COMPARING TWO SPECIES DISTIBUTION MODELS USING SATELLITE ONLY AND READY-MADE ENVIRONMENTAL VARIABLES FOR THE DAKOTA SKIPPER (*Hesperia dacotae*), INTERLAKE REGION OF MANITOBA, CANADA

by

Colin Murray

The University of Winnipeg

Department of Biology

Winnipeg, Manitoba, Canada

A thesis submitted to the Faculty of Graduate Studies

in partial fulfillment of the requirements for a

Master of Science in Bioscience, Technology, and Public Policy.

April 2024

ABSTRACT

The Dakota skipper, *Hesperia dacotae* (Skinner 1911) [Hesperiidae, Lepidopterida] is a rare prairie obligate butterfly with an affinity for anthropogenically undisturbed, grassland habitat with diverse native flora. Persistent threats include habitat fragmentation, destruction, and degradation. These and other threats have caused precipitous population declines and local extirpation across its range. Consequentially, the Dakota skipper is currently listed as Endangered in Canada and Threatened in the United States, and the province of Manitoba. Species distribution models (SDM) are a well-known technique which attempt to predict a species distribution on a landscape. These predictions can then be used to inform conservation actions such as guiding survey effort, land acquisitions, and reintroductions. The objectives of this project were to: 1) Compare Dakota skipper models using freely available high resolution remotely sensed products to those using more traditional environmental predictors. 2) Field validate both models to identify the most accurate model using efficient and economical methods. 3) Address issues of modelling rare species to produce a robust SDM for the Dakota skipper in Manitoba. I found that SDMs built from environmental variables generated from satellite imagery performed comparably to one produced from readily available geospatial information. I also found that field validation was more accurate for evaluating SDMs than purely statistical methods. I also produced usable SDMs for the Dakota skipper in the Interlake. Implications from this study are that the advantages of satellite imagery can be leveraged to create useable SDMs to guide conservation actions. This study also further supports the need to field validate an SDM over relying on model statistical output which can be misleading.

ACKNOWLEDGEMENTS

Thank you to my committee! Dr. Richard Westwood, Dr. Katherine Dearborn, Dr. Joni Stori and to Bill Watkins, Manitoba Wildlife Branch (retired), and Dr. Andy Park (retired) for their encouragement, support, guidance, and valuable feedback. Thank you to Dr. Ryan Danby, Department of Geography and Planning, Queen's University for generously agreeing to be my external examiner.

Thank you to Chris Friesen, Manitoba Conservation Data Centre, for his understanding and support during this project.

Special thanks to Karin Newman for her encouragement and taking care of the kids when it wasn't her turn. And for convincing me to start this endeavour- not to pass up a great opportunity.

All my love to Heather for their unsurpassed love, patience, understanding, and support through all of this. And for gently telling me to "shut the heck up and get to work" when I was wandering round the house in my robe and toque all stressed out about a new set of revisions that just came in.

Special thank you and my love to Sacha and Mae for your patience and teasing while dealing with a distracted Dad. Now that this is done, I can really get involved in your lives again. Like really involved.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

CHAPTER 1: INTRODUCTION

The Dakota skipper, *Hesperia dacotae* (Skinner 1911) [Hesperiidae, Lepidoptera] is a small prairie obligate butterfly which prefers anthropogenically undisturbed habitat with diverse native flora (Royer and Marrone 1992; Swengel and Swengel 1999). It occurs in Minnesota, North and South Dakota, Saskatchewan and Manitoba, and is extirpated from Illinois and Ohio. In Saskatchewan, Dakota skipper was first recorded in 2001. Subsequent surveys found Dakota skipper present between Glen Ewan, Oxbow, and Roche Percee in the Souris River valley in Saskatchewan (Hooper 2003; COSEWIC 2014). In southern Manitoba, Dakota skipper is found in the Interlake and Oak Lake regions. It is likely extirpated from southeast Manitoba near Tolstoi, where it was last collected in 1987, and presumably last sighted in 2000 (Britten and Glasford 2002; COSEWIC 2014; US Fish and Wildlife Service 2014).

Habitat destruction and degradation, and extreme weather events are primary persistent threats that have resulted in the majority of population loss, followed by extreme fragmentation of remaining suitable habitat (Dana 1991; Environment Canada 2007; Henderson and Koper 2014), resulting in dramatic population declines with local and provincial/state level extirpation (COSEWIC 2014; US Fish and Wildlife Service 2014). As a result, Dakota skipper is now listed globally Endangered, Endangered in Canada, and Threatened in the United States, and the Province of Manitoba (Manitoba 1998; Canada 2018; Royer 2019; US Fish and Wildlife Service 2014, 2021; Canada 2023).

The Canadian Federal Recovery Strategy for the Dakota skipper (Environment Canada 2007) sets recovery objectives and planning priorities for federally listed species. A high priority objective for the Dakota skipper is to create species distribution models to aid in identifying potential habitat to guide survey activities and counter habitat loss and degradation.

Species distribution models (SDMs), also called element distribution models or habitat suitability models, relate species occurrence locations to a selection of biophysical variables at those locations and then use those relationships to predict distribution across a land region of interest (Elith and Graham 2009). The relative probability surface model output for the study area is used to predict occurrences. Other model outputs include relative contribution of variables to the model and model performance statistics (Phillips *et al.* 2006; Phillips and Dudík 2008; Merow *et al.* 2013).

Species distribution models are used in conservation science to more efficiently guide survey and conservation efforts to discover new occurrences, locate suitable habitat to support reintroduction efforts. They can also be used to help evaluate land acquisitions and prioritize habitat at risk of anthropogenic disturbance. They are also increasingly used to screen development projects for species at risk, and distribution changes due to climate change (Pielke 2003; Anderson and Martinez-Meyer 2004; Guisan *et al.* 2013; Lipsey *et al.* 2015; Fourcade 2016; Smeraldo *et al.* 2017; Villero *et al.* 2017; Hunter-Ayad *et al.* 2020; Bellis *et al.* 2024; NatureServe 2024) and interestingly, to inform invasive species control (Jiménez-Valverde *et al.* 2011), and paleoecology research (Fløjgaard *et al.* 2009).

Most modern SDMs leverage simultaneous advances in computer hardware, geographic information systems (GIS) technology, and greater access to digital data, and simultaneous advances in more sophisticated modelling software to produce more complex models (Guisan and Thuiller 2005; Beauvais *et al.* 2006; Lahoz-Monfort *et al.* 2010; Guisan *et al.* 2013). The general approach is to use SDM software with species occurrences and digital geospatial biophysical layer variable inputs to construct the species proximate niche space and extrapolate the average of the proximate niche across a study area (Elith and Leathwick 2009; Merow *et al.* 2013).

A common approach to obtain these model variables is to use “pre-made” readily available physical/environmental spatial products the user is familiar with (e.g. Land use/land cover, soil, Forest Resource Inventory, digital elevation model, and climatic variables) (Rabe *et al.* 2002; Wintle *et al.* 2005; Royer *et al.* 2008; Elith and Leathwick 2009; Lahoz-Monfort *et al.* 2010; Lapin *et al.* 2013; Westwood *et al.* 2019; Dearborn *et al.* 2022). A different, or complementary approach, which is gaining traction with the greater accessibility of no-cost satellite imagery, is to create model variables solely from satellite imagery (Lahoz-Monfort *et al.* 2010).

Advantages to variables created from satellite imagery include: (1) Imagery acquisition intervals for an area is often weeks or even days where common model inputs are often not updated for years and at considerable expense, (2) Smaller refresh windows and an extensive archive make it easier to temporally match species observations with modelling variables, (3) Imagery often has comparable or better spatial resolution, (4) satellite imagery offers superior spatial coverage which is consistent across jurisdictional boundaries and available for areas missed by common model variables, (5) many variables can be generated from satellite imagery which cannot be created from common model inputs, and (6) these variables tend to be finer continuous scale variables (i.e. NDVI) than more coarsely classified common model variables (i.e. land use) which may not inform the model as well (Bellis *et al.* 2008; St-Louis *et al.* 2009; Lahoz-Monfort *et al.* 2010; Duro *et al.* 2014; St-Louis *et al.* 2014).

Previous studies have used satellite imagery as the sole generator of model variables (Bellis *et al.* 2008; Lahoz-Monfort *et al.* 2010; St-Louis *et al.* 2014; Halford *et al.* 2024). However, they lacked provisions to compare the resulting model to a model using common variables. Indeed, their study area was deliberately chosen because it was identified as an environmental layer "data-poor region" (Lahoz-Monfort *et al.* 2010). The commonly used land cover variable was also criticized primarily for removing within-class habitat variability which may degrade species-habitat relationship information needed for modelling. Two studies attempted to address the criticisms of using land use derived model variables by comparing them to NDVI (Duro *et al.* 2014) and texture (Culbert *et al.* 2012) variables derived from Landsat imagery. However, in both studies, the focus was on narrowly comparing categorical land use variables to continuous NDVI and texture variables to describe overall species richness in a relatively large area, instead of creating rigorous SDMs for a species in a defined region.

Past Dakota skipper research in Manitoba, Saskatchewan and the northern U.S. has generally focused on describing local biotic and abiotic site characteristics (Dana 1991; Rigney 2013; Seidle *et al.* 2018) as well as some preliminary SDM studies to determine potential habitat range of Dakota skipper and Poweshiek skipperling, *Oarisma poweshiek* (Westwood *et al.* 2019;

Post van der Burg *et al.* 2020; Seidle *et al.* 2020; Dearborn *et al.* 2022) including additional modelling response to climate change (Barnes *et al.* 2024).

This study compares the utility of these two different modelling approaches. I constructed and compared the performance of two SDMs within the same regional study area to identify suitable habitat for the Dakota skipper: One model made use of commonly available "traditional" geospatial model variables (the Common Model) while the other model used satellite imagery to create the variables (the Satellite Model).

As part of this study, I also validated the models by ground truthing model results, which has been shown to be a better indicator of model performance than by evaluating only model output statistics (Westwood *et al.* 2019; Dearborn *et al.* 2022) and which the other studies did not perform. The objectives of this project were to: 1) Compare Dakota skipper models using freely available high resolution remotely sensed products to those using more traditional environmental predictors. 2) Field validate both models to identify the most accurate model using efficient and economical methods. 3) Address issues of modelling rare species to produce a robust SDM for the Dakota skipper in Manitoba.

CHAPTER 2: LITERATURE REVIEW

2.1 Study Organism

2.1.1 Description

The Dakota skipper (*Hesperia dacotae* Skinner 1911) is a butterfly from the Skipper family (Hesperiidae) named for their distinctive fast, low zig-zag flight pattern above the grass canopy, described as “skipping”, much like a stone on water (Klassen *et al.* 1989; Layberry *et al.* 2015). Dakota skipper was previously considered a subspecies of *H. sassacus* [Harris 1862] (Skinner 1911) but was subsequently elevated to species level (Skinner and Williams 1924)

There are 70 species of Hesperiidae in Canada which can be distinguished from other butterflies by their often dull colouration, thicker thorax, hooked antennae and relatively short wings and body size (Klassen *et al.* 1989; Layberry *et al.* 2015). The Hesperiidae is currently divided into six subfamilies, of which three are found in Canada. The Dakota skipper is one of 45 skippers that are part of the Branded skippers subfamily (Hesperiinae) and differentiated by their orange-brown tawny colouration, larger bodies and small wings. When resting, adults hold their forewings in the vertical plane and hindwings horizontally giving them an “X-Wing” fighter jet appearance (*Star Wars: Episode IV - A New Hope* 1977). Females and males in this subfamily have different wing colouration with the females usually being darker. The males also have a dark “brand” of scent scales, called a “stigma” found on the forewing, which the females lack. Adults have a wing-span between 21 and 33mm. The males tend to be pale yellow-orange above with a brown border and relatively short stigma. The females can be greyish to brown above. Both are greyish brown below, with the male slightly more orange, with relatively less pronounced medial spots (Klassen *et al.* 1989; Layberry *et al.* 2015).

Dakota skipper eggs are just over 1 mm in diameter, semi-hemispherical, and translucent white (Dana 1991). Larvae are 19 to 22 mm in length when fully grown, with a brown body, and a black head, spiracles, and legs. Lateral pits on the head casing distinguishes them from other Hesperia. Pupae are reddish-brown in colour (McCabe 1981).

Similar species with range and flight period overlap in Manitoba include Ottoe skipper (*Hersperia ottoe* WH Edwards, 1866), Long Dash skipper (*Polites mystic* WH Edwards, 1863), and Tawny-edged skipper (*Polites. themistocles* Latreille, [1824]) (Klassen *et al.* 1989; COSEWIC 2014; Layberry *et al.* 2015). Other similar species not within the range or flight period in Manitoba include Plains skipper (*Hersperia assiniboia* Lyman, 1892); Leonard's skipper (*Hersperia leonardus* Harris, 1862), and Indian skipper (*Hersperia sassacus* Harris 1862) (Klassen *et al.* 1989; COSEWIC 2014; Layberry *et al.* 2015).

2.1.2 Biology and Life Cycle

The flight period in Manitoba and Saskatchewan is generally between late-June to mid-July (Klassen *et al.* 1989; Dearborn and Westwood 2014; Layberry *et al.* 2015) and seems to be either earlier or comparable with populations further south in the U.S. depending on the year (McCabe 1981). It may fly one or two weeks later in the western portion of its range in North Dakota (McCabe and Post 1977), possibly due to spring time variation (McCabe 1981). Observations made in late July 2002 in Saskatchewan may have been due to an unusually cool spring (Hooper 2003).

Adult males emerge about five days earlier than females, possibly due to longer larval development time for females due to egg formation (McCabe 1981). Mating occurs on the emergence date and the female will continually lay eggs throughout the two to four-week lifespan. Eggs hatch in about 10 days (range 7 to 20) depending on temperature (McCabe 1981). From July to October larvae will continue to develop with noticeably slower activity in the fourth or fifth instar as they enter diapause to overwinter. Diapause appears obligate as death occurred in most larvae which were artificially forced to develop further (Dana 1991). Post-diapause development occurs between May and June with two additional instars until pupation (Dana 1991).

2.1.3 Behavior

Newly hatched larvae crawl to the base of grass plants and make shelters on the stem at, or just below ground level out of silk and grass clippings. Shelters will be enlarged two to three times as the larvae grows. After diapause, one or two shelters are constructed horizontally on the ground and can reach several centimeters in length. Pupation usually occurs in specially constructed silk chambers the size of the larva with a sealed entrance (Dana 1991). Most feeding occurs within the shelter with the larva harvesting food at night (McCabe 1981; Dana 1991) or at sunset or during overcast conditions (Dana 1991).

During the day, adult males and females will regularly obtain flower nectar while pursuing other activities such as mating or oviposition. Adult males tend to perch on flowers waiting for females to pass by or will slowly fly around mating sites just above the grass canopy. A male will pursue a passing female often flying slightly above and ahead. They will alight on vegetation and the female will mate if receptive (McCabe 1981; Dana 1991). Males will also tend to pursue other males, different skipper species, and other insects such as moths. Engaged Dakota skipper males will tend to twirl around each other ascending two to three metres and often more laterally (McCabe 1981; Dana 1991).

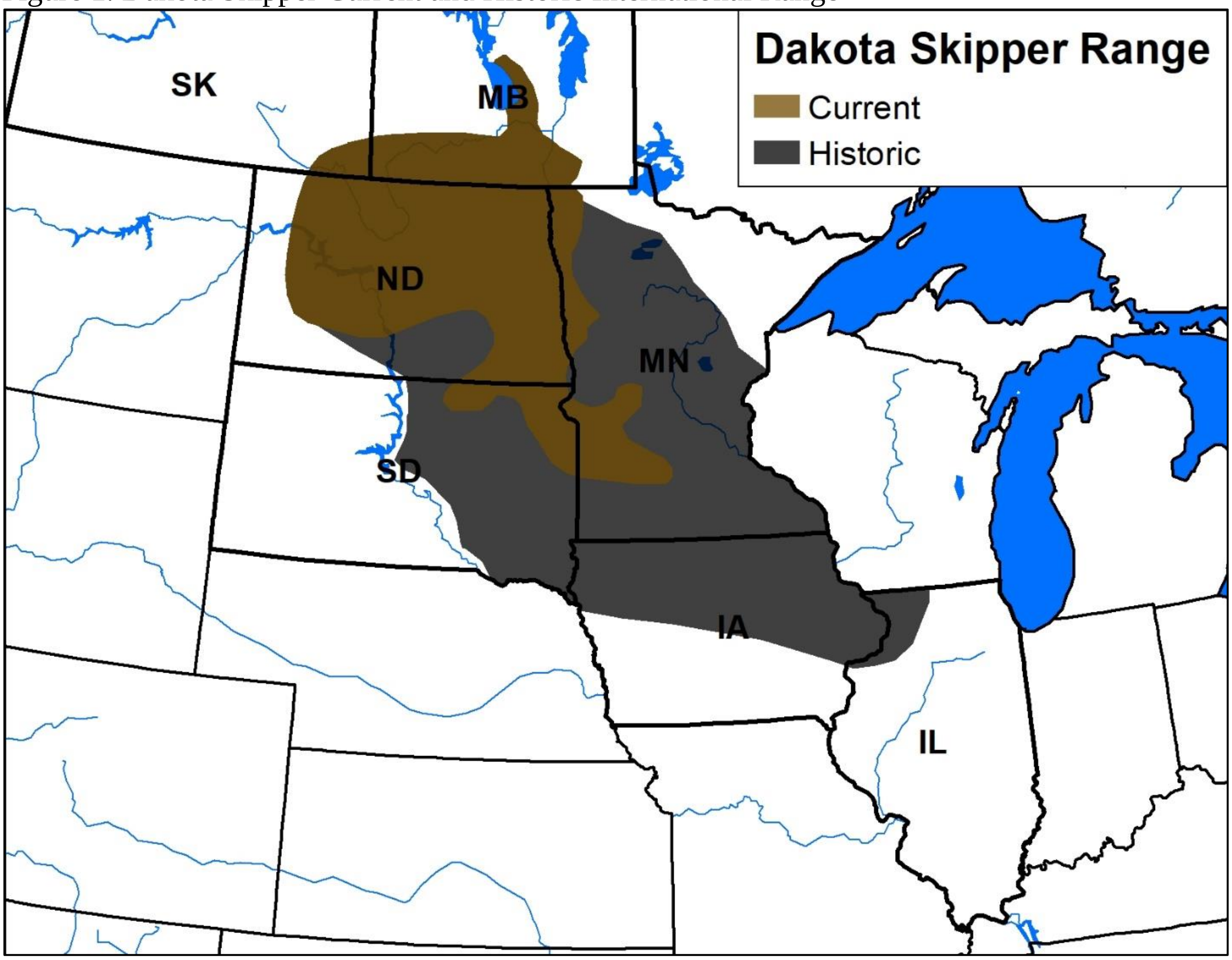
Perching and pursuit of other males led McCabe (1981) to conclude that Dakota skipper males were aggressively defending territories. However, Dana (1991) argued that, among other evidence, male perch fidelity is low, available perches are abundant and tend to give the same advantage, and that male pursuits tend to be more investigative than aggressive.

Female oviposition occurs during the day. Females fly just above the grass layer and land in a bare patch of ground or on a rock. The female will then climb to a suitable blade of grass and affix a single egg usually to the underside of the blade, about two to four centimeters above the ground. After oviposition, the female climbs up the grass to fly (Dana 1991).

2.1.4 Distribution and Population

Global range of Dakota skipper (Figure 1) extends into west Minnesota, east North and South Dakota, southeastern Saskatchewan along the Souris River, and southern Manitoba. It is presumed extirpated from Illinois and Iowa (US Fish and Wildlife Service 2014). Within Manitoba there are two existing population areas, one in the Interlake region by Lundar, and another in the southwest by Oak Lake. A third population was once located in southeast Manitoba in the Manitoba Tall Grass Prairie Preserve near Tolstoi, but is now considered extirpated (COSEWIC 2014). Historically, Dakota skipper was likely distributed across the northern prairie landscape but subsequent habitat destruction and extreme fragmentation has led to smaller isolated populations, which in turn are vulnerable to extirpation due to the inability of migration and establishment from other populations (Dana 1991). This idea is supported by recent genetic work which found current populations are isolated from each other but were likely connected in the recent past (Britten and Glasford 2002).

Figure 1. Dakota Skipper Current and Historic International Range



2.1.5 Status and Threats

Globally the Dakota skipper is listed “G2-Imperiled” (NatureServe 2022) and uplisted from “Vulnerable” (WCMC 1996) to “Endangered” on the International Union for Conservation of Nature and Natural Resources (IUCN) Red List (Royer 2019). In Canada it is listed federally under the Species at Risk Act as Endangered (Canada 2018, 2023) uplisted from Threatened (COSEWIC 2014; Canada 2017). In Manitoba it is ranked as “S2-Rare throughout its range - may be vulnerable to extirpation” (MBCDC 2023) and is listed as Threatened under the provincial Endangered Species and Ecosystems Act (Manitoba 1998). Saskatchewan does not have a formal provincial status but it is ranked “S1-Critically Imperiled” (SKCDC 2023).

The United States Fish and Wildlife Service listed Dakota skipper “Threatened” under the Endangered Species Act (US Fish and Wildlife Service 2014, 2021) In Minnesota it was listed “Threatened” in 1984 and updated to “Endangered” in 2013 (MDNR 2018). In North Dakota it became listed as “Threatened” in 2014 via the federal Endangered Species Act (US Fish and Wildlife Service 2014; North Dakota Fish and Game 2021). It does not have a state status in South Dakota but is ranked “S2- Imperiled” by the South Dakota Natural Heritage Program (SDGFP 2022). Iowa lists Dakota skipper as “Endangered” however it is likely extirpated (US Fish and Wildlife Service 2014; Iowa Natural Resource Commission 2022). It is also likely extirpated from Illinois (US Fish and Wildlife Service 2014).

These listings are due to a measured decline in range extent, number of sites, and numbers of individuals at sites (COSEWIC 2014; US Fish and Wildlife Service 2014, 2021). Identified primary threats contributing to these declines include conversion of habitat to cropland, overgrazing and trampling from cattle, and pesticide drift from agriculture crop fields. Some losses are also attributed to an increase in severity and frequency of flooding in the skipper’s low relief habitat. (McCabe 1981; Dana 1991; Swengel and Swengel 1999; Environment Canada 2007; Rigney 2013; US Fish and Wildlife Service 2021). Inappropriate application of management techniques such as timing, frequency, and extent of prescribed burning, and haying (McCabe 1981; Dana 1991; Swengel and Swengel 1999) are also threats. Management techniques and natural disturbance events can be especially damaging when not scaled to the inherently small habitat patch sizes. This reduces undisturbed habitat refugia which supports individuals both during and post-disturbance (McCabe 1981; Dana 1991; Schilcht and Saunders 1995; Dana 1997; COSEWIC 2014). However, appropriate management including prescribed burning and haying can benefit the species (McCabe 1981; Dana 1991; Swengel and Swengel 1999; Bates 2006).

2.1.6 Habitat Affinity

Dakota skipper is considered a prairie obligate found in undisturbed (untilled, intact, remnant) and undegraded (little or no grazing or burning) prairie with high plant diversity (Royer and Marrone 1992; Swengel and Swengel 1999) where they have available appropriate adult nectar and larval food and shelter plants and microhabitat physical conditions. Royer and Marrone (1992) and Royer *et al.* (2008) roughly generalise Dakota skipper habitat into "Type A lowland" and "Type B upland" classes based on topography, soil, and hydrology. Type A is described as wet-mesic tall grass prairie characterised by very little topographic change of a few meters over near-shore glacial lake deposits. Type B is described as dry-mesic mixed grass prairie on rolling hills over gravel moraines. Both types are further, and confusingly, subdivided into relatively drier "upland prairie" and wetter "lowland prairie" portions. The relatively dry upland prairie portions are overwhelmingly where Dakota skipper is found (Braker *et al.* 1985; Dana 1997; Swengel and Swengel 1999; Cochrane and Delphey 2002; Webster 2003; US Fish and Wildlife Service 2014).

In Type A prairies, drier upland prairies are in a patchy mosaic with wetter lowland prairie and non-prairie communities including lower relief wet meadows, grading down slope to, marshes, then open water. Shrub, Oak-Aspen stands tend to fill in any larger openings. The low relief and poor drainage can cause flooding from seasonal spring melts, but the upland prairie "islands" tend to protect the Dakota skipper to some degree (Royer *et al.* 2008). However, some upland portions can be inundated during prolonged precipitation events, or multi-year wet cycle (Rigney 2013). Soils tend to be calcareous and alkaline, suitable for cattle pasture or haying agricultural activities and less suitable for cropland (McCabe 1981; Webster 2007).

Type B prairies have drier upland mixed grass or tall grass prairies on well drained hill sides, with the wetter lowland prairies occurring between where water can accumulate (Royer *et al.* 2008) on alkaline calcareous or sandy loam soils (Dana 1997). Type B prairie can be further classed into three variants when considering upland and lowland arrangement. Type B Variant 1 is described as steep river valley hills (Selby and Glenn-Lewin 1989; Cochrane and Delphey 2002; Webster 2007; Royer *et al.* 2008) created by the "dendritic growth of tributary ravines" (Dana 1991, 1997). Soils tend to be calcareous to sandy loam (Dana 1997).

Variant 2 occurs on well drained flat upland areas adjacent to the valley slopes (Selby and Glenn-Lewin 1989; Dana 1997) and consist of loams (Dana 1991) and smaller exposed or near surface gravel (Selby and Glenn-Lewin 1989). Variant 1 and 2 on level uplands and valley slopes do not seem to have a complementary lowland prairie component except when viewed at a larger scale where lowland prairie and wet meadows are found on the valley bottom (Selby and Glenn-Lewin 1989; Dana 1991, 1997).

Variant 3 is described by (McCabe 1981; Braker *et al.* 1985; Dana 1997) as close calcareous gravel beach ridges creating a ridge and swale topography with mixed grass upland prairie on well drained ridges, and lowland prairie on poorly drained depressions.

The steep nature and poor soils of the Variant 1 hills tend to protect these areas from destruction by crop agriculture and generally grazing is regulated to low and moderate levels (Braker *et al.* 1985; Dana 1991; Royer and Marrone 1992; Webster 2007). Variant 2 on flat uplands have tended to be destroyed by crop agriculture or severe grazing pressure (Selby and Glenn-Lewin 1989; Dana 1991, 1997).

In Manitoba, almost all known Dakota skipper occurrences occur in Type A prairie (Webster 2003, 2007; Rigney 2013; MBCDC 2017) and characterised by Webster (2003, 2007) as tall grass prairie, or tall grass transitioning to mixed grass prairie (Rigney 2013).

Webster (2007) and Rigney (2013) noted that survey sites were similar floristically between the Interlake and south-west Manitoba regions, but both did not make comparisons to the south-east Manitoba region. Morden (2006) compared the Interlake and south-east Manitoba region and found plant species composition was significantly different between the Interlake and the south-east. The Interlake species abundance was more evenly distributed where the south-east was generally dominated by a few plant species. Adult and larval host plants were also more abundant in the Interlake than the south-east. Shrubs were also commonly dispersed throughout the south-east prairies but generally sparse in the Interlake making the Interlake prairies relatively open.

