

# Bedre Beslutninger med Bayesianske Netværk

Dina Research Report No. 72 • July 1998

Erik Jørgensen<sup>1</sup> & Steffen L. Lauritzen<sup>2</sup>

Denne artikel er udgivet i Statens Jordbrugs- og Veterinærvidenskabelige Forskningsråds særnummer af *Naturens Verden* og bedes refereret som:

Jørgensen, E. & S. L. Lauritzen (1998).  
Bedre Beslutninger med Bayesianske Netværk. *Naturens Verden*. 7. p. 280-287.

Rapporten er tilgængelig som en PostScript fil på World Wide Web med  
URL: <http://www.sp.dk/~ej/BayesNet/Bayes.ps>

---

<sup>1</sup>Biometry Research Unit  
Danish Institute of Agricultural Sciences  
P.O. Box 50  
DK-8830 Tjele  
Tel: + 45 8999 1230  
Fax: + 45 8999 1669  
E-mail: [ej@dina.sp.dk](mailto:ej@dina.sp.dk)  
WWW: <http://www.sp.dk/~ej>

<sup>2</sup>Department of Mathematics,  
Aalborg University  
Fredrik Bajers Vej 7E,  
DK-9220 Aalborg, Denmark  
Tel: +45 9635 8858  
Fax: +45 9815 8129  
E-mail: [steffen@math.auc.dk](mailto:steffen@math.auc.dk)  
WWW: <http://www.math.auc.dk/people/staff/steffen.html>

Bayesianske net har et væld af praktiske anvendelser. De anvendes til at finde årsagen til printerfejl på din PC, til overvågning af styrerakter i rumfærgen Challenger og til medicinske diagnoser. Også inden for jordbruget er mulighederne oplagte. Vi forsøger at åbne et vindue til nettens hemmeligheder. Bayesianske net er udviklet sidst i firserne baseret på klassisk sandsynlighedsregning, kombineret med moderne grafteoretiske algoritmer.

Danske jordbrugsforskere var blandt de første til at interessere sig for teknikken igennem forskningssamarbejdet DINA. Microsoft Research har forlængst oprettet en forskningsafdeling, som udvikler nettens anvendelsesmuligheder i computerteknologien.

Hewlett-Packard har valgt at samarbejde med danske forskere: HP har oprettet et forskningslaboratorium på Aalborg Universitet og har investeret i Aalborg-firmaet HUGIN Expert A/S.

## 1 Er soen med grise?



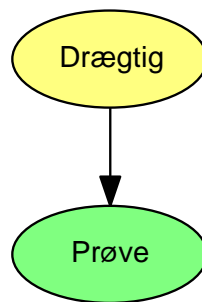
Figur 1: Er denne so med grise? *Foto E. Keller Nielsen*

En landmand med husdyrhold skal så hurtigt som muligt konstatere, om et dyr, som er blevet parret (løbet) nu også er blevet drægtigt.

Hvis dyret ikke er drægtigt, skal han observere dyret ekstra omhyggeligt, så han opdager, når det kommer i brunst igen. Hvis dyret er drægtigt, skal landmanden tilpasse fodringen, så fostrenes behov opfyldes. Hos køer skal mælkeproduktionen standses i god tid inden fødslen. I svineholdet skal man sikre sig, at der er plads til soen når den skal føde (fare), enten i et nyligt rengjort staldafsnit eller i en ledig farehytte.

Man foretager derfor ofte en drægtighedsprøve, der svarer helt til en graviditetsprøve hos mennesker. Der er flere forskellige metoder til at foretage prøven. Metoderne omfatter hormonmålinger i blod eller urin og ultralydsscanninger.

Fig. 2 viser et bayesiansk net til hjælp for drægtighedsundersøgelsen. Formålet med nettet er at fortælle landmanden, hvor sandsynligt det er, at soen eller koen er drægtig, når han kender prøvens udfald.



Figur 2: Bayesiansk net til drægtighedsundersøgelse af husdyr. Knuden *Drægtig* kan være i en af tilstandene *Nej* eller *Ja*, og knuden *Prøve* kan være *Positiv* eller *Negativ*.

### 1.1 Nettets elementer

Bayesianske net er baseret på såkaldte *grafiske modeller*. De kan modellere meget komplekse systemer selv om de er bygget af enkle 'byggeklodser'. En fordel ved den simple opbygning er, at det er let at forstå modellerne. Byggeklodserne præsenteres nedenfor.

