partendo da informazioni Shuttle Radar abbiamo ricavato le curve di livello della regione ed un modello di elevazione digitale DEM; questi dati sono stati poi trasformati in un indicatore di potenzialità agricola, che varia in funzione dell'altitudine, come spiegato nel capitolo 4.2.1 (Fig.4-25).

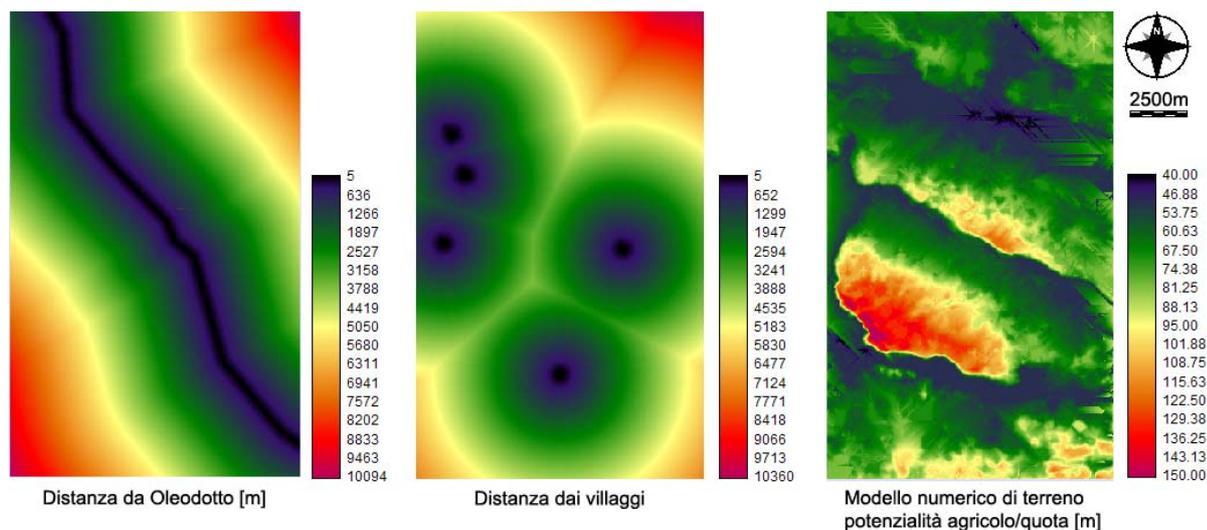


Fig 4-25: Esempi di variabili esplicative introdotte nei modelli predittivi per simulare le dinamiche di uso del suolo nella Joyanca. Per l'analisi spaziale abbiamo utilizzato Idrisi Andes (ClarkLab, USA); per la costruzione del DEM e del modello numerico di terreno abbiamo utilizzato ArcGis 9.

Riassumendo, l'input ai modelli predittivi sarà così composto: 10 variabili esplicative (mappe di distanza dalle strade nel 1998 e 2000, mappe di distanze dalle aree deforestate nel 1998 e 2000, mappa di distanza dai villaggi, mappa di distanza dal Rio San Pedro, mappa distanza dai laghi e dal fiume a sud del sito archeologico, mappa di distanza dall'oleodotto e DEM) e 3 mappe classificate di uso del suolo (carte binarie del 1998, 2000 e 2003). Ulteriori informazioni saranno poi applicate alle mappe di probabilità come incentivi-disincentivi alla deforestazione (zonazioni basate sulla opinione di esperti e del modellizzatore): disincentivo alla deforestazione per la zona nord del Rio San Pedro, data la presenza del narcotraffico, ed incentivo alla riforestazione per l'area centrale, le terre alte di *mesetas*, dove si pratica una agricoltura mobile più sostenibile rispetto alle regioni di *bajo* soggette a pastorizia.

4.2.4.3 Analisi esplorativa e scelta delle variabili

La scelta delle variabili esplicative si basa anzitutto sulle conoscenze del modellizzatore, sulla sua capacità di

rimettersi in gioco confrontandosi con studiosi di altre discipline utili alla comprensione del fenomeno in esame e sulla revisione delle precedenti pubblicazioni su argomenti simili. Ma è anche un processo empirico basato sulla visualizzazione delle distinte variabili sovrapposte alla mappa reale di uso del suolo, per identificare quelle più significative per spiegare i differenti tipi di cambi (Fig.4-26).

Sovrapposizione variabile esplicativa / mappa uso del suolo
Analisi visuale

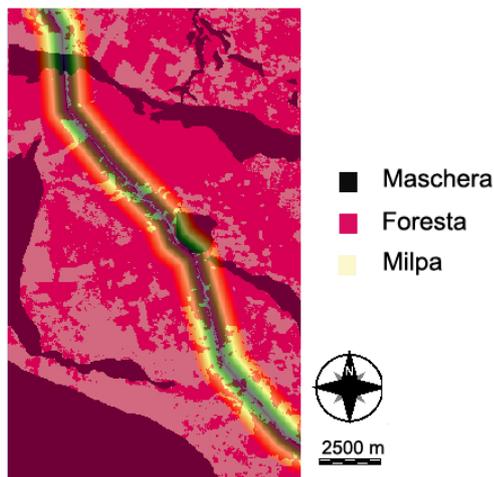


Fig 4-26: Sovrapposizione della variabile esplicativa "distanza dall'oleodotto", alla mappa di uso del suolo del 2003. L'analisi visuale permette di stabilire l'importanza della variabile e la sua significatività per spiegare i differenti tipi di cambio osservati.

Per un'analisi quantitativa dell'importanza delle variabili esplicative scelte, relativamente alle dinamiche di deforestazione studiate, abbiamo utilizzato la funzione *Test Explanatory Power* di Idrisi Andes [Eastman, 2006]. La procedura è basata sull'analisi della tabella di contingenza fra coppie di mappe, ed utilizza il coefficiente di Cramer (V)²⁸ come misura quantitativa del grado di associazione. Un alto valore

28 Per una descrizione dettagliata del metodo di Cramer si rimanda ad [Agresti, 2002].

(comunque sempre minore di 1) del coefficiente indica una buona significatività esplicativa della variabile esaminata, mentre un valore uguale a 0 indica che le due mappe sono indipendenti [Bonham-Carter, 1994].

Per il calcolo di "V" è necessario ottenere i valori di area nella tabulazione incrociata fra coppie di mappe in esame. La tabella di area fra la carta A e la carta B è la matrice \mathbf{T} , formata da elementi T_{ij} , dove $i=1, \dots, n$ classi di B (linee della tabella) e $j=1, \dots, m$ classi di A (colonne della tabella). I totali marginali di T sono definiti da T_i (somma dell' i -esima riga) e T_j (somma j -esima colonna), mentre $T_{..}$ identifica la somma totale di linee e colonne. Quando le due mappe sono indipendenti l'una dall'altra, senza alcuna correlazione fra loro, l'area attesa in ogni categoria di sovrapposizione, è data dal prodotto fra i totali marginali diviso la somma totale. Dunque l'area attesa per l' i -esima riga e j -esima colonna sarà:

$$T_{ij}^* = \frac{T_i * T_j}{T_{..}} \quad (15)$$

Definiamo il coefficiente statistico chi-quadrato come:

$$\chi^2 = \sum_i^n \sum_j^m \frac{(T_{ij} - T_{ij}^*)^2}{T_{ij}^*} \quad (16)$$

cioè l'espressione " $(\text{osservato} - \text{atteso})^2 / \text{atteso}$ ", che ha un limite inferiore di 0 quando l'area osservata è uguale all'area attesa e dunque le due mappe sono completamente indipendenti. Mano a mano che l'area osservata si discosta dall'area attesa, il chi-quadrato aumenta. Il coefficiente di Cramer V è definito come [Bonham-Carter, 1994]:

$$V = \sqrt{\frac{\chi^2}{T_{..} * M}} \quad (17)$$

dove "M" rappresenta il minimo di $(n-1, m-1)$.

[Eastman, 2006] suggerisce di ritenere nella modellizzazione le variabili con valori superiori a 0,15. Nella nostra ricerca, solo la variabile "mnt-potenzialità agricola" mostra un coefficiente di Cramer inferiore al limite consigliato (Tab.4-3); la deforestazione negli ultimi anni, si verifica in maniera piuttosto caotica (incendi incontrollati, disperazione dei contadini) ed interessa tanto le terre alte quanto le aree di *bajo*. Non è più evidente dunque la differenza precedentemente osservata fra attività agricole sostenibili sulle regioni più elevate delle mesetas e la pastorizia persistente nelle terre basse. Abbiamo comunque deciso di mantenere nello studio modellizzativo anche questa variabile, sperando che sia comunque utile per sottolineare, in futuro, le dinamiche di uso del suolo legate alle due diverse attività antropiche.

	D-Milpa(00)	D-Milpa(98)	D-oleodotto	D-strade(00)	D-strade(98)	D-Rio S.P	D-fiume sud	D-laghi	D-villaggi	mnt
CRAMER V	0,73	0,52	0,19	0,26	0,25	0,31	0,3	0,31	0,19	0,11

Tab 4-3: Coefficienti di Cramer per analizzare il potere esplicativo delle variabili scelte nello studio delle dinamiche di deforestazione nella Joyanca.

4.2.4 Considerazioni finali

Abbiamo visto come la costruzione dei dati geografici che alimentano la simulazione, sia un lavoro estremamente complesso ed oneroso. Il preprocessing delle immagini satellitari è stato caratterizzato da un grande numero di procedure operative per migliorare la qualità dei dati originali e facilitare l'estrazione di informazioni utili alla ricerca. L'importanza di tutte queste operazioni viene spesso sottovalutata nell'analisi dei risultati finali della simulazione. Modellizzare un fenomeno non significa solo costruire un modello matematico ed implementarlo in un linguaggio informatico; l'attività di studio iniziale dei

dati, la scelta delle variabili e dei parametri da utilizzare, rappresentano una parte importante del lavoro svolto, al pari, e forse più, della simulazione vera e propria.

CAPITULO No. 5

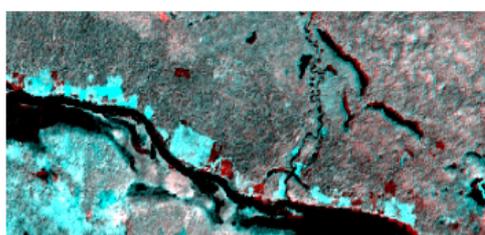
Validazione ed analisi comparativa dei modelli predittivi

*“E o esplendor dos mapas, caminho abstracto para a imaginação
concreta,
Letras e riscos irregulares abrindo para a maravilha.
(Fernanda Pessoa, 1933)*

5.1 Evoluzione del fenomeno e dei processi-pilota

*"Models will have poor predictive power if the driving processes of LUC change over the time. However even if there are no fundamental changes in these driving processes, the relationship between the independent variables and the propensity for disturbance will change for some simple mathematical reasons. After a certain area of land has been disturbed it can no longer become newly disturbed. Therefore the relationship between the independent variables and propensity for disturbance must change"*¹ [Pontius et al., 2004b, Pag.333].

Dall'analisi dei dati multispettrali e delle carte classificate abbiamo evidenziato un notevole cambiamento dei processi-pilota fra i periodi 1988-1998, 1998-2000 e 2000-2003. Nel primo intervallo considerato le dinamiche di deforestazione sono chiaramente concentrate al nord del Rio San Pedro, che ha rappresentato per secoli l'unica via di accesso e spostamento nella regione. Queste terre sono più elevate di quelle a sud, periodicamente inondate durante la stagione delle piogge. Le parcelle deforestate presentano una forma geometrica regolare e sono adibite alla milpa o alla pastorizia di sussistenza (Fig.5-1).



RGB serie multitemporale (zoom su Rio San Pedro)

R: TM 4 1988

GB: TM 4 1992

■ Primi insediamenti (forme geometriche al nord del Rio S.Pedro)
■ Sedimenti alluvionali (area irregolare al sud)

Fig 5-1 Analisi multitemporale dei cambi fra il 1988 ed il 1992. Primi insediamenti nella regione a nord del Rio San Pedro.

¹ I modelli possono avere un potere predittivo limitato se i processi-pilota dell'uso del suolo cambiano nel tempo. Comunque anche quando non ci sono cambiamenti importanti nei processi-pilota, le relazioni fra variabili indipendenti e probabilità di deforestazione, potrebbe cambiare per semplici ragioni matematiche. Dopo che un'area è stata deforestata, essa non può più essere nuovamente disturbata per un lungo periodo (durante il quale ho rigenerazione della vegetazione; non posso deforestare una parcella senza foresta). Quindi le relazioni fra variabili indipendenti e probabilità di deforestazione devono cambiare.

I processi di deforestazione sono guidati anzitutto dall'espansione di parcelle già esistenti, e secondariamente dalla nascita di nuovi poligoni deforestati. Dal 1994 al 1998 la colonizzazione nel sud della regione inizia a creare i primi spazi per l'agricoltura e la pastorizia in prossimità dei bacini idrici. Dal 1998 al 2000 gli effetti delle attività antropiche nelle zone a sud diventano evidenti lungo il perimetro delle strade e dei villaggi. Nelle terre comunitarie, le attività agricole si concentrano sulle parti più elevate delle *mesetas* e sono rappresentate da piccoli poligoni più o meno regolari. Dal 2000 si nota una rottura nei precedenti trends di occupazione della foresta. Le zone a nord del Rio San Pedro sono ora controllate dai narcotrafficanti che ostacolano l'attività degli immigranti. La deforestazione cresce vertiginosamente al sud, attorno ai villaggi, lungo il perimetro delle strade e dei sentieri, ed in prossimità dei grandi laghi. Nello stesso periodo si osserva l'arrivo di un nuovo fronte di deforestazione dalle regioni del sud, escluse dalla nostra ricerca. L'aumento demografico e la scarsità di terre produttive spinge i nuovi arrivati a bruciare la vegetazione in zone poco adatte alle attività agricole e pastorizie; la disperazione, e non un'obiettiva logica di produttività, guida questi processi di deforestazione che danno origine a numerosi poligoni irregolari sparpagliati su tutta l'area. La loro localizzazione è aleatoria e dipende dall'arbitrarietà del contadino. "*L'évolution est créatrice et non seulement logique*"² [Le Moigne, 1994, Pag.263]. È dunque impossibile definire una chiara relazione fra le variabili indipendenti ed i cambiamenti osservati, al fine di ottenere una loro quantificazione e collocazione spaziale precisa.

L'evoluzione delle condizioni sociali, economiche ed ambientali registrate a partire dal 1998 determina un rapido

² L'evoluzione è creatrice e non solo logica.

incremento delle aree deforestate (Fig.5-2), rispetto al decennio precedente.

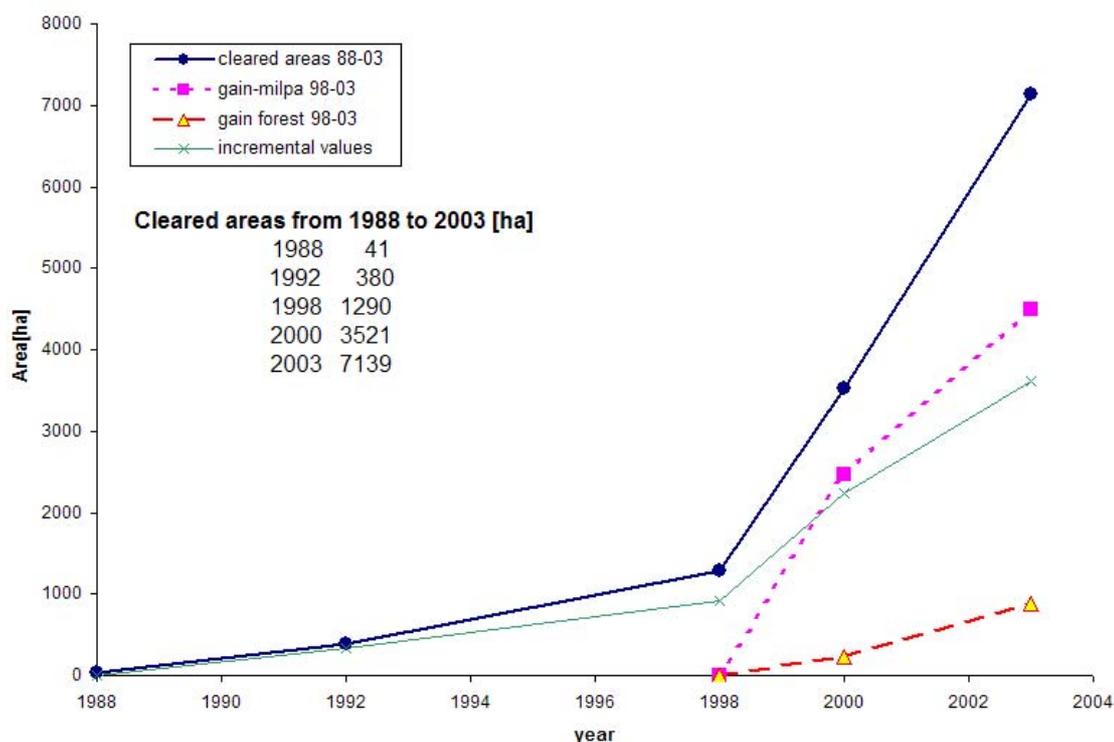


Fig 5-2: Trend di deforestazione dal 1988 al 2003. Guadagno e perdita per le due classi Foresta e Milpa durante il periodo simulato. Dati ottenuti dal processing di immagini SPOT e TM Landsat, classificate e successivamente filtrate per eliminare i "pixel persi" o piccoli gruppi di 2-3 pixel (la milpa ha dimensioni > 1ha).

Fra il 1998 ed il 2000 registriamo una perdita di foresta media di 1115 ettari/anno, di poco inferiore a quella registrata nei tre anni successivi (1206 ettari/anno); a livello quantitativo le informazioni utilizzate in fase di calibrazione dovrebbero dunque garantire una stima approssimativamente corretta della deforestazione nel periodo simulato. I cambiamenti nei processi-pilota, evidenziati dopo il 2000, determineranno probabilmente un errore maggiore nell'allocazione spaziale dei pixels deforestati.

Un modello può produrre risultati non soddisfacenti se i processi, gli agenti e le variabili cambiano nel tempo, o se i

dati di allenamento non catturano la reale complessità delle dinamiche di LUCC [Pontius et al., 2006].

I limiti di un modello sono legati tanto alla debolezza del fondamento teorico e della metodologia adottati, quanto alla scarsa conoscenza del contesto studiato; è spesso difficile separare questi due criteri nella valutazione del risultato finale. Per questo ribadisco, ancora una volta, come sia importante soffermarsi a lungo sulla costruzione del database che alimenterà il modello. Il tempo perso in questa prima fase sarà poi facilmente recuperato durante l'analisi dei risultati ottenuti [Costanza et al., 1998].

Le prime simulazioni offriranno delle mappe grossolane, imperfette; queste apportano, ciononostante, una nuova informazione sul sistema rappresentato, utile sia per migliorare le nostre conoscenze del fenomeno in esame, sia per affinare la parametrizzazione del modello.

*"Il y a des moments où nous devons simplement agir, en toute connaissance de notre ignorance des conséquences possibles, et nous devons toujours nous donner la possibilité de reconnaître nos erreurs passées et de changer le cours de notre action"*³ [Le Moigne, 1994, Pag.].

5.2 Applicazione dei modelli al caso della Joyanca

I quattro modelli descritti nel capitolo 3 sono stati alimentati con i dati prodotti nel capitolo 4. La fase di calibrazione utilizza le carte classificate del 1998 (inizio) e 2000 (fine) per prevedere l'evoluzione dei patterns spaziali nel 2003, scelta come data di riferimento per la validazione (Fig.5-3).

³ Ci sono momenti in cui bisogna semplicemente agire, essendo consapevoli della nostra ignoranza sulle conseguenze possibili, e dobbiamo sempre darci la possibilità di riconoscere i nostri errori passati e di cambiare il corso delle nostre azioni.

