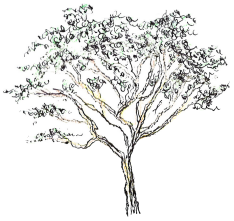


Modelling rhino presence with Bayesian networks

Maryn van der Laarse - 14051428

A dissertation submitted in partial fulfilment of the requirements for the degree

Master of Engineering (Industrial Engineering)

in the
Faculty of Engineering, Built Environment, and Information Technology

University of Pretoria

July, 2019

Abstract

Modelling complex systems such as how the white rhinoceros *Ceratotherium simum simum* uses a landscape requires innovative and multi-disciplinary approaches. Bayesian networks have been shown to provide a dynamic, easily interpretable framework to represent real-world problems. This, together with advances in remote sensor technology to easily quantify environmental variables, make non-intrusive techniques for understanding and inference of ecological processes more viable than ever. However, when modelling an animal's use of a landscape we only have access to presence locations. These data are also extremely susceptible to both temporal and spatial sampling bias in that animal presence locations often originate from aerial surveys or from individual rhinos fitted with tracking collars. In modelling species' presence, little recognition is given to finding quantifiable drivers and managing confounding variables. Here we use presence-unlabelled modelling to construct Bayesian networks for rhino presence with remotely sensed covariates and show how it can provide an understanding of a complex system in a temporal and spatial context. We find that strategic unlabelled data sampling is important to counter sampling biases and discretisation of covariate data needs to be well considered in the tradeoff between computational efficiency and data accuracy. We show how learned Bayesian networks can be used to not only reveal interesting relations between drivers of rhino presence, but also to perform inference. Having temporally aware environmental variables such as soil moisture and distance to fire, allowed us to infer rhino presences for the following time step with incomplete evidence. We confirmed that in general, white rhinos tend to be close to surface water, rivers and previously burned areas with a preference for warm slopes. These relationships between drivers shift notably when modelling for individuals. We anticipate our dissertation to be a starting point for more sophisticated models of complex systems specifically investigating its use to model individual behaviour.

Keywords: *Bayesian network, animal presence modelling, remote sensing, modelling for inference, spatial modelling*

Acknowledgements

What a journey this has been! This dissertation properly introduced me to the wonderful sphere of spatial data science applied in conservation and taught me how to think deeper about concepts we so often take for granted.

My sincere gratitude goes to my supervisor, Prof. Johan Joubert, not only for his guidance and knowledge during the execution of this project, but also for his ongoing support in my journey into cross disciplinary research. I would also like to thank Dr. Sam Ferreira from South African National Parks (SANParks) who went out of his way to support me and teach me the fundamentals of large mammal ecology. Likewise, I would like to acknowledge SANParks Scientific Services for making datasets available.

I am grateful for my colleagues at African Parks who have been extremely accommodating in my research venture. Thank you to my friends who stuck around in the difficult times and all my love goes to my family whose support seems inexhaustible.

There were many days that were not easy, and this dissertation has been both my survival and escape for many months. Even though I will never stop dreaming up solutions for our world's problems nor cease my adventures into all of Africa's remote places, I have learned when to fight and when to accept.

Contents

List of Figures

List of Tables

Acronyms

AIC	Akaike Information Criterion
API	Application Programming Interface
BIC	Bayesian Information Criterion
CPT	Conditional Probability Table
DAG	Directed Acyclic Graph
DBN	Dynamic Bayesian Network
DEM	Digital Elevation Model
DN	Digital Numbers
EM	expectation-maximization
FIRMS	Fire Information for Resource Management System
GEE	Google Earth Engine
GPS	Global Positioning System
HMM	Hidden Markov Model
KNP	Kruger National Park
LAI	Leaf Area Index
LiDAR	Light Detection and Ranging
MAR	missing-at-random
MCAR	missing-completely-at-random
MNAR	missing-not-at-random
MI	Moisture Index
MNDWI	Modified Normalised Difference Water Index
MODIS	Moderate Resolution Imaging Spectroradiometer
NASA	National Aeronautics and Space Administration
NBRT	Normalized Burn Ratio Thermal

NDMI Normalized Difference Moisture Index

NDVI Normalised Difference Vegetation Index

NIR Near infrared

NP non-deterministic polynomial-time

OCC One Class Classification

SANParks South African National Parks

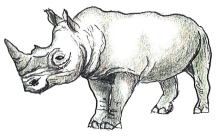
SDM Species Distribution Model

SVM Support Vector Machine

TOA Top-Of-Atmosphere

Chapter 1

Introduction



The unicorn, through its intemperance and not knowing how to control itself, for the love it bears to fair maidens forgets its ferocity and wildness; and laying aside all fear it will go up to a seated damsel and go to sleep in her lap, and thus the hunters take it.

Leonardo da Vinci (1452-1519)

Billions have been spent to combat the trade in illegal wildlife and still, rhino numbers continue to plummet. The cost of saving a species has become extremely high – both in terms of expenditure and in terms of lives lost at the front line. We need to optimise existing resources and innovate by drawing approaches from disciplines other than our own. This chapter gives background on the broader poaching problem, how modelling for insight can be beneficial and introduces Bayesian networks. We provide a research design, research methodology and outline of the dissertation.

1.1 Our rhino conundrum

The white rhinoceros, *Ceratotherium simum simum*, disappeared from the then Transvaal as early as 1896, at the hands of humans (Bigalke, 1963). Not only did English sportsmen revel in the slaughter of hundreds of animals, but the Boer settlers hunted for trade and exported an estimated 90 000 kg of ivory in 1855 alone, the equivalent of approximately 1000 elephants. Still, the white rhino managed to endure, as Hluhluwe-Umfolozi Game Reserve restocked Kruger National Park (KNP), reintroducing over a thousand white rhinos over a 12-year period, commencing in 1961 (Pienaar, 1970). Laws and protected areas seem to have initially kept animals safe. However, a sophisticated demand for rhino horn prevails and, today, poaching is fuelled by transnational organised crime enabled by complex supply chain networks. The white rhino is a keystone species, a source of tourism revenue and has become a symbol of humankind's ability to save our own future.

Haas and Ferreira (2016) identified nodes in the illegal chain most vulnerable to disruption: eliminate the demand, dismantle organised crime syndicates, increase anti-poaching

measures and improve quality of life for people living near protected areas. Using this knowledge, authorities should adopt a multi-faceted strategy, spending the bulk of its budget on sustainable, long-term interventions and smaller, cost-efficient changes. Existing anti-poaching measures should be used as effective as possible to disrupt the supply chain at its source by making poaching for opportunists unappealing.

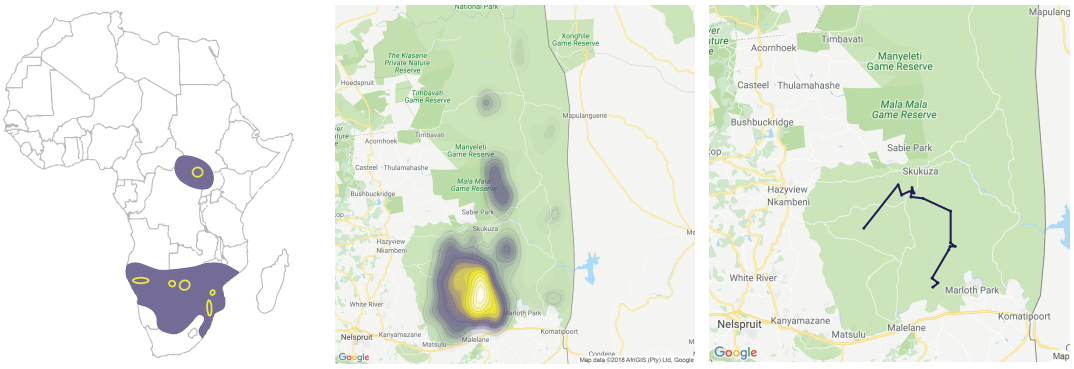
One way to increase the cost of obtaining rhino horn is to increase the risk for a poacher to get caught. If park managers and rangers are enabled to anticipate poaching hotspots, the chances of apprehending would-be offenders can increase. Some models in literature follow the logic that a poaching event will occur if both a poacher and a rhino are present in the absence of a protector (Haas and Ferreira, 2017; Koen et al., 2017; Park et al., 2015). Consequently, it is easier to protect something if its protectors have a sense of where to find it. However, locations of rhinos are unknown or, at best, limited to a few datasets obtained from rhinos fitted with Global Positioning System (GPS) tracking collars or from aerial surveys. Without investing in more expensive equipment, how can we better estimate how rhinos are distributed in space during a given time?

1.2 Modelling animal behaviour

It remains challenging to model something as complex as animal behaviour as it is driven by a combination of external factors and internal traits. It also changes over time. It will be unjust not to reference Box (1979)'s famous concept of *all models are wrong, but some are useful* as it is to date impossible to have any system existing in the real world to be exactly represented by a model. However, any chosen model often provide remarkably useful insight and approximations.

Likewise, attempting to model animal presence, which is most certainly a complex system existing in the real world, we do not strive to achieve the full truth. Rather, we attempt to build illuminating and useful models. The modelling of any real world problem involves decisions on assumptions and selection of suitable modelling approaches. In environmental systems, an appropriate modelling scale often needs to be considered, where scale is conveyed separately as the extent of the study area and the resolution (pixel size). Thus, a large area can be modelled at either at a *coarse* or *fine* resolution. The following hierarchy places modelling approaches into three levels, broadly aligning with modelling at various scales.

Level 1: Large-scale distribution This level deals with species dynamics for conservation planning at a national or global scale. For example, decision makers may need to identify potential areas best suited to be converted into future protected areas. Expert opinion, together with large-scale environmental data such as weather maps are usually acquired to construct global distribution maps. Figure 1.1a shows the historical distribution of white rhino.



(a) Large scale distribution. (b) Population distribution. (c) Individual movement.

Figure 1.1: Hierarchy of animal behaviour modelling.

Level 2: Population distribution Models at this level view a population as an aggregated unit, usually focus on a specific protected area and are rarely concerned with individual differences. Population dynamics, which involves statistics of a population such as size, age, immigration and emigration and modelling techniques concerned with capturing species' habitat such as Species Distribution Models (SDMs) in a spatial manner are included here. Data are often obtained from surveys such as aerial censuses (Figure 1.1b) or camera traps and environmental data consisting of expert opinion, field samples and remotely sensed data.

Level 3: Individual movement Individual modelling approaches are often concerned with how an animal moves and how it spends most of its time, without giving specific attention to population traits. Examples of individual modelling approaches include random walks, forward particle filters and state-space models (Jonsen et al., 2005). Recent advances in animal tag technology allow researchers to sample individual animals' locations more frequently at finer temporal and spatial resolutions (Figure 1.1c). Although electronic tags are becoming more popular, it remains an intrusive form of data capturing and its data often have time irregularities and location errors. Such data are also temporally autocorrelated.

Depending on the particular application, it is important to know which assumptions are justified and which should be worried about more. Different modelling approaches are better suited to each modelling level and are also dependent on data availability. As a start, this dissertation is not concerned with the distribution and dynamics of white rhino across the globe as illustrated by Figure 1.1a, but rather interested in population dynamics whilst realising the underlying individual traits influencing larger scale distributions. The above hierarchy is useful to select a suitable modelling level, identify shortfalls within each modelling level, but also to determine gaps between the various levels.

1.2.1 Challenges

Whichever modelling level is selected, uncertainty needs to be mitigated, simplicity needs to be realised and the most appropriate modelling technique should be selected. Uncertainty is not a new phenomenon and certainly not unique to our problem.

Where they are not

The first evident form of uncertainty in our real world problem, is contained within obtainable field data. As the world stands on the peril of the Fourth Industrial Revolution, technological advances are integrating the biological, physical and digital worlds. More often than not, conservation operates in remote areas and data transmission and technology maintainability are more challenging than in other industries. This and the limits of non-intrusive technologies to monitor animals, introduce a number of biases to existing data. Data are usually available either from surveys, providing a peek into population distributions, or from animal tags that record paths of only a small number of individuals.

Another challenge in modelling animal presence is that available data consist of *presence-only* samples. For example, spotting a rhino during an aerial survey, only locations where rhinos have been present can be labelled as such. We do not know for certain where a rhino will never be. In the more general machine learning literature, this problem is referred to as the One Class Classification (OCC) problem, where the negative class is either not properly sampled or not available (Khan and Madden, 2009). One approach in dealing with this is by sampling unlabelled points, also referred to as *background points*. Unlabelled points are not attempting to necessarily be a proxy for absence locations, but rather to represent the environment within the study area (Phillips et al., 2009). Conversely, pseudo-absences are explicitly attempting to estimate locations where species will not be. Still, it cannot be said with certainty that those points will always be absent of the species.

More than habitat models

Generally, models give a static depiction of a very dynamic system. One of the biggest shortfalls of available modelling techniques, especially at a population level, is that actual presence distributions are rarely accurately modelled. Instead, a species' habitat is modelled and not whether animals actually use that habitat. To understand this better we need to grasp the niche concept as encapsulated by Environmental Sciences: The range of environmental conditions in which a species can survive is known as the *fundamental niche*. As biotic interactions and dispersal may exclude certain areas of the fundamental niche to be used, the *realised niche* is the area which a species is really found within. For example, the fundamental niche of white rhinos will include the most part of southern Africa, but it can only be found in certain protected areas such as in KNP, as borders prohibit animals to migrate even though suitable habitat exists elsewhere. This also holds true at smaller scales, as white rhinos will rarely be seen in the north of KNP even though they can in fact survive there. Thus, the white rhino's realised niche is much smaller than its fundamental niche. An SDM is one of the most popular tools within Ecology to model a species' habitat, with its output mostly seen as a quantification of the realised niche of a species being that observations are already constrained by limited resources and biotic interactions, Guisan and Thuiller (2005). This dissertation is specifically interested in how this realised niche changes over time and in reaction to a dynamic environment.

Deeper than the population

Another important shortfall within existing methods of modelling animal behaviour lies not so much within a specific level of modelling, but rather between the linkage of different modelling scales. Population-level consequences, which are often observed and analysed as isolated events, essentially emerge from decisions made by unique individuals (Morales and Ellner, 2002). Only a handful of studies attempted to bridge the gap between population

dynamics and individual movement (Morales et al., 2010). Imagine a white rhino bull is fitted with a GPS tracking collar that records his location every hour. These locations are really only observed events as a result of an underlying process: the bull's behaviour. It is difficult to understand why the points of presences were located where it was, if we are not interested in the bull's behaviour. Disregarding the underlying process *causing* the observable events may impair behavioural models. Important drivers of movement could appear insignificant without accounting for heterogeneity in animal behaviour. For example, imagine a dataset containing observations of a few white rhino. If we fit a general statistical model to the data set based on a few environmental variables, we may overlook the fact that one bull was in fact experiencing must—a periodic condition associated with high testosterone levels and aggressive behaviour—and was travelling further than usual to reach a cow in oestrus.

We need to consider an animal's changing behaviour over time and specific to their unique traits. Species distribution models, population dynamics and other machine learning algorithms most often aim to quantify behaviour at population level. But two individuals will most certainly behave differently. They may have different phenotypes (condition, size), different past experiences and may experience different environments or even have different personalities. Individual behaviour should not be ignored even when population level predictions are sought.

1.2.2 Problem statement

This dissertation attempts to answer the question

> *"How can we better estimate where all the white rhinos within a protected area are at a given time?"*

Animal presence is a complex process, and models in literature that have attempted to capture it are either (i) impaired by data uncertainty, (ii) producing static models or (iii) over-simplified for population level applications.

Having an answer to the dissertation's question will allow us to answer more practical questions. Park managers and rangers concerned with ecological management may ask the following questions: *"Which area is suitable for the re-introduction of white rhino? Where should rangers patrol this week? If it rains this week, where will most of the white rhinos be next week? If we had burn a certain landscape, how would rhinos have responded differently?"*

1.3 The genesis of Bayesian networks

The questions in the previous subsection can be grouped within Judea Pearl's Three Layer Causal Hierarchy which captures the classification of causal information in terms of types of questions that each level is capable of answering (Pearl and Mackenzie, 2018). The first level is *association*, invoking purely statistical relationships defined by raw data. For example, observing a patch of lush vegetation makes it more likely for a rhino to be present. Such an association can be inferred directly from data using conditional expectations.

The second level, *intervention*, involves not only observing what is, but also intervening what we see. For example, what will happen if we ignite a fire, will we still observe a rhino? Such questions cannot be answered from data alone, as they involve a change in the rhino's behaviour in reaction to a new threat.

Finally the uppermost level, *counterfactuals*, will ask questions such as what if we had not ignited a fire, would we have observed a rhino? If we can answer questions at

the *counterfactual* level, we can answer questions at lower levels. This hierarchy explains why artificial intelligence systems based only on associations are still not able to do what thinkers, like us, can do.

The question *"can machines think?"* was first asked by Alan Turing as early as 1950 appearing in the British journal, *Mind*. It later became clear that Turing did not consider whether or not machines can think, but rather whether machines can do what thinkers can do (Turing, 2009). Before this, the main approach to artificial intelligence was driven by Boole (1854) and his followers using basic inference logic: if A is equal to B and B is equal to C, then C is equal to A. Although theoretically plausible, this approach quickly reached limits and forced statisticians to include many exceptions. For a realistic representation of a rule, the number of exceptions quickly grew beyond the ability to keep track of them all, especially in modelling systems where the rules are unclear. In response to this, the first "expert system" was created by the Turning Award winner Edward Feigenbaum, in the 1970s, attempting to emulate a human expert's decision-making ability (Feigenbaum and McCorduck, 1984). Early expert systems such as an oil-exploration application and a medical diagnostic programme were enthusiastically implemented, but by 1980 it became clear that these systems were struggling under uncertainty. In the late 1970s, "How to deal with uncertainty?" was a question on the minds of many. Attempts ranged from Zadeh (1988)'s fuzzy logic, to Bell (1988)'s belief functions.

Finally in 1982, Judea Pearl published a paper illustrating an efficient belief-updating algorithm using Bayesian probabilities. Three years later Pearl dubbed this approach *Bayesian networks*. His idea was not to model the expert but nature, and in doing so, to model nature to be compatible with scientific knowledge using a hierarchy of cause-effect relationships. He insisted on coherence with acceptable probability theory and Bayesian reasoning. Bayesian networks offer an easily interpretable model and evolved to become capable of integrating expert knowledge when limited data are available and can be extended within a spatial, temporal and agent-based context. The Bayesian network approach gets closest in answering more than just a question from the *association* level in Pearl's Three Layer Causal Hierarchy. The Bayesian network approach gets closest in answering "if we had burn a certain landscape, how would rhinos have responded differently?"

1.4 Research design

The following dissertation statement is formulated:

> *Rhino presence is a complex process and should be captured using methods capable of dealing not only with uncertainty, but also with changes across time and space.*

To evaluate this statement, the following objectives have to be achieved:

1. manage uncertainty and limited data availability;
2. produce models with temporally and spatially aware outputs; and
3. provide an understanding of the complexity of modelling animal presence.

1.5 Research methodology

In response to the surge in rhino poaching, researchers from different disciplines applied unique theories to the situation. Ecologists created species distribution models to map

a rhino's habitat. Supply chain analysts identified nodes in the illegal trade network most vulnerable to disruption (Haas and Ferreira, 2016). Sociologists realised the key role politics and communities play in the problem and urge us to shift our focus towards communities living adjacent protected areas and in fact use them as change agents (Hubschle, 2018). Additionally, criminologists applied theories such as the Routine Activities Theory which is based on the principle that a crime will be committed when the following three factors align: (1) the presence of a motivated offender; (2) the presence of a suitable target; and (3) the absence of capable guardians (Cohen and Felson, 2016). Herbig and Warchol (2011) identified the dense communities with high unemployment and crime rates surrounding KNP provide a pool of motivated offenders and thus, focusing on equipping guardians may require less effort than defusing motivated offenders.

Operations Research provides a framework for relevant disciplines to work cooperatively to find a solution enabling researchers to address specific aspects of a problem, while considering the big picture. We aimed to develop new and improved models of rhino behaviour and new and improved tools to better equip decision makers for protected areas. According to Manson (2006) this qualifies as doing research into Operations Research, which the Institute for Operations Research and the Management Sciences (INFORMS) describe as "the discipline of applying advanced analytical methods to help make better decisions".

We adopted a Design Research perspective which Manson (2006) describes within an Operations Research viewpoint as "a process of using knowledge to design and create useful artefacts, and then using various rigorous methods to analyse why, or why not, a particular artefact is effective." Ultimately, any understanding gained feeds back into the body of knowledge of the discipline. Figure 1.2 shows that a researcher should first become aware of a problem. In this dissertation we became *aware* of the limitations within available rhino presence models in the sense that most (i) are impeded by uncertainty and data availability, (ii) produce static models and are (iii) over-simplified for population level applications. The *awareness* of the potential of Bayesian networks lead to the *suggestion* to apply Bayesian networks to the problem. This lead to the development of three artefacts: (i) a framework to map behavioural drivers of white rhino presence to environmental variables, (ii) a methodology to cleverly sample unlabelled and remotely sensed dynamic variables and (iii) multiple Bayesian network formulations for rhino presence. These artefacts were *evaluated* using statistical measures on a test data set as well as cross-checking with animal behaviour experts. The research was *concluded* through the formulation of findings and discussions of new awarenesses and suggestions.

1.6 Outline of the dissertation

The dissertation is structured as follows: Chapter 2 reviews literature on Bayesian networks, establishes knowledge about white rhino behaviour and how to quantify it through remotely sensed data. Chapter 3 evaluates available tools most suited for the analysis, introduces our study area, discusses how to best manage the uncertainty in the data, and describes the process of sampling environmental covariates. Chapter 4 gently introduces Bayesian networks applied to rhino presence focusing on three variables, and then expands through the inclusion of more variables. Chapter 5 explores the addition of time-aware variables, introduces an approach for missing data and constructs Bayesian networks for individual rhinos. Chapter 6 concludes the dissertation, providing an overview of our biggest findings and proposes future research areas.

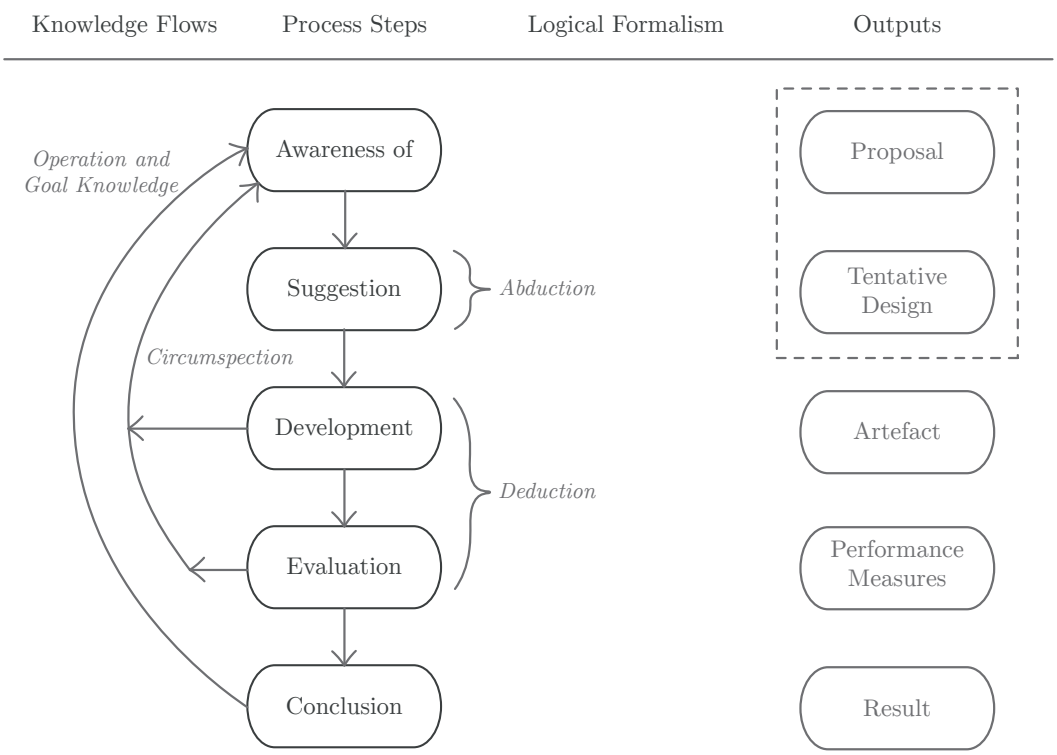


Figure 1.2: Design research methodology as reproduced from Manson (2006).

Chapter 2

Literature study

When Henry Ford made cheap, reliable cars, people said, "Nah, what's wrong with a horse?" That was a huge bet he made, and it worked.

Elon Musk on Twitter

This chapter takes a plunge into Bayesian networks, introducing concepts such as structure and parameter learning and inference. We then provide a description of factors influencing rhino behaviour and how to quantify it from remotely sensed data.

