

Un sistema di certificazione risk-based per i controlli in agricoltura biologica: un'applicazione tramite Bayesian networks

Danilo Gambelli*, Francesco Solfanelli*, Raffaele Zanolì*

INTRODUZIONE

La competitività dei prodotti biologici poggia in modo sostanziale su una diversificazione qualitativa rispetto ai prodotti convenzionali, perseguita tentando di mantenere un differenziale di prezzo contenuto. Il processo di certificazione in questo contesto gioca un ruolo cruciale: da un lato rappresenta l'elemento di verifica e garanzia della differenziazione del prodotto biologico, costituendone quindi un elemento imprescindibile ed essenziale; dall'altro costituisce un elemento di costo transazionale che riduce la competitività dei prodotti biologici determinando diversi elementi di costo, sia diretti che indiretti, che si distribuiscono lungo tutta la filiera del prodotto.

In linea generale, la certificazione è un sistema attraverso il quale è possibile stabilire la conformità di un prodotto o servizio rispetto a un determinato standard (per l'agricoltura biologica il Reg. CE 834/2007). L'elemento che caratterizza tutti i sistemi di certificazione dei prodotti biologici adottati nei vari stati membri dell'Unione Europea è il fatto che la certificazione, e di conseguenza il rilascio della licenza, è condotta da un soggetto di parte terza (Organismo di Controllo e Certificazione: ODC) in ottemperanza agli standard previsti. Gli ODC, a seguito di ispezioni e procedure di controllo, verificano il rispetto della normativa in materia di prodotti biologici, e nel caso di riscontro di non conformità ai sensi del Reg. CE 834/2007 sanzionano gli operatori.

Il Reg. CE 834/07 classifica genericamente le non conformità come irregolarità e infrazioni, senza offrirne una definizione particolarmente precisa. Nel presente studio, si è quindi fatto diretto riferimento alle indicazioni in materia del regolamento tecnico redatto da ACCREDIA. Questo, recependo la normativa europea, fornisce infatti delle precise linee guida per la definizione delle non conformità e dei provvedimenti sanzionatori.

L'obiettivo generale del presente lavoro è quello di analizzare gli aspetti normativi e procedurali del processo di controlli previsti nei sistemi di certificazione per le aziende agricole biologiche, e di individuare con un modello probabilistico i fattori strutturali e gestionali aziendali che in questo contesto maggiormente influenzano il rischio di commissione di infrazioni ed irregolarità da parte degli operatori. In questa sede ci si limita ad un'analisi di un caso studio riferito ai produttori agricoli.

* Dipartimento di Scienze Agrarie, Alimentari e Ambientali – Università Politecnica delle Marche.

L'approccio metodologico per il modello probabilistico si basa sull'applicazione delle Reti Bayesiane (*Bayesian Networks*, qui di seguito BN); il principale vantaggio di questo metodo risiede nel fatto che una forte semantica causale rende possibile l'identificazione di relazioni non note a priori partendo da un set di dati a disposizione. In questo contesto le BN offrono la possibilità di condurre inferenze mirate da un lato all'individuazione dei principali elementi di impatto sulle probabilità di riscontro di non conformità, e dall'altro alla simulazione del rischio di non conformità relativo a di varie combinazioni di elementi strutturali e gestionali aziendali.

Il presente studio è stato condotto analizzando i dati di un campione di aziende agricole italiane certificate dall'IMC (Istituto Mediterraneo di Certificazione) nell'anno 2007. Sulla base delle informazioni strutturali e gestionali delle aziende biologiche, unitamente a quelle relative al numero di controlli per azienda effettuati e ai rispettivi esiti in termini di sanzioni emesse, è possibile calibrare un sistema esperto in grado di simulare da un lato i fattori strutturali e gestionali che rappresentano i principali elementi di rischio di violazione del disciplinare, e dall'altro di considerare i corrispondenti livelli di rischiosità per diverse tipologie aziendali.

La struttura del lavoro è la seguente: nel primo paragrafo viene sviluppata un'introduzione del sistema di certificazione in agricoltura biologica, riportando la situazione attuale e le problematiche di efficienza che interessano maggiormente il settore. Nel secondo paragrafo viene descritta la metodologia utilizzata per il modello probabilistico, mentre il terzo è dedicato alla descrizione del reperimento dei dati. Nel quarto paragrafo vengono riportati i risultati della ricerca; le considerazioni conclusive terminano il lavoro.

1 STATO ATTUALE DEL SISTEMA DI CERTIFICAZIONE BIOLOGICA IN ITALIA

Con l'entrata in vigore del Reg. CEE 2092/91, sostituito a gennaio del 2009 dal nuovo regolamento del Consiglio relativo alla produzione e all'etichettatura dei prodotti biologici (Reg. CE 834/2007 del 28 giugno 2007), l'agricoltura biologica è stato il primo metodo di produzione agricola a dotarsi di un sistema di controllo e certificazione di processo esteso a tutta la filiera. La certificazione è l'elemento chiave dell'intero sistema agricolo biologico odierno, sia per quanto riguarda la garanzia della qualità del prodotto e la rigorosità del processo produttivo, sia per permettere al consumatore di individuare con certezza i prodotti biologici stessi (Schmid et.al., 2008). Il compito principale dell'intero sistema di certificazione in agricoltura biologica è quello di garantire l'assoluta integrità del prodotto; tale competenza è resa indispensabile dalla presenza, lungo tutta la filiera dei prodotti agricoli biologici, di imperfezioni nell'informazione rivolta al consumatore nonché di comportamenti ingannevoli da parte degli operatori.

Il sistema normativo di riferimento per la certificazione della rispondenza di un prodotto biologico allo standard previsto è il Reg CE 834/2007, che attribuisce agli ODC le funzioni concernenti le fasi di ispezione, certificazione e rilascio delle licenze. Relativamente all'attività di controllo, la normativa europea ha lasciato agli stati