In Saskatchewan, all occurrences have been discovered on Type B Variant 1 prairie (Webster 2003, 2007; SKCDC 2017) characterised as steep river valley hillsides under some grazing pressure

(Webster 2007). Interestingly, Webster (2007) only found adults at the top and within “low relief” gully patches usually less than 200m^2 and containing bluestem grasses (*Schizachyrium scoparium* (Michaux) Nash) and (*Andropogon gerardi* Vitman) and not on the more relatively exposed intervening mixed grass prairie, suggesting Dakota skipper was found in remnant tall grass prairie patches or transitioning to mixed grass prairie as found by Rigney (2013). Royer and Marrone (1992) observed similar use of tall grass prairie “microsites” on hill slopes in western North Dakota. Henderson and Koper (2014) provide an updated reconstruction of historical tall grass prairie in western Canada and conclude that “Fingers of tall-grass prairie extended further west along the lower slopes and floodplains of major melt-water valleys radiating into adjacent Saskatchewan and North Dakota” occupying mesic sites. They add that these prairies have been greatly disturbed and difficult to distinguish from mixed grass prairie. Herbel and Anderson (1959) also noted that sloped areas can be more resilient to grazing disturbance and Dana (1991) observed on sloped terrain, where grazing had ceased, tall grasses began re-establishing in mixed grass areas and that after five years developed a “tall grass character”. He also noted a greater tall grass component in areas on slopes that were originally excluded from grazing.

2.1.7 Nectar and Larval Plants and Indicator Species

Dakota skipper adults are likely opportunistic feeders and have been observed nectaring on a wide variety of forbs throughout their range (see McCabe 1981; Dana 1991; Swengel and Swengel 1999; Webster 2003; Rigney 2013; US Fish and Wildlife Service 2014). These differences observed across their range and between studies is likely due to seasonal availability of favoured nectar plants and annual variability in Dakota skipper densities requiring, in some years, individuals to seek alternative, less favoured nectar sources (Swengel and Swengel 1999; Webster 2003).

In Manitoba, Dakota skipper have been observed nectaring primarily on Black-eyed susan (*Rudbeckia hirta* L. or *R. serotina* Nutt.), Wood lily (*Lilium philadelphicum* L.), Hairbell (*Campanula rotundifolia* L.), Dogbane (*Apocynum* sp.) (Webster 2003; Rigney 2013) and Smooth camas (*Anticlea elegans* Pursh)(Webster 2007). In Saskatchewan Webster (2003, 2007) notes Purple cone flower (*Echinacea angustifolia* (DC) (Heller) as the primary nectar source.

Observations of wild larvae *in situ* in their habitat have never been recorded (Nordmeyer *et al.* 2021) however, Dana (1991) studied larvae hatched from collected eggs and placed on potted plants located in prairie habitat to show that Dakota skipper larvae can feed on a variety of grasses, and were most frequently observed eating: Little bluestem (*Schizachyrium scoparium* (Michx.) Nash), Big bluestem (*Andropogon gerardii* Vitman), Prairie dropseed (*Sporobolus heterolepis* (A. Gray) A. Gray), Sideoats grama (*Bouteloua curtipendula* (Michx.) Torr.), to some extent on Sand millet (*Dichanthelium wilcoxianum* (Vasey) Freckmann), and Kentucky blue grass (*Poa pratensis* L.).

Nordymeyer *et al.* (2021) studied larvae in the laboratory who were fed one of seven grass species from hatching to adulthood. Survivorship was greatest for larvae fed one of the five native grass species (Big bluestem *Andropogon gerardii*, Sideoats grama *Bouteloua curtipendula*, Porcupine grass *Hesperostipa spartea,* Little bluestem *Schizachyrium scoparium*, Prairie dropseed *Sporobolus heterolepis*) and poorest for larvae fed one of the two invasive grass species (Smooth brome *Bromus inermis*, and Kentucky bluegrass *Poa pratensis*) offered in the study. Smooth brome, which is pervasive in prairie habitat, had the lowest survivorship, smallest larval mass, and longest to reach maturity than larvae fed on the other six grass species. Additionally, of the five native grasses, Big bluestem and Little bluestem are generally associated as ideal larval hosts. However, in this study, larvae had poorer survivorship than the larvae fed the other three native grasses.

Indicator plant species in Dakota skipper habitat could include the above nectar and larval hosts. Rigney (2013) conducted vegetation surveys at known Dakota skipper sites in the Interlake region finding five indicator plant species associated with Dakota skipper sites: Stiff goldenrod *Solidago rigida*, Creeping bentgrass *Agrostis stolonifera*, Creeping spikerush *Eleocharis palustris*, Rigid sedge *Carex tetanica*, and Trembling aspen *Populus tremuloides*. Webster (2007) used evidence of *A. elegans*, *L. philadelphicum*, *R. serotina*, and *C. rotundifolia* along with *S. scoparium* and *A. gerardii* to locate potential habitat in Manitoba, and *E. angustifolia* with *S. scoparium* and *A. gerardii* in Saskatchewan as they were easy to spot from the road. McCabe (1981) noted *A. elegans* as an “extremely good indicator” of alkaline prairie, the preferred habitat in North Dakota. Notably, McCabe (1981) also considered hayed areas, lack of cattle grazing, and areas close to gravel quarries as indicators of potential habitat.

Similar species with range and flight period overlap in Manitoba include Ottoe skipper (*Hersperia ottoe* WH Edwards, 1866), Long Dash skipper (*Polites mystic* WH Edwards, 1863), and Tawny-edged skipper (*Polites. themistocles* Latreille, [1824]) (Klassen *et al.* 1989; COSEWIC 2014; Layberry *et al.* 2015). Other similar species not within the range or flight period in Manitoba include Plains skipper (*Hersperia assiniboia* Lyman, 1892); Leonard's skipper (*Hersperia leonardus* Harris, 1862), and Indian skipper (*Hersperia sassacus* Harris 1862) (Klassen *et al.* 1989; COSEWIC 2014; Layberry *et al.* 2015). Garita Skipper (*Oarisma garita* (Reakirt)) and Long Dash skipper are present at most sites in Manitoba and Saskatchewan (COSEWIC 2003).

In Minnesota and North Dakota documented predators include: Ambush bugs (*Phymata* sp.), flower spiders, (*Misumena vatia* Clerck and *Misumenops carletonicus* Dendale and Redner), orb weaver spiders (Araneidae: *Misumenops* spp.), and ants capturing larvae (Cochrane and Delphey 2002, McCabe 1981; Dana 1991). The parasite, *Ooencyrtus* sp., has also been found in lab reared eggs from field collections. Potential predators include: Robber flies, dragonflies, birds, small mammals (Lederhouse et al. 1987).

2.1.8 Habitat Suitability Modelling

Niche Space Model Concept

Hutchinson (1957) described the niche space concept as one of the "fundamental niche" and "realized niche" of a species. The fundamental niche contains all the environmental variables required for survival and propagation. The fundamental niche is generally regarded as theoretical since, due to species competition, disturbance, and biological or physical barriers, all the species' environmental space cannot be occupied. The realized niche is then this redacted subset. Niche space models attempt to approximate the realized niche space by using species occurrences and a subset of realized niche environmental variables, called environmental predictors, (Elith *et al.* 2006) and then project this modelled realized niche space to a geographic area (Phillips *et al.* 2006). Model output will generally consist of statistics which measure model performance and a “response surface” or map showing the model’s prediction of species presence variability across an area. Output may also provide the relative contribution of each variable to constructing the model. Variable response curves are generated which describe the where along the full range of a variable’s values a species is likely to occur.

Model Data Types

Guillera-Arroita *et al.* (2015) placed model type occurrence data into four classes: presence only (PO), presence-background (PB), presence-absence (PA), and occupancy-detection (DET). Westwood (2017) includes species biology (SB), and abundance-absence (AA). Species biology data is information about a species physiology or ecology (e.g. temperature, habitat) usually obtained from expert opinion and research. It is then used to create a species profile for modelling.

As the names imply, PO data consists of only presence occurrences. PO data is frequently used with a random background of samples taken from the geographic area, which changes the data to PB. PA contains both presence and absence occurrences. AA data is a variant of PA but contains measures of abundance instead of only presence. DET consists of detection and non-detection records generated from repeat visits to the same site or time-to-detection in a single visit. DET can be used to calculate probability of presence for different habitat types across the geographic region.

Guillera-Arroita *et al.* (2015) noted that DET data is the most information rich followed by PA then PO. They reported that the most common data type (57%) used in SDM studies is PB. This has been attributed to the increased availability of digitized species occurrence data from museums, herbaria, and citizen science efforts (Elith *et al.* 2006; Westwood 2017) and better model performance than using PO data (Elith *et al.* 2006). Even when PA data is available, absences may be of suspect value (Anderson and Martinez-Meyer 2004) due to temporal or spatial surveying limitations.

Model Types

Burgman *et al.* (2005) locate habitat suitability models within the broader field of landscape ecology. They use the term “model specification” where Barry and Elith (2006) use the term “model frame” to recognize which model class is being used, and the associated errors and assumptions inherent in that class. If the model frame is appropriate, then a good approximate relationship or linkage can be developed between the species of interest and the landscape.

Model frame selection depends heavily on available data type. A mismatch can cause poor model performance. Given the vast number of modelling frames available for use (34 types in 100 randomly chosen studies) and the preponderance of PO datasets available, it is perhaps best to focus initial efforts on models capable of using PO data (Guillera-Arroita *et al.* 2015).

Model evaluation studies using standardized geographic regions and datasets have shown that multiple linear regression and machine learning types perform better than more established approaches using PB data including general linear models (GLMs), general additive models (GAMs), multivariate regression (MARs), and genetic algorithms (GARPs) (Elith *et al.* 2006; Hernandez *et al.* 2006). This success is due to the newer methods' increased capability in modelling non-linear complex environmental relationships (Barry and Elith 2006). Evaluation of this subset of model types ultimately found that machine learning approaches using maximum entropy models, and multivariate approaches using boosted regression trees (Elith *et al.* 2006; Hernandez *et al.* 2006; Phillips *et al.* 2009) performed the best.

2.1.9 Environmental Predictors

Predictor Types

Modern environmental predictors are geospatial and include both vector (e.g. a soils layer) and raster data (e.g. satellite imagery) types. Conventional predictors usually include climate data, soils and land use/cover layer, and a digital elevation model. However, use of no-cost, high resolution, multispectral remotely sensed imagery is increasingly becoming an attractive addition or alternative option for SDM work, especially in areas lacking more traditional predictors (Lahoz-Monfort *et al.* 2010).

Derived Predictors

Derived predictors are produced from "raw" predictor data and can be thought of as extracting certain information from the source data. Many varieties of derived predictors can be generated from digital elevation surfaces, satellite imagery (Lischke *et al.* 1998) and vector layers. For example, slope and aspect layers can be generated from a digital elevation surface, and distance from nearest road can be derived from a road layer. Wintle et al. (2005) note that derived predictors can perform better than the raw products.

Proximal and Distal Predictors

Environmental predictors can be divided into "proximal" and "distal" types (Austin 2002). Proximal predictors are more physiologically meaningful as they are directly related to the actual environmental variable they represent, while distal predictors are relatively indirect, merely correlated to proximal predictors (Lischke *et al.* 1998; Austin 2002). For example, an average temperature surface may be considered more proximal than a total solar insolation surface. In most cases though, distal environmental predictors are more frequently used in modelling, as they are more readily available (Austin 2002).

Evaluation of Environmental Predictors

Model output usually shows the relative amount of information each variable contributes to creating a model as a percentage. Poorer performing variables, which are not contributing much to the model could be dropped in favour of other variables which may perform better. Additionally, to reduce model overfit, variable selection is often limited to one variable for every ten point observations. Therefore, removing poorly contributing variables will free up more points which could be allocated to the test sample which is used in cross-validation analysis (see section on model evaluation below). (Phillips *et al.* 2006; Phillips and Dudík 2008). Contribution scores could also be used to tease out variable correlations by selectively removing one of two or more variables which are suspected of being correlated before running the model building again and comparing overall model performance and variable contribution (Lahoz-Monfort *et al.* 2010).

Model output also generates variable response curves which shows which range of values for a particular variable are predicted as being suitable for the species. Response curves can be used to describe the environmental conditions which might be present at a location with a high or low relative likelihood of occurrence (Phillips *et al.* 2006; Phillips and Dudík 2008).

2.1.10 Model Evaluation

Evaluation via Model Output

A common method of evaluating overall model performance is to use model statistics generated during output. The area under the receiver operator curve (ROC or AUC) is generally used as a measure of model performance (Elith *et al.* 2006; Phillips and Dudík 2008) and cross-model comparison (Elith *et al.* 2006; Phillips *et al.* 2009).

AUC was borrowed from medical science used to evaluate the quality of a diagnostic test (Elith *et al.* 2006). i.e. the ability to successfully class true disease (true positives) from truly non-disease (true negatives). However, as a diagnostic is never able to successfully classify all true positives as positives, nor all true negatives as negative, there will be some overlap resulting in false positives and false negatives. The AUC is the probability that a randomly chosen positive and negative pair are classed successfully. The curve plots a test's true positive rate (sensitivity) by false positive rate (1-specificity) at changing thresholds cross the range of possible values. In a test that shows perfect diagnosis ability (perfect discrimination), the AUC value will be 1 (100% sensitivity and 100% specificity). The closer the curve passes through the upper left corner of the graph, the better its performance, while an AUC value of 0.5 shows that discrimination is no better than random (Hanley and McNeil 1982).

In the case of species modelling, the AUC is the probability that the model will successfully class a randomly chosen positive occurrence and a randomly chosen negative (absent) occurrence. However, some authors have criticized using the AUC for evaluating SDM model performance (Lobo *et al.* 2008; Jiménez and Soberón 2020). Suggesting that instead of reflecting good model performance, due to the spatial nature of species data and the uncertainty of pseudo-absences, the high AUC values may indicate that the species is restricted to a portion of the available environmental envelope offered from the predictors in comparison to a generalist species which may have lower AUC values (Elith et al. 2006; Lobo et al 2008).

For most species, observation data will include presence but will likely lack reliable absences or if recorded it can be difficult to interpret (Brotons *et al.* 2004; Guillera-Arroita *et al.* 2015).

Therefore, most SDMs that use presence only data also add a set of random background occurrences, also known as pseudo-absences, which replace the true absences (Phillips *et al.* 2006). However, research shows that adding too many background points can artificially increase AUC values (Lobo *et al.* 2008) since background points can be a mix of both presences and absences.

Increase in study area size also tends to inflate the AUC value (Elith and Burgman 2002). Pseudo-absences which are more geographically distant are less likely to occupy the target species' environmental niche space which generates a higher number of true absences which will be discriminated by the model as such (Lobo *et al.* 2008; Jiménez and Soberón 2020). Similarly, drawing pseudo-absences from ever expanding background sampling areas, within the overall study area, also progressively inflated AUC values (VanDerWal *et al.* 2009) And related, using ever larger pseudo-absence sample sets also tends to increase AUC values (Phillips and Dudík 2008) likely by behaving similarly to the larger size of the study area or background sample area: More environmentally distant locations can be sampled.

Evaluation of AUC Derived from Cross-Validation

AUC values derived from a cross-validation procedure where a portion of the species occurrences are retained to test model performance, is considered more accurate than values calculated from fitting the training data alone (Merow *et al.* 2013). However, cross-validation will tend to overfit the model leading to inflated AUC values (Elith and Burgman 2002; Anderson and Raza 2010). Overfit and AUC inflation occur because it can be difficult to find spatially independent training and testing datasets when working with species occurrences (Hijmans 2012).

Evaluation Using Continuous Likelihood Values

AUC values are considered rank based where the higher the AUC value the better the model performs. However, a model's performance can also be tested using the correlation between the known occurrences and their predicted relative likelihood values. The Pearson correlation coefficient is the generally used correlation test (Elith *et al.* 2006; Phillips and Dudík 2008).

Evaluation Using Thresholding Model Results

Categorizing the continuous probability surface produced by the model into two or more classes based on a threshold value is a common practice to evaluate model performance (Scherrer *et al.* 2020). Areas are classed into species presence/absence or categories such as high and low suitability for the species on either side of the threshold. A confusion matrix is then generated and a statistical test such as Cohen's Kappa is applied (Merow *et al.* 2013). Thresholding has been criticized for strictly placing likelihood values into either/or bins which reduces model information and are often not reflective of a real division (Liu *et al.* 2005; Jiménez-Valverde and Lobo 2007; Merow *et al.* 2013; Scherrer *et al.* 2020). Threshold values are often semi-arbitrarily assigned for example, at 0.5 where values above are considered suitable, and values below are considered unsuitable. Larger values are also selected to increase confidence that suitable areas are in fact suitable at the expense of excluding some areas (Liu *et al.* 2005). Model output results have also been used to apply a more objective threshold with some degree of success (Jiménez-Valverde and Lobo 2007), however some question the utility of thresholding and instead suggest using tests which incorporates the continuous nature of the relative likelihood surface (Merow *et al.* 2013; Dearborn *et al.* 2022).

Evaluation via Field Ground Truthing

Field validation involves gathering a test data set by surveying sites within the study area and assigning a species presence likelihood or suitability score to the sites based on site attributes which are correlated to species presence. Similar to model statistic evaluation, the test data is then compared to the relative likelihood values from the response surface by applying thresholds and also by using the continuous relative likelihood values (Dearborn *et al.* 2022).

Field validation can be rightly ignored in studies where the objectives are to simply evaluate modelling techniques and responses (Araújo and Guisan 2006; Elith and Leathwick 2009; Warren *et al.* 2021). However, field validation is considered the gold standard for model evaluation for real world applicability (Araújo and Guisan 2006; Elith and Leathwick 2009; Jiménez and Soberón 2020; Westwood *et al.* 2019; Dearborn *et al.* 2022; Draper *et al.* 2019). Statistical evaluation or partial field validation where only highly suitable areas are surveyed should not be considered an appropriate substitute (Barry and Elith 2006; Elith and Leathwick 2009; Boria *et al.* 2014). Testing the model to an independent survey data set however is rarely done due to extra expense from field surveys, data management, and analysis (Araújo and Guisan 2006; Elith and Leathwick 2009; Westwood *et al.* 2019; Dearborn *et al.* 2022).

In contrast to a static one-time evaluation, an iterative workflow can also be set up, in collaboration with conservation managers where models are improved upon by informing field data capture which is used to inform modelling (Pielke 2003; Burgman *et al.* 2005; Barry and Elith 2006).

Cross-model comparisons using AUC Values

Cross-model comparisons are usually done to compare different modelling techniques, or assess different modelling parameters or decisions on model performance.

The same statistical tests used to evaluate a model can be used to make cross-model comparisons. Models can be ranked based on AUC values. Also, non-parametric tests such as Pearson’s correlation coefficient and Wilcoxon tests are used by comparing known occurrences to predicted with the resulting test statistic used to compare models against each other. And lastly, Kappa type statics are generated for each model from confusion matrixes by thresholding known and predicted occurrences (Elith *et al.* 2006).

However, cross-model comparisons tend to be highly constrained in space and time since AUC values are influenced by study area geographic extent and resolution, and target species among other model parameters (Araújo and Guisan 2006; Randin *et al.* 2006; Anderson and Raza 2010; Owens *et al.* 2013; Sousa-Silva *et al.* 2014).

2.1.11 Using a Model

As noted above, an AUC value higher than 0.5 indicates the model discriminates or classifies better than random (Hanley and McNeil 1982). Therefore, the higher the AUC value the better the model performs making it more useful to inform conservation management actions. However, there is no consensus on how high the AUC should be before model performance is considered adequate for application. An AUC of 0.8 is generally viewed that the model is useful, however arbitrary, that limit is assigned (Jiménez and Soberón 2020). Others note cut-offs at 0.75 (Elith and Burgman 2002;

Elith *et al.* 2006; Sousa-Silva *et al.* 2014) and 0.7 (Dearborn *et al.* 2022) above which the model is informative enough to be used for decision making.

When using a SDM the focus is usually on the relative likelihood surface generated for the species. This layer is usually central to using the model for species-at-risk conservation because it shows locations where the species potentially occurs, and does not occur, within the study area. However, the continuous likelihood values are difficult to understand (Jiménez-Valverde and Lobo 2007). Therefore, despite model evaluation generally discouraging thresholding, it is used here to make sense of the relative likelihood surface and be more useful for conservation work (Liu *et al.* 2005; Lobo *et al.* 2008).

Classification of the surface can be binary presence/absence or low, medium, and high classes (Phillips *et al.* 2006; Lahoz-Monfort *et al.* 2010; Anderson and Gonzalez 2011). These classes can also be quantified by area, as another method to cross-compare model results (Lahoz-Monfort *et al.* 2010).

2.1.12 Model Error, Bias, and Assumptions

Models invariably simplify the linkages between a species and how it distributes on a landscape. There will always be some mismatch or error between the predicted distribution and what is real (Veregin 1989; Barry and Elith 2006). Producing a highly accurate model is the goal of most studies, so model error is generally maligned. Although error is not necessarily a bad thing. Burgman *et al.* (2005) state that “All models are false” and a proper understanding of model error informs the level of uncertainty or risk in conservation actions so that managers can develop robust plans. Model error can also inform error in ecological knowledge, which can subsequently be used to correct model error using an iterative approach.

Species Occurrence Error and Bias

Habitat suitability modelling relies heavily on the accuracy of the species occurrence data used to construct the species “realized niche” or the species “environmental envelope”, which is subsequently mapped across a landscape to predict suitable habitat locations (Guisan and Zimmermann 2000). However, many authors discuss sample error (Veregin 1989; Hijmans *et al.*

1999; Rowe 2005; Elith *et al.* 2006; Graham *et al.* 2008) and sample bias (Hijmans *et al.* 2000; Reese *et al.* 2005; Araújo and Guisan 2006; Phillips *et al.* 2006), especially for modelling rare species, since occurrence data is usually sparse and extracted from surveys with other objectives in mind (Elith *et al.* 2006). Sample error generally refers to error made when collecting a sample. In the case of a Dakota skipper observation, examples of sample error could include: taxonomic, geographic location, location description, transcription, lack of absence data, and small sample size (Veregin 1989; Hijmans *et al.* 1999; Rowe 2005; Graham *et al.* 2008). Sample bias refers to unequal sampling effort and can be due to infrastructure (e.g. surveys confined near roadways), habitat (disproportional sampling in some habitats), hotspot (more sampling in areas of high species diversity), and species-area (over or under sampling in an area a species occurs) (Hijmans *et al.* 2000; Reese *et al.* 2005; Phillips *et al.* 2009). Other biases include a currency bias, where too much time has elapsed since collection occurrence/absence data, or where there have been multiple negative observations since initial discovery, or reduced confidence that the habitat is still suitable. The occurrence would then provide false information to the model (Pearson *et al.* 2004; Phillips and Dudík 2008). Absence data bias refers to non-detection of a species at a location due to under sampling or previous disturbances (Hirzel and Guisan 2002).

Both sample error and bias can degrade a model's predictive performance (Pearson *et al.* 2004; Reese *et al.* 2005). While sample error tends to affect performance directly, bias in geographic space can cause bias in environmental space and lead to model overfit to those biases (Phillips *et al.* 2009; Boria *et al.* 2014). Model overfit may show excellent performance but could effectively be incomplete by emphasizing sample effort through bias, and underemphasizing results related to the species environment and distribution (Phillips *et al.* 2009; Boria *et al.* 2014).

Some researchers (Guisan and Zimmermann 2000; Reese *et al.* 2005) suggest survey design should be developed with the expressed objective of modelling. However, most acknowledge that this is usually not realistic nor practical, (Lischke *et al.* 1998; Hijmans *et al.* 2000; Rowe 2005; Elith *et al.* 2006; Graham *et al.* 2008). Work is then usually done to mitigate these errors and biases (see Elith *et al.* 2006) or to at least acknowledge their existence and realize model performance may suffer

(Pearson *et al.* 2004; Reese *et al.* 2005). Others have tested spatial bias and found simulating close proximity to roads mimics a systematic sample design (Reese *et al.* 2005). Still others (Graham *et al.* 2008) compared models using actual occurrences to the same occurrences with simulated locational error and found model prediction was degraded but robust enough to still be useful. Other studies suggest averaging the environmental predictor values around a point occurrence to account for spatial inaccuracies (Graham *et al.* 2008) or partition the dataset based on location accuracy (Pearson *et al.* 2004).

To reduce spatial bias, Boria et al. (2014) employed *a priori* spatial filtering which they showed reduced model overfitting, and increased model performance. Phillips et al. (2009) found spatially biased presence-absence data still outperformed presence only or pseudo-absence data. They then modelled absence data by using occurrence data from a "target group" of species that share similar survey methods and spatial bias to increase sample size.

Static and Dynamic Models and Equilibrium

Static models assume equilibrium or quasi-static state between the target species and the environment where dynamic models simulate a changing environment over time (Lischke *et al.* 1998). Equilibrium can hold true in relatively short time frames, especially if there are no strong natural or anthropogenic disturbances, or succession (Lees and Ritman 1991; Brzeziecki *et al.* 1993), or if the species reacts slowly to change (Guisan and Zimmermann 2000). Occurrence currency error as noted above can also violate the equilibrium assumption (Pearson *et al.* 2004; Phillips and Dudík 2008).

Despite these drawbacks, static models can be more simply constructed, represent larger study regions, and require less detailed knowledge about the species (Guisan and Zimmermann 2000). Also, in some cases, environmental predictors which account for dis-equilibrium can be included in static modelling (Lees and Ritman 1991).

Dynamic models are generally considered more realistic but more complex to model because they include both space and time (Lischke *et al.* 1998). However relatively recent increases in computation and storage capacity in parallel with advances in a machine learning have made dynamic modelling easier to implement. Advances include developing sophisticated “Digital Twins” models (Botín-Sanabria *et al.* 2022) which could be implemented in SDM research.

Environmental Predictor Error

Just as there are numerous environmental predictors available for modelling, there are also numerous environmental predictor errors which should be addressed, if possible, to avoid poor model performance (Reese *et al.* 2005). Predictor errors include: Resolution, spatial error, currency, intercorrelation, and relevancy (Lischke *et al.* 1998; Austin 2002; Barry and Elith 2006; Phillips *et al.* 2006; Lahoz-Monfort *et al.* 2010; Fourcade *et al.* 2014).

Resolution error takes the form of geographic and environmental niche resolution (Lischke *et al.* 1998; Austin 2002; Barry and Elith 2006; Beauvais *et al.* 2006; Phillips *et al.* 2006; Lahoz-Monfort *et al.* 2010). Predictors need to interact with the species at the geographic scale of the study (Lees and Ritman 1991). For example, applying climatic variables for a regional level study is likely inappropriate. At the study scale, too high or low spatial resolution of the predictor can also cause error and render a model less accurate (Lischke *et al.* 1998; Pradervand *et al.* 2014). For example, 1 km grid size at a regional level may be too coarse a grain size, resulting in the elimination of smaller features (Barry and Elith 2006). Alternatively, high resolution predictors, coupled with highly accurate occurrence data and a vagile species may cause information loss, necessitating other approaches such as averaging predictors across a larger area (Graham *et al.* 2008).