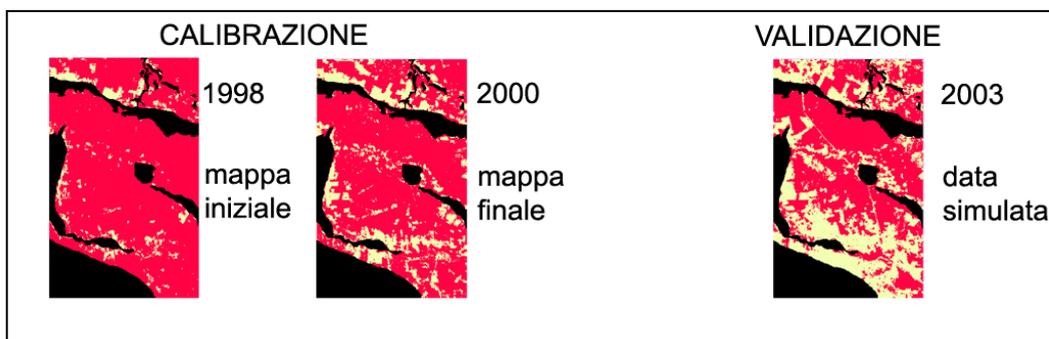
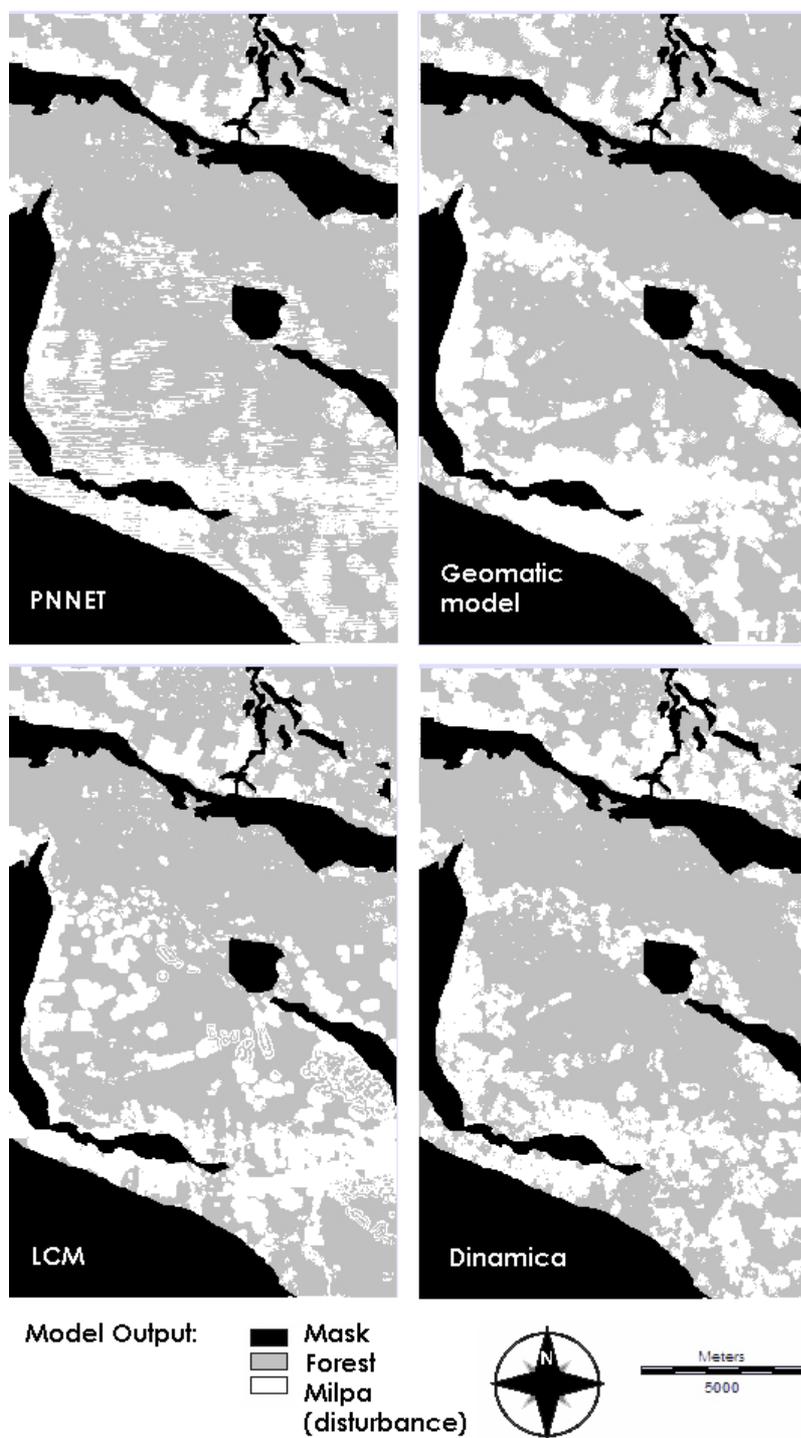


Fig 5-3: Periodo di calibrazione 1998-2000, e data di riferimento per la validazione (reale 2003 - simulato 2003).

Come già detto, abbiamo semplificato la complessità dei LUCC considerando solamente due dinamiche: deforestazione e riforestazione, senza suddividere i differenti stadi transitivi che le caratterizzano. I modelli predittivi sono stati alimentati con lo stesso input e permettono di proiettare l'evoluzione del paesaggio, relativamente allo stesso scenario. Quest'ultimo descrive in generale un mondo futuro possibile, caratterizzato da differenti ipotesi sulle condizioni economiche, politiche, demografiche, etc. [De Castro et al., 2007; Hauglustaine et al., 2004]. Lo scenario utilizzato ricalca le tendenze e le informazioni osservate negli ultimi anni del periodo di allenamento (scenario storico).

I modelli PNNET, CA_Markov (modello geomatico), LCM e Dinamica EGO, hanno dunque prodotto quattro mappe simulate di uso del suolo nel 2003 (Fig.5-4).



Simulated Land Cover in 2003

Fig 5-4: Output dei modelli predittivi. Carte di uso del suolo nel 2003, partendo da informazioni raccolte nel 1998 e 2000. [Immagine da Follador et al., 2008]

5.3 Validazione e discussione dei risultati

*"Il est ainsi fréquent, lors de l'exposé d'un modèle, que la question de la validation, qui se veut être la question piège dans ce domaine, soit posée, plongeant le conférencier dans un embarras bien visible. Il est alors répondu que le travail présenté est inachevé et que la phase de validation doit être abordée prochainement ou que le modèle est pour l'instant trop abstrait et devrait être raffiné avant que de s'intéresser à la validation. La meilleure réponse consisterait à retourner la question : « Qu'entendez-vous par validation ? ». Il serait alors régulièrement répondu qu'un modèle est valide lorsque ses valeurs de sortie sont proches des données observées..."*⁴ [Amblard et al., 2006b, Pag.103].

La validazione⁵ classica è dunque basata su un raffronto fra la mappa simulata e la mappa reale, o più in generale fra l'output di un modello e un insieme di osservazioni empiriche.

Questa fase della modellizzazione verifica l'adeguatezza fra il modello ed il sistema rappresentato, in funzione degli obiettivi della ricerca. La semplice analisi quantitativa dei risultati deve essere affiancata dalla valutazione sulle capacità del modello di riprodurre le caratteristiche

4 Capita spesso che dopo l'esposizione di un modello, si ponga una domanda sulla validazione, punto debole di questo dominio, lasciando il conferenziere in un imbarazzo visibile. Spesso si risponde affermando che il lavoro non è ancora terminato e che la fase di validazione verrà affrontata in un futuro prossimo, oppure che per ora il modello è troppo astratto perché si possa pensare alla validazione. La migliore risposta comunque sarebbe quella di domandare: "cosa intende lei per validazione?". Sempre vi sarà risposto che un modello è validato quando i suoi output sono simili ai valori osservati nella realtà.

5 In alcune pubblicazioni si suddivide la validazione in due fasi distinte: interna ed esterna. La prima verifica la conformità fra il programma implementato ed il modello teorico (il mio modello possiede le proprietà volute? È esattamente questo che volevo fare?); parallelamente verifico se il programma sia stato costruito rispettando i criteri di robustezza e flessibilità che garantiscono la riproducibilità delle esperienze condotte. Infine effettuo un test per riscontrare eventuali errori informatici o logici di base (cicli infiniti, divisioni per zero, etc.). La validazione esterna invece analizza i risultati dei modelli e li confronta con i dati reali di riferimento; in questa tesi utilizziamo il termine "validazione" esclusivamente per identificare la validazione esterna. Il termine "verifica" è preferito all'espressione "validazione interna" per caratterizzare le fasi iniziali della modellizzazione (vedi Fig.3 del Capitolo 2).

evolutive dei processi spaziali (forma ed isometria delle parcelle deforestate, ad es.).

Negli ultimi anni sono stati sviluppati numerosi approcci statistici per misurare la concordanza fra due mappe categoriche [Rodrigues et al., 2007; Pontius et al., 2006, 2004 & 2001; Boots et al., 2006; Kleijnen, 1999]. Nessun metodo offre comunque una risposta esaustiva al problema della validazione, in particolare per quello che riguarda l'analisi della distribuzione spaziale delle discordanze.

5.3.1 Analisi visuale

Il primo approccio adottato si basa su una semplice analisi visuale dei risultati di ogni singolo modello in relazione alla mappa reale di uso del suolo nel 2003 (da Fig.5-5 a Fig.5-8). Sebbene riduttivo questo metodo permette di ottenere una prima indicazione sulle prestazioni dei modelli, evidenziando eventuali grandi incongruenze in termini quantitativi e spaziali. Parliamo qui di valutazione generale dei risultati, piuttosto che di validazione.

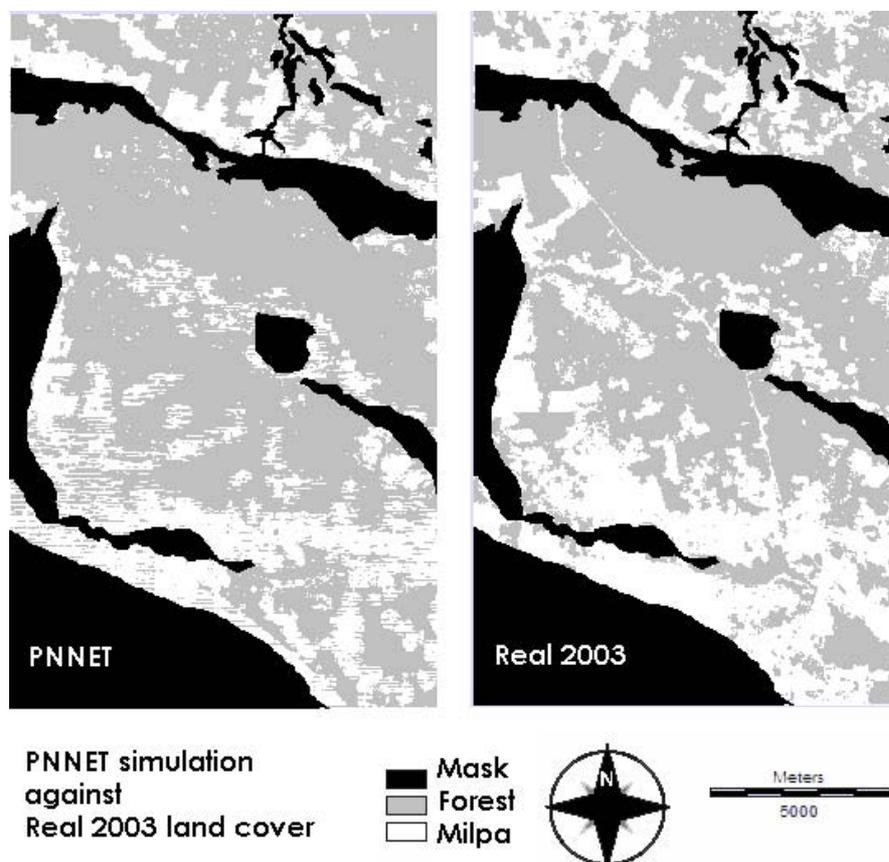


Fig 5-5: Carta di uso del suolo simulata da PNNET e carta reale nel 2003.

Possiamo notare come PNNET sottostimi la magnitudine della deforestazione nella zona sud della Joyanca; la performance migliora per le zone centrali e al nord (Fig.5-5). La localizzazione dei pixels deforestati è abbastanza realistica, ma le parcelle risultano troppo frammentate, probabilmente come risultato delle approssimazioni adottate nella rete neurale (pixels frontiera).

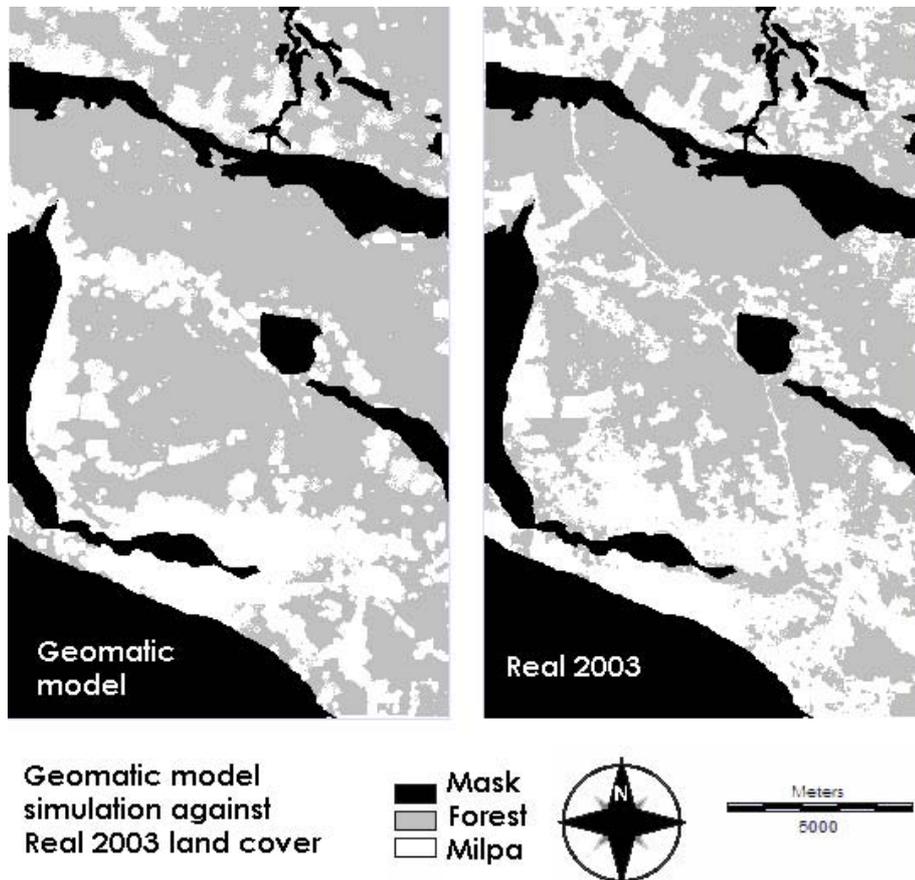


Fig 5-6: Carta di uso del suolo simulata da CA-Markov e carta reale nel 2003.

Il modello CA_Markov aumenta la dimensione delle aree deforestate rispetto a PNNET (Fig.5-6). Parallelamente notiamo un'aggregazione compatta delle parcelle nelle zone centro-meridionali, poco rappresentativa della reale frammentazione del paesaggio.

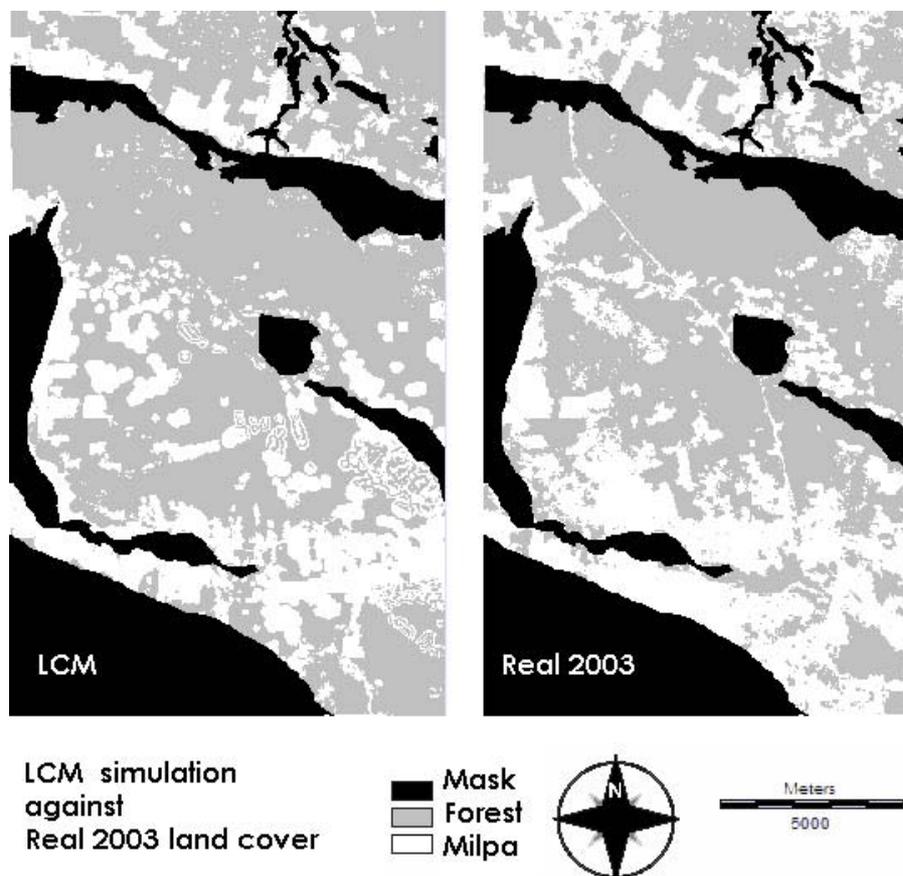


Fig 5-7: Carta di uso del suolo simulate da LCM e carta reale nel 2003.

Land Cover Modeler (LCM) produce una distribuzione spaziale delle aree deforestate poco soddisfacente, in cui risaltano tanti poligoni irregolari più o meno assemblati nella zona centrale (Fig.5-7). Alcuni di questi hanno una forma irrealistica caratterizzata da un'alternanza concentrica di foresta e milpa.

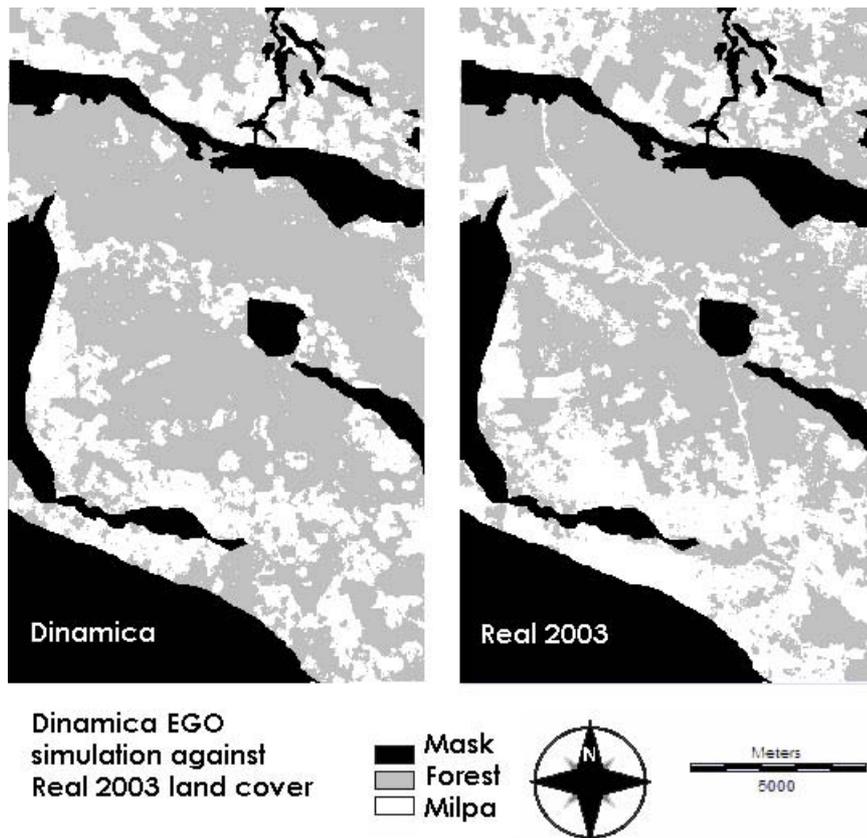
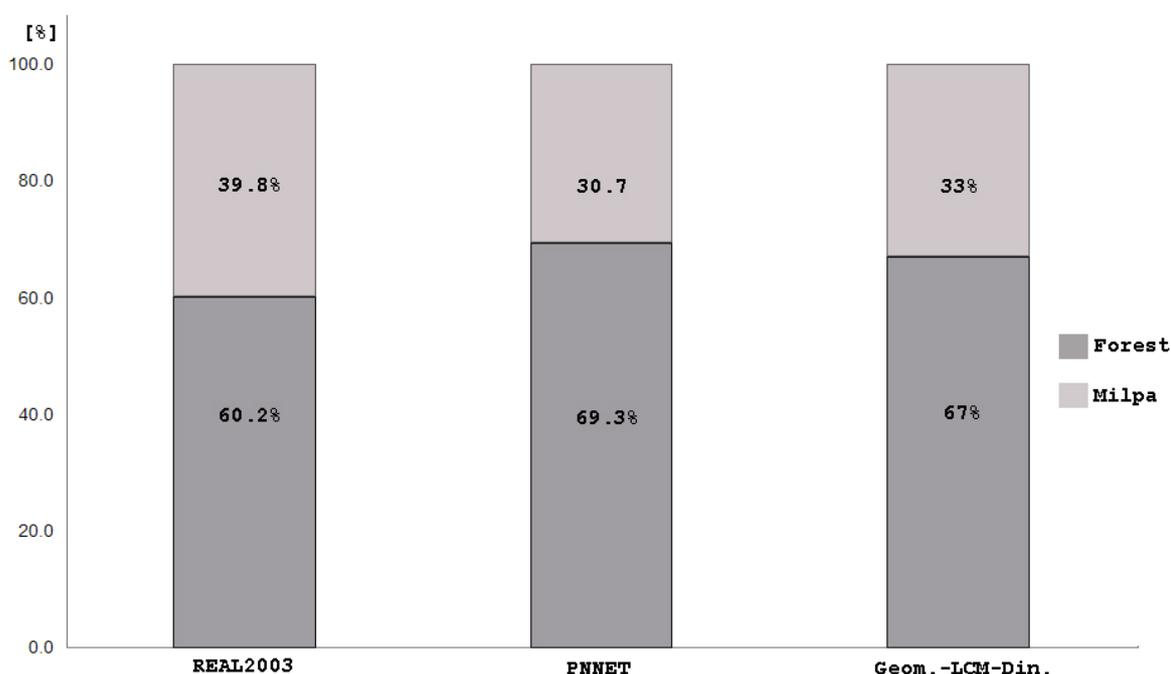


Fig 5-8: Carta di uso del suolo simulate da Dinamica Ego e carta reale nel 2003.