membri ampia facoltà di decidere l'organizzazione del sistema sul proprio territorio nazionale; di fatto, pur essendo la struttura base del sistema di certificazione la stessa in tutta Europa, ogni singolo paese ha la possibilità di attuare il sistema apportando delle specifiche modifiche. Nell'ambito dell'Unione Europea esistono due principali approcci. Nella maggior parte dei paesi l'organizzazione e il funzionamento del sistema è demandato ad ODC privati, che possono a loro volta essere domestici o provenire da altri paesi. In alcuni stati come la Danimarca e la Finlandia, sia la certificazione che il monitoraggio sono invece sotto l'egida della pubblica autorità (Seppänen, 2004); un simile sistema è adottato anche in Olanda, dove l'autorità pubblica demanda ad un solo ODC il compito di effettuare le ispezioni. Per quanto riguarda il primo approccio, la normativa europea prevede che gli ODC operanti nel territorio siano riconosciuti da una specifica autorità preposta. In Italia l'autorità competente per il coordinamento dell'intero sistema di controllo in agricoltura biologica è il Ministero delle Politiche Agricole, Alimentari e Forestali (MIPAF)¹, mentre sono le regioni ad essere coinvolte nel processo di autorizzazione degli ODC privati e di conseguenza ad attuare la sorveglianza su di essi nei territori di rispettiva competenza². L'attività di sorveglianza e autorizzazione da parte dell'autorità competente risulta essenziale, sia per garantire il corretto e uniforme funzionamento del sistema di controllo e certificazione, affidato a un numero crescente di ODC privati, sia per restituire dati e informazioni utili al miglioramento dell'intero sistema normativo e organizzativo di settore. Al momento in Italia sono autorizzati ad operare 18 ODC privati, di cui 5 sono autorizzati ad operare esclusivamente nel territorio della provincia autonoma di Bolzano; va evidenziato comunque che circa il 95% degli operatori è certificato da dieci ODC³.

Il Reg. CE 834/2007 introduce una importante novità rispetto alla precedente normativa in materia, imponendo l'obbligo di accreditamento degli ODC alla normativa UNI CEI EN 45011. Tale obbligo risponde alla necessità di sopperire al problema della mancanza di uniformità di interpretazione della normativa rispetto alle specificità del settore biologico e la molteplicità degli ODC coinvolti nel processo di accreditamento. In Italia gli ODC hanno l'obbligo di adottare un procedimento codificato per le attività di certificazione e ispezione, sulla base del rapporto tecnico RT ACCREDIA n. 16⁴. Le linee guida ACCREDIA recepiscono la normativa europea individuando due livelli di non conformità, irregolarità ed infrazione, sopperendo di

¹ Per le attività di monitoraggio e controllo delle produzioni biologiche il MIPAF ha demandato il compito all'Istituto Nazionale Controllo Qualità e Repressione Frodi (ICQRS).

² Con la recente legge finanziaria 2007 si è tuttavia avviata la riorganizzazione del sistema di sorveglianza, le cui competenze a livello ministeriale sono state affidate all'Ispettorato centrale per il controllo della qualità mentre; assieme alle Regioni, è in corso di definizione una nuova normativa che disciplini l'attività di sorveglianza.

³ Per maggiori dettagli si veda www.sinab.it.

⁴ Per maggiori dettagli si veda www.accredia.it.

fatto ad una certa vaghezza del Reg. CE 834/2007 in merito alla definizione delle irregolarità e infrazioni, già sottolineata da Measures e Padel (2006).

L'irregolarità viene definita come il mancato rispetto di aspetti formali e della documentazione prevista; tali inadempienze non comportano effetti prolungati e ad ogni modo non alterano l'affidabilità dell'operatore. Con riferimento a quanto sopraesposto circa il principio di gradualità di applicazione delle sanzioni, le irregolarità sono suddivise nell'ambito del RT 16 in lievi e gravi.

L'infrazione consiste nel mancato rispetto degli obblighi previsti dal Reg. CE 834/2007. Tuttavia, rispetto alle irregolarità, è una tipologia di inadempienza che può originare sia da reiterate e gravi irregolarità, sia da violazioni non formali della normativa, generando effetti prolungati sulla certificabilità del prodotto e alterando l'affidabilità dell'operatore. Anche per le infrazioni il RT 16 introduce la distinzione tra infrazioni lievi e gravi.

A queste due tipologie di inadempienze, sulla base del livello di gravità, vengono applicati cinque diversi provvedimenti sanzionatori specifici: richiamo, diffida, soppressione, sospensione della certificazione, sospensione dell'azienda.⁵

Il riscontro di irregolarità o infrazioni origina dall'attività ispettiva degli ODC che si compone di visite ordinarie e straordinarie. Nell'ambito di tali ispezioni si eseguono sopralluoghi aziendali, monitoraggi d'ufficio nonché eventuali campionamenti per l'individuazione di situazioni non conformi alle disposizioni in materia di agricoltura biologica. Sia le visite aziendali che l'eventuale prelievo di campioni sono programmate dagli ODC previa redazione di un piano di controllo annuale con il quale, facendo riferimento all'analisi del rischio di violazione, si stabiliscono il numero e le tipologie di aziende soggette al controllo. Gli approcci attualmente utilizzati dagli ODC operanti nel biologico per la valutazione del rischio di non conformità si basano su una procedura che, dopo aver rilevato la presenza e lo stato dei fattori di rischio per ciascuna azienda in esame, assegna dei punteggi di rischio derivanti da semplici somme algebriche. Tale metodologia presenta alcuni limiti: in primo luogo si considera una ridotta gamma di fattori di rischio, che possono risultare troppo aggregati e scarsamente rappresentativi di tutte le realtà presenti a livello nazionale. In secondo luogo non si sfrutta adeguatamente la possibilità di utilizzare i dati storici circa la presenza di non conformità impedendo di fatto un aggiornamento dei parametri di rischio.

L'obiettivo generale del presente lavoro è quello di individuare, sulla base dei riscontri ottenuti durante le fasi ispettive della certificazione, un approccio per valutare la probabilità di commettere irregolarità o infrazione da parte degli operatori, alla luce dei fattori strutturali e gestionali che li caratterizzano. In particolare i modelli costruiti mediante l'utilizzo di BN possono fornire un valido supporto per la programmazione del piano delle visite ispettive. In situazioni di incertezza, o comunque di conoscenza non completa, le BN possono infatti offrire la possibilità di

⁵ Per maggiori dettagli si rimanda il lettore alla consultazione del rapporto tecnico R.T. 16 – www.accredia.it.

condurre inferenze deduttive ed induttive, su cui basare analisi causali nonché valutazioni di scenario.

2 IL MODELLO PROBABILISTICO

Le applicazioni di modelli basati su BN sono relativamente recenti, e tra i lavori pionieristici possono essere ricordati quelli di Howard e Matheson, (1981), Cooper (1984), Pearl (1986,1988), Andersen et al. (1987), Lauritzen e Spiegelhalter (1988). Horvitz et al. (1988) definiscono una BN come un modello probabilistico utilizzabile per la gestione dell'incertezza in sistemi esperti. In ambito economico agrario peraltro le applicazioni sono state relativamente scarse, e prevalentemente riferite a modelli di supporto decisionale a scopi agronomici. Ad esempio Kristensen et al. (2002) e Tari (1996) hanno applicato le BN per verificare le potenziali implicazioni di metodi di coltivazione a basso impiego di pesticidi o fungicidi per i cereali, mentre Cain et al (2003) ha utilizzato il metodo come supporto per la pianificazione delle produzioni agricole in un'area dello Sri Lanka. Tutte queste applicazioni di fatto impiegano le BN come sistema di supporto decisionale, in grado di gestire in termini probabilistici le diverse opzioni potenzialmente percorribili in un sistema complesso. Gambelli e Bruschi (2010) utilizzano invece un approccio basato sulle BN al fine di individuare i fattori che maggiormente influenzano la probabilità di aziende agricole ad abbandonare il settore biologico.