Spatial error refers to locational accuracy, or how well located a predictor feature is in space (Barry and Elith 2006). For example, Landsat satellite images boast 30 m pixel resolution, however, Landsat 8 and 9 location error can be approximately 17 m to 30 m. Depending on the correction data available for Landsat 7, spatial error is generally greater (USGS 2024). Therefore, a pixel may actually represent one of eight 30 m patches of land adjacent to its current location. Vector soil layers are another example where spatial accuracy of the soil polygon features varies depending on the accuracy of original soils maps and the resolution they were digitized. Lower resolution results in coarser polygon borders that do not match actual physical locations of the soil they are meant to represent, as compared to higher resolution.

Currency or temporal error is also a concern (Barry and Elith 2006; Phillips *et al.* 2006). Predictors which are older or newer than the species occurrence may violate the species-environment equilibrium assumption.

Intercorrelated predictors present another error type potentially contributing to model overfit (Fourcade *et al.* 2014) because they convey similar information. An example is two vegetation indices derived from the same imagery bands. Derived predictors like these are particularly vulnerable and intercorrelation should be measured and controlled by removing highly correlated predictors (Brotons *et al.* 2004; Lahoz-Monfort *et al.* 2010; Fourcade *et al.* 2014; Merow *et al.* 2014).

A predictor also may not have information to contribute to modelling a species environmental niche (Austin 2002; Barry and Elith 2006). For example, distal predictors, such as slope or interpolated soil moisture data, may not adequately characterize a species proximal physiological trait such as moisture, nor have adequate resolution to inform a model.

Finally, models construct an approximate niche space from a selection of predictors and, if the selection is satisfactory, will produce a model with usable predictive power. However, adequate predictors may simply not exist for a species, or when available lack of ecological knowledge about the species prevents their use. Also, given the vast number of predictor options, modelers may be ignorant of available predictors. These are considered "missing" predictors and their absence may reduce model predictive power (Barry and Elith 2006). Building a useful model may still be limited by missing predictors despite leveraging the best available remote imagery and geospatial data and implementing the best current practices to create other novel predictor variables. For example, despite knowing that the Dakota skipper is a mixed and tall grass prairie obligate, it is currently difficult to detect tall grass prairie from remotely sensed imagery.

2.1.13 Past and Current Work

Previous studies have used satellite imagery as the sole generator of model variables (Bellis *et al.* 2008; Lahoz-Monfort *et al.* 2010; Culbert *et al.* 2012; St-Louis *et al.* 2014; Halford *et al.* 2024). However, they lacked provisions to compare the resulting model to a model using common variables. Indeed, their study area was deliberately chosen because it was identified as an environmental layer "data-poor region" (Lahoz-Monfort *et al.* 2010). The commonly used land cover variable was the focus for two of the studies (Culbert *et al.* 2012; Duro *et al.* 2014) which was criticized primarily for removing within-class habitat variability which may degrade species-habitat relationship information needed for modelling. Furthermore, classification removes the fuzzy boundary ecotone between habitats in favour of a hard-edge. And finally general concerns over classification standards, temporal currency, and expense in production of these layers (Bellis *et al.* 2008; St-Louis *et al.* 2014). These two studies attempted to address the criticisms of using land use derived model variables by comparing them to NDVI (Duro *et al.* 2014) and texture (Culbert *et al.* 2012) variables derived from Landsat imagery. However, in both studies, the focus was on narrowly comparing categorical land use variables to continuous NDVI and texture variables to pattern overall species richness in a relatively large area, instead of creating rigorous SDMs for a species in a defined region.

My study area has many years of coverage for commonly used model variable layers which will be used to test the utility of the two quite different model approaches to identify Dakota skipper habitat. Also, none of the studies validated their models by direct field ground truthing the results as done in my study.

My study also leverages newer sensor platforms (Landsat 8 and Sentinel 2) which offers higher spatial resolution and solves a previous issue with Landsat 7TM associated with degraded image quality (ESA (European Space Agency) 2015; USGS and NASA 2019). Both Common and Satellite models were generated using Maximum Entropy (MaxEnt) software (Phillips *et al.* 2006). MaxEnt is a "machine learning" modelling application which was chosen since it is freely available, commonly used by government and industry and relatively easy to operate (Fourcade *et al.* 2014). Studies which evaluated various modelling approaches found that machine learning models outperformed other types such as generalized additive models (Elith and Graham 2009; Dearborn *et al.* 2022). MaxEnt also can use presence-only species data points, which is preferred given studies have shown that presence-only models also outperform models requiring presence and absence species points (Elith *et al.* 2006; Dearborn *et al.* 2022). Relying only on presence-only data also eliminates the need for absence data which can be highly problematic to verify (Guillera-Arroita *et al.* 2015).

The Canadian Federal Recovery Strategy for the Dakota skipper (Environment Canada 2007) sets out recovery objectives and recovery planning priorities. SDM modelling is considered a high priority task to counter habitat loss and degradation and identify potential habitat to survey.

Previous research has described local biotic and abiotic site/habitat characteristics during surveys (Schilcht and Saunders 1995; Swengel and Swengel 1999; Cochrane and Delphey 2002; Webster 2003, 2007) and as the main objective of their study (Dana 1991, 1997; Bates 2006; Morden 2006; Royer *et al.* 2008; Rigney 2013). Four recent SDMs have been created for the Dakota skipper, however, all four used a mixture of common and satellite imagery derived environmental variables (Table 1). Three of the studies were situated in regions outside of the Manitoba Interlake area (Post van der Burg *et al.* 2020; Seidle *et al.* 2020; Barnes *et al.* 2024). The study by Seidle *et al.* (2020) was created for a region in south-western Saskatchewan and the study by Barnes *et al*. (2024) Post van der Burg *et al.* (2020) was applied to three and four US states, respectively. In both cases the habitat is likely substantially different than that found in the Interlake area of Manitoba and therefore they cannot be directly comparable. Dearborn *et al.* (2022) generated models for two areas in Manitoba, one area being the Interlake using MaxEnt modelling software. However, they used a mixture of common and satellite imagery derived environmental variables which have relatively low spatial resolutions compared to the Sentinel 2 imagery used in this study.

Table 1. Previous Dakota skipper SDMs with Environmental Variables

Study	Environmental Variables
Post van der Burg et al. 2020	1. Land cover 2. Soil series- Topographic slope 3. Soil series- Drainage
Seidle et al. 2020	1. Soil series- Class, bulk density, organic carbon, present sand, percent silt, percent clay, A horizon depth, ammonium. 2. Grasslands- Custom developed 3. BIOCLIM- Variables BIO1 to BIO19
Dearborn et al. 2022	1. Landuse- Distance to Cropland 2. Landuse- Distance to Wetland 3. Landuse- Grassland percent (moving window) 4. Landuse- Distance to Forest 5. Landuse- Grassland shared edge 6. Mean spring surface temperature- MODIS 7. Mean snowmelt date- MODIS 8. Topographic wetness model- SRTM 9. Soil series
Barnes *et al.* 2024	1. Rangeland Analysis Platform-Landuse 2. Potentially Undisturbed Lands Layer-Landuse 3. GeoMorph90 4. SoilGrids250v2.0 5. AdaptWest-BioClim 6. Sentinel 2-Indices

CHAPTER 3: METHODS

3.1 Study Area

The study area is located in the Interlake area of Manitoba, Canada centered on (97.60N 50.70W WGS84). The Interlake Study Area (IL) combines the Ecodistricts of Ashern, Gimli, Lundar, and the northern portions of Winnipeg and Portage (within the Interlake Plain and Lake Manitoba Plain Ecoregions) comprising approximately 24,800 km^2 (ESWG (Ecological Stratification Working Group) 1995; Marshal *et al.* 1999) (Appendix 3). Ecodistricts are "distinctive assemblages of landform, relief, surficial geological material, soil, water bodies, vegetation, and land uses" (ESWG (Ecological Stratification Working Group) 1995). I chose to use Ecodistricts to define the study area since they describe relatively natural boundaries and homogeneous environments in contrast to a sized rectangle. Ecodistricts were used to constrain the areas to be modelled relatively close to the "geographic domain" of the modelling data. In other words, extending the study area outside of this domain, into more heterogenous environments, can degrade model performance by introducing variation in unknown environmental variables or disturbance regimes (Elith *et al.* 2006; Randin *et al.* 2006). The study region contains of most of the known Dakota skipper Interlake occurrences and encompasses what appears to be the Type A habitat present in the Interlake. There is also additional area which superficially may contain suitable habitat and also a portion of area which is modified by agriculture but may contain smaller pockets of overlooked suitable habitat. I chose to exclude most of the southern portion of the Winnipeg Ecodistrict because this area of unsuitable habitat would likely reduce model accuracy and the extensive crop agriculture has likely eliminated any remaining suitable habitat (Jiménez and Soberón 2020).

Climate influences of the study region include: The Lake Manitoba Plain Ecoregion has mean annual temperatures of 2 to 3 °C (mean summer 16 °C; Mean winter -12.5 °C) and mean annual precipitation of 450-700 mm. The Interlake Plain Ecoregion measures annual average temperature of 1 °C (mean summer 15.5 °C and mean winter -14.5 °C) and mean annual precipitation is between 425-575 mm) (ESWG (Ecological Stratification Working Group) 1995).

The southern Interlake is heavily modified by agricultural land use and generally transitions from aspen parkland in the south to boreal forest to the north. The south and east portions of the study area are made up of predominately flat agricultural cropland with smaller aspen and oak woodlots which steadily transitions northward to a mosaic of cattle pasture, hayland meadows, lowland sedge meadows, and wetland with aspen forest. This area then transitions to mixedwood forests, and finally to predominately conifer forest and peat bogs (Manitoba Remote Sensing Centre 2006). While tall grass prairie habitat historically comprised the grassland component of this mosaic, the majority of the prairie in the study area has either been converted to cropland, hayland, degraded by pasturing animals, or has been subject to the introduction of invasive plant species (Henderson and Koper 2014).

3.1.1 Modelling Approach

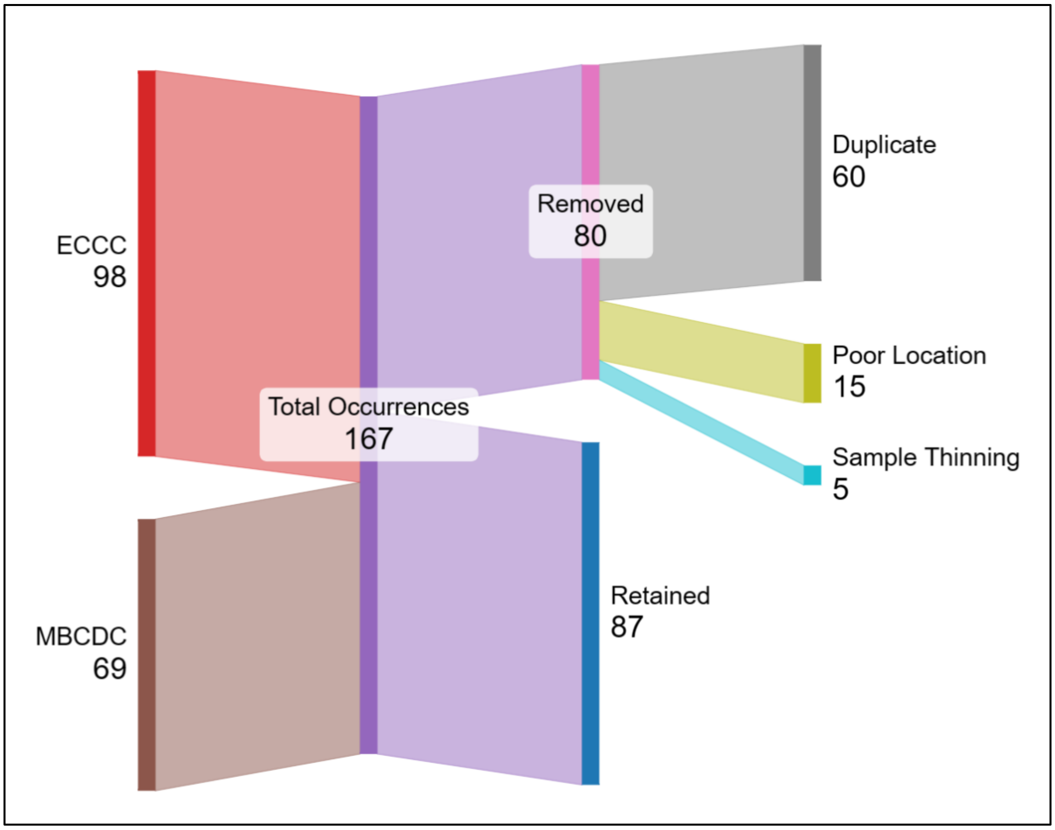
MaxEnt species distribution modelling software was chosen to model Dakota skipper distribution in Manitoba. Previous studies which evaluated various model types found that machine learning models such as MaxEnt, Boosted Regression Trees, and Random Forest outperformed non-machine learning models such as generalized additive models and multivariate adaptive regression splines (Brotons *et al.* 2004; Elith *et al.* 2006; Phillips *et al.* 2009). Machine learning models have also been found to perform better with presence-only data compared to non-machine learning modelling options (Elith *et al.* 2006; Phillips *et al.* 2009; Dearborn *et al.* 2022). Additionally, MaxEnt is user-friendly and is one of the most commonly employed tools in species distribution modelling research (Merow *et al.* 2013).

MaxEnt predicts species occurrences by creating a relative probability surface of occurrence for a species within a study area from a set of species presence point locations. It does this by determining the mean values of a chosen number of biotic and abiotic environmental variables of existing species locations to create a set of environmental constraints. The chosen environmental variables are transformed to so-called features or functions which best describe the variable data. The software then compares the set of environmental constraints to mean values found at a set of random background point locations taken across the study area and determines how well they agree with the constraints. The resulting model is validated with a retained set of species presence points. From all the model variants which satisfy the environmental constraints, MaxEnt selects the model of maximum entropy or the model which is least constrained. This tends to strike a balance between selecting an under-specified model which may be biased or lack predictive power and an over-specified model which may overfit the data leading to over-confident predictions and a model specific to only the input dataset. MaxEnt improves the discrimination of a set of randomly selected sites to the presence location constraints through an iterative approach. The selected model will then classify the entire study area creating the relative probability surface for a particular species (Phillips *et al.* 2006; Phillips and Dudík 2008).

3.2 Species Occurrence Selection

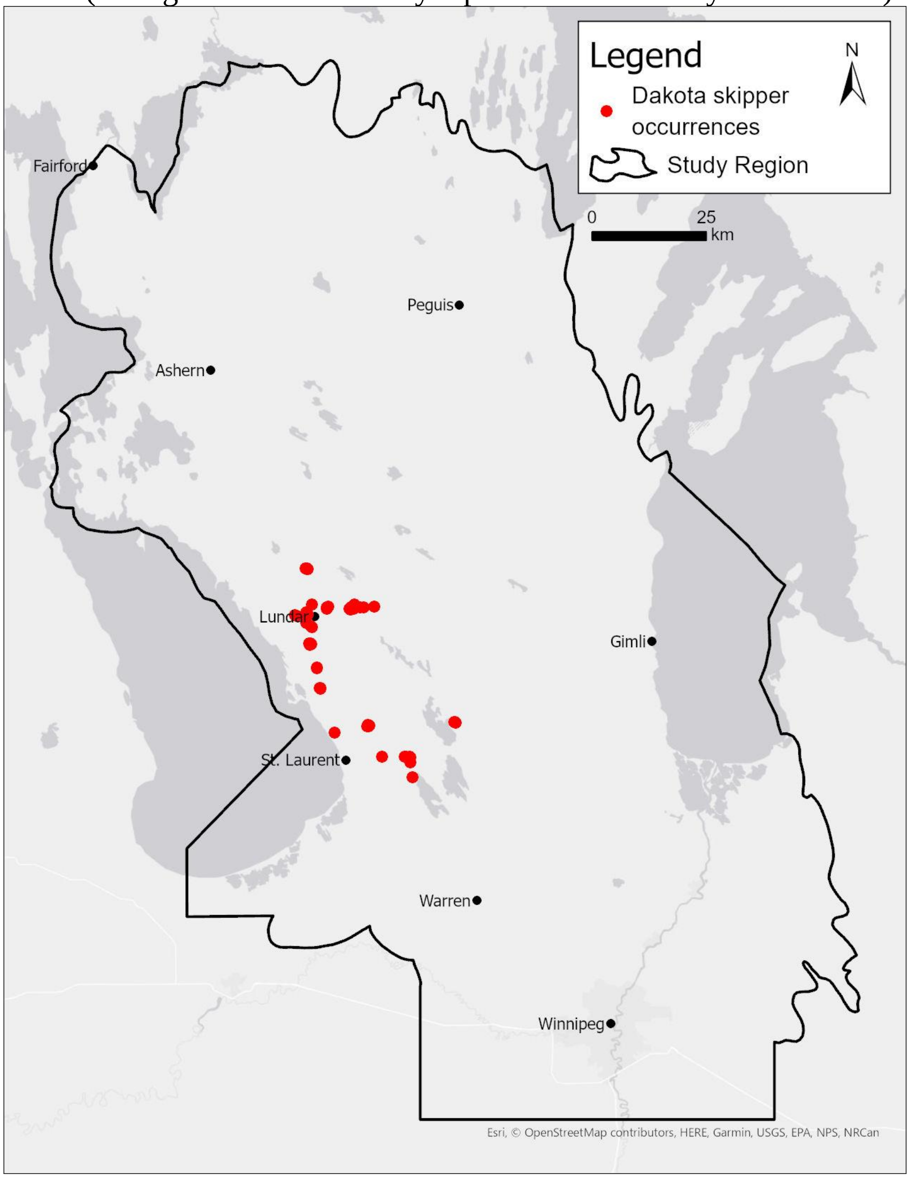
Dakota skipper occurrences were sourced from the Manitoba Conservation Data Centre and Environment and Climate Change Canada (Manitoba Conservation Data Centre 2018; Environment and Climate Change Canada 2021). These sources were chosen as they contained the best available current and historic occurrence information for Manitoba based on all known Dakota skipper surveys (Appendix 1. Dakota skipper survey references. Occurrences were compiled into geospatial vector point, polyline, and polygon feature representations using an internationally recognized data entry standard (NatureServe 2002). Point features describe low locational uncertainly (i.e. location captured using a GPS) and polyline and polygon features were used to describe observations with higher location uncertainty. For example, a buffer maybe applied to an observation with a poor location description or imprecise geographic coordinates to illustrate the uncertainty. Since point features are required for the analysis, the point feature occurrences were merged with the polyline and polygon features after first converting to centroid points. The locational uncertainty of the observations was retained in the attributes. Occurrences from both sources, generally located in the Interlake area were extracted regardless of year when observed (N=167 Year range: 1987 – 2020). Location and attribute information errors including taxonomic, geographic location, location description, and transcription (Veregin 1989; Hijmans *et al.* 1999) were searched for and corrected where possible. Figure 2 illustrates the selection process. Specifically, duplicate observations both within and between datasets were first removed. Then, observations with a locational uncertainty of 100 m or greater were removed unless the centroid or a location within the uncertainty buffer could be found in habitat that looked reasonably suitable when viewed using current satellite or air photos. Remaining observations were then removed if the habitat was obviously unsuitable. For example, the point was retained if located in hay land or native pasture but was not retained in a woodlot, cropland or if the area was obviously disturbed which would indicate a land use change since the observation. Visit notes supplied by surveyors were used when available to inform habitat type and quality. Sample thinning was then employed to reduce spatial autocorrelation and potential model overfit which can occur if observations are spatially close (Boria *et al.* 2014). The older and/or less locationally accurate point within 30 m of another observation was removed (Lischke *et al.* 1998; Boria *et al.* 2014; Fourcade *et al.* 2014). Based on this screening process, of the initial 167 Interlake occurrences, 87 Dakota skipper occurrences were accepted into the study for the Interlake (Year range: 1995 - 2020; Average: 2009; 40% from 2002, 23% from 2020, 6% 2018, and 31% from other years). Figure 3 shows that the general locations of the retained occurrences are concentrated in the centre and east of the study region, near the towns of Lundar and St. Laurent.

Figure 2. Dakota skipper occurrence location selection process, (Manitoba Conservation Data Centre 2018; Environment and Climate Change Canada 2021)



Produced using SankeyMATIC.com

Figure 3. General locations of Dakota skipper occurrences in the Interlake study region used in model construction (each general location may represent one to many occurrences)



3.3 Variable Creation and Selection

3.3.1 Common Model Variable Creation

I chose four geospatial layers commonly used in SDMs. Three of which were used in previous Dakota skipper SDMs and are listed in Table 2. Common model variables and displayed in Appendix 2. Common Model and Satellite Model variable maps including: Land use/Land cover, Soils, and the Shuttle Radar Topology Mission digital elevation model (SRTM DEM). The fourth layer, the Forest Resource Inventory (FRI) is also commonly used in SDMs where available. All four layers are freely available on Manitoba Government websites (Manitoba Land Initiative 2017; Manitoba Government 2022). Of these four layers, I only used the original Land use and SRTM DEM layers in the models. I did not use the FRI or Soils layers directly, however, from the FRI I created: Closest Forest Stand Area and Distance to Nearest Deciduous Forest. From the soils layer I created: Soil Ph, Soil Deposition Mode, and Soil Drainage (Table 2. Common model variables). Despite being the most current versions available, the Common Model variables are older compared to the oldest satellite imagery from 2016. The FRI is the oldest at 37 years with the most current being the land use layer at 10 years. Some portions of the soil survey layer are even older at 59 years. This supports the concern that Common Model variables lack timely updates.

Table 2. Common model variables

Variable Number	Variable Abbreviation	Variable Name	Layer Used	Resolution (m^2)	Software	Acquired	Source
1	Landuse	Land use	Provincial land use layer	30	ARC	2004 to 2006	(Manitoba Remote Sensing Centre 2006)
2	Closest_stand_area	Area of closest deciduous stand	Provincial FRI layer	30	ARC	1979 to 1992	(Manitoba Conservation 1996)
3	Dist_Deciduous	Distance to closest deciduous stand	Provincial FRI layer	30	ARC	1979 to 1992	(Manitoba Conservation 1996)
4	Soil_PH	Soil pH	Provincial soil laye	30	ARC	1953 to 1975	(Agriculture and Agri-Food Canada 2017)
5	Soi_mode_deposit	Soil deposition mode	Provincial soil layer	30	ARC	1953 to 1975	(Agriculture and Agri-Food Canada 2017)
6	Drainage	Soil drainage	Provincial soil layer	30	ARC	1953 to 1975	(Agriculture and Agri-Food Canada 2017)
7	DEM_SRTM30	SRTM DEM	SRTM DEM	30	ARC	2000	(Farr *et al.* 2007)

Land Use Variable

A land use raster mosaic covering the Interlake study area was created from combining portions of seven of the most current land use tiles covering various extents acquired from the Manitoba Remote Sensing Centre. The land use layers were created by the Centre from Landsat 7TM+ satellite imagery acquired from 2004 to 2006 with 30 m pixel resolution. Appendix 4 lists the Manitoba Land use classification system includes 18 coarse land cover/land use classes representing natural vegetation and anthropogenic phenomena (Manitoba Remote Sensing Centre 2006). The land use mosaic was added as a model variable "as is" with no secondary variables produced from the layer. An approximate 17 km^2 portion of the upper north-east corner of the study area had no available coverage representing 6.8% of the total study area.

Forest Resource Inventory (FRI)

I created a vector mosaic of the Manitoba FRI encompassing the study region using portions of the most recent coverages of nine Forest Management Units: 1, 2, 10, 23, 40, 41, 43, and 45. The FRI was created from interpreting air photos at a scale of 1:15,840 production date range between 1979 and 1992 (Manitoba Conservation 1996). Preliminary analysis suggested that known Dakota skipper occurrences are associated with adjacent deciduous forest (i.e. Trembling aspen) stands next to grassland. Therefore, to model this relationship, a deciduous forest layer, defined as containing less than 25% conifer trees, was created from the FRI and the model variables, area of closest deciduous forest stand, and distance to closest deciduous forest stand were generated. To create the deciduous forest layer, I initially considered all FRI habitat classes including Forested and Non-Forested Land. Non-Forested Land was removed from the study even though some habitat types contained forest, tree species were not provided. The FRI further distinguishes Forested Land as either "Productive Forested Land" from "Non-Productive Forested Land". I only selected deciduous forest from Productive Forested Land since Non-Productive Forested Land habitat does not support deciduous trees or the deciduous habitat type was not considered applicable i.e. "Hardwood Treed Rock class." (Manitoba Conservation 1996).

Soil variables

Soil influences water and nutrient availability for plants and provides variable substrates for growth. I therefore wanted to represent soil variables in the SDM.

A Manitoba Detailed Soil Survey, freely available from Agriculture and Agri-Foods Canada website (Agriculture and Agri-Food Canada 2017) was selected for the study. The polygons were digitized at 1:100 000 scale from various soil surveys conducted in Manitoba to create a comprehensive provincial soil layer. Just under half of the soil polygons are from detailed soil surveys with a 1:20,000 scale where the other polygons are soil reconnaissance surveys with scales at 1:100,000, 1:250,000, and 1:126,720. The detailed soil polygons are primarily in the lower southern portion of the study area particularly in the southeast. The study area contains four large area soil surveys ranging in year and scale of 1953 to 1975 and 1:100,000 to 1:126,720 respectively (Agriculture and Agri-Food Canada 2017).

The attributes soil pH, drainage, and mode of soil deposition were selected from the soil attribute table for use in the model and the layer clipped to the study area.

Digital Elevation Model variable

Elevation is thought to influence water drainage and retention on the landscape including floristic composition, soil development, and solar radiation absorption (Brown 1994; Rich and Fu 2000; Farr *et al.* 2007; Elith and Leathwick 2009). For these reasons elevation was chosen as a model variable.

I chose the Shuttle Radar Topography Mission (SRTM) Version 3-1. arcsecond data product captured in 2000 during a NASA Space Shuttle mission, which covered between 60N and 60S latitude with a horizontal spatial resolution of approximately 30 m at the equator (Farr *et al.* 2007) and is freely available as 1 x 1 arcsecond of latitude and longitude tiles or approximately 70 x 110 km tiles with 27 m horizontal resolution in the study area (study area mean elevation = 264 m; range: 212 m – 313 m) (United States Geological Survey 2015). The individual tiles were merged then clipped to the study area. The SRTM DEM was used because it had the best available resolution and coverage for the study region. Two other DEMs created for producing digital orthophotos were not used because both had limited coverage for only southern Manitoba. The older DEM used for a 1998 ortho imagery series had 100 m cell resolution while a more recent ortho refresh, ending in 2014, produced an inferior DEM (Murray 2013 unpublished data). The SRTM DEM may better be classified as a digital surface model (DSM) because it will tend to overestimate ground elevation because vegetation and other features will differentially absorb the radar pulse then reflecting a portion of the beam back to the sensor before it reaches the ground. Research has been done to correct for this overestimate by determining to what extent various vegetation types absorb the radar beam (Carabajal and Harding 2006). However, the techniques were either too general (Zhao *et al.* 2018) or were specific to other habitats (Su *et al.* 2015).