Infine analizziamo il risultato di Dinamica EGO. Osserviamo ancora una sottostima della deforestazione nella zona a sud, ed una sovrastima nelle zone a nord del Rio San Pedro. La forma e la distribuzione delle parcelle deforestate è abbastanza realistica, in particolare in prossimità del sito archeologico (Fig.5-8).

5.3.2 Analisi quantitativa

In ogni mappa simulata calcoliamo il numero di pixels predetti per ogni categoria di copertura. Il confronto con le quantità registrate nella carta reale del 2003 offre un'idea, spaziale, della congruenza fra osservato e simulato (Tab.5-1).



	Forest [pixels]	Milpa [pixels]	Tot. [pixels]	Forest [ha]	Milpa [ha]	Forest (%)	Milpa (%)
Real2003	270080	178492	448572	10803.2	7139.6	60.2	39.8
PNNET	311019	137553	448572	12440.8	5502.1	69.3	30.7
Geom,LCM,Din	300767	147805	448572	12030.6	5912.2	67.0	33.0

Tab 5-1: Concordanza quantitativa fra mappa reale e simulata nel 2003.

PNNET mostra una peggiore previsione quantitativa rispetto ai tre metodi che utilizzano l'analisi prognostica di Markov per calcolare il numero di pixels in ogni transizione. I modelli non riescono a catturare l'estensione della deforestazione che si registra fra il 2000 ed il 2003; ciò è probabilmente legato sia alla poca rappresentatività dei dati

storici utilizzati, sia a limiti intrinseci alla modellizzazione, come osservato in [Pontius et al., 2004c].

Rileviamo un'incongruenza quantitativa del 9% circa per la mappa predetta da PNNET, e del 7% circa per gli altri modelli, con un'evidente sottostima delle aree deforestate.

5.3.3 Analisi pixel-a-pixel

Due carte possono avere la stessa percentuale di pixel per ogni classe di uso del suolo, ma diverse allocazioni spaziali di queste quantità. Lo studio della incongruenza spaziale è abbastanza complesso; essa deriva da uno slittamento di posizione fra pixel di categorie differenti [Pontius, 2002]. È molto importante separare l'analisi statistica quantitativa dei cambi di uso del suolo dall'analisi spaziale, con la localizzazione delle differenze, per meglio comprendere le dinamiche ambientali in atto e per migliorare i punti deboli della modellizzazione.

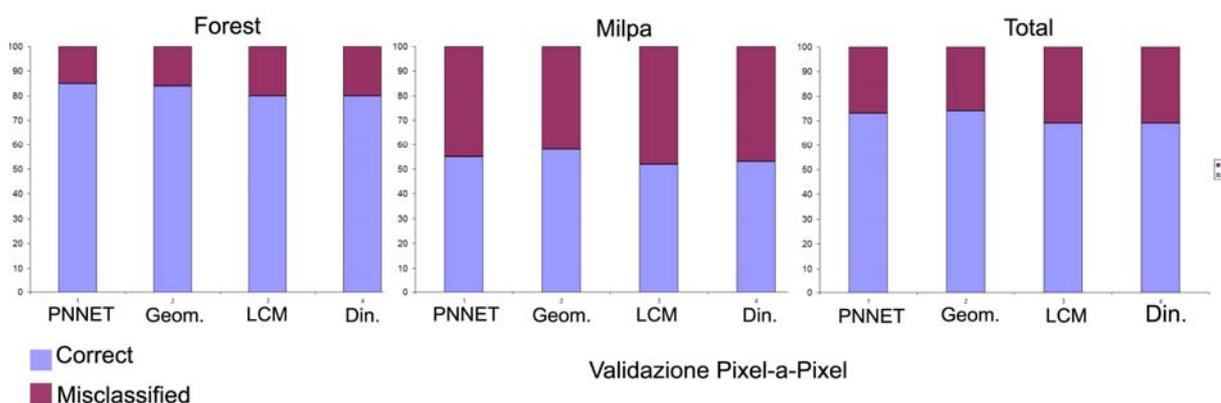
La validazione pixel-a-pixel (Tab.2) è un metodo molto semplice, ma abbastanza limitato poiché non considera la prossimità spaziale della congruenza fra pixel reale e pixel simulato. *"If a cell is misclassified, then the entire cell counts as an error, even if the correct classification is found in the adjacent cell"*⁶ [Pontius et al., 2004b, Pag.330].

La validazione pixel-a-pixel sottolinea ancora una volta come tutti i modelli abbiano una scarsa capacità di prevedere la deforestazione, con una precisione media del 54,5%. Questa percentuale aumenta fino al 82,25% per la classe foresta. In base a (Tab.5-2) possiamo concludere che il modello CA_Markov ha una migliore performance per la previsione delle parcelle deforestate (58%), mentre PNNET offre una miglior stima per le aree forestali (85%).

⁶ Se il pixel non è classificato correttamente, l'intera cella conta come un errore, anche se la classificazione corretta si trova nella cella adiacente.

Cell-by-Cell Agreement

	Forest (pixels)		Milpa (pixels)		Total (pixels)	
REAL 2003	270080		178492		448572	
	Correct	Mis	Correct	Mis	Correct	Mis
PNNET	230656 (85%)	39424 (15%)	98129 (55%)	89363 (45%)	328785 (73%)	119787 (27%)
Geomatic	226902 (84%)	43178 (16%)	104627 (58%)	73865 (42%)	331529 (74%)	117043 (26%)
LCM	215852 (80%)	54228 (20%)	93589 (52%)	84903 (48%)	309441 (69%)	139131 (31%)
Dinamica	216848 (80%)	53232 (20%)	94580 (53%)	83912 (47%)	311428 (69%)	137144 (31%)



Tab 5-2: Validazione pixel-a-pixel fra mappa reale e simulata nel 2003.

5.3.4 Land Use Cover Change Budget

L'analisi della matrice di transizione ci permette di evidenziare quantitativamente le più importanti dinamiche avvenute e di calcolare il cambio netto fra due date, per ogni categoria di uso del suolo. Questo valore, sebbene fornisca un'informazione utile, non è sempre rappresentativo della reale evoluzione di un fenomeno. Una diminuzione del cambio netto non è sempre associabile ad una diminuzione delle dinamiche di LUCC. Queste possono determinare uno spostamento delle diverse categorie fra l'istante 1 e l'istante 2 all'interno dell'area di studio, mentre la quantità totale rimane inalterata. Per esempio una parte di bosco in Peten può essere deforestata, ma allo stesso tempo un'area di milpa

abbandonata di uguale dimensione può riforestarsi. Questo tipo di cambio è detto "swap"⁷. Il calcolo del cambio netto non permette di catturare questo tipo di informazione e, ritornando all'esempio del bosco, non evidenzia nessuna dinamica di LUCC (deforestato-riforestato=0). "The concept of swap is particularly important because some of the most common sources of land-cover data are available in a form that gives only the quantity of each land-cover type over the time...the danger is that the net change can dramatically underestimate the local change on the landscape"⁸ [Pontius et al., 2004c, Pag.253].

Aggiungiamo inoltre che la dimensione del pixel adottato nella modellizzazione deve essere in grado di catturare la distanza che caratterizza ogni transizione⁹; la scelta di una risoluzione troppo grossolana può mascherare alcuni tipi di cambi che avvengono a scala locale e non sono evidenziati quantitativamente dal cambio netto. [Pontius et al., 2004a & 2004c; Soares-Filho et al., 2002] evidenziano l'utilità di un'analisi multi-risoluzione dei LUCC per esaminare le differenti distanze su cui avvengono le transizioni di ogni categoria. Questa validazione è basata sulla raffronto di due mappe a differenti risoluzioni, ottenute aggregando celle limitrofe in blocchi via via maggiori. Quest'ultimi hanno un grado di appartenenza ad ogni classe (*fuzzy membership*) che dipende dalla porzione di pixels con risoluzione fine nel blocco, appartenenti ad ogni categoria .

Il calcolo del LUCC budget [Pontius et al., 2004c] permette di considerare il guadagno e la perdita per ogni classe e l'importanza dei cambi avvenuti dal 2000 al 2003, in relazione alle dinamiche di uso del suolo reali (Tab.5-3). Utilizziamo

7 Swap: cambio di posizione.

8 Il concetto di swap è particolarmente importante, poiché le principali fonti di dati sulle classi di copertura riportano delle informazioni solamente sulla quantità di ogni categoria nel tempo...il pericolo è che il calcolo del cambio netto possa drammaticamente sottostimare le reali dinamiche di LUCC del paesaggio.

9 Questa distanza è comunemente detta *swapping distance*.

dunque questi valori per confrontare le prestazioni dei modelli, focalizzando l'attenzione sulla loro capacità di rappresentare gli *swaps*.

Land Use Cover Change Budget

Milpa	GAIN%	LOSS%	TOT. Change%	abs (net change)%	SWAP (Tot-abs)%
REAL2003-REAL 2000	24.01	4.92	28.93	19.10	9.83
PNNET2003-REAL2000	9.97	0.00	9.97	9.97	0.00
GEOM2003-REAL2000	14.02	1.76	15.78	12.26	3.52
LCM2003-REAL2000	13.37	1.12	14.49	12.26	2.23
DIN2003-REAL2000	15.05	2.79	17.84	12.26	5.59

Tab 5-3: LUCC budget. Confronto fra dinamiche di deforestazione reali e simulate.

L'evoluzione reale della deforestazione dal 2000 al 2003 mostra una percentuale di cambi totali considerevole (28,93%), notevolmente sottostimata da tutti i modelli.

PNNET non prevede nessuna riforestazione nel periodo simulato ignorando questa transizione; la performance è insoddisfacente ed il valore ottenuto è dunque il più basso (9,97%). Il modello stima che non ci sarà nessun tipo di ricrescita della vegetazione nelle parcelle precedentemente deforestate, che al contrario continueranno ad aumentare in dimensione.

Dinamica EGO ha la migliore capacità di riprodurre le dinamiche reali fra il 2000 ed il 2003, come indicato dalla *ratio* "Swap/net-change" uguale a 0,46, molto simile al valore reale di 0,51. Nonostante ciò la quantità di cambi totali prevista (17,84%), resta manifestamente sottostimata.

I modelli CA_Markov e LCM hanno un comportamento intermedio, mostrando, come PNNET, una maggior propensione a prevedere la persistenza delle categorie di uso del suolo.

5.3.5 Complementarietà dei modelli

Verifichiamo ora la consistenza fra modelli intersecando la mappa di uso del suolo reale con le mappe simulate (Tab.5-4). La sovrapposizione delle 5 carte mostra una percentuale di pixel correttamente predetti del 52,24%, evidenziando la divergenza di comportamento dei 4 modelli, che singolarmente hanno una precisione media attorno al 70%. Questa differenza è dovuta alla diversità degli approcci teorici (automatico-ottimizzato, semiautomatico, supervisionato) delle metodologie utilizzate per l'allocazione spaziale dei pixels.

Per ogni intersezione la percentuale di foresta è sempre maggiore di quella per la milpa, dovuto principalmente alla persistenza ed alle notevoli dimensioni ed omogeneità delle aree forestali (condizione ottimale). La deforestazione, al contrario, evolve in maniera rapida e caotica, interessando parcelle di dimensioni variabili e particolarmente frammentate.

		Accurate prediction scores %			
		Milpa	Forest	Total	Improvement
Intersection of 4 models		13.2	39.04	52.24	-
Intersection of 3 models	1)PNNET \cap Geom \cap LCM	15.59	42.89	58.48	6.24
	2)PNNET \cap Geom \cap Din	14.96	42.88	57.84	5.60
	3)PNNET \cap LCM \cap Din	14.31	40.32	54.63	2.39
	4)Geom \cap LCM \cap Din	13.84	40.1	53.94	1.70
Intersection of 2 models	1)PNNET \cap Geom	18.51	47.31	65.82	13.58
	2)PNNET \cap LCM	17.3	45.28	62.58	10.34
	3)PNNET \cap Din	16.64	45.24	61.88	9.64
	4)Geom \cap LCM	17.39	44.54	61.93	9.69
	5)Geom \cap Din	17.09	44.6	61.69	9.45
	6)LCM \cap Din	15.26	41.8	57.06	4.82
Single model's prediction	1)PNNET	21.88	51.42	73.3	21.06
	2)Geom	23.32	50.58	73.9	21.66
	3)LCM	20.86	48.12	68.98	16.74
	4)Din	21.08	48.34	69.42	17.18

Tab 5-4: Accuratezza della simulazione (%) in funzione dell'intersezione fra i risultati dei modelli. L'incremento totale della precisione è calcolato rispetto al valore ottenuto dalla sovrapposizione dei 4 modelli.

La combinazione fra PNNET-CA_Markov-LCM è la migliore fra le possibili alternative con 3 modelli, con un aumento del 6,24% rispetto al valore precedente. Utilizzando PNNET o CA_Markov separatamente con altri modelli otteniamo degli incrementi di prestazione più ridotti.

La stessa conclusione può essere fatta analizzando le differenti combinazioni fra 2 modelli, dove la maggior consistenza è quella della coppia PNNET-CA_Markov (13.58%). Il risultato è interessante, se consideriamo che essi utilizzano approcci metodologici opposti; PNNET è un modello automatico-ottimizzato, mentre CA_Markov si basa sulla supervisione del modellizzatore. L'incremento minore (4,82%) si registra utilizzando LCM-Dinamica EGO.

Analizziamo infine il comportamento di ogni singolo modello. CA_Markov e PNNET realizzano la miglior percentuale di previsione corretta (73,9% e 73,3%) con un incremento della concordanza del 21% circa rispetto al valore ottenuto dall'intersezione dei quattro modelli.

5.3.6 Simulazione e frequenza del cambio osservato

Abbiamo visto come le performance di ogni modello siano influenzate dalla natura delle classi di copertura e dalle caratteristiche evolutive dei loro patterns spaziali. PNNET mostra una maggior capacità di rappresentare le aree forestali (85,4% della foresta è simulato correttamente), mentre CA_Markov risulta migliore per le dinamiche della milpa (58,62%). Vogliamo ora analizzare il comportamento dei singoli modelli in funzione del numero di cambi verificatesi in un certo pixel, dal 1998 al 2003 (Fig.5-9). La modellizzazione è basata su una discretizzazione temporale; nel periodo simulato posso dunque osservare un massimo di 2 transizioni fra classi di uso del suolo.

PNNET ha la miglior performance (93,74%) per la simulazione delle aree statiche. Nella Joyanca la persistenza caratterizza principalmente le zone forestali e, in maniera meno importante, alcune parcelle deforestate adibite alla pastorizia. Parallelamente notiamo come questo modello abbia prestazioni ridotte (4,47%) per la simulazione delle aree più dinamiche, soggette ad una rapida alternanza di usi del suolo.

Queste ultime sono meglio rappresentate da CA_Markov (23,85) e Dinamica EGO (21,99%); la migliore capacità di catturare i processi che determinano numerosi cambi di copertura è facilitata dalla costante supervisione del modellizzatore o dell'esperto. Parliamo, in questo caso, di modellizzazione basata sui processi, piuttosto che di modellizzazione basata sui dati storici. PNNET e LCM, al contrario, sono dei metodi semi-automatici, in cui le relazioni fra dinamiche spaziali e variabili esplicative sono definite dal perceptrone sulla base dell'input e della topologia della rete. La possibilità di influenzare questi risultati, in base alla propria esperienza, è più ridotta.

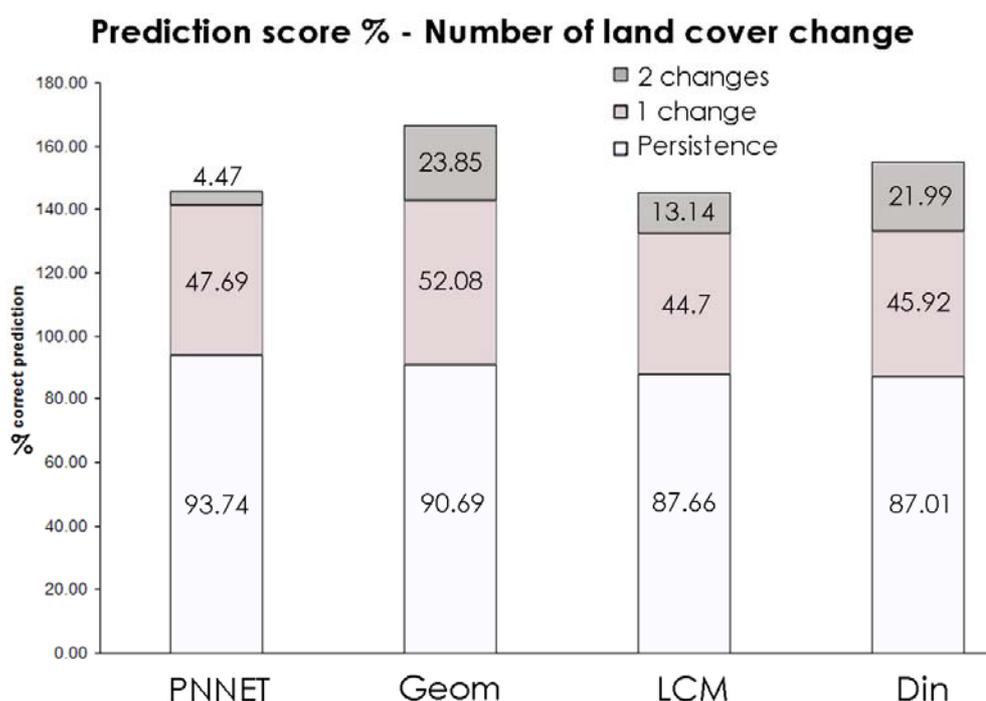


Fig 5-9: Accuratezza della simulazione (%) in funzione del numero di cambi di uso del suolo registrati in ogni pixel.

Tutti i metodi mostrano una grande capacità di rappresentare la persistenza delle classi di copertura, dovuta alla semplicità di questa non-transizione; *"if the model predicts persistence, then the model is usually accurate"*¹⁰ [Pontius et al., 2004c, Pag.251]. Molte volte una elevata congruenza fra reale e simulato è più influenzata dalla staticità delle classi di uso del suolo presenti nella regione di studio, piuttosto che dall'approccio metodologico utilizzato.

I modelli si comportano in maniera simile nelle aree con una sola transizione fra le classi di copertura; il valore più elevato si registra per CA_Markov (52%), mostrando una brusca perdita di prestazioni rispetto alla persistenza. Questi valori si abbassano ulteriormente quando il numero di cambi avvenuti in un'area aumenta a scapito della staticità della copertura.

5.3.7 Validazione Fuzzy

Abbiamo già accennato che il confronto pixel-a-pixel è una tecnica semplice, ma limitata, per valutare la congruenza fra la mappa reale e quella simulata, poiché non include nell'analisi la prossimità spaziale della congruenza reale-simulato.

Facciamo ora una validazione all'interno di un certo vicinaggio di un pixel; due mappe che non corrispondono esattamente cella-a-cella, possono presentare dei patterns spaziali simili, e la corretta classificazione si potrebbe trovare nei pixel immediatamente adiacenti. Per realizzare questa analisi, abbiamo utilizzato il *"vicinity based comparison method"* di Dinamica EGO [Rodrigues et al., 2007], basato sul concetto di *fuzzy location* [Hagen_Zanker, 2006 & 2003] in cui il valore di ogni pixel è influenzato da se

10 Se il modello prevede persistenza, in genere il modello è preciso.

stesso e dai pixels che si trovano nella finestra di vicinaggio. La scelta di quest'ultima dipende dalla dimensione del pixel e dagli obiettivi; noi abbiamo utilizzato il vicinaggio di Moore 5X5, considerando come accettabile un errore di allocazione delle celle deforestate in un intorno di 100 metri¹¹. Il nostro scopo, infatti, non è prevedere "quale albero sarà deforestato in t_1 ", ma individuare i futuri trends di deforestazione e le aree più vulnerabili (hot spots). Quando la dimensione del pixel è maggiore oppure il problema studiato richiede una mappa spaziale più dettagliata (mappa di rischio di frane, ad es.), la finestra di vicinaggio dovrà essere ridotta, oppure questo tipo di validazione abbandonato.

