Ecological Modelling 313 (2015) 127–136

Contents lists available at ScienceDirect

Ecological Modelling

journal homepage: www.elsevier.com/locate/ecolmodel

Using Gaussian Bayesian Networks to disentangle direct and indirect associations between landscape physiography, environmental variables and species distribution

Eric Meineri*, C. Johan Dahlberg, Kristoffer Hylander

Department of Ecology, Environment and Plant Sciences, Stockholm University, Sweden

ARTICLE INFO

Article history:
Received 3 March 2015
Received in revised form 16 June 2015
Accepted 17 June 2015

Keywords:
Species distribution model
Topoclimate
Bayesian Network
Path analyses
Topography
Climate

ABSTRACT

Landscape physiography affects temperature, soil moisture and solar radiation. In turn, these variables are thought to determine how species are distributed across landscapes. Systems involving direct and indirect associations between variables can be described using path models. However, studies applying these to species distribution modelling are rare.

Bayesian Networks are path models designed to represent associations across observed variables. Here, we demonstrate the use of Bayesian Networks to disentangle the direct and indirect associations between landscape physiography, soil moisture, solar radiation, temperature and the distribution patterns of four plants at their northern range limit in Sweden.

Fine scale variations in maximum temperatures were associated with variations in elevation, distance to coast and solar radiation. In contrast, fine scale variations in minimum temperature were associated with distance to coast, cold air drainage and soil moisture. These associations between landscape physiography and minimum and maximum temperature were predicted, furthermore, to be associated with growing season length, growing degree day and ultimately species distributions. All species were indirectly associated with aspect through their responses to either solar radiation or temperature.

The models demonstrated strong indirect associations between landscape physiography and species distributions. The models suggested that local variation in light can be as important as temperature for species distributions. Disentangling the direct and indirect associations between landscape physiography, environmental variables and species distribution can provide new and important insights into how landscape components are linked to species distributions.

1. Introduction

Several studies have highlighted the importance of topographic heterogeneity and small scale climatic gradients on species' resilience to climate changes (e.g. Keppel et al., 2012; Lenoir et al., 2013; Loarie et al., 2009). It has been argued that species may only need to migrate over short distances to keep pace with climate change in landscapes with large topographic heterogeneity because of considerable microclimatic variations over such distances (Loarie et al., 2009). Similarly, it has been proposed that several taxa may have persisted in climatic refugees outside their climatic range during periods with unsuitable regional climates (Hampe et al., 2013; Pearson, 2006) and that these climatic refugees occurred in places with specific topographic combinations (Dobrowski, 2011). Therefore, a clear understanding of how landscape physiography (here defined as the range of topographic variation encountered within a focal landscape) affects fine scale species distributions through variations in microclimatic conditions is necessary to predict climate change impacts on range dynamics better.

Landscape physiography has been described as indirectly regulating landscape scale species distributions through its effects on temperature, soil moisture and solar radiation (Austin, 2007; Austin and Van Niel, 2011). However, the indirect associations between landscape physiography and species distribution are numerous and can be complex since temperature, soil moisture and solar radiation are interconnected. For instance, maximum temperature is generally driven by elevation and solar radiation (Dobrowski, 2011;

* Corresponding author at: Department of Ecology, Environment and Plant Sciences, Stockholm University, Lilla Frescati, SE – 106 91, Stockholm, Sweden. Tel.: +46 739162288.
E-mail address: meinerieric@gmail.com (E. Meineri).

http://dx.doi.org/10.1016/j.ecolmodel.2015.06.028
0304-3800/

Lookingbill and Urban, 2004; Vanwalleghem and Meentemeyer, 2009), of which the latter is a function of physiographic variables such as aspect, slope and elevation (Fu and Rich, 2002). Similarly, variations in minimum temperatures have been linked to variations in soil moisture and to variables related to cool air pooling, for example relative elevation, topographic convergence and vicinity to water bodies (Dobrowski, 2011; Lookingbill and Urban, 2004; Vanwalleghem and Meentemeyer, 2009). If we are to understand the links between landscape physiography and species distribution, we need to disentangle the network of direct and indirect associations between physiographic variables, environmental variables and species distribution. This will help us to move from simple correlations to a more mechanistic understanding of species distribution across landscape.