2.1 Bayesian networks

Bayesian networks equip data scientists with a manner in which to translate their human understanding of the world. They are especially effective in dealing with uncertainty and missing data, which are often characteristics of environmental studies. For example, imagine we want to know if there is a drought in an area. Often when there is a drought, we will observe dry grass. However, the grass may be dry because of a normal seasonal change. An animal may die when there's a drought or it can perhaps be killed by a poacher. Finally, if an animal has died, we will probably observe vultures circling the carcass. Figure 2.1 illustrates this example, which is partially inspired by Pearl (2014)'s earthquake example. We may want to use such a causal graph to predict what will happen (if there is a drought, an animal will die) or to infer causes from observations (if there are dead animals and dry grass, then there is probably a drought). The

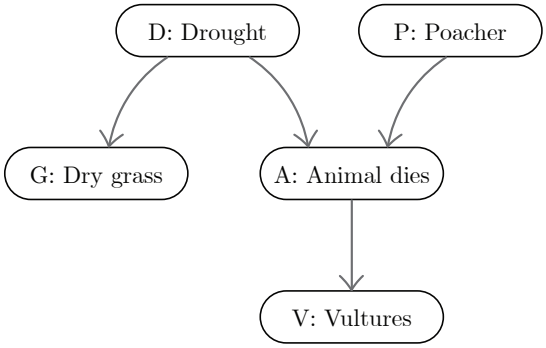


Figure 2.1: A causal graph: Observing vultures.

arcs are not necessarily absolute, for example, animals often die with neither a drought occurring nor a poacher present in the area.

Scientists' need to capture their understanding of the human world into a model can either be seen as a constraint or as the exact reason for using Bayesian networks. When predictive accuracy is the sole objective, the data scientist may be steered to apply other machine learning algorithms. However, Bayesian networks are easy to interpret, diagnose and tweak and may be more applicable to *model for understanding* as well as for inference. Bayesian networks provide a natural and efficient method for portraying probabilistic dependences among a set of variables. During the last two decades, Bayesian networks have been used in a variety of settings to study causal inference and probabilistic prediction, from modelling traffic flows in Beijing (Sun et al., 2006), to revealing relations between genes, the environment and disease (Su et al., 2013). Still, its application in Environmental Sciences remains scarce (Aguilera et al., 2011).

Even though the drought example described above is over-simplified, Bayesian networks provide a modular and interpretable framework to model complex systems by marrying the language of data on one side and that of diagrams on the other.

A Bayesian network is formally defined as a type of *graphical model* representing the probabilistic dependencies among a set of random variables, $\boldsymbol{x} = \{x_1, x_2, ..., x_p\}$ and consists of the following two components:

1. a network structure referred to as a Directed Acyclic Graph (DAG), $G = (\boldsymbol{v}, A)$, where each node $v_i \in \boldsymbol{v}$ represents a random variable of interest, x_i and each arc or directed edge represents a stochastic dependence between variables; and

2. a set of parameters as conditional probability distributions contained within a Conditional Probability Table (CPT) associated with each node, namely $P\left(x_i|pa(x_i)\right)$, where x_i represents a node and $pa(x_i)$ its parents. Each CPT represents the stochastic dependence represented by the arcs.

If there exists a directed edge in a DAG from node D (*drought*) to node G (*grass*), G is said to be a *child* of D and D the *parent* of G. In Figure 2.1 A (*animal dies*) is the parent of V (*vultures*) and likewise V is the child of A. The importance of differentiating between *child* and *parent* nodes becomes apparent when we introduce Bayesian networks' independence characteristic. Bayesian networks have an important characteristic in that each node is assumed to be conditionally independent of the set of all its precursors, given the values of their parents. For example, are the variables D (*drought*) and V (*vultures*) independent? Intuitively not, since when there is a drought, an animal will probably die and we will probably observe vultures. However, when we know for certain that an animal has died and we wonder if we will see vultures, whether there is a drought or not is of no concern. The absence of a directly connecting arc between two nodes is understood as two variables being independent given the values of intermediate nodes.

This important independence feature, known as the *Markov property* allows complex systems to be modelled using a limited number of parameters for each node (Charniak, 1991). The *Markov property* enables the representation of the full joint probability distribution of the variables $\boldsymbol{x}$ as a product of conditional probabilities associated with each variable x_i using the the chain rule in (2.1) (Korb and Nicholson, 2010)

$$p(\boldsymbol{x}|\theta) = \prod_{i=1}^{n} p\left(x_i|pa(x_i), \theta_i\right) \tag{2.1}$$

where $\boldsymbol{x} = x_1, ...x_n$ are the random variables and $\boldsymbol{\theta} = \theta_1, ...\theta_n$ are the parameters of the variable x_i given its parents $pa(x_i)$. The prior probabilities of all nodes with no

predecessors and the conditional probabilities of all other nodes given their predecessors are required. Together these CPTs compactly represent the full joint distribution.

Figure 2.1 portrays a hypothetical DAG of the *drought* example described above, assuming there are no hidden or latent variables. There is an arc shown from *animal dies* to *vultures* indicating that observing vultures is understood to be dependent on an animal dying. There are also arcs from *drought* to *dry grass*, from *drought* to *animal dies* and from *poacher* to *animal dies*, indicating dependence of children nodes on their parents. In this DAG there is no arc from either *drought* or *poacher* to *vultures*, indicating that knowing the state of *animal dies* leaves the *vultures* independent of *drought* and *poacher*. The full joint probability distribution the variables can be calculated according to (2.2).

$$P(D, P, G, A, V) = P(V|A).P(A|D, P).P(G|D).P(D).P(P) \tag{2.2}$$

If we assume all the variables in Figure 2.1 are understood to be binary, the number of parameters required for the conditional probabilities are reduced from 32 parameters to 10 under the *Markov property*, emphasising its significance. The variables do not necessarily have to be binary or even discrete for the independence assumption to hold, but the computational efficiency is improved when applying discretisation to continuous variables.

2.1.1 Learning a Bayesian network

Fitting a Bayesian network involves two distinct steps of learning, the first corresponding to learning the underlying graphical structure and the second to estimating the conditional probability parameters associated with the random variables. This problem, finding the exact structure and parameters of a Bayesian network, has been demonstrated to be an non-deterministic polynomial-time (NP)-hard problem (Cooper, 1990), and becomes part of a class of problems for which there is no known polynomial algorithmic solution. The time to solve NP-hard problems increases exponentially with problem size.

Structure learning

Structure learning entails the identification of the graph structure of a Bayesian network best suited to the dependence structure of the data. Another option is to define the structure manually using expert knowledge in areas where data are challenging to obtain. Not surprisingly given the limited data in the environmental modelling sphere, researchers often favour the use of domain knowledge of a human expert in building the graph structure (Bromley et al., 2005; Koen et al., 2017; Uusitalo et al., 2005). However, with the upsurge in open-source availability of high resolution remotely sensed data, environmental indices can be computed and sampled for the use of learning structures.

Literature equips us with two main categories for learning the structure of a Bayesian network from data: constraint-based methods and score-based methods. Constraint-based methods work by using the observed data to identify conditional independence relationships between variables which are used in return to constrain the possible underlying network structure. Directed separation (d-separation) between nodes indicate conditional independence relations between the corresponding variables and plays a critical part in adding directions to arcs. Having three disjoint nodes, X, Y and Z, it is understood that Y will d-seperate X and Z if Y blocks every path from a node in X to a node in Z. In the *Observing vultures* example in Figure 2.1, *animal dies* d-separates *vultures* from both *drought* and *poacher* as it is **serial**. However, being part of a **convergent** path, *animal dies* does not d-separate *drought* from *poacher*. The **divergent** path allows for *drought* to d-separate *dry grass* from *animal dies*. Unique conditional dependence relations for **serial**,

divergent and **convergent** are crucial in constraint-based methods. The Growth-Shrink algorithm, based on the concept of the *Markov blanket* is most widely used in this approach. The *Markov blanket* for a node, Y, contains all the nodes shielding Y from the rest of the network, thus consisting of its own parents, children and its children's other parents. In Figure 2.1, as node *drought* does not have a parent, its Markov blanket includes its children, *dry grass* and *animal dies*, and the other parent of its children, *poacher*.

The second approach to learning the structure of a Bayesian network, namely score-based methods, return the single structure which maximises the score that measures how well it fits the data. Most of these methods use heuristic search techniques such as simulated annealing or hill-climbing as Bayesian networks often have too many possible structures to allow for a comprehensive search.

A number of scoring methods are employed for Bayesian networks. The full marginal likelihood is sometimes impractical for larger networks favouring the use of approximate scores such as the Bayesian Information Criterion (BIC) which is identical to the *Minimum Description Length*. The BIC is known to result in simpler graphs as its calculation involves the penalisation of networks with many arcs. If a more complex graph is desired, the Akaike Information Criterion (AIC) may be used instead.

Parameter learning

Learning the parameters in a Bayesian network is a significant task in itself, but is also important in the process of structure learning. Many structure learning algorithms such as score-based methods estimate parameters as part of the operation. Parameter learning of a given structure G consists of estimating the conditional probability distributions over all variables matching the empirical dataset with the highest likelihood.

Data scientists often have to deal with the *curse of dimensionality*. As the dimensionality of the feature space increases, the number of configurations grow exponentially and thus the number of configurations covered by an observation decreases. A large feature space requires not only high computational power but also larger training datasets to create meaningful models. In this regard, parameter learning becomes substantially simplified through the employment of the *Markov property* once the network structure has been learned from the data. Local distributions involve only a small number of variables and their dimension usually does not scale with the number of features, relieving the *curse of dimensionality*.

Literature presents two principal approaches to learning parameters of Bayesian networks, one method based on *Bayesian estimation* and the other on the use of *maximum likelihood estimation*. In this dissertation we focus on the use of *maximum likelihood estimation*.

The likelihood of the model $M(G, \boldsymbol{\theta})$ with structure G and conditional probability distributions $\boldsymbol{\theta}$, given the dataset D is given by (2.3),

$$L(M|D) = \prod_{d \in D} p(d|M) \tag{2.3}$$

where $P(d|M)$ is the likelihood of any sample d given the model M. Parameter learning then involves finding parameters maximising the likelihood function in (2.3) and is expressed in (2.4).

$$\boldsymbol{\theta} = \arg\max_{\theta} L(M_\theta|D) \tag{2.4}$$

Maximising the likelihood function is also equivalent to maximising the log-likelihood function.

2.1.2 Inference in Bayesian network

Much work on Bayesian networks focused on the inference problem. Imagine we have a Bayesian network representing some probability distribution and have to answer a probabilistic query. For example, in the *Observing vultures* example in Figure 2.1, we can query the probability of observing vultures in an area knowing that an animal has recently died there. This can be answered by computing the probability distribution over some random variable $\boldsymbol{X}$, given some evidence $\boldsymbol{E} = \boldsymbol{e}$, such that $P(\boldsymbol{X}|\boldsymbol{E} = \boldsymbol{e})$.

In calculating these queries, there always exist a tradeoff between accuracy and speed. Seeing that the Bayesian network problem is NP-hard (Cooper, 1990), obtaining exact inference, by giving exact posterior predictions computed based on the query, from Bayesian network can become computationally intractable. An alternative approach to exact inference is settling for approximate inference, even though approximate inference in Bayesian network is also NP-hard (Dagum and Luby, 1993).

Exact inference often makes use of either the variable elimination algorithm or clique tree algorithm. Approximate inference often make use of sampling-based approaches by randomly picking assignments of the random variables and then estimating all properties of the joint distribution from the samples. In this dissertation, we performed exact inference where computationally feasible and shifted to approximate inference for more complicated networks.

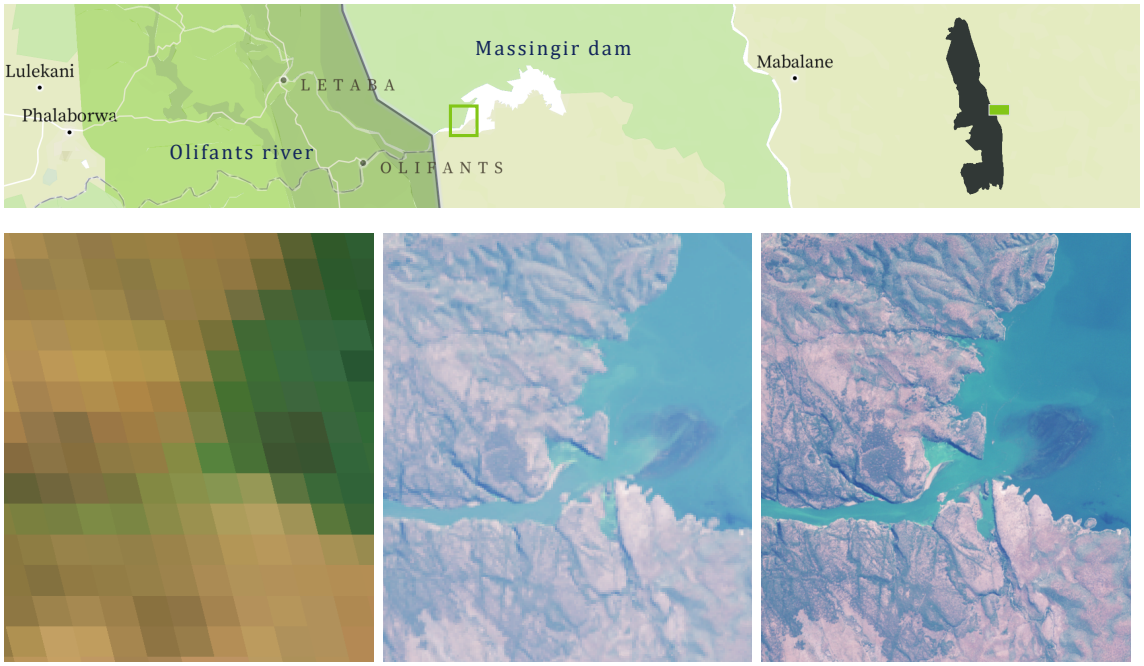
2.2 Eyes in the sky

To construct any sort of model, we need to identify variables of interest related to what we are trying to model. The simplest rhino presence model will have food and water as variables, as an animal wants to maximise its energy by increasing its foraging time in productive areas (Charnov, 1976). Literature provides a guide for identifying possible variables and we evaluated its quantification capability against available remotely sensed indices.

The field of earth observation is rapidly evolving, with more satellites than ever equipped with sensors capturing a wide range of wavelengths. Elon Musk's SpaceX and other companies are busy launching tens of thousands of low orbiting satellites into space with the sole purpose of providing affordable, reliable and fast connectivity to remote places with internet speeds of at least 10 Mbps (Wells, 2019). In the next five years our experience of the night sky may be very different. Satellites are not only transforming the way we interact with the world, but also how we view it.

The Moderate Resolution Imaging Spectroradiometer (MODIS) was launched in 1999 and provides of the most used payloads in earth observation, supplying up to 250 m resolution every 16 days from as early as 2000. The MODIS instruments capture information in 36 spectral bands with wavelengths varying from 0.4 µm to 14.4 µm. Landsat 8, a collaboration between National Aeronautics and Space Administration (NASA) and the United States Geological Survey (USGS) was launched in 2013 providing nine spectral bands of the entire earth every 16 days at an improved spatial resolution of 30 m. The EU Copernicus Programme launched the Sentinel-2 satellite in 2015 providing worldwide coverage every sixth day of 10 m pixels consisting of 13 Top-Of-Atmosphere (TOA) bands of which three are in the red-edge region, centered at 705 nm, 740 nm and 783 nm. These red-edge bands provide significant improvements for vegetation mapping, green leaf area index, chlorophyll content (Delegido et al., 2011) and burn severity mapping (Fernández-Manso et al., 2016).

We considered spatial and temporal extent and resolution in identifying sensors from which to obtain remotely sensed imagery. Seeing that we required data from as far back as 2013, whilst ensuring satisfactory spatial and temporal resolution, we used Landsat 8 for most remotely sensed indices. We argued that the larger temporal range is of more value than the improved 10 m spatial resolution, as we anticipated rhino presence data to contain spatial errors of as big as 30 m. Figure 2.2 shows the Olifants river flowing into the Massingir Dam in Mozambique, comparing the spatial resolution of products from MODIS, Landsat 8 and Sentinel-2. The MOD09Q1.006 Terra Surface Reflectance 8-Day Global 500 m product is shown in Figure 2.2a, with the visual bands of Landsat 8 and Sentinel-2 in Figure 2.2b and 2.2c, respectively. The significant improvement in feature detection from MODIS to Landsat 8 sustained our decision in the use of Landsat 8 imagery.



(a) MODIS. (b) Landsat 8. (c) Sentinel-2.

Figure 2.2: Comparison of available remotely sensed products.

2.3 White rhino behaviour

Environmental indices calculated from remotely sensed imagery provide exclusive access to a dynamic digital landscape of the entire earth. This section describes what is known about the white rhino's interactions with its environment and attempts to map these behavioural drivers against available remotely sensed indices in literature.

2.3.1 The study area

The focus area of this dissertation was the Kruger National Park (KNP) which falls within the low-lying savanna of South Africa, covering a staggering 19 485 km^2 with annual rainfall exceeding 450 mm (Venter et al., 2003). KNP comprises landscape types from granite and gneiss deposits separated by Karoo sediment. In the southern parts, wooded savanna comprising *Sclerocarya caffra* and *Acacia nigrescens* dominates the basalts, while mixed

Combretum spp. and *Acacia spp.* are the key species on granites. Northern parts comprise primarily *Colophospermum mopane* woodlands (Gertenbach, 1983).

Figure 2.3: The study area of this dissertation: Kruger National Park.

2.3.2 Food and water

Not unlike other organisms, white rhinos move. They move either to achieve a short-term goal such as feeding, drinking, reproduction and escaping threats, or longer-term goals such as avoidance of inbreeding and kin interaction (Holyoak et al., 2008).

The white rhino is possibly the largest pure grazer that ever lived (Estes, 1991) and feeds selectively and avoids forbs. In KNP the white rhino actively sought out shade-loving grasses such as *Panicum maximum* along riverbands, especially in the mornings (Pienaar, 1994). They also prefer short grass species such as *Dactyloctenium aegyptium*, *Digitaria* spp., *Panicum coloratum Sporobolus nitens* and *Urochloa mosambicensis*. As grazers generally prefer higher quality vegetation, Normalised Difference Vegetation Index (NDVI) (Tucker, 1979) a widely used vegetation greenness index can be used to quantify vegetation quality with values ranging from -1 to 1, with 1 relating to healthy vegetation all from a satellite's spectral bands, calculated by (2.5),

$$\text{NDVI} = \frac{\text{NIR} - \text{Red}}{\text{NIR} + \text{Red}} \tag{2.5}$$

with *red* being visible red light and NIR near-infrared light. The accuracy of NDVI can be expanded by incorporating a red edge band in the index. A few studies have successfully identified tree species from multispectral Sentinel-2 imagery albeit at a scale of 10m (Immitzer et al., 2016; Karasiak et al., 2017).

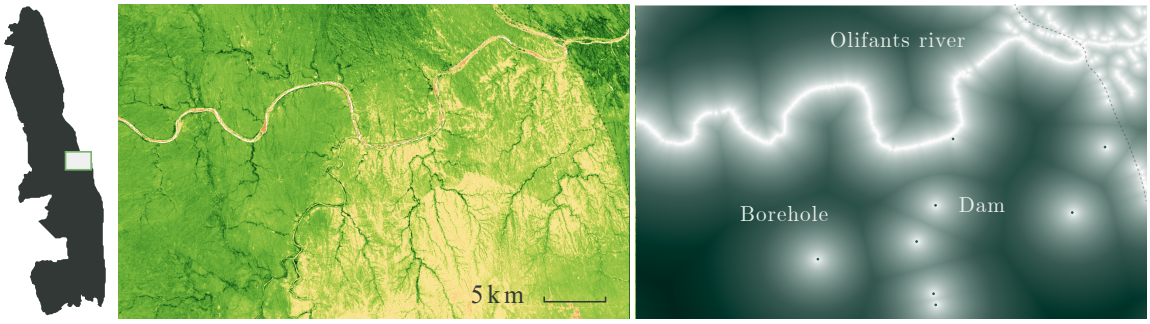
Grazing gradients can also often be observed using distance from rivers as a proxy. KNP has four perennial rivers flowing west to east including the Sabie and Crocodile rivers in the south. A number of ephemeral rivers flow through the park during the wet season.

Other than food, the white rhino also needs water for survival. Redfern et al. (2003) showed that forage quality and quantity should be viewed together when analysing grazer distributions. Two daily drinks will satisfy the rhino, but they can go up to four days without water (Pienaar, 1994). KNP has an interesting history of introducing artificial waterholes into the park. These waterpoints and rivers act as activity nodes, influencing the landscape-scale distribution of herbivores (Smit et al., 2007). Specifically, most grazers (including white rhino) have been shown to associate with artificial waterpoints, whereas browsers associated positively with rivers.

Artificial waterholes and remotely sensed surface water should be included when quantifying water as driver. Modified Normalised Difference Water Index (MNDWI) is often used to accurately differentiate between non-water and water features and is calculated using (2.6)

$$\text{MNDWI} = \frac{\text{Green} - \text{MIR}}{\text{Green} + \text{MIR}} \tag{2.6}$$

with *Green* referring to visible green light and *MIR* to middle infrared reflection. MNDWI values greater than zero are classified as water bodies (Han-Qiu, 2005) and the Euclidean distance can be calculated form each pixel to the nearest water source. Figure 2.4a shows a map of NDVI surrounding the Olifants river in February 2019. Figure 2.4b shows distance to the nearest water source, where surface water was identified from Landsat 8 images and enhanced with a artificial water layer of KNP.



(a) NDVI. (b) Distance to water.

Figure 2.4: Vegetation quality and distance to the nearest water source from Landsat 8 images and artificial waterholes (source: self produced).

Even at a resolution of 30 m, which Landsat 8 data provides, smaller water bodies remain undetected as many of the artificial waterholes were left undetected. Animals rarely need a water source of larger than 30 m^2 to satisfy their thirst. To account for this limitation, a moisture index can be used to capture not only surface water but also underground water as elephants often dig holes in riverbeds providing access to water otherwise unreachable. The moisture index will also be helpful to identify muddy areas, as white rhino enjoy wallowing in these hallows for up to several hours.

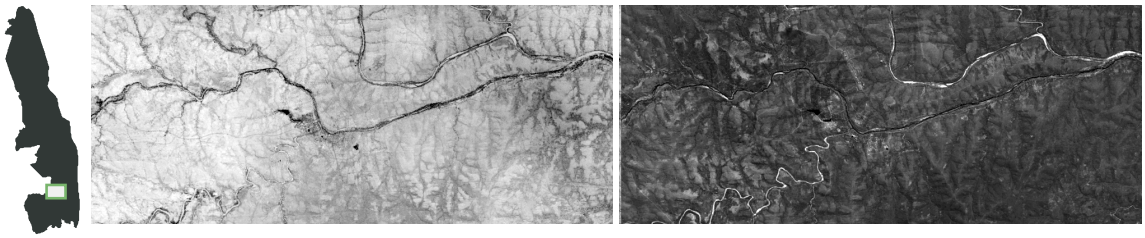
The Normalized Difference Moisture Index (NDMI) is calculated from the Near-Infrared (NIR) and Short Wave Infrared (SWIR) using (2.7). NIR reflectance is influenced by leaf dry matter content and leaf internal structure whereas SWIR reflectance is influenced by the spongy mesophyll structure in vegetation and the vegetation water.

$$\text{NDMI} = \frac{\text{NIR} - \text{SWIR}}{\text{NIR} + \text{SIWR}} \tag{2.7}$$

Ideally we do not only want to capture vegetation moisture, but soil moisture as that will be a better proxy for rainfall. Zhan et al. (2007) developed a simple method, expressed in (2.8), to estimate soil moisture with moderate accuracy not using a thermal band

$$\text{SMMRS} = 1 - \frac{1}{\sqrt{M^2 + 1}}(\text{NIR} + M * \text{RED}) \tag{2.8}$$

where the slope is $M = 1.40426$. Figure 2.5 shows soil moisture and vegetation moisture layers in the Skukuza area highlighting the Sabie river.



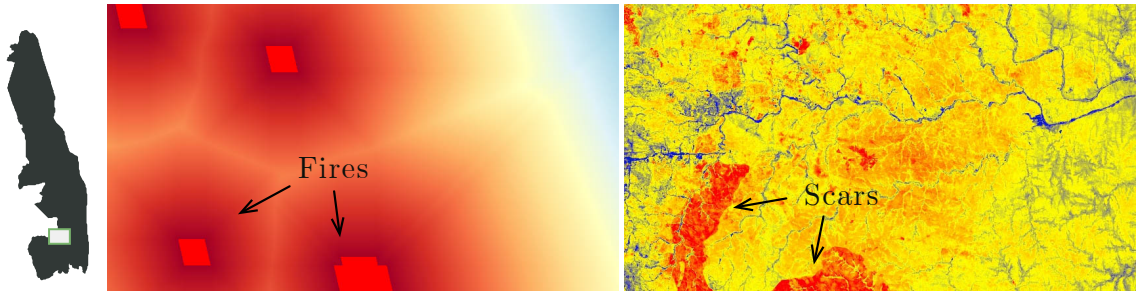
(a) Soil moisture. (b) Vegetation moisture Moisture Index (MI).

Figure 2.5: Moisture layers of KNP in the Skukuza area (source: self produced).