**Grafen** Selv om nettet er simpelt, indeholder det alligevel mange elementer af modellerne. Ellipserne i Fig. 2 viser såkaldte *knuder* i grafen. Til hver knude knyttes et sæt af forskellige tilstande. Knuden *Drægtig* har tilstandene *Nej* og *Ja*. Tilsvarende har knuden *Prøve* tilstandene *Positiv* og *Negativ*, svarende til drægtighedsprøvens udfald. De to knuder er forbundet med en linje, en såkaldt *kant*. Pilen angiver kantens retning.

En knude med indgående pile fra andre knuder er *barn* af disse knuder, og disse er til gengæld *forældre* til knuden. Knuden *Drægtig* er den eneste forælder til knuden *Prøve*, der er barn af *Drægtig*. Kantens retning er udtryk for en årsagssammenhæng. Udfaldet af drægtighedsprøven er en konsekvens af dyrets tilstand og ikke omvendt.

**Sandsynligheder** Ud over selve grafen skal man angive sandsynligheder for, at knuderne befinder sig i de forskellige tilstande. For knuder, der ikke har nogen forældre (som *Drægtig*), angives disse blot direkte, baseret på ekspertviden eller tidligere observationer. Eksempelvis bliver omkring 85% af søerne drægtige, mens det for køer er en noget lavere procentdel (omkring 50%). I en grafisk model for løbning af en sø vil man derfor angive sandsynligheden 0.15 for tilstanden *Nej* og 0.85 for tilstanden *Ja*.

For knuder med forældre (som *Prøve*) angives sandsynligheder for de forskellige tilstande når man kender tilstandene for de knuder, som er forældre. I vores eksempel skal sandsynligheden for, at prøven viser et positivt udfald, angives henholdsvis når dyret er drægtigt, og når dyret ikke er drægtigt. Fra andre drægtighedsundersøgelser af søer kendes typiske værdier for disse sandsynligheder. Prøven giver positivt udslag i 95% af tilfældene, hvis soen er drægtig. Hvis soen ikke er drægtig vil der kun være positivt udslag i 30% af tilfældene. Sandsynlighederne afhænger selvfølgelig af, hvilken prøvemethode man anvender.

## 1.2 Brug af nettet

Når grafen og sandsynlighederne er angivet, er nettet klar til brug. En landmand foretager en drægtighedsprøve med negativt udfald og vil kende dyrets virkelige tilstand. Han er så interesseret i at drage slutninger imod pilens retning, fra barn til forælder. Ved specifikation af nettet fulgte vi hele tiden pilens retning. Takket være Bayes formel (se boks om Bayes) er det dog muligt at 'vende' beregningerne og slutte i modsat retning af pilen.

Det bayesianske net foretager beregningerne. Selv med et negativt udfald af prøven er der en sandsynlighed på 29% for at soen er drægtig. Er prøvens udfald positivt, stiger sandsynligheden til gengæld til omkring 95%.

### BAYES og hans formel

Thomas Bayes (1702–1761) var engelsk teolog og matematiker. Han var den første som anvendte sandsynlighedsregning til at drage slutninger om fremtidige begivenheder på basis af erfaringer. I den forbindelse anvendte han blandt andet en formel, der nu er kendt som *Bayes formel*. I moderne udgave ser den således ud:

$$P(D|R) = \frac{P(R|D)P(D)}{P(R|D)P(D) + P(R|\bar{D})(1 - P(D))}.$$

I eksemplet med drægtighedsprøven angiver  $P(R|D)$  og  $P(R|\bar{D})$  sandsynligheden for prøvens resultat,  $R$ , i de tilfælde, hvor soen er hhv. drægtig eller ikke, og  $P(D)$  angiver sandsynligheden for at soen er drægtig *før* prøveresultatet er kendt. På venstre side af lighedstegnet angiver  $P(D|R)$  den ønskede sandsynlighed for at soen er drægtig *efter* at prøvens resultat er kendt.

Et bayesiansk net udnytter en version af formlen til at beregne tilsvarende sandsynligheder i mere indviklede situationer.