3.3.2 Satellite Model Imagery Sources

Sentinel 2 remotely sensed satellite imagery and Shuttle Radar Topology Mission (refer to the DEM section under Common Model Variable Creation) imagery were the two sources for creating candidate variables for the Satellite model (See above section for details about the SRTM DEM).

At the time of the study, Sentinel 2A multi-spectral imager (MSI) was one of a proposed pair of satellites (Sentinel 2A and Sentinel 2B) operated by the European Space Agency. Sentinel 2A became operational in 2016 and, using a multi-spectral imager (MSI) "push broom" sensor, acquires 110 x 110 km^2 tiles of 13 bands (4 10 m, 6 20 m and 3 60 m resolution) with, at the time, a ten-day return period between latitudes 56S and 83N. It was selected because the imagery is freely available, provides extensive coverage, a short return period, has 13 image bands, and relatively high spatial resolution.

Easy access, province-wide coverage, and short return window makes it more likely that suitable cloud-free imagery can be acquired for a large study area and allows for image processing workflows to be easily transferred to creating SDMs to other areas of the province. Large 110 x 110 km^2 tiles reduce the number of tiles for a study area reducing potential edge issues and 13 bands allows the creation of numerous indices. The more commonly used visible and near infrared bands have 10 m pixel resolution and the other 20 m and 60 m bands can be resampled to 10 m.

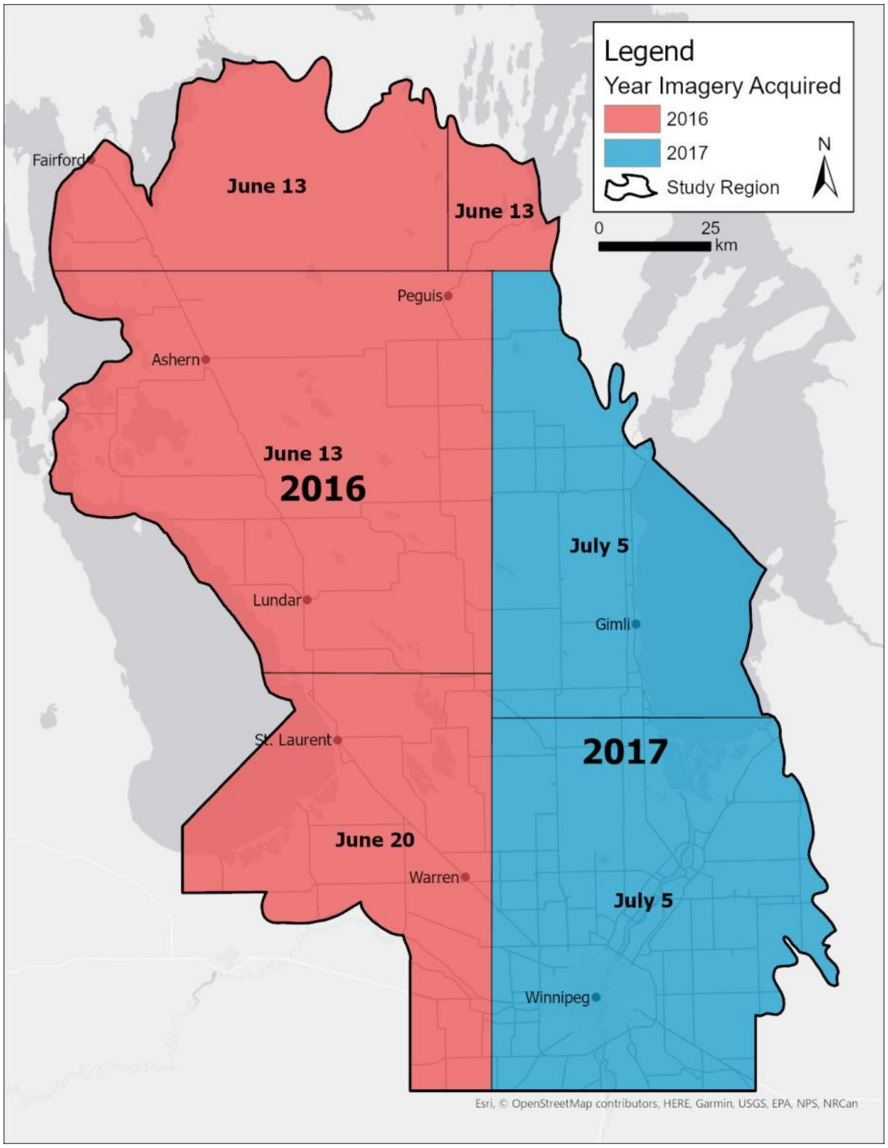
Higher resolution makes this imagery ideal for the Interlake study area where the mosaic of Aspen forest and grass meadows can be closely spaced and using higher resolution imagers can reduce "pixel bleed" which can lead to spurious interpretation by the modelling software algorithm.

3.3.3 Satellite Model Imagery Processing

Sentinel 2 tiles covering the study area were downloaded (European Space Agency 2023) which contained minimal cloud cover and acquired during the grassland growing season, between mid-June and mid-August.

At the time of the study approximately two years of imagery was acquired due to the recent launch of the sensor. I sought to have complete coverage for the 2016 season, the earliest imagery available. However, two tiles from 2017 were substituted due to extensive cloud cover in the matching 2016 tiles. In all, eight tiles completed coverage of the study area consisting of six acquired in mid-June of 2016 and two acquired in early July of 2017. Figure 4 shows the spatial extent of imagery acquired for both years the date individual tiles were acquired. The 2016 acquisition was in the west and north portion of the study region and the 2017 acquisition was in the eastern and more of the south portion. The one-year difference in acquisition time was less of a concern than ensuring close season dates. There was also likely negligible land use change during this short time period.

Figure 4. Extent of Sentinel 2 imagery for the study region. Information acquired in 2016 (red area) and 2017 (blue area) and the spatial extent and date individual tiles (delineated by thin black lines) were acquired.



ESA SNAP software (European Space Agency 2017) designed for Sentinel satellite imagery was used for viewing and processing the imagery. Since image processing was generally computer intensive I distributed work tasks across four virtual computers on Microsoft Azure cloud services (Microsoft Corporation 2018). Imagery is generally offered pre-processed to level “1C” which provides top of atmosphere (TOA) reflectance. Recognizing that atmospheric distortions may negatively affect modelling, I used the SNAP “sen2cor” plug-in (Mueller-Wilm *et al.* 2018) to obtain level “2A” or bottom of atmosphere (BOA) reflectance imagery. Level 2A BOA was used to create model variables when possible however some bands are dropped during the 1C to 2A processing which are needed to produce some indices which necessitated using Level 1C.

The bands for the 1C TOA and 2A BOA tiles were resampled to 10 m resolution using the SNAP resample tool offered with the main software package designated “simple resolution”. The 2A tiles were also resampled using the SNAP “super resolution plug-in “sen2res” (Brodu 2018; European Space Agency 2018). This plug-in produces a more sophisticated 10 m resolution product by using the existing 10 m bands to determine how best to unmix the 20 m and 60 m pixels. The process is loosely analogous to using a higher resolution panchromatic image (e.g. the Landsat panchromatic band) to pan-sharpen lower resolution bands. Three separate 10 m resolution mosaics were then created by merging the differently processed tiles: Level 1C TOA 10 m Simple Resolution, Level 2A BOA 10 m Simple Resolution, Level 2A BOA 10 m Simple Resolution, and Level 2A TOA Super Resolution.

I attempted a cloud free Level 2A BOA 10 m resolution mosaic for 2016 by creating a synthetic image using the SNAP sen2three processor (European Space Agency 2017, 2018). This processor takes a time-series stack of tiles, in this case all from the 2016 acquisition window, and systematically replaces “bad” pixels, such as cloud, with “good” pixels from another overlapping image acquired at a different time. Processing was computer and time intensive and upon visual inspection revealed that the processor replaced shadow as no-data pixels. This was particularly pronounced for cloud shadow. These errors made this imagery product unsuitable for the analysis and therefore was not used in the model.

SRTM Version 3 1 arc-second tiles of the study area were obtained from EarthExplorer (United States Geological Survey 2018). ArcGIS Desktop (ESRI (Environmental Systems Research Institute) 2018) was used to merge the tiles to make a 30 m resolution mosaic. The commonly used DEM interpolator, Inverse Distance Weighted (IDW) was applied to resample from 30 m to 10 m pixel resolution to match the resolution of the Sentinel 2 imagery. The resulting 30 m and 10 m SRTM DEMs mosaics were then clipped, maintaining a one-kilometer buffer from the study area boundary to reduce interpolation errors close to the boundary edge.

In all, five mosaics were created (Table 3. Imagery mosaics created from Sentinel 2A and SRTM tiles used to create model variables.**Error! Reference source not found.**) which were then used to produce the imagery only model variables.

3.3.4 Variable Processing and Creation

Variables derived from the Sentinel 2A imagery were created from functions using various band combinations from the same mosaic. The SRTM DEM was used with specific geoprocessing formulas to generate the other DEM related variables. In all, forty (40) candidate variables were created in total. Twenty-two (22) from the three Sentinel 2A mosaics (Level 1C TOA 10 m Simple Resolution, Level 2A BOA 10 m Simple Resolution, and Level 2A BOA Super Resolution). Ten variables were created from the two SRTM mosaics, SRTMDEMRes30m and SRTMDEMRes10m. Eight additional variables were created from the NDVI8a candidate variable which was initially created from the Level 2A SuperRes10m mosaic. These 40 candidate variables were then grouped into seven classes based on variable similarity (Appendix 5. .

Table 3. Imagery mosaics created from Sentinel 2A and SRTM tiles used to create model variables.

Remote Sensing Platform	Input imagery Tiles	Initial Resolution (m^2)	Output Mosaic	Bands	Final Resolution (m^2)	Processor
Sentinel 2A	2A (BOA)	10m to 60m	Level 2A SuperRes10m		10	SNAP super resolution resampling
Sentinel 2A	1C (TOA)	10m to 60m	Level 1C TOA Simple Res10m		10	SNAP regular resampling
Sentinel 2A	2A (BOA)	10m to 60m	Level 2A BOA Simple Res10m		10	SNAP regular resampling
SRTM	SRTM	30m	SRTMDEMRes30m		30	NA
SRTM	SRTM	30m	SRTMDEMRes10m		10	ArcGIS IDW geoprocessor

3.3.5 Variables Derived from Sentinel Imagery

Vegetation Group

Eight vegetation indices were generated from the Sentinel 2 super resolution mosaic using the SNAP software (European Space Agency 2017). Three index formulas offered by the SNAP software were used to generate six vegetation indices. Each index was developed twice using the red and the red edge band respectively. The red band was simply swapped with the red edge band. These three indices include the common Normalized Difference Vegetation Index (NDVI) (Rouse *et al.* 1974; Lange *et al.* 2017) as well as two modified indices which improve on the original NDVI. The Atmospherically Resistant Vegetation Index (ARVI) (Huete and Jackson 1988; Kaufman and Tanre 1992) which reduces atmospheric noise the NDVI is prone toward and the Modified Soil Adjusted Vegetation Index (MSAVI) which reduces the influence of bare soil where vegetation is sparse (Huete 1988; Huete and Jackson 1988; Qi *et al.* 1994).

Nonlinear Vegetation Index (NVI) applies a nonlinear relationship between the index and vegetation greenness instead of the linear relationship of other vegetation indices. The NVI tends to reduce saturation of dense canopies which is problematic with the NDVI without adding bands or constants to the equation. A simple NVI used in this study squares the near infra-red (NIR) band in the index formula (Goel and Qin 1994; Feng *et al.* 2019).

Enhanced Vegetation Index (EVI) is similar to the NDVI however it corrects for atmospheric and background canopy noise to which the NDVI is susceptible to. The EVI is also more sensitive in lush vegetation which reduces oversaturation of greenness which can be problematic for the NDVI (Huete *et al.* 1994; United States Geological Survey 2018).

Biophysical Group

The biophysical group consisted of three candidate variables: Leaf Area Index (LAI), Fraction of Absorbed Photosynthetically Active Radiation (FAPAR), and Fraction of Vegetation Cover (FVC). These variables are used to describe landscape biophysical properties for example, carbon, nutrient, hydrological, and energy cycles. They quantify vegetation cover, land degradation, and erosion (Liang and Wang 2019) and are therefore considered promising candidates for the SDM to differentiate vegetation and land use in the study area. The variables were developed using the Sentinel 2 simple sample 10 m resolution mosaic in SNAP (Weiss and Baret 2016; European Space Agency 2017).

The LAI attempts to measure top-down two-dimensional leaf area coverage in a single pixel. This index is used to measure biophysical processes such as carbon and nutrient cycles and productivity. (Goel and Qin 1994; Bréda 2003; Weiss and Baret 2016; Feng *et al.* 2019).

FAPAR measures the portion of photosynthetically active radiation generally within the visible light spectrum range absorbed by vegetation and is generated from a vegetation index. FPAR is used to quantify various biophysical processes related to energy transfer, climate change, and hydrology cycle (Goward and Huemmrich 1992; Liu *et al.* 2015; Weiss and Baret 2016)

FVC measures the amount of alive vegetation cover in a pixel. It is derived from a vegetation index (NDVI) and is used to measure land degradation, erosion, and desertification (Purevdorj *et al.* 1998; Weiss and Baret 2016).

Brightness

Surface Albedo measures the total radiative reflectance of a surface from the total incident solar radiation. It influences surface energy budget which in turn affects heat exchange (temperature) and hydrology which affects and is affected by biophysical processes (Zhang *et al.* 2022). Efforts have been made to model fine scale Albedo measurements using moderately high resolution satellite imagery like Landsat and Sentinel 2 (Lin *et al.* 2022). I used a simple Albedo formula which sums pixel values across each band (Lahoz-Monfort *et al.* 2010) compare to (Naegeli *et al.* 2017). The resulting greyscale image shows fine scale albedo heterogeneity which may be useful when applied to an SDM for the above reasons.

Brightness is related to soil indices and is used to modify vegetation indices to reduce soil background influences (Huete *et al.* 1985; Huete and Jackson 1988). Here I used two simple Brightness indices generated from the green, red, and near-infrared bands. Lighter shades indicate dry exposed soils and salt evaporites and more darker values tend to indicate increase in healthy vegetation and soil moisture (Escadafal 1989; Escadafal *et al.* 1989)

Water Indices

Normalized Difference Water Index 2 (NDWI2) (McFeeters 1996) detects surface water using the difference in reflectance between the green and near-infrared bands. Infrared radiation is absorbed while green is reflected. Raster cell values greater than 0 are more likely surface water where <0 values are more likely dry land. NDWI2 originated from modifying the Normalized Difference Water Index (NDWI) which is used to measure water moisture in vegetation (Gao 1996). The green and near-infrared reflective characteristics of urban buildup are similar to that of water which causes the NDWI2 to misclassify urban areas as surface water. The Modified Normalized Difference Water Index (MNDWI) (Du *et al.* 2016) attempts to improve on the NDWI2 by swapping the near-infrared band with a mid-infrared band which is not as absorptive to urban buildup. In testing, the MNDWI tends to reduce misclassification of urban areas (Du *et al.* 2016; Murray 2016) however the NDWI2 tended to overall sense surface water more effectively (Murray 2016). The MNDWI use of a mid-infrared band increases the pixel cell size to 30 m which needs to be resampled down to 10 m to be used with the green Sentinel 2A band. The resampling may result in a loss of information or incorrect 10 m pixel classification.

Both the NDWI2 and the MNDWI were selected as candidate variables since they are relatively easy to generate and would inform the model for areas of surface water and therefore likely unsuitable habitat. These indices would also inversely identify drier upland areas which are associated with native tallgrass prairie and Dakota skipper occurrences but are generally a few meters elevated above the more flood-prone lowland areas and so are not easily detected with current DEM layers. The two variables were generated from the Sentinel 2A SuperRes10m mosaic in SNAP using the green (band 3), near infra-red (8), and shortwave infrared (band 11).

Tasselled Cap

Tasselled Cap is a relatively older analysis which produces three products: Brightness, greenness which is related to vegetation, and blueness or wetness related to water. High values for all three indices indicate brighter surface features, denser or more vigorously growing vegetation, and wetter areas (Kauth and Thomas 1976).

Brightness, Vegetation, and Wetness candidate variables were generated from a 6 band and a 13 band mosaics respectively, for a total of six Tasselled Cap candidate variables. SNAP was used to generate the 6 band and 13 band mosaics from the Sentinel 2A C1_simpsamp10m res mosaic and Tasselled Cap coefficients specific to the Sentinel 2A sensor (Shi and Xu 2019) were used to construct the raster formulas needed to make the six candidate variables.

Tasselled Cap Brightness, Vegetation, and Wetness offer characterization of local biotic and abiotic site conditions and therefore were selected as candidate variables. The six variables were included in the Brightness group, Vegetation group, and Water/Moisture group respectively.

NDVI8a GLCM variables.

I chose Grey Level Co-occurrence Matrix (GLCM) texture analysis as it is an established texture analysis technique and is shown to generally improve image classification (Hall-Beyer 2017, 2017). The analysis also produces image products which could indirectly characterize favourable Dakota skipper environmental conditions based on texture, thereby better informing the model.

I used R GLCM package (Zvoleff 2019) to generate eight GLCM candidate variables from the NDVI8a which itself is a candidate variable and which was produced from the Superres 10 m^2 mosaic. Other indices and imagery could also have been used to generate their corresponding GLCM variables however this would have been labour and time prohibitive therefore only one index was analyzed with GLCM. The NDVI8a was chosen over the other variables because vegetation is an important indicator of Dakota skipper habitat and is a well-known vegetation index. The Sentinel 2A "red-edge" or Band 8a was used because it may provide more information about vegetation health or type to the index than other infra-red bands used in other NDVIs since the Band 8a spectrum range more closely matches the reflectance wavelength of plant chlorophyl.

No *a priori* selection of texture variables was done since I assumed that all were favourable to inform the model and any subsequent correlation would be dealt with during variable selection. The eight variables are listed in Appendix 5. Model candidate variables created from Sentinel 2A imagery and SRTM DEM. and include: Contrast, Dissimilarity, Homogeneity, Second Momentum, Entropy, Mean, Variance, and Correlation.

GLCM requires consideration of many parameters before analysis can begin (Hall-Beyer 2017, 2017). These include choosing: The GLCM texture analysis to perform, the imagery band to analyze, the size of the moving window applied to the image, and the offset direction and distance from the reference pixel to the neighbour pixel. I chose parameters which I thought would aid the model. I selected the generally applied 5-bit image rescaling or 32 grey levels for the analysis. Rescaling to less grey levels is used to reduce the number of potential pixel pairs and so reduce the number of zero pair matches in a window. This increases statistical validity. Thirty-two (32) or 5 bit or 4 bit are generally used in software and so 32 was selected to keep the image relatively complex. A 9 x 9 (90 x 90 m^2) window was a compromise between a smaller window being more homogeneous and a larger window having too much heterogeneity. Visual inspection suggested that a 9 x 9 window would be suitable to detect local variability where the heterogeneity of the hill and swale mosaic can be experienced over relatively short distances from upland treed areas to grassland/hayland/pasture to moist sedge meadows to wetland and then open water. This window may also accurately model Dakota skipper adult behavior where local heterogeneity is more important than lower scale landscape variability. The (4, 4) pixel offset selected the partnering pixel four pixels to the northeast from the reference pixel. This shift moved the neighbour pixel away from the reference a bit so there might be some difference as would also increase variability due to the general physiography of the natural north-west to south-east oriented narrow bands of “hill and swale” patterning indicative to much of the study area.

Variables Derived from SRTM 3-1 Arcsecond Imagery

Elevation is thought to influence water drainage and retention on the landscape including floristic composition, soil development, and solar radiation absorption (Brown 1994; Fu and Rich 2000; Rich and Fu 2000; Farr *et al.* 2007). For these reasons elevation and nine derived products were included as candidate variables generated from the two SRTM 3 mosaics.

The SRTM 3 10 m resolution mosaic was included as a candidate variable and also used to generate five other candidate variables using elevation surface analysis tools in ArcGIS (ESRI (Environmental Systems Research Institute) 2018). They included: slope, aspect and three curvature layers. Slope measures the steepness of each raster cell in the SRTM 3 mosaic. The higher the slope value the steeper the slope. Aspect measures the direction of slope measured as a bearing from 0 to 359 degrees. Curvature describes the shape of the slope. Positive values indicate the slope surface is convex or curved upward where a negative value indicates that the surface is concave. Parallel curvature describes the curvature parallel to the direction of slope and perpendicular curvature models the curvature perpendicular to the direction of slope (Moore *et al.* 1991). Slope accuracy was estimated to be within 10% of actuals measured along the Amazon river (Hendricks and Alsdorf 2004; LeFavour and Alsdorf 2005; Farr *et al.* 2007).

Solar insolation affects Dakota skipper emergence and facilitates flight period activities (Dearborn and Westwood 2014). To model this requirement I generated four solar insolation candidate variables using ArcGIS Solar Radiation toolset (Fu and Rich 2000; Rich and Fu 2000; ESRI (Environmental Systems Research Institute) 2018).

I used the SRTM 3 30m resolution mosaic to model total solar radiation instead of the SRTM3 10 m resolution mosaic to remove any error generated from the IDW resampling from 30 m to 10 m propagating to the solar radiation variables. Solar radiation is quantified in watt hours/meter squared for each mosaic raster cell calculated across the study area. Total radiation includes: Direct radiation from the sun, reflected radiation from objects, and diffuse radiation which is direct radiation scattered by the atmospheric water vapour (clouds) and dust.

Total radiation was modelled for the four solar insolation variables differing only by temporal period. The variable Annual Radiation modelled total radiation for the entire 12 month year, summer radiation variable spanned six summer months (May to October), winter radiation for five months, from November to March, and flight time radiation variable included the months of June and July.

Since the study area spans more than one degree latitude, it was necessary to divide the study area into south and north regions, generate the four insolation layers separately for each region, then merge (Fu and Rich 2000; ESRI (Environmental Systems Research Institute) 2018). The four insolation variables were then resampled using IDW interpolation to 10 m^2 resolution and clipped to the study area to remove the 1km buffer.

3.3.6 Variable Selection

To correct for model overfit in MaxEnt, one variable is usually included for every 10 occurrence points (Harrell Jr. *et al.* 1996; Guisan and Zimmermann 2000). My goal was to run a maximum of seven variables to retain a sufficient portion of sample points (n=17) to cross-validate the model (Westwood *et al.* 2019). To achieve this, I performed correlation analysis on the 40 candidate variables and subsequentially removed candidate variables which were highly correlated (between -0.7 and 0.7) within group or between groups (Elith *et al.* 2006; Dearborn *et al.* 2022). Following Lahoz-Monfort *et al.* (2010) then ran the remaining eleven variables together in MaxEnt and subsequentially removed the variable which contributed the least to the model. I then ran the remaining variables and once again removed the least contributing variable. I continued to do this variable attrition until model performance, measured by AUC, was worse than the previous model. I then chose the five variables and model from the second last run as the representative for the Satellite Model (Table 4. Satellite Model predictor variables final selection and variable maps in Appendix 2. Common Model and Satellite Model variable maps.

Table 4. Satellite Model predictor variables final selection

Variable Number	Variable Abbreviation	Variable Name	Remote Sensing Platform	Imagery Used	Software	Source
1	TcapWET06	Tasselled Cap Wetness 6 Bands	Sentinel 2	C1_simpsamp10mres	SNAP	(Shi and Xu 2019)
2	ARVI8a	Atmospherically resistant	Sentinel 2	SuperRes10m	SNAP	(Kaufman and

		Vegetation Index				Tanre 1992)
3	TcapGVI06	Tasselled Cap Greeness or vegetation 6 Bands	Sentinel 2	C1_simpsamp10mres	SNAP	(Shi and Xu 2019)
4	glcm_ent	GLCM Entropy	Sentinel 2	NDVI8a	R	(Zvoleff 2019)
5	SRTM DEM	SRTM DEM	STRMDEMRes10m	STRMDEMRes10m	ArcGIS	(Farr *et al.* 2007)

3.4 Model Selection

3.4.1 Model Thresholding

Thresholding classifies the continuous relative probability model output into two or more species suitability classes (e.g. Suitable/Unsuitable; High, Medium, Low). It generally aids in the interpretation of model results by clearly depicting areas as either: suitable or unsuitable, or indicating the species is present or absent which can make the model more useful in research and conservation applications than presenting the continuous relative likelihood layer. However, thresholding can lose model information, potentially introducing modeller bias and arbitrary non-biologically meaningful categories (Liu *et al.* 2005; Jiménez-Valverde and Lobo 2007; Merow *et al.* 2014; Guillera-Arroita *et al.* 2015).

Merow *et al.* (2014) is highly critical of using thresholds and recommends to only use the original continuous model relative probability or likelihood outputs. Liu *et al.* (2005) and Jiménez-Valverde and Lobo (2007) are also critical of some threshold techniques but have attempted to find more robust thresholding procedures. Guillera-Arroita *et al.* (2015) conclude that thresholding is rarely needed and detrimental for most modelling purposes. However, if necessary, thresholds and the procedure should be selected based on conservation objectives.

There are different techniques to threshold model likelihood surface outputs including assigning arbitrary breaks, using a threshold algorithm (Liu *et al.* 2005; Jiménez-Valverde and Lobo 2007; Merow *et al.* 2014) and classifying from *a priori* training areas from computer based techniques (Dearborn *et al.* 2022). Regardless of thresholding method, the objective should be to limit commission and omission errors so that the chosen threshold discriminates "good" from "bad" as accurately as possible (Jiménez-Valverde and Lobo 2007).

Following these broad guidelines, I chose to use the "Equal test of sensitivity and specificity" algorithm. This test finds the threshold where the difference between sensitivity and specificity is minimum which would presumably minimize commission/omission errors. This threshold is shown to perform relatively well compared to other thresholding methods (Liu *et al.* 2005; Jiménez-Valverde and Lobo 2007; Phillips and Dudík 2008) and is an option available in the MaxEnt program. The thresholds produced from this test seemed rather low (Common = 0.32, Satellite = 0.23) out of a relative likelihood range from 0 to 1. Therefore, a higher threshold of 0.80 for both models was also added to test model performance at a very high probability class (Table 5) despite possibly higher risk of commission (false-positive) errors.

Table 5. Suitability category thresholds for the Common and Satellite models.