Formalmente, le BN sono rappresentabili come modelli grafici che descrivono le interazioni tra un set di variabili, o nodi, collegate attraverso una rete di archi direzionali, o frecce, che rappresentano i legami probabilistici tra le variabili stesse. Le BN sono costruite rispettando la condizione di aciclicità del grafo, ovvero in modo che non esistano mai successioni di archi orientati che partono da un nodo e tornino allo stesso nodo.

Le variabili possono essere sia continue che discrete; anche nel caso di variabili continue è però possibile “discretizzate” i valori, definendo ogni nodo come una variabile casuale discreta che assuma un limitato numero di stati mutuamente esclusivi. A ciascun nodo A viene associata una corrispondente tabella di probabilità condizionata:

$P(A|B_1, \dots, B_n)$, dove B_1, \dots, B_n sono i nodi “figli” del nodo originario A , cioè nodi che risultano dipendenti da A .

La logica di funzionamento delle BN è basata sul teorema delle probabilità condizionate, noto come teorema di Bayes. Dati due eventi, A e B , si definisce la probabilità dell'evento A , posto che si verifichi l'evento B , come segue:

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$$

Data la loro struttura, le BN possono essere utilizzate per determinare la probabilità che una variabile casuale assuma uno dei valori che la definiscono, posto che le altre variabili del sistema abbiano assunto determinati valori. Ad esempio, è possibile stimare la probabilità che una variabile assuma determinati valori, una volta definite le realizzazioni campionarie di altre variabili del modello. Tali realizzazioni campionarie, vengono generalmente denominate “evidenze”. Tali evidenze possono essere inserite quindi utilizzate per simulare la probabilità di un determinato stato dei valori assunti dalle variabili della rete. Se per semplicità ci limitiamo ad un caso con due soli nodi, A, B, dato un insieme di evidenze e possiamo definire la distribuzione di probabilità di probabilità che A come segue:

$$P(A|e) = \frac{P(A,e)}{P(e)}$$

dove

$$P(e) = \sum_A P(A,e)$$

Dato un set di variabili V di cui sia ignota la realizzazione campionaria, e un insieme e di evidenze, è possibile determinare la più probabile configurazione di V date le realizzazioni campionarie note e .

La propagazione delle probabilità condizionate può anche essere utilizzata a ritroso, nel caso in cui si sia interessati a individuare quale dovrebbero essere le condizioni che potrebbero con maggior probabilità determinare lo stato $A=a$ e $B=b$. Per una trattazione approfondita dell'argomento si veda, tra gli altri Jensen (1996).

La costruzione di una BN si articola sostanzialmente in due fasi: la costruzione della struttura del grafo aciclico che rappresenta la struttura di connessione tra le variabili, o nodi, e la determinazione delle rispettive probabilità condizionate. Le BN possono essere sviluppate sia a partire da dati statistici sia utilizzando informazioni qualitative o *expert assessments* (Hogeveen et al., 1994). Nel presente studio il modello bayesiano è stato costruito a partire da un database di aziende biologiche, utilizzando la procedura di apprendimento strutturale denominata *Path Condition* (PC). Questo metodo utilizza una procedura statistica basata sul test chi-quadro per testare l'indipendenza condizionale tra le variabili (Spirites et al, 2000). Sostanzialmente il metodo procede iterativamente, ipotizzando dapprima l'ipotesi di un grafo completo in cui cioè tutti i nodi del *network* siano connessi a ciascun altro; successivamente inizia a testare statisticamente l'indipendenza delle relazioni, eliminando i legami tra i nodi che siano risultati indipendenti. L'algoritmo Il metodo PC è utilizzato dal software *Hugin Expert 7* come metodo base; tale algoritmo si è dimostrato efficiente ed affidabile anche nel caso di reti sparse e grafi di grandi dimensioni (Kalisch and Beuhlmann, 2007).

Una volta individuata la struttura del *network*, vale a dire la presenza o assenza di legami tra le variabili, la fase successiva consiste nella determinazione della forza del legame stesso, attraverso il calcolo delle tavole di probabilità condizionata. Anche in questo caso è possibile “apprendere” le probabilità sulla base della distribuzione delle variabili del database o, nel caso in cui questo non fosse disponibile determinarle sulla base di valutazioni qualitative, letteratura, *expert assessment*, ecc. Nel nostro caso ci siamo avvalsi del primo approccio, sfruttando l’algoritmo di apprendimento delle probabilità *Expectation Maximisation* (EM). L’algoritmo EM è basato sul metodi di stima di massima verosimiglianza; in concreto l’algoritmo ricerca il vettore dei parametri che massimizza la probabilità che l’insieme di dati completo possa aver generato il *set* di dati incompleto ed osservato (Dempster et al. 1977). Il modello BN è stato applicato utilizzando il *software Hugin Expert 7*.

3 I DATI UTILIZZATI

Il presente studio è stato realizzato analizzando un campione di aziende agricole italiane assoggettate al Reg. CEE 2092/91 al 31/12/2007; tali dati sono stati forniti dall’Istituto Mediterraneo di Certificazione, ODC delegato dall’autorità competente e accreditato secondo la norma EN 45011. Il campione, pur costituendo circa il 10% delle aziende agricole biologiche italiane, non è da intendersi come statisticamente rappresentativo dell’universo. La maggior parte delle aziende del campione opera infatti nel centro Italia e pertanto presenta caratteristiche strutturali e manageriali tipiche di queste regioni.

La prima fase di reperimento dei dati ha consentito l’individuazione del numero e della tipologia di non conformità commesse da ogni singola azienda. Le aziende prese in esame in questa ricerca sono 3.455; il numero totale di aziende esaminate che commettono almeno un’irregolarità o infrazione è di 1.334, pari al 38,6% del campione. In Tabella 1 vengono riportate le frequenze delle non conformità riscontrate, divise per tipologia.