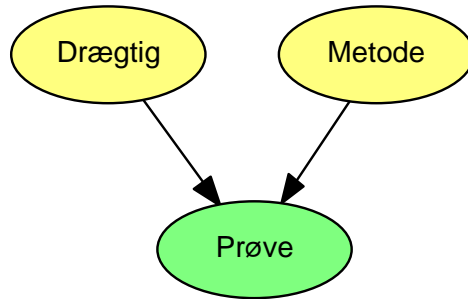
Thomas Bayes udgav aldrig selv sin afhandling "*Essay Towards Solving a Problem in the Doctrine of Chances*". Den blev først udgivet i 1763 efter hans død.

### 1.3 Udvidelsesmuligheder

Det viste eksempel er meget forenklet, men det er let at udvide nettet. Som nævnt afhænger prøvens nøjagtighed af den anvendte metode. Dette kan inddrages i modellen ved at tilføje en ekstra knude i diagrammet, som vist i Fig. 3.

Knuden *Metode* kan have en tilstand for hver af de mulige metoder, det vil sige en for hormonmåling og en for ultralydsscanning. Ved brug af nettet kan man både inddrage viden om den anvendte metode og om prøvens udfald.

Det udvidede net ville være relevant også selvom kun en enkelt metode blev anvendt. For selv med samme metode vil forskellige personer kunne opnå forskellig nøjagtighed. Så kan tilstandene i *Metode* betegne den person, som foretager undersøgelsen. Nettet kan så benyttes til at lære, hvor gode personer er til at anvende metoden. Dette kræver dog at nettet kombinerer undersøgelser på flere søer.

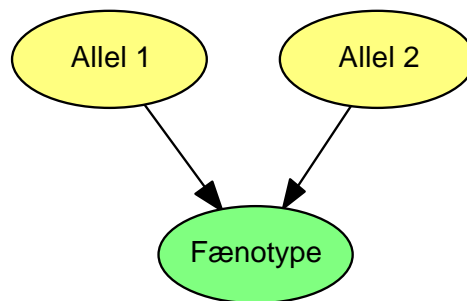


Figur 3: Bayesiansk net for en drægtighedsprøve med to prøvemethoder. Knuden *Metode* kan for eksempel have tilstandene *Hormonmåling* og *Ultralyd*.

## 2 Øjenfarve hos en familie

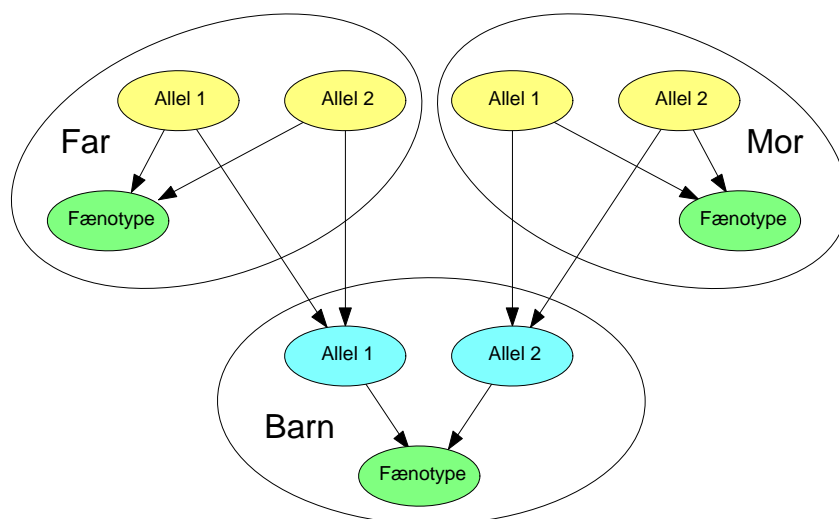
Et andet område, hvor de grafiske modeller har været anvendt, er det genetiske. For hvert af vore arvelige anlæg (gener) har vi 2 kopier (alleller), et der stammer fra vores far og et fra vores mor. Som eksempel kan vi lade  $B$  betegne et gen for brune øjne og  $b$  et gen for blå øjne. De tre mulige parvise kombinationer ( $BB$ ,  $Bb$  og  $bb$ ) af disse to arvelige anlæg kaldes *genotyper*. *Fænotypen* er den synlige konsekvens af genotypen, nemlig at øjnene er blå eller brune. Hvis genotypen er  $bb$ , er øjenfarven *Blå*. Hvis genotypen er  $BB$ , er øjenfarven *Brun*. Da arveanlægget for brune øjne er dominant, vil også genotypen  $Bb$  have fænotype *Brun*.