La mappa di similarità *fuzzy* evidenzia la congruenza spaziale fra mappa reale e simulata; questa quantità varia da 0 (discordanza) ad 1 (perfetta sovrapposizione). I valori intermedi sono calcolati utilizzando una funzione di decadimento con la distanza dal pixel centrale della finestra. Il metodo *fuzzy* permette dunque di ampliare la validazione booleana (errore/corretto), ponderando la distribuzione dello stato di ogni cella attorno al pixel centrale [Godoy et al., 2008]. Avremmo dunque una scala di valori continua [0,1], dove lo 0 indica l'errore, 1 la perfetta sovrapposizione, ed i valori intermedi il grado di congruenza fuzzy, all'interno della finestra di vicinaggio (Fig.5-10). Questi ultimi si collocano principalmente sui bordi delle aree deforestate, poiché l'errore commesso riguarda quasi esclusivamente i processi di deforestazione. Le zone persistenti presentano valori simili nella validazione hard e fuzzy.

Consideriamo come parzialmente corretti (relativamente al nostro obiettivo) i pixels con un grado di congruenza maggiore di 0; l'incremento di precisione ottenuto rispetto

¹¹ La finestra di Moore ha un lato di 5 pixel. Ogni pixel ha una risoluzione di 20m.

alla semplice sovrapposizione cella-a-cella, può essere utilizzato per confrontare i risultati dei modelli (Tab.5-5).

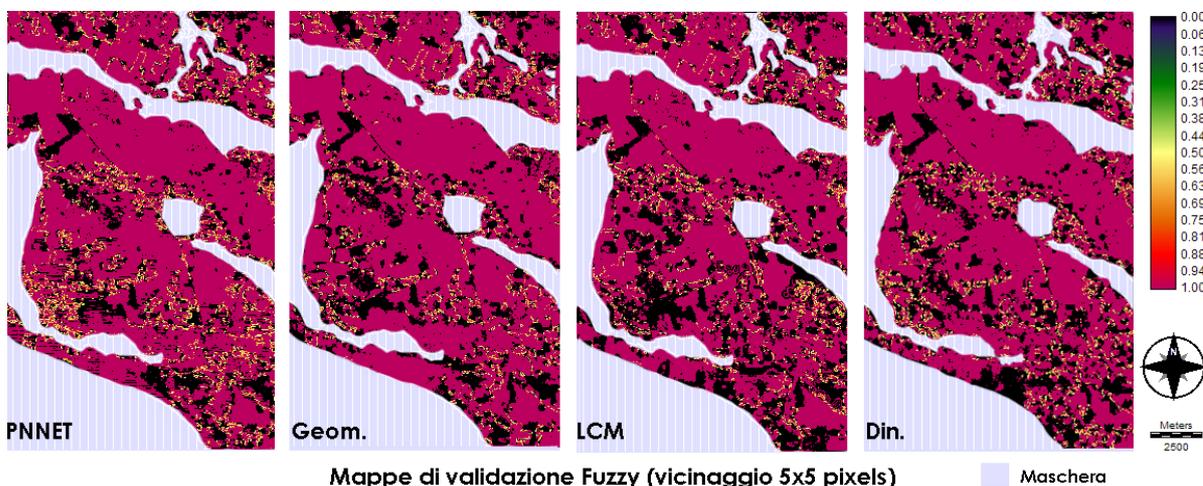


Fig 5-10: Validazione fuzzy ottenuta con la funzione *Reciprocal Similarity Map* di Dinamica EGO. A differenza della validazione hard (corretto/errore), questo approccio considera la prossimità spaziale della congruenza all'interno di una finestra di vicinaggio. Il valore 0 (nero) indica errore, 1 la corrispondenza pixel-a-pixel. Valori intermedi indicano pixel corretti ma spostati di poco rispetto alla reale posizione.

Cell-by-cell vs. Fuzzy validation

	Cell-by-cell Error (Pixel)	Fuzzy Error (Pixel)	Improvement (Pixel)	Improvement (%)
PNNET	119787	84083	35704	7.96
Geomatic	117043	93482	23561	5.25
LCM	139131	113178	25953	5.79
Dinamica	137144	109768	27376	6.10

Tab 5-5: Confronto fra validazione hard (pixel-a-pixel) e validazione fuzzy.

Notiamo una ovvia diminuzione delle celle classificate come errate, con un incremento medio della precisione superiore al 6%. Il modello CA_Markov mostra il minore margine di miglioramento della precisione (5,25%), probabilmente dovuto all'uso di un filtro di contiguità, nell'ultima fase di MOLA, che abbassa l'idoneità dei pixels isolati e crea delle parcelle compatte. PNNET al contrario, forma delle zone deforestate estremamente frammentate, dovute all'approssimazione iniziale dei pixels-frontiera; la

validazione fuzzy permette di aumentare l'accuratezza di simulazione dal 73% (hard) a circa 81%, con un incremento del 7,96%. Questi dati indicano che il modello ha ancora ampi margini di miglioramento, sia a livello quantitativo che spaziale, legati all'abbandono delle approssimazioni utilizzate per economizzare memoria e tempo di calcolo.

La validazione fuzzy riflette meglio la congruenza dei patterns spaziali dei processi fra carta reale e simulata, come emerge dall'analisi visuale [Godoy et al., 2008; Follador et al., 2008]. La scelta di questo metodo e la sua parametrizzazione devono essere sempre fatti in funzione degli obiettivi e della dimensione del supporto utilizzato. Nel nostro caso non interessa conoscere con precisione se il pixel deforestato (400 m²) è ben localizzato o spostato di alcune decine di metri; l'obiettivo è individuare le zone più danneggiate, evidenziate da blocchi di pixels contigui.

5.4 Conclusioni

È molto difficile comparare le performances di differenti modelli predittivi, perché troppi sono gli aspetti della modellizzazione da tenere in considerazione. Quest'ultima è spesso applicata su aree test con dinamiche chiare e rappresentate da una grande quantità di informazioni. I modelli possono dunque avere performance molto buone in queste condizioni, ma risultare totalmente inadatti a lavorare in contesti differenti. Un'ulteriore complicazione riguarda la qualità e la rappresentatività dei dati storici utilizzati, spesso incapaci di catturare la complessità antropica e spaziale del sistema studiato, o semplicemente indicativi di processi-guida che cambiano fra la fase di calibrazione ed il periodo simulato.

In questo capitolo abbiamo confrontato i risultati di quattro modelli predittivi alimentati con lo stesso input ma caratterizzati da un fondamento teorico e strutturale

differente. Abbiamo sottolineato come tutti i metodi presentino una scarsa capacità di riprodurre l'evoluzione dei patterns spaziali dei processi di deforestazione, mentre la previsione delle zone statiche è molto buona. Questi limiti sono dovuti sia all'evoluzione dei processi-guida osservati fra il periodo 1998-2000 ed il periodo 2000-2003, sia a caratteristiche intrinseche alla modellizzazione. È dunque impossibile distinguere esattamente le debolezze del modello dalla complessità delle dinamiche del paesaggio e dalla qualità dei dati.

Le mappe predittive prodotte per il 2003, sono state calcolate rispetto allo stesso scenario storico, che registra le condizioni demografiche e socio-economiche osservate nel periodo di allenamento 1998-2000.

L'analisi quantitativa fra reale e simulato ha evidenziato una sottostima sia per PNNET (77%), sia per i metodi che utilizzano l'analisi prognostica di Markov (82%). La validazione pixel-a-pixel mostra dei risultati molto simili, diminuendo la percentuale di previsione per le parcelle deforestate a circa 55% rispetto alla realtà. PNNET ha la miglior capacità di rappresentare le zone forestali, mentre il modello geomatico CA_Markov offre una previsione migliore per la milpa (58%).

Il calcolo del LUCC budget dal 2000 al 2003 mostra i limiti di PNNET, il cui output non prevede nessuna riforestazione in questo periodo. Il suo swap è dunque nullo. Il modello Dinamica EGO approssima in maniera abbastanza buona le dinamiche osservate dal 2000 al 2003, con una ratio (swap/cambio-netto) di 0,46 prossima al valore reale 0,5.

L'analisi delle performances dei modelli in relazione alla frequenza di cambio registrata in un pixel, ha fornito dei valori simili per CA_Markov (23,85%) e Dinamica EGO (22%), che meglio si adattano alle aree più dinamiche, mentre PNNET prevede con maggior precisione la persistenza (93,74%).

CONCLUSION

“L'avenir dépend de notre clairvoyance et de choix tant politiques qu'individuels. Le meilleur moyen de prédire l'avenir est de l'inventer.”(Eric Lambin, 2004)

c.I. Conclusioni generali

Il ritmo, l'estensione e la portata dell'impatto antropico sulla Terra sono senza precedenti e per questo spesso sottovalutati. La trasformazione umana degli ecosistemi e del paesaggio terrestre sono la fonte principale di cambiamento del nostro Pianeta, ed influenzano la sua capacità di sostenere la vita. L'intensificazione e la diversificazione dell'uso del suolo, l'evoluzione tecnologica, la crescita demografica e la richiesta vertiginosa di prodotti alimentari, il consumismo occidentale che cammina mano nella mano con lo sviluppo economico dei paesi emergenti, la scarsità di terre produttive ed un'ineguale distribuzione delle risorse naturali, hanno causato rapidi squilibri del ciclo biogeochimico, idrologico e delle dinamiche del paesaggio in natura. Un'analisi approfondita dei cambiamenti globali passa attraverso una rappresentazione semplificata del sistema uomo-ambiente, in grado di distillare le principali relazioni fra i processi di uso del suolo ed il contesto geografico, politico, sociale ed economico in cui essi si realizzano. Una profonda comprensione di come l'attività antropica influenzi le dinamiche della biosfera terrestre è necessaria per valutare i rischi che concernono una società ed il suo ambiente, e per pianificare uno sviluppo alternativo che permetta di ridurre tale rischio. L'acronimo LUCC (Land Use Cover Change) è qui utilizzato per indicare sia i cambiamenti fisici del paesaggio sia la forma di gestione politico-economica che lo caratterizza (Capitolo 1).

In questa tesi abbiamo sottolineato l'importanza dei nuovi sistemi d'informazione, quali il telerilevamento, i SIG ed i modelli predittivi, per lo studio dei LUCC. In particolare la nostra ricerca si è concentrata in un'area di foresta tropicale, detta la Joyanca, scelta all'interno della Biosfera

Maya in Peten, Guatemala, che ha conosciuto, negli ultimi decenni, una rapida colonizzazione e degradazione (Capitolo 4.1). Questa regione non è descritta da attributi numerici o cartografici dettagliati; le poche informazioni con cui abbiamo iniziato il nostro lavoro provengono da precedenti pubblicazioni in campo sociologico e fitogeografico.

Presentiamo qui una metodologia di analisi e previsione dei LUCC, applicabile in condizioni non ottimali, per offrire un'alternativa alla modellizzazione tradizionale che troppo spesso si basa su una grande quantità di informazioni per descrivere dinamiche semplici, lente ed omogenee. I modelli, in questo contesto operativo, offrono ottime performance, ma risultano inefficienti, e dunque poco pratici, se applicati in aree con informazioni limitate o con dinamiche paesaggistiche più complesse. Abbiamo deciso di verificare i risultati ottenuti nella rappresentazione di un sistema estremamente frammentato, ed a volte caotico, con l'ausilio dei soli dati ricavati dal processamento di immagini multispettrali. Questi serviranno successivamente ad alimentare i modelli sviluppati e riparametrizzati per descrivere l'evoluzione dei patterns spaziali dei processi di LUCC in Peten.

La possibilità di seguire l'evoluzione di un fenomeno attraverso una vasta scala di risoluzioni spazio-temporali, rende l'utilizzo dei dati multispettrali indispensabile per una completa analisi dello stato di salute del nostro Pianeta. L'informazione satellitare permette di lavorare in aree con condizioni ambientali estreme o di difficile accesso, con grande risparmio di tempo e denaro. Le immagini TM Landsat (P20-R48) e SPOT (P606, P607-R315) raccolte dal 1988 al 2003, sono state "calibrate" ed interpretate con l'aiuto di esperti della regione e delle problematiche studiate. In questa maniera siamo riusciti a sopperire all'impossibilità di

realizzare, all'interno del nostro progetto di ricerca, un lavoro *in situ*, per ovvie ragioni economiche e temporali.

Il processamento delle immagini (Capitolo 4.2) ha permesso di migliorare la qualità dei dati originari ed eliminare eventuali ridondanze. La creazione di un nuovo indicatore NDMI, derivato dalla fusione degli indici vegetali NDVI e di stress idrico NDII, ed il suo utilizzo nella tecnica di analisi multitemporale RGB, ne ha evidenziato l'utilità e la praticità nello studio delle dinamiche di LUCC durante il periodo esaminato. Le carte dei cambi avvenuti fra il 1988 ed il 2003, ci hanno aiutato a visualizzare i principali processi di uso del suolo e la loro relazione con le variabili indipendenti scelte. Questa fase rappresenta la parte più delicata di tutta l'opera di modellizzazione; comprendere bene il fenomeno studiato, rappresentarlo attraverso i formalismi più idonei a raggiungere i propri obiettivi, evidenziare gli agenti, capirne il comportamento e le motivazioni, separare i fattori che maggiormente influenzano i LUCC, integrare dati raccolti su differenti scale spazio-temporali, è un lavoro oneroso che richiede un dialogo costante fra differenti discipline. Il tempo perso in questa prima fase, sarà poi facilmente recuperato durante la simulazione e nell'accettazione generale dei risultati da parte del mondo scientifico.

Attraverso le funzioni di analisi spaziale presenti in differenti SIG, abbiamo estrapolato dalle immagini, le matrici rappresentative delle variabili indipendenti selezionate e creato un database omogeneo (Capitolo 4.2.4). Alcune di queste carte, come la mappa di produttività agricola, sono state costruite sulla base delle pubblicazioni fitogeografiche esistenti e delle opinioni di esperti, cercando così di integrare informazione numerica digitale ed informazione

socio-culturale (socializzazione del pixel o pixelizzazione del sociale?).

Modellizzare un sistema complesso è un obiettivo difficile e spesso utopico, ed ogni sua descrizione fissa risulta incompleta ed instabile. I limiti del processo modellizzativo risiedono nelle numerose semplificazioni adottate per rappresentare il sistema, le sue dinamiche e le sue proprietà emergenti su varie scale spazio-temporali. L'estensione di queste ultime non può essere catturata attraverso delle aggregazioni o delle riduzioni, senza che una parte della loro ricchezza informativa non vada persa. I dati su cui si basa la modellizzazione vengono raccolti ed utilizzati in funzione di un'idea semplicistica del sistema, che è fortemente influenzata dalle nostre conoscenze ed dalla nostra esperienza. Il risultato sarà, per conseguenza, impregnato di una soggettività che ne limita la significatività rappresentativa.

In questa tesi abbiamo voluto dimostrare l'importanza della funzione pedagogica dei modelli predittivi per sopperire ai limiti del nostro sapere cognitivo, che tende a semplificare ciò che in realtà semplice non è. Un modello permette innanzitutto di esplorare il sistema studiato, la validità delle ipotesi e delle scelte effettuate, e ci consente di evidenziare comportamenti e rapporti causali prima ignorati o mal interpretati. In questa fase, che chiamiamo di apprendimento, si stabiliscono progressivamente, sulla base dei risultati delle prime simulazioni, le relazioni fra le variabili indipendenti, i processi studiati e le caratteristiche evolutive di questi ultimi.

Partendo da queste considerazioni abbiamo sviluppato un modello predittivo, PNNET, basato sulle reti neurali, che

hanno in passato dimostrato una grande capacità di catturare e rappresentare le relazioni nascoste, spesso non lineari, di un sistema complesso. Il modello è implementato in R[®] e liberamente scaricabile da internet assieme ad un piccolo tutoriale. L'utilizzo di un approccio automatico dovrebbe inoltre metterci al riparo dalle critiche sulla soggettività del risultato. Le performance di PNNET sono state confrontate con quelle di altri 3 modelli in commercio, scelti per il loro differente approccio teorico e per i costi limitati. I modelli CA_Markov e Land Cover Modeler sono due moduli di Idrisi; il nostro lavoro si è dunque limitato ad una loro riparametrizzazione con i dati a disposizione. La costruzione del modello in Dinamica EGO, un ambiente freeware di simulazione, ha portato alla creazione di una sequenza di 4 programmi facilmente adattabili a differenti problematiche e contesti geografici. Questo modello è fornito come allegato a questa tesi, che può essere utilizzata come manuale d'uso (Capitolo 3).

I differenti metodi predittivi sono stati alimentati con lo stesso input, riferito al periodo 1998-2000, per poterne verificare limiti e pregi in condizioni operative uguali. La fase di validazione ha confrontato i risultati di ogni singolo modello con l'immagine reale del 2003, quantificando le incongruenze (Capitolo 5).

Tutte le mappe predittive hanno sottostimato la reale portata e magnitudine dei cambi realizzatesi fra il 2000 ed il 2003. I processi-pilota registrano un'evoluzione fra il periodo di calibrazione ed il periodo simulato, determinando un'errata allocazione spaziale dei cambi; è dunque difficile separare i limiti del modello dalla complessità del fenomeno o dalla rappresentatività dei dati.

PNNET ha mostrato un'ottima capacità di prevedere la persistenza delle grandi aree forestali, mentre la performance si è notevolmente ridotta nelle zone soggette a cambi frequenti. Le parcelle deforestate presentano un'irreale frammentazione, con un'alternanza di linee di pixel deforestati e non, dovuta alle approssimazioni iniziali introdotte per risparmiare memoria e tempo di calcolo. Questa sorta di campionamento basato sui pixel frontiera, non è rappresentativo della reale distribuzione delle categorie di uso del suolo. L'abbandono di questa approssimazione e l'utilizzo di tutta l'informazione disponibile, dovrebbe garantire un ampio margine di miglioramento delle performance di PNNET.

I metodi CA_Markov e Dinamica EGO hanno mostrato una migliore capacità di visualizzare il cambiamento ed i processi di deforestazione. Questo risultato è probabilmente dovuto all'approccio metodologico utilizzato, in cui la supervisione del modellizzatore e la sua esperienza giocano un ruolo fondamentale nella rappresentazione della complessità del sistema.

Tutti i modelli dimostrano una buona precisione nel descrivere la persistenza, mentre le performance calano bruscamente nelle aree più dinamiche.

Riassumendo possiamo dire che questa tesi ha cercato di portare un contributo alla teoria di modellizzazione, uscendo dai binari tradizionali che ne limitano l'applicabilità in condizioni ottimali. Abbiamo provato a rispondere alle domande: "è possibile modellizzare un sistema complesso, frammentato e caotico, in assenza di informazioni dettagliate? Come sopperire a questa mancanza? Come costruire una metodologia che offra dei risultati utili senza un grande

dispendio di soldi e tempo, utilizzando invece informazioni facilmente disponibili? Qual è la significatività di questi risultati e come valutare l'affidabilità di un modello?"

Abbiamo visto come sia difficile ed a volte impossibile catturare la complessità spaziale, temporale ed antropica dei processi in atto nella regione della Joyanca. I modelli non rappresentano una "answer machine" in grado di dare una risposta certa a qualsiasi questione. Sono semplicemente uno strumento per testare delle ipotesi e per meglio comprendere il funzionamento del sistema uomo-ambiente. Se correttamente parametrizzato ed alimentato, un metodo predittivo è sicuramente un aiuto indispensabile alla pianificazione politica, economica ed ambientale di una regione. Un modello di LUCC può essere utilizzato per valutare gli effetti sul breve periodo di una scelta amministrativa o di un piano di sviluppo, prevenendo eventuali effetti indesiderati e rischi per la società.

c.2. Conclusioni "locali"

Nel caso della Joyanca, le mappe predittive che rappresentano l'evoluzione dei patterns spaziali di uso del suolo, hanno evidenziato le carenze ed i limiti dell'attuale forma di gestione della Biosfera Maya. Le aree protette sono da tempo soggette a fenomeni di deforestazione destinati ad intensificarsi con l'aumento della popolazione nella regione. La volontà di preservare il bosco tropicale dimostrata dal governo con la creazione della RBM, si scontra con le politiche che incentivano la migrazione dei ceti più poveri e dei senza terra in queste aree di frontiera, con lo scopo di risolvere la questione sociale, di colonizzare un'area ancora poco abitata ed attirare nuovi capitali stranieri.

Sulla base dei dati raccolti fino al 2003, abbiamo creato una mappa predittiva del mese di giugno 2008, utilizzando per l'apprendimento le mappe classificate del 1998 (inizio) e 2003 (fine). La simulazione (Fig.c-1) mostra una progressiva coesione delle parcelle deforestate a sud. Le zone centrali, all'interno dei confini comunitari, riescono solo parzialmente ad arginare l'avanzata di questo fronte pioniere. A nord del Rio San Pedro, zona nucleo di parco naturale protetto, si registrano nuovi fenomeni di deforestazione, evidenziando l'inefficacia delle misure preventive adottate.