Herbivory and fire often compete in consuming biomass (Smit and Archibald, 2019). Pienaar (1994) occasionally observed white rhinos feeding on fresh burnt stubble the day after an area burnt. Not only do white rhino feed on burnt stubble, they also enjoy freshly sprouting *Themeda triandra* a few days after a burn. Holden et al. (2005) proposed the Normalized Burn Ratio Thermal (NBRT) index to map historical fires using (2.9), dividing the Thermal Infrared band by 10000

$$\text{NBRT} = \frac{\text{NIR} - \text{SWIR} * \text{TIRS}}{\text{NIR} - \text{SWIR} * \text{TIRS}} \tag{2.9}$$

where NIR is Near-Infrared band (band 5 for Landsat 8) and SWIR is the Thermal Infrared band (Band 11 for Landsat 8), using a threshold of 0.93 to separate scars from burned areas. Fire is not only an important influence on vegetation, but also adds to the landscapes of fear an animal experiences. Google Earth Engine (GEE) provides a Fire Information for Resource Management System (FIRMS) dataset containing near-real time LANCE fire detection images from MODIS MOD14/MYD14 at a resolution of 1 km.



(a) Fires and distance to fires. (b) Scars from NBRT1 16 days after burning.

Figure 2.6: Remotely sensed fire data in the south of KNP for June 2018 (source: self produced).

The white rhino is rarely predated, except for calves that are vulnerable to hyenas and lions (Estes, 1991). Other than environmental threats such as fire and disease, the true

threat to rhinos is man. We can quantify this by incorporating human settlements and roads, and calculate the distance to the nearest human presence.

Intrinsic factors may influence an individual to disperse from a suitable habitat patch to another when the fitness benefits outweighs the cost of movement (Bowler and Benton, 2005). Dispersal is a decision process happening at individual level and is often driven by kin interactions, inbreeding avoidance or habitat variability. Although individual dispersal affects population dynamics, only a fraction of the population will usually disperse.

Adult bulls are solitary, except when associating with females in oestrus. Sub-adults and cows are rarely solitary as they associate in pairs, usually the female with her most recent offspring. At two-to-three years of age, a juvenile may be rejected but will join up with another juvenile or female without calves. Demographic studies revealed that two calfless cows may also be seen in pairs. Herds of up to six can be observed and larger groups may be formed temporarily.

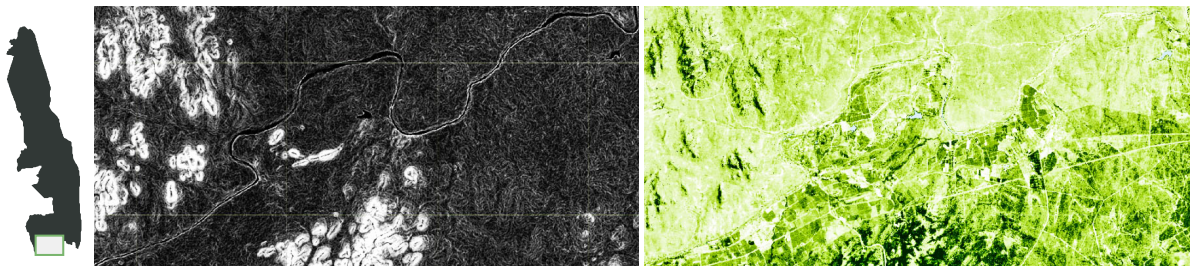
2.3.3 Underlying structure

The white rhino typically feed on the bottomlands, riverbanks and brackish areas next to water courses during early mornings. They will also wallow in the mud ponds adjacent water courses during the wet season. During mid morning the white rhino will typically move uphill towards ridgecrests and midslopes where it is cooler than the low-lying areas. They choose to be in the low-lying woodland or thicket valley bottoms on cool and windy days (Pienaar, 1994). Later afternoon, white rhinos can be found moving downwards and feeding on the footslopes, continuing feeding throughout the night (Owen-Smith and Smith, 1973).

Gertenbach (1983) distinguished between these landtypes in KNP as areas with a specific fauna and flora, macroclimate, soil type and geomorphology. These landforms are also quantified by GEE's landcover product, Global ALOS Landforms as produced by Theobald et al. (2015). They developed a comprehensive landform classification based on dominant physical processes and hillslope position. First, they distinguished four hillslope positions from topographic position index (TPI) along the cantena, namely ridges/peaks or summits, upper slopes or shoulders, lower slopes or foot slopes and valley bottoms or toe slopes. Second, features were identified at gradient extremes differentiating between flat and steep. Third, they also used the continuous heat load index (CHILI), distinguishing between warm, neutral and cool hillslope positions . Each classification is shown in Table A.1 in the Appendix.

Slope can also be calculated from the Shuttle Radar Topography Mission's digital elevation data providing Digital Elevation Models (DEMs) on a near-global scale at 30 m resolution. Slope as shown by Figure 2.7a might be of interest as animals often tend to take the path of least cost to resources.

Animals tend to rest in the shades of trees and Landsat provides Vegetation Continuous Fields (VCF) tree cover layers containing the percentage of horizontal ground covered by woody vegetation taller than five meters in each 30 m pixel. We used a layer compiled from the NASA/USGS Global Land Survey (GLS) collection of Landsat data from 2010. Figure 2.7b shows a tree cover sample in the south of KNP, with dense canopies of cultivated lands just outside the border of the park.



(a) Slope. (b) Percentage tree cover.

Figure 2.7: Trees and slope in the south of KNP (source: self produced).

All of these environmental variables can be understood as driving white rhino presence in some way. These indices, mapping the underlying environmental structure and capturing food and water availability, provide a quantitative means of mapping our understanding of white rhino behaviour.

2.4 Conclusion

This chapter introduced Bayesian networks as an approach to model white rhino behaviour and reviewed various environmental indices from remotely sensed data that can act as proxies for drivers of white rhino presence. The next chapter reviews data sources which represent historic white rhino distributions in KNP. These data sources usually only contain presence points and not absence records (where white rhinos will never be) and we propose a methodology to overcome this challenge and mitigate sampling biases.

Chapter 3

Data sampling

We can only see a short distance ahead, but we can see plenty there that needs to be done.

Alan Turing (1912 – 1954)

In modelling complex systems in the real world, uncertainty is unavoidable. It is impossible to capture all possible states and exceptions that underlie a finite number of observations. Probabilistic inference deals with the problem of finding the posterior probabilities of unobserved variables given the observations of other variables. This chapter gives a brief overview of available analysis tools and then strategically samples unlabelled data. We then provide a framework to sample environmental covariates at scale.

3.1 The right tools

This dissertation aimed to use open-source software as far as possible, and luckily, there are many capable tools to choose from. `Python` by Rossum (1995) and `R` by R Core Team (2019) are widely used programming languages, both able to provide a general purpose computing environment for applied statistics with extensive spatial analysis libraries.

Even though these environments support spatial data computations and machine learning algorithms, one of the biggest limitations in working with raster data is its intensive processing requirements. Doing analysis over a temporal spectrum of a couple of years quickly becomes insupportable as big raster files need to be downloaded and analysed. Google Earth Engine (GEE), courtesy of Gorelick et al. (2017), provides a planetary scaled remote sensing analysis platform with both a `JavaScript` and `Python` Application Programming Interface (API). The API enables data scientists to use the powerful libraries of GEE through their programming environment of choice. This makes GEE not only versatile but also scalable, allowing for on-the-fly data collection and correction at a global scale. We utilised these powerful tools by using GEE for efficient raster analysis, `Python` for repetitive and modular sampling tasks and, ultimately, `R` for building Bayesian networks. Specifically, the `bnlearn` package by Scutari (2010) was used which provides an

implementation of the most commonly used Bayesian network structure and parameter learning algorithms. The resulting Bayesian networks were easily stored as `.net` files, ready to be imported in other networking software. The `bnlearn` package seamlessly integrates with the `gRain` package by Højsgaard (2012) for inference and with the `raster` package by Hijmans (2019) to process all spatial data.

3.2 Positive and unlabelled data

To capture the white rhino's use of the landscape, Scientific Services from Kruger National Park (KNP) provided data sets containing rhino observation records for the period from 2013 to date.

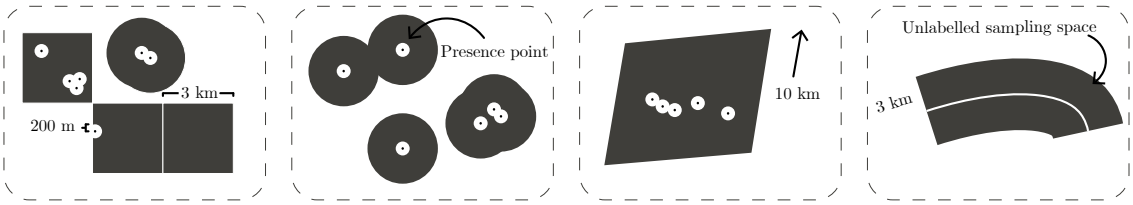
3.2.1 Unlabelled data

The modelling of rhino presence can be viewed as a classification problem attempting to classify a pixel as either containing a rhino presence or not. However, one will only ever have information of rhino observations and are thus faced with the One Class Classification (OCC) problem where it is assumed that only information for one of the classes is obtainable, and the negative class is either not available or not properly sampled.

The most common approach of dealing with the OCC problem, especially in a species habitat modelling scenario, is to generate unlabelled points (also referred to as *pseudo-absences* or *background points*). Unlabelled point sampling design strongly influences models' predictive performance while the number of samples taken are less important (Stokland et al., 2011). The relationship between the environmental extent from which pseudo-absences are sampled and that of the presence observations is important in a model's performance. Common sampling procedures include random selection (Stockwell, 1999), random selection with environmental-weighted exclusion (Zaniewski et al., 2002), random selection with geographic-weighted exclusion (Hirzel et al., 2001) and using locations where species were not present on visit (Elith and Leathwick, 2007).

Borrowing from the general machine learning community, sampling of unlabelled data should aim to mimic the same bias as the labelled data. Most of our presence data sources introduced both temporal and spatial sampling biases. Presence data are either obtained from surveys, susceptible to availability, observer and detectability bias or obtained from animal tags capturing information about specific individuals. Some presence records are obtained from social media platforms resulting in the data being biased to only be close to publicly accessible areas such as roads.

Imitating these biases, a background point was sampled for each presence point with a corresponding date and time, limiting the sampling space depending on the nature of the data generating mechanism as shown in Figure 3.1. This sampling approach is described in more detail in the following sub-sections.



(a) Aerial surveys (b) Aerial demographics (c) Tracking (d) Social media

Figure 3.1: Sampling approach for unlabelled points.

3.2.2 Aerial data

During the dry season each year South African National Parks (SANParks) launches aerial surveys across KNP retrieving population estimates for some species. We obtained data from annual aerial rhino surveys conducted from 2013 to 2018 resulting in a total of 23244 white rhino presence points. Bi-annual aerial demographic studies conducted in February and September each year added another 3013 points to our database. Each record contained the species name, number of animals observed, sex of animals observed, longitude and latitude. The total number of rhinos observed at each sighting was extracted, as multiple rhinos were often observed in one sighting.

Aerial surveys

Aerial surveys conducted in KNP are based on block-based surveys through the intensive searching of 878 blocks of 3 km^2 covering 41.5 % of the park (Ferreira et al., 2015). Surveys take place at the end of the dry season every year, using a Eurocopter Squirrel helicopter surveying blocks randomly distributed throughout the park, with a slight bias towards the high-density areas south of the Sabie River. Flying at an average speed of 65 Knots. 45 m above the ground, a 200 m search area on either side of the aircraft is covered by two observers sitting on either side of the helicopter. Three biases are introduced to the presence points in Figure 3.2a using this sampling technique, namely *availability bias* where animals are unaivalable to observers, for example, animals standing under trees, *observer bias* where observers have different capabilities and *detectability bias*, where animals may be present and available, but there exists variation in detecting them.

(a) Presence records. (b) Sampling space. (c) Unlabelled records.

Figure 3.2: Generating an aerial survey sampling space.

Attempting to mimic these biases in the presence data, all 878 searching blocks were included in the sampling space, but masked by a 200 m buffer around each presence points, assuming the environment was suitable for the whole of the 200 m radius. Some presence points fell outside the searching blocks, as they were probably observed when flying between blocks and a 3 km buffer was generated around these points, masking it with the smaller 30 m buffer. The sampling space in Figure 3.2b was then used to sample 23244 unlabelled points as shown in Figure 3.2c.

Demographic studies

Demographic studies are carried out annually by SANParks' Scientific Services to get a better understanding of population dynamics. Rhino demographics are noted on ad-hoc flights, not necessarily sampling specific search blocks. 3013 presence points was obtained from this, Figure 3.3a, and followed a similar approach than before to sample unlabelled data. A 3 km buffer around each presence point was masked with a 200 m buffer, resulting in the sampling space in Figure 3.3b. A unlabelled point was sampled for each presence point to correspond the date and time as shown in Figure 3.3c.

(a) Presence records. (b) Sampling space. (c) Unlabelled records.

Figure 3.3: Generating an aerial demographic survey sampling space.

3.2.3 GPS collar data

A handful of white rhinos (eleven cows and three bulls) were fitted with Global Positioning System (GPS) tracking collars for the period of February 2016 to January 2017, recording a location about every eight hours. One of these rhino cows was poached in May 2016, another in June 2016 and one of the bull's during January 2017. Some rhinos were only fitted with a collar later in 2016 with another's collar removed resulting in a total of 6296 presence locations.

GPS tracking data were cleaned by removing any duplicate points, as well as masking the area surrounding the Scientific Services' office, as collars were sometimes still recording positions while under maintenance.

One approach in creating a suitable unlabelled data sampling space is to draw a buffer around the path of a tracked rhino. However, this will introduce unnecessary constraints on the sampling space as the line connecting the tracking points might be inaccurately passing through locations a rhino will never use. Instead, we created a 200 m buffer around each tracking point to account for possible spatial errors as noted in telemetry technology as illustrated by Figure 3.1c. Seeing that white rhino bulls rarely leave their home range of about 2 km^2 and cows and calves of about 12-15 km^2 (Owen-Smith, 1971), we assumed a rhino had the choice of moving as far as 10 kilometres if they really wanted to. This possible space was then masked with the 200 m error buffer, resulting in the sampling space in Figure 3.4b. Figure 3.4c shows the corresponding 6296 unlabelled data points.

(a) Presence records. (b) Sampling space. (c) Unlabelled records.

Figure 3.4: Generating a tracking survey sampling space.

3.2.4 iNaturalist

In a world where social media is becoming the norm, we consulted iNaturalist (iNaturalist.org, 2019), an online social network platform for naturalists, citizen scientists, and biologists used to map and share observations of biodiversity across the globe. Although only 97 white rhino observations were obtained, it remained useful to include multiple sources, especially one that will only grow in its popularity. This demonstrated the approach for sampling points constrained by biases often included with social media sources. Specifically data obtained from social media platforms are constrained by tourist access points and visibility from roads. Animal-scale objects were assumed to be resolvable as extended objects from a distance of just under three kilometres. We drew a 3 km buffer around the public tarred roads in KNP as shown in Figure 3.1d and merged it with a layer buffering 3 km around iNaturalist presence points as no layer covering the entire public road network in KNP was available.

(a) Presence records. (b) Sampling space. (c) Unlabelled records.

Figure 3.5: Generating an iNaturalist sampling space.

3.3 Covariate sampling

For all the environmental indices identified as possible drivers of white rhino behaviour, we obtained locally available data layers together with imagery from the remote sensors, Landsat 8 and Moderate Resolution Imaging Spectroradiometer (MODIS). From these data sources, the following environmental covariates were calculated:

- Distance to the nearest tourist camp (km) as calculated from a point theme referencing the extent of KNP's main tourist rest camps obtained from the SANParks data repository.
- Distance to the nearest public tar road (km) as calculated from a line layer showing the publicly accessible roads within KNP obtained from the SANParks data repository.
- Distance to the nearest river as calculated from a line layer of the major rivers in KNP obtained from the SANParks data repository.
- Distance to the nearest fire (km) in the current time step (t=16 days), calculated as the euclidean distance from the Fire Information for Resource Management System (FIRMS) dataset.
- Distance to the nearest fire (km) in the previous time step (t=16 days) using the FIRMS dataset. The fire raster layer was converted to a vector layer after which the the euclidean distance to the nearest fire polygon was calculated.
- Distance to the nearest scar or burnt area (km) using Normalized Burn Ratio Thermal (NBRT) calculated by Equation 2.9 on Landsat 8 imagery. Scars were identified under a threshold of having an NBRT value smaller than 0.93, and converted to a vector layer. The euclidean distance was then calculated to the nearest scar polygon.
- Landform from Global ALOS Landforms as produced by Theobald et al. (2015). (See Table A.1 in the Appendix for a detailed classification of 15 different landforms).
- Landtype as described by Gertenbach (1983). (See Table A.2 in the Appendix for a detailed description of the 100 landtype classes).
- Soil type as described by Gertenbach (1983). (See Table A.2 in the Appendix for a reference of soil types).
- Soil moisture in the previous timestep (t = 16 days) calculated by Equation 2.8 on Landsat 8 imagery.
- Vegetation moisture using Normalized Difference Moisture Index (NDMI) calculated by Equation 2.7 on Landsat 8 images.
- Distance to the nearest surface water (km). An Modified Normalised Difference Water Index (MNDWI) layer was calculated using 2.6 on Landsat 8 imagery. Surface water was identified as having MNDWI values greater than zero. This layer was transformed to a vector and augmented by an artificial waterhole layer provided by SANParks Scientific Service, where after the euclidean distance to each point or polygon was calculated.
- Vegetation quality using Normalised Difference Vegetation Index (NDVI) calculated with Equation 2.5 on Landsat 8.
- Slope from a Digital Elevation Model (DEM) provided by the Shuttle Radar Topography Mission.
- Percentage tree canopy cover from the National Aeronautics and Space Administration (NASA)/USGS Global Land Survey (GLS) using Landsat images.

3.3.1 Processing covariates

Before indices were calculated from remotely sensed data, preprocessing of Landsat 8 layers commenced. Specifically, the Top-Of-Atmosphere (TOA) reflectance as calibrated by Google Earth Engine following the methodology of Chander et al. (2009) from Landsat 8's Collection 1 Tier 1 was used.

All data layers were also re-projected from the sinusoidal (SIN) projection, to the Hartebeesthoek94/Lo33 projection specified for South African areas east of 32°E.

Clouds cover around 67% of the Earth's surface at any given time, regulating the planet's energy levels (King et al., 2013). Inevitably, remotely sensed environmental indices will be inaccurate for pixels covered by clouds. Using the collection quality band from Landsat 8, all cloud pixels were masked using algorithm 1.

Algorithm 1: Mask clouds using the quality band of Landsat 8.

```
Result: cloud masked image
initialization;
for each image in collection do
    select BQA band;
    for each element in BQA band do
        turn the cloud bit off
    end
    mask image with cloudmask
end
return image
```

Each index was calculated using equations provided in Section 2.3 by executing the code snippets as illustrated in the Appendix. These environmental covariates are extremely cyclic as illustrated by Figure 3.6. Specifically, fires are more prevalent in dry season, from February to September, resulting in a decrease in the mean distance to fire and distance to scar layers. Soil moisture and NDVI increases over the wet season, January to May. More surface water is available during the wet season and as a result, the distance to the nearest water source decreases in the wet season, beginning in January.

For each presence and unlabelled record in our dataset, the layers listed in the previous section were sampled systematically using a combination of `Python` and GEE. This is where the power of GEE was truly realised in doing on-the-fly raster calculations for each date in our dataset. Algorithm 2 was used to sample covariates.

As mentioned earlier, cloud cover makes the sampling of some environmental indices inaccurate. Out of our total of 65 300 records, 17 644 records with missing values for period *t*, and 27 190 records with missing values for period *t-1*, roughly relating to 30 % of records. This is in line with King et al. (2013)'s estimate of percentage clouds cover of the Earth's surface at any given time. In an attempt to reduce the number of missing variables, the searching period was increased for each incomplete record and resampled available environmental indices. This decreased the number of records with missing values from 17 644 to 10 801 for period *t* and from 27 190 to 3242 for period *t-1*.

This approach delivered a full data set consisting of remotely sensed variables mapped against possible drivers of white rhino presence. In the following chapters, we used this data set to build Bayesian networks.

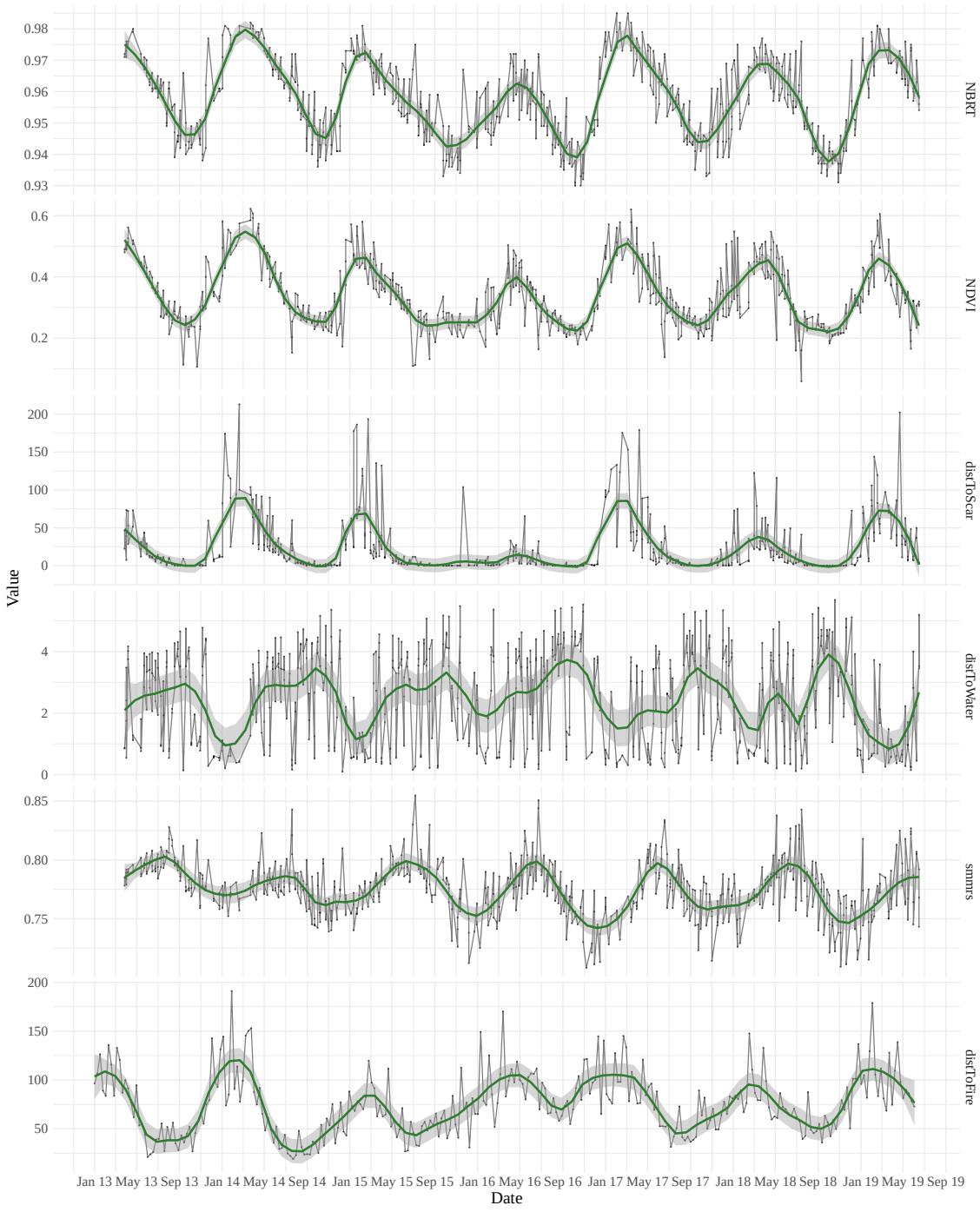


Figure 3.6: Remotely sensed indices over time.

Algorithm 2: Sample covariates for each record.

```
Result: dataframe of records
period = p;
get static indices;
for each date, t, in data do
    select points;
    convert to GEE feature collection on client side;
    for t and t-1 do
        startdate = subtract p/2 days;
        enddate = add p/2 days;
        get fire indices;
        append indices;
        if Landsat 8 image not available for period p then
            period=p*2
        end
        get Landsat 8 images for period;
        mask clouds;
        calculate indices;
    end
    append indices to raster brick;
    for each point in data sample do
        sample values from raster brick;
    end
    Convert samples to dataframe;
    if fire image for sample contained no fires then
        set sample's distance to fire to maximum;
    end
    Append results;
end
return dataframe
```

Chapter 4

Exploring the Bayesian network

...quantify processes in the real world, not just patterns in the data.

Judea Pearl in *"The Book of Why: The New Science of Cause and Effect"*

This chapter introduces Bayesian networks applied to the problem of capturing rhino presence. We discuss different discretisation methods and evaluation approaches. We then construct our first Bayesian network using three variables, after which we introduce more variables to the network.