**Simpel genetisk model** Fig. 4 viser den simple genetiske model som et bayesiansk net. Genotypen er angivet med to knuder, en for hver allel, og fænotypen har begge disse alleller som forældre. Hver allel har to tilstande  $b$  og  $B$ . Knuden *Fænotype* har tilstandene *Blå* og *Brun*.



Figur 4: Bayesiansk net som angiver sammenhæng mellem arveanlæg og fænotype.

**Model for nedarvning** Det simple net kan let udbygges til et bayesiansk net for en familie med far, mor og barn ved at lave et net for hver af de tre personer og hæfte nettene sammen som vist i Fig. 5.



Figur 5: Bayesiansk net for nedarvning af arveanlæg fra far og mor til et barn.

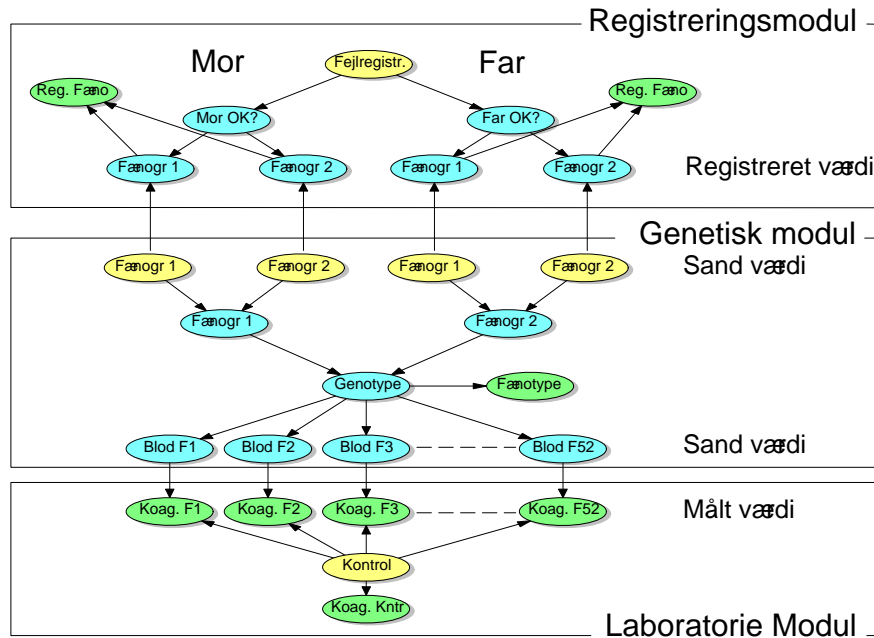
I forhold til Fig. 4 er der tilføjet kanter mellem faderens to alleller og barnets *Allel 1* samt mellem moderens alleller og barnets *Allel 2*. De tilhørende sandsynligheder bestemmes af, at den allel, som barnet arver fra faderen, er valgt tilfældigt blandt faderens to alleller, og tilsvarende for den allel, der arves fra moderen.

Ved at sammenbygge disse byggeklodser i hele stamtræer kan man anvende bayesianske net til at identificere sygdomsgener, til genetisk rådgivning og til planlægning af undersøgelsesprogrammer for arvelige sygdomme.

### 3 BOBLO

Det netværk, der blev beskrevet i det foregående afsnit er kernen i en af de første udnyttelser af bayesianske net inden for landbruget, i det såkaldte **BOBLO**-system (navnet står for **BO**vine **BLO**odtyping, hvor bovint angiver, at det vedrører kvæg).

Inden for mælkeproduktion gennemføres et intensivt avlsarbejde. Anvendelse af inseminering med nedfrosset sæd betyder, at hver tyr får mange kalve. Derved kan man med stor sikkerhed udvælge de bedste tyre som fædre til den næste generation af avlstyre. Et sådant avlsarbejde



Figur 6: Forenklet grafisk model for et af blodtypekomplekserne i BOBLO.

kræver stor nøjagtighed i registreringerne, så man er absolut sikker på dyrenes afstammingsforhold.

Denne sikkerhed opnås ved at undersøge blodtyper hos afkommet og sammenligne med blodtyperne hos de dyr, der er registreret som forældre. Hvis ikke blodtyperne stemmer overens, må det skyldes en registreringsfejl og oplysningerne om det pågældende dyr må udelades fra avlsværdiberegningerne.