Uso del suolo nel giugno 2008 sulla base dei dati del 1998 e 2003

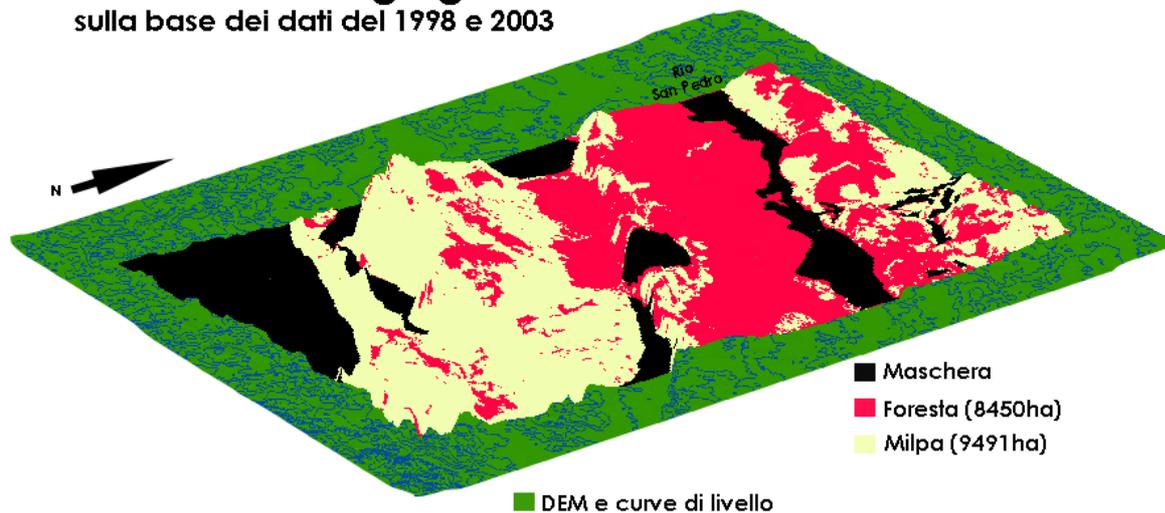


Fig c-1: Mappa predittiva di uso del suolo nel giugno 2008. Nella simulazione sono stati utilizzate le mappe classificate del 1998 e del 2003 (periodo 5 anni) e le variabili raccolte fino al 2003. Immagine sovrapposta ad un DEM ricavato da informazioni Shuttle Radar (esagerazione verticale=2).

In questa simulazione non sono state integrate nuove informazioni (socio-economiche, demografiche, politico-amministrative) riferite al periodo 2003-2008, presupponendo uno scenario evolutivo uguale a quello registrato nei dati storici. La mappa visualizza dunque la situazione che avremmo se le condizioni osservate fra il 1998 ed il 2003 dovessero persistere. Esiste comunque la probabilità che i processi-pilota, le variabili e la pianificazione territoriale del Peten, abbiano subito delle modifiche e siano cambiati durante

gli ultimi 5 anni, riducendo conseguentemente la significatività rappresentativa della predizione.

c.3. Limiti e prospettive future

I limiti di questa tesi si confondono spesso con i limiti della teoria modellistica.

Lo studio dei LUCC richiede una profonda comprensione dei drivers biofisici e socio-economici, nonché del contesto ambientale e geografico in cui essi operano. Ciò domanda una collaborazione stretta fra differenti discipline ed una capacità di rimettere in gioco le proprie convinzioni e metodologie, presupposti spesso assenti in ambito accademico e lavorativo.

Lo studio dei LUCC si fonda su un'analisi che passa attraverso molteplici risoluzioni spaziali per evidenziare comportamenti locali ed effetti emergenti. L'integrazione di dati raccolti a differenti scale e *l'upscaling* da micro a macro, sono stati risolti attraverso l'utilizzo dei SIG e dei metodi spazialmente espliciti, ma ancora oggi rappresentano un tema di discussione e di critica all'interno della teoria modellistica. *Come passare da un'informazione puntuale, come la rendita di un contadino, ad un'informazione estesa su tutta un'unità amministrativa, come le regole di mercato e le politiche agricole nazionali? Ed ancora, come integrare dati privi di dimensione spaziale, come i fattori culturali e religiosi, in un'analisi spazialmente esplicita, senza cadere in una banalizzazione e semplificazione eccessiva dell'informazione? Come utilizzare dati che propongono visioni della realtà da prospettive apparentemente opposte (economia ed ecologia, ad es.)?* Una chiara risposta a queste domande non è ancora stata fornita ed i tentativi realizzati fanno spesso

riferimento ad un dominio o ad una problematica particolare, e dunque non sono generalizzabili.

La sfida maggiore è riuscire a catturare e modellizzare la complessità dei sistemi antropici o antropizzati, dove l'imprevedibilità del libero arbitrio dà origine ad un caos "apparentemente sorprendente". Il programma mondiale Global Land Project (GLP), creato per studiare le dinamiche dei sistemi terrestri ed i loro effetti sul Pianeta, ha evidenziato la necessità di utilizzare un modello multi-agente (ABM/LUCC) per includere i processi decisionali all'interno della ricerca sui cambiamenti globali. L'ABM/LUCC ha dimostrato grandi potenzialità sia come laboratorio sperimentale, per esplorare in maniera deduttiva i legami fra le micro-interazioni ed i macro-effetti, sia per rappresentare la coevoluzione del sistema uomo-ambiente. La qualità più importante dell'approccio multi-agente sembra essere la sua capacità di rappresentare le proprietà emergenti di un sistema complesso, partendo dall'analisi di ogni agente e ogni variabile attraverso un ampio spettro di risoluzioni spaziali.

Questa tesi ha cercato di creare dei ponti fra differenti discipline, fra cui l'informatica, la statistica, la geografia, la sociologia, l'ingegneria, le scienze forestali e la botanica, per ottenere una rappresentazione più veritiera possibile dei LUCC in Peten. Probabilmente l'approccio metodologico utilizzato e le semplificazioni adottate, non hanno permesso di catturare tutta la ricchezza informativa di ogni disciplina. Lo studio dell'evoluzione dei patterns spaziali nella Joyanca ha avuto più una funzione esploratoria e di apprendimento, piuttosto che di previsione e pianificazione futura. L'analisi si è concentrata sulle dinamiche a livello locale, mentre manca un'integrazione (*upscaling*) con le informazioni a macro-scala (piani di

sviluppo economico, ridistribuzione delle terre, etc.), che permetta una migliore comprensione dei LUCC e dei loro effetti emergenti in tutta la regione.

Una piccola critica deve essere fatta anche a riguardo dei dati multispettrali utilizzati. Le immagini TM Landsat e SPOT sono state acquisite dal GEODE/CNRS, prima dell'inizio di questa ricerca, seguendo un criterio di economicità e di buone condizioni meteo. Poca attenzione è stata dedicata all'intervallo temporale fra una scena e la successiva; i dati si riferiscono a volte a periodi dell'anno ed a stagioni differenti, il che rende impossibile eliminare eventuali errori di omissione o commissione, legati al ciclo fenologico e alla variabilità scena-a-scena. La scelta delle immagini satellitari e l'assenza di un intervallo regolare fra una data e la successiva, ha determinato anche la differente distanza temporale utilizzata fra calibrazione e simulazione, diminuendo probabilmente le performance dei modelli.

La validazione ha cercato di confrontare mappa reale e mappa simulata, evidenziando differenti tipi di incongruenze (quantità, dinamiche, frequenza dei cambi, etc.). Un ulteriore sforzo, a mio avviso, dovrebbe essere fatto per studiare la forma delle parcelle deforestate e la frammentarietà dei processi in atto. Quest'ultima, infatti, determina una rapida degradazione dell'ecosistema tropicale ed un'elevata mortalità della vegetazione. Alcune mappe (CA_Markov, ad es.) offrono buoni risultati durante la fase di validazione ma rappresentano parcelle deforestate troppo compatte e coese. L'analisi della frammentarietà del paesaggio è una tappa necessaria per studiare il reale impatto dell'attività antropica sulla foresta tropicale ed è un'informazione necessaria per tutti gli studi che riguardano la produzione di gas ad effetto serra lungo i fronti di deforestazione.

Ignorare questa caratteristica può portare ad una sovrastima del grado di affidabilità del modello e limitare la significatività dei risultati ottenuti.

*"...il ne faut pas penser que le but est d'offrir le modèle le plus beau... on ne fait pas de la modélisation pour la modélisation mais pour comprendre le fonctionnement du Monde et comprendre les impacts de certaines perturbations sur le Monde"*¹ [Ramat, 2006, Pag.50].

*"des modèles des phénomènes perçus ou conçus dans le temps et dans l'espace, nous disons volontiers qu'ils sont des représentations... représenter c'est aussi re-présenter, présenter à nouveau, après qu'un temps soit écoulé, en acceptant donc que la re-présentation ne soit pas nécessairement un duplicata, une copie conforme. Re-présenter, c'est proposer une représentation qui porte en elle-même sa légitimation : elle a mémoire et projet, et elle fonde sa légitimité sur sa cohérence avec son histoire et ses devenirs"*² [La Moigne, 1994, Pag.266].

1 Non dobbiamo pensare che lo scopo sia offrire il modello più "bello"... non si fa della modellizzazione per il semplice gusto di modellizzare ma per comprendere il funzionamento del Mondo e capire l'impatto che certe perturbazioni possono avere sul mondo.

2 ...diremmo che i modelli percepiti o concepiti nello spazio e nel tempo sono delle rappresentazioni...rappresentare è anche ri-presentare, presentare di nuovo dopo un certo periodo, accettando dunque che non sia esattamente un duplicato, una copia perfetta. Ri-presentare è proporre una rappresentazione che ha una legittimità in se: ha memoria ed un fondamento teorico, basa la sua legittimità sulla coerenza fra dati storici e divenire.

CONCLUSIONS

“L'avenir dépend de notre clairvoyance et de choix tant politiques qu'individuels. Le meilleur moyen de prédire l'avenir est de l'inventer.”(Eric Lambin, 2004)

c.I. Conclusions Générales

Le rythme, l'étendue et la portée de l'impact anthropique sur la Terre sont sans précédents et, pour cette raison, souvent sous-estimés. La transformation humaine des écosystèmes et du paysage est la source principale des changements de la Planète, et influence sa capacité à assurer la pérennité des systèmes vivants. L'intensification et la diversification des formes d'utilisation du sol, l'évolution technologique, la croissance démographique et la demande vertigineuse de produits alimentaires, le modèle de consommation occidental qui accompagne le développement économique des pays émergents, la pénurie de terres productives et une inégale distribution des ressources naturelles, toutes ces circonstances ont produit de rapides déséquilibres des cycles biogéochimique et hydrologique et des dynamiques paysagères. Une analyse approfondie des changements globaux passe à travers une représentation simplifiée du système homme-environnement, de façon à rendre compte des principales relations entre l'utilisation du sol et le contexte géographique, politique, social et économique dans lequel elles se manifestent. Une profonde compréhension des influences anthropiques sur les dynamiques de la biosphère terrestre est nécessaire pour évaluer les risques qui concernent une société donnée et son environnement et pour prendre les mesures destinées à réduire ce risque. L'acronyme LUCC (Land Use & Cover Change) est ici utilisé pour désigner aussi bien les changements physiques du paysage que les modalités de gestion politico-économique auxquelles il est soumis (Chapitre 1).

Dans cette thèse, nous avons souligné l'importance des nouveaux systèmes d'information, comme la télédétection, les SIG et les modèles prédictifs, pour étudier les LUCC. Nous les avons mis en œuvre sur notre terrain de recherche, situé dans

une région de forêt tropicale, la *Joyanca*, au coeur de la Biosphère Maya du Petén, au Guatemala. Cette région a connu, dans les dernières décennies, une rapide colonisation et une dégradation marquée (Chapitre 4.1). Nous n'avons pas disposé d'attributs numériques ou cartographiques détaillés; le peu d'informations à partir desquelles nous avons entamé nos recherches proviennent de publications précédentes issues des champs de recherche sociologique et phytogéographique.

Nous présentons ici une méthodologie d'analyse et de prévision des LUCC, applicable dans des conditions non-optimales. Nous espérons ainsi offrir une alternative aux modélisations traditionnelles, trop souvent basées sur une grande quantité d'information pour décrire des dynamiques simples, de longue durée et homogènes. Ces modèles traditionnels, lorsqu'ils sont appliqués en ces conditions opérationnelles, offrent d'excellentes performances, mais se révèlent inefficients, et donc peu pratiques, dès lors qu'ils sont appliqués sur des terrains déficients en informations ou caractérisés par des dynamiques paysagères plus complexes. Nous avons décidé de vérifier les résultats obtenus dans la représentation d'un système extrêmement fragmenté et parfois chaotique, avec l'unique secours des données obtenues par traitement d'images multispectrales. Celles-ci serviront successivement à alimenter les modèles développés et à les re-paramétrer pour décrire l'évolution des *patterns* spatiaux associés aux processus de LUCC au Petén.

La possibilité de suivre l'évolution d'un phénomène à travers une vaste échelle de résolutions spatio-temporelles, rend l'utilisation des données multispectrales indispensable pour une complète analyse de l'état de santé de notre Planète. L'information satellitaire permet de travailler sur des espaces dotés de conditions environnementales extrêmes ou d'accès difficile, pour un faible coût et en un temps limité. Les images acquises par Landsat TM (P20-R48) et SPOT

(P606, 607-R315) de 1988 à 2003, ont été « calibrées » et interprétées, avec l'aide d'experts de la région et des problématiques étudiés. Nous avons donc pu conduire notre recherche malgré l'impossibilité de réaliser un travail de terrain, pour d'évidentes raisons matérielles.

Le traitement des images (Chapitre 4.2) a permis d'améliorer la qualité des données originelles et d'éliminer d'éventuelles redondances. La création d'un nouvel indicateur NDMI, dérivé de la fusion des index de végétation NDVI et du stress hydrique NDII, et son application dans la technique d'analyse multitemporelle RGB, a mis en évidence son utilité et son caractère pratique dans l'étude des dynamiques de LUCC pendant la période examinée. Les cartes des changements intervenus entre 1988 et 2003 nous ont permis de visualiser les principaux processus d'utilisation du sol et leur relation avec les variables explicatives choisies. Cette phase représente la partie plus délicate de toute la modélisation. En effet, bien comprendre le phénomène étudié, le représenter à travers les formalismes les plus aptes à atteindre les objectifs visés, mettre en évidence les agents, en comprenant le comportement et les motivations, distinguer les facteurs dont l'influence sur les LUCC est prépondérante, intégrer des données recueillies à différentes échelles spatio-temporelles, ceci représente un travail ardu qui demande d'établir un dialogue constant entre différentes disciplines. Le temps perdu pour cette première partie est facilement récupéré ensuite au cours de la simulation et par l'accueil fait aux résultats par le monde scientifique.

À travers les fonctions d'analyse spatiale présentes en différents SIG, nous avons extrapolé à partir des images les matrices représentatives des variables explicatives sélectionnées, et avons créé une base de données homogène. Certaines de ces cartes, comme celle de productivité agricole, ont été construites en se basant sur les publications

phytogéographiques existantes et sur l'opinion des experts, cherchant ainsi à intégrer l'information numérique et l'information socioculturelle (socialisation des pixels ou pixellisation du social?).

Modéliser un système complexe est un objectif difficile et souvent utopique et chacune de ses descriptions fixes apparaît incomplète et instable. Les limites du processus de modélisation résident dans les nombreuses simplifications adoptées pour représenter le système, ses dynamiques et ses propriétés émergentes à diverses échelles spatio-temporelles. L'étendue de ces dernières ne peut pas être saisie à travers des associations ou des réductions, sans risquer de perdre une partie de leur richesse informative. Les données sur lesquelles se fondent les modélisations sont recueillies et utilisées en fonction d'une idée simpliste du système, qui est fortement influencée par nos connaissances et expériences. Le résultat sera, par conséquent, imprégné d'une subjectivité qui en limite la signification représentative.

A travers cette thèse nous avons voulu montrer l'importance de la fonction pédagogique des modèles prédictifs pour dépasser les limites de notre savoir cognitif, qui tend à simplifier ce qui, en réalité, ne l'est pas. Un modèle permet avant tout d'explorer le système étudié, la validité des hypothèses et des choix effectués, et il rend possible la mise en évidence des comportements et des rapports causaux d'abord ignorés ou mal interprétés. Dans cette phase, que nous appelons l'apprentissage, s'établissent progressivement, sur la base des résultats des premières simulations, les relations entre les variables explicatives, les processus étudiés et les caractéristiques évolutives de ces derniers.

A partir de ces considérations, nous avons développé un modèle prédictif, PNNET, basé sur les réseaux de neurones, qui ont déjà démontré leur grande aptitude à saisir et représenter les relations masquées, souvent non-linéaires, d'un système

complexe homme-environnement. Le modèle est implémenté en R[®] et librement téléchargeable sur Internet avec un petit « tutorial ». L'utilisation d'une approche automatique devrait en outre nous mettre à l'abri des critiques portant sur la subjectivité du résultat. Les performances de PNNET ont été confrontées avec celles de trois autres modèles du commerce, choisis pour leurs différentes approches théoriques et pour leurs coûts limités. Les modèles CA_Markov et Land Cover Modeler sont deux modules d'Idrisi ; notre travail s'est donc limité à la re-paramétrisation avec les données disponibles. La construction du modèle en Dinamica EGO, une plateforme freeware de simulation, a conduit à la création d'une séquence de quatre programmes facilement adaptables à différentes problématiques et différents contextes géographiques. Ce modèle est fourni en annexe dans cette thèse, qui peut être utilisée comme manuel (Chapitre 3).

Les différentes méthodes prédictives ont été alimentées avec le même input, relatif à la période 1998-2000, pour en déterminer les limites et les défauts en conditions opérationnelles identiques. La phase de validation a confronté les résultats de chaque modèle avec l'image réelle de 2003, en quantifiant les divergences. (Chapitre 5).

Toutes les cartes prédictives ont sous-évalué la portée réelle et l'importance des changements qui se sont produits entre 2000 et 2003. Les processus-pilotes enregistrent une évolution entre la période de calibrage et la période simulée, en déterminant une allocation spatiale erronée des changements; il est donc difficile de distinguer ce qui est dû aux limites du modèle de ce qui est dû à la complexité du phénomène ou à la représentativité des données.

Le PNNET a montré une excellente capacité de prévoir la persistance des grandes surfaces forestières, cependant cette performance s'est considérablement réduite dans les zones sujettes à des changements fréquents. Les parcelles déboisées

présentent une fragmentation non réaliste, avec une alternance de lignes de pixels déforestés et non-déforestés, due aux approximations initiales introduites pour limiter la place mémoire et le temps de calcul. Cette sorte d'échantillonnage basée sur le pixel frontalier, n'est pas représentative de la distribution réelle des catégories d'utilisation du sol. L'abandon de ces approximations et l'emploi de toute l'information disponible, devraient garantir une bonne marge d'amélioration des performances du PNNET.

Les méthodes CA_Markov et Dinamica EGO ont montré une meilleure capacité à visualiser les changements et les processus de déforestation. Ce résultat est probablement dû à l'approche méthodologique utilisée, dans laquelle la supervision du modélisateur et son expérience jouent un rôle fondamental dans les représentations de la complexité du système.

Tous les modèles montrent une bonne précision dans la description de la persistance (stabilité), cependant les performances baissent brusquement dans les aires plus dynamiques.

En résumé, notre thèse a tenté d'apporter une contribution à la théorie de la modélisation, en sortant des voies traditionnelles qui en limitent l'applicabilité à des situations en conditions optimales. Nous avons répondu aux questions suivantes : *« Est-il possible de modéliser un système complexe, fragmenté et chaotique, en absence d'informations détaillées ? Comment dépasser cette contrainte ? Comment construire une méthodologie qui débouche sur des résultats utiles mais qui, grâce à l'utilisation d'informations facilement disponibles, reste économe en temps et en argent ? Quelle est la signification de ces résultats et comment évaluer la fiabilité d'un modèle ? »*

Nous avons constaté la difficulté et parfois l'impossibilité de saisir la complexité spatiale, temporelle et anthropique des processus en vigueur dans la région de la *Joyanca*. Les modèles ne représentent pas une « *answer machine* » en mesure de donner une réponse correcte à n'importe quelle question. Ils sont simplement un moyen de tester une hypothèse et de mieux comprendre le fonctionnement du système homme-environnement. Correctement paramétrée et alimentée, une méthode prédictive peut certainement apporter une aide déterminante à la planification politique, économique et environnementale d'une région. Un modèle de LUCC peut être utilisé pour évaluer les effets à brève échéance d'un choix administratif ou d'un plan de développement, en prévenant les éventuels risques et effets non désirés pour la société.

c.2. Conclusions "locales"

Dans le cas de la *Joyanca*, les cartes prédictives, qui représentent l'évolution des patterns spatiaux d'utilisation du sol, ont mis en évidence les carences et les limites de mode de gestion actuel de la Biosphère Maya. Les aires protégées sont depuis longtemps sujettes à des phénomènes de déforestation destinés à s'intensifier avec l'augmentation de la population dans la région. La volonté de préserver la forêt tropicale, manifestée par le gouvernement avec la création de la RMB, entre en contradiction avec les politiques qui stimulent la migration des catégories les plus pauvres et des « *sans terre* », dans ces zones de frontière, afin de résoudre la question sociale, de coloniser une région encore peu habitée et d'attirer de nouveaux capitaux étrangers.