Suitability Category	Common Model	Satellite Model
Low	0 – 0.25	0 – 0.32
Medium	>0.25 – 0.80	>0.32 – 0.80
High	>0.80 - 1	>0.80 - 1

3.4.2 Field Validation

Site Selection

Candidate point sites (n=240) for field validation were randomly generated from within the study area stratified by 120 sites per model and 40 sites for each of the three model probability threshold classes (see Model Thresholding section). Point sites were the centroid location of a pixel square either the 30 m^2 pixel for the Common Model or 10 m^2 for the Satellite Model. Forty (40) sites per category ensured that 20 sites could be achieved to adequately perform statistical analysis while accounting for likely accessibility issues (I.e. landowner refusal/unable to contact or inaccessible due to terrain). I was blind to the condition or habitat suitability of a site before each location visit, only that the models predicted the locations as one of three (Low, Medium, High) threshold categories (blind sampling procedure).

Site Evaluation

I then evaluated the location for suitability to support Dakota skipper to gauge the accuracy of the model prediction. To evaluate the suitability of a site I used a list of indicator plant species (developed by Rigney 2013; Dearborn *et al.* 2022) that are indicative of high quality habitat for Dakota skipper and a series of physical and environmental factors present at the location (protocol also in Dearborn *et al.* 2022) to evaluate applicability to one of the three probability classes described above. A detailed indicator plant list and scores are listed in Appendix 6. Dakota skippers were not surveyed for at sites since detection or non-detection would not provide enough information to evaluate model performance. The data would likely be unreliable given its brief flight period is variable by time of emergence and duration and may also still go undetected at a site given its cryptic nature and circumstances during a survey or the entire field season. Instead, plant suitability scores measure habitat information at a site and allow ranking sites by quality to support Dakota skippers which can be compared to the relative likelihood predictive layers generated by the models.

Site Survey Design

Site evaluation consisted of employing either a detailed or rapid assessment protocol at each point location (For more information about sample protocol refer to Appendix 7). Detailed assessments were conducted when initial impression indicated that a site had high native floristic diversity and was not severely disturbed by cattle or not agriculture cropland. Detailed and rapid assessments were modified between Common and Satellite approaches due to the difference in pixel size. The intent was to establish a smaller quadrat square within the larger pixel to mitigate influence from the adjacent pixels when sampling for indicator plant species (Appendix 7 – Figure 6)

3.4.3 Model Validation

I evaluated the environmental variables for each model by how much they contributed to model performance. This was done in two ways: By reviewing the percent contribution tables generated as model results output and also by the model output graphs which showed how each variable affected (increased or decreased) model gain when the variable was isolated or omitted from the other model variables when run.

To evaluate individual model performance, and also to compare models, I rank compared the AUC values and also generated confusion matrixes for each model based on the threshold values assigned to each model to describe classification errors of omission and commission. Generally, a confusion matrix will display how two classifiers or raters classify cases or observations into the respective classes. One classifier is usually considered the reference that the other classifier, the candidate, is evaluated against. In this study the ground survey and the model are the reference and predictor classifiers respectively. The ground survey plant scores and the model relative likelihood values for each site are classified into the Low, Medium, and High prediction classes of the confusion matrix. If all sites are in the diagonal table cells then the two classifiers are in perfect agreement meaning the candidate classifier performs as well as the reference standard. Errors of omission occur when the candidate incorrectly classifies reference sites and errors of commission occur when the candidate incorrectly classifies candidate sites.

In addition, I used the confusion matrices to calculate Cohen's Kappa (Cohen 1960) k-statistic for each model which provided a means to rank compare one model against the other and describe overall model performance. Cohen's Kappa measures the proportion of agreement between two evaluators after chance agreement is removed, also referred to as a chance corrected measure of agreement. In this study the two evaluators are the model relative likelihood surfaces and the corresponding ground survey plant indicator scores thresholded and added to a confusion matrix. Kappa ranges from 0 to +1. A value of 0 indicates rater agreement was by chance with increasing values from zero meaning less of the agreement was due to chance with +1 meaning perfect agreement (Rosenfield and Fitzpatrick-Lins 1986).

Kappa values from 0-0.2 is considered slight agreement between the classifiers while 0.2-0.4 indicates fair agreement, 0.4-0.6 moderate agreement, 0.6 and above indicating substantial agreement. Other Kappa indexes have been developed to interpret the strength of agreement between candidate and reference classifier, however they have been criticized for being somewhat arbitrary (Landis and Koch 1977). However, Kappa can be used to rank two or more models based on their relative values.

Due to the inherent issues with thresholding we also chose to use Pearson's correlations coefficients to compare the relative likelihood surface pixel values at sample locations to the corresponding sample site vegetation suitability scores (Westwood *et al.* 2019; Dearborn *et al.* 2022). Testing continuous data instead of arbitrarily classifying results into present-absent, high-low, good-poor classes is shown to produce more accurate evaluations. Pearson's correlation coefficient measures degree of agreement and provides a statistic to measure individual model performance and also rank order comparisons similar to Kappa (Elith *et al.* 2006; Phillips and Dudík 2008). A paired z-test was also applied to measure statistical difference between the two models if any differences exist.

Statistical analyses were performed using R Statistical Software v4.1.2 (R Core Team 2021). The paired z-test on the Pearson correlations was calculated using the paired.r function in the psych R package (Revelle 2023). The Pearson correlations and Cohen's Kappa were calculated using cor.test and kappa2 functions respectively in the irr R package (Gamer *et al.* 2019). Geospatial data management and analysis was performed using SNAP (European Space Agency 2017) and ArcGIS Desktop (ESRI (Environmental Systems Research Institute) 2018). Models were created using MaxEnt (Phillips *et al.* 2006).

CHAPTER 4: RESULTS

4.1 Generation of Relative Likelihood Maps by Threshold Category

Table 6 shows the percent of total area covered by the three relative likelihood classes calculated from the continuous relative likelihood surfaces produced from the Common Model and Satellite Model. The two modelling approaches differ in area assigned between the Medium and Low suitability categories of less than 5 percent, while each model has a similar area in the High category: 1.1 and 1.3 percentages, respectively. However, the distribution of suitable habitat (High likelihood category) in the Common Model is quite concentrated in the west central part of the study area while the distribution of suitable habitat in the Satellite Model is dispersed across the west and central part of the study area (Figure 4). The Common Model locates most of the suitable habitat to the northwest, of the Shoal Lakes, whereas the Satellite Model shows suitable spread over two prominent strips oriented southeast to northwest mostly in the central region of the study area.

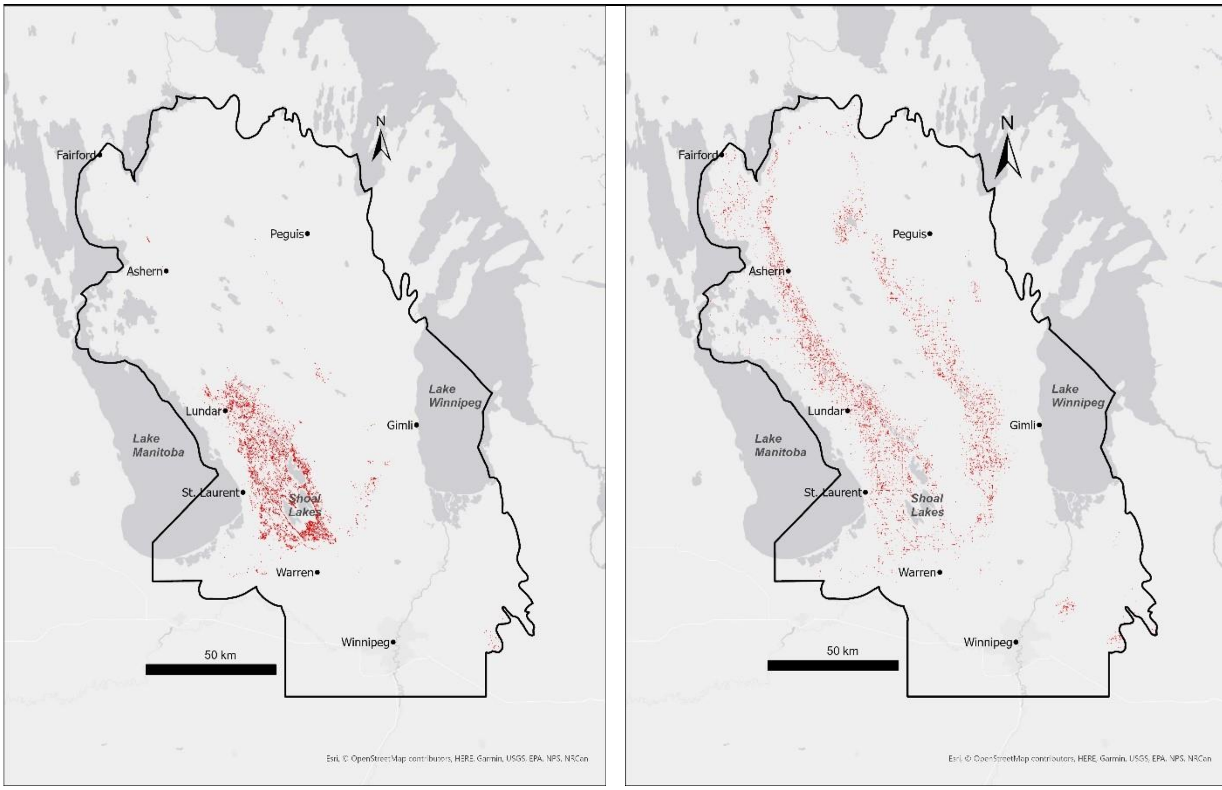
Table 6. Percent of total area and area (km sq) for the Common Model and Satellite Model by Relative Likelihood Category.

Relative Likelihood Category	Common Model	Satellite Model
High	1.1 (227)	1.3 (320)
Medium	6.4 (1591)	9.3 (2291)
Low	92.5 (22926)	89.4 (22126)
Total Area (km sq)	24795	24737

Common Model relative likelihood categories (High: 1 to 0.8; Medium: <0.8 to 0.25; Low: <0.25 to 0). Satellite Model: relative likelihood categories (High: 1 to 0.8; Medium: <0.8 to 0.32; Low: <0.32 to 0).

Figure 5. Distribution of High relative likelihood category in the Common and Satellite species distribution models in the Interlake region of Manitoba

Common Model	Satellite Model



4.2 Comparison of Species Distribution Model Performance

With an average of 25 runs produced for each model type the average AUC values were quite high (Common Model 0.978 SD 0.001 and Satellite Model: 0.961 SD 0.006).

The percent contribution of environmental variables used in the Common Model (Appendix 8 –Table 12) were: Land use (34%), soil deposition mode (21%), SRTM DEM (17%), soil drainage (14%), distance to deciduous (8%), soil Ph (5%), and area of closest deciduous forest stand (2%).

The Jackknife test for the Common Model (Appendix 8 – Figure 7Table 12), shows that specifically, the SRTM DEM, land use, and soil drainage variables provide the most useful information to the model: When run individually SRTM DEM and land use contribute most to the model based on increased model gain. However, model gain decreases the most when land use and soil drainage variables are omitted followed by the SRTM DEM.

Review of the Common Model individual response curves (Appendix 8 – Figure 9Table 12) for the environmental variables in isolation shows that the largest likelihood of species presence occurs where elevation (SRTM DEM) = 264 m (study area mean = 264 m; range: 212 m – 313 m), and Land use class is predominately 4 (grassland/rangeland). The likelihood of finding Dakota skipper is also linked (but to a lesser degree) to: 6 (marsh), 16 (roads/trails), and 11 (open deciduous forest); soil pH of 7.8, soil deposition mode of 11 (undifferentiated mineral) and a second lower peak for 7 (till), distance to deciduous near 100 m, soil drainage class of 5 (Poor), and area of closest forest stand (approximately 1000 ha).

The percent contribution of environmental variables (Appendix 8 – Table 13) used in the Satellite Model include the SRTM DEM (45%), NDVI GLCM entropy (25%), tasseled cap wet (20%), tasseled cap green (10%), and ARVI8a (1%).

The Jackknife test for the Satellite Model (Appendix 8 – Figure 8**Error! Reference source not found.**Table 12) shows that the SRTM DEM provides the most useful information to the model. Model gain is greatest when the SRTM DEM is the only variable and the model gain is reduced considerably when it is removed.

Satellite model individual response curves (Appendix 8 – Figure 10**Error! Reference source not found.Error! Reference source not found.**Table 12) show highest likelihood of species presence is where elevation (SRTM DEM) = 264 m, NDVI GLCM entropy = 2.4, tasseled cap wet = -0.16, tasseled cap green = 0.02, and ARVI8a = 0.4. Excluding elevation, interpreting these values are not as straightforward as with the Common Model. The Entropy measures the amount of disorder, or in the context of GLCM analysis, variability in an image. Therefore, higher Entropy values indicates greater local variability (Hall-Beyer 2017). The GLCM entropy raster range is 0 – 4.2 and an average value 1.0. Therefore, local site variability where Dakota skipper is predicted is moderately high at 2.4. This may be described by the hill and swale topography at most known Dakota skipper sites where variability would be high due to the close arrangement of different vegetation types: Hayland-pasture grassland, deciduous or mixed forest, and wet meadow and wetland. The tasseled cap wet values across the study area range from -1.26 to 0.60 mean -

0.08. Higher values indicates wetter areas (Kauth and Thomas 1976). The value -0.16 indicates predicted Dakota skipper areas are likely slightly drier than average in the study area which may describe the relatively drier upland Dakota skipper sites on hayland and pasture where the study area average would be lower due to wetlands and meadows. Tasseled cap greenness values across the study area range from -1.20 to 0.32 (mean 0.04). Higher values indicates denser or more vigorously growing vegetation (Kauth and Thomas 1976). The value of 0.02 indicates areas where Dakota skipper is predicted are slightly less green than average and may be indicative of the less densely vegetated haylands/ grasslands and pastures. A visual inspection of the greenness tasseled cap raster also indicates that the mean and maximum greenness values may also been boosted by the vigorous growth of plants in the agricultural cropland areas of the study. ARVI8a ranges from -0.7 to 0.9 average 0.48 for the study area. Low values can be interpreted as areas with no or sparse vegetation such as water or rock. Higher values generally indicate more dense, green vegetation such as temperate rainforest. Values of 0.2 to 0.4 generally indicate grassland or shrub (Weier and Herring 2000). The value of 0.4 indicates Dakota skipper are predicted to be found in grassland (hayland/pasture) or shrub areas. The higher value may be due to locations being in close proximity to deciduous or mixed wood stands which is characteristic of the close arrangement of forest, grassland, and wet meadows discussed earlier and is common at known Dakota skipper locations.

4.3 Ground Validation Results

I surveyed a total of 128 sites between 2 July and 17 September 2021. From the 120 candidate sites generated randomly for each model, I sampled 66 Common Model sites and 62 of the Satellite Model sites. The number of sites by model type and suitability categories are shown in

Table 8.

Table 7. Average site suitability score and plant species count by model and suitability threshold. Site suitability scores = sum of indicator plant species scores found at each site shows site plant suitability score totals averaged 11.3 (range 0 – 49). The Common Model suitability

scores averaged 12.2 (range 0 – 49) and the Satellite Model scores averaged 10.3 (range 0-49) and indicator plant suitability scores averaged 5.3. The Common Model averaged 5.7 (range 0 – 49) and the Satellite Model scores averaged 5.0 (range 0-49). I classified survey sites into High, Medium, and Low suitability categories based on the total plant suitability score obtained from the ground survey. A site with High suitability was considered to have a total plant score of > = 30, while Medium sites had scores of <30 >=10 and Low sites were considered to have scores of <10 (Table 7. Average site suitability score and plant species count by model and suitability threshold. Site suitability scores = sum of indicator plant species scores found at each site). High category sites were interpreted as generally containing a large portion of native plants, some of which are nectar and larval hosts for Dakota skipper. Medium category sites although having some plants present which support or are characteristic of typical Dakota skipper habitat lacked certain species or the quality or quantity of host plants was lower which was considered reflective of less desirable habitat. Sites in the Low Category lacked most of the larval or adult host plants capable of supporting or sustaining a long-term population of Dakota skipper (Dearborn *et al.* 2022). Table 7. Average site suitability score and plant species count by model and suitability threshold. Site suitability scores = sum of indicator plant species scores found at each site shows that the average number of plant indicator species per site match the average plot suitability scores per site by their respective suitability category. A detailed list of plant indicator species by site from the field survey is available on request.

Table 7. Average site suitability score and plant species count by model and suitability threshold. Site suitability scores = sum of indicator plant species scores found at each site.

Model Type	Suitability Category	Average Plot Suitability Score	Average Number of Indicator Plant Species Per Plot
Common Model	Low	1.8	1.1
	Medium	17.5	8.5
	High	39.3	16.6
	All Categories	12.2	5.7
Satellite Model	Low	1.9	1.3
	Medium	19.2	9.4
	High	39.2	16.3
	All Categories	10.3	5.0

Table 8. Survey sites by model type and suitability categories.

Model	Suitability Category			Total
	Low	Medium	High	
Common Model	23	23	20	66
Satellite Model	22	23	17	62
Total	45	46	37	128

Table 9. Survey sites by model type and plant score suitability categories.

Model	Suitability Category			Total
	Low	Medium	High	
Common Model	36	20	10	66
Satellite Model	39	17	6	62
Total	75	37	16	128

Pearson's correlation coefficient estimates calculated from continuous model relative likelihood values and field sampled vegetation suitability scores at sample sites were strong to moderate: Common Model 0.64 (0.47-0.76; 95% CI) and Satellite Model 0.57 (0.36-0.71; 95% CI). A paired z-test found no statistical difference between these two model correlation coefficients ($z = 0.62$, $p = 0.54$). The z-test was used for these two independent datasets by transforming values to standard z-scores using the z-transformation.

Confusion matrixes were generated for each model rating the relative likelihood categories to the reference plant score categories for each site (Table 10 and Table 11). The accuracy of the Common model was 0.58 (0.44 – 0.70 95% CI) and Satellite model was 0.52 (0.39 – 0.66 95% CI). The results of the confusion matrix show both models over-assigned a Low rank plant site as a Medium category (13 sites classified as Medium likelihood of presence in both models were classified as Low quality sites by the plant scoring carried out during ground validation). However, both models performed better when discriminating Low and Medium plant score categories where there was agreement with 20 sites in the Low category and 9 and 8 sites respectively in the Medium category. Both models also classed Low and Medium sites as High suitability rank in some cases.

Table 10. Confusion matrix for Common Model.

Relative Likelihood Categories	Plant Score Categories		
	Low	Medium	High
Low	20	3	0
Medium	13	9	1
High	3	8	9

Table 11. Confusion matrix for Satellite Model.

Relative Likelihood Categories	Plant Score Categories		
	Low	Medium	High
Low	20	2	0
Medium	13	8	2
High	6	7	4

Cohen's Kappa tests indicated both models were in fair agreement with the reference suitability categories. However, the Common Model performed better (k = 0.36) than the Satellite Model (k = 0.25).

CHAPTER 5: DISCUSSION

I compared two SDMs for the Dakota skipper using different ways to generate model variables: One approach used readily available geospatial layers to create model variables while the other approach used only satellite imagery to generate variables. Although the Common Model performed nominally better than the Satellite Model, model performance metrics suggest that the two approaches were comparable based on statistics generated with the model (i.e. AUC) and other analysis performed including Accuracy, Kappa, and Pearson Correlation. The area of the study allocated by each model for each relative likelihood category (High, Medium, Low) were similar; AUC values for each model were high but also comparable between models; Confusion matrix accuracy measurements from the thresholded categories were also comparable including overall accuracy and Kappa statistics. Finally, Pearson's correlation estimates from the continuous relative likelihood values were statistically similar.

The Common Model high suitability category is concentrated around the Shoal Lakes north to Lundar which is in stark contrast to the more dispersed nature of the high suitability category for the Satellite model. A visual inspection of the environmental predictor land use classes show that the primary grassland/rangeland class and the three secondary predictor classes are not organized solely in that area. The soil variables show a topological error to the south indicated by a stark change in soil attributes. This error shows in the relative likelihood layer as an abrupt linear bound to the High suitability category running east-west which matches the error in the soil layer. However, the error likely only limits the extent of the High suitability category slightly to the south and has no influence on other areas. This is supported by the soil and other predictor variables that are relatively dispersed and do not show a particular concentration around the high suitability category coverage. The threshold level may also influence the spatial distribution of the study area classified as part of the High suitability category- higher thresholds may spatially constrain the distribution where a lower threshold may create a more dispersed pattern. This could be ruled out for this study since both models had the same High suitability thresholds set at 0.80.

The remarkably high AUC values for both models are likely from using cross-validation and large background sample of pseudo-absences both of which are known to inflate AUC values and lead to model overfit (Elith and Burgman 2002; Phillips and Dudík 2008; Anderson and Raza 2010). The large ratio of withheld test data and the relatively high number of background points (10 000 points) may have exaggerated this effect. Apart from measuring model performance, the high AUC values may indicate that Dakota skipper occupies a restricted environmental space within the defined study area and the provided environmental predictors compared to a more generalist species which would have lower AUC scores (Lobo *et al.* 2008).

The inflated AUC values for both models underscores the need to field validate SDMs as done in this study and advocated by others (Westwood *et al.* 2019; Dearborn *et al.* 2022) to get a more realistic measure of model performance.

It is also worth noting this study demonstrates that variable response curves derived from satellite imagery can also be translated relatively easily into useful values to describe predicted suitable site conditions. Interpreting these site conditions shows that, despite the differences in how the environmental variables were constructed for each model, the interpretation for both are generally in agreement: the SRTM variable, the only one used in both models, has the same likelihood of occurrence value for both models (264 m); the predominantly predicted grassland/rangeland class in the Land use variable is analogous to the ARVI8a ,entropy, tasseled cap wet and green values which indicate drier grassland/pasture areas as predominant but with some variability which can be associated with the secondary land use categories noted as being important in the Common Model.

5.1 Study Limitations

The study produced two usable SDMs for the Dakota skipper. However, there were some limitations encountered during the research. First, the modelling software required all predictor variables to have the same raster cell size resolution (30 m for the Common Model and 10 m for the Satellite Model). This necessitated up sampling some of the Sentinel 2A bands from 20 m and 60 m resolution to 10 m resolution, and the SRTM DEM resampling from 30 m to 10 m. Error from resampling low resolution to high resolution was reduced by generating most variables from the Super Resolution mosaic which resampled the 20 m and 60 m bands using a more sophisticated modelling technique than the regular resampling algorithm. The Super Resolution processor used the information in the existing 10 m bands (blue, green, red, and near infra-red) to inform how the lower resolution bands were resampled.

The SRTM DEM derivatives evaluated in the Satellite Model were not evaluated in the Common Model. This could arguably have changed the predictor variables used to build the Common Model and subsequentially, performance. However, correlation analysis performed during Satellite Model variable selection showed that the STRM DEM and its derivatives were highly correlated resulting in the STRM DEM being selected for the model with the derivatives being were dropped. This would likely have been a similar outcome if the STRM DEM derivatives were included as candidate variables in the Common Model.

Land cover was not available for the northeast corner of the study area. The lack of coverage supports the need for using remotely sensed imagery to create predictor variables. The missing portion was relatively small and covers what is unsuitable boreal and peatland habitat, according to visual inspection of satellite and airphoto imagery. Therefore, the missing land cover layer in the area likely did not appreciably affect model performance.

SDM modelling is a sophisticated and highly technical research field requiring expertise in geomatics, statistics, data management, modelling, ecology, and field work. These specialized skills can be a barrier to conservation researchers interested modelling. Generating SDMs solely from remotely sensed imagery may be extra burdensome requiring additional technical knowledge to manage and manipulate remotely sensed imagery, and knowledge of indices and specialized software to produce model variables.

5.2 Conservation Management Implications

This study supports using satellite imagery as the sole source of predictor variables for creating SDMs for conservation management actions. The advantages of remotely sensed imagery can be leveraged to generate SDMs for species include large spatial extent of available data in a consistent format. For example, it can be used in areas which do not have readily available geo-spatial products. Reasons for geo-spatial information poor regions include: Administrative areas which do not have active geomatics programmes, or no economic incentives (i.e. forestry, agriculture, mining) to invest in products, and remote locations. Furthermore, creating SDMs may be hampered by geopolitical boundaries. For example, geo-spatial data maybe available in one jurisdiction but not the other. Or available in both but not compatible due to differences in methods used to create the products or differences in classification systems. Another advantage for conservation management is the relatively low cost of imagery and derive predictor variables. Many imagery products used for SDM work are freely available and easy to acquire. Deriving predictor variables can be done rather quickly, using minimal labour, on a mid-level computer, with opensource software. Frequent imagery acquisition intervals can also be advantageous for conservation management. New SDMs can be made using current environmental predictors and older models can be updated faster without having to wait for a pre-made predictor to be released.

5.3 Future Study Recommendations

Future work could involve continuing to build and compare both model types but with different species in other geographic areas. This would further validate using remotely sensed imagery, to create SDMs is a viable option.

Experimentation with new remote sensing platforms, different indices, and other techniques could be done to create novel predictor variables. Through this process, new predictor variables which generate better performing models may be identified.

Investigations generating a hybrid Dakota skipper model using the top performing variables from both models could be undertaken. This research could take advantage of work already carried out in this study, including the developed predictor variables and ground validation surveys and may produce a model that performs better than the two parent models.

To further test and validate the developed model, work could be conducted to survey for Dakota skipper in the Interlake.

Work could be conducted to enhance and refine the indicator species list with further vegetation surveys at known Dakota skipper sites in the Interlake and incorporate the findings from (Nordmeyer *et al.* 2021) which found that some known larvae hosts were less healthy for larvae to feed on than other native plants.

Investigations of the ground validation plot data and/or the predictor variables from this study and Dearborn *et al.* (2022) could be merged and if this merged dataset could be used to build other Dakota skipper models or SDMs for other species.

REFERENCES

Agriculture and Agri-Food Canada (2017) Detailed soil survey (DSS) compilations. Version 3. Available at: http://sis.agr.gc.ca/cansis/nsdb/dss/v3/index.html.

Anderson, R.P. and Gonzalez, I., Jr. (2011) Species-specific tuning increases robustness to sampling bias in models of species distributions: An implementation with Maxent, Ecological Modelling, 222(15), pp. 2796–2811. Available at: https://doi.org/10.1016/j.ecolmodel.2011.04.011.

Anderson, R.P. and Martinez-Meyer, E. (2004) Modeling species' geographic distributions for preliminary conservation assessments: an implementation with the spiny pocket mice (Heteromys) of Ecuador, Biological Conservation, 116(2), pp. 167–179. Available at: https://doi.org/10.1016/S0006-3207(03)00187-3.

Anderson, R.P. and Raza, A. (2010) The effect of the extent of the study region on GIS models of species geographic distributions and estimates of niche evolution: preliminary tests with montane rodents (genus Nephelomys) in Venezuela, Journal of Biogeography, 37(7), pp. 1378–1393. Available at: https://doi.org/10.1111/j.1365-2699.2010.02290.x.

Araújo, M.B. and Guisan, A. (2006) Five (or so) challenges for species distribution modelling, Journal of Biogeography, 33(10), pp. 1677–1688. Available at: https://doi.org/10.1111/j.1365-2699.2006.01584.x.