Tabella 1: Frequenze delle non conformità riscontrate per tipologia:

Tipologia Non Conformità	Frequenza Non Conformità					
	0	1	2	3	4	> 4
Irregolarità lieve	2349	687	271	79	38	31
Irregolarità grave	2907	420	95	19	8	6
Infrazione lieve	3409	42	4	0	0	0
Infrazione grave	3420	33	2	0	0	0
TOTALE NC	2121	710	335	143	58	88

Un ulteriore elemento di interesse emerso durante l’analisi del campione è costituito dal fatto che per la totalità delle imprese che hanno compiuto una infrazione, lieve o grave, è stato anche riscontrato il compimento di almeno una irregolarità. In altre parole, nel campione di imprese a disposizione, l’elemento

dominante in termini di probabilità di riscontro di infrazione è costituito dal riscontro di irregolarità.

Stante la ridotta numerosità di casi relativi a infrazioni gravi (Tabella 1), nel modello probabilistico si è preferito aggregare tali tipologie con le infrazioni lievi in una unica variabile che consideri il riscontro di infrazioni, sia lievi che gravi (INFRAZIONE). Le due variabili irregolarità lieve e irregolarità grave sono invece riportate nel modello bayesiano in maniera distinta (irregolarità lieve= IRR LIEVE; irregolarità grave = IRR GRAVE).

Per ciascuna azienda presa in esame è stato inoltre possibile estrarre diverse variabili aziendali utilizzate nel modello come fattori di rischio di non conformità al regolamento. Di seguito si riportano i fattori di rischio con i relativi acronimi utilizzati nel modello probabilistico.

1) Fattori di rischio strutturali/imprenditoriali.

- età dell'imprenditore (ETÀ);
- superficie agricola utilizzabile (SAU);
- dimensione economica aziendale espressa in termini di Unità di Dimensione Europea (UDE). La dimensione economica è stata determinata attraverso l'applicazione dei redditi standard forniti dalla banca dati INEA. (UDE);
- conduzione dell'azienda. Tale variabile è di tipo dicotomico ed identifica l'eventuale ricorso da parte dell'azienda a terzi per le operazioni colturali (CONTO TERZI);
- presenza del piano delle misure concrete e precauzionali. Il piano delle misure concrete e precauzionali viene redatto da tutte le aziende che, a causa di svariati motivi, risultano particolarmente esposte a rischi di contaminazione del prodotto (es. vicinanza a strade molto trafficate, fabbriche inquinanti, aziende non biologiche che generalmente adottano sistemi di produzione impattanti) (AREE A RISCHIO);
- ottenimento del certificato (licenza) per la vendita dei prodotti. Gli operatori licenziatari sono autorizzati dall' ODC all'immissione sul mercato dei prodotti identificati come "biologici" (LICENZA).

2) Fattori di rischio legati all'ordinamento colturale. Attraverso l'analisi del Piano Annuale delle Produzioni (PAP) è stato possibile identificare la tipologia di colture praticate dalle aziende nonché la superficie di terreno coltivabile destinato alle stesse.

- Fattori di rischio colturali – seminativi e orticole: FRUMENTO DURO, FARRO, ORZO, AVENA, FRUMENTO TENERO, MAIS, GIRASOLE, SOIA, LINO, LENTICCHIA, CECE, PISELLO, FAVA, CARCIOFO, CORIANDOLO, ORTICOLE, AROMATICHE E OFFICINALI;
- Fattori di rischio colturali – colture permanenti: POMACEE, FRUTTETO MISTO, DRUPACEE, KIWI, AGRUMI, OLIVO, VITE DA VINO, VITE DA TAVOLA, NOCE, NOCCIOLO, CASTAGNO, MANDORLO;
- Fattori di rischio colturali – foraggiere: FORAGGERE GRAMINACEE, FORAGGERE LEGUMINOSE, ALTRE FORAGGERE, PASCOLO.

I dati sopra illustrati sono stati utilizzati come variabili (nodi) per la costituzione del modello probabilistico; tutte le variabili sono state rese binarie (0; 1), eccezione

fatta per quelle relative alla superficie aziendale (SAU), all'età dell'imprenditore (ETÀ) e alla dimensione economica aziendale (UDE), per le quali si sono invece previsti diversi stati.

4 RISULTATI E DISCUSSIONE

La rete risultante dalle elaborazioni dell'algoritmo PC utilizzato per l'apprendimento della struttura e delle relative probabilità condizionate viene riportata in Figura 1. Al fine di rendere più agevole l'interpretazione grafica, i nodi sono stati raggruppati rispettando la classificazione delle variabili proposta nel paragrafo precedente. Sulla base della rete così determinata è stato possibile eseguire due tipi di analisi. Nel primo caso si considerano gli effetti sul rischio di infrazione e irregolarità di ciascuna variabile singolarmente considerata, al fine di individuare i fattori con i maggiori impatti marginali sul rischio di non conformità. Nel secondo caso si considerano gli effetti congiunti di un gruppo di variabili, al fine di simulare il rischio di diverse tipologie aziendali. Il numero di possibili combinazioni è ovviamente molto elevato, e in questa sede ci limitiamo a considerare le principali categorie aziendali riscontrabili nel territorio nazionale. In entrambi i casi si valuta come la probabilità media di non conformità è modificata quando si considerano "evidenze" introdotte per singole variabili o gruppi di esse. La probabilità media di non conformità è quella che ci si può attendere dal complesso delle aziende considerate, nel caso in cui non si consideri alcuna specifica caratterizzazione delle variabili del modello (nessuna evidenza). Essa misura quindi il valore atteso di rischio di non conformità derivante dalla distribuzione media delle variabili considerate nel modello.