En nous basant sur les données recueillies jusqu'en 2003, nous avons établi une carte prédictive pour le mois de juin 2008, en utilisant pour l'apprentissage les cartes classifiées

de 1998 (début) et 2003 (fin). La simulation (Fig.c-1) montre une cohésion progressive des parcelles déforestées au sud. Les zones centrales, à l'intérieur des frontières communautaires, ne réussissent que partiellement à endiguer l'avancée de ce front pionnier. Au nord du fleuve San Pedro, zone-noyau du parc naturel protégé, on enregistre de nouveaux phénomènes de déforestation, signe de l'inefficacité des mesures préventives adoptées.

Uso del suelo nel giugno 2008 sulla base dei dati del 1998 e 2003

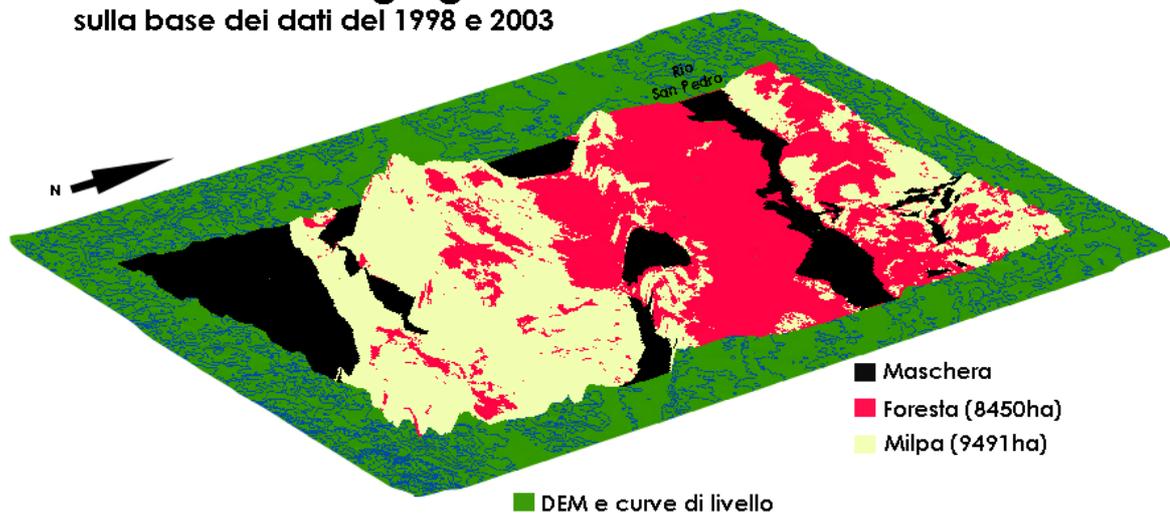


Fig c-1: Carte prédictive d'utilisation du sol en Juin 2008. Dans la simulation ont été utilisées les cartes classifiées de 1998 et de 2003 (période 5 ans) et les variables récoltées jusqu'à 2003. Image drapée sur un MED construit d'après les informations du Shuttle Radar (exagération verticale=2).

Dans cette simulation, les nouvelles informations (socio-économiques, démographiques, politiques) relatives à la période 2003-2008 n'ont pas été assimilées : on a présumé un scénario identique à celui enregistré par les données historiques. La carte visualise donc la situation qui prévalerait si les conditions observées entre 1998 et 2003 devaient persister. Il existe cependant une certaine probabilité que les processus-pilotes, les variables et la planification territoriale du Petén, aient subi des modifications et se soient développés pendant les cinq

dernières années, réduisant ainsi la précision des prédictions.

c.3. Limites et perspectives futures

Les limites de cette thèse se confondent souvent avec les limites propres à la théorie de la modélisation.

L'étude des LUCC demande une profonde compréhension des *drivers* biophysiques et socio-économiques, ainsi que du contexte environnemental et géographique dans lequel ils agissent. Cet objectif nécessite une collaboration étroite entre différentes disciplines et une capacité à remettre en jeu ses convictions et habitudes de pensée, conditions souvent absentes aussi bien dans le domaine académique que professionnel.

L'analyse des LUCC s'effectue selon de multiples résolutions spatiales pour mettre en évidence des comportements locaux et des effets émergents. L'intégration de données recueillies à différentes échelles et le passage du local au global (*upscaling*), ont été résolus par l'utilisation des SIG et des méthodes spatialement explicites, mais aujourd'hui encore ils représentent un sujet de controverse à l'intérieur de la théorie de la modélisation. *Comment passer d'une information ponctuelle, comme le revenu d'un paysan, à une information plus étendue sur toute une unité administrative, comme les règles de marché et les politiques agricoles nationales ? De même, comment articuler des données dépourvues de dimension spatiale, comme les facteurs culturels et religieux, à des analyses spatialement explicites, sans tomber dans une banalisation et une simplification excessives de l'information ? Comment utiliser des données qui proposent des visions de la réalité selon des perspectives apparemment opposées (économie et écologie, par exemple.) ?* Une réponse précise à ces questions n'a pas encore été fournie, les essais

déjà réalisés font souvent référence à un domaine ou à une problématique spécifique, et ils ne sont donc pas généralisables.

Le défi majeur est de réussir à saisir et modéliser la complexité des systèmes anthropiques ou anthropisés, où l'imprévisibilité du libre arbitre conduit à un chaos "apparemment surprenant". Le programme mondial Global Land Project¹ (GLP), créé pour étudier les dynamiques des systèmes terrestres et leurs effets sur la Planète, a mis en évidence la nécessité d'utiliser un modèle multi-agents (ABM/LUCC) pour inclure les processus décisionnels au cœur de la recherche sur les changements globaux. L'ABM/LUCC a montré de grandes potentialités soit comme laboratoire expérimental, pour explorer de façon déductive les liens entre les micro-interactions et les macro-effets, soit pour représenter les co-évolutions du système homme-environnement.

Cette thèse a tenté d'établir un pont entre différentes disciplines, comme l'informatique, la statistique, la géographie, la sociologie, l'ingénierie, les sciences forestières et la botanique, pour obtenir la représentation la plus véridique possible des LUCC du Petén. Probablement l'approche méthodologique utilisée et les simplifications adoptées, n'ont pas permis de capturer toute la richesse informative de chaque discipline. L'étude de l'évolution des patterns spatiaux dans *la Joyanca* a eu plutôt une fonction d'exploration et d'apprentissage, que de prévision et de planification. L'analyse s'étant focalisée sur les dynamiques de niveau local, il a donc manqué une intégration (*upscaling*) avec les informations à macro-échelle (des plans de développement économique, incitations fiscales, etc.), qui

¹ Le GLP est co-financé par le IGBP (International Geosphere-Biosphere Programme) et par le IHDP (International Human Dimension Programme on global environmental change).

autorise une meilleure compréhension des LUCC et de leurs effets émergents dans toute la région.

Une observation doit également être faite à l'égard des données multispectrales. Les images TM Landsat et SPOT ont été acquises par le laboratoire GEODE/CNRS avant le début de nos recherches, conformément aux critères d'économie et de bonnes conditions météorologiques. Peu d'attention a ainsi été dédiée à l'intervalle temporel entre une scène et la suivante : les données se réfèrent parfois à différentes périodes et saisons de l'année, ce qui rend impossible l'élimination d'éventuelles erreurs d'omission ou de commission, liées au cycle phénologique et à la variabilité inter-scènes. Le choix des images satellitaires et l'absence d'un intervalle régulier entre une donnée et la suivante, ont aussi déterminé l'écart temporel différent utilisé entre le calibrage et la simulation, en diminuant probablement les performances des modèles.

La validation a consisté à confronter la carte réelle et la carte simulée, en mettant en évidence les différents types de divergences (quantité, dynamiques, fréquence des changements, etc.). Un effort ultérieur, à notre avis, devrait être fait pour étudier la forme des parcelles déforestées et la fragmentarité des processus en cours. Cette dernière caractéristique, en effet, détermine une rapide dégradation de l'écosystème tropical et une mortalité élevée des espèces constituant la végétation. Quelques cartes (CA_Markov, par exemple) offrent de bons résultats pendant la phase de validation mais elles représentent des parcelles déforestées trop compactes et de petite taille. L'analyse de la fragmentarité du paysage est une étape nécessaire pour étudier l'impact réel de l'activité anthropique sur la forêt tropicale. Cette information est indispensable aux études qui concernent la production de gaz à effet serre, au long des fronts de déforestation. Ignorer cette caractéristique peut

conduire à une sous-estimation du degré de fiabilité du modèle et limiter la signification des résultats obtenus.

"...il ne faut pas penser que le but est d'offrir le modèle le plus beau... on ne fait pas de la modélisation pour la modélisation mais pour comprendre le fonctionnement du Monde et comprendre les impacts de certaines perturbations sur le Monde" [Ramat, 2006, Pag.50].

"des modèles des phénomènes perçus ou conçus dans le temps et dans l'espace, nous disons volontiers qu'ils sont des représentations... représenter c'est aussi re-présenter, présenter à nouveau, après qu'un temps soit écoulé, en acceptant donc que la re-présentation ne soit pas nécessairement un duplicata, une copie conforme. Re-présenter, c'est proposer une représentation qui porte en elle-même sa légitimation : elle a mémoire et projet, et elle fonde sa légitimité sur sa cohérence avec son histoire et ses devenirs" [Le Moigne, 1994, Pag.266].

In Musica...

Sono ancora aperte come un tempo le osterie di fuori porta,
 ma la gente che ci andava a bere fuori o dentro è tutta morta:
 qualcuno è andato per età, qualcuno perchè già DOTTORE
 e insegue una maturità, si è sposato, fa carriera ed è una morte un
 po' peggiore...

Cadon come foglie o gli ubriachi sulle strade che hanno scelto,
 delle rabbie antiche non rimane che una frase o qualche gesto,
 non so se scusano il passato per giovinezza o per errore,
 non so se ancora desto in loro, se m' incontrano per forza, la
 curiosità o il timore...

Io ora mi alzo tardi tutti i giorni, tiro sempre a far mattino,
 le carte poi il caffè della stazione per neutralizzare il vino,
 ma non ho scuse da portare, non dico più d'esser poeta,
 non ho utopie da realizzare: stare a letto il giorno dopo è forse
 l'unica mia meta...

Si alza sempre lenta come un tempo l'alba magica in collina,
 ma non provo più quando la guardo quello che provavo prima.
 Ladri e profeti di futuro mi hanno portato via parecchio,
 il giorno è sempre un po' più oscuro, sarà forse perchè è storia,
 sarà forse perchè invecchio...

Ma le strade sono piene di una rabbia che ogni giorno urla più
 forte,

son caduti i fiori e hanno lasciato solo simboli di morte.

Dimmi se son da lapidare se mi nascondo sempre più,
 ma ognuno ha la sua pietra pronta e la prima, non negare, me la
 tireresti tu...

Sono più famoso che in quel tempo quando tu mi conoscevi,
 non più amici, ho un pubblico che ascolta le canzoni in cui credevi
 e forse ridono di me, ma in fondo ho la coscienza pura,
 non rider tu se dico questo, ride chi ha nel cuore l'odio e nella
 mente la paura...

Ma non devi credere che questo abbia cambiato la mia vita,
 è una cosa piccola di ieri che domani è già finita.

Son sempre qui a vivermi addosso, ho dai miei giorni quanto basta,
 ho dalla gloria quel che posso, cioè qualcosa che andrà presto,
 quasi come i soldi in tasca...

Non lo crederesti ho quasi chiuso tutti gli usci all'avventura,
 non perchè metterò la testa a posto, ma per noia o per paura.

Non passo notti disperate su quel che ho fatto o quel che ho avuto:
 le cose andate sono andate ed ho per unico rimorso le occasioni che
 ho perduto...

Sono ancora aperte come un tempo le osterie di fuori porta,
 ma la gente che ci andava a bere fuori o dentro è tutta morta:
 qualcuno è andato per formarsi, chi per seguire la ragione,
 chi perchè stanco di giocare, bere il vino, sputtanarsi ed è una
 morte un po' peggiore...

[Canzone delle osterie di fuori porta. Francesco Guccini, 1974].

In Poesia...

*Estou cansado, é claro,
 Porque, a certa altura, a gente tem que estar cansado.
 De que estou cansado, não sei:
 De nada me serviria sabê-lo,
 Pois o cansaço fica na mesma.
 A ferida dói como dói
 E não em função da causa que a produziu.
 Sim, estou cansado,
 E um pouco sorridente
 De o cansaço ser só isto –
 Uma vontade de sono no corpo,
 Um desejo de não pensar na alma,
 E por cima de tudo uma transparência lúcida
 Do entendimento retrospectivo...
 E a luxúria única de não ter já esperanças?
 Sou inteligente; eis tudo.
 Tenho visto muito e entendido muito o que tenho visto,
 E há um certo prazer até no cansaço que isto nos dá,
 Que afinal a cabeça sempre serve para qualquer coisa¹.*

[Fernando Pessoa, 1935].

¹ Sono stanco, é ovvio. Perché , arrivati a questo punto, bisogna essere stanchi. Di cosa non lo so. A nulla mi servirebbe saperlo. Perché la stanchezza rimarrebbe la stessa, la ferita fa male in ogni caso e non in funzione di ciò che l'ha prodotta. Sì, son proprio stanco, ed un poco sorridente, che la stanchezza sia solo questo – una voglia di sonno nel corpo, un desiderio di non pensare all'anima, soprattutto una tranquillità lucida, di una valutazione retrospettiva... E la lussuria cambia quando non si ha più speranza? Sono intelligente: questo é tutto. Ho visto molte cose e capito molto di ciò che ho visto, e c'è perfino nella stanchezza che questo mi dà, un certo piacere, Che alla fine la testa sempre serve per ogni cosa.

RI NGRAZI AMENTI

Premetto che non è facile ringraziare tutti coloro che mi hanno accompagnato in quella *andatura vacillante che alcuni chiamano esperienza*. Ed aggiungo, inoltre, che sono veramente negato per questo tipo di cose.

Ringrazio anzitutto il Prof. Roberto Bruno, mio tutore presso il DICMA, per i numerosi consigli, professionali e non, per la pazienza e l'amicizia dimostrata negli ultimi 10 anni della mia "onorabile" carriera accademica.

Je remercie le Professeur Martin Paegelow, mon co-Directeur au sein du laboratoire GEODE/CNRS, pour m'avoir présenté avec enthousiasme le monde de la géomatique et pour avoir supporté, avec patience, mes limites linguistiques dus à ma condition de perpétuel immigré. Une pensée va à la famille Paegelow pour la courtoisie et l'affection montrée.

Je remercie le Professeur Nathalie Villa qui a été fondamentale pour le développement et compréhension du modèle basé sur les réseaux de neuraux. Sa disponibilité et rapidité en répondre mes incessantes questions, sont admirables. Je la remercie encore pour avoir lu toute ma thèse en italien pour le prix modique "d'une pizza géante et d'une glace au chocolat".

Obrigado Professeur Jacques Hubschman, pour m'avoir ouvert les yeux sur l'importance d'une approche multidisciplinaire dans l'étude des changements globaux, et pour m'avoir constamment stimulé à ne pas m'enfermer dans une sévérité mathématico statistique. Une pensée aussi pour la famille Hubschman pour la grande gentillesse démontre pendant tous ces années.

Ringrazio Ezio Crestaz per l'interesse che ha sempre manifestato verso il lavoro svolto, per l'aiuto e la pronta disponibilità a discutere questioni riguardanti i SIG e la modellizzazione. Lo ringrazio anche, e soprattutto, per l'amicizia dimostrata da quando ci conosciamo.

Ringrazio il Prof. Michel Maignan per aver partecipato con interesse alla discussione di questa tesi di dottorato.

Je remercie le Professeur Jean Paul Metaillié, directeur du GEODE/CNRS, pour l'excellent accueil au près du laboratoire, pour m'avoir aidé à mieux comprendre les problématiques de la région d'étude, au Petén, et enfin, pour les nombreuses discussions sur cuisine et plantes (super romarin).

Je remercie tous les amis du GEODE/CNRS pour l'accueil, la compréhension, la patience et la disponibilité. Un spécial merci à la *Hugues' Band* pour l'amitié et l'accueil toujours très chaleureux ! Je remercie également Super J.F pour l'aide dans les situations d'urgence, et Sandrine, pour le soutien "bureaucratique". J'embrasse très fort tous ceux que je n'ai pas cités ici.

Ringrazio la mia Fernanda, fidanzata e collega, per il sostegno durante questi ultimi 3 anni di dottorato, e per le altre, troppe, cose che non posso qui sintetizzare.

Ringrazio mamma, papà ed Enri, per essere sempre il mio punto di riferimento, anche se lontani in quel di Provagna. E ringrazio Provagna.

Ringrazio di cuore gli amici "di bevute" di Tolosa, Andrea, Jaco S., Elena, Midu, Francesca Federico e Ghiom, per il sostegno extra-lavorativo e l'amicizia.

Infine mi congratulo con Marco Follador, per aver terminato una lunga carriera universitaria con ancora tutti i capelli in testa (anche se bianchi), ed aver mantenuto una discreta salute mentale.

BI BLI OGRAFI A

BIBLIOGRAFIA.

Agarwal C., Green G.M., Grove J.P., Evans T.P. & Schweik C.M., (2002). A review and assessment of land use change models. *CIPES collaborative report series No.1*.

Agresti A., 2002. *Categorical data analysis*, 2nd edition. Wiley-series in probability & statistics.

Agterberg F.P & Bonham-Carter G.F., (1990). Deriving weights of evidence from geosciences contour maps for the prediction of discrete events. *Proceedings of XXII international symposium AP-COM*, Pag.381-395.

Al-juaidi F., Millington A.C. & McLaren S.J., (2003). Merged remotely sensed data for geomorphological investigation in desert: examples from central Saudi Arabia. *The geographical journal*, vol.169, I.2, Pag.117-130.

Almeida C.M., (2003). Modelagem da dinâmica espacial como uma ferramenta auxiliar ao planejamento: simulação de mudanças de uso da terra em áreas urbanas para as cidades de Bauru e Piracicaba (SP), Brasil. *Tese de doutorado do curso da Pós-graduação em sensoriamento remoto*, INPE, São José dos Campos.

Alonso D. & Sole R.V., (2000). The DivGame simulator: a stochastic cellular automata model of rainforest dynamics. *Ecological Modeling*, vol.133, Pag.131-141.

Amblard F. & Phan L., (2006, a). *Modélisation et simulation multi-agents*. Lavoisier, Paris.

Amblard F., Rouchier J. & Bommel P., (2006, b). Evaluation et validation de modèles multi-agents. *Modélisation et simulation multi-agents*. Lavoisier, Paris.

Andretta M., Serra R. & Villani M., (2006). A new model for polluted soil risk assessment. *Computers & Geosciences*, vol.32, Pag. 890-896.

Anselin L., (2002). Under the hood: issue in the specification and interpretation of spatial regression models. *Agricultural economics*, vol.27, Pag.247-267.

Anwar S.M., Jeanneret C.A., Parrott L. & Marceau D.J., (2007). Conceptualization and implementation of a multi-agent model to simulate whale-watching tours in the St.Lawrence estuary in Quebec, Canada. *Environmental modeling & software*, vol.22, Pag. 1775-1787.

Arnauld C., Métaillié J.P. & Breuil-Martinez V., (2004). La Joyanca en la region de las masetas y en el Petén noroccidental. *La Joyanca (la Libertad, Guatemala): antigua ciudad maya del noroeste del Petén*, Pag.37-53. Centro francés de estudios mexicanos y centroafricanos & Asociación Tikal, Ciudad de Mexico.

Assoun M.H., (1995). *Fundamentals of artificial neural networks*. MIT press, Cambridge, Massachusetts.

Atkinson P.M. & Lewis P., (2004). Resolution manipulation and sub-pixel mapping. *Remote sensing image analysis*, Springer.

Avolio M.V., Crisci G.M., Di Gregorio S., Rongo R., Spataro W. & Tronfio G.A., (2006). SCIARA γ 2: an improved cellular automata model for lava flow and applications to the 2002 Etnean crisis. *Computers & Geosciences*, vol.32, Pag.876-889.

Baker W.L., (1989). A review of models of landscape change. *Landscape ecology*, vol.2, No.2, Pag.111-133.

Barredo J. & Montserrat G., (2007). Draft copy: Modeling urban land use change in Europe: toward a set of IPCC sres urban land use scenarios. *Modeling environmental dynamics*, Book In Press. Springer International.

Batty M. & Torrens P.M., (2001). Modeling complexity: the limits to prediction. *Paper 36*. CASA, UCL, London.

Baudelle G. & Regnauld H., (2004). *Echelles et temporalités en géographie*. Sedes, Paris.