Austin, M.P. (2002) Spatial prediction of species distribution: an interface between ecological theory and statistical modelling, Ecological Modelling, 157(2–3), pp. 101–118. Available at: https://doi.org/10.1016/S0304-3800(02)00205-3.

Barnes, K.W., Toso, L.B. and Niemuth, N.D. (2024) Dakota skipper distribution model for North Dakota, South Dakota, and Minnesota aids conservation planning under changing climate scenarios, Frontiers in Ecology and Evolution, 12:1304748 Available at: https://www.frontiersin.org/articles/10.3389/fevo.2024.1304748.

Barry, S. and Elith, J. (2006) Error and uncertainty in habitat, Journal of Applied Ecology, 43(3), pp. 413–423. Available at: https://doi.org/10.1111/j.1365-2664.2006.01136.x.

Bates, L.M. (2006) The influence of prescribed burning and grazing on the Dakota Skipper, Hesperia dacotae, habitat in south-eastern Manitoba. University of Winnipeg, pp. 56.

Beauvais, G. et al. (2006) Element distribution modelling: A primer, Wyoming Natural Diversity Database. Wyoming: University of Wyoming, pp. 42.

Bellis, J.M. et al. (2024) Climate change threatens the future viability of translocated, Diversity and Distributions, 30(2), pp. Available at: https://doi.org/10.1111/ddi.13795.

Bellis, L.M. et al. (2008) Modeling habitat suitability for Greater rheas based on satellite image, Ecological Applications, 18(8), pp. 1956–1966. Available at: https://doi.org/10.1890/07-0243.1.

Boria, A.B. et al. (2014) Spatial filtering to reduce sampling bias can improve the performance of ecological niche, Ecological Modelling, 275, pp. 73–77. Available at: https://doi.org/10.1016/j.ecolmodel.2013.12.012.

Botín-Sanabria, D.M. et al. (2022) Digital Twin Technology Challenges and Applications: A Comprehensive, Remote Sensing, 14(6). Available at: https://doi.org/10.3390/rs14061335.

Braker, N. et al. (1985) Felton Prairie, Natural Heritage Program. St. Paul, Minnesota: Department of Natural Resources, pp. 81+20maps. Available at: http://files.dnr.state.mn.us/eco/nongame/projects/consgrant_reports/1985/1985_braker_figures.pdf.

Bréda, N.J.J. (2003) Ground-based measurements of leaf area index: a review of methods, instruments and current controversies, Journal of Experimental Botany, 54(392), pp. 2403–2417. Available at: https://doi.org/10.1093/jxb/erg263.

Britten, H.B. and Glasford, J.W. (2002) Genetic population structure of the Dakota skipper (Lepidoptera : Hesperia dacotae): A North American native prairie obligate, Conservation Genetics, 3(4), pp. 363–374.

Brodu, N. (2018) Nicolas Brodu Research: Superres. Available at: https://nicolas.brodu.net/recherche/sen2res/ (Accessed: 23 June 2018).

Brotons L. et al. (2004) Presence-absence versus presence-only modelling methods for predicting bird habitat, Ecography, 27(4), pp. 437–448. Available at: https://doi.org/10.1111/j.0906-7590.2004.03764.x.

Brown, D.G. (1994) Predicting Vegetation at Treeline using Topography and Biophysical Disturbance, Journal of Vegetation Science, 5(5), pp. 641–656. Available at: https://doi.org/10.2307/3235880.

Brzeziecki, B., Kienast, F. and Wildi, O. (1993) A simulated map of the potential natural forest vegetation of Switzerland, Journal of Vegetation Science, 4(4), pp. 499–508. Available at: https://doi.org/10.2307/3236077.

Burgman, M.A., Lindenmayer, D.B. and Elith, J. (2005) Managing landscapes for conservation under uncertainty, Ecology, 86(8), pp. 2007–2017. Available at: https://doi.org/10.1890/04-0906.

Canada (2017) Species at Risk Act. Schedule 1 (Subsections 2(1), 42(2) and 68(2)). List of Wildlife Species at Risk. Part 3. Threatened Species. Available at: http://www.sararegistry.gc.ca/species/schedules_e.cfm?id=1.

Canada (2018) Canada Gazette: Part II. Ottawa, Wednesday, February 21, Canada Gazette, 152(4), pp. 126–337.

Canada (2023) Species at Risk Act. Schedule 1 (Subsections 2(1), 42(2) and 68(2)). List of Wildlife Species at Risk. Part 2. Endangered Species. Available at: https://laws.justice.gc.ca/eng/acts/S-15.3/page-10.html#h-435647.

Carabajal, C.C. and Harding, D.J. (2006) SRTM C-band and ICESat Laser Altimetry Elevation Comparisons as a Function of Tree Cover and Relief, Photogrammetric Engineering and Remote Sensing, 72, pp. 287–298.

Cochrane, J.F. and Delphey, P. (2002) Status assessment and conservation guidelines. Dakota Skipper Hesperia Dacotae (Skinner) (Lepidoptera:Hesperiidea). Iowa, Minnesota, North Dakota, South Dakota, Manitoba, and Saskatchewan, US Fish and Wildlife Service. Twin Cities Field Office, Minnesota: US Fish and Wildlife Service, pp. 84.

Cohen, J. (1960) A Coefficient of Agreement for Nominal, Educational and Psychological Measurement, 20(1), pp. 37–46. Available at: https://doi.org/10.1177/001316446002000104.

COSEWIC (2014) COSEWIC assessment and status report on the Dakota skipper, Hesperia dacotae, in Canada, Committee on the Status of Endangered Wildlife in Canada. Ottawa: Ottawa, pp. xii + 61. Available at: www.registrelep-sararegistry.gc.ca/default_e.cfm.

Culbert, P. et al. (2012) Modeling broad-scale patterns of avian species richness across the Midwestern United States with measures of satellite image, Remote Sensing of Environment, 118, pp. 140–150. Available at: https://doi.org/10.1016/j.rse.2011.11.004.

Dana, R. (1991) Conservation management of the skippers Hesperia dacotae and ottoe, Minnesota Agricultural Experiment Station. St. Paul, Minnesota: University of Minnesota, pp. 74.

Dana, R. (1997) Characterization of three Dakota Skipper sites in Minnesota, Natural Heritage and Nongame Research Program. St. Paul, Minnesota: Minnesota Department of Natural Resources, pp. 18. Available at: http://files.dnr.state.mn.us/eco/nongame/projects/consgrant_reports/1997/1997_dana.pdf.

Dearborn, K. and Westwood, R. (2014) Predicting adult emergence of Dakota skipper and Poweshiek skipperling (Lepidoptera: Hesperiidae) in Canada, Journal of Insect Conservation, 18(5), pp. 875–884. Available at: https://doi.org/10.1007/s10841-014-9695-8.

Dearborn, K.D., Murray, C. and Westwood, R. (2022) Ground Validation Reveals Limited Applicability of Species Distribution Models for Dakota Skipper (Hesperia dacotae, Lepidoptera: Hesperiidae) Recovery Efforts in Canada, Environmental Entomology, 51(6), pp. 1249–1261. Available at: https://doi.org/10.1093/ee/nvac084.

Draper, D., Marques, I. and Iriondo, J.M. (2019) Species distribution models with field validation, a key approach for successful selection of receptor sites in conservation translocations, Global Ecology and Conservation, 19, pp. 13. Available at: https://doi.org/10.1016/j.gecco.2019.e00653.

Du, Y. et al. (2016) Water Bodies' Mapping from Sentinel-2 Imagery with Modified Normalized Difference Water Index at 10-m Spatial Resolution Produced by Sharpening the SWIR, Remote Sensing, 8(4), pp. 19. Available at: https://doi.org/10.3390/rs8040354.

Duro, D.C. et al. (2014) Predicting species diversity in agricultural environments using Landsat TM imagery, Remote Sensing of Environment, 144, pp. 214–225. Available at: https://doi.org/10.1016/j.rse.2014.01.001.

Elith, J. et al. (2006) Novel methods improve prediction of species' distributions from occurrence data, Ecography, 29(2), pp. 129–151. Available at: https://doi.org/10.1111/j.2006.0906-7590.04596.x.

Elith, J. and Burgman, M. (2002) Predictions and their validation: Rare plants in the Central Highlands, Victoria, Australia, in M.J. Scott et al. (eds) Predicting Species Occurrences: Issues of Accuracy and Scale. Washington: Island Press. pp. 303-313.

Elith, J. and Graham, C.H. (2009) Do they? How do they? WHY do they differ? On finding reasons for differing performances of species distribution models, Ecography, 32(1), pp. 66–77. Available at: https://doi.org/10.1111/j.1600-0587.2008.05505.x.

Elith, J. and Leathwick, J.R. (2009) Species Distribution Models: Ecological Explanation and Prediction Across Space and Time, Annual Review of Ecology, Evolution, and Systematics, 40(1), pp. 677–697. Available at: https://doi.org/10.1146/annurev.ecolsys.110308.120159.

Environment and Climate Change Canada (2021) Dakota skipper (Hesperidia dakotae) occurrences comprehensive dataset, (Unpublished Material).

Environment Canada (2007) Recovery Strategy for the Dakota Skipper (Hesperia dacotae) in Canada. Species at Risk Act Recovery Strategy Series, Environment Canada. Ottawa: Environment Canada, pp. vi–25.

ESA (European Space Agency) (2015) Sentinel-2 User Handbook, European Space Agency. Darmstadt, Germany: European Space Agency, pp. 64.

Escadafal, R. (1989) Remote sensing of arid soil surface color with Landsat thematic mapper, Advances in Space Research, 9(1), pp. 159–163. Available at: https://doi.org/10.1016/0273-1177(89)90481-X.

Escadafal, R., Girard, M.C. and Courault, D. (1989) Munsell soil color and soil reflectance in the visible spectral bands of landsat MSS and TM data, Remote Sensing of Environment, 27(1), pp. 37–46. Available at: https://doi.org/10.1016/0034-4257(89)90035-7.

ESRI (Environmental Systems Research Institute) (2018) ArcGIS Desktop, 10.5 (Computer Program).

ESWG (Ecological Stratification Working Group) (1995) A national ecological framework for Canada. Ottawa, Ontario: Minister of Supply and Services Canada (Book, Whole).

European Space Agency (2017) Sen2three: Software installation and user manual (SUM), Sen2three: Software installation and user manual (SUM). Available at: http://step.esa.int/thirdparties/sen2three/1.1.0/sen2three-1.1.0.htmldoc/sum.html#installation-and-setup (Accessed: 24 June 2018).

European Space Agency (2017) SNAP (SeNtinel's Application Platform 5 (Computer Program).

European Space Agency (2018) SNAP Supported Plugins: Sen2Res, SNAP Supported Plugins: Sen2Res, Available at: https://step.esa.int/main/snap-supported-plugins/sen2res/ (Accessed: 21 June 2018).

European Space Agency (2018) SNAP Supported Plugins: Sen2three, SNAP Supported Plugins: Sen2three, Available at: https://step.esa.int/main/snap-supported-plugins/sen2three/ (Accessed: 21 June 2018).

European Space Agency (2023) Copernicus Open Assess Hub, Copernicus Open Assess Hub. Available at: https://scihub.copernicus.eu/ (Accessed: 5 July 2023).

Farr, T.G. et al. (2007) The shuttle radar topography mission, Reviews of Geophysics, 45(2), pp. 33. RG2004. Available at: https://doi.org/10.1029/2005RG000183.

Feng, W. et al. (2019) An optimized non-linear vegetation index for estimating leaf area index in winter wheat Precision Agriculture, 20(6), pp. 1157–1176. Available at: https://doi.org/10.1007/s11119-019-09648-8.

Fløjgaard, C. et al. (2009) Ice age distributions of European small mammals: insights from species distribution, Journal of Biogeography, 36(6), pp. 1152–1163. Available at: https://doi.org/10.1111/j.1365-2699.2009.02089.x.

Fourcade, Y. et al. (2014) Mapping Species Distributions with MAXENT Using a Geographically Biased Sample of Presence Data: A Performance Assessment of Methods for Correcting Sampling, PloS one, 9(5), pp. e97122. Available at: https://doi.org/10.1371/journal.pone.0097122.

Fourcade, Y. (2016) Comparing species distributions modelled from occurrence data and from expert-based range maps. Implication for predicting range shifts with climate, Ecological Informatics, 36, pp. 8–14. Available at: https://doi.org/10.1016/j.ecoinf.2016.09.002.

Fu, P. and Rich, P. (2000) The solar analyst 1.0. User, 1.0 (Computer Program).

Gamer, M. et al. (2019) irr: Various Coefficients of Interrater Reliability and Available at: https://CRAN.R-project.org/package=irr (Accessed: 12 January 2024).

Gao, B. (1996) NDWI—A normalized difference water index for remote sensing of vegetation liquid water from, Remote Sensing of Environment, 58(3), pp. 257–266. Available at: https://doi.org/10.1016/S0034-4257(96)00067-3.

Goel, N.S. and Qin, W. (1994) Influences of canopy architecture on relationships between various vegetation indices and LAI and Fpar: A computer, Remote Sensing Reviews, 10(4), pp. 309–347. Available at: https://doi.org/10.1080/02757259409532252.

Goward, S.N. and Huemmrich, K.F. (1992) Vegetation canopy PAR absorptance and the normalized difference vegetation index: An assessment using the SAIL, Remote Sensing of Environment, 39(2), pp. 119–140. Available at: https://doi.org/10.1016/0034-4257(92)90131-3.

Graham, C.H. et al. (2008) The influence of spatial errors in species occurrence data used in distribution, Journal of Applied Ecology, 45(1), pp. 239–247. Available at: https://doi.org/10.1111/j.1365-2664.2007.01408.x.

Guillera-Arroita, G. et al. (2015) Is my species distribution model fit for purpose? Matching data and models to, Global Ecology and Biogeography, 24(3), pp. 276–292. Available at: https://doi.org/10.1111/geb.12268.

Guisan, A. et al. (2013) Predicting species distributions for conservation, Ecology Letters, 16(12), pp. 1424. Available at: https://doi.org/10.1111/ele.12189.

Guisan, A. and Thuiller, W. (2005) Predicting species distribution: offering more than simple habitat, Ecology Letters, 8(9), pp. 993. Available at: https://doi.org/10.1111/j.1461-0248.2005.00792.x.

Guisan, A. and Zimmermann, N.E. (2000) Predictive habitat distribution models in, Ecological Modelling, 135(2–3), pp. 147–186. Available at: https://doi.org/10.1016/S0304-3800(00)00354-9.

Halford, G. et al. (2024) Can species distribution models using remotely sensed variables inform reintroductions? Trialling methods with Carterocephalus palaemon the Chequered Skipper, Journal of Insect Conservation [Preprint], pp. 14. Available at: https://doi.org/10.1007/s10841-024-00555-6.

Hall-Beyer, M. (2017) GLCA Texture: A Tutorial. v. 3.0 March 2017. University of Calgary. Department of Geography, pp. 76. Available at: https://prism.ucalgary.ca/server/api/core/bitstreams/8f9de234-cc94-401d-b701-f08ceee6cfdf/content.

Hall-Beyer, M. (2017) Practical guidelines for choosing GLCM textures to use in landscape classification tasks over a range of moderate spatial scales, International Journal of Remote Sensing, 38(5), pp. 1312–1338. Available at: https://doi.org/10.1080/01431161.2016.1278314.

Hanley, J.A. and McNeil, B.J. (1982) The Meaning and use of the Area Under a Receiver Operating Characteristic (Roc) Curve, Radiology, 143(1), pp. 29–36. Available at: https://doi.org/10.1148/radiology.143.1.7063747.

Harrell Jr., F.E., Lee, K.L and Mark, D.B. (1996) Multivariable prognostic models: Issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors, Statistics in Medicine, 15(4), pp. 361–387. Available at: https://doi.org/10.1002/(SICI)1097-0258(19960229)15:4<361::AID-SIM168>3.0.CO;2-4.

Henderson, D. and Koper, N. (2014) Historic Distribution and Ecology of Tall-Grass Prairie in Western Canada, in Proceedings of the North American Prairie Conference. Prairie Naturalist, pp. 40–49.

Hendricks, G.A. and Alsdorf, D.E. (2004) Channel Slopes on Amazon Basin Rivers from the SRTM DEM, in. AGU Fall Meeting Proceedings.

Herbel, C.H. and Anderson, K.L. (1959) Response of True Prairie Vegetation on Major Flint Hills Range Sites to Grazing Treatment, Ecological Monographs, 29(2), pp. 171–186. Available at: https://doi.org/10.2307/1942202.

Hernandez, P.A. et al. (2006) The effect of sample size and species characteristics on performance of different species distribution modeling methods, Ecography, 29(5), pp. 773–785. Available at: https://doi.org/10.1111/j.0906-7590.2006.04700.x.

Hijmans, R.J. et al. (1999) Using GIS to check co-ordinates of genebank accessions, Genetic Resources and Crop Evolution, 46(3), pp. 291–296.

Hijmans, R.J. et al. (2000) Assessing the geographic representativeness of genebank collections: the case of Bolivian wild potatoes, Conservation Biology, 14(6), pp. 1755–1765. Available at: https://doi.org/10.1046/j.1523-1739.2000.98543.x.

Hijmans, R.J. (2012) Cross-validation of species distribution models: removing spatial sorting bias and calibration with a null model, Ecology, 93(3), pp. 679–688. Available at: https://doi.org/10.1890/11-0826.1.

Hirzel, A. and Guisan, A. (2002) Which is the optimal sampling strategy for habitat suitability modelling, Ecological Modelling, 157(2–3), pp. 331–341. Available at: https://doi.org/10.1016/S0304-3800(02)00203-X.

Hooper, R. (2003) Dakota skipper (Hesperia dacotae) in Saskatchewan, Blue Jay, 6 (Journal Article), pp. 124–125.

Huete, A., Justice, C. and Liu, H. (1994) Development of vegetation and soil indices for MODIS-EOS, Remote Sensing of Environment, 49(3), pp. 224–234. Available at: https://doi.org/10.1016/0034-4257(94)90018-3.

Huete, A.R. (1988) A soil-adjusted vegetation index (SAVI), Remote Sensing of Environment, 25(3), pp. 295–309. Available at: https://doi.org/10.1016/0034-4257(88)90106-X.

Huete, A.R. and Jackson, R.D. (1988) Soil and atmosphere influences on the spectra of partial canopies, Remote Sensing of Environment, 25(1), pp. 89. Available at: https://doi.org/10.1016/0034-4257(88)90043-0.

Huete, A.R., Jackson, R.D. and Post, D.F. (1985) Spectral response of a plant canopy with different soil backgrounds, Remote Sensing of Environment, 17(1), pp. 37–53. Available at: https://doi.org/10.1016/0034-4257(85)90111-7.

Hunter-Ayad, J. et al. (2020) Reintroduction modelling: A guide to choosing and combining models for species reintroductions, Journal of Applied Ecology, 57(7), pp. 1233–1243. Available at: https://doi.org/10.1111/1365-2664.13629.

Hutchinson, G.E. (1957) Concluding Remarks, Cold Spring Harbor Symposia on Quantitative Biology, (22), pp. 415–427.

Iowa Natural Resource Commission (2022) Endangered and Threatened Plant and Animal Species, Iowa Administrative Code. Available at: https://www.legis.iowa.gov/docs/iac/chapter/571.77.pdf (Accessed: 27 November 2023).

Jiménez, L. and Soberón, J. (2020) Leaving the area under the receiving operating characteristic curve behind: An evaluation method for species distribution modelling applications based on presence-only data, Methods in Ecology and Evolution, 11(12), pp. 1571–1586. Available at: https://doi.org/10.1111/2041-210X.13479.

Jiménez-Valverde, A. et al. (2011) Use of niche models in invasive species risk assessments, Biological Invasions, 13(12), pp. 2785–2797. Available at: https://doi.org/10.1007/s10530-011-9963-4.

Jiménez-Valverde, A. and Lobo, J.M. (2007) Threshold criteria for conversion of probability of species presence to either–or presence–absence, Acta oecologica (Montrouge), 31(3), pp. 361–369. Available at: https://doi.org/10.1016/j.actao.2007.02.001.

Kaufman, Y.J. and Tanre, D. (1992) Atmospherically resistant vegetation index (ARVI) for EOS-MODIS, - IEEE Transactions on Geoscience and Remote Sensing, 30(2), pp. 261–270. Available at: https://doi.org/10.1109/36.134076.

Kauth, R.J. and Thomas, G.S. (1976) The Tasselled Cap - A Graphic Description of the Spectral-Temporal Development of Agricultural Crops as Seen by LANDSAT. LARS Symposia, paper 159. Purdue University, Indiana. Available at: https://docs.lib.purdue.edu/lars_symp/159/.

Klassen, P.. Westwood, R., Preston, W. and McKillop, B. (1989) The Butterflies of Manitoba. Canada: Manitoba Museum (Book, Whole), pp. 290. Available at: http://catalog.hathitrust.org/Record/009190710.

Lahoz-Monfort, J.J. et al. (2010) Satellite imagery as a single source of predictor variables for habitat suitability modelling: how Landsat can inform the conservation of a critically endangered lemur, Journal of Applied Ecology, 47(5), pp. 1094–1102. Available at: https://doi.org/10.1111/j.1365-2664.2010.01854.x.

Landis, J.R. and Koch, G.G. (1977) The Measurement of Observer Agreement for Categorical Data, Biometrics, 33(1), pp. 159–174. Available at: https://doi.org/10.2307/2529310.

Lange, M. et al. (2017) Validating MODIS and Sentinel-2 NDVI Products at a Temperate Deciduous Forest Site Using Two Independent Ground-Based Sensors, Sensors, 17(8), pp. 18. Available at: https://doi.org/10.3390/s17081855.

Lapin, C.N., Etterson, M.A. and Niemi, G.J. (2013) Occurrence of the Connecticut Warbler Increases With Size of Patches of Coniferous Forest, The Condor, 115(1), pp. 168–177. Available at: https://doi.org/10.1525/cond.2013.110202.

Layberry, R., Hall, P. and Lafontaine, D. (2015) The Butterflies of Canada. Toronto: University of Toronto Press (Book, Whole).

Lees, B.G. and Ritman, K. (1991) Decision-Tree and Rule-Induction Approach to Integration of Remotely Sensed and GIS Data in Mapping Vegetation in Disturbed or Hilly Environments, Environmental management, 15(6), pp. 823–831. Available at: https://doi.org/10.1007/BF02394820.

LeFavour, G. and Alsdorf, D.E. (2005) Water slope and discharge in the Amazon River estimated using the shuttle radar topography mission digital elevation model, Geophysical Research Letters, pp. 32.

Liang, S. and Wang, J. (2019) Advanced Remote Sensing: Terrestrial Information Extraction and Applications. Elsevier Science, pp. 19. Available at: https://books.google.com/books?id=1-y_DwAAQBAJ.

Lin, X. et al. (2022) Estimating 10-m land surface albedo from Sentinel-2 satellite observations using a direct estimation approach with Google Earth Engine, ISPRS Journal of Photogrammetry and Remote Sensing, 194, pp. 1–20. Available at: https://doi.org/10.1016/j.isprsjprs.2022.09.016.

Lipsey, M.K. et al. (2015) One step ahead of the plow: Using cropland conversion risk to guide Sprague’s Pipit conservation in the northern Great Plains, Biological Conservation, 191, pp. 739–749. Available at: https://doi.org/10.1016/j.biocon.2015.08.030.

Lischke, H. et al. (1998) Vegetation responses to climate change in the Alps-Modeling studies, in P. Cebon et al. (eds) Views From the Alps: Regional Perspectives on Climate Change. Boston: MIT Press, pp. 309–305. Available at: http://libproxy.uwinnipeg.ca/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=nlebk&AN=1353.

Liu, C. et al. (2005) Selecting thresholds of occurrence in the prediction of species distributions, Ecography (Copenhagen), 28(3), pp. 385–393. Available at: https://doi.org/10.1111/j.0906-7590.2005.03957.x.

Liu, R. et al. (2015) Modelling of fraction of absorbed photosynthetically active radiation in vegetation canopy and its validation, Biosystems Engineering, 133, pp. 81–94. Available at: https://doi.org/10.1016/j.biosystemseng.2015.03.004.

Lobo, J.M., Jiménez-Valverde, A. and Real, R. (2008) AUC: a misleading measure of the performance of predictive distribution models, Global Ecology and Biogeography, 17(2), pp. 145–151. Available at: https://doi.org/10.1111/j.1466-8238.2007.00358.x.

Manitoba Conservation (1996) Manitoba Conservation Forest Inventory Manual 1.3: 1996-1997, (Unpublished Material), pp. 22.

Manitoba Conservation Data Centre (2018) Department of Natural Resources and Northern Development. Available at: https://gov.mb.ca/nrnd/fish-wildlife/cdc/index.html.

Manitoba, G. of (1998) Manitoba 1998 Manitoba Endangered Species Act Regulation Listing Threatened Endangered Extirpated Species, 1 Threatened species(Laws/Statutes). Available at: http://web2.gov.mb.ca/laws/regs/current/_pdf-regs.php?reg=25/98.

Manitoba Government (2022) Data MB, Data MB. Available at: https://geoportal.gov.mb.ca/ (Accessed: 2 January 2022).

Manitoba Land Initiative (2017) Manitoba Natural Resources and Northern Development. Available at: https://mli.gov.mb.ca/ (Accessed: 2 January 2022).

Manitoba Remote Sensing Centre (2006) Land Use/ Land Cover classes for Manitoba. Manitoba Department of Natural Resources and Northern Development., (Unpublished Material), pp. 2.

Marshal, I.B., Schut, P.H. and Ballard, M. (1999) A National Ecological Framework for Canada: Attribute Data. Agriculture and Agri-Food Canada, Research Branch, Centre for Land and Biological Resources, and Environment Canada, State of the Environment Directorate, Ecozone Analysis Branch, Ottawa/Hull. Available at: https://sis.agr.gc.ca/cansis/nsdb/ecostrat/1999report/index.html.

MBCDC (2017) Conservation status ranks of vertebrate animals in Manitoba, Manitoba Conservation Data Centre. Winnipeg, Manitoba: Manitoba Conservation Data Centre. Available at: http://www.gov.mb.ca/sd/cdc/pdf/animal_ranks.pdf.

MBCDC (2023) Conservation status ranks of vertebrate animals in Manitoba, Manitoba Conservation Data Centre. Winnipeg, Manitoba: Manitoba Conservation Data Centre. Available at: https://gov.mb.ca/nrnd/fish-wildlife/cdc/index.html (Accessed: 27 November 2023).

McCabe, T. (1981) The Dakota Skipper, Hesperia dacotae (SKINNER): Range and biology, with special reference to North Dakota, Journal of the Lepidopterists Society, 35(3), pp. 179–193.

McCabe, T. and Post, R. (1977) Skippers (Hesperioidea) of North Dakota, Agricultural Experimental Station. North Dakota Insect Publication No. 11. North Dakota: North Dakota State University, pp. 70.

McFeeters, S.K. (1996) The use of the normalized difference water index (NDWI) in the delineation of open water features, International Journal of Remote Sensing, 17(7), pp. 1425–1432. Available at: https://doi.org/10.1080/01431169608948714.