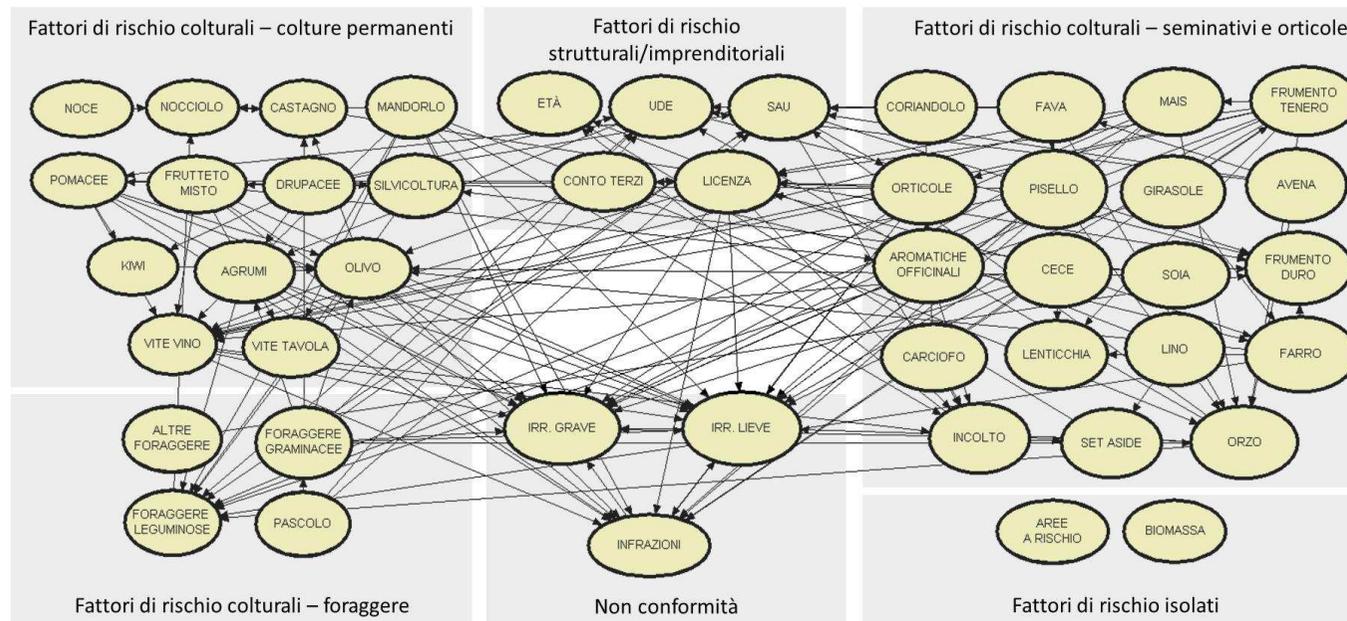


Figura 1 – Bayesian network risultante dal processo di apprendimento strutturale.

Dall'analisi del modello probabilistico emerge con chiarezza che tutte le imprese per le quali sia stata riscontrata un'infrazione abbiano anche compiuto almeno una irregolarità, grave o lieve. Questo primo risultato comporta delle notevoli implicazioni da un punto di vista interpretativo del ruolo delle caratteristiche strutturali delle imprese in termini di impatto sulla probabilità di commettere infrazioni e irregolarità. Di fatto, nel contesto del database a disposizione, gli elementi strutturali e gestionali che risultano avere maggiore impatto sulla probabilità di riscontro di irregolarità, sono necessariamente gli stessi che presentano il maggiore impatto sulla probabilità di riscontro delle infrazioni. In altre parole, gli effetti sulla variazione delle probabilità di commettere infrazione sono "dominati" dalle variabili IRR LIEVE e IRR GRAVE: ipotizzando che un operatore abbia commesso una irregolarità, la probabilità che si riscontri un'infrazione presso lo stesso operatore aumenta, e viceversa.

In particolare, la probabilità media di riscontro di infrazione è pari al 2,17 % (Tabella 3). Sulla base delle stime delle probabilità condizionate applicate al modello, tale probabilità quasi quadruplica nel caso in cui si ipotizzi che un'azienda abbia anche commesso una irregolarità di tipo grave.

L'inserimento nel modello di tre livelli di non conformità (irregolarità lievi, gravi e infrazioni) rende possibile cogliere con maggiore precisione le interazioni tra i tipi di non conformità, per analizzare se e come il fatto di commettere un'irregolarità lieve comporti una variazione della probabilità di commettere un'irregolarità grave, oltre che valutarne l'impatto sulla probabilità di riscontrare un'infrazione.

Accanto alla valutazione delle forti interazioni reciproche evidenziate dalle diverse tipologie di non conformità, si è proceduto alla misurazione degli impatti sul rischio di non conformità derivanti anche dagli aspetti strutturali e manageriali delle imprese considerate.

Per meglio valutare gli impatti dei diversi fattori di rischio singolarmente considerati, è utile confrontarli con lo schema dei fattori di rischio indicati da ACCREDIA nel rapporto tecnico RT16. Il confronto appare interessante poiché i due schemi sono il risultato di due diversi approcci: il rapporto RT 16 delinea sulla base di valutazioni tecnico-agronomiche una graduatoria di elementi che possono incrementare o ridurre rischio atteso di non conformità; il modello probabilistico segue invece un approccio empirico determinando il rischio di non conformità sulla base delle effettive violazioni riscontrate. Nella Tabella 2 si propone una sintesi comparativa tra le due classificazioni. I risultati derivanti dal modello probabilistico sono stati tradotti in termini qualitativi, convertendoli in una scala qualitativa che indica l'impatto positivo (+), nullo (0) o negativo (-) sul rischio di non conformità. Dal confronto emerge come il modello probabilistico confermi la rischiosità di alcuni elementi strutturali e gestionali indicati come critici nel RT16, in particolare la presenza di un operatore licenziatario, (che è emerso nettamente come un elemento critico dal modello) la presenza di produzioni a rischio OGM, che nel nostro modello sono di fatto rappresentate da mais e soia, e una elevata superficie aziendale. Peraltro altri elementi indicati nel RT 16 come ad impatto positivo sul rischio, quali il ricorso

al contoterzismo, l'ubicazione in area a rischio, sono risultati nel nostro modello come del tutto ininfluenti. Venendo alle singole produzioni, va precisato che il modello probabilistico è stato strutturato in termini molto più dettagliati rispetto alle indicazioni del RT16, ma alcune categorie generali sono comunque confrontabili. Tra le produzioni emergono le maggiori differenze tra i due approcci: ad eccezione della concordanza per quanto riguarda l'impatto positivo sul rischio delle produzioni frutticole, viticole e delle erbe aromatiche, per le altre produzioni in particolare cerealicole e foraggere i risultati del modello probabilistico sono opposti rispetto alle indicazioni del RT16, evidenziando per entrambi un aumento del rischio associato a tali colture. Una parziale concordanza tra i due approcci è riscontrabile per quanto riguarda le produzioni olivicole, indicate come ad impatto negativo sul rischio anche dal RT16 e alle quali il modello probabilistico associa una riduzione del rischio di non conformità. E' infine interessante notare come la presenza di produzioni di agrumi e ceci riduca il rischio stimato nel modello probabilistico.

Tabella 2 – Fattori di rischio di non conformità: confronto tra risultati del modello probabilistico e indicazioni RT 16.

Variabile	Modello probabilistico	RT16
<i>Aspetti strutturali e gestionali</i>		
Operatore licenziatario	+	+
Irregolarità lieve	+	0
Irregolarità grave	+	+
Irregolarità gravi anno prec.	nd	0
Infrazioni anno prec.	nd	+
Dimensione economica	+	nd
Superficie aziendale >50 ha	+	+
Superficie aziendale tra 15 e 50 ha	0	0
Superficie aziendale < 15	0	-
Ricorso a contoterzismo	0	+
Ubicazione in area a rischio	0	+
Presenza produzioni a rischio OGM	+(mais e soia)	+
<i>Produzioni</i>		
Foraggiere	+	-
Erbe aromatiche e officinali	+	+
Frutteto	+	+
Pisello proteico	+	nd
Soia	+	nd
Orzo	+	nd
Grano tenero	+	nd
Mais	+	nd
Cereali	+	-
Vite da vino	+	+
Orticole	+	0
Olive	-	0
Cece	-	nd
Agrumi	-	nd

+: incremento di rischio; 0: nessun incremento di rischio; -: riduzione di rischio; nd: non disponibile.