Bergen K.M., Brown D.G., Rutherford J.F. & Gustafson E.J., (2005). Change detection with heterogeneous data using

ecoregional stratification, statistical summaries and a land allocation algorithm. *Remote sensing of environment*, vol.97, Pag.434-446.

Berger T., Couclelis H., Manson S.M. & Parker D.C., (2001a). Part 1: Introduction and conceptual overview. *Agent based models of LUCC, LUCC Report series No.6*, Pag.1-6.

Berger T., Goodchild M., Janssen M.A., Manson S.M., Najlis R. & Parker D.C., (2001b). Part 2: Methodological consideration for agent-based modeling of LUCC. *Agent based models of LUCC, LUCC Report series No.6*, Pag.7-26.

Berlan J.P., (2004). *La guerra al vivente*. Bollati Boringhieri editore.

Bishop C., (1995). *Neural networks for pattern recognition*. Oxford University Press, NY.

Bonham-Carter G.F., (1994). *Geographic information systems for geoscientists*. Pergamon.

Boots B. & Csillag F., (2006). Categorical maps, comparison and confidence. *Journal of geographical systems*, vol.8, Pag.109-118.

Briassoulis H., (2000). Analysis of land use change: theoretical and modeling approaches. *The Web book of regional science*, WVU Regional Research Centre. <http://www.rri.wvu.edu/WebBook/Briassoulis/contents.htm>.

Brown D.G., Walker R., Manson S. & Seto K., (2004). Modeling land use and land cover change. *Land change science: observing, monitoring and understanding trajectories of change on the Earth's surface*, chapter 23. Springer, New York.

Bruno R., Follador M., Paegelow M., Renno F. & Villa N., (2006). Integrating remote sensing, GIS and prediction models to monitor the deforestation and erosion in Peten reserve, Guatemala. *Proceedings of XIth International Congress of IAMG*, Liège, Belgium. <http://w3.geode.univ-tlse2.fr/doctorants/follador.htm>.

Casti J.L., (1992). *Reality Rules: I, Picturing the world in mathematics*. Wiley-Interscience.

Chavez P.S., Guptill S.C. & Bowell L.B., (1984). Image processing techniques for thematic mapper data. *Technical papers, 50th annual meeting of the American society of photogrammetry, vol.2*, Pag.728-742.

Chen X., Vierling L. & Deering D., (2005). A simple and effective radiometric correction method to improve landscape change detection across sensor and across time. *Remote sensing of environment, vol.98*, Pag.63-79.

Chica-Olmo M. & Abarca-Hernandez F., (2006). Variogram derived image texture for classifying remotely sensed images. *Remote sensing image analysis: including spatial domain*, Pag.93-111. Springer, Netherlands.

Chica-Olmo M. & Abarca-Hernandez F., (2000). Computing geostatistical image texture for remotely sensed data classification. *Computer & Geosciences, vol.26*, Pag.373-383.

Chiles J.P. & Delfiner P., (1999). *Geostatistics. Modelling spatial uncertainty*. Wiley serie in Probability and statistics, John Wiley & Sons, New York.

Cohen W.B., Fiorella M., Gray G., Helmer E. & Anderson K., (1998). An efficient and accurate method for mapping forest clearcuts in the Pacific Northwest using Landsat Imagery. *Photogrammetric and remote sensing, vol.64*, Pag.293-300.

Chomitz K.M. & Gray D.A., (1996). Roads, land use and deforestation: a spatial model applied to Belize. *The World Bank economic review, vol.10, No.3*, Pag.487-512.

Chowdhury R.R. & Schneider L.C., (2004). Land cover and land use: classification and change analysis. *Integrated land-change science and tropical deforestation in southern Yucatan*, Pag. 105-141. Oxford University Press, USA.

Christakos G., Bogaert P. & Serre M.L., (2002). *Temporal GIS: Advanced function for field based application*. Springer.

Christie M.A., Glimm J., Grove J.W, Higdon D.M., Sharp D.H & Wood-Schultz M., (2005). Error analysis and simulation of complex phenomena. *Los Alamos Science*, NO.29, Pag.6-25.

Chung C.F., Fabbri A.G. & Chi K.H., (2002). A strategy for sustainable development of nonrenewable resources using spatial prediction models. Draft copy in *Geoenvironmental Deposit Models for Resources Exploitation and Environmental Security*, Dordrecht, Kluwer Academic publishers.

Chuvieco-Salineró E., (2002). *Teledetección ambiental*. Ariel Ciencia, Barcelona.

Coppin P.R. & Bauer M.E., (1996). Change detection in forest ecosystems with remote sensing digital imagery. *Remote sensing reviews*, vol.13, Pag. 207-234.

Coquillard P. & Hill D.R.C., (1997). *Modélisation et simulation d'écosystèmes*. MASSON, Paris.

Costanza R. & Ruth M., (1998). Using dynamic modeling to scope environmental problems and build consensus. *Environmental management*, vol.22, No.2, Pag. 183-195.

Courier International, (2007). Ecologie, la Chine s'alarme enfin! *Courier International*, N°853, Pag.32-38.

Cruz R. De la, (1976). *Mapa de zonas de vida de Guatemala*, 1:500000. Instituto Nacional Forestal (INAFOR), Ministerio de Agricultura, Guatemala.

Cuevas G. & Mas J.F., (2008). Land use scenarios: a communication tool with local communities. Draft copy in *Modelling environmental dynamics*, Springer series "Environmental science and engineering". Springer, Heidelberg, Germany.

D'Ambrosio D., Spataro W. & Iovine G., (2006). Parallel genetic algorithms for optimizing cellular automata models of natural complex phenomena: an application to debris flow. *Computers & Geosciences*, vol.32, Pag.861-875.

Dean C., (2007). The Problems in Modeling Nature, With Its Unruly Natural Tendencies. *New York Times, Science - Books on science*. A review of "Unless Arithmetic, why environmental scientists can't predict the future".

De Castro F.V.F., Soares Filho B.S. & Mendoza E., (2007). Modelagem de cenários de mudanças na região de Brasília aplicada ao zoneamento ecológico econômico do Estado do Acre. *Anais XIII Simpósio de Sensoriamento Remoto*, INPE, Pag. 5135-5142.

DeFries R., Achard F., Brown S., Herold M., Murdiyarso D., Schlamadinger B. & Sousa C.de, (2007). Earth observations for estimating greenhouse gas emissions from deforestation in developing countries. *Environmental science & policy, vol.10*, Pag.385-394.

Di Gregorio S. & Serra R., (1999). An empirical method for modelling and simulating some complex macroscopic phenomena by cellular automata. *Future generation computer systems, vol.16*, Pag.259-271.

Dubos-Paillard E., Langlois P. & Guermont Y., (2003). Analyse de l'évolution urbaine par automate cellulaire: le modèle SpaCelle. *Espace géographique, vol. 32, No. 4*, Pag. 357-378.

Earth Resource Mapping, (1998). *ER Mapper 6.0, Tutorial*.
©Earth Resource Mapping Pty Ltd.

Eastman R.J, (2006). *IDRISI Andes, Tutorial*. Clark Labs, Clark University, MA, USA.

Eastman R.J, (2003). *IDRISI Kilimanjaro, Tutorial*. Clark Labs, Clark University, MA, USA.

Eckert N., Parent E., Bélanger L. & Garcia S., (2007). Hierarchical Bayesian modeling for spatial analysis of the number of avalanche occurrences at the scale of the township. *Cold region science and technology, vol.50*, Pag.97-112.

Eddy A.J., (1993). Environmental research: what we must do. *Environmental Modeling with GIS*, Pag. 3-7. Oxford University Press, New York.

Effantin-Touyer R., (2006). De la frontière agraire à la frontière de la nature. *Thèse de doctorat de l'INA Paris-Grignon*.

Effantin-Touyer R. & Hubert B., (2004). Politique de stabilisation foncière et conservation de la biodiversité: le cas de la Réserve de Biosphère Maya, Petén, Guatemala. *Autrepart*, vol.31.

Engenhofer M.J., Clementini E. & Di Felice P., (1994). Evaluating inconsistencies among multiple representations. *Sixth International Symposium on Spatial data Handling*, Edinburgh, Scotland. Pag.901-920.

FAO, (2005). Global Forest Resources Assessment, *FRA 2005*. FAO, Roma.

FAO, (2002). Mapeo mundial. *Evaluación de los recursos forestales mundiales 2000 - Informe principal*, Capitulo 47. FAO, Roma.

Farina A., (2001). *L'ecologia del paesaggio*. UTET, Torino.

Farmer M., Swales V., Kristensen L., Nitsch H., Osterburg B. & Poux X., (2007). A research paper of the cross compliance network. *Sixth framework programme*, SSPE-CT-2005-022727.

Ferrand N., (2006). Modéliser avec et pour les acteurs de terrain. *Modélisation et simulation multi-agents*, Pag.219-238. Lavoisier, Paris.

FIPA/AID-CONAP, (2002). *Plan estratégico Zona de Usos Múltiples RBM*. Guatemala.

Follador M., Villa N., Paegelow M., Renno F. & Bruno R., (2008). Tropical deforestation modelling: a comparative analysis of different predictive approaches. The case study of Peten, Guatemala. Draft copy in *Modelling environmental*

dynamics, Springer series "Environmental science and engineering". Springer, Heidelberg, Germany.

Follador M. & Renno F., (2006). Sustainable planning of non-renewable resources using remote sensing and GIS analysis. *Proceedings of International symposium: Nature e société, analyse et modèles*, La Baule, France.

http://letg.univ-nantes.fr/colloque/pdf/PO_FOLLADOR.pdf.

Follador M., (2005). Using remote sensing for mineral exploration in tropical forest area of Brazil. *GIS and Spatial Analysis*, vol. 1, Pag. 127-133. *Proceedings of X International Congress of IAMG*, Toronto, Canada.

<http://w3.geode.univ-tlse2.fr/doctorants/follador.htm>.

Follador M., (2004). A caracterização mineral pelo sensoriamento remoto. *Monografia de Pos-graduação em Geoprocessamento*, Departamento de Cartografia, Instituto de Geociências, UFMG, Belo Horizonte, Brasil.

<http://www.cgp.igc.ufmg.br/centrorecursos/2004/MarcoFollador2004.pdf>.

Fonseca T.F., (2001). *Ontology-driven geographic information systems*. Thesis at University of Maine.

Foody G.M., (2006). Sub-pixel methods in remote sensing. *Remote sensing and image analysis: including the spatial domain*, Pag. 37-49. Springer, Netherlands.

Foody G.M., (1997). Fully fuzzy supervised classification of land cover from remotely sensed imagery with an artificial neural network. *Neural computing & applications*, vol.5, Pag.238-247.

Fournier J.M., (2003). Nouvelles temporalités, changements spatiaux et inégalités sociales. L'exemple des villes pétrolières du Casanare (Colombie). *Cybergeo*, No.238, Pag.1-17.

Gabrielli A., (2007). *Il grande Italiano 2008*. Hoepli, Milano

- Gaiser T., Krol M., Frischkorn H. & de Araujo J.C., (2003). *Global change and regional impacts*. Springer.
- Galton A., (2001). Space time and the representation of Geographical Reality. *Topoi*, vol.20, Pag.173-187.
- Genderen C. & Pohl J.L.Van, (1998). Review article multisensor image fusion in remote sensing: concepts, methods and applications. *International journal of remote sensing*, vol.19, No.5, Pag.823-854.
- Geoghegan J., Villar S.C., Keplis P., Mendoza P.M., Ogneva-Himmelberger Y., Chowdhury R., Turner B.L. & Vance C., (2001). Modeling tropical deforestation in the southern Yucatan peninsular region: comparing survey and satellite data. *Agriculture, ecosystems and environment*, vol.85, Pag.25-46.
- GLP, (2005). Science plan and implementation strategy. *IBGP report NO.53 and IHDP report NO.19*, Stockholm.
- Godoy M.M.G. & Soares-Filho B.S., (2008). Modelling intra-urban dynamics in the Savassi neighbourhood, Belo Horizonte city, Brazil. Draft copy in *Modelling environmental dynamics*, Springer series "Environmental science and engineering". Springer, Heidelberg, Germany.
- Gomez I. & Méndez E., (2005). Asociacion de Comunidades Forestales de Petén, Guatemala: contexto, logras y desafios. *El caso de la Asociacion de Comunidades Forestales de Petén (ACOFOP)*. Fundacion PRISMA, San Salvador.
- Gregory P.J. & Ingram J.S.I., (2000). Food and forestry: global change and global challenges. *Agriculture, Ecosystems and Environment*, vol.82, Pag.1-2.
- Guyot G. & Gu X., (1994). Effect of radiometric correction on NDVI-determined from SPOT-HRV and Landsat-TM data. *Remote sensing of environment*, vol.49, No.3, Pag.169-180.

Haack B. & Jampoler S., (1995). Colour composite comparison for agricultural assessment. *International journal of remote sensing*, vol. 16, No.9, Pag.1589-1598.

Hagen-Zanker A., (2006). Map comparison methods that simultaneously address overlap and structure. *Journal of geographical system*, vol.8, Pag.165-185.

Hagen-Zanker A., (2003). Fuzzy set approach to assessing similarity of categorical maps. *International journal of geographical information science*, vol.17, Pag.235-249.

Hame T., (1991). Spectral interpretation of changes in forest using satellite scanner images. *Acta Forest. Fennica*, vol.222, Pag 1-111.

Hayes D.J. & Sader S.A., (2001). Change detection techniques for monitoring forest clearing and regrowth in a tropical moist forest. *Photogrammetric engineering and remote sensing*, vol.67, I.9, Pag.1067-1075.

Hauglustaine D., Jouzel J. & Le Treut H., (2004). *Climat : chronique d'un bouleversement annoncé*. Le Pommier, Paris.

Hazelton N.W.J., Bennet L.M. & Masel J., (1992). Topological structures for 4-D geographic information system. *Computers, Environment and Urban System*, vol.16, Pag.227-237.

Hegselmann R. & Flache A., (1998). Understanding complex social dynamics: a plea for cellular automata based modeling. *Journal of artificial societies and social simulation*, vol.1, No.3, <http://jasss.soc.surrey.ac.uk/1/3/1.html>.

Hengl, T., (2006). Finding the right pixel size. *Computer & Geosciences*, vol. 32, Pag. 1283-1298.

Hong G. & Zhang Y., (2008). A comparative study on radiometric normalization using high resolution satellite images. *International journal of remote sensing*, vol.29, I.2, Pag.425-438.

INPE, (2004). *Tutorial do Spring*. INPE, São Paulo.

Iovine G., Sheridan M. & Di Gregorio S., (2006). Computer simulation of natural phenomena for hazard assessment. *Computers & Geosciences*, vol. 32, I.7, Pag 859-861.

Irwin E.G. & Geoghegan J., (2001). Theory, data, methods: developing spatially explicit economic models of land use change. *Agriculture, ecosystems and environment*, vol.85, Pag.7-23.

Jensen J.R., (1986). *Introductory digital image processing - a remote sensing perspective*. Prentice-Hall, New Jersey.

Jorgensen S.E., (1994). *Models as instruments for combination of ecological theory and environmental practice*. *Ecological modeling*, vol. 75-76, Pag. 5-20.

Joshi C., De Leeuw J., Skidmore A.K., van Duren I.C. & van Oosten H., (2006). Remotely sensed estimation of forest canopy density: a comparison of the performance of four methods. *International journal of applied earth observation and geoinformation*, vol.8, Pag.84-95.

Junker F., Wieck C., Jansson T. & Pérez I., (2003). Policy description in the CAPRI model. *CAPRI working papers series*, 03-03. Bonn University.

Kanevski M. & Maignain M., (2004). *Analysis and modeling of spatial environmental data*. EPFL press, Lausanne.

Kanungo D.P., Arora M.K., Sarkar S. & Gupta R.P., (2006). A comparative study of conventional, ANN black box, fuzzy and combined neural and fuzzy weighting procedures for landslide susceptibility zonation in Darjeeling Himalayas. *Engineering Geology*, vol.85, Pag.347-366.

Kashanipour R.A. & McGee R.J., (2004). Northern Lacandon Maya medicinal plant use in the communities of Lacanja Chan Sayab and Naha', Chiapas, Mexico. *Journal of ecological anthropology*, vol.8, Pag.47-59.

Kaufman Y.J., (1993). Aerosol optical thickness and atmospheric path radiance. *Journal of geophysical research*, vol.98, No.D2, Pag.2677-2692.

Kavouras M., (2001). Understanding and modeling spatial change. Draft version in *Life and motion of socio-economic units, GISDATA series 8*, Taylor & Francis, London.

Kavzoglu T. & Mather P.M., (2003). The use of backpropagation artificial neural networks in land cover classification. *International journal of remote sensing*, vol.24, I.23, Pag.4907-4938.

Kleijnen J.P.C., (1999). Validation of models: statistical techniques and data availability. *Proceeding of the 1999 Winter Simulation Conference*, Pag.647-654.

Kohonen T., (2006). Self-organizing neural projections. *Neural Networks*, vol.19, Pag.723-733.

Koomen E., Rietveld P. & de Nijs T., (2007). Modelling land use change for spatial planning support. *The annals of regional science*, Editorial. Springer, Berlin.

Kristensen L. & Primdahl J., (2004). Potential for environmental cross-compliance to advance agri-environment objectives. *Report of EU concerted Action Project "developing cross-compliance in the EU - background, lessons and opportunities"*. Danish centre for forest, landscape and planning, the royal veterinary and agricultural university, Denmark.

Lambin E., (2004). *La Terre sur un fil*. Le Pommier, Paris.

Lambin E., Geist H.J. & Lepers E., (2003). Dynamics of land use and land cover change in tropical regions. *Annual review of environment and resources*, vol.28, Pag.205-241.

Lambin E., et al., [2001]. The cause of land use and land cover change: moving beyond the myths. *Global environmental change*, vo.11, Pag.261-269.

Lambin E., Baulies X., Bockstael N., Fischer G., Krug T., Leemans R., Moran E.F., Rindfuss R.R., Sato Y., Skole D., Turner B.L. & Vogel C., (2000). LUC Implementation strategy. *IGBP Report 48 and IHDP Report 10*. Stockholm.

Lambin E., (1997). Modelling and monitoring land-cover change processes in tropical regions. *Progress in physical geography*, vol.21, No.3, Pag.375-393.

Langran G., (1993). *Time in geographic information systems*. Taylor & Francis, London.

Laurence W.F., Laurence S.G & Delamonica P., (1998). Tropical forest fragmentation and greenhouse gas emissions. *Forest ecology and management*, vol.110, Pag.173-180.

Lee V.C.S. & Wong H.T., (2007). A multivariate neuro-fuzzy system for foreign currency risk management decision making. *Neurocomputing*, vol.70, Pag.42-951.

Le Moigne J.L., (1994). *La théorie du système general*. 4° édition, Puf, Paris.

LeSage J.P., (1999). *Spatial econometrics*. Webbook, <http://www.rri.wvu.edu/WebBook/LeSage/spatial/wbook.pdf>

Li C., (2007). Quantifying greenhouse gas emissions from soils: scientific basis and modeling approach. *Soil science and plant nutrition*, vol.53, Pag.344-352.

Lillesand T.M. & Kiefer R.W., (1979). *Remote sensing and image interpretation*. John Wiley & Sons, NY.

Lovelock J., (2006). *A vingança de Gaia*. Intrinseca, Rio de Janeiro.

Manca D., (2005). Reti neurali artificiali. *Documentazione CIP 2005*, Santa Margherita di Pula (CA).

Mas J.F & Flores J.J., (2007). A review of the application of artificial neural networks to the analysis of remotely-sensed data. Draft copy for *International Journal of Remote Sensing*, In press.

Matheron G., (1978). Estimer et choisir. *Cahiers du centre de morphologie mathématique de Fontainebleau, Fasc.7*, Ecole de Mines de Paris.

Matheron G., (1989). *Estimating and choosing - an essay on probability in practice*. Springer, Berlin.

McKenzie J.W., Williamson I.P. & Hazelton N.W.J., (2000). 4-D adaptive GIS: justification and methodologies. Dept. of Geomatics, Univ. of Melbourne: http://www.geom.unimelb.edu.au/research/SDI_research/publications/ianpubli.htm .

Métaillié J.P, Veliz M, Galop D. & Carozza J.M., (2000). Estudios geograficos, botanicos y paleoambientales. *Informe No.2, Segunda temporada de campo en el sitio arqueologico de la Joyanca y su region*, Capitulo 11. Guatemala.

Mertens K.C., Verbeke L.P.C., Westra T. & de Wulf R.R., (2004). Sub-pixel mapping and sub-pixel sharpening using neural network predicted wavelet coefficients. *Remote sensing of environment, vol.91, I.2*, Pag.225-236.

Mezzadri-Centeno T., (1998). La modélisation et la projection spatio-temporelle dans les SIG. *Thèse de doctorat de l'université Paul Sabatier de Toulouse (Science)*.

Mirzavand G., (2003). Using principal techniques on ETM+ 2002 for arid and semi-arid environment Central Iran. @GISdevelopment.net.

Munroe D.K & Muller D., (2007). Issue in spatially explicit statistical LUCC models: examples from western Honduras and central highlands of Vietnam. *Land use policy, vol.24*, Pag.521-530.