For at hver tyr skal være entydigt bestemt udfra blodtyperne, er det nødvendigt at inddrage en række forskellige blodtypesystemer. Nedarvningen af de forskellige blodtyper er mere kompleks end illustreret i eksemplet. Ofte vil der være mere end to niveauer for hver allel og målingen af fænotypen (blodtypen) er langt mere kompliceret. Den kan kun måles indirekte gennem en kombination af laborietest. Princippet er, at man undersøger blodet for ialt 52 forskellige faktorer.



I laboratoriet testes om blodet koagulerer, når en af disse faktorer er til stede. Ved at kombinere de 52 koagulationsmålinger kan blodtyperne i princippet bestemmes. Som ved alle andre målinger er der en vis fejlmulighed ved disse koagulationsmålinger. Tidligere var det nødvendigt at lade eksperter vurdere koagulationsmålingerne for at tage højde for disse mulige fejl.

BOBLO er opbygget så det indeholder alle disse elementer. Det bayesianske net kombinerer den genetiske model med en model for de forskellige fejlmuligheder. Systemet kræver altså ikke ekspertvurdering af måleresultaterne. Ved brug af systemet indberettes forældrenes blodtyper og resultatet af koagulationsmålinger for selve dyret. Det bayesianske net beregner derefter sandsynligheden for, at forældrene er fejlindberettet. En simplificeret udgave af BOBLO er vist i Fig. 6.

Systemet er endnu ikke taget i direkte praktisk anvendelse, og i givet fald vil det være nødvendigt at inkludere muligheden for brug af DNA-fingeraftryk til slægtskabsbestemmelse. Disse DNA-fingeraftryk giver en mere sikker bestemmelse af individets genotype, men mange af de fejlmuligheder, der er skitseret ovenfor, eksisterer stadig.

Et sådant system til håndtering af DNA-fingeraftryk, vil også kunne finde anvendelse i retsplejen, hvor DNA-fingeraftryk benyttes til at vurdere skyldsspørgsmål hos mennesker.

#### **HUGIN og MUNIN.**

Bayesianske net blev første gang indført i kunstig intelligens af amerikaneren Judea Pearl omkring 1985. Metodologien brød afgørende igennem med ekspertsystemet **MUNIN** (**MU**scle and **Nerve Inference Network**) til elektromyografisk diagnose af sygdomme i muskel- og nervesystemet, se Fig. 7. MUNIN blev udviklet på Aalborg Universitet som en del af et ESPRIT projekt. Gennembruddet blev muliggjort af en effektiv beregningsalgoritme, udviklet af Steffen Lauritzen fra Aalborg i samarbejde med David Spiegelhalter fra Cambridge.

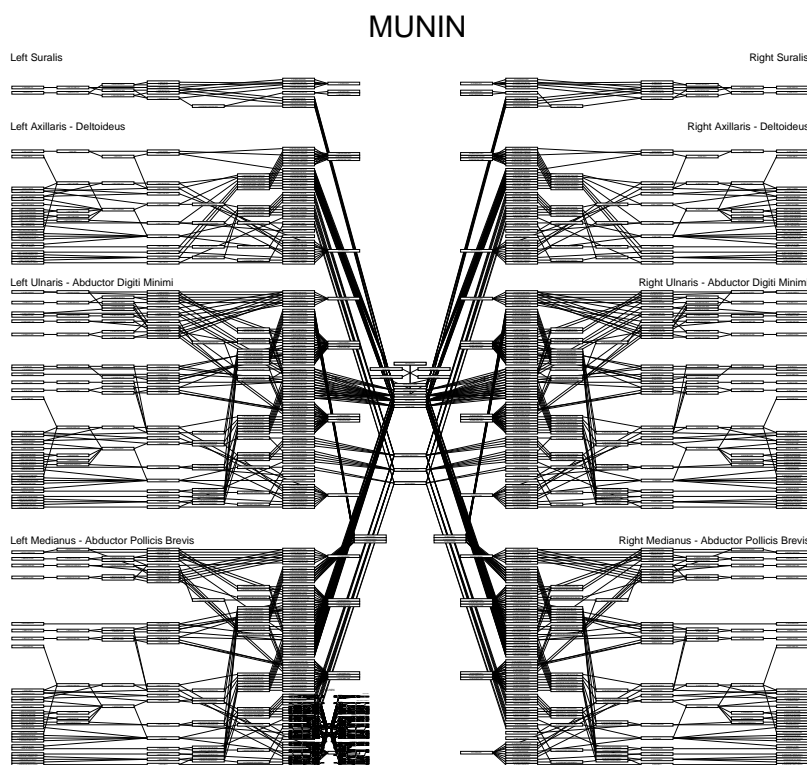
Så kom der fart i sagerne. Algoritmen blev hurtigt forbedret, og et generelt program til opbygning af bayesianske net kom til verden og fik navnet **HUGIN** (**H**andling **U**ncertainty in **G**eneral **I**nference **N**etworks). HUGIN er siden 1989 produceret, udviklet og markedsført af HUGIN Expert A/S, som blev oprettet i samme forbindelse.