Va rilevato che nel RT16 accanto agli aspetti sopra considerati, vengono citati tra gli operatori con più alto rischio quelli per i quali siano state riscontrate precedenti infrazioni, e tra quelli a rischiosità invariata quelli per i quali siano state riscontrate precedenti irregolarità gravi. La non disponibilità di dati storici non ha permesso di considerare nel modello probabilistico il ruolo di eventuali precedenti infrazioni, ma solo la contemporanea presenza di irregolarità lievi o gravi nello stesso anno.

Il modello probabilistico, oltre a fornire un valido supporto per l'individuazione dei principali fattori che influiscono sulla probabilità di commettere non conformità, permette anche la valutazione del livello di rischio di non conformità specifico per diverse tipologie di aziende, considerando l'effetto congiunto di diverse combinazioni di variabili sul rischio di infrazione. In tabella 3 vengono presentati i risultati di cinque diverse simulazioni rappresentative dei principali comparti produttivi biologici presenti in Italia. Per ciascuna simulazione si è voluto analizzare in dettaglio l'effetto degli specifici fattori di amplificazione del rischio di non conformità individuati (Tabella 2), quali la presenza di operatore licenziatario e il riscontro di irregolarità lievi e/o gravi. In quanto segue vengono brevemente descritte le tipologie aziendali utilizzate per le simulazioni riportate in Tabella 3:

- Azienda cerealicola: azienda cerealicola di 24 ha condotta da un imprenditore di 60 anni in zona collinare che adotta una rotazione comprendente cereali autunno-vernini, colture primaverili e foraggiere permanenti (grano duro, farro, girasole, erba medica, sovescio). La meccanizzazione è prevalentemente basata su mezzi aziendali, eccezione fatta per le operazioni di raccolta della granella che è affidata a contoterzisti.
- Azienda zootecnica: non avendo a disposizione i dati relativi alle UBA, al fine di individuare un'ipotetica azienda zootecnica ubicata nelle colline pedemontane del centro Italia, si è qui ricostruito una tipologia di azienda zootecnica sulla base delle sole variabili culturali. Oltre alla presenza di pascolo, che tra tutte risulta la variabile che meglio funge da indicatore per le aziende zootecniche, è stata considerata la presenza di colture foraggiere leguminose e graminacee, cereali autunno-vernini (orzo e avena) nonché proteiche da granella (favino e pisello proteico) per garantire l'autosufficienza alimentare del bestiame allevato. L'azienda risulta condotta da un imprenditore di 50 anni e non utilizza conto terzi per lo svolgimento delle operazioni culturali.
- Azienda frutticola: azienda di 5 ha condotta da un imprenditore di 45 anni specializzata nella coltivazione di frutta biologica in pianura. Coltivazione di specie drupacee e pomacee allevate con ausilio di sola meccanizzazione aziendale per la potatura, trattamenti e raccolta del prodotto.
- Azienda olivicola: azienda di 15 ha specializzata nella produzione di olive da olio con sesti di impianto intensivi e con meccanizzazione aziendale per la potatura e la raccolta del prodotto. L'imprenditore è di 45 anni e l'azienda è ubicata in zona pianeggiante.
- Azienda non specializzata (policultura): azienda di 20 ha condotta da un imprenditore di 65 anni, ubicata in zona collinare e con prevalente ordinamento cerealicolo – industriale (erba medica, favino/pisello, grano duro, girasole). L'azienda presenta anche vite e olivo in misura marginale rispetto ai seminativi. L'azienda utilizza conto terzi per le operazioni di semina delle colture industriali e per le operazioni di raccolta delle granelle.

I risultati delle simulazioni riportati in Tabella 3 evidenziano che quella zootecnica è la tipologia aziendale alla quale si associa la maggior probabilità di commettere

irregolarità lieve. L'influenza di alcuni fattori di rischio come un'elevata SAU e dimensione economica, uniti alla presenza di foraggiere nelle rotazioni, incrementa la probabilità di irregolarità lieve di circa due volte rispetto alla media; da notare tuttavia che per questa simulazione non si riscontrano aumenti considerevoli della probabilità di commettere irregolarità grave. Anche il rischio di infrazione è superiore di circa il 25% rispetto alla media di tutte le aziende. Le aziende olivicole e non specializzate mostrano d'altro canto i valori di minor rischio complessivo. Al contrario di quanto atteso, anche le aziende frutticole e le aziende non specializzate, al pari di quelle olivicole, non evidenziano un particolare aumento rischio di non conformità rispetto alla media del campione; nonostante la complessità della gestione aziendale tali tipologie aziendali dimostrano rischi relativamente elevati di commettere irregolarità lieve ma solo in concomitanza della variabile relativa all'utilizzo della licenza per la vendita dei prodotti (licenziatario). Per quanto riguarda l'azienda cerealicola, non si riscontrano particolari scostamenti di rischio di non conformità rispetto alla situazione media nel caso non si ipotizzi la presenza di operatore licenziatario. La situazione si modifica però radicalmente nel caso si consideri la presenza di operatori licenziatari. Confermando quanto indicato nella Tabella 2 e nel RT 16, gli operatori licenziatari infatti dimostrano di costituire un fattore di rischio particolarmente rilevante per la prevalenza delle tipologie aziendale, ad eccezione del caso della azienda di tipo misto. La Tabella 3 illustra le variazioni di probabilità dei diversi tipi di non conformità in corrispondenza di 3 ipotesi, distinguendo per ciascuna il caso di imprese gestite da operatori licenziatari e non. Nella prima ipotesi (H1) si considerano gli effetti sulle probabilità di non conformità dovute esclusivamente alle diverse tipologie aziendali, considerando che non si siano riscontrati nessuno dei diversi tipi di non conformità; nella seconda ipotesi (H2) si considerano le probabilità di IRR GRAVE e INFR ipotizzando che si sia riscontrata una IRR LIEVE; infine nella terza ipotesi (H3) si considerano le probabilità di INFR nel caso in cui si siano riscontrate IRR LIEVI e IRR GRAVI.