Munroe D.K., Southworth J. & Tucker C.M., (2004). Modeling spatially and temporally complex land cover change: the case of western Honduras. *The professional geographer, vol.56 (4)*, Pag. 544-559.

Nelson G.C & Geoghegan J., (2002). Deforestation and land use change: sparse data environments. *Agricultural economics*, vol.27, Pag.201-216.

Nemmour H. & Chibani Y., (2006). Multiple support vector machines for land cover change detection: an application for mapping urban extension. *ISPRS journal of photogrammetry & remote sensing*, vol.61, Pag.125-133.

Nittler J. & Tschinkel H., (2005). Manejo comunitario del bosque en la reserve de la Biosfera Maya de Guatemala. Documento presentado a l'Agencia de los EE.UU para el desarrollo international y SANREM. University of Georgia.

Orzak S.H. & Sober E., (1993). A critical assessment of Levins's "the strategy of model building in population biology" 1966. *The quarterly review of biology*, vol.68, No.4, Pag.533-546.

Paegelow M. & Camacho O.M.T, (2005). Possibilities and limits of prospective GIS land cover modeling-a compared case study: Garrotex (France) and Alpujarra Granadina (Spain). *International Journal of GIS*, vol.19, I.6, Pag.697-722.

Paradella W.R., Santos A.R.Dos, Veneziani P. & Morais M.C.De, (2001). O sinergismo de imagens standard descendente do RADARSAT-1 e thematic mapper do Landsat-5, Aplicação geologica na Amazonia. *Anais X SBSR*, INPE, Pag.315-322, Foz do Iguaçu.

Parker D.C., Manson S.M., Janssen M.A., Hoffmann M.J. & Deadman P., (2003). Multi-agent systems for the simulation of LUCC: a review. *Annals of the Association of American Geographers*, vol.93(2), Pag.314-337.

Parks B.O., (1993). The need for integration. *Environmental Modeling with GIS*, Pag.31-34. Oxford University Press, New York.

Passos M.M.Dos, (2007). *De Estrada dos colonos a corredor de exportação*. Editora Massoni, Maringà-PR.

Pathak H., Li C. & Wassmann R., (2005). Greenhouse gas emissions from Indian rice fields: calibration and upscaling using the DNDC model. *Biosciences*, vol.2, Pag.113-123.

Pelekis N., Theodoulis B., Kopanakis I. & Theodoridis Y., (2004). Literature review of spatio-temporal database models. *The Knowledge Engineering Review*, vol.19, Pag.235-274.

Pessoa F., (2006). *Poesia completa da Alberto Caeiro*, 2a reimpressão. Companhia de bolso, São Paulo.

Pessoa F., (2004). *Alvaro de Campos*. Companhia das letras, São Paulo.

Peuquet D.J., (1994). It's about time: a conceptual framework for representation of temporal dynamics in geographic information systems. *Annals of the Association of American Geographers*, vol. 84, Pag. 441-461.

Pontius R.G.J. & Chen H., (2006). Land use & cover change modeling. *Land change modeling with GEOMOD*, Idrisi Andes tutorial, Clark University.

Pontius R.G.J., Huffaker D. & Denman K., (2004, a). Useful techniques of validation for spatially explicit land-change models. *Ecological modeling*, vol.179, Pag. 445-461.

Pontius R.G.J. & Pacheco P., (2004, b). Calibration and validation of a model of forest disturbance in the western Ghats, India 1920-1990. *Geojournal*, vol.61, Pag.325-334.

Pontius R.G.J., Shusas E. & McEachern M., (2004, c). Detecting important categorical land changes while accounting for persistence. *Agriculture, ecosystems and environment*, vol.101, Pag.251-268.

Pontius R.G.J. & Schneider L.C., (2001). Land cover change model validation by an ROC method for the Ipswich watershed, Massachusetts, USA. *Agriculture, ecosystems & environment*, vol.85, I.1-3, Pag.239-248.

Printz A. & Lang R., (2003). The GIS-based model for sustainable development of land use (MOSDEL). *Global change and regional impacts*, Pag. 57-78. Springer.

Raffestin C., (2005). *Dalla nostalgia del territorio al desiderio di paesaggio*. Alinea editrice, Firenze.

Ramat E., (2006). Introduction à la modélisation et à la simulation à événements discrets. *Modélisation et simulation multi-agents*, Pag. 49-71. Lavoisier, Paris.

Renno F., (2007). *Le Sertão Mineiro: un territoire à la recherché de ses paysages*. Version provisoire de la thèse doctorale en environnement et paysage. GEODE/CNRS, Université de Toulouse-Le Mirail, Toulouse.

Rindfuss R.R., Walsh S.J., Turner B.L., Fox J. & Mishra V., (2004). Developing a science of land change: challenges and methodological issues. *PNAS*, vol.101, Pag. 13976-13981. The national Academy of sciences of the USA.

Rodriguez H.O., Soares-Filho B.S. & Souza Costa W.L. de, (2007). Dinamica EGO, uma plataforma para modelagem de sistemas ambientais. *Anais XIII simposio brasileiro de sensoriamento remoto*, Pag.3089-3096. INPE.

Roger A., (1997). *Court traité du paysage*. Gallimard.

Rokityanskiy D., Benitez P.C., Kraxner F., McCallum I., Obersteiner M., Rametsteiner E. & Yamagata Y., (2007). Geographically explicit global modeling of land use change, carbon sequestration and biomass supply. *Technological forecasting & social change*, vol.74, Pag. 1057-1082.

Rouchier J., Bousquet F., Dejardins M.R. & Antona M., (2001). A multi-agent model for describing transhumance in North Cameroon: comparison of different rationality to develop a routine. *Journal of economic dynamics and control*, vol.25, Pag.527-559.

Saaty, T.L., (1977). A scaling method for priorities in hierarchical structures. *Journal of mathematical psychology*, vol.15, Pag.234-281.

Sabins F.F., (1987). *Remote sensing, principles and interpretation*. 2° Edition, Freeman and Company, NY.

Sader S.A., Bertrand M. & Wilson E.H., (2003). Satellite change detection of forest harvest patterns on an industrial forest landscape. *Forest science*, vol.49, I.3, Pag.341-353.

Schott J.R., Salvaggio C. & Volchok W.J., (1988). Radiometric scene normalization using pseudoinvariant features. *Remote sensing of environment*, vol.26, Pag.1-16.

Serneels S. & Lambin E., (2001). Proximate causes of land use change in Narok district, Kenya: a spatial statistical model. *Agriculture, ecosystems and environment*, vol.85, Pag 65-81.

Shamshad A., Bawadi M.A., Wan Hussin W.M.A., Majid T.A. & Sanusi S.A.M., (2005). First and second order Markov chain models for synthetic generation of wind speed time series. *Energy*, vol.30, Pag.693-708.

Shiva V., (2003). *Le guerre dell'acqua*. Feltrinelli, Milano.

Soares-Filho B., Coutinho Cerqueira G. & Lopes Pennachin C., (2002). DINAMICA - a stochastic cellular automata model designed to simulate the landscape dynamics in an Amazonian colonization frontier. *Ecological modeling*, vol.154, I.3, Pag.217-235.

Soares-Filho B., (1998). Modelagem da Dinamica de paisagem de uma regioao fronteira de colonizaco amazonica. *Tese de doutorado* na escola Politcnica da Universidade de So Paulo.

Song C., Woodcock C.E., Seto K.C., Lenney M.P. & Maccombers S.A., (2001). Classification and change detection using Landsat TM data: when and how to correct atmospheric effects? *Remote sensing of environment*, vol.75, Pag.230-244.

Steyaert L.T, (1993). A perspective on the state of environmental simulation modeling. *Environmental Modeling with GIS*, Pag. 16-30. Oxford University Press, New York.

Torrens P.M., (2002). Cellular automata and multi-agent systems as planning support tools. *Planning support system in practice*, Pag. 205-222. Springer, London.

Tso B. & Mather P.M., (2001). *Classification methods for remotely sensed data*. Taylor & Francis, NY.

Turner II B.L., Geoghegan J. & Foster D.R., (2004). *Integrated land-change science and tropical deforestation in southern Yucatan*. Oxford University Press, USA.

Veldkamp A. & Lambin E.F., (2001). Predicting land-use change. *Agriculture, ecosystems and environment*, vol.85, Pag.1-6.

Veldkamp A. & Fresco L.O., (1996). CLUE: a conceptual model to study the conversion of land use and its effects. *Ecological Modeling*, vol.85, Pag.253-270.

Venables W.N. & Ripley B.D., (1999). *Modern applied statistics with S-plus, 3^oedition*, Springer, New York.

Verburg P.H., Schulp C.J.E., Witte N. & Veldkamp A., (2006). Downscaling of land use change scenarios to assess the dynamics of European landscapes. *Agriculture, ecosystems and environment*, vol.114, Pag.39-56.

Verburg P.H. & Veldkamp A., (2005). Introduction to the special issue on spatial modeling to explore land use dynamics. *International journal of GIS*, vol.19, No.2, Pag.99-102.

Verburg P.H. & Veldkamp A., (2001). The role of spatially explicit models in land-use change research: a case study for cropping patterns in China. *Agriculture, ecosystems and environment*, vol.85, Pag..177-190.

Villa N., Paegelow M., Camacho O.M.T., Cornez L., Ferraty F., Ferré L. & Sarda P., (2007). Various approaches for

predicting land cover in mountain areas. *Communication in Statistics-simulation and computation*, vol.36, Pag.73-86.

Villa N., (2005). Eléments d'apprentissage en statistique fonctionnelle - classification et régression fonctionnelle par réseaux de neurones et support vector machine. *Thèse de doctorat* (mathématiques) de l'Université Toulouse-Le Mirail.

Villa N., (2002). *Réseaux de neurones*. Mémoire de DEA, Université Paul Sabatier, Toulouse.

Wilson E.H. & Sader S.A., (2002). Detection of forest harvest type using multiple dates of Landsat TM imagery. *Remote sensing of environment*, vol.80, Pag.385-396.

Wolfram S., (1984). Cellular automata as models of complexity. *Nature*, vol.311, Pag.419-424.

Wolfram S., (1983). Statistical mechanics of cellular automata. *Review of modern physics*, vol.55, Pag.601-643.

Yang G.J., Vounatsou P., Zhou X.N., Tanner M. & Utzinger J., (2005). A Bayesian-based approach for spatio-temporal modeling of county level prevalence of schistosoma japonicum infection in Juangsu province, China. *International journal of parasitology*, vol.35, Pag.155-162.

Lista delle figure

Fig 1-1: Global Land Programme (GLP)	4
Fig 1-2: Limite di sostenibilità del sistema terrestre	13
Fig 1-3: Esempio di studio multi-temporale	17
Fig 2-1: Complessità di un modello	27
Fig 2-2: Lo schema concettuale del Earth System Science Concept	30
Fig 2-3: Fasi della modellizzazione	32
Fig 2-4: Fasi principali della pianificazione spaziale	43
Fig 2-5: Complessità di un sistema	46
Fig 2-6: Tassonomia dei metodi di modellizzazione proposti da Coquillard e Hill	48
Fig 2-7: Differenti regole di vicinaggio	58
Fig 2-8: Il modello SCIARA γ 2-regole di vicinaggio	60
Fig 2-9: Modello DNDC-schema concettuale	65
Fig 2-10: Procedimento di upscaling del DNDC	65
Fig 2-11: Modello DIMA	66
Fig 2-12: Esempio di approccio Multi-agente	70
Fig 3-1: Rete neurale con un solo strato nascosto	77
Fig 3-2: Neurone biologico e neurone artificiale	78
Fig 3-3: Schema generale di una rete neurale	80
Fig 3-4: Funzione logistica sigmoide	80
Fig 3-5: Apprendimento supervisionato	83
Fig 3-6: Auto-apprendimento	83
Fig 3-7: Regole di vicinaggio adottate in PNNET	89
Fig 3-8: Struttura di PNNET	90
Fig 3-9: Approccio metodologico adottato in PNNET	93
Fig 3-10: Andamento dell'errore reale-simulato in PNNET	94
Fig 3-11: Matrice variables.quant.temps	97
Fig 3-12: Diagramma di flusso di Dinamica.EGO	102
Fig 3-13: Modello 01-determine_transition_matrix.xml	104
Fig 3-14: Modello 02-WoE_range_coeff.xml	109
Fig 3-15: Modello 03-WoE_correlazione.xml	110
Fig 3-16: Modello 04-Run_simulazione.xml	112
Fig 3-17: Strategia utilizzata in MCE-OWA	117
Fig 4-1: Riserva della Biosfera Maya e localizzazione dell'area di studio	126
Fig 4-2: La Joyanca-foto delle lagune	129
Fig 4-3: DEM dell'area di studio	130
Fig 4-5: L'agricoltura taglia-brucia-foto	132
Fig 4-6: Esempio di agricoltura mobile-foto	134
Fig 4-7: Attività pastorizie-foto	135
Fig 4-8: Mappa degli incendi in Peten nel 2005	135

Fig 4-9: Rettangolo involvente l'area di studio	141
Fig 4-10: Maschera binaria	142
Fig 4-11: Aumento del contrasto digitale	144
Fig 4-12: Problema di saturazione (overflow)	145
Fig 4-13: Espansione dei toni di grigio	146
Fig 4-14: Correzione radiometrica DOS	150
Fig 4-15: Parametri di un'immagine SPOT	151
Fig 4-16: Colour Composite ed OIF, dati 27/03/00	154
Fig 4-17: Cause principali dei pixels misti	156
Fig 4-18: Dimensione del pixel e frammentarietà del paesaggio	157
Fig 4-19: Contrasto spettrale della vegetazione	160
Fig 4-20: Indici vegetali NDVI	162
Fig 4-21: Diagramma di dispersione dei pixels	165
Fig 4-22: Analisi multitemporale RGB (NDMI)	167
Fig 4-23: Classificazione delle immagini satellitari	171
Fig 4-24: Immagini classificate del 98-00-03	172
Fig 4-25: Matrici delle variabili esplicative	174
Fig 4-26: Analisi dell'importanza di una variabile	175
Fig 5-1: Cambi fra il 1988 ed il 1992	181
Fig 5-2: Deforestazione 1988-2003	183
Fig 5-3: Periodo di calibrazione e validazione	185
Fig 5-4: Output dei modelli predittivi	186
Fig 5-5: Carta PNNET e carta reale nel 2003	189
Fig 5-6: Carta CA-Markov e carta reale nel 2003	190
Fig 5-7: Carta LCM e carta reale nel 2003	191
Fig 5-8: Carta Dinamica Ego e carta reale nel 2003	192
Fig 5-9: Precisione e frequenza del cambiamento	200
Fig 5-10: Validazione fuzzy	203
Fig c-1: Simulazione dell'uso del suolo nel 2008	216
Fig c-1: Carte prédictive d'utilisation du sol en 2008	230

Li sta Tabel l e

Tab 1-1: Esempio di modellizzazione predittiva multi-scenario	20
Tab 1-2: Principali cause di LUCC	22
Tab. 3-1: Matrici di Markov per il Petén	106
Tab 3-2: Criteri ambientali nello studio dei LUCC in Petén	116
Tab 3-3: Pesi utilizzati nel modello CA_Markov	117
Tab 4-1: Lista delle immagini satellitari utilizzate	138
Tab 4-2: Calcolo dell'OIF per i dati ETM del 27/03/00	153
Tab 4-3: Coefficiente. di Cramer per le variabili esplicative	177
Tab 5-1: Analisi quantitativa reale-simulato nel 2003	193
Tab 5-2: Analisi pixel-a-pixel reale-simulato nel 2003	195
Tab 5-3: LUCC budget	197
Tab 5-4: Complementarietà dei modelli	198
Tab 5-5: Validazione pixel-a-pixel e validazione fuzzy	203

Riassunto

L'uso frequente dei modelli predittivi per l'analisi di sistemi complessi, naturali o artificiali, sta cambiando il tradizionale approccio alle problematiche ambientali e di rischio. Il continuo miglioramento delle capacità di elaborazione dei computer facilita l'utilizzo e la risoluzione di metodi numerici basati su una discretizzazione spazio-temporale che permette una modellizzazione predittiva di sistemi reali complessi, riproducendo l'evoluzione dei loro patterns spaziali ed calcolando il grado di precisione della simulazione. In questa tesi presentiamo una applicazione di differenti metodi predittivi (Geomatico, Reti Neurali, Land Cover Modeler e Dinamica EGO) in un'area test del Petén, Guatemala. Durante gli ultimi decenni questa regione, inclusa nella Riserva di Biosfera Maya, ha conosciuto una rapida crescita demografica ed un'incontrollata pressione sulle sue risorse naturali. L'area test può essere suddivisa in sotto-regioni caratterizzate da differenti dinamiche di uso del suolo. Comprendere e quantificare queste differenze permette una migliore approssimazione del sistema reale; è inoltre necessario integrare tutti i parametri fisici e socio-economici, per una rappresentazione più completa della complessità dell'impatto antropico. Data l'assenza di informazioni dettagliate sull'area di studio, quasi tutti i dati sono stati ricavati dall'elaborazione di 11 immagini ETM+, TM e SPOT; abbiamo poi realizzato un'analisi multitemporale dei cambi uso del suolo passati e costruito l'input per alimentare i modelli predittivi. I dati del 1998 e 2000 sono stati usati per la fase di calibrazione per simulare i cambiamenti nella copertura terrestre del 2003, scelta come data di riferimento per la validazione dei risultati. Quest'ultima permette di evidenziare le qualità ed i limiti per ogni modello nelle differenti sub-regioni.

Parole chiave: Modelli predittivi; Discretizzazione spazio-temporale; Telerilevamento; Reti neurali; Catene di Markov; Dinamica; Gestione del rischio; Deforestazione; Petén; Guatemala

Abstract

The frequent use of predictive models for analyzing of complex, natural or artificial phenomena is changing the traditional approaches to environmental and hazard problems. The continuous improvement of computer performance allows for more detailed numerical methods, based on space-time discretisation, to be developed and run for a predictive modeling of complex real systems, reproducing the way their spatial patterns evolve and pointing out the degree of simulation accuracy. In this contribution we present an application of several methods (Geomatics, Neural Networks, Land Cover Modeler and Dinamica EGO) in the tropical training area of Petén, Guatemala. During the last few decades this region, included in the Biosphere Maya reserve, has seen a fast demographic raise and a subsequent uncontrolled pressure on its own geo-resources. The test area can be divided into several sub-regions characterized by different land use dynamics. Understanding and quantifying these differences permits a better approximation of a real system; moreover we have to consider all the physical, socio-economic parameters, which will be of use for representing the complex and sometimes random human impact. Because of the absence of detailed data from our test area, nearly all the information was derived from the image processing of 11 ETM+, TM and SPOT scenes; we studied the past environmental dynamics and we built the input layers for the predictive models. The data from 1998 and 2000 were used during the calibration to simulate the land cover changes in 2003, selected as reference date for the validation. The basic statistics permit to highlight the qualities or the weaknesses for each model on the different sub-regions.

Keywords: Predictive Models; Space-time discretisation; Remote Sensing; Neural Networks; Markov Chains; MCE; Dinamica; Risk management; Deforestation; Petén; Guatemala

Résumé

L'utilisation chaque jour plus fréquent des modèles prédictifs pour décrire des systèmes complexes, naturels ou artificiels, est en train de changer progressivement les approches traditionnelles des problématiques environnementales et de gestion du risque. Le remarquable potentiel des ordinateurs actuels rend possible le développement de modèles numériques de calcul basés sur une discrétisation spatio-temporelle afin de simuler le comportement de phénomènes complexes ainsi que de prévoir des scénarios futures, avec différents degrés d'approximation. On présente ici une application de différents modèles prédictifs (géomatique, réseaux de neurones, LCM Idrisi Andes et Dinamica), dans une région de forêt tropicale au Petén, Guatemala. Cette région, inscrite dans la réserve de biosphère Maya, a connu dans les dernières décades, une croissance démographique et une pression non contrôlée sur les ressources naturelles. Dans la région d'étude nous pouvons signaler des sous-régions avec différentes dynamiques d'utilisation du sol. Comprendre et quantifier de telles différences permet l'obtention d'une bonne approximation de la situation réelle et il est nécessaire d'intégrer les paramètres physiques, sociaux et économiques essentiels pour décrire la complexité, parfois aléatoire, de l'action anthropique. La première phase d'étude est basée exclusivement sur l'analyse d'informations satellites, car il n'existe pas d'autres données utiles disponibles sur cette étendue. Le traitement de 11 images ETM+, TM et SPOT a rendu possible l'observation des dynamiques environnementales passées ainsi que la construction de l'Input pour les différentes approches prédictives. Les résultats obtenus, en partent des informations collectées en 1998 et 2000, seront validés avec une image réelle de 2003; les statistiques de base vont permettre de souligner les points forts et les points faibles de chaque modèle sur les différentes sous-régions.

Mots clefs: modèles prédictifs; discrétisation spatio-temporelle; télédétection; réseaux de neurones; chaîne de Markov; MCE; Dinamica; gestion du risque; déforestation; Petén; Guatemala.