HUGIN og MUNIN kendes fra den nordiske mytologi. Det er to ravne på Odins skuldre, som hvisker i hans øre og danner grundlag for Odins alvidenhed. Der er lang tradition for beslutningsstøtte

...

## 4 Andre muligheder

Udvidede versioner af de bayesianske net indeholder flere 'byggeklodser'. En vigtig udvidelse er de såkaldte *influensdiagrammer*. I standardversionen af bayesianske net er resultatet af beregningerne en sandsynlighedsfordeling over tilstande for de enkelte knuder, for eksempel en sandsynlighed for at soen er drægtig eller ej. Dette er ofte tilstrækkeligt til at understøtte en beslutning. Men i andre situationer er der ikke kun én, men flere beslutninger involveret, og de enkelte beslutninger afhænger af hinanden. For at håndtere dette, skal de mulige beslutninger inddrages direkte i den grafiske model. Hertil anvendes influensdiagrammer, der er bayesianske net, suppleret med to typer knuder: *beslutningsknuder*, der i stedet for tilstande indeholder mulige beslutninger, og *nytteknuder*, der er børn af de øvrige knuder og knuden knytter en nytteværdi til hver tilstand i deres forældreknuder.



Figur 7: MUNIN er verdens første større bayesianske net. Det er beregnet til diagnose af sygdomme i muskler og nerver. Den viste version omfatter omkring 1000 knuder.

## 5 Perspektiver for jordbruget

Udviklingen af de metoder, der er beskrevet i det foregående, har inspireret til et omfattende forsknings- og udviklingsarbejde inden for jordbrugsforskning og -rådgivning. Arbejdet foregår hovedsageligt inden for rammerne af **DINA** (**D**anish **I**nformatics **N**etwork in **A**gricultural **R**esearch).

Arbejdet med anvendelsen af de grafiske modeller foregår primært i Danmarks Jordbrugsforskning og på Landbohøjskolen med tætte kontakter til det datalogiske og statistiske miljø på Aalborg Universitet og til rådgivningssektoren på Landbrugets Rådgivningscenter i Skejby.

I dette afsnit beskrives nogle af de udviklingsprojekter, der har været gennemført, og som forventes gennemført i fremtiden.

### 5.1 Bekæmpelse af meldug i vinterhvede

Landbrugsproduktion er en dynamisk proces, hvor dyrkningen løbende overvåges. Hvis forløbet afviger fra det forventede vil landmanden måske skulle gribe ind.

Hvis planterne bliver angrebet af sygdom, kan det være nødvendigt at foretage indgreb inden sygdommen spreder sig, for at begrænse skaden på afgrøden. I det konventionelle landbrug vil man kunne benytte et svampemiddel, mens en anden mulighed kunne være at fjerne afgrøden fra den angrebne del af marken, og dermed forhindre at sygdommen spreder sig. Sådanne beslutninger baseres på en afvejning af omkostningerne ved ikke at foretage sig noget mod omkostningerne ved at gennemføre indgrebet.

Beslutningen kræver viden om afgrødens og sygdommens øjeblikkelige status, for eksempel afgrødens vækststadium og sygdommens nuværende udbredelsesgrad. Denne viden baseres på observationer af marken kombineret med forhold som klima og dyrkningsmetode. Ud over den nuværende status kræves en forudsigelse af den videre udvikling for sygdom og afgrøde, både hvis der foretages indgreb nu og hvis indgreb udelades eller udsættes til senere. En sådan forudsigelse må nødvendigvis være usikker, alene på grund af usikkerhed om klimaet i resten af vækstsæsonen.

I et PhD projekt udvikledes et bayesiansk net til behandling af svampesydommen meldug i vinterhvedemarker. Nettet viste sig velegnet til at håndtere denne problemstilling. Her er beslutningen om indgreb sammensat af en række beslutninger. Selvom man beslutter at undlade indgreb i denne uge, har man jo stadig muligheden for at gribe ind i næste uge. For at håndtere dette aspekt udnyttede man den ekstra facilitet med beslutnings- og nytteværdiknuder i de såkaldte influensdiagrammer.