L'impatto di questo fattore sul rischio di non conformità risulta peraltro non uniforme rispetto alle diverse tipologie aziendali e alle diverse combinazioni di non conformità. Ad esempio mentre per le aziende cerealicole l'effetto congiunto di irregolarità e operatore licenziatario è marcato e presente per tutte le diverse combinazioni considerate, ponendo la tipologia cerealicola come quella caratterizzata da maggiore rischiosità, nel caso delle aziende zootecniche l'effetto della variabile "licenziatario" è di fatto più limitato e circoscritto alle irregolarità di tipo lieve e grave. Considerazioni analoghe possono riferirsi anche alla tipologia di azienda non specializzata. Nel caso dell'azienda di tipo frutticolo, in presenza dell'operatore licenziatario in concomitanza con la presenza di irregolarità lievi o gravi già riscontrate, il rischio di infrazione subisce una seppur lieve riduzione.

Anche per quanto riguarda le simulazioni dei casi aziendali, per tutti i modelli considerati, il rischio di infrazione si conferma essere fortemente connesso con il riscontro di irregolarità. In linea generale, un operatore che abbia commesso un'irregolarità lieve ha un rischio doppio di commettere infrazione, mentre coloro che

hanno commesso una irregolarità grave vedono quadruplicato il rischio di infrazione. Sebbene tale effetto sia tendenzialmente trasversale a tutti i casi aziendali considerati, in alcuni gli effetti sono particolarmente evidenti specie se in concomitanza con la presenza della qualifica di operatore licenziatario. Ad esempio nel caso dell'azienda non specializzata si passa da un rischio di infrazione medio inferiore alla media, ad un valore di rischio particolarmente elevato in caso di presenza di operatore licenziatario che abbia anche commesso una irregolarità di tipo grave. Effetti proporzionalmente meno marcati della presenza di irregolarità sul rischio di infrazione sono riscontrabili invece per le aziende frutticole.

Tabella 3 – Probabilità di non conformità per tipologia aziendale e tipologia di operatore; valori percentuali.

		Prob media	Operatore licenziatario	Probabilità di non conformità per tipologia aziendale				
				Az. Cerealicola	Az. Zootecnica	Az. Olivicola	Az. Frutticola	Az. Mista
H 1: assenza non conformità riscontrate	Prob. IRR LIEVE	31,70	No	40,15	69,38	22,99	26,35	37,61
			Si	84,63	72,50	48,41	47,68	37,65
	Prob. IRR GRAVE	15,86	No	15,91	15,78	9,93	9,95	6,83
			Si	71,97	17,27	20,87	22,85	6,00
	Prob. INFR	2,17	No	2,37	3,13	1,61	1,74	1,50
			Si	7,39	3,30	2,89	2,86	1,58
H 2: IRR LIEVE riscontrata	Prob. IRR GRAVE	26,55	No	26,67	22,75	24,00	27,59	12,94
			Si	78,09	23,82	31,12	25,89	14,26
	Prob. INFR	3,88	No	4,55	4,26	4,35	4,62	2,58
			Si	8,35	4,34	4,88	4,49	3,93
H 3: IRR LIEVE e IRR GRAVE	Prob. INFR	7,85	No	8,22	9,97	7,59	8,53	7,04
			Si	9,53	9,97	8,48	7,50	9,84

CONSIDERAZIONI CONCLUSIVE

Il modello probabilistico sviluppato nel presente contributo offre la possibilità di condurre inferenze mirate da un lato all'individuazione dei principali elementi d'impatto sulle probabilità di riscontro di non conformità, e dall'altro alla simulazione del rischio di non conformità relativo a varie combinazioni di elementi strutturali e gestionali aziendali. Esso permette di individuare a priori, sulla base di ipotesi in merito alle caratteristiche strutturali ed economiche delle aziende, il livello di rischio relativo a ciascuna azienda tipo presa in esame.

L'analisi del campione di aziende biologiche prese in considerazione in questo contributo ha messo in luce diversi fattori di rischio capaci di influenzare la probabilità di commettere irregolarità e infrazioni alla vigente normativa. Tra le variabili considerate nel modello, sono emerse tra le più rilevanti sia alcune legate ad aspetti "manageriali" come in particolare la presenza di operatori licenziatari, sia altre legate alla presenza in azienda di particolari ordinamenti produttivi, in particolare frutteto e orticole, mais e leguminose. Inoltre è emerso con chiarezza il particolare effetto moltiplicativo del rischio di irregolarità gravi e/o di infrazioni, dovuto al

riscontro di irregolarità lievi. In particolare emerge come tutte le imprese per le quali è stata riscontrata una infrazione, hanno anche compiuto almeno una irregolarità lieve o grave. Un confronto con i fattori di rischio indicati nel RT16 ha evidenziato come se per gli elementi gestionali e strutturali vi sia una sostanziale omogeneità di indicazione, per quanto riguarda la valutazione della rischiosità di infrazione legata alle singole produzioni il modello probabilistico abbia invece ottenuto risultati spesso opposti alle indicazioni del RT16.

La flessibilità del modello probabilistico adottato ha anche permesso di simulare la rischiosità di particolari tipologie aziendali, ottenute combinando i diversi fattori di rischio. Le tipologie caratterizzate da maggior rischio di non conformità sono risultate le aziende di tipo zootecnico, mentre le aziende olivicole e frutticole si attestano su valori medi di rischio bassi. In tutti i casi simulati, ad eccezione dell'azienda non specializzata, viene confermato il ruolo cruciale della presenza di operatore licenziatario come fattore di amplificazione del rischio di non conformità. E' peraltro interessante notare come l'effetto congiunto di irregolarità e operatore licenziatario non sia però uniforme tra le diverse tipologie aziendali: mentre nel caso delle aziende frutticole non si notano effetti cumulativi sulla rischiosità generale, nel caso delle aziende cerealicole l'impatto diviene invece particolarmente marcato. Quest'ultimo risultato in particolare appare particolarmente rilevante sia perché tale tipologia aziendale risulta piuttosto diffusa, sia perché le aziende genericamente classificate come cerealicole sono considerate nel RT 16 di ACCREDIA a basso rischio di non conformità.

Il modello delle BN utilizzato in questo lavoro si è dimostrato uno strumento potente e flessibile, potenzialmente utilizzabile dagli ODC per mettere in atto un sistema d'ispezione efficiente e basato sul rischio effettivo di non conformità, ma anche dalle autorità pubbliche di controllo (in Italia il dipartimento ICQRF del MIPAF) per l'analisi dell'efficacia e dell'affidabilità di detti ODC. La possibilità di condurre inferenze deduttive ed induttive, su cui basare analisi di scenario e valutazioni di tipo diagnostico, unita alla capacità di aggiornare in maniera automatica il *database* utilizzato per l'apprendimento, rendono tale metodologia particolarmente interessante. L'applicazione empirica dei modelli ha mostrato come le classi di rischio utilizzate al momento dagli ODC accreditati ACCREDIA siano ragionevoli. Tuttavia, l'uso combinato di un modello bayesiano costruito su dati storici costantemente aggiornati da nuova evidenza empirica consistente nelle ispezioni condotte annualmente, potrebbe permettere un aggiornamento dinamico dei fattori di rischio e delle classi di rischiosità degli operatori, in modo da rendere il sistema più efficace e meno costoso.