MDNR (2018) Rare Species Guide: Hesperia dacotae. Available at: www.dnr.state.mn.us (Accessed: 27 November 2023).

Merow, C. et al. (2014) What do we gain from simplicity versus complexity in species distribution models?, Ecography, 37(12), pp. 1267–1281. Available at: https://doi.org/10.1111/ecog.00845.

Merow, C., Smith, M.J. and Silander, J.A. (2013) A practical guide to MaxEnt for modeling species' distributions: what it does, and why inputs and settings matter, Ecography (Copenhagen), 36(10), pp. 1058–1069. Available at: https://doi.org/10.1111/j.1600-0587.2013.07872.x.

Microsoft Corporation (2018) Microsoft Azure Cloud Computing, Azure: Invent with purpose. Available at: https://azure.microsoft.com/en-ca/ (Accessed: 20 June 2018).

Moore, I.D., Grayson, R.B. and Ladson, A.R. (1991) Digital terrain modelling: A review of hydrological, geomorphological, and biological applications, Hydrological Processes, 5(1), pp. 3–30. Available at: https://doi.org/10.1002/hyp.3360050103.

Morden, C.-J. (2006) Potential reintroduction of the Dakota skipper (Hesperia dacotae) into south-eastern Manitoba. University of Winnipeg, pp. 31.

Mueller-Wilm, U., Devignot, O. and Pessiot, L. (2018) Sentinel 2 (S2 MPC): Sen2cor Configuration and User manual. Issue 2. European Space Agency, pp. 64.

Murray, C. (2013) Digital elevation model raster and TIN, 3D scene and fly thru for the Delta Marsh area of Manitoba. Brandon, Manitoba: Assiniboine Community College, pp. 4.

Murray, C. (2016) Testing of NDWI2 and MNDWI to measure surface water of Whitewater Lake, Manitoba. And a comparison of the two indices.

Naegeli, K. et al. (2017) Cross-Comparison of Albedo Products for Glacier Surfaces Derived from Airborne and Satellite (Sentinel-2 and Landsat 8) Optical Data, Remote Sensing, 9(2), pp. 22. Available at: https://doi.org/10.3390/rs9020110.

NatureServe (2002) Element Occurrence Data Standard, NatureServe. Arlington, Virginia: NatureServe, pp. 201.

NatureServe (2022) Hesperia dacotae. Available at: explorer.natureserve.org (Accessed: 27 November 2023).

NatureServe (2024) NatureServe Explorer Pro, NatureServe Explorer Pro. Available at: https://natureserve-explorer-pro (Accessed: 9 March 2024).

Nordmeyer, C.S., Runquist, E., and Stapleton, S. (2021) Invasive grass negatively affects growth and survival of an imperiled butterfly, Endangered Species Research, 45, pp. 301–314.

North Dakota Fish and Game (2021) Threatened and Endangered Species: North Dakota ESA Listed Species. Available at: https://gf.nd.gov/wildlife/endangered.

Owens, H.L. et al. (2013) Constraints on interpretation of ecological niche models by limited environmental ranges on calibration areas, Ecological Modelling, 263, pp. 10–18. Available at: https://doi.org/10.1016/j.ecolmodel.2013.04.011.

Pearson, R.G., Dawson, T.P. and Liu, C. (2004) Modelling species distributions in Britain: a hierarchical integration of climate and land-cover data, Ecography, 27(3), pp. 285–298. Available at: https://doi.org/10.1111/j.0906-7590.2004.03740.x.

Phillips, S.J. et al. (2009) Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data, Ecological Applications, 19(1), pp. 181–197. Available at: https://doi.org/10.1890/07-2153.1.

Phillips, S.J. et al. (2009) Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data, Ecological Applications, 19(1), pp. 181–197. Available at: https://doi.org/10.1890/07-2153.1.

Phillips, S.J., Anderson, R.P. and Schapire, R.E. (2006) Maximum entropy modeling of species geographic distributions, Ecological Modelling, 190(3–4), pp. 231–259. Available at: https://doi.org/10.1016/j.ecolmodel.2005.03.026.

Phillips, S.J. and Dudík, M. (2008) Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation, Ecography, 31(2), pp. 161–175. Available at: https://doi.org/10.1111/j.0906-7590.2008.5203.x.

Pielke, R. (2003) The role of models in prediction for decision, in C. Canham, J. Cole, and W. Lauenroth (eds) Models in Ecosystem Science. Princeton, New Jersey: Princeton University Press, pp. 111–135.

Post van der Burg, M. et al. (2020) Capturing Spatiotemporal Patterns in Presence-Absence Data to Inform Monitoring and Sampling Designs for the Threatened Dakota Skipper (Lepidoptera: Hesperiidae) in the Great Plains of the United States, Environmental Entomology, 49(5), pp. 1252–1261. Available at: https://doi.org/10.1093/ee/nvaa081.

Pradervand, J.-N. et al. (2014) Very high resolution environmental predictors in species distribution models: Moving beyond topography?, Progress in Physical Geography, 38(1), pp. 79–96. Available at: https://doi.org/10.1177/0309133313512667.

Purevdorj, TS. et al. (1998) Relationships between percent vegetation cover and vegetation indices, International Journal of Remote Sensing, 19(18), pp. 3519–3535. Available at: https://doi.org/10.1080/014311698213795.

Qi, J. et al. (1994) A modified soil adjusted vegetation idex, Remote Sensing of Environment, 48(2), pp. 119–126. Available at: https://doi.org/10.1016/0034-4257(94)90134-1.

R Core Team (2021) R Statistical Software. Version: 4.1.2.

Rabe, M.L. et al. (2002) Use of a GIS-based habitat model to identify potential release sites for Michell’s satyr (Neonympha m. mitchellii) in Michigan. 2002–07. Michigan State University Extension: US Fish and Wildlife Service, pp. 13.

Randin, C.F. et al. (2006) Are niche-based species distribution models transferable in space?, Journal of Biogeography, 33(10), pp. 1689–1703. Available at: https://doi.org/10.1111/j.1365-2699.2006.01466.x.

Reese, G.C. et al. (2005) Factors Affecting Species Distribution Predictions: A Simulation Modeling Experiment, Ecological Applications, 15(2), pp. 554–564. Available at: https://doi.org/10.1890/03-5374.

Revelle, W. (2023) psych: Procedures for Psychological, Psychometric, and Personality Research. Northwestern University, Evanston, Illinois. Available at: <https://CRAN.R-project.org/package=psych> (Accessed: 12 January 2024).

Rich, P. and Fu, P. (2000) Topoclimatic Habitat Models. Proceedings of the Fourth International Conference on Integrating GIS and Environmental Modeling. Banff, Alberta, Canada, pp. 4.

Rigney, C. (2013) Habitat characterization and biology of the threatened Dakota Skipper (Hesperia dacotae) in Manitoba. University of Winnipeg, pp. 259.

Rosenfield, G.H. and Fitzpatrick-Lins, K. (1986) A coefficient of agreement as a measure of thematic classification accuracy., Photogrammetric Engineering and Remote Sensing, 52, pp. 223–227.

Rouse, J.W. et al. (1974) Monitoring the vernal advancement and retrogradation (Greenwave effect) of natural vegetation. Final Report. Texas A&M University Remote Sensing Center, pp. 390.

Rowe, R.J. (2005) Elevational gradient analyses and the use of historical museum specimens: a cautionary tale, Journal of Biogeography, 32(11), pp. 1883–1897. Available at: https://doi.org/10.1111/j.1365-2699.2005.01346.x.

Royer, E. (2019) Hesperia dacotae. The IUCN Red List of Threatened Species 2019, Dakota Skipper. Available at: https://dx.doi.org/10.2305/IUCN.UK.2019-2.RLTS.T9968A122963341.en (Accessed: 27 November 2023).

Royer, R. and Marrone, G. (1992) Conservation status of the Dakota Skipper (Hesperia dacotae) in North and South Dakota, US Fish and Wildlife Service. Denver, Colorado: US Fish and Wildlife Service, pp. 50.

Royer, R., McKenney, R. and Newton, W. (2008) A characterization of non-biotic environmental features of prairies hosting the Dakota Skipper (Hesperia Dacotae, Hesperiidae) across its remaining US Range, Journal of the Lepidopterists Society, 62(1), pp. 1–17.

Scherrer, D., Mod, H.K. and Guisan, A. (2020) How to evaluate community predictions without thresholding?, Methods in Ecology and Evolution, 11(1), pp. 51–63. Available at: https://doi.org/10.1111/2041-210X.13312.

Schilcht, D. and Saunders, M. (1995) Completion of status surveys for the dakota skipper (Hesperia dacotae) and the poweshiek skipper (Oarisma poweshiek) in Minnesota with data on the Regal Fritillary (Speyeria idalia). Final report., Minnesota Nongame Wildlife Program. Minnesota Department of Natural Resources, pp. 43. Available at: http://files.dnr.state.mn.us/eco/nongame/projects/consgrant_reports/1993/1993_schlicht.pdf.

SDGFP (2022) Rare Animals of South Dakota. Available at: https://gfp.sd.gov/rare-animals/ (Accessed: 27 November 2023).

Seidle, K.M. et al. (2018) Environmental associations of Hesperia dacotae (Lepidoptera: Hesperiidae) in southeastern Saskatchewan, Canada, The Canadian Entomologist. 2018/07/31 edn, 150(5), pp. 652–662. Available at: https://doi.org/10.4039/tce.2018.33.

Seidle, K.M. et al. (2020) Extent of Dakota skipper, Hesperia dacotae, distribution in Southeastern Saskatchewan, Canada, Journal of Insect Conservation, 24(6), pp. 1073–1081. Available at: https://doi.org/10.1007/s10841-020-00276-6.

Selby, G. and Glenn-Lewin, D.C. (1989) A systematic inventory, population monitoring program, and ecological study of rare Lepidoptera at the Prairie Coteau Scientific and Natural Area (SNA), Pipestone County, Minnesota. Final report submitted to the Minnesota Department of Natural Resources., Minnesota Department of Natural Resources. Minnesota Department of Natural Resources, pp. 49. Available at: http://files.dnr.state.mn.us/eco/nongame/projects/consgrant_reports/1989/1989_glenn-lewin_selby.pdf.

Shi, T. and Xu, H. (2019) Derivation of Tasseled Cap Transformation Coefficients for Sentinel-2 MSI At-Sensor Reflectance Data, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 12(10), pp. 4038–4048. Available at: https://doi.org/10.1109/JSTARS.2019.2938388.

SKCDC (2017) Saskatchewan Invertebrate Taxa List, Saskatchewan Conservation Data Centre. Regina, Saskatchewan: Saskatchewan Conservation Data Centre, pp. 124. Available at: http://www.biodiversity.sk.ca/SppList.htm.

SKCDC (2023) Saskatchewan Invertebrate Taxa List, Saskatchewan Conservation Data Centre. Regina, Saskatchewan: Saskatchewan Conservation Data Centre, pp. 15. Available at: http://biodiversity.sk.ca/TaxaList/sk-taxa-invertebrate-track.pdf.

Skinner, H. (1911) New species or subspecies of North American butterflies (Lepid.), Entomological News, 22(9), pp. 412–413.

Skinner, H. and Williams, R.C. (1924) On the Male Genitalia of the Hesperiidae of North America: Paper VI, Transactions of the American Entomological Society (1890-), 50(3), pp. 177–208.

Smeraldo, S. et al. (2017) Species distribution models as a tool to predict range expansion after reintroduction: A case study on Eurasian beavers (Castor fiber), Journal for Nature Conservation, 37, pp. 12–20. Available at: https://doi.org/10.1016/j.jnc.2017.02.008.

Sousa-Silva, R. et al. (2014) Improving the assessment and reporting on rare and endangered species through species distribution models, Global Ecology and Conservation, 2, pp. 226–237. Available at: https://doi.org/10.1016/j.gecco.2014.09.011.

Star Wars: Episode IV - A New Hope (1977). Beverly Hills, CA: Lucasfilm and Twentieth Century Fox.

St-Louis, V. et al. (2009) Satellite image texture and a vegetation index predict avian biodiversity in the Chihuahuan Desert of New Mexico, Ecography, 32, pp. 468–480. Available at: https://doi.org/10.1111/j.1600-0587.2008.05512.x.

St-Louis, V. et al. (2014) Modelling avian biodiversity using raw, unclassified satellite imagery, Philosophical Transactions of the Royal Society B: Biological Sciences, 369(1643), pp. 20130197. Available at: https://doi.org/10.1098/rstb.2013.0197.

Su, Y. et al. (2015) SRTM DEM Correction in Vegetated Mountain Areas through the Integration of Spaceborne LiDAR, Airborne LiDAR, and Optical Imagery, Remote Sensing, 7, pp. 11202–11225. Available at: https://doi.org/10.3390/rs70911202.

Swengel, A. and Swengel, S. (1999) Observations of prairie skippers (Oarisma Poweshiek, Hesperia dacotae, H. ottoe, H. leonardus pawnee, and Atrytone arogos iowa) (Lepidoptera: Hesperiidae) in Iowa, Minnesota and North Dakota during 1988-1997, The Great Lakes Entomologist, 32(4), pp. 267–292.

United States Geological Survey (2015) The Shuttle Radar Topology Mission (STRM) Collection User Guide. Version 3. United States Geological Survey. Available at: https://lpdaac.usgs.gov/documents/179/SRTM_User_Guide_V3.pdf (Accessed: 5 July 2023).

United States Geological Survey (2018) EarthExplorer: USGS: Science for a changing world, EarthExplorer. Available at: https://earthexplorer.usgs.gov/ (Accessed: 26 June 2018).

United States Geological Survey (2018) Landsat Missions: Landsat Enhanced Vegetation Index, Landsat Enhanced Vegetation Index. Available at: https://www.usgs.gov/landsat-missions/landsat-enhanced-vegetation-index.

US Fish and Wildlife Service (2021) Recovery Plan for the Dakota Skipper (Hesperia dacotae). U.S. Fish and Wildlife Service.

US Fish and Wildlife Service (2014) Endangered and threatened wildlife and plants; Threatened species status for Dakota Skipper and Endangered species status for Poweshiek Skipperling; Final Rule, 50 CFR Part 17(Laws/Statutes).

USGS (2024) Landsat Levels of Processing, Landsat Levels of Processing. Available at: https://www.usgs.gov/landsat-missions/landsat-levels-processing#L1GS (Accessed: 18 February 2024).

USGS and NASA (2019) Landsat 7 (L7) Data Users Handbook Version 2.0. LSDS-1927. Sioux Falls, South Dakota, pp. 151.

VanDerWal, J. et al. (2009) Selecting pseudo-absence data for presence-only distribution modeling: How far should you stray from what you know?, Ecological Modelling, 220(4), pp. 589–594. Available at: https://doi.org/10.1016/j.ecolmodel.2008.11.010.

Veregin, H. (1989) Taxonomy of error in spatial databases. Santa Barbara, Calif: NCGIA Publ (Book, Whole), pp. 109.

Villero, D. et al. (2017) Integrating species distribution modelling into decision-making to inform conservation actions, Biodiversity and Conservation, 26, pp. 21. Available at: https://doi.org/10.1007/s10531-016-1243-2.

Warren, D.L. et al. (2021) The effects of climate change on Australia's only endemic Pokémon: Measuring bias in species distribution models, Methods in Ecology and Evolution, 12(6), pp. 985–995. Available at: https://doi.org/10.1111/2041-210X.13591.

WCMC (1996) Hesperia dacotae. Available at: http://dx.doi.org/10.2305/IUCN.UK.1996.RLTS.T9968A13028912.en.

Webster, R. (2003) 2002 survey of the Dakota Skipper, Hesperia dacotae (Skinner), Canadian Wildlife Service, Prairie and Northern Region. Edmonton, Alberta: Environment Canada, p. 20+separate data sheets.

Webster, R. (2007) Dakota Skipper, Hesperia dacotae (Skinner), survey in southeast Saskatchewan and southwest Manitoba during 2007, Canadian Wildlife Service, Prairie and Northern Region. Edmonton, Alberta: Environment Canada, p. 27.

Weier, J. and Herring, D. (2000) Measuring vegetation (NDVI & EVI), Measuring vegetation (NDVI & EVI). Earth Observatory. NASA. Available at: https://earthobservatory.nasa.gov/features/MeasuringVegetation (Accessed: 13 January 2024).

Weiss, M. and Baret, F. (2016) S2ToolBox Level 2 products: LAI, FAPAR, FCOVER Version 1.1. European Space Agency. Available at: https://step.esa.int/docs/extra/ATBD_S2ToolBox_L2B_V1.1.pdf.

Westwood, A. (2017) Choosing a species distribution model for conservation, (Journal Article), p. 50.

Westwood, R. et al. (2019) A field-validated species distribution model to support management of the critically endangered Poweshiek skipperling (Oarisma poweshiek) butterfly in Canada, Conservation Science and Practice, 2(3), p. n/a. Available at: https://doi.org/10.1111/csp2.163.

Wintle, B.A., Elith, J. and Potts, J.M. (2005) Fauna habitat modelling and mapping: A review and case study in the Lower Hunter Central Coast region of NSW, Austral Ecology, 30(7), pp. 719–738. Available at: https://doi.org/10.1111/j.1442-9993.2005.01514.x.

Zhang, X. et al. (2022) Review of Land Surface Albedo: Variance Characteristics, Climate Effect and Management Strategy, Remote Sensing, 14(6). Available at: https://doi.org/10.3390/rs14061382.

Zhao, X. et al. (2018) A global corrected SRTM DEM product for vegetated areas, Remote Sensing Letters, 9, pp. 393–402. Available at: https://doi.org/10.1080/2150704X.2018.1425560.

Zvoleff, A. (2019) Calculating image textures with GLCM, Calculating image textures with GLCM. Available at: http://azvoleff.com/articles/calculating-image-textures-with-glcm/ (Accessed: 7 June 2019).

APPENDICES

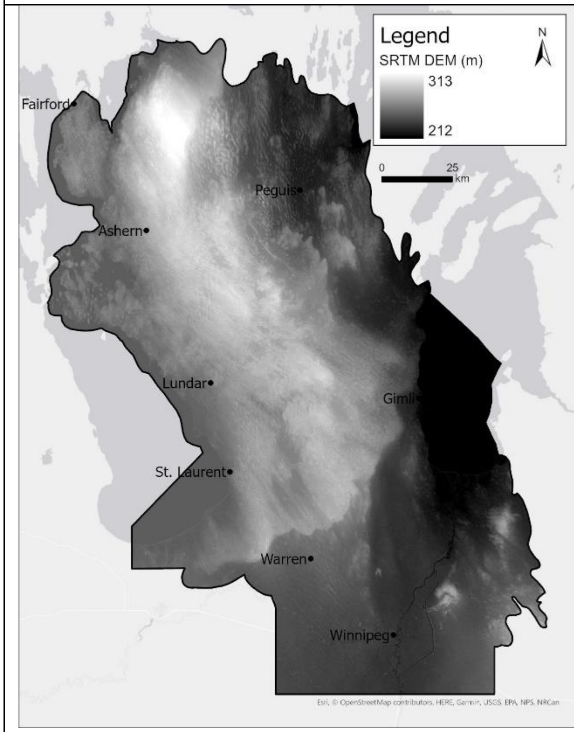
Appendix 1. Dakota skipper survey references

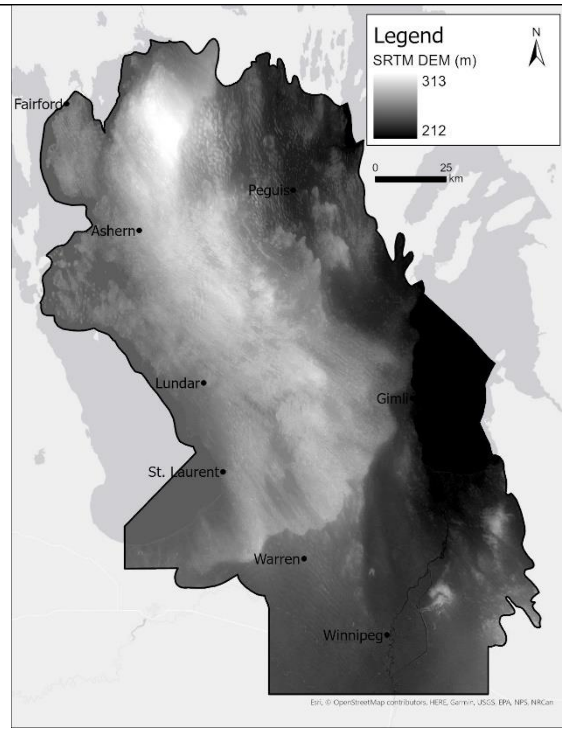
Year	Reference
1973	Hooper, R.R. 1973. The Butterflies of Saskatchewan. Museum of Natural History, Regina, Saskatchewan. 216 pp.
1989	Klassen,P.,Westwood, A.R., Preston. W.B. and W.B. McKillop. 1989. The Butterflies of Manitoba. Manitoba Museum of Man and Nature. Winnipeg. 290 pp.
1995	David Delf-personal communication, Nov 7, 1995.
1995	Klassen, P. 1995. Personal communications (by phone) re locations and dates of Powesheik and Dakota Skipper occurrence (collection) records. Mimeo. 10 pp.
2001	Britten, Hugh. 2001. Conservation genetics of the Dakota Skipper (Hesperia dacotae). Department of Natural Resources. Natural Heritage and Nongame Research Program. 26 pp.
2002	Britten, H. B., and J. W. Glasford. 2002. Genetic population structure of the Dakota Skipper (Lepidoptera: Hesperia dacotae): a North American native prairie obligate. Conservation Genetics 3:363-374.
2002	Walleyn, A. 2002. Habitat Use By Prairie Bird Species.
2002	Britten, Hugh. 2002(?). Personal Communication to Manitoba Conservation Data Centre. Numbers and locations of Dakota Skipper (Hesperia dacotae) captured and released near Tolstoi, Manitoba.
2003	Webster, Reginald, 2003. 2002 survey of the dakota skipper, Hesperia dacota (Skinner). Microsoft Word 20pp. + report and data tables in Microsoft Excel.
2006	Morden, C-Jae. 2006. Potential reintroduction of the Dakota skipper (Hesperia dacotae) into south-eastern Manitoba. Honours Thesis, Dept. of Biology, University of Winnipeg.
2006	Richard Westwood. 2006. Survey of butterflies and skippers on W30-17-3W.
2007	Bates, Lara. 2007. The influence of prescribed burning and grazing on the Dakota Skipper, Hesperia dacotae, habitat in south-eastern Manitoba. Honors Thesis, Department of Biology, University of Winnipeg.
2007	Webster, Reginald P. 2007. Dakota Skipper, Hesperia dacotae (Skinner), survey in southeast Saskatchewan and southwest Manitoba during 2007. Prepared for Canadian Wildlife Service, Prairie and Northern Region, Environment Canada, Edmonton, AB.
2009	Manitoba Conservation Data Centre. 2009. Skipper records identified from netted specimens collected by the Manitoba Conservation Data Centre. 2 records. In Microsoft Excel format.
2012	C. Rigney. 2010-2012. Dakota skipper observeations from 2010 to 2012.
2013	Richard Westwood. 2013. Field surveys for Dakota Skipper during 2013.
2016	Westwood, Richard. 2016. Personal communication to Colin Murray regarding Dakota Skipper surveys in Riding Mountain National Park. Email correspondence.
2018	Canadian National Collection of Insects, Arachnids, and Nematodes. 2018. Search of the CNC database for Hesperia dacotae (dakota skipper) specimens. Conducted on: 2018MAY25
2018	Murray, C., Carla Church, and Chris Friesen. 2018. Observations from 2018 field season. Unpublished data. Manitoba Conservation Data Centre, Wildlife Branch, Manitoba Conservation.

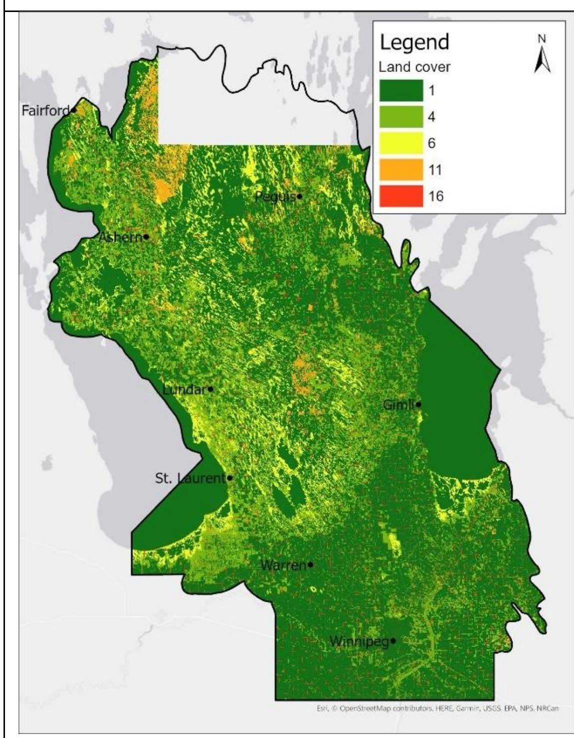
2020	Church, C., and Alanah Jones. 2020. Photos of skippers observed during Dakota Skipper Survey. Manitoba Conservation Data Centre, Wildlife Branch, Manitoba Conservation.
2020	Church, C., Alanah Jones, Kathryn Yarchuk, Colin Murray and Chris Friesen. 2020. Observations from 2020 field season: GPS points ESRI shapefiles, and field notes. Unpublished data. Manitoba Conservation Data Centre, Wildlife Branch, Manitoba Conservation.
2020	Westwood, R. 2020. Personal communications, skipper identification of photos taken during Dakota skipper surveys.
2021	Henault, Justis. 2021. Lepidopteran surveys Manitoba, 2021 report, data, locations, and tracks. PDF, excel, and kml files.