4.1 Discretisation

To improve computational efficiency, we made all continuous covariate data discrete. A discretisation of a real-value vector, $\boldsymbol{v} = (v_1, ...v_n)$ is an integer-value vector $\boldsymbol{d} = (d_1, ...d_n)$ with the following properties:

1. each element of $\boldsymbol{d}$ is in the set $0, 1...D-1$ for some small positive integer D, known as the degree of discretisation; and
2. for all $1 \leq i, j \leq n$ there is $d_i < d_j$ if and only if $v_i \leq v_j$.

One approach to discretisation is to use set intervals, derived from possible values the variables can take in its existence. *Interval discretisation* divide the interval v_n into k equal size bins with k being a user-defined value. Another approach is using *quantile discretisation*, placing n/k values in each bin. Both these two approaches are extremely sensitive to outliers and may produce a strongly skewed range (Catlett, 1991).

Hartemink's Information-preserving Discretisation (IPD) is an entropy-based discretisation method depending on minimizing loss of mutual information between each two real-valued random variables (Hartemink, 2001).

The mutual information between two variables X and Y with joint probability distribution $p(X, Y)$ and marginal distributions $p(x)$ and $p(y)$ is defined as

$$I(X, Y) = \sum_{y \epsilon Y} \sum_{x \epsilon X} p(X, Y) \log \frac{p(x, y)}{p(x)p(y)}. \tag{4.1}$$

If X and Y are independent, per the independence definition $p(x, y) = p(x)p(y)$, and thus $I(X; Y) = 0$. When modelling environmental networks and having variables such as Normalised Difference Vegetation Index (NDVI) or distance to the nearest surface water, the joint distribution function is rarely known and it is difficult to know whether variables are independent or not.

We explored both Hartemink's approach and interval discretisation, sidestepping the problem of defining a probabilistic model for the data. This represents a tradeoff between computational efficiency and the unavoidable loss in accuracy of the discrete representation of the original data. The reduction in accuracy also decreases a model's ability to generalise over values not necessarily contained in the data. We should especially take care, as our scenario is constrained by various temporal and spatial biases.

4.2 Model evaluation

Interestingly, most algorithms used to learn the structure of a Bayesian network are not explicitly targeted at classification problems but rather seek to reduce the discrepancy between the true dependence structure and that of the estimated one. Another important note is that the concept of *target variable* is central in classification problems but alien in Bayesian networks (Nagarajan et al., 2013). Fundamental classification measures may only provide some indication of model performance, but should not be seen as the ultimate and only metrics. This is especially true in our dealing with data more susceptible to struggle with generalisation.

Cross validation

Cross validation is a popular approach to validate statistical models and has been applied to a broad range of models (Hastie et al., 2009). We used the five-fold cross-validation method for model validation (Geisser, 2017). We randomly partitioned the data into $k = 5$ parts and created k cross-validation samples $\boldsymbol{X}^*_{-1}, ...\boldsymbol{X}^*_{-k}$. A Bayesian network is then learned in parallel, independently, from each $k - 1$ split. The corresponding classification error was computed in parallel and then averaged to calculate the *cross-validated classification error*.

Structure scoring

We exerted the log likelihood scoring method as well as the Bayesian Information Criterion (BIC) to evaluate the Bayesian network's structure against the data.

The formulated test data

Seeing that presence data were generated through processes that introduced a number of biases to it, a test data set was formulated and deliberately sampled to manage the biases in the presence data. We sampled 60 presence and 60 unlabelled points from each data source — from aerial surveys and aerial demographic studies, tracking data and iNaturalist data — resulting in a small test set of 480 records.

We used the learned Bayesian networks to infer presence values for each of the records in the test set, calculating the sensitivity, specificity and accuracy with a 95% confidence interval. This test dataset became the test for generalisation, as a model learned on single source data (i.e. tracking-only data) may perform well in the cross-validation test on data from the same source, but perform poorly when tested against data from other sources.

Computational time

We also evaluated the computational time to learn a network and parameters for a Bayesian network on a Dell Latitude 7280 with 8 GB RAM, an Intel(R) Core(TM) i5-6200 CPU @ 2.30GHz 2.40GHz processor operating on a 64-bit Windows operating system.

4.3 Our first Bayesian network

As a gentle introduction in the construction of Bayesian networks, we included NDVI, `ndvi_t`, and distance to the nearest surface water, `distToWater_t`, as variables changing over time. We introduced `landform` as a static variable. The simplest animal behavioural model will include food and water variables as an animal wants to maximise its energy by increasing its foraging time in productive areas (Charnov, 1976). We followed different attempts in revealing network structures. We used constraint-based (Growth-Shrink) and score-based (hill-climbing) algorithms to reveal insights into underlying structures, but we also investigated the manual definition of structures steered by literature knowledge and domain expertise. Figure 4.3 shows the structure learned using the Growth-Shrink algorithm on data that were discretised following Hartemink's approach.

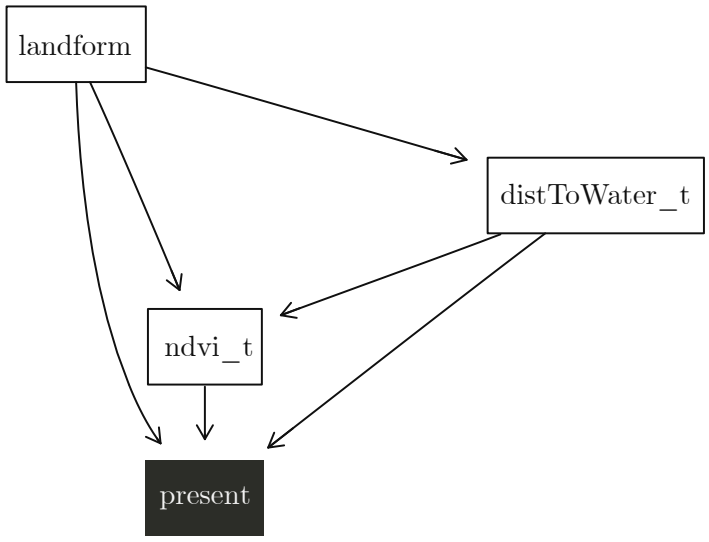


Figure 4.1: A simple network structure learned using the Growth-Shrink algorithm.

This structure containing four nodes and six arcs, captured an expected structure showing that landform influences how far from water an area will be, and also affects the NDVI or quality of the vegetation. Ultimately, all three variables, `landform`, `distToWater_t` and `ndvi_t` showed to affect whether or not a white rhino will be observed in a given cell. This model achieved a cross validation accuracy of 69.62 % and likewise an accuracy of 72.87 % (95 % CI (0.6434, 0.8032)) on the formulated test set.

The directed arc from `landform` to `present` is a rather interesting one. Recall that landforms distinguish between hillslope position, gradient extremes and the continuous heat load index (CHILI). Figure 4.2 explores this relationship by plotting landform values for both presence and unlabelled points.

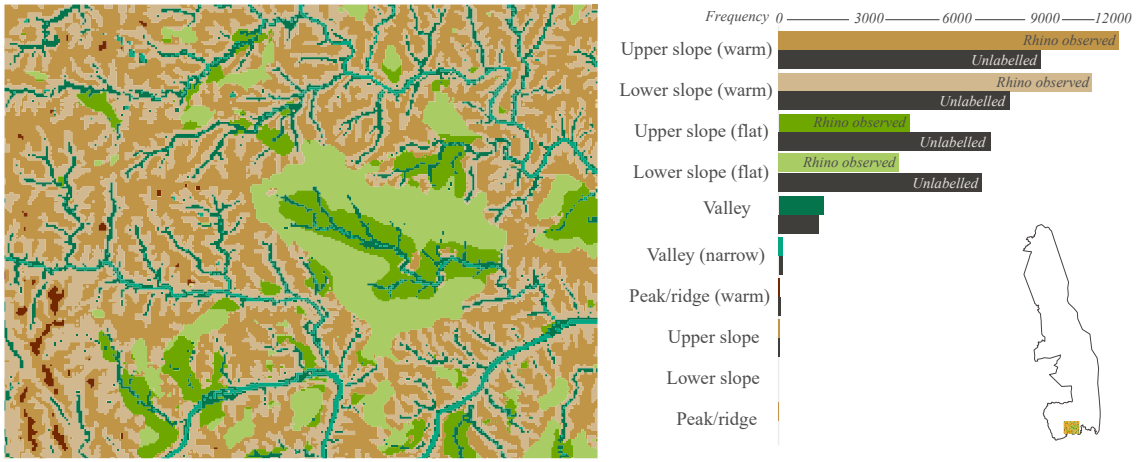


Figure 4.2: The relationship between landform and rhino presence.

The presence points were most dominantly associated with *upper slope (warm)* and *lower slope (warm)* and less so with *upper slope (flat)* and *lower slope (flat)*. This suggests that it is not so much hillslope position or gradient extremes affecting the probability of observing a rhino or not, but rather the landform's continuous heat load index (CHILI). White rhinos prefer warmer slopes above flat slopes.

4.4 Giving it structure

Expanding on the Bayesian network in Figure 4.1, we introduced a number of novel variables as described in the Table 4.1.

Table 4.1: Variables used in Bayesian network learning.

Label	Variable name	Values discretised by Hartemink's approach
`ToCamp`	Distance to camp (km)	(0.175,10]; (10,19.8]; (19.8,27.2]; (27.2,49.3]
`distToRiver`	Distance to river (km)	[0,1.09]; (1.09,5.43]; (5.43,10.9]; (10.9,21.7]
`distToRoad`	Distance to road (km)	(0,3.48]; (3.48,10.4]; (10.4,17.4]; (17.4,34.8]
`distToScar_t`	Distance to scar (km) as derived from NBRT	[0,7.97]; (7.97,15.9]; (15.9,39.8]; (39.8,159]
`distToWater_t`	Distance to water (km) as derived from MNDWI and artificial waterpoints	(0,3.47]; (3.47,5.79]; (5.79,9.26]; (9.26,23.2]
`landform`	Landforms by Theobald et al. (2015)	11, Peak/ridge (warm); 12, Peak/ridge; 21, Upper slope (warm); 22, Upper slope; 24, Upper slope (flat); 31, Lower slope (warm); 32, Lower slope; 34, Lower slope (flat); 41, Valley; 42, Valley (narrow)
`landtype`	Landtypes by Gertenbach (1983)	See Appendix
`ndmi_t`	NDMI as vegetation moisture	(-0.279,-0.135]; (-0.135,-0.064]; (-0.064,0.15]; (0.15,0.437]
`ndvi_t`	NDVI as vegetation greenness	(-0.0763,0.218]; (0.218,0.302]; (0.302,0.554]; (0.554,0.765]
`slope`	Slope	[0,0.535]; (0.535,1.07]; (1.07,1.6]; (1.6,10.7]
`soil`	Soil type	1, Alluvial soils; 2, Basalt soils; 3, Ecca shale soils, 4, Gabbro soils; 5, Granitic soils; 6, Rhyolite soils; 7, Sandy soils
`treeCover`	Percentage tree cover	(0.978,4.38]; (4.38,5.5]; (5.5,8.88]; (8.88,23.5]
`present`	Rhino observed	1, Rhino observed; 0, Unlabelled
`type`	Presence data source	1, Aerial survey; 2, Aerial demography; 3, Tracking; 4, iNaturalist

Using the variables in Table 4.1, discretised with Hartemink's approach, the hill-climbing algorithm learned a structure as shown in Figure 4.3 with 13 nodes and 19 arcs. We reversed arcs from `distToCamp` to `distToRiver` and from `distToCamp` to `landform`. With a sensitivity threshold of 0.45, this model achieved a cross validation accuracy of 68.50 % and an accuracy of 70.08 % (95 % CI (0.6132, 0.7788)) on the formulated test set.

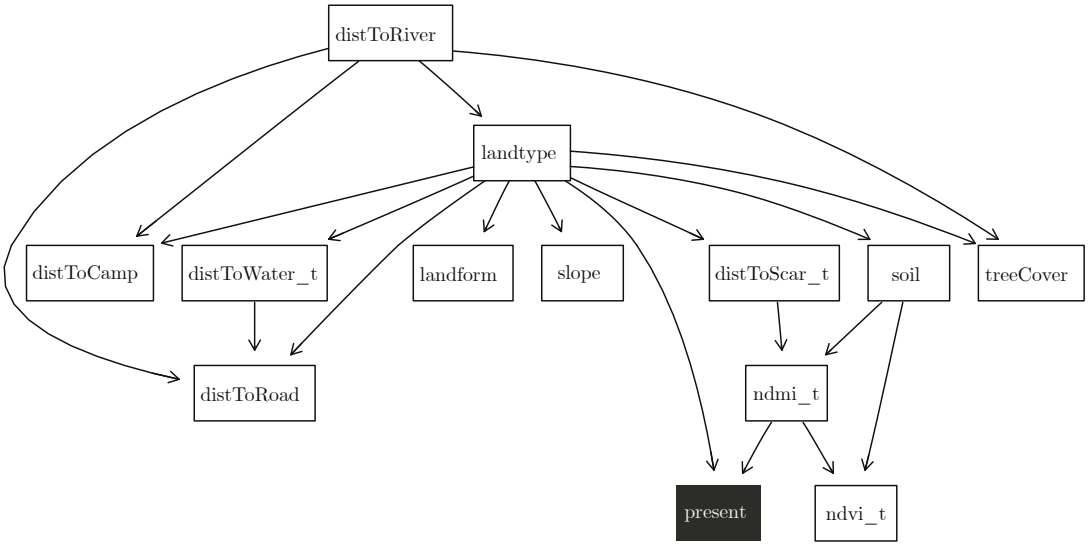


Figure 4.3: Model following Hartemink's discretisation method using a hill climbing algorithm.

A consistent network structure is achieved in Figure 4.3 with anticipated relationships, for example the distance to rivers influences the distance to camps, distance to roads and percentage tree cover. Already in 1928, the first so-called rest huts, namely Satara, Pretoriuskop and Skukuza were built. Most of the camps were constructed on river banks so as to overlook a stream and it thus makes sense that there exists a directed arc from `distToRiver` to `distToCamp`. The first tourist services introduced in 1923 were exclusively limited to rail transport with the first roads constructed from 1927 onwards (Joubert, 2015). Roads were often constructed to follow the path of a river to provide a more enjoyable game-viewing experience to tourists, consistent with the directed arc from `distToRiver` to `distToRoad`. Large trees (taller than 10 m) have generally been found to be closer to the main river courses (Shannon et al., 2008) most probably as soil moisture and nutrient levels are higher (Scholes and Archer, 1997) accounting for the directed arc from `distToRiver` to `treeCover`.

The structure in Figure 4.3 suggests that vegetation moisture (Normalized Difference Moisture Index (NDMI)) is influenced by soil type and distance to the nearest previously burnt area. In turn, vegetation moisture affects vegetation greenness (NDVI) and the chances of observing a rhino. Although the distance to the nearest scar does not directly impact the chances of observing a rhino, it does impact vegetation moisture, which in turn influences rhino observations.

Although many relationships exist between nodes with the same data source, such as `distToScar_t`, `ndmi_t` and `ndvi_t` which are all derived from Landsat-8 products, relationships between nodes from different data origins were also formed. For example, soil type as per Gertenbach (1983)'s definition influenced both vegetation moisture and greenness derived from remotely sensed imagery. This gives confidence in our data set

as the hill-climbing algorithm successfully detected interactions between variables with different data origins.

4.5 Inference in Bayesian network

The `bnlearn` package in R provide algorithms for both exact and approximate inference. Exact inference is obtained by giving exact posterior predictions computed using the closed-form formulas for naive Bayes, whereas approximate inference is achieved using either Logic Sampling or Likelihood Weighting.

Using the the network in Figure 4.3, the current probability to observe a rhino in a given cell is 0.5075 and observing an unlabelled cell is 0.4925. Although we sampled the same number of unlabelled points as presence points, missing values were introduced resulting in the unequal probability. It is also useful to note that this 50/50 probability is not representative of the real world. There are only an estimated 7235 (95% Cl: 6649–7830) white rhinoceroses living in Kruger National Park (KNP) during 2016 (Ferreira et al., 2018). Under the forgiving assumption that each rhino will be alone in a 1 km^2 block, the distribution translates to a probability of 0.3713 to observe a rhino.

Nonetheless, imagine we observe a piece of land within four kilometres of water. In the network in Figure 4.3, there exists no direct arc from `distToWater_t` to `present`, but by virtue of its membership in the Markov blanket, providing the network with evidence of being within 3.47 km of water, the probability of observing a rhino increases from 0.5075 to 0.5383. The divergent path ensures that the confounding variable `landtype` d-seperates `distToWater_t` and `present`. As `landtypes` is not instantiated, evidence is propagated from `distToWater_t` to `present`.

Providing evidence for `distToWater_t`, the probability of `present` decreases stepwise all the way to 0.3304 for the interval (9.26, 23.2] showing that white rhinos prefer to be closer to surface water. This is consistent with Smit et al. (2007)'s finding that high-density grazers prefer to be within 4 km of waterholes.

Again, imagine we observe a piece of land and know to which landform it belongs. In fact, we are observing a *warm lower slope*. If we present this evidence to the Bayesian network in Figure 4.3, the probability of observing a rhino changes from 0.5075 to 0.5772. On the other hand, imagine we observe a piece of land, standing in a *flat lower slope*, the probability decreases drastically from 0.5075 to 0.3889. These changes in the probability of observing a rhino operate indirectly as `landform` is not depicted in the network with an arc, but influences the probability of `presence` through the relationship with `landtype`. Evidence from `landform` is propogated through the uninstantiated `landtype` to `present`, illustrating a distinctive feature of Bayesian network modelling.

When we change the observed soil type to *rhyolite soils*, the probability of observing a rhino changes from 0.5075 to 0.2069. Observing *gabbro soils*, *granitic soils*, *sandy soils*, *basalt soils* and *ecca soils* change the probability of observing a rhino to 0.5450, 0.5671, 0, 0.3706 and 0.3602, respectively.

When we provide the network with new evidence of distance to the nearest scar being in the interval of [0, 7.97], a slight increase from 0.5075 to 0.5130 is observed. The interval (7.97, 15.9] of distance to the nearest scar reduced the probability of `present` to 0.4301 and the interval (15.9, 39.8] reduced the probability of `present` to 0.3762. This suggests that generally white rhinos prefer to be closer to previously burnt areas.

4.5.1 More inference

Using the network in Figure 4.3, imagine we observe a white rhino in a given area. What does this new evidence tell us about the rest of the network? Setting `present` to positive, evidence propagates through the entire network to provide the most probable states of each node as illustrated in Figure 4.4. The *Orpen - SA05* becomes the state of landtype with the highest probability, suggesting that we might observe Timbavati gabbro with moderately dense *Combretum apiculatum* and *Combretum zeyheri*. We are likely within 5.43 km from the nearest river and within 3.47 km of water. The nearest rest camp is probably between 10 and 19.8 km from us and the nearest road between 3.48 and 10.4 km from us. We are either standing on a warm upper slope or a warm lower slope, but presumably not on a flat slope. The percentage of tree cover is most likely between 5.50 and 8.88 % and we are within 7.97 km from the nearest previously burnt area. Vegetation moisture is most probably the lowest possible, between -0.279 and -0.135 and the NDVI the second lowest possible (between 0.218 and 0.302).

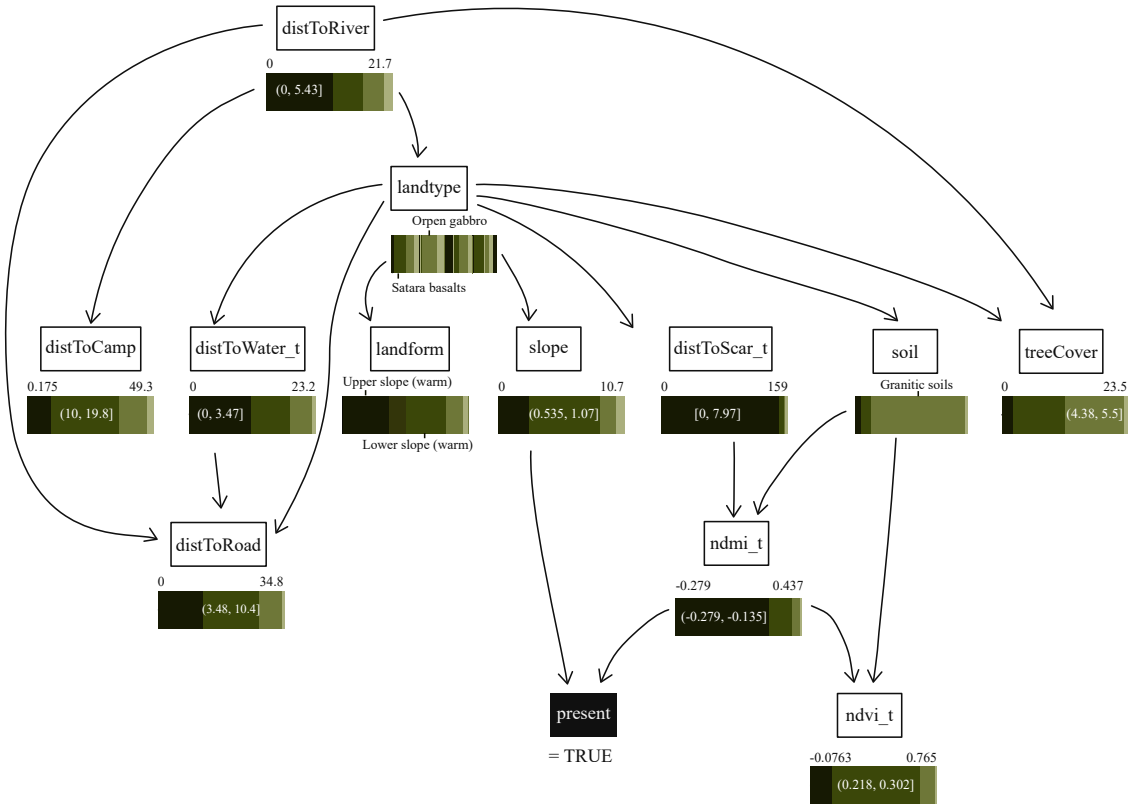


Figure 4.4: Inference when providing evidence of observing a rhino to the network.

4.5.2 Interval discretisation

Thus far, we followed Hartemink's discretisation approach. Experimenting with different discretisation techniques, we deployed *interval discretisation* with 22 intervals resulting in a model learned as shown in Figure 4.5. Using the exact same variables and structure learning algorithm as the model in Figure 4.3, this model preformed similar, achieving a cross validation accuracy of 68.81 % and an accuracy on the formulated test set of 70.16 %. This illustrated a slight increase in the generalisation of models. We reversed an arc from `present` to `distToScar_t` and from `distToCamp` to `landform`.

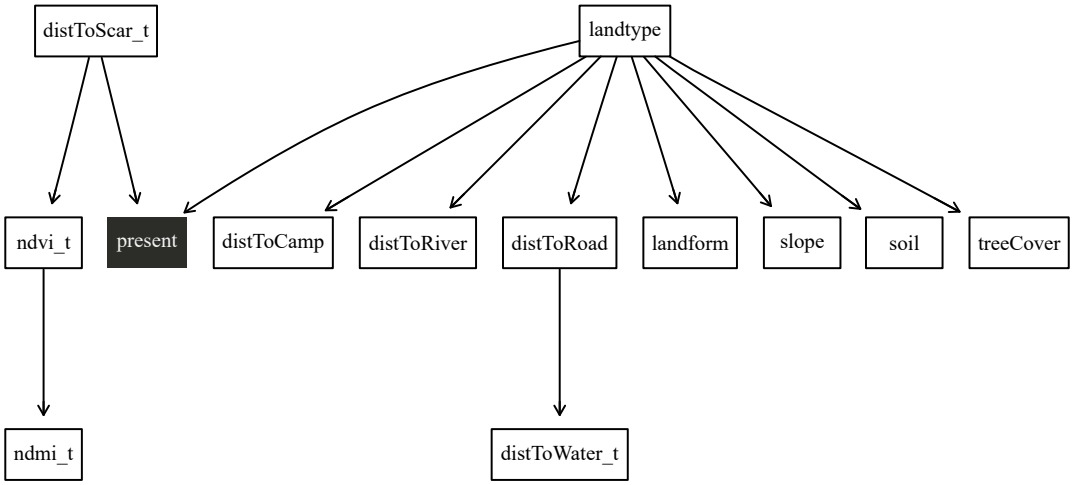


Figure 4.5: Model learned following an interval discretisation method.

Even though the structure in Figure 4.3 learned using Hartemink's approach and the structure in Figure 4.5 using interval discretisation differ substantially, similar accuracies were achieved. For any single directed graph selected as the most optimal according to a score-based algorithm, there will in general be many other possible structures consistent with the same conditional independencies (Chickering, 2002).

4.6 Summary

Table 4.2 summarises all measures for (i) the basic Bayesian network with four variables, (ii) the Bayesian network constructed on data discretised following Hartemink's approach and (iii) the Bayesian network following interval discretisation. Here, we found that the hill climbing algorithm, a score-based method resulted in more expected structures than the growth-shrink algorithm, a constraint-based method. Interesting relations were revealed, not only between environmental variables the probability of observing a rhino, but also among different environmental variables. We successfully illustrated how to perform simple inferences and illustrated that discretisation of continues covariates can be performed using either Hartemink's approach or a interval method.

Table 4.2: A summary of simple networks.