Nonostante la carenza di letteratura al riguardo, è infatti ragionevole credere che una quota significativa dei costi di certificazione potrebbe di fatto dipendere da una ridotta efficienza dello stesso sistema di certificazione, specie nelle fasi di ispezioni lungo la filiera (Commins, 2004).

Future possibili evoluzioni dell'approccio proposto nel presente lavoro potranno orientarsi verso la quantificazione dei potenziali incrementi di efficienza di un sistema

di controllo *risk based*, offrendo una valutazione della potenziale riduzione di costi transazionali legati al processo di certificazione.

RIFERIMENTI BIBLIOGRAFICI

Andersen S.K., Jensen F.V., Olesen K.C. (1987), "The HUGIN core preliminary considerations on systems for fast manipulations of probabilities", in atti del workshop on inductive reasoning: managing empirical information in AI-Systems, Roskilde (Denmark), aprile 1987.

Cain J.D., Jinapala K., Makin I.W., Somaratna P.G., Ariyaratna B.R., Perera L.R. (2003), "Participatory decision support for agricultural management. A case study from Sri Lanka", in *Agricultural Systems*, 76, 457-482.

Commins K. (2004), "Overview of current status of standards and conformity assessment systems", Working Paper, UNCTAD/IFOAM/FAO, International Task Force on Harmonization and Equivalence in Organic Agriculture. Available at: <http://orgprints.org/00003109>.

Cooper G.F. (1984), "NESTOR: A computer based medical diagnostic aid that integrates causal and probabilistic knowledge", Technical report HHP 84-48, *Medical Computer Science Group, Stanford University*, Stanford, California, 1987.

Dempster P., Dempster N., Laird M., Rubin D. B. (1977), "Maximum likelihood from incomplete data via the EM algorithm", in *Journal of Royal Statistical Society*, 39, 1-38.

Gambelli D., Bruschi V. (2010), "A bayesian network to predict the probability of organic farms' exit from the sector: a case study from Marche, Italy", in *Computer and Electronics in Agriculture*, 71, 1, 22-31.

Hogeveen H., Thiesson B., Andersen S.K., Schukken Y.H., Noordhuizen-Stassen E.N. (1994), "Approaches to the building of bayesian belief networks for disease diagnosis", in: Rowlands G.J. – Kyule M.N. – Perry B.D. (1994): Proceedings of the 7th international symposium on veterinary epidemiology and economics, Nairobi, Kenya, in the *Kenya Veterinarian* 18, 2, 411-414.

Howard R.A., Matheson J.E. (1981), "Influence diagrams", in: Howard R.A., Matheson J.E. (1981), "Readings in decision analysis". Strategic Decisions Group, Menlo Park, California, pp. 763-771.

Horvitz E.J., Breese J.S., Henrion M. (1988), "Decision theory in expert systems and artificial intelligence", in *International Journal of Approximate Reasoning*, 2, 247-302.

Jensen F. (1996), *An introduction to bayesian networks*, New York, Springer, 113- 120.

Kalisch M., Beuhlmann P. (2007), "Estimating high dimensional directed acyclic graphs with the PC-algorithm", *Journal of Machine Learning Research*, 8, pp 613-636.

Kristensen K., Rasmussen I.A. (2002), "The use of a bayesian network in the design of a decision support system for growing malting barley without use of pesticides", in *computers and electronics in agriculture*, 33, 197-217.

Lauritzen S.L., Spiegelhalter D.J. (1988), "Local computations with probabilities on graphical structures and their application to expert systems", in *Journal of Royal Statistical Society, Series B*, 50(2), 157-224.

Measures M., Padel S. (2006), Comments on commission proposal for total revision of the EU regulation 2092/91, disponibile su www.organicadvice.org.uk/EU2092/91.doc.

Pearl J. (1986), "Fusion, propagation, and structuring in belief networks", in *Artificial Intelligence*, 29(3), 241-288.

Pearl J. (1988), *Probabilistic reasoning in intelligent systems: networks of plausible inference*, San Francisco, Morgan Kaufmann Publishers.

Schmid O., Dabbert S., Eichert C., González V., Lampkin N., Michelsen J., Slabe A., Stokkers R., Stolze M., Stopes C., Wollmuthová P., Vairo D., Zanolini R. (2008), Organic action plans. Development, implementation and evaluation. A resource manual for the organic food and farming sector. Research Institute of Organic Agriculture (FiBL); CH-Frick and European Union Group of the International Federation of Organic Agriculture Movements (IFOAM), Brussels, disponibile su <http://orgprints.org/13481>.

Seppänen L. (2004), "Local and organic food and farming around the Baltic Sea", in *Ekologiskt lantbruk*, 40, 93-97, disponibile su <http://orgprints.org/6185/>.

Spirites P., Glymour C., Scheines R. (2000), *Causation, prediction and search*, 2nd ed., New York, MIT Press.

Tari F. (1996), "A bayesian network for predicting yield response of winter wheat to fungicide programmes", in *Computers and Electronics in Agriculture*, 15, 111-121.

Summary

Risk-based inspection system for the organic certification: a bayesian networks approach.

The existing method of certification in organic agriculture system, which requires a periodical inspection for all the operators, is inefficient due to the high cost of these controls. A risk based decision support system, that could assist the inspection body during the planning of the annual inspection visits, is advocated to be more cost-effective and efficient. The risk based decision support system is constructed as a Bayesian network; the models incorporate the factors that influence risk of irregularity and analyse their effects by determining probability of non-compliance. Empirical findings using a sample of Italian data on inspection of organic farms, support the idea that the current risk categories used by control bodies in Italy are reasonable, but could be recursively updated by using a Bayesian network model and incremental inspection evidence.

JEL Code: C41

Key words: Organic Certification, Risk-based Inspections, Risk Modelling, Discrete-choice Models, Bayesian Networks.

Corresponding Author: Raffaele Zanoli - Dipartimento di Scienze Agrarie, Alimentari e Ambientali (Università Politecnica delle Marche) - via breccie bianche 60131 Ancona (Italy) – Qualifica: professore ordinario – tel. (+39) 071-2204929 - e-mail: zanoli@agrecon.univpm.it.