Erfaringer fra projektet vil fremover blive inddraget i et varslingsystem for bladlus. Systemet vil blive en del af det internet baserede rådgivningssystem *PI@nteinfo*. Gennem internettet kan den enkelte landmand indberette observationer fra sine egne marker. Disse observationer vil blive kombineret med observationer fra andre brugere af systemet i samme lokalområde. Derved forventes en betydeligt forbedret varslings.

## 5.2 Dyrkning af maltbyg uden pesticider

Der er udviklet et bayesiansk net, som samler ekspertviden og forskningsresultater om dyrkning af maltbyg. Brugeren af systemet kan vælge forskellige dyrkningsmetoder og det bayesianske net forudsiger så udbytte og kvalitet af maltbyggen. For eksempel kan man vælge såtid, såmængde samt lokalitet i landet. Af andre faciliteter kan nævnes beregning af effekten af mekanisk ukrudtsbekæmpelse på udbredelsen af svampeangreb. Systemet er i øjeblikket på prototypestadiet, og der søges midler til at videreudvikle systemet.

## 5.3 Prognosesystem for malkekvæg

Hver enkelt malkekvægsbesætning er tildelt en mælkekvote, der fastlægger et loft for, hvor meget mælk, der må produceres pr. år i besætningen.

Dette giver nogle bindinger i tilrettelæggelse af produktionen, som er vanskelige at overskue. For eksempel, hvis en ko ikke er blevet drægtig ved en løbning, giver det en forsinkelse på mindst 3 uger i hendes mælkeproduktion. Hvis en ko bliver syg og må erstattes af en ung ko, er der usikkerhed om, hvornår den nye ko kan starte produktionen af mælk og om størrelsen af hendes mælkeydelse.

For at kunne planlægge produktionen så præcist som muligt, skal al information, der findes om besætningens dyr, udnyttes. Herved fås en mere præcis prognose for mælkeproduktionen, og dermed også en bedre mulighed for tidligt at justere denne.

Derfor er et prognosesystem ved at blive opbygget på Landbrugets Rådgivningscenter, hvor informationer om løbninger og drægtighedsprøver kombineres med oplysninger om mælkeydelse, så prognosen løbende kan opdateres.

## 5.4 Øvrige initiativer

De nævnte eksempler er kun et udsnit af den brede vifte af initiativer, hvor de bayesianske net finder anvendelse inden for jordbruget. Andre områder er nedarvning af kødkvalitetsgenskaber, årsagsfaktorer for forekomst af yverbetændelse hos malkekøer, værdi af foderanalyser ved beregning af fodertildeling, beslutningsstøtte under potteplanteproduktion, optimale behandlingsstrategier for lungesygdomme hos kalve og slagtesvin o.s.v.

Også på det mere metodeorienterede område samarbejdes mellem jordbrugsforskningen og den grundlæggende forskning vedrørende bayesianske net.

### Internetadresser:

Vil man vide mere om bayesianske net, er internettet en god kilde til oplysninger. Adressen til *HUGIN Experts* hjemmeside er <http://www.hugin.dk/> og herfra kan man få fat i en prøveversion af HUGIN og meget andet. Der er også nyttige oplysninger fra *Association of Uncertainty in Artificial Intelligence* på adressen <http://www.auai.org/> eller fra *Microsoft Research* med adressen <http://www.research.microsoft.com/research/dtg/>.

Aktiviteterne i *DINA* findes omtalt på adressen <http://www.dina.dk/>.

De viste eksempler på bayesianske netværk i de første to afsnit kan hentes fra <http://www.dina.dk/NaturensVerden.html>. Eksemplerne kan benyttes i prøveversionen af HUGIN.

## Referencer

- Charniak, E. (1991). Bayesian networks without tears. *AI magazine* **12**, 50–63.
- Jensen, F. V. (1996). *An Introduction to Bayesian Networks*. University College London Press, London.
- Lauritzen, S. L. & Spiegelhalter, D. J. (1988). Local computations with probabilities on graphical structures and their application to expert systems (with discussion). *Journal of the Royal Statistical Society, Series B* **50**, 157–224.
- Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems*. Morgan Kaufmann Publishers, San Mateo.