	Basic (Figure 4.1)	Hartemink (Figure 4.3)	Interval (Figure 4.5)
Algorithm	Growth shrink	Hill-climb	Hill-climb
Data included	All	All	All
Discretisation	Hartemink	Hartemink	Interval
Number of nodes	4	13	12
Number of arcs	6	19	11
Log Likelihood	-268791.8	-569938.4	-854575.2
BIC	-277074.7	-597388	-894063.3
Cross validation classification error	0.3038	0.3150	0.3119
Cross validation classification accuracy	0.6962	0.6850	0.6881
Accuracy on formulated test set	0.7287	0.7008	0.7016
95 % CI on formulated test set	(0.6434, 0.8032)	(0.6132, 0.7788)	(0.6129, 0.7804)
Sensitivity on formulated test set	0.7564	0.7273	0.7467
Specificity on formulated test set	0.6863	0.6600	0.6327
Computational time (seconds)	0.9744	0.5336	0.1924

Chapter 5

Extending the Bayesian network

And yet it moves.

Galileo Galilei (1564-1642)

This chapter extends on the Bayesian networks we constructed in the previous chapter by introducing variables operating in the previous timestep. We challenge the relationship between fire and rhino presence by only looking at tracking data. We then explore networks constructed using missing data after applying the EM-algorithm. Finally, we construct Bayesian networks for individual rhinos.

5.1 Adding time

The Landsat 8 satellite collects images of the Earth with a 16-day repeat cycle. This enables us to derive environmental indices in timesteps of 16 days. In the previous chapter, we included the variable `distToScar_t` in an attempt to capture a white rhino's possible preference for grazing on burnt stubble and freshly sprouting grasses. Including variables from a previous timestep may allow us to make inference for the following timestep. We introduced the distance to the nearest fire in the previous timestep ($t = 16$ days) and the soil moisture in the previous timestep ($t = 16$ days). We also include distance to the nearest fire in the current timestep, expecting it to capture something about white rhino's fear response to burning fires.

Table 5.1: Added variables used in Bayesian network learning.

Label	Variable name	Values discretised by Hartemink's spproach
`distToFire_t`	Distance to fire in this timestep (t=16 days)	[0,50]; (50,100]; (100,950]; (950,1000]
`distToFire_t1`	Distance to fire in the previous timestep (t=16 days)	[0,50]; (50,300]; (300,950]; (950,1000]
`soilMoisture_t1`	Soil Moisture in the previous timestep (t=16 days)	(-0.0736,0.726]; (0.726,0.776]; (0.776,0.826]; (0.826,0.927]

Again following Hartemink's discretisatisation approach, we constructed a Bayesian network learned using the hill climb algorithm. To be consistent with our temporal

assumptions, we reversed an arc from `distToFire_t` to `distToFire_t1` and also from `distToCamp` to `distToRiver` and `landtype`. Figure 5.1 shows the learned network with 14 nodes and 22 arcs achieving a cross validation accuracy of 72.04 % and an accuracy of 69.75 % on the formulated test set using a sensitivity threshold of 0.48.

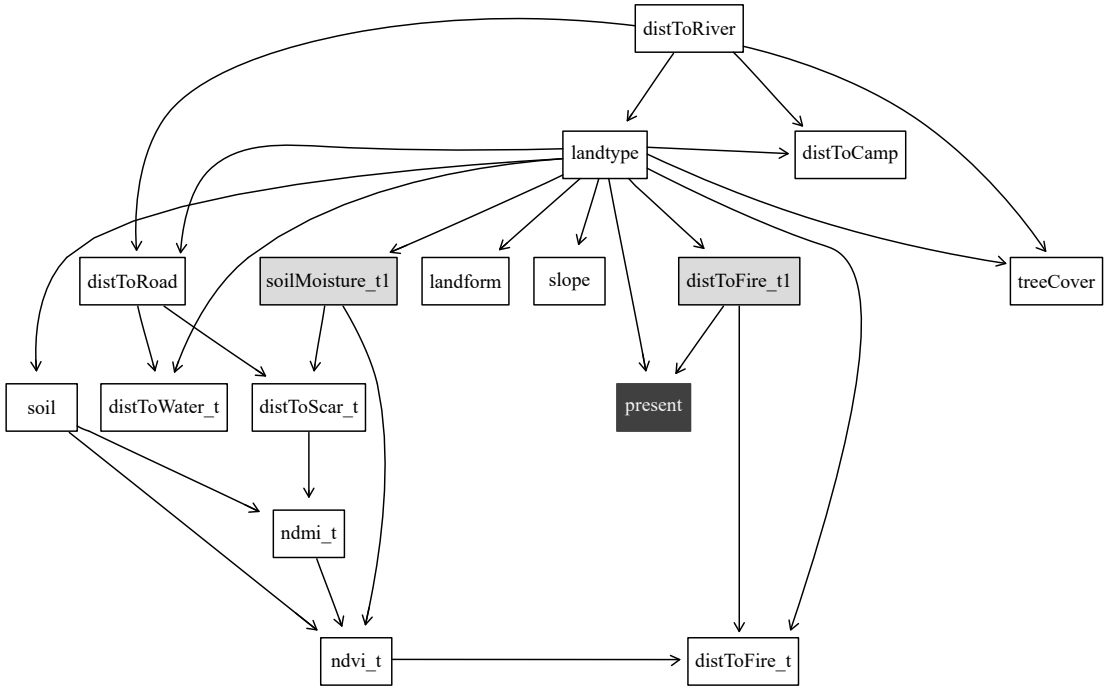


Figure 5.1: Hill climbing with more variables.

Presenting evidence to the network in Figure 5.1 of being within 50 km of a historic fire, the probability of observing a rhino changes from 0.5081 to 0.6124. When between 50 and 300 km the probability decreases to 0.2621. When between 950 to 1000 km from a historic fire the probability of observing a rhino shifts to a more random value of 0.4920 representative of our data, as we set each pixel's value to 1000 km when there was no fire in Kruger National Park (KNP).

Providing evidence to the network of being within 50 km of the nearest fire in the current timestep (t = 16 days) increased the probability of observing a rhino from 0.5081 to 0.5372 not entirely suggestive that rhino's prefer to be further from fires. More likely, to fully capture the expected fear of fire a higher temporal resolution is required, as many fires burn shorter than 16 days. A fire burning the first five days will still be observed as a fire in the 16-day period, even though it may already be smouldering on the sixth day and be cold the days thereafter (Cahoon Jr et al., 1992).

5.1.1 No aerial data

Data from an aerial origin falls victim to a number of biases. Specifically, availability bias, observer bias, and detectability bias. Most previous models, like the one in Figure 5.1, showcased a strong relationship between `distToFire_t1` and `present`. Suspecting that this might be a result of visibility sampling bias during aerial surveys, we constructed another Bayesian network, following the exact same approach as for the model in Figure 5.1, except we limited the data to exclude all aerial survey records. This model, shown

by Figure 5.2 shows a model trained by excluding all aerial survey data with 15 nodes and 23 arcs achieved the highest cross validation accuracy of 81.66 %. However, the same model achieved an accuracy on the formulated test set of only 60.23 % indicating that the model had difficulty generalising and making inference for data from sources other than tracking.

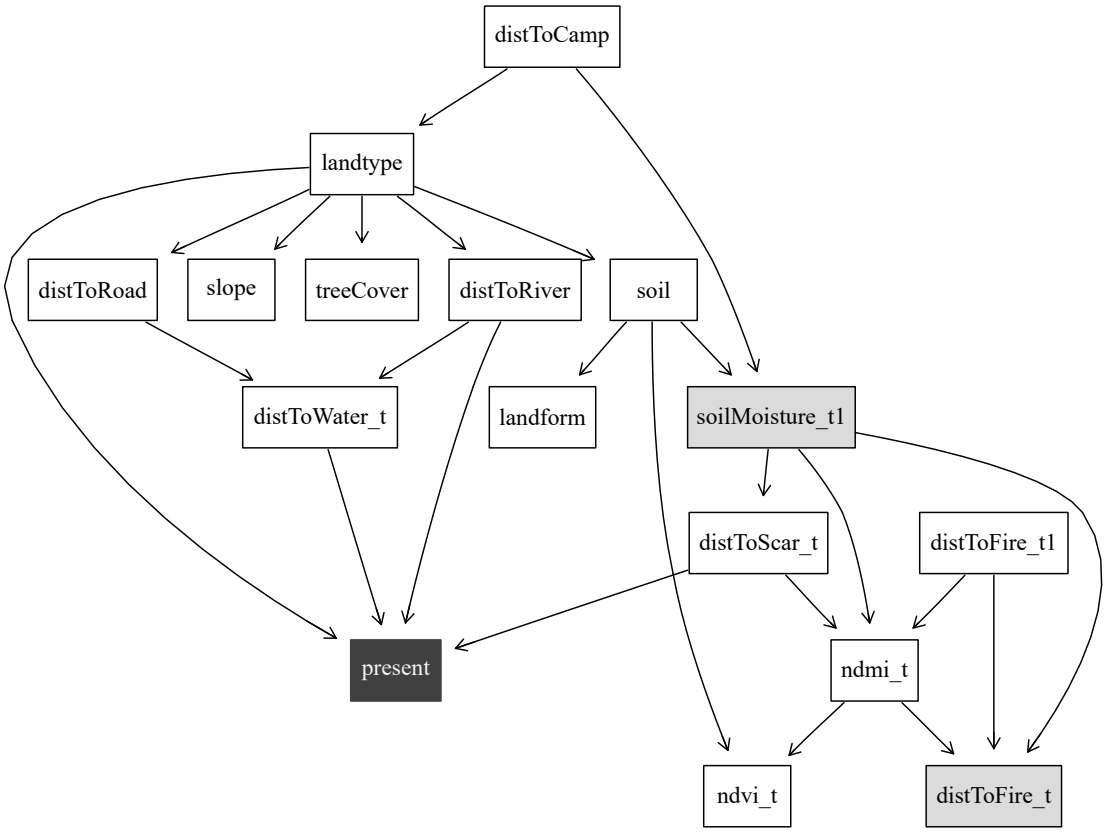


Figure 5.2: Model learned excluding all data from an aerial source.

Interestingly, the strong relationship between `present` and `distToFire_t1` does not exist, but there is well a relationship between `present` and `distToScar_t` indicating that historic fires might well play a significant role.

5.1.2 Only aerial data

We also constructed a model only using aerial survey data, namely aerial survey data from annual surveys and from demographic studies. Even though these data were temporally constrained as it only contained values for February, September and October it showed to be best generalising across the test set, achieving a cross validation error of 71.83 % and an accuracy of 74.44 % on the formulated test set.

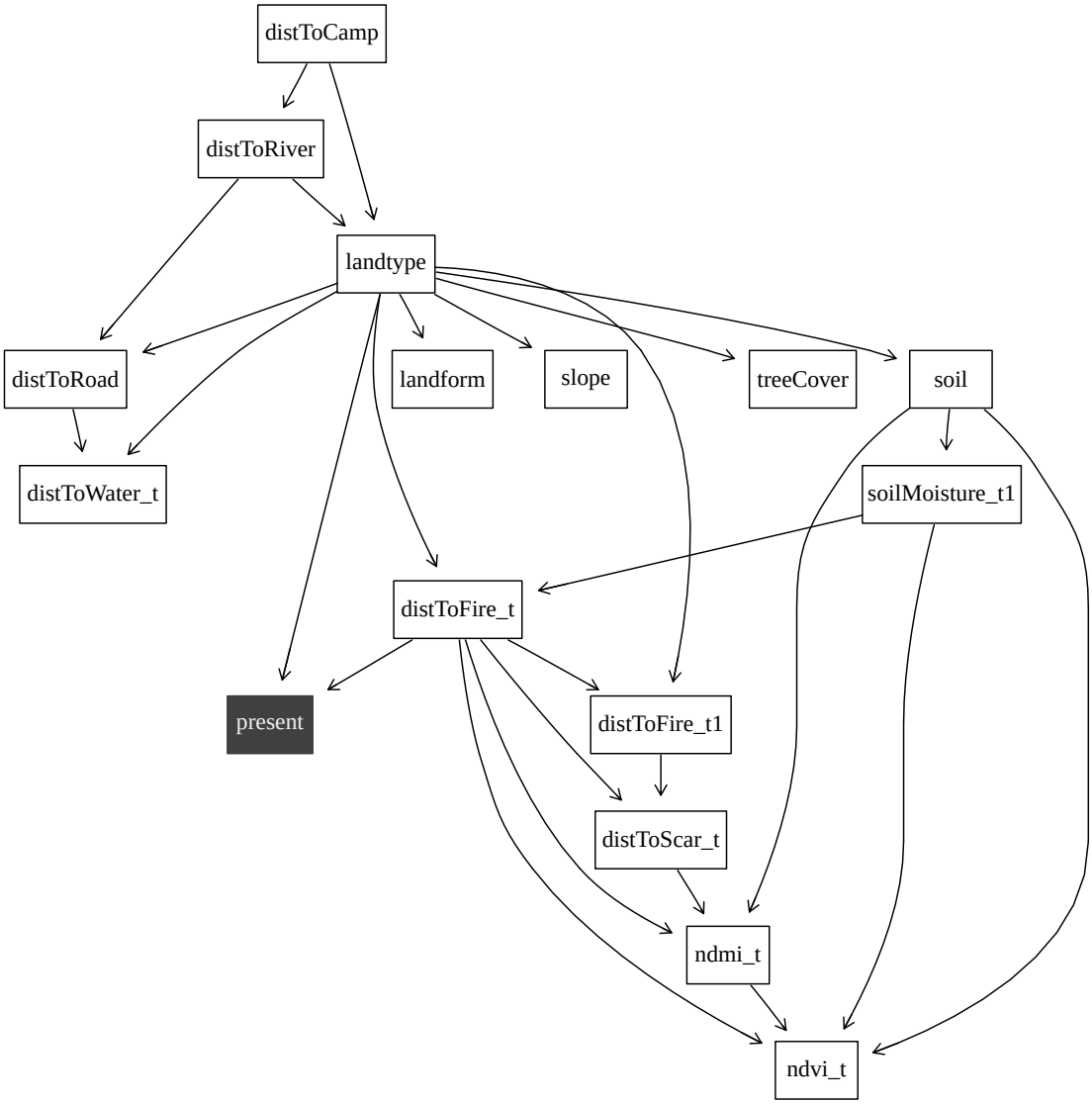


Figure 5.3: Model learned only using data from an aerial source.

5.1.3 Summary thus far

Table 5.2 summarises performance measures for (i) the Bayesian network with the inclusion of two variables operating in the previous timestep, (ii) the Bayesian network constructed by excluding all aerial data and (iii) the Bayesian network constructed on only aerial data.

Table 5.2: A summary from networks with many covariates from different presence sources.

	Adding time (Figure 5.1)	No aerial (Figure 5.2)	Only aerial (Figure 5.3)
Algorithm	Hill-climb	Hill-climb	Hill-climb
Data included	All	Track and iNat	Aerial
Discretisation	Hartemink	Hartemink	Hartemink
Number of nodes	16	16	14
Number of arcs	26	24	21
Log Likelihood	-705908	-127088.8	-552977.8
BIC	-740999	-148647.4	-583290.7
Cross validation classification error	0.2856	0.1834	0.2817
Cross validation classification accuracy	0.7138	0.8166	0.7183
Accuracy on formulated test set	0.7250	0.6023	0.7444
95 % CI on formulated test set	(0.6360, 0.8025)	(0.4923, 0.7051)	(0.6416, 0.8306)
Sensitivity on formulated test set	0.7123	0.5593	0.7455
Specificity on formulated test set	0.7447	0.6897	0.7429
Computational time (seconds)	0.0867	0.0887	0.0910

Why is Normalised Difference Vegetation Index (NDVI) less important?

In most of the preceding experiments `ndvi_t` was rarely found to directly impact the chances of observing a rhino. Why is NDVI less important once we introduce more variables? To answer this question, we need to revisit the fact that the white rhino is a bulk grazer, feeding selectively on types of grass. Even though NDVI is an acceptable measure of leaf area, especially in savannas or other open areas where saturation values are rarely if ever reached (Archibald and Scholes, 2007), it does not distinguish between woody and herbaceous biomass. A pixel with a high NDVI value might just have dense tree cover and will not quantify the availability of quality food for the white rhino.

In Figure 4.3 where we introduced only three covariates to the model, `ndvi_t` does impact the chances of observing a rhino suggesting that NDVI still warrants its inclusion as a variable in networks. The inclusion of other variables may configure the relationships between NDVI for grasses and trees. Soil moisture and rainfall allow cues responsible for leaf flush in grasses (Archibald and Scholes, 2007), whereas relative humidity, day length, temperatures and soil moisture cue leaf flush in savanna trees (Do et al., 2005). Archibald and Scholes (2007) also found that grasses have a higher Leaf Area Index (LAI) in the peak of the growing season, but trees turn green earlier, and stay green for longer. Specifically, grass green-up occurs on average 53 days after tree green-up in the Skukuza region with trees flushing on average on the 100th day of the year, ranging 7.10 days ($\sigma = 9$) and grass ranging 30.11 days ($\sigma = 29$).

Parker and Witkowski (1999) suggested that extensive high grazing near waterholes homogenises the herbaceous layer. Animals who tolerate sparse surface water lived happily in their arid landscapes, until artificial waterholes were introduced. Water-dependent antelope such as Burchell's zebra, *Equus burchelli*, were then able to stay for longer periods in arid landscapes with lions and other predators following them. A combination of increased competition and higher number of predators ensured that roan and other rare antelope's numbers drastically declined (Grant et al., 2002). Although we hoped to

demonstrate this relationship of poorer vegetation quality close to surface water, we could not find a satisfactory relationship between distance to surface water and vegetation quality. This can be that we operate at a spatial resolution to coarse to spot these differences. It is also possible that NDVI is just not a good indicator of herb homogenises, and canopy height may rather be evaluated.

5.1.4 Defining the structure

Literature showed that using prior knowledge drawn from literature and expert knowledge can improve model efficiency. Borrowing from all previous model structures learned and consulting literature and expert knowledge, we built another Bayesian network, by first using the hill climbing algorithm to learn the structure and then manually changing and adding arcs manually.

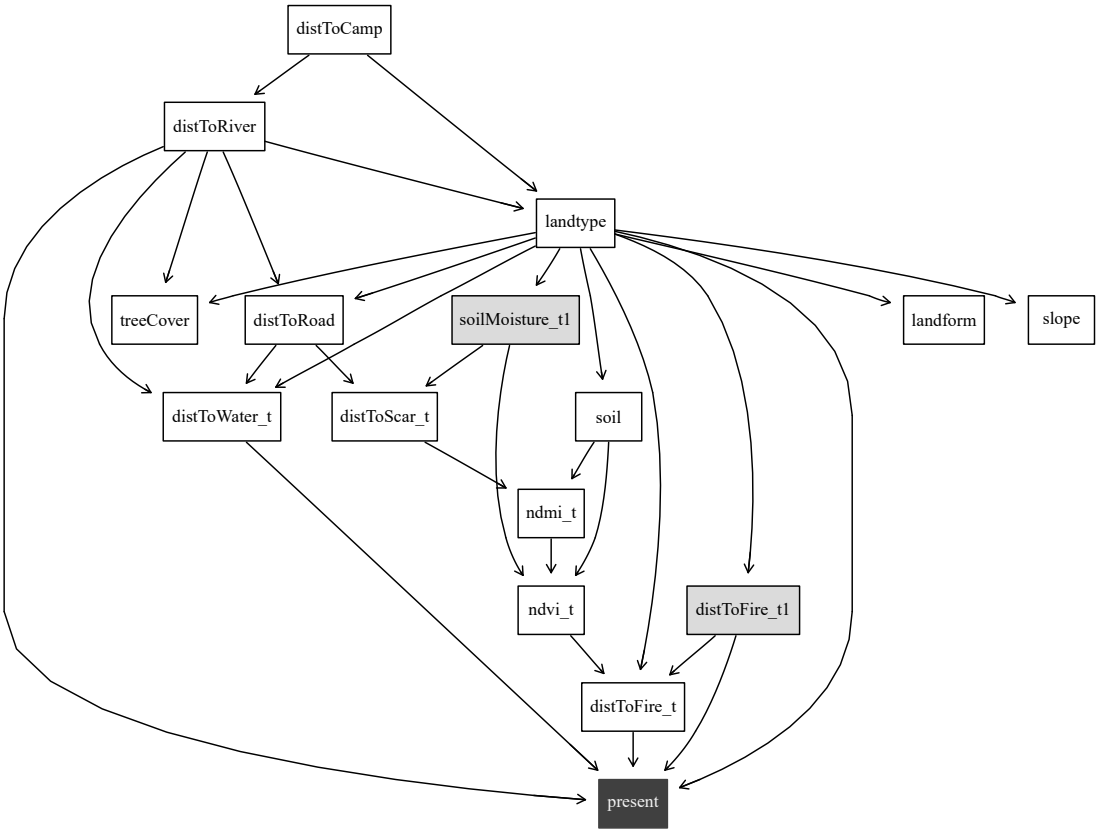


Figure 5.4: Defining the structure.

The model in Figure 5.4 achieved a cross validation accuracy of 70.87 % and an accuracy of 65.29 % on the formulated test set, illustrating that manual expert input is valuable to build networks.

5.2 Missing data

The *Markov property* of Bayesian networks enforces modularity facilitating the learning of the underlying structure. The likelihood function can be decomposed into a product of terms depending only on the conditional probability parameters and a node's parents.

This allows for small changes such as a change in direction for an arc to be made locally without the need to re-evaluate the entire network. However, this decomposition is only possible when data are complete. One problem we have ignored so far in learning Bayesian networks, is how we dealt with missing data.

We assumed the data adhered to being missing-completely-at-random (MCAR), in that we supposed the mechanism resulting in missing values depends neither on the observed data nor on the missing data and we simply discarded any records containing missing values.

This MCAR assumption may not hold as many variables in our data were unavailable for sampling during the introduction of Algorithm 1 to mask clouds on the Landsat 8 image collection. This suggests that the mechanism resulting in missing values in our case do depend on the data as it is more cloudy during seasonal changes. Rather, our data follow a missing-at-random (MAR) assumption where the missing value mechanism indeed depends on the observed data. When we discarded incomplete data, we inherently did so non-arbitrarily making values variables can take unavailable to be part of the learning process.

The MAR assumption also means missing values can be estimated from the observed data. In fact it is even possible that our data adhere to the missing-not-at-random (MNAR) assumption where the missing value mechanism depends on both the observed and the missing data.

To regulate the planet's energy levels, clouds cover around two thirds of the Earth's surface at any given time (King et al., 2013). Inevitably, clouds obscure the accurate capture of the earth's surface and any remotely sensed index will return missing values. With almost two thirds of our data discarded, statistical power is lost and ultimately the joint distribution may be skewed. We also found that the Landsat 8 cloud mask performs well in excluding clouds, but the Modified Normalised Difference Water Index (MNDWI) falls victim to picking out cloud shadows as water sources.

Learning parameters from partial records is an easier problem compared to finding the underlying structure from partial records. The popular expectation-maximization (EM) algorithm can be used to find maximum likelihood estimates of parameters from incomplete data under the MAR assumption. Friedman et al. (1997) extended the EM-algorithm (Dempster et al., 1977), as a method to learn the structure of a network from incomplete data. Deterministic approaches such as the EM-algorithm often tend to only find local maxima.

Imputed values are computed by plugging new values for the parents of node in the local probability distribution of the node extracted from the fitted network. The imputed values were computed by averaging likelihood weighting simulations using all the relevant variables as evidence. 500 random samples were averaged for each new observation, with the predicted level taking the highest conditional probability. The random seed was set to ensure reproducible results as this imputation method is based on a stochastic simulation and may produce different structures.

Using the EM-algorithm we trained a model on a dataset containing partial records, discretised through the interval method. Computational time increased significantly from seconds to 4.36 minutes. The resulting network in Figure 5.5 consisted of 14 nodes and 21 arcs achieving a cross validation accuracy of 70.87 % and an accuracy of 65.29 % on the formulated test set using a sensitivity threshold of 0.41. To ensure the temporal validity of the model we reversed an arc from `ndmi_t` to `soilMoisture_t1`.

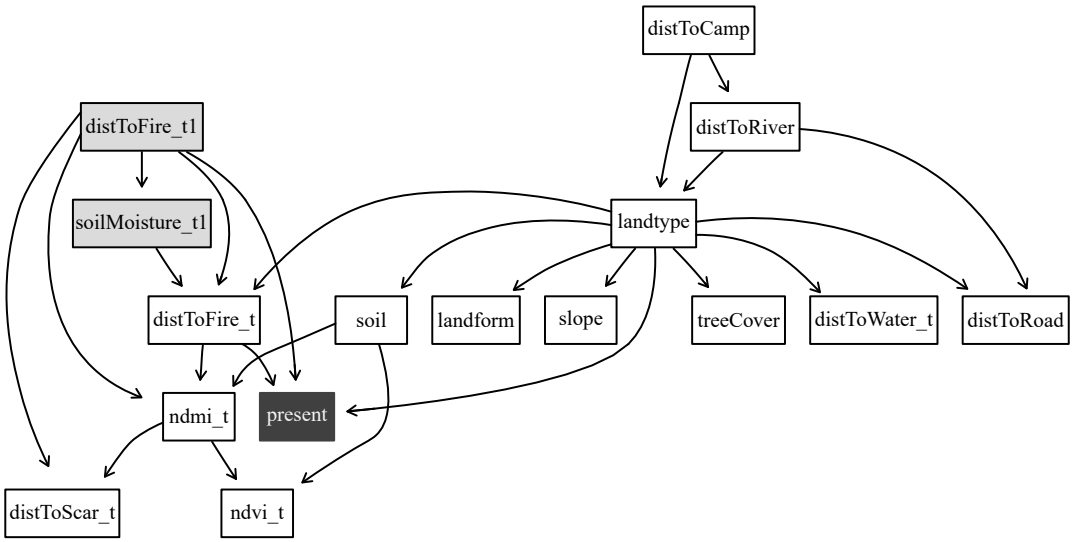


Figure 5.5: Structure learnt from incomplete data using the EM-algorithm.