Path models are well suited for studying direct and indirect associations between environmental variables within natural systems. Of the models available, path analyses or Structural Equation Models (SEM) are often applied. These models are designed to fit a user-specified path diagram and test whether the data are "faithful" to (fit) the path diagram (Grace, 2006). Consequently, these models are better suited for confirmatory than for exploratory analyses. Having strict assumptions about the existing associations between variables is not always straightforward. In contrast to SEM, Bayesian Networks are path models that can be derived from the data using learning algorithms. Thus such models can be used as an exploratory method (Alameddine et al., 2011; Chen and Pollino, 2012; Højsgaard et al., 2012; Nagarajan et al., 2013). These models have been shown to retrieve meaningful direct and indirect associations in highly dimensional datasets and can be fitted from rather small datasets (Aguilera et al., 2011; Margaritis, 2003; Meyer et al., 2014; Nagarajan et al., 2013). Because of their learning ability, we believe that Bayesian Networks are well suited to identify the complex associations between physiographic variables, environmental variables and species distributions. Most applications of Bayesian Network in environmental studies rely on categorical or discretized continuous variables, giving the impression that the use of pure continuous variables is not possible within the framework of Bayesian Networks and suggesting that they require a level of discretization (Aguilera et al., 2011, 2010; Alameddine et al., 2011; Chen and Pollino, 2012; Hamilton et al., 2015; Meyer et al., 2014; Uusitalo, 2007). However Gaussian versions of Bayesian Networks together with their specific learning algorithms do exist and are well described in the literature (e.g. Højsgaard et al., 2012; Kelly et al., 2013; Margaritis, 2003; Nagarajan et al., 2013).

In this study, we demonstrate the use of Gaussian Bayesian Networks to disentangle the direct and indirect associations between landscape physiography, environmental variables known to affect plants directly, and species distribution patterns of four plants at their northern range limit. We specifically applied the Gaussian Bayesian Network approach to (i) assess direct associations between species occurrences and temperature variables, soil moisture and solar radiation, (ii) quantify the relative strengths of these associations and (iii) assess how landscape physiography indirectly links to species distributions within the focal landscape.

2. Methods

2.1. Study area

The study area is located in the coastal region of the "High Coast" (Höga Kusten) in central Sweden and covers approximately 7000 km^2 from the Baltic Sea to 70 km inland (latitude 62.5–63.3°, longitude 16.8–18.9°, decimal degrees, Fig. 1). This landscape is highly diverse in terms of slope and aspect. The variation in elevation encountered within the landscape is relatively low, ranging from 0 to 470 m. The study area is located in the "middle boreal subzone" (Sjörs, 1999), a transition zone in which many southern species are at their upper range limit. The study area is characterized by a mosaic of managed coniferous forests of different ages. In this paper, we only consider mature forests.

2.2. Environmental variables

2.2.1. Temperature variables

In spring 2011, Maxim 1922L iButton temperature loggers (Hubbart et al., 2005) were set up at 57 forested sites in the study area. Sites were selected to represent the different slope, elevation and aspect combinations occurring within mature forests in the study area. At each site, near ground temperature was recorded every hour for a period of one year. Temperature was recorded at height of 1 m as a compromise between the meteorological and ecological interests within the research project (see Vercauteren et al.,

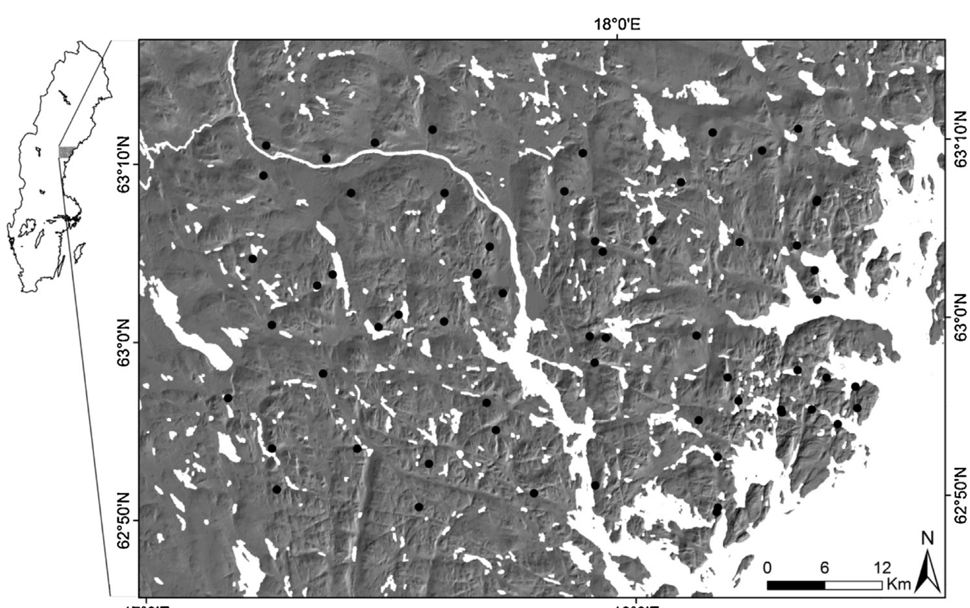

Fig. 1. Hillshade map of the study area within Sweden. Black circles represent 57 forested study sites with on-site temperature measurements.