The subsequent incorporation of missing data through application of the EM algorithm provides increased possibilities to detect underlying relations. Specifically, the direct arc from `distToFire_t1` to `distToScar_t` was revealed for the first time. This illustrates that keeping incomplete records, kept more values for `distToScar_t` that were previously discarded through the cloud masking algorithm on Landsat 8.

Table 5.3 summarises the performance results of (i) the Bayesian network with a manually defined structure and (ii) of the Bayesian network constructed with the EM-algorithm on missing data.

Table 5.3: A summary from simple networks using different structure learning algorithms and discretisation approaches.

	Defining the structure (Figure 5.4)	Missing data (Figure 5.5)
Algorithm	Hill-climb and manual	EM
Data included	All	All with incomplete records
Discretisation	Hartemink	Interval
Number of nodes	16	14
Number of arcs	30	21
Log Likelihood	-694820.5	-501648
BIC	-804505.5	-538144
Cross validation classification error	0.2250	0.2913
Cross validation classification accuracy	0.7750	0.7087
Accuracy on formulated test set	0.6833	0.6529
95 % CI on formulated test set	(0.5922, 0.7652)	(0.5610, 0.7371)
Sensitivity on formulated test set	0.7260	0.6800
Specificity on formulated test set	0.6170	0.6087
Computational time (seconds)	0.0867	261.75

5.3 Inference in space

Here, we demonstrate the practical application of using a Bayesian network for inference purposes. Specifically, we aim to provide a map of pixels with the highest probability of

having a rhino present to allow decision makers to optimise their patrol routes. Similarly, we aim to highlight areas most likely to observe rhinos and areas least likely to observe rhinos.

First, we obtained covariate data for a test period. Our test period coincided with 16 August 2018, overlapping with the periods as show in Table 5.4.

Table 5.4: Dates for test period.

	Start	Target date	End
t	2018/08/08	2018/08/16	2018/08/24
t-1	2018/07/23	2018/08/01	2018/08/08

Using the model produced by the EM algorithm shown in Figure 5.5, the following covariates were included and obtained for the periods: `present`, `landform`, `landtype`, `soil`, `ndvi_t` , `distToWater_t` , `distToCamp` , `distToRoad` , `distToScar_t`, `slope` , `treeCover`, `distToFire_t` and `distToFire_t1`, `soilMoisture_t1`.

First, we translated the raster stack into vector data and discretise each variable's values using the same discretisation parameters used during model training. Realising the computational cost of inference, we cropped the raster stack to be within -23.0 and -22.8 south and between 31.0 and 31.2 east in KNP. Providing this discrete data as evidence to the model, we inferred the probabilities for the node `present`. After inference took place, we transformed the resulting vector back into a raster. Even for this relative small area, inference took 6.538 hours to compute. Figure 5.6 shows the raster evidence provided to each variable (except `present`) in the network to ultimately infer the raster of most likely values for `present`. There were no fires for this period.

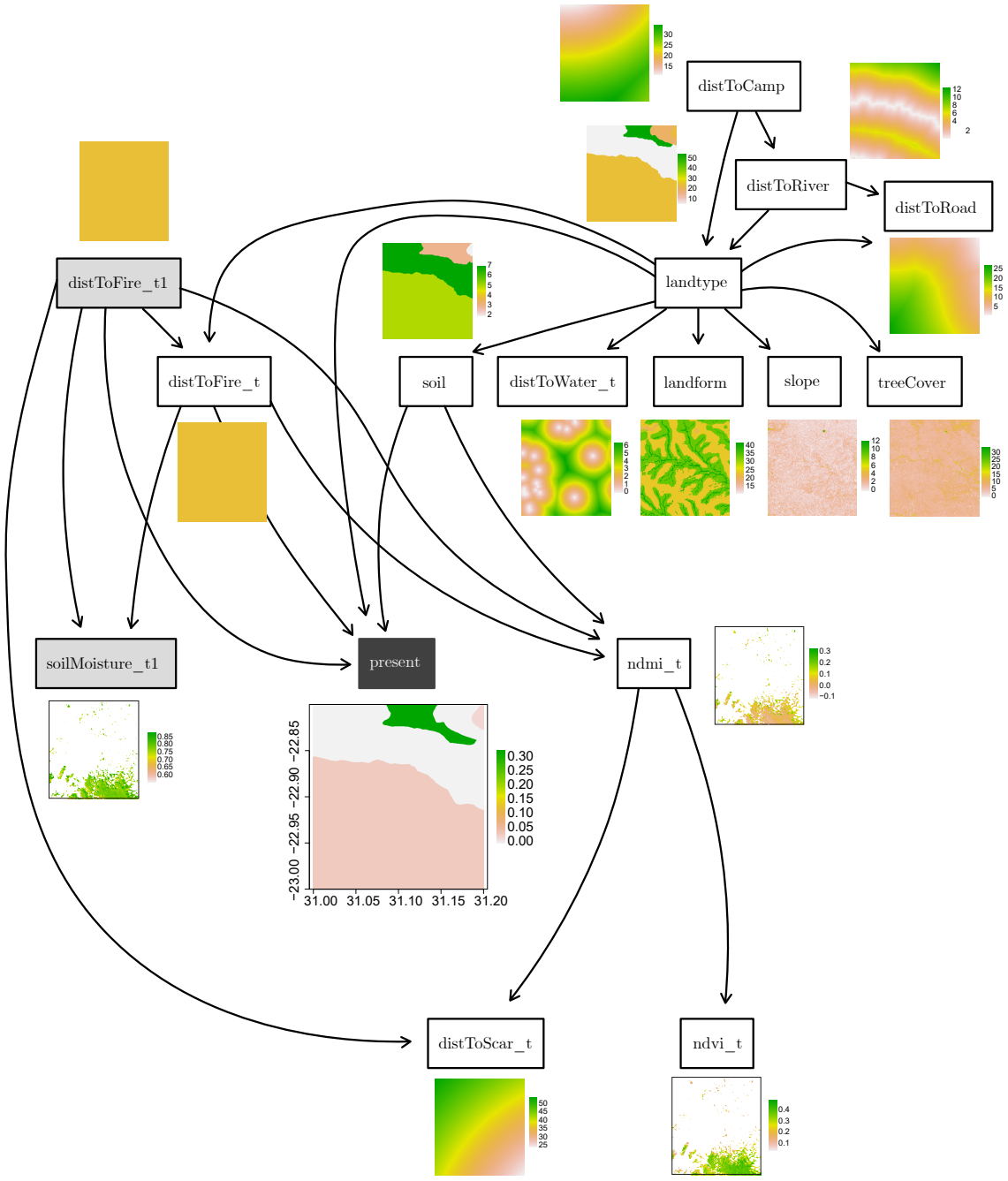


Figure 5.6: Spatial inference after providing a Bayesian network with new evidence.

5.4 Causality

Causal semantics suggest each directed arc can be seen as representing a causal association from each parent to its children. It remains tempting to claim that all Bayesian networks represent causal relationships as we have implied thus far. A causal network does not only provide mechanistic insight into a system but also allows for effective interventions (e.g. burning a landscape, patrolling a specific landform) to be accurately predicted.

Unfortunately, the possibility of unobserved variables, also known as *latent* or *hidden*

variables usually preclude causal interpretation. For a Bayesian network to be interpreted as representing causality, the following needs to be assumed: (i) The Markov condition of Equation (2.1) extends to the notion of causality: Given the affects of a variable of its immediate cause, that variable is independent of all earlier causes; (ii) if any two variables are correlated, then it must be assumed that one is the causal parent of the other or there is a third variable causing both; and (iii) there must be no variables missing from the network that are causal parents of two or more of the variables contained in the network.

However, it is not necessary to interpret a Bayesian network as causal to extract meaningful information from the learned structures. In fact, researchers studying the interactions between genes, the environment and disease interpreted learned structures not as causal but rather allowed counter-intuitive parent-child relationships. Departing from a strict causal interpretation enhanced their ability to discover complex associations between genes and the environment (Su et al., 2013). This type of structure may more accurately represent the mechanisms that generated the data in which covariates are chosen primarily because a rhino was observed in the cell (despite our attempts at unlabelled sampling) and not at random. We will refer to networks built without deliberately reversing arcs as being *non-causal*.

5.5 Individuals

Thus far, we focused primarily on population distributions in the hierarchy of animal behaviour modelling in Figure 1.1. In the previous section, we obtained some insight in the influence of individual choices on population level behaviour when we constructed models excluding all aerial survey data. Individual differences such as age, sex, phenotype and experience can significantly influence decision making and generalising across differences in populations can mask important behaviour (Holyoak et al., 2008). However, studies linking individual movement behaviour to population dynamics are rare in literature. Here, we explore this notion further by learning *non-causal* Bayesian network from individual rhino's tracking data. This type of structure is of specific interest in modelling individual behaviour as the tracking data generating mechanism is controlled by the animal walking on a set path. Recall that an unlabelled point for each tracking record was sampled assuming the rhino had a choice of walking 10 km in one setting. We learned Bayesian networks using the hill climbing algorithm using the tracking data of six individuals with the covariates discretised following Hartemink's approach. The results of each individual network is summarised in Table 5.5.

Table 5.5: Performance of Bayesian networks learned on individual rhino data.

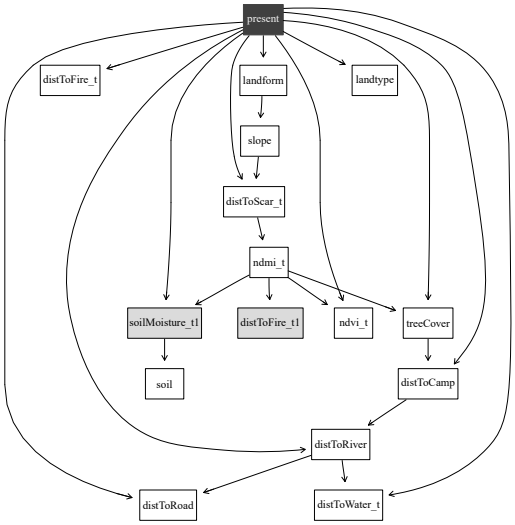
	Bull A (Figure 5.7a)	Bull C (Figure 5.7c)	Bull E (Figure 5.7e)
Sex	Male	Male	Male
Number of nodes	16	16	16
Number of arcs	23	23	17
Log Likelihood	-10785.54	-19081.29	-13436.4
BIC	-12079.18	-20382.37	-15079.81
Cross validation classification error	0.5295	0.404	0.3752
Cross validation classification accuracy	0.4705	0.596	0.6248
Accuracy on formulated test set	0.4405	0.4694	0.3908
95 % CI on formulated test set	(0.3322, 0.5530)	(0.3678, 0.5729)	(0.2879, 0.5013)
Sensitivity on formulated test set	0.1786	0.1905	0.2069
Specificity on formulated test set	0.9643	0.9714	0.7586

	Cow B (Figure 5.7b)	Cow D (Figure 5.7d)	Cow F (Figure 5.7f)
Sex	Female	Female	Female
Number of nodes	16	16	16
Number of arcs	21	20	22
Log Likelihood	-19326.43	-20534.74	-18534.19
BIC	-21577.17	-21838.96	-20821.5
Cross validation classification error	0.457	0.4289	0.2572
Cross validation classification accuracy	0.5430	0.5711	0.7428
Accuracy on formulated test set	0.5484	0.4947	0.5000
95 % CI on formulated test set	(0.4417, 0.6519)	(0.3905, 0.5993)	(0.3939, 0.6061)
Sensitivity on formulated test set	0.3871	0.2540	0.3000
Specificity on formulated test set	0.8710	0.9688	0.8750

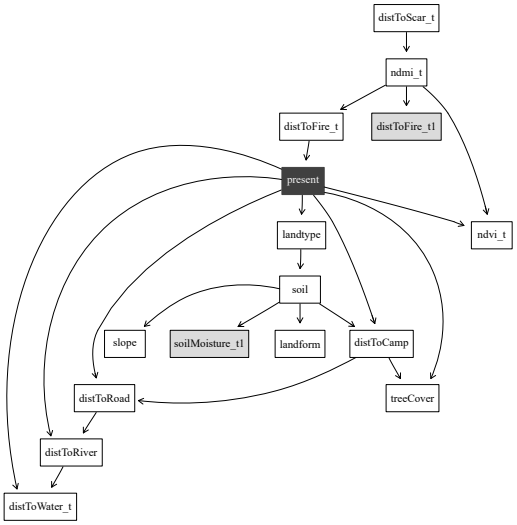
It is important to note that the covariates were not discretised with new thresholds, but instead the same bins were used that were used for all other models. This would have created bins for which there might not have been values present in the tracking dataset. The number of available tracking points possibly also affected the performance of the model, as we drastically reduced the number of records when sub-setting the tracking data per individual.

Interestingly, very unique structures were learned illustrating the internal traits and different environmental preferences as shown in Figure 5.7.

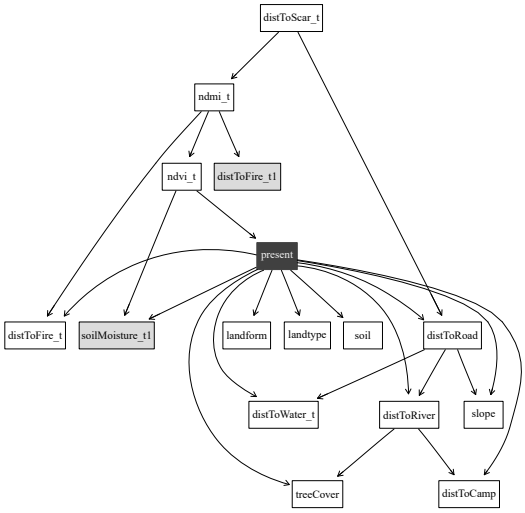
Although the sample size is extremely small, there existed a slightly higher predictability for female rhinos, with one cow achieving the highest cross validation accuracy being 74.28 %. Profoundly, all individual models performed extremely poor on the formulated test set with balanced accuracies close to 50 %. This poor generalisation stresses the importance of individual traits in ultimately understanding the behaviour of a population.



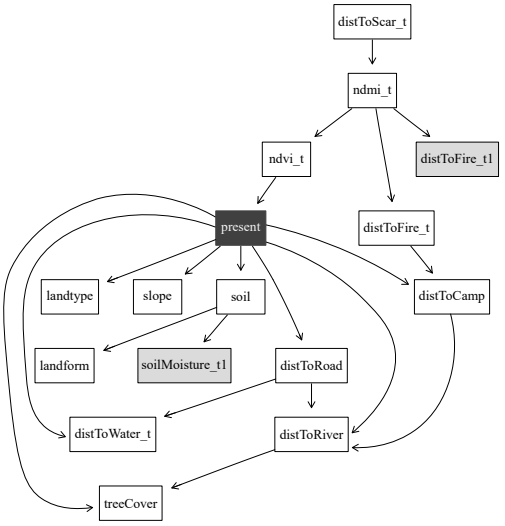
(a) Bull A.



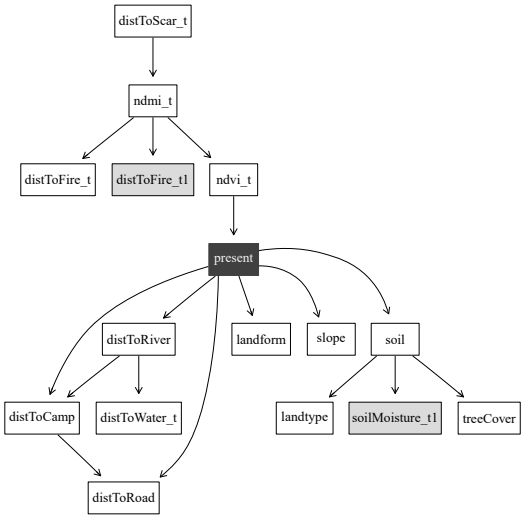
(b) Cow B.



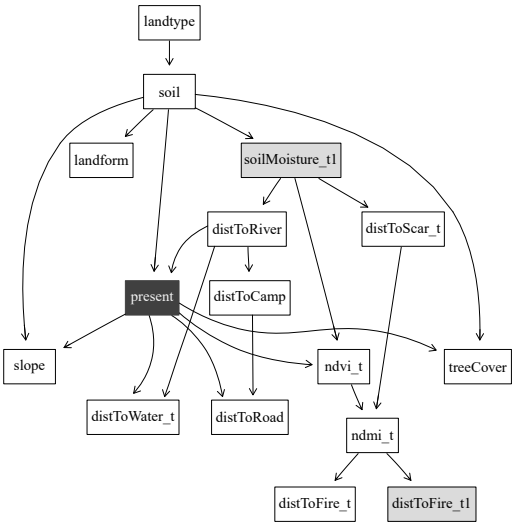
(c) Bull C.



(d) Cow D.



(e) Bull E.



(f) Cow F.

Figure 5.7: Bayesian networks for individuals.

Chapter 6

Conclusion

To be informed is to know simply that something is the case. To be enlightened is to know, in addition, what it is all about: why it is the case, what its connections are with other facts, in what respects it is the same, in what respects it is different, and so forth.

Mortimer J. Adler in *"How to Read a Book: The Classic Guide to Intelligent Reading"*

This chapter gives an overview of the findings of this dissertation, answering the research question proposed in the beginning of this dissertation. We conclude this study by discussing possible future research in the field.

6.1 Brief summary of findings

This dissertation provides a starting framework to capture white rhino presence using a method not only capable of dealing with uncertainty, but also with changes across time and space. We produced Bayesian networks providing new insight into the complexity of modelling animal presence, but also revealed relationships between underlying variables orchestrating environmental processes and ultimately influences white rhino's use of the landscape. These networks can be used to provide decision makers with a deeper understanding of a complex environmental system but also allow inference under uncertainty when new evidence is exposed. Specifically, understanding the relationship between a threatened species in Kruger National Park (KNP) and its dynamic environment, management can make more informed decisions of where to deploy anti-poaching resources.

Literature shaped our known understanding of white rhino behaviour and specifically their response to changes in the environment. We quantified our newly learned knowledge by mapping known facts to remotely sensed indices, producing our *first artefact* in the

form of a framework that maps behavioural drivers of white rhino presence to environmental variables. Rhino presence data were obtained and processed from different sources, specifically from aerial surveys, aerial demographic studies, tracking data and the social platform, iNaturalist. Through strategic sampling, we produced a rich dataset containing presences, unlabelled points and fifteen cloud-free remotely sensed variables from five different sources. This methodology to cleverly sample unlabelled and remotely sensed dynamic variables, became our *second artefact*, and can be used to sample covariates for other use cases.

We then delivered our *third artefact*, namely multiple Bayesian network formulations of rhino presence. These included networks ranging from a simple Bayesian network formulation with only four nodes and six arcs to more complicated models with up to 16 nodes. Most models obtained cross validation accuracies of around 70 %. This demonstrated how to learn the graphical underpinning of Bayesian network from quantifiable data, where previously this was completed through expert opinion only. We did not only compare different structure learning approaches, namely the hill climbing and grow-shrink algorithms, but also defined a network structure ourselves. We found, in agreement with Cheng et al. (2002), that score-based methods are predominantly preferred, especially when working with small sample sizes and noisy data. Using hill climbing, a score-based method, graph learning revealed relationships between variables otherwise missed while also providing more realistic structures than growth-shrink, a constraint-based algorithm.

In these first sets of Bayesian networks, we learned that the probability of observing a white rhino increases on warm slopes compared to flat slopes. Rhinos also tend to be closer to rivers (within 5.43 km), closer to surface water (within 3.47 kilometres) and closer to previously burned areas (within 7.97 km).

We showed how these networks can be used to make inference when new evidence is presented. This may be especially useful for park managers and rangers concerned with ecological management. For example, *"should we send a patrol on the flat slopes close to a river this month?"* or *"if we burn a certain landscape, will the probability of finding a rhino there increase?"*

The problem of discretization and to what extent it limited our models' performance is still questionable. Continuous variables can be kept as such, representing the conditional probability distributions as conditional density functions. However, when variables are continuous, structural learning becomes significantly more difficult as the number of possible dependence relations becomes infinite. Nonetheless, we illustrated that Bayesian networks of similar accuracies, albeit with different structures can be constructed using different discretisation methods. Both Hartemink's discretisation and interval discretisation provided accurate models, but Hartemink's discretisation resulted in a more interpretable network structure.

We illustrated how to learn a network with variables operating in the previous timestep and to then perform inference without presenting complete evidence to the network, specifically making inference about the next timestep.

Missing data is still a challenge in earth observation, as clouds prevent sensors to observe the surface. Much information is lost when records with missing values are discarded and the expectation-maximization (EM)-algorithm provides a practical manner to estimate model parameters while learning the structure without disregarding incomplete records. In our case, this greatly increased our sample size and allowed for the discovery of the relationship between distance to the nearest fire in the previous timestep and the distance to a previously burned area in the current timestep. This gave us confidence in the network's ability to detect relationships between variables computed from different

sensors.

Although computationally expensive, we demonstrated how to do inference in space creating heatmaps of the most probable area for white rhinos to be observed.

Individual behaviours play a significant role in population dynamics and we found very unique structures and predictability when constructing Bayesian networks for individual rhinos. This reiterates that the process of how a species uses a landscape is extremely complex. On the other hand, we experienced first hand how Bayesian networks are easy to interpret, diagnose and tweak. All models within this dissertation were constructed as to be easily packaged as a modular and scalable tool that updates automatically as new data becomes available as it is built entirely with open-source software and remotely sensed environmental data.

As we find ourselves nestled within the Information Age, cross disciplinary research is more pertinent than ever to allow the integration of expertise from unique industries. We anticipate that this dissertation will be a learning experiment for all parties involved, delving into behavioural ecology literature, mapping that to remotely sensed environmental variables and ultimately exploring Judea Pearl's novel Bayesian networks within an interesting setting. We believe this dissertation will welcome more sophisticated models of complex environmental systems operating in a spatial and temporal context investigating approaches to bridge the gap between individual behaviours and population level dynamics.

6.2 Proposed research

During the working of this dissertation, many areas were identified with the potential to increase the quality of this dissertation. Each of these areas are briefly discussed next.

6.2.1 Uncertainty in the data

This dissertation addressed the One Class Classification (OCC) problem through means of positive-unlabelled sampling. Support vector machines have been used to address the OCC problem and some researchers have suggested investigating the development of a Bayesian network capable of dealing with one-class data without the need to sample unlabelled data. It is possible that these alternate approaches in dealing with the OCC problem may produce more accurate results.

6.2.2 Space and time

Aiming to understand the distribution of rhinos, we operated not only within a spatial context, but also within a temporal context. In this dissertation, we possibly oversimplified both the spatial and temporal context.

Bayesian networks in time

Recall the *Observing vultures* example in figure 2.1 where we modelled whether or not vultures will be present given evidence on whether an animal has died in an area. We assumed an animal may die if there has either been a drought in the area or a poacher has attempted to kill it. We knew that when there has been a drought, dry grass would have been observed. This model disregards any temporal knowledge about the process. However, an animal will surely only die some time after a drought has occurred and grass will only dry when it has been deprived from water for some time.

We included some temporally-aware variables in our networks, simply through the inclusion of variables operating in the previous time step. Another approach is by defining a timestep, and construct a variant of an ordinary Bayesian network to capture temporal changes as shown in Figure 6.1. This is known as a Dynamic Bayesian Network (DBN) where each variable of interest has one copy for each time step of interest. Links exist between nodes in the same time step and between nodes in different time steps. Feedback loops between nodes within a normal Bayesian network are not allowed although many real world situations require that. Hypothetically, let's say an animal will be chased off by vultures circling in the air and are thus less likely to die. This creates a feedback loop in our causal graph, as we already established a causal relationship between *vultures* and *animal dies*. This is where DBN comes in handy as it can be used to explicitly model feedback functions. We could add another link between *vultures* in time step $t-1$ and *animal dies* in time step t.

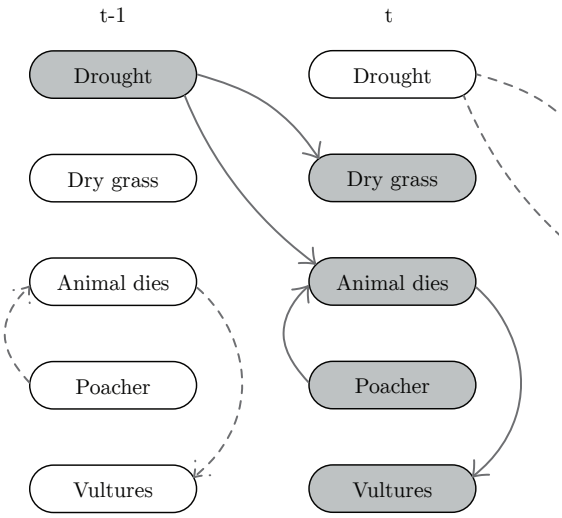


Figure 6.1: A temporal causal graph.

Bayesian networks in space

Tobler's first law of Geography states *"Everything is related to everything else, but near things are more related than distant things"*. In this dissertation, we ignored the spatial correlation between pixels and merely viewed each pixel as a separate observation uninfluenced by its neighbours. We assumed any spatial correlation would have been contained within the environmental variables itself, and we constructed Bayesian networks as normal. (Tobler, 1970). Disregarding spatial correlation may be an oversimplification and Bayesian networks capturing probabilistic relationships between spatial features should be considered. In this context every domain variable will be a node (e.g. vegetation quality) and edges will connect attributes directly dependent on each other.

6.2.3 Confounding variables

In our data, observable variables only describe some relevant aspects of the world. Consider all the underlying structures we have learned thus far. An unobserved quantity, *rhino present*, results in a rhino actually being *observed*. Knowledge of rhino presence makes the environmental variables independent of the actual observation. On the other hand, if we do not include *rhino present* all observable environmental variables seem related to each other. By introducing hidden or latent variables, simpler models can be learnt, models will be less prone to overfitting and more efficient at inference (Friedman et al., 1997). Hidden variables are only useful if they are connected to other nodes in the network.

Have we missed variables? White rhinos tend to defecate in communal dung heaps, or *middens*, as large as 30 m in diameter (average diameter at 7 ± 0.29 m) (Marneweck et al., 2018). The dominate male defecates in the centre and can have as many as 30 middens in his home range, potential challenging males on the edge of the middens and females on the periphery, advertising their oestrous state (Owen-Smith and Smith, 1973).

With the high resolution imagery available, middens can possibly be identified through object-based identification and used as an environmental variable in modelling. Instead of using a static human settlement layer, object-based identification of buildings can be completed as new high resolution imagery becomes available as this will give a more realistic depiction especially in areas struggling with human encroachment. Although some studies have accurately identified unique vegetation species from hyperspectral imagery, as each species has a unique reflectance print, it is often expensive as specialised airborne sensors. Adding to the rhino's landscape of fear, information regarding poacher presence can also be considered to be included.

6.2.4 Focussing on the individual

So far, we focused on the application of Bayesian networks to model population-level behaviour. However, as in the traffic world, roads do not decide to become congested, but instead congestion is a result of the decisions of individual commuters. The same is true for rhino distributions: a landscape does not choose to have a rhino present within it, but rather its presence is a result of the decisions of that individual rhino. Mechanisms relying on unique personalities of animals shape behaviour of individuals which ultimately results in population-level consequences (Morales and Ellner, 2002; Spiegel et al., 2017). For example, individual gender and vegetation structure influences lion hunting behaviour. Male lions prefer ambush hunting, killing in landscapes with much shorter line-of-sight (16.2 m), whereas females prefer social hunting in the open (Loarie et al., 2013). White rhino cows and bulls of different ages will behave differently and we should not disregard these individual differences in animal modelling.

Agent-based modelling provides a means to capture an individual's interactions with its environment, with other individuals and biotic components. Bayesian networks can be adapted to become an *engine* within a unique agent, which is known as a Hidden Markov Model (HMM). Underlying behavioural states can be inferred from uppermost observations such as walking speed and walking direction. Step lengths and angles can be calculated from Global Positioning System (GPS) collar data. For example, slow and variable movement may indicate foraging while faster, directed movement may indicate searching behaviour as demonstrated in Figure 6.2a.

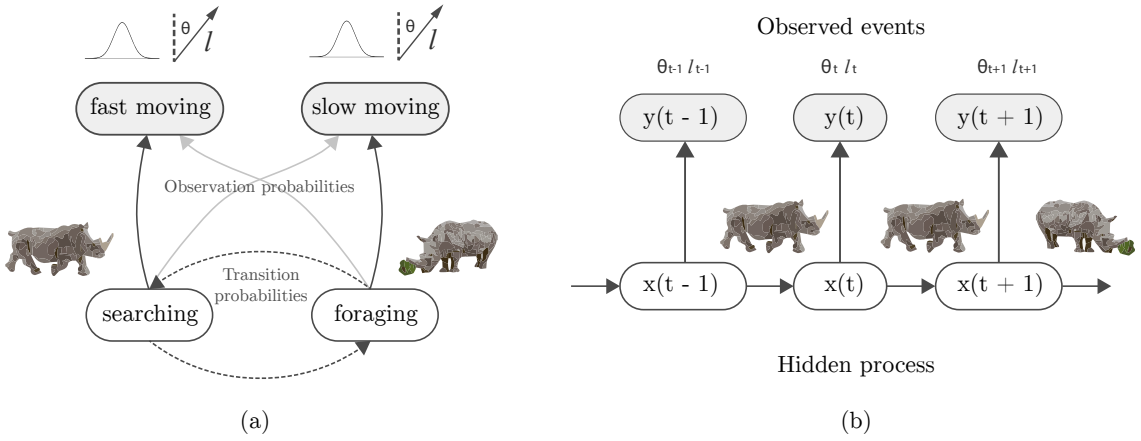


Figure 6.2: A HMM with its *observation model* and *transition model*.

This is in effect an HMM consisting of an *observation model* relating to step lengths and

turning angles and a *transition model* relating to the possible hidden states, i.e. *foraging* and *searching*. At each discrete time point t_i parameters of the step length distribution and parameters of the turning angle distribution are determined by an underlying unobserved state, as illustrated in Figure 6.2b. This allows us to realise individual animal behaviour. Population behaviour emerges not only from complex interactions between individual animals but also between animals and their environments (van der Vaart et al., 2016). For example, a rhino will be more likely to switch from *searching* to *foraging* when it reaches more productive areas. Luckily, environmental covariates can be included in an HMM, so that the transitional probabilities of shifting between states become functions of the covariates.

Appendix A

Environmental variables

Table A.1: Landform categories.

ID	Landform	Position	TPI	Slope	CHILI
11	Peak/ridge (warm)	Summit	(0.0 <mTPIs <1.0) and (30 <(E0-En) <300)		Warm
12	Peak/ridge	Summit	(0.0 <mTPIs <1.0) and (30 <(E0-En) <300)		Neutral
13	Peak/ridge (cool)	Summit	(0.0 <mTPIs <1.0) and (30 <(E0-En) <300)		Cool
14	Mountain/divide	Summit	(0.0 <mTPIs <1.0) and ((E0-En) >= 300)		
15	Cliff	Summit		>50	
21	Upper slope (warm)	Upper slope	(0.0 <mTPIs <1.0) and ((E0-En) <= 30)		Warm
22	Upper slope	Upper slope	(0.0 <mTPIs <1.0) and ((E0-En) <= 30)		Neutral
23	Upper slope (cool)	Upper slope	(0.0 <mTPIs <1.0) and ((E0-En) <= 30)		Cool
24	Upper slope (flat)	Upper slope	(0.0 <mTPIs <1.0) and ((E0-En) <= 30)	<2	
31	Lower slope (warm)	Lower slope	(-0.75 <mTPIs <1.0) and ((E0-En) <-5)		Warm
32	Lower slope	Lower slope	(-0.75 <mTPIs <1.0) and ((E0-En) <-5)		Neutral
33	Lower slope (cool)	Lower slope	(-0.75 <mTPIs <1.0) and ((E0-En) <-5)		Cool
34	Lower slope (flat)	Lower slope	(-0.75 <mTPIs <1.0) and ((E0-En) <-5)	<2	
41	Valley	Valley bottom	(mTPIs <-0.75)		
42	Valley (narrow)	Valley bottom	(mTPIs <-1.2) and ((E0-En) <= -5)		

Table A.2: Landtype categories.

ID	Label	Geology	Woody cover	Soil
1	Malelane - MA01	Banded & layered porphyritic gneiss & migmatite of Nelspruit granite suite & schist of Barberton sequence	C.appiculatum bush savanna with several rare & extraneous chromophytic species	Granitic soils
2	Orpen - SA05	Timbavati gabbro (quartz gabbro, gabbro and olivine gabbro) occur as differentiated plates in Basement complex	Isolated patches with moderately dense C. apiculatum/C. zeyheri bush savanna on granitic inliers occur.C.hereroense/C.apiculatum bush savanna on	Gabbro soils
3	Pretoriuskop - SK01	Granite/gneiss of Nelspruit granite suite	Open-dense Terminalia sericea/Dichrostachys cinereatree savanna. Pterocarpus angolensis is often a prominent tree	Granitic soils
4	Stolsnek - MA02	Porphyritic gneiss & granite (domes) of Nelspruit granite suite	Open to moderately dense C.appiculatum bush savanna with T.sericea frequenly dominant on upper footslope	Granitic soils
5	Napi - SK02	Granite/gneiss of Nelspruit granite suite	Moderately dense T. sericea bush savanna with mixed Combretum on crest	Granitic soils
6	Makhuthwanini - SK05	Granite/gneiss of Nelspruit granite suite	Moderately dense mixed Combretum spp./D. cinerea bush savanna	Granitic soils
7	Lwakahle - SK04	Granite/gneis of Nelspruit granite suite intryded by numerous dolerite dykes	Moderately dense C. apiculatum/G.bicolor bush savanna with A..exuvialis & A. nigrescens	Granitic soils
8	Lwakahle - SK04	Granite/gneis of Nelspruit granite suite intryded by numerous dolerite dykes	Moderately dense C. apiculatum/G.bicolor bush savanna with A..exuvialis & A. nigrescens	Granitic soils
9	Napi - SK02	Granite/gneiss of Nelspruit granite suite	Moderately dense T. sericea bush savanna with mixed Combretum on crest	Granitic soils
10	Makhuthwanini - SK05	Granite/gneiss of Nelspruit granite suite	Moderately dense mixed Combretum spp./D. cinerea bush savanna	Granitic soils
11	Vutome - VU01	Ecca shale/mudstone;Clarens sandstone/dolerite/basalt;Colluvium from sandstone/gneiss over shale/mustone	A.welwitschii/E.diinorum tree savanna with Spirostachys africana near drainage channels.	Ecca shale soils
12	Satara - SA01	Olivine-poor basalt of Sabie R basalt formation intruded by dolerite dykes	Acacia nigrescens/Sclerocarya birrea tree savanna.Dichrostachys cinerea prominent shrub	Basalt soils
13	Skukuza - SK07	Granite/gneis of Nelspruit granite suite intruded by numerous mafic dykes of pre- & post- Karoo age. Migmatite occurs frequentl	Moderately dense C. apiculatum/G.bicolor bush savanna with A..exuvialis & A. nigrescens	Granitic soils
14	Pretoriuskop - SK01	Granite/gneiss of Nelspruit granite suite	Open-dense Terminalia sericea/Dichrostachys cinereatree savanna. Pterocarpus angolensis is often a prominent tree	Granitic soils
15	Pretoriuskop - SK01	Granite/gneiss of Nelspruit granite suite	Open-dense Terminalia sericea/Dichrostachys cinereatree savanna. Pterocarpus angolensis is often a prominent tree	Granitic soils
16	Renosterkoppies - SK06	Granite/gneiss of Nelspruit granite suite	Moderately dense T. sericea bush savanna with C.zeheri dominant on crest	Granitic soils
17	Randspruit - SK03	Granite/gneiss of Nelspruit granite suite	Moderately dense T. sericea bush savanna with C.zeheri dominant on crest	Granitic soils
18	Orpen - SA05	Timbavati gabbro (quartz gabbro, gabbro and olivine gabbro) occur as differentiated plates in Basement complex	Isolated patches with moderately dense C. apiculatum/C. zeyheri bush savanna on granitic inliers occur.C.hereroense/C.apiculatum bush savanna on	Gabbro soils
19	Sabiepoort - SP01	Mainly rhyolite & dacite intruded by dolerite dykes. Large intrusive granophyre bodies occur on W	Dense-moderately dense C.apiculatum/P.rotundifolius bush savanna. A.nigrescens on clayey soils associated with dolerite dykes & D.africana	Rhyolite soils
20	Rietpan - SP02	Mainly rhyolite & interbedded basalt intruded by many dolerite dykes	Open Pterocarpus rotundifolius/Albizia harveyi shrub savanna	Rhyolite soils
21	Sabiepoort - SP01	Mainly rhyolite & dacite intruded by dolerite dykes. Large intrusive granophyre bodies occur on W	Dense-moderately dense C.apiculatum/P.rotundifolius bush savanna. A.nigrescens on clayey soils associated with dolerite dykes & D.africana	Rhyolite soils
22	Nhlanguleni - SK08	Orpen gneis with scarce/absent Mafic dykes	Moderately dense T. sericea bush savanna with C.zeheri dominant on crest	Granitic soils
23	Orpen - SA05	Timbavati gabbro (quartz gabbro, gabbro and olivine gabbro) occur as differentiated plates in Basement complex	Isolated patches with moderately dense C. apiculatum/C. zeyheri bush savanna on granitic inliers occur.C.hereroense/C.apiculatum bush savanna on	Gabbro soils
24	Renosterkoppies - SK06	Granite/gneiss of Nelspruit granite suite	Moderately dense T. sericea bush savanna with C.zeheri dominant on crest	Granitic soils
25	Salitje - SA06	Granite/gneiss of the Nelspruit granite suite intruded by a dense swarm of dolerite dykes	Isolated patches with moderately dense C. apiculatum/C. zeyheri bush savanna on granitic inliers occur.A.nigrescens tree savanna on midslope	Gabbro soils
26	Orpen - SA05	Timbavati gabbro (quartz gabbro, gabbro and olivine gabbro) occur as differentiated plates in Basement complex	Isolated patches with moderately dense C. apiculatum/C. zeyheri bush savanna on granitic inliers occur.C.hereroense/C.apiculatum bush savanna on	Gabbro soils
27	Shidyanamani - PH02	Gneiss & migmatite of Makhutswi gneiss intruded by dolerite dykes but gneiss is dominant over migmatite	C.apiculatum bush savanna	Gabbro soils
28	Orpen - SA05	Timbavati gabbro (quartz gabbro, gabbro and olivine gabbro) occur as differentiated plates in Basement complex	Isolated patches with moderately dense C. apiculatum/C. zeyheri bush savanna on granitic inliers occur.C.hereroense/C.apiculatum bush savanna on	Gabbro soils
29	Nhlanguleni - SK08	Orpen gneis with scarce/absent Mafic dykes	Moderately dense T. sericea bush savanna with C.zeheri dominant on crest	Granitic soils
30	Rabelais - SK10	Orpen gneis with intrusive gabbro	Moderately dense T. sericea bush savanna with mixed Combretum on crest	Granitic soils
31	Houtboschrand - PH01	Gneiss & migmatite of Makhutswi gneiss intruded by dolerite dykes	C.mopane/C.apiculatum bush to tree savanna	Granitic soils
32	Shidyanamani - PH02	Gneiss & migmatite of Makhutswi gneiss intruded by dolerite dykes but gneiss is dominant over migmatite	C.apiculatum bush savanna	Granitic soils
33	Bulweni - BU01	Mainly fine grained sandstone & shale of Karoo sequence	Chomophytes on rock outcrops with T.sericea/C.zeyheri/C.collinum bush savanna on deep sand. C.mopane/E.divinorum tree/bush savanna on shale/mudstone	Ecca shale soils

ID	Label	Geology	Woody cover	Soil
34	Shidyanamani - PH02	Gneiss & migmatite of Makhutswi gneiss intruded by dolerite dykes but gneiss is dominant over migmatite	C.apiculatum bush savanna	Granitic soils
35	Shilawuri - LE07	Timbavati gabbro (quartz gabbro, gabbro and olivine gabbro)	Dense-moderately dense C. mopane shrub savanna	Gabbro soils
36	Timbavati - SK11	Makhutswi gneis intruded by dolerite & gabbro dykes on low intensity	Moderately dense C. apiculatum/A. exuvialis bush savanna	Granitic soils
37	Muzandzeni - SK09	Orpen gneis intruded in some places by numerous dolerite dykes	Moderately dense mixed Combretum spp./A. nigrescens bush savanna	Granitic soils
38	Muzandzeni - SK09	Orpen gneis intruded in some places by numerous dolerite dykes	Moderately dense mixed Combretum spp./A. nigrescens bush savanna	Granitic soils
39	Mavumbye - SA02	Sabie river basalt formation with interlayers of olivine-rich basalt (picrite) of Letaba basalt formation. Dolerite dykes occu	Dense to open A. nigrescens bush savanna	Basalt soils
40	Muzandzeni - SK09	Orpen gneis intruded in some places by numerous dolerite dykes	Moderately dense mixed Combretum spp./A. nigrescens bush savanna	Granitic soils
41	Balule - SA04	Rhyolitic dykes occur in low densities.Due to shallowness of soils, dolerite dykes more prominently as rocky ridge	Moderately dense to dense C. apiculatum/A. nigrescens bush savanna with T.pruniodes/G.bicolor bush savanna on midslope	Basalt soils
42	Bangu - SA03	Sabie river basalt formation with interlayers of olivine-rich basalt (picrite) of Letaba basalt formation. Dolerite dykes occu	Dense A. nigrescens bush savanna	Basalt soils
43	Mavumbye - SA02	Sabie river basalt formation with interlayers of olivine-rich basalt (picrite) of Letaba basalt formation. Dolerite dykes occu	Dense to open A. nigrescens bush savanna	Basalt soils
44	Satara - SA01	Olivine-poor basalt of Sabie R basalt formation intruded by dolerite dykes	Acacia nigrescens/Sclerocarya birrea tree savanna.Dichrostachys cinerea prominent shrub	Basalt soils
45	Nwanetsi - SP04	Mainly interbedded rhyolite & basalt	Moderately dense C. apiculatum/ C. hereroense/ T. prunioides bush savanna	Rhyolite soils
46	Sabiepoort - SP01	Mainly rhyolite & dacite intruded by dolerite dykes. Large intrusive granophyre bodies occur on W	Dense-moderately dense C.apiculatum/P.rotundifolius bush savanna. A.nigrescens on clayey soils associated with dolerite dykes & D.africana	Rhyolite soils
47	Pumbe - SP03	Rhyolite capped by layer of well rounded unconsolidated quartzitic pebbles, cobblestones & sand	Moderately dense - open C.apiculatum/ C. zeyheri/ T.sericea/ Pseudolachnostylis maprouneifolia bush savanna	Rhyolite soils
48	Tsheri - PH03	Pegmatite of Makhutswi gneiss dominant.& Gneiss,migmatite,dolerite,amphibolite & schist	C.apiculatum/c.mopane/T.pruniodes bush to tree savanna.Chomophytes:Kirkia acuminata,Ficus abutilifolia & Euphorbia cooperi occur	Granitic soils
49	Phalaborwa - PH04	Gneiss & migmatite of Makhutswi gneiss with relicts of Gravelotte group prominent in W part.Dolerite dykes & syenite plugs occu	C.apiculatum/T.sericea bush savanna	Granitic soils
50	Letaba - LE02	Both olivine-poor (Sabie river formation) & olivine-rich (Letaba formation) basalt. Extensive alluvial deposits flank Letaba R	Open - moderately dense C. mopane bush savanna	Basalt soils
51	Shilawuri - LE07	Timbavati gabbro (quartz gabbro, gabbro and olivine gabbro)	Dense-moderately dense C. mopane shrub savanna	Gabbro soils
52	Malopeni - PH06	Mainly amphribolite of Gravelotte group	C.mopane/C.apiculatum bush savanna	Granitic soils
53	Shivhulani - PH05	Mainly amphibolite of Gravelotte group & migmatite of Makhutswi gneiss.Dolerite dykes occur & scattered syenite plugs occur as	C.apiculatum/C.mopane bush savanna	Granitic soils
54	Shivhulani - PH05	Mainly amphibolite of Gravelotte group & migmatite of Makhutswi gneiss.Dolerite dykes occur & scattered syenite plugs occur as	C.apiculatum/C.mopane bush savanna	Granitic soils
55	Klipkoppies - KL02	Mainly rhyolite & dacite	Open to moderately dense C.appiculatum/C.mopane bush savanna	Rhyolite soils
56	Mooiplaas - LE03	Mainly olivine-rich basalt of Letaba formation	Dense-moderately dense C. mopane shrub savanna.Dominant grasses:Bothriocloa radicans & Setaria woodii	Basalt soils
57	Mahlangeni - PH07	Mainly gneiss & migmatite of Gougplaas gneiss with scaterred occurance of amphibolite & schist of Gravelotte & Giyani groups. I	C.apicualtum/C.mopane/T.prunioides bush to tree savanna	Granitic soils
58	Malopeni - PH06	Mainly amphribolite of Gravelotte group	C.mopane/C.apiculatum bush savanna	Granitic soils
59	Shivhulani - PH05	Mainly amphibolite of Gravelotte group & migmatite of Makhutswi gneiss.Dolerite dykes occur & scattered syenite plugs occur as	C.apiculatum/C.mopane bush savanna	Granitic soils
60	Tsende - PH08	Mainly gneiss & migmatite of Gougplaas gneiss.Amphibolite as relicts of Gravelotte group in S & clu dolerite dykes occur	C.apiculatum bush savanna with C.mopane tree savanna on metalavas	Granitic soils
61	Shilawuri - LE07	Timbavati gabbro (quartz gabbro, gabbro and olivine gabbro)	Dense-moderately dense C. mopane shrub savanna	Gabbro soils
62	Mooiplaas - LE03	Mainly olivine-rich basalt of Letaba formation	Dense-moderately dense C. mopane shrub savanna.Dominant grasses:Bothriocloa radicans & Setaria woodii	Basalt soils
63	Olifants - LE01	Mainly olivine-poor basalt of Sabie R formation;intrusive rocks occur mainly dolerite and rhyolite	Moderately dense Terminalia prunioides/ Colophospermum mopane/Combretum apiculatum bush savanna	Basalt soils
64	Sabiepoort - SP01	Mainly rhyolite & dacite intruded by dolerite dykes. Large intrusive granophyre bodies occur on W	Dense-moderately dense C.apiculatum/P.rotundifolius bush savanna. A.nigrescens on clayey soils associated with dolerite dykes & D.africana	Rhyolite soils
65	Gorge - KL01	Mainly rhyolite & dacite with largely intrusive granophyre bodies on W	Open-moderately dense C.apiculatum/T.prunioides bush savanna - dense Androstachys johnsonii/Euphorbia confinalis forest	Rhyolite soils
66	Mooiplaas - LE03	Mainly olivine-rich basalt of Letaba formation	Dense-moderately dense C. mopane shrub savanna.Dominant grasses:Bothriocloa radicans & Setaria woodii	Basalt soils

ID	Label	Geology	Woody cover	Soil
67	Manyeleti - LE04	Mainly olivine-poor basalt of Sabie R formation	Dense-moderately dense C. mopane shrub savanna. Dominant grass:Themeda triandra	Basalt soils
68	Shilawuri - LE07	Timbavati gabbro (quartz gabbro, gabbro and olivine gabbro)	Dense-moderately dense C. mopane shrub savanna	Gabbro soils
69	Mahlangeni - PH07	Mainly gneiss & migmatite of Gougplaas gneiss with scaterred occurance of amphibolite & schist of Gravelotte & Giyani groups. I	C.apicualtum/C.mopane/T.prunioides bush to tree savanna	Granitic soils
70	Nalatsi - PH09	Mainly gneiss of Gougplaas gneiss & schist of Giyani group. Dolerite dykes occur	C.mopane/A.nicrescens tree savanna with isolated tracts of T.sericea/C.apiculatum bush savanna	Granitic soils
71	Bububu - PH10	Mainly amphibolite of Giyani group & migmatised gneiss of Goudplaas gneiss. Dolerite dykes occur	C.mopane/Pterocarpus rotundifolius/C.apiculatum bush savanna	Granitic soils
72	Bulweni - BU01	Mainly fine grained sandstone & shale of Karoo sequence	Chomophytes on rock outcrops with T.sericea/C.zeyheri/C.collinum bush savanna on deep sand. C.mopane/E.divinorum tree/bush savanna on shale/mudstone	Ecca shale soils
73	Bububu - PH10	Mainly amphibolite of Giyani group & migmatised gneiss of Goudplaas gneiss. Dolerite dykes occur	C.mopane/Pterocarpus rotundifolius/C.apiculatum bush savanna	Granitic soils
74	Shilawuri - LE07	Timbavati gabbro (quartz gabbro, gabbro and olivine gabbro)	Dense-moderately dense C. mopane shrub savanna	Gabbro soils
75	Shingwedzi - LE05	Mainly olivine-rich basalt of Letaba formation.Extensive alluvial deposits flank major rivers	Open-moderately dense C. mopane bush savanna with A.exuvialis	Basalt soils
76	Mooiplaas - LE03	Mainly olivine-rich basalt of Letaba formation	Dense-moderately dense C. mopane shrub savanna.Dominant grasses:Bothriocloa radicans & Setaria woodii	Basalt soils
77	Mooiplaas - LE03	Mainly olivine-rich basalt of Letaba formation	Dense-moderately dense C. mopane shrub savanna.Dominant grasses:Bothriocloa radicans & Setaria woodii	Basalt soils
78	Mooiplaas - LE03	Mainly olivine-rich basalt of Letaba formation	Dense-moderately dense C. mopane shrub savanna.Dominant grasses:Bothriocloa radicans & Setaria woodii	Basalt soils
79	Masokosa - NW02	A thin (<1.5m) sand layer,unconsolidated conglomerate consisting of pebbels & boulders,&basalt of the Letaba basalt formation	Complex associations of a variety of communities associated with different soil types	Sandy soils
80	Mooiplaas - LE03	Mainly olivine-rich basalt of Letaba formation	Dense-moderately dense C. mopane shrub savanna.Dominant grasses:Bothriocloa radicans & Setaria woodii	Basalt soils
81	Mphongolo - PH11	Gneiss & migmatite of Goudplaats.Dolerite dykes occur in dense swarms	C.mopane/C.apiculatum/P.rotundifolius tree to bush savanna	Granitic soils
82	Marithenga - BU02	Mainly shale & maudstone of Karoo sequence	C.mopane/E.divinorum tree savanna with S.aficana prominent in localized areas	Ecca shale soils
83	Tsotsi - BU03	Mainly sandstone & shale of Soutpansberg group	T.sericea/C.collinum/C.zeyheri bush savanna with Burkea africana/Pseudolachnostylis maprouneifolia tree savanna on very deep soils.C.mopane/E.divinorum tre	Ecca shale soils
84	Dothole - PH12	Mainly andesite & basalt of Sibasa formation	C.collinum/P.rotundifolius/C.apiculatum bush savanna	Sandy soils
85	Punda - PA01	Reddish quartzite, sandstone & conglomerate with intercalated andesite of Soutpansberg group	Moderately dense C. apiculatum/ K.acuminata tree savanna. Dense Androstachys johnsonii forests occur on very steep slopes characterized by rock outcrops &	Ecca shale soils
86	Manyeleti - LE04	Mainly olivine-poor basalt of Sabie R formation	Dense-moderately dense C. mopane shrub savanna. Dominant grass:Themeda triandra	Basalt soils
87	Lanner Gorge - PA03	Fine grained sandstone of Clarens formation (Karoo sequence)	Open-dense C. mopane tree savanna along mid- & footslopes where talus accumulated. Dense A. johnsonii forest along steep & rocky slopes	Sandy soils
88	Madzaringwe - PA02	Shale, mudstone, grid, conglomerate & sandstone of Ecca group (Karoo sequence)	Open-moderately dense C. mopane/ T.prunioides bush-tree savanna	Ecca shale soils
89	Lanner Gorge - PA03	Fine grained sandstone of Clarens formation (Karoo sequence)	Open-dense C. mopane tree savanna along mid- & footslopes where talus accumulated. Dense A. johnsonii forest along steep & rocky slopes	Sandy soils
90	Punda - PA01	Reddish quartzite, sandstone & conglomerate with intercalated andesite of Soutpansberg group	Moderately dense C. apiculatum/ K.acuminata tree savanna. Dense Androstachys johnsonii forests occur on very steep slopes characterized by rock outcrops &	Sandy soils
91	Baobab Hill - PA04	Nephelinite lavas of Mashikiri formation & olivine-rich lavas of Letaba basalt formation	Moderately dense-dense C. mopane/Commiphora glandulosa tree/bush savanna. Adansonia digitata conspicuous tree. T. prunioides & Euphorbia confinalis occasi	Basalt soils
92	Mashikiri - LE06	Nepheline-rich (Mashikiri formation) & olivine-rich (Letaba formation) basalt.Scattered sandstone koppies occur along W boundar	Open-moderately dense C. mopane/T. prunioides bush - shrub savanna with Adansonia digitatais	Basalt soils
93	Nwambiya - NW01	A thick layer(up to 6m) of unconsolidated sand of quaternary age	Baphia massaiensis/Guibourtia conjugata bush thickets with Xeroderris stuhlmanni-i/C.apiculatum/C.zeyheri bush savanna on midslope	Sandy soils
94	Malonga - PA06	Conglomerate & coarse-grained & gritty sandstone with limestone & marl intercalations of Malvernia formation, & picrite basalt	Open C. mopane/ Euclea schimperi/ T. prunioides shrub savanna with dense A. johnsonii forest dominant in certain localities on midslope	Basalt soils
95	Nwambiya - NW01	A thick layer(up to 6m) of unconsolidated sand of quaternary age	Baphia massaiensis/Guibourtia conjugata bush thickets with Xeroderris stuhlmanni-i/C.apiculatum/C.zeyheri bush savanna on midslope	Sandy soils
96	Malonga - PA06	Conglomerate & coarse-grained & gritty sandstone with limestone & marl intercalations of Malvernia formation, & picrite basalt	Open C. mopane/ Euclea schimperi/ T. prunioides shrub savanna with dense A. johnsonii forest dominant in certain localities on midslope	Basalt soils
97	Pafuri - PA05	Alluvial sediments	Dense -tall A. albida/F. sycomorus/X. zambesiaca riverine forest. A. xanthophloea around pans. Dense to open A. tortilis bush savanna in Luvuvhu floodplain	Alluvial soils
98	Baobab Hill - PA04	Nephelinite lavas of Mashikiri formation & olivine-rich lavas of Letaba basalt formation	Moderately dense-dense C. mopane/Commiphora glandulosa tree/bush savanna. Adansonia digitata conspicuous tree. T. prunioides & Euphorbia confinalis occasi	Sandy soils
99	Lanner Gorge - PA03	Fine grained sandstone of Clarens formation (Karoo sequence)	Open-dense C. mopane tree savanna along mid- & footslopes where talus accumulated. Dense A. johnsonii forest along steep & rocky slopes	Sandy soils
100	Orpen - SA05	Timbavati gabbro (quartz gabbro, gabbro and olivine gabbro) occur as differentiated plates in Basement complex	Isolated patches with moderately dense C. apiculatum/C. zeyheri bush savanna on granitic inliers occur.C.hereroense/C.apiculatum bush savanna on	Granitic soils

Appendix B

Calculating environmental indices

```
import ee

# ndvi
def NDVI(image, t):
    try:
        ndvi = image.normalizedDifference(['B5', 'B4'])
        bandname = 'ndvi_' + t
        ndvi = ndvi.select(['nd'], [bandname])
        return ndvi
    except:
        pass

# moisture index
def NDMI(image, t):
    ndmi = image.normalizedDifference(['B5', 'B6'])
    bandname = 'ndmi_' + t
    ndmi = ndmi.select(['nd'], [bandname])
    return ndmi

# burnt area (scar)
def nbrt(image):
    return image.expression(
        '(NIR - 0.0001 * SWIR * Temp) / (NIR + 0.0001 * SWIR * Temp)',{
            'NIR': image.select('B5'),
            'SWIR': image.select('B7'),
            'Temp': image.select('B11')
        }
    )

# Distance to scar
def distToScar(image, t, fc):
    try:
        burn = nbrt(image)
        scar = burn.lte(0.93).gte(0.90)
        #noscar = burn.gte(0.93);
```

```
        scar_mask = scar.updateMask(scar)

        scar_mask = scar_mask.clipToCollection(fc)

        # Convert scar layer from raster to vector
        scar_vector = scar_mask.reduceToVectors(
            geometry=fc,
            scale=60,
            geometryType='polygon',
            eightConnected=False,
            labelProperty='scar',
        )

        # Calculate distance
        distToScar = scar_vector.distance(500000, 100).clipToCollection(fc)

        bandname = 'distToScar_' + t

        distToScar = distToScar.select(['distance'], [bandname])
        return distToScar
    except:
        pass

def MNDWI(image):
    return image.normalizedDifference(['B3', 'B5'])

def distToWater(image, t, fc, waterpoints):
    try:
        waterindex = MNDWI(image)
        water = waterindex.gt(0)
        water_mask = water.updateMask(water)
        water_mask = water_mask.clipToCollection(fc)

        # Convert water layer from raster to vector
        water_vector = water_mask.reduceToVectors(
            geometry=fc,
            scale=60,
            geometryType='polygon',
            eightConnected=False,
            labelProperty='water',
        )

        # Join with waterpoints
        surfacewater = ee.FeatureCollection([water_vector, waterpoints]).flatte

        # Calculate distance
        distToWater = surfacewater.distance(500000, 100).clipToCollection(fc)

        bandname = 'distToWater_' + t
```

```
        # Rename band
        distToWater = distToWater.select(['distance'], [bandname])

        return distToWater

    except:
        pass

# Fires
def distToFire(fire_img, t, fc):
    try:
        # Convert fire layer from raster to vector
        fire_vector = fire_img.reduceToVectors(
            geometry=fc,
            scale=60,
            geometryType='polygon',
            eightConnected=False,
            labelProperty='fire',
        )

        distToFire = fire_vector.distance(500000, 100).clipToCollection(fc)

        bandname = 'distToFire_' + t

        distToFire = distToFire.select(['distance'], [bandname])

        return distToFire

    except:
        pass

# Soil moisture
def smmrs(image, t):
    smmrs = image.expression(
        '(1-(1/sqrt(1.9719+1))*(NIR+1.40426*RED))',{
            'NIR': image.select('B5'),
            'RED': image.select('B4')
        }
    )
    bandname = 'soilMoisture_' + t
    soilMoisture = smmrs.select(['constant'], [bandname])
    return soilMoisture

# Function to mask clouds using the quality band of Landsat 8
def maskL8(image):
    qa = image.select('BQA')
    # Check that the cloud bit is off
    mask = qa.bitwiseAnd(1 << 4).eq(0)
```

```
return image.updateMask(mask)
```

Bibliography

Aguilera, P., Fernández, A., Fernández, R., Rumí, R., and Salmerón, A. (2011). Bayesian networks in environmental modelling. *Environmental Modelling & Software*, 26(12):1376–1388.

Archibald, S. and Scholes, R. (2007). Leaf green-up in a semi-arid african savanna-separating tree and grass responses to environmental cues. *Journal of Vegetation Science*, 18(4):583–594.

Bell, D. E. (1988). *Decision making: Descriptive, normative, and prescriptive interactions*. cambridge university Press.

Bigalke, R. (1963). Die uitroeiing van breelip- of witrenoster (ceratotherium simum simum (burch.)) in transvaal en sy hervestiging: 'n historiese en kritiese oorsig. *Fauna and Flora*, 14(5).

Boole, G. (1854). The laws of thought (1854).

Bowler, D. E. and Benton, T. G. (2005). Causes and consequences of animal dispersal strategies: relating individual behaviour to spatial dynamics. *Biological Reviews*, 80(2):205–225.

Box, G. E. (1979). All models are wrong, but some are useful. *Robustness in Statistics*, 202:549.

Bromley, J., Jackson, N. A., Clymer, O., Giacomello, A. M., and Jensen, F. V. (2005). The use of Hugin ® to develop Bayesian networks as an aid to integrated water resource planning. *Environmental Modelling & Software*, 20(2):231–242.

Cahoon Jr, D. R., Stocks, B. J., Levine, J. S., Cofer III, W. R., and O'Neill, K. P. (1992). Seasonal distribution of african savanna fires. *Nature*, 359(6398):812.

Catlett, J. (1991). Megainduction: machine learning on very large databases. *PhD thesis, Basser Department of Computer Science, University of Sydney*.

Chander, G., Markham, B. L., and Helder, D. L. (2009). Summary of current radiometric calibration coefficients for landsat mss, tm, etm+, and eo-1 ali sensors. *Remote sensing of environment*, 113(5):893–903.

Charniak, E. (1991). Bayesian networks without tears. *AI magazine*, 12(4):50.

Charnov, E. L. (1976). Optimal foraging: attack strategy of a mantid. *The American Naturalist*, 110(971):141–151.

Cheng, J., Greiner, R., Kelly, J., Bell, D., and Liu, W. (2002). Learning Bayesian networks from data: an information-theory based approach. *Artificial intelligence*, 137(1-2):43–90.

Chickering, D. M. (2002). Learning equivalence classes of Bayesian-network structures. *Journal of machine learning research*, 2(Feb):445–498.

Cohen, L. E. and Felson, M. (2016). Social change and crime rate trends: A routine activity approach (1979). In *Classics in Environmental Criminology*, pages 203–232. CRC Press.

Cooper, G. F. (1990). The computational complexity of probabilistic inference using Bayesian belief networks. *Artificial intelligence*, 42(2-3):393–405.

Dagum, P. and Luby, M. (1993). Approximating probabilistic inference in Bayesian belief networks is NP-hard. *Artificial intelligence*, 60(1):141–153.

Delegido, J., Verrelst, J., Alonso, L., and Moreno, J. (2011). Evaluation of Sentinel-2 red-edge bands for empirical estimation of green LAI and chlorophyll content. *Sensors*, 11(7):7063–7081.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, 39(1):1–22.

Do, F. C., Goudiaby, V. A., Gimenez, O., Diagne, A. L., Diouf, M., Rocheteau, A., and Akpo, L. E. (2005). Environmental influence on canopy phenology in the dry tropics. *Forest Ecology and Management*, 215(1-3):319–328.

Elith, J. and Leathwick, J. (2007). Predicting species distributions from museum and herbarium records using multiresponse models fitted with multivariate adaptive regression splines. *Diversity and distributions*, 13(3):265–275.

Estes, R. (1991). *The behavior guide to African mammals*, volume 64. University of California Press Berkeley.

Feigenbaum, E. A. and McCorduck, P. (1984). *The fifth generation*. Pan Books London.

Fernández-Manso, A., Fernández-Manso, O., and Quintano, C. (2016). Sentinel-2a red-edge spectral indices suitability for discriminating burn severity. *International journal of applied earth observation and geoinformation*, 50:170–175.

Ferreira, S. M., Greaver, C., Knight, G. A., Knight, M. H., Smit, I. P., and Pienaar, D. (2015). Disruption of rhino demography by poachers may lead to population declines in Kruger National Park, South Africa. *PLoS One*, 10(6):e0127783.

Ferreira, S. M., Greaver, C., Nhleko, Z., and Simms, C. (2018). Realization of poaching effects on rhinoceroses in Kruger National Park, South Africa. *African Journal of Wildlife Research*, 48(1).

Friedman, N. et al. (1997). Learning belief networks in the presence of missing values and hidden variables. In *ICML*, volume 97, pages 125–133.

Geisser, S. (2017). *Predictive inference*. Routledge.

Gertenbach, W. D. (1983). Landscapes of the Kruger National Park. *Koedoe*, 26(1):9–121.

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., and Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*.

Grant, C., Davidson, T., Funston, P., and Pienaar, D. (2002). Challenges faced in the conservation of rare antelope: a case study on the northern basalt plains of the Kruger National Park. *Koedoe*, 45(2):45–66.

Guisan, A. and Thuiller, W. (2005). Predicting species distribution: offering more than simple habitat models. *Ecology letters*, 8(9):993–1009.

Haas, T. C. and Ferreira, S. M. (2016). Combating rhino horn trafficking: The need to disrupt criminal networks. *PLOS ONE*, 11:e0167040.

Haas, T. C. and Ferreira, S. M. (2017). Optimal patrol routes: interdicting and pursuing rhino poachers. *Police Practice and Research*, pages 1–22.

Han-Qiu, X. (2005). A study on information extraction of water body with the Modified Normalized Difference Water Index (MNDWI). *Journal of Remote Sensing*, 5:589–595.

Hartemink, A. J. (2001). *Principled computational methods for the validation discovery of genetic regulatory networks*. PhD thesis, Massachusetts Institute of Technology.

Hastie, T., Tibshirani, R., and Friedman, J. (2009). The elements of statistical learning: data mining, inference, and prediction.

Herbig, F. J. and Warchol, G. (2011). South african conservation crime and routine activities theory: a causal nexus? *Acta Criminologica: Southern African Journal of Criminology*, 24(2):1–16.

Hijmans, R. J. (2019). *`Raster`: Geographic data analysis and modeling*. R package version 2.8-19.

Hirzel, A. H., Helfer, V., and Metral, F. (2001). Assessing habitat-suitability models with a virtual species. *Ecological modelling*, 145(2-3):111–121.

Højsgaard, S. (2012). Graphical independence networks with the `gRain` package for R. *Journal of Statistical Software*, 46(10):1–26.

Holden, Z., Smith, A., Morgan, P., Rollins, M., and Gessler, P. (2005). Evaluation of novel thermally enhanced spectral indices for mapping fire perimeters and comparisons with fire atlas data. *International Journal of Remote Sensing*, 26(21):4801–4808.

Holyoak, M., Casagrandi, R., Nathan, R., Revilla, E., and Spiegel, O. (2008). Trends and missing parts in the study of movement ecology. *Proceedings of the National Academy of Sciences*, 105(49):19060–19065.

Hubschle, A. (2018). Ending wildlife trafficking. http://globalinitiative.net/ending-wildlife-trafficking.

Immitzer, M., Vuolo, F., and Atzberger, C. (2016). First experience with Sentinel-2 data for crop and tree species classifications in central europe. *Remote Sensing*, 8(3):166.

iNaturalist.org (2019). iNaturalist Research-grade Observations, Occurrence dataset https://doi.org/10.15468/ab3s5x accessed via gbif.org. www.inaturalist.org.

Jonsen, I. D., Flemming, J. M., and Myers, R. A. (2005). Robust state–space modeling of animal movement data. *Ecology*, 86(11):2874–2880.

Joubert, S. (2015). The history on the development of the Sabie and Shingwedzi Reserves and the Kruger National Park, 1898 to 1946. *Available at: www. san-parks. org, DoA*, 19.

Karasiak, N., Sheeren, D., Fauvel, M., Willm, J., Dejoux, J.-F., and Monteil, C. (2017). Mapping tree species of forests in southwest france using Sentinel-2 image time series. In *2017 9th International Workshop on the Analysis of Multitemporal Remote Sensing Images (MultiTemp)*, pages 1–4. IEEE.

Khan, S. S. and Madden, M. G. (2009). A survey of recent trends in one class classification. In *Irish conference on artificial intelligence and cognitive science*, pages 188–197. Springer.

King, M. D., Platnick, S., Menzel, W. P., Ackerman, S. A., and Hubanks, P. A. (2013). Spatial and temporal distribution of clouds observed by modis onboard the terra and aqua satellites. *IEEE Transactions on Geoscience and remote sensing*, 51(7):3826–3852.

Koen, H., de Villiers, J. P., Roodt, H., and de Waal, A. (2017). An expert-driven causal model of the rhino poaching problem. *Ecological Modelling*, 347:29–39.

Korb, K. B. and Nicholson, A. E. (2010). *Bayesian artificial intelligence*. CRC press.

Loarie, S. R., Tambling, C. J., and Asner, G. P. (2013). Lion hunting behaviour and vegetation structure in an african savanna. *Animal Behaviour*, 85(5):899–906.

Manson, N. (2006). Is operations research really research? *Orion*, 22(2):155–180.

Marneweck, C., Jürgens, A., and Shrader, A. M. (2018). The role of middens in white rhino olfactory communication. *Animal Behaviour*, 140:7 – 18.

Morales, J. M. and Ellner, S. P. (2002). Scaling up animal movements in heterogeneous landscapes: the importance of behavior. *Ecology*, 83(8):2240–2247.

Morales, J. M., Moorcroft, P. R., Matthiopoulos, J., Frair, J. L., Kie, J. G., Powell, R. A., Merrill, E. H., and Haydon, D. T. (2010). Building the bridge between animal movement and population dynamics. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 365(1550):2289–2301.

Nagarajan, R., Scutari, M., and Lèbre, S. (2013). Bayesian networks in R. *Springer*, 122:125–127.

Owen-Smith, N. (1971). Territoriality in the white rhinoceros (ceratotherium simum) burchell. *Nature*, 231(5301):294–6.

Owen-Smith, R. N. and Smith, R. N. O. (1973). *The behavioural ecology of the white rhinoceros*. PhD thesis, University of Wisconsin Madison.

Park, N., Serra, E., Snitch, T., and Subrahmanian, V. S. (2015). APE: A data-driven, behavioral model-based anti-poaching engine. *IEEE Transactions on Computational Social Systems*, 2(2):15–37.

Parker, A. and Witkowski, E. (1999). Long-term impacts of abundant perennial water provision for game on herbaceous vegetation in a semi-arid african savanna woodland. *Journal of Arid Environments*, 41(3):309–321.

Pearl, J. (2014). *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Elsevier.

Pearl, J. and Mackenzie, D. (2018). *The book of why: The new science of cause and effect*. Allen Lane, an imprint of Penguin Books.

Phillips, S. J., Dudík, M., Elith, J., Graham, C. H., Lehmann, A., Leathwick, J., and Ferrier, S. (2009). Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. *Ecological applications*, 19(1):181–197.

Pienaar, D. (1994). Habitat preferences of the white rhino in the Kruger National Park. In *Proceedings of Symposium on Rhinos as Game Ranch Animals. Onderstepoort*.

Pienaar, U. d. V. (1970). The recolonisation history of the square-lipped (white) rhinoceros Ceratotherium simum simum (Burchell) in the Kruger National Park (October 1961-november 1969). *Koedoe*, 13(1):157–170.

R Core Team (2019). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Redfern, J. V., Grant, R., Biggs, H., and Getz, W. M. (2003). Surface-water constraints on herbivore foraging in the Kruger National Park, South Africa. *Ecology*, 84(8):2092–2107.

Rossum, G. (1995). `Python` reference manual.

Scholes, R. and Archer, S. (1997). Tree-grass interactions in savannas. *Annual review of Ecology and Systematics*, 28(1):517–544.

Scutari, M. (2010). Learning Bayesian Networks with the `bnlearn` R package. *Journal of Statistical Software*, 35(3):1–22.

Shannon, G., Druce, D. J., Page, B. R., Eckhardt, H. C., Grant, R., and Slotow, R. (2008). The utilization of large savanna trees by elephant in southern Kruger National Park. *Journal of Tropical Ecology*, 24(3):281–289.

Smit, I. P. and Archibald, S. (2019). Herbivore culling influences spatio-temporal patterns of fire in a semiarid savanna. *Journal of Applied Ecology*, 56(3):711–721.

Smit, I. P., Grant, C. C., and Devereux, B. J. (2007). Do artificial waterholes influence the way herbivores use the landscape? Herbivore distribution patterns around rivers and artificial surface water sources in a large African savanna park. *Biological Conservation*, 136(1):85–99.

Spiegel, O., Leu, S. T., Bull, C. M., and Sih, A. (2017). What's your move? Movement as a link between personality and spatial dynamics in animal populations. *Ecology letters*, 20(1):3–18.

Stockwell, D. (1999). The garp modelling system: problems and solutions to automated spatial prediction. *International journal of geographical information science*, 13(2):143–158.

Stokland, J. N., Halvorsen, R., and Støa, B. (2011). Species distribution modelling—effect of design and sample size of pseudo-absence observations. *Ecological Modelling*, 222(11):1800–1809.

Su, C., Andrew, A., Karagas, M. R., and Borsuk, M. E. (2013). Using Bayesian networks to discover relations between genes, environment, and disease. *BioData mining*, 6(1):6.

Sun, S., Zhang, C., and Yu, G. (2006). A Bayesian network approach to traffic flow forecasting. *IEEE Transactions on intelligent transportation systems*, 7(1):124–132.

Theobald, D. M., Harrison-Atlas, D., Monahan, W. B., and Albano, C. M. (2015). Ecologically-relevant maps of landforms and physiographic diversity for climate adaptation planning. *PloS one*, 10(12):e0143619.

Tobler, W. R. (1970). A computer movie simulating urban growth in the Detroit region. *Economic geography*, 46(sup1):234–240.

Tucker, C. J. (1979). Red and photographic infrared linear combinations for monitoring vegetation. *Remote sensing of Environment*, 8(2):127–150.

Turing, A. M. (2009). Computing machinery and intelligence. In *Parsing the Turing Test*, pages 23–65. Springer.

Uusitalo, L., Kuikka, S., and Romakkaniemi, A. (2005). Estimation of Atlantic salmon smolt carrying capacity of rivers using expert knowledge. *ICES Journal of Marine Science*, 62(4):708–722.

van der Vaart, E., Johnston, A. S., and Sibly, R. M. (2016). Predicting how many animals will be where: How to build, calibrate and evaluate individual-based models. *Ecological modelling*, 326:113–123.

Venter, F. J., Scholes, R. J., and Eckhardt, H. C. (2003). The abiotic template and its associated vegetation pattern. *The Kruger experience: Ecology and management of savanna heterogeneity*, 83:129.

Wells, C. (2019). Elon Musk lays out vision for Starlight satellite project. https://imherald.com/2019/05/19/elon-musk-lays-out-vision-for-starlight-satellite-project.

Zadeh, L. A. (1988). Fuzzy logic. *Computer*, 21(4):83–93.

Zaniewski, A. E., Lehmann, A., and Overton, J. M. (2002). Predicting species spatial distributions using presence-only data: a case study of native New Zealand ferns. *Ecological modelling*, 157(2-3):261–280.

Zhan, Z., Qin, Q., Ghulan, A., and Wang, D. (2007). NIR-red spectral space based new method for soil moisture monitoring. *Science in China Series D: Earth Sciences*, 50(2):283–289.