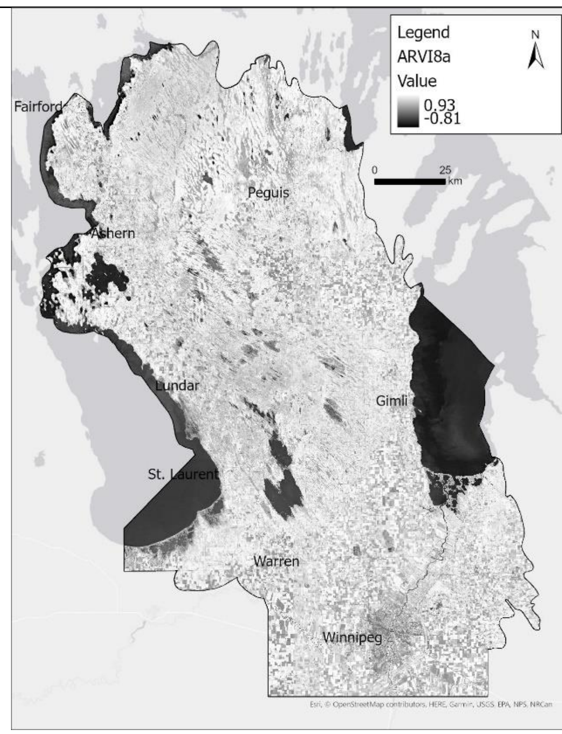
Appendix 2. Common Model and Satellite Model variable maps

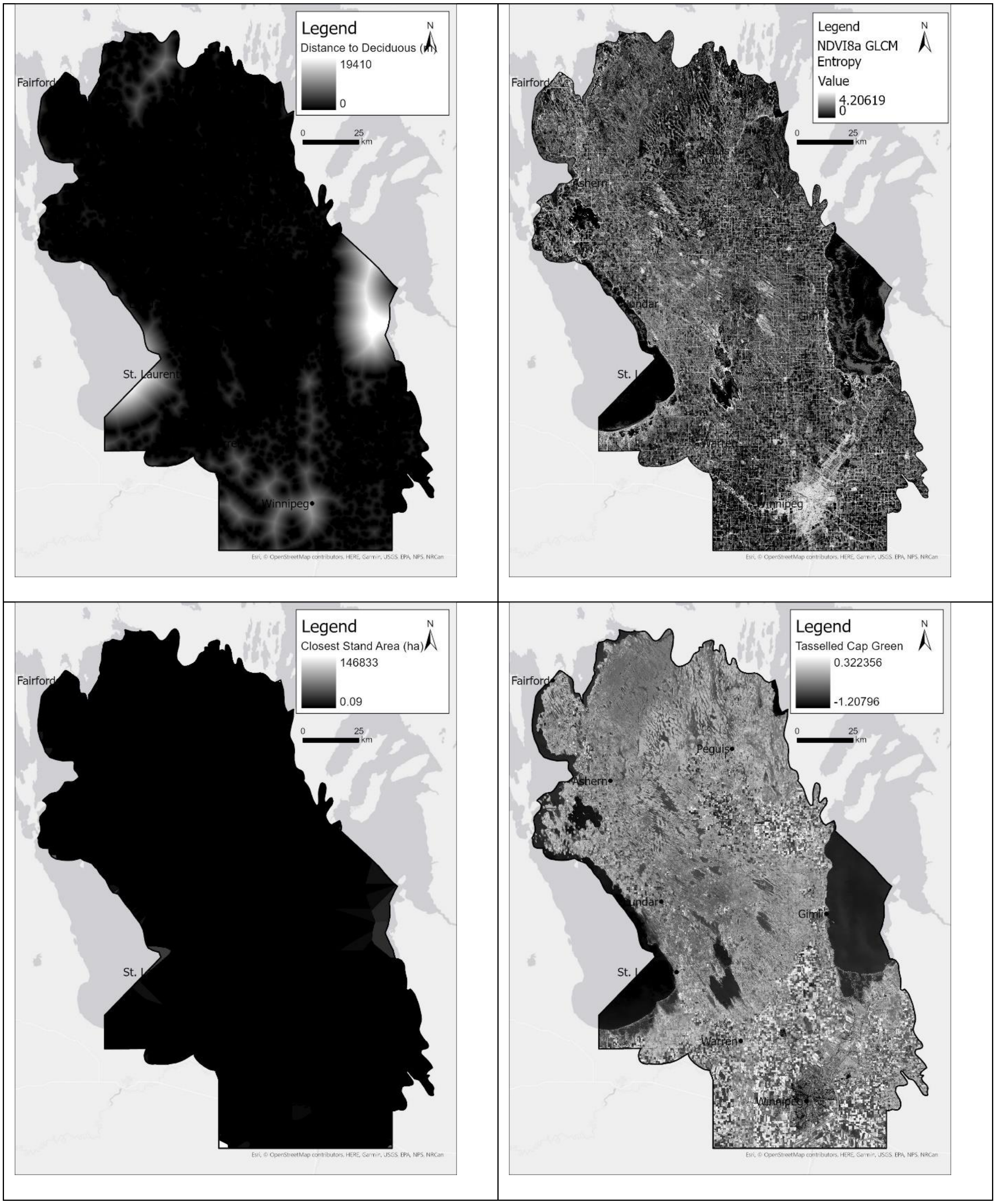
Common Model Variables	Satellite Model variables

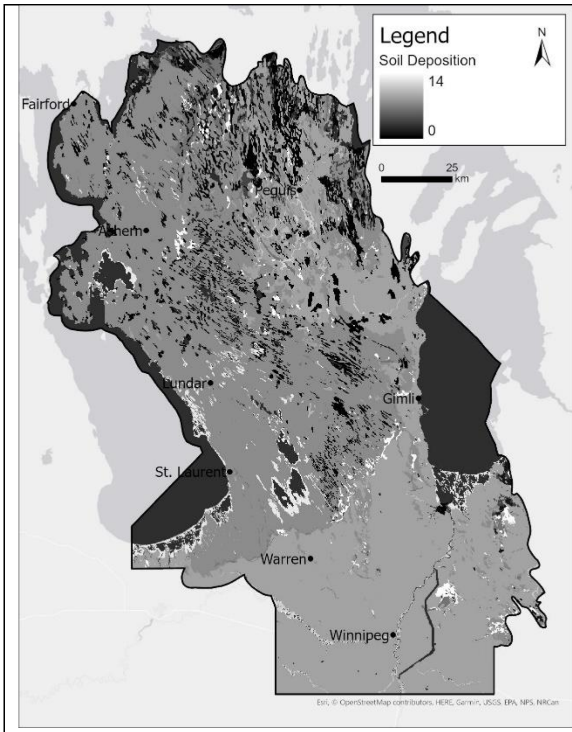


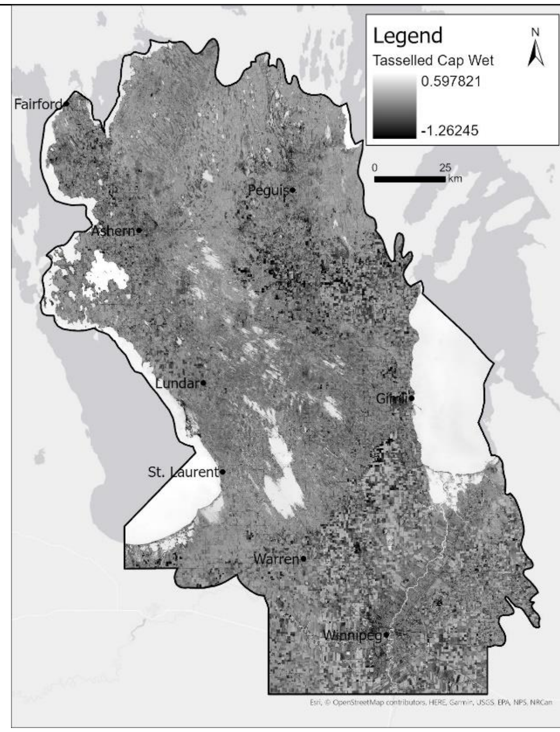


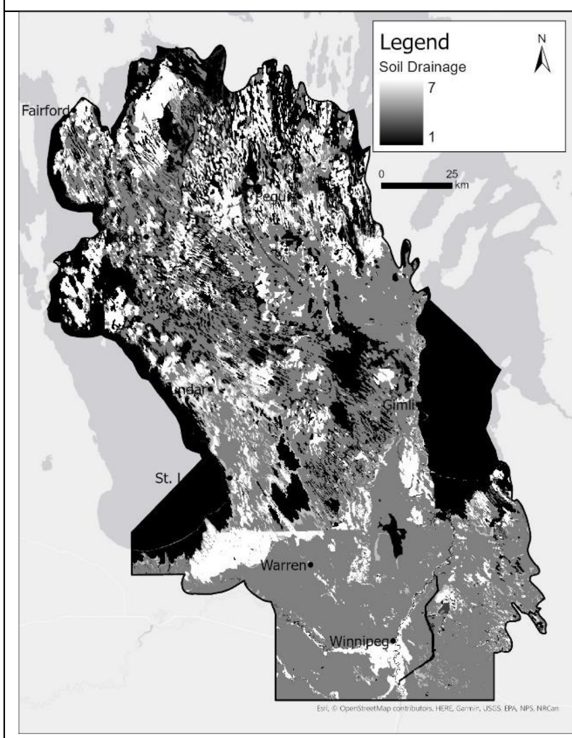


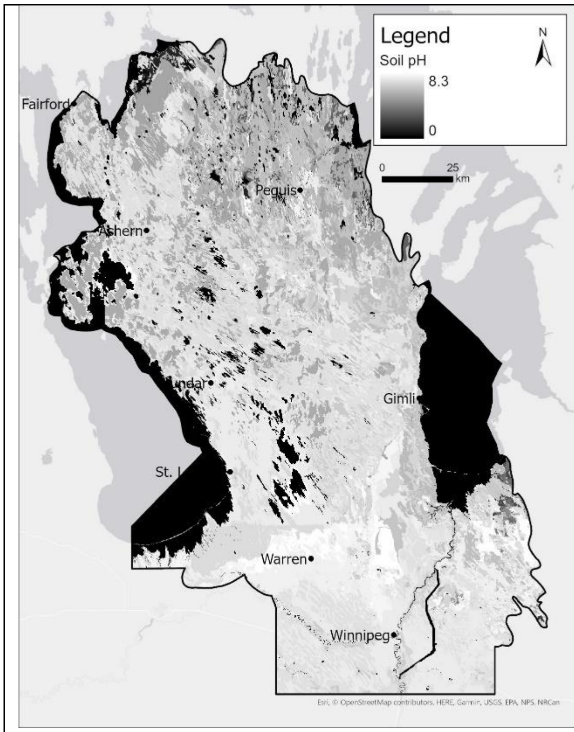



Legend
Distance to Deciduous (m)
19410
0
0 25 km
Fairford
St. Laurent
Winnipeg
Legend
NDVI8a GLCM Entropy
Value
4.20619
0
Fairford
Legend
Closest Stand Area (ha)
146833
0.09
Fairford
Legend
Tasselled Cap Green
0.322356
-1.20796
Fairford
Peguis
Ashern
Gimli
Warren
Esri, © OpenStreetMap contributors, HERE, Garmin, USGS, EPA, NPS, NRCan

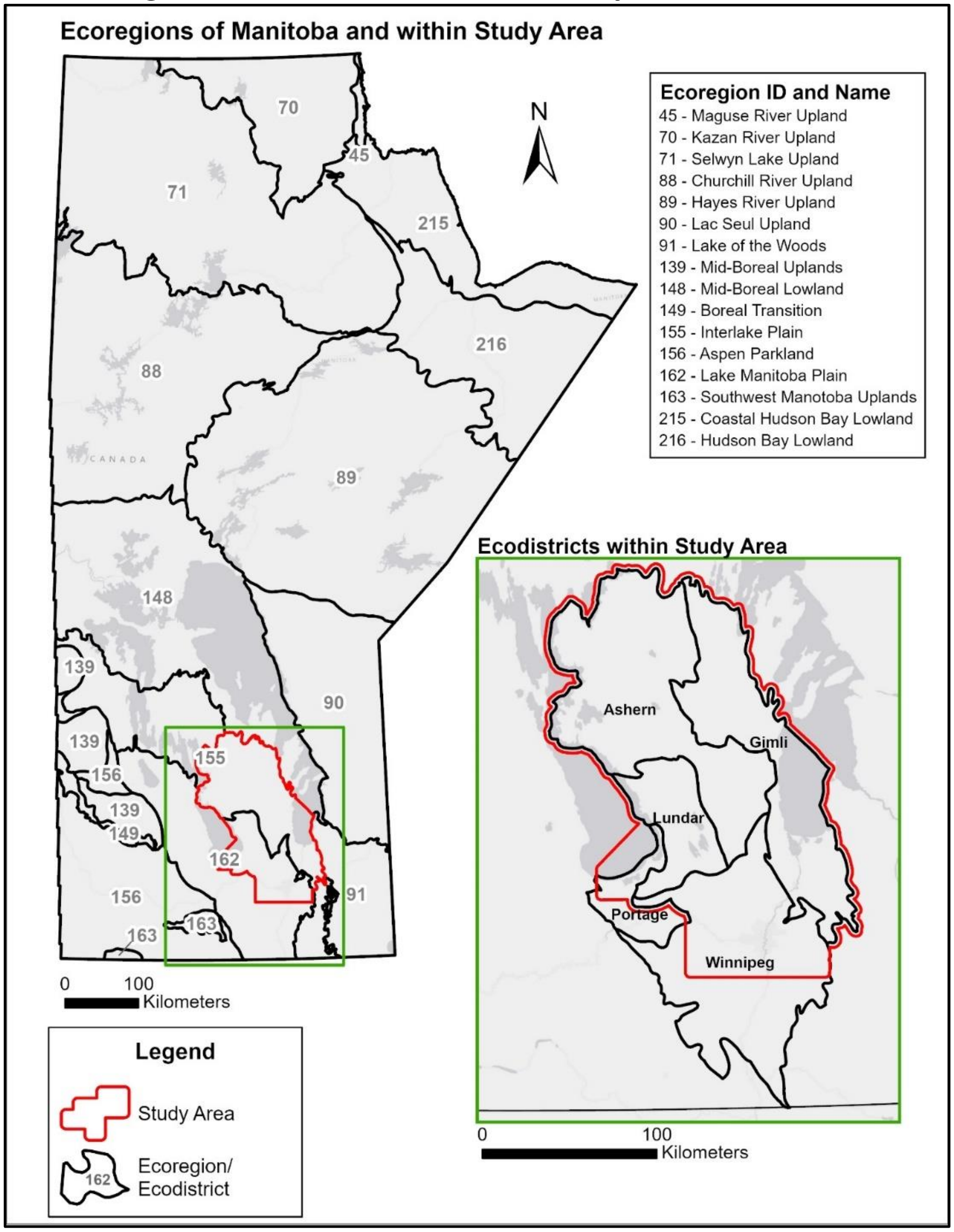

Legend
Soil Deposition
14
0
N
0
25
km
Fairford
Peguis
Ashern
Lundar
Gimli
St. Laurent
Warren
Winnipeg
Esri, © OpenStreetMap contributors, HERE, Garmin, USGS, EPA, NPS, NRCan


Legend
Tasselled Cap Wet
0.597821
-1.26245
N
0
25
km
Fairford
Peguis
Ashern
Lundar
Gimli
St. Laurent
Warren
Winnipeg
Esri, © OpenStreetMap contributors, HERE, Garmin, USGS, EPA, NPS, NRCan


Legend
Soil Drainage
7
1
N
0
25
km
Fairford
Peguis
Gimli
Warren
Winnipeg
Esri, © OpenStreetMap contributors, HERE, Garmin, USGS, EPA, NPS, NRCan


Legend
Soil pH
8.3
0
N
0
25
km
Fairford
Peguis
Ashern
Gimli
Warren
Winnipeg
Esri, © OpenStreetMap contributors, HERE, Garmin, USGS, EPA, NPS, NRCan

Appendix 3. Ecoregions and Ecodistricts within the study area



Appendix 4. Manitoba land use land cover classification system

1. Agricultural Cropland	Lands dedicated to the production of annual cereal, seed and specialty crops. These lands would normally be cultivated on an annual basis.
2. Deciduous Forest	Forest lands where 75% to 100% of the tree canopy is deciduous. Dominant species include trembling aspen (Populus tremuloides), balsam poplar (Populus balsamifera), and white birch (Betula papyrifera). May encompass small patches of grassland, marsh or fens less than two hectares in size. Dense forest canopy (>60%), open canopy (26-60%), sparse canopy (10-25%).
3. Water Bodies	Consists of all open water – lakes, rivers, streams, ponds, lagoons
4. Grassland/ Rangeland	Lands of mixed native and/or tame prairie grasses and herbaceous vegetation. May also include scattered stands of shrub such as willow (Salix spp.), choke-cherry (Prunus spp.), saskatoon (Amelanchier spp.) and pincherry (Prunus spp.). Lands may be used for the harvesting of hay while others may be grazed. Both upland and lowland meadows fall into this class. There is normally (<10%) shrub or tree canopy.
5. Mixedwood Forest	Forest lands where 26% to 74% of the tree canopy is coniferous. May encompass treed bogs, marsh or fens less than two hectares in size. Dense forest canopy (>60%), open canopy (26-60%), sparse canopy (<26%).
6. Marsh	Wetlands comprised of various herbaceous species. Wetlands range from intermittent inundated (temporary, seasonal, semi-permanent) to permanent depending on the current annual precipitation. Common vegetation species include; sedge (Carex spp.), whitetop (Scolochloa festucacea), giant reed grass (Phragmites australis), prairie cordgrass (Spartina pectinata), mannagrass (Glyceria spp.), slough grass (Beckmannia spp.), cattail (Typha spp.), and bulrush (Scirpus spp.).
7. Bogs	Wetlands dominated by bryoid-mosses (ie. Spaghnum spp.) and ericaceous shrubs such as labrador tea (Ledum spp.). Tamarack (Larix larcina) and black spruce (Picea mariana) are also found with a sparse to dense (10 – 100%) canopy.
8. Treed Rock	Lands of exposed bedrock with less than 60% tree canopy. The dominant tree species include jack pine (Pinus banksiana), and/or black spruce (Picea mariana) with shrub cover such as alder (Alnus spp.). Open canopy (26-60%), sparse canopy (10-25%).
9. Coniferous Forest	Forest lands where 75-100% of the tree canopy is coniferous. Jack Pine, white spruce (Picea glauca) and black spruce are the dominant species in this class. May include patches of treed bog, marsh and/or fens less than two hectares in size. Dense forest canopy (>60%), open canopy (26-60%), sparse canopy (10-25%).

10. Wildfire Areas	Forest lands that have been recently burnt (< 5 years) with sporadic regeneration and can include pockets of unburned trees.
11. Open Deciduous Forest/Shrub	Lands characterized by shallow soils and/or poor drainage which support mainly a cover of shrubs such as willow (Salix spp.), alder (Alnus spp.), Saskatoon (Amelanchier spp.) and/or stunted trees such as trembling aspen, balsam poplar and birch with a sparse (10-25%) to open canopy (26-60%).
12. Forage Crops	Agricultural lands used in the production of forage such as alfalfa and clover or blends of these with tame species of grass. Fall seeded crops such as winter wheat or fall rye may be included here.
13. Cultural Features	Cities, towns, villages and communities with place names. Also includes; cemeteries, shopping centres, large recreation sites, autowreck yards, airports, cottage areas, race tracks and rural residential.
14. Forest Cutovers	Forest lands where commercial timber have been completely or partially removed by logging operations.
15. Bare Rock/Gravel/Sand	Lands of exposed bedrock, gravel and/or sand dunes and beaches with less than 10% vegetation. Also includes gravel quarry/pit operations, mine tailings, borrow pits and rock quarries.
16. Roads/Trails	Highways, secondary roads, trails and cut survey lines or right-of-ways such as railways and transmission lines.
17. Fen	Wetlands with nutrient-rich, minerotrophic water, and organic soils composed of the remains of sedges (Carex spp.) and/or mosses (Drepanocladus spp.), where sedges, grasses, reeds and moss predominate but could include shrub and sparse tree canopy of black spruce and/or tamarack. Much of the vegetative cover of fens would be similar to the vegetation zones of marshes.
18. Lichen Heath	Lands characterized by an abundance of lichen (C. alpestris, C. mitis, C. rangerferina) and heath vegetation (L. decumbens, V. vitis-idaea, V. uglinosum, E. nigrum) located on well drained summits and upper slopes. The forest canopy is sparse (< 10%) with the dominant tree being black spruce. Lichen heath is typically found in the taiga shield ecozone.

(Manitoba Remote Sensing Centre 2006)

Appendix 5. Model candidate variables created from Sentinel 2A imagery and SRTM DEM.

Var. Num.	Group Run Num.	Group Run Name	Candidate Var. Name Abbrev.	Candidate Var. Name	Remote Sensing Platform	Imagery Used	Software	Source
1	1	Water/Moisture	NDWI2	Normalized Difference Water Index	Sentinel 2	SuperRes10m	SNAP	(Du *et al.* 2016)
2	1	Water/Moisture	MNDWI	Modified Normalized Difference Water Index	Sentinel 2	SuperRes10m	SNAP	(Du *et al.* 2016)
3	1	Water/Moisture	TcapWET06	Tasselled Cap Wetness 6 Bands	Sentinel 2	C1_simpsamp10mres	SNAP	(Shi and Xu 2019)
4	1	Water/Moisture	TcapWET13	Tasselled Cap Wetness 13 Bands	Sentinel 2	C1_simpsamp10mres	SNAP	(Shi and Xu 2019)
5	2	Vegetation	NDVI8	Normalized Difference Vegetation Index	Sentinel 2	SuperRes10m	SNAP	(Lange *et al.* 2017)
6	2	Vegetation	NDVI8a	Normalized Difference Vegetation Index	Sentinel 2	SuperRes10m	SNAP	(Lange *et al.* 2017)
7	2	Vegetation	ARVI8	Atmospherically resistant Vegetation Index	Sentinel 2	SuperRes10m	SNAP	(Kaufman and Tanre 1992)
8	2	Vegetation	ARVI8a	Atmospherically resistant	Sentinel 2	SuperRes10m	SNAP	(Kaufman and Tanre 1992)

				Vegetation Index				
9	2	Vegetation	NLI	NonLinear Vegetation Index	Sentinel 2	SuperRes10m	SNAP	Indexdatabase.de and Sentinel hub custom scripts
10	2	Vegetation	TcapGVI06	Tasselled Cap Greeness or vegetation 6 Bands	Sentinel 2	C1_simpsamp10mres	SNAP	(Shi and Xu 2019)
11	2	Vegetation	TcapGVI13	Tasselled Cap Greeness or vegetation 13 Bands	Sentinel 2	C1_simpsamp10mres	SNAP	(Shi and Xu 2019)
12	2	Vegetation	EVI	Enhanced Vegetation Index	Sentinel 2	SuperRes10m	SNAP	(Huete *et al.* 1994)
13	2	Vegetation	MSAVI2B8	Modified Soil Adjusted Vegetation Index 2	Sentinel 2	SuperRes10m	SNAP	(Qi *et al.* 1994)
14	2	Vegetation	MSAVI2B8a	Modified Soil Adjusted Vegetation Index 2	Sentinel 2	SuperRes10m	SNAP	(Qi *et al.* 1994)
15	3	DEM	SRTM DEM	SRTM DEM	SRTM	SRTMDEMRes10m	ARC	(Farr *et al.* 2007)
16	3	DEM	SRTM DEM Slope	SRTM DEM Slope	SRTM	SRTMDEMRes10m	ARC	(Farr *et al.* 2007)
17	3	DEM	SRTM DEM Aspect	SRTM DEM Aspect	SRTM	SRTMDEMRes10m	ARC	(Farr *et al.* 2007)

18	3	DEM	SRTM DEM Curve	SRTM DEM Slope Curvature	SRTM	SRTMDEMRes10m	ARC	(Farr *et al.* 2007)
19	3	DEM	SRTM DEM SCPara	SRTM DEM Slope Curvature Parallel	SRTM	SRTMDEMRes10m	ARC	(Farr *et al.* 2007)
20	3	DEM	SRTM DEM SCPerp	SRTM DEM Slope Curvature Perpendicular	SRTM	SRTMDEMRes10m	ARC	(Farr *et al.* 2007)
21	4	Brightness	TcapSBI06	Tasselled Cap Brightness 6 Bands	Sentinel 2	C1_simpsamp10mres	SNAP	(Shi and Xu 2019)
22	4	Brightness	TcapSBI13	Tasselled Cap Brightness 13 Bands	Sentinel 2	C1_simpsamp10mres	SNAP	(Shi and Xu 2019)
23	4	Brightness	Albedo	Albedo	Sentinel 2	SuperRes10m	SNAP	(Lahoz-Monfort *et al.* 2010)
24	4	Brightness	BI	Brightness Index	Sentinel 2	SuperRes10m	SNAP	(Escadafal *et al.* 1989)
25	4	Brightness	BI2	Second Brightness Index	Sentinel 2	SuperRes10m	SNAP	(Escadafal *et al.* 1989)
26	5	Insolation	TRADFLY	Total Radiation Flight Time	SRTM	SRTMDEMRes30m to 10m	ARC	(Fu and Rich 2000)
27	5	Insolation	TRADSUM	Total Radiation Summer time	SRTM	SRTMDEMRes30m to 10m	ARC	(Fu and Rich 2000)

28	5	Insolation	TRADAN	Total Radiation Total Annual	SRTM	SRTMDEMRes30m to 10m	ARC	(Fu and Rich 2000)
29	5	Insolation	Winter	Total Radiation Winter time	SRTM	SRTMDEMRes30m to 10m	ARC	(Fu and Rich 2000)
30	6	Biophysical Processor	LAI	Leaf Area Index	Sentinel 2	L2ASimpSamp	SNAP	(Weiss and Baret 2016)
31	6	Biophysical Processor	FAPAR	Fraction of Absorbed Photosynthetically Active Radiation	Sentinel 2	L2ASimpSamp	SNAP	(Weiss and Baret 2016)
32	6	Biophysical Processor	FVC	Fraction of Vegetation Cover	Sentinel 2	L2ASimpSamp	SNAP	(Weiss and Baret 2016)
33	7	GLCM	glcm_ent	Entropy	Sentinel 2	NDVI8a	R	(Zvoleff 2019)
34	7	GLCM	glcm_secm	Second Momentum	Sentinel 2	NDVI8a	R	(Zvoleff
35	7	GLCM	glcm_homo	Homogeneity	Sentinel 2	NDVI8a	R	(Zvoleff
36	7	GLCM	glcm_dis	Dissimilarity	Sentinel 2	NDVI8a	R	(Zvoleff
37	7	GLCM	glcm_con	Contrast	Sentinel 2	NDVI8a	R	(Zvoleff
38	7	GLCM	glcm_mean	Mean	Sentinel 2	NDVI8a	R	(Zvoleff
39	7	GLCM	glcm_var	Variance	Sentinel 2	NDVI8a	R	(Zvoleff
40	7	GLCM	glcm_cor	Correlation	Sentinel 2	NDVI8a	R	(Zvoleff

Appendix 66. Detailed list of indicator plants and suitability scores.

Type	Common Name	Scientific Name	Highly abundant excellent habitat	Indicator of excellent habitat	Adult nectar or larval food species	Native to North America	Suitability Score
Forb	Yarrow	*Achillea millefolium*	0	0	1	1	2
Forb	False dandelion	*Agoseris glauca*	0	1	1	1	3
Graminoid	Quackgrass	*Elymus repens*	0	0	0	0	0
Graminoid	Slender wheatgrass	*Elymus trachycaulus*	0	1	0	1	2
Graminoid	Creeping bentgrass	*Agrostis stolonifera*	1	0	0	0	1
Graminoid	Big bluestem	*Andropogon gerardi*	1	1	1	1	4
Shrub	Kinnikinnick	*Arctostaphylos uva-ursi*	0	1	0	1	2
Forb	Milkweed	*Asclepias* spp.	0	0	1	1	2
Forb	Milkvetch	*Astragalus* spp.	0	0	1	1	2
Graminoid	Smooth brome	*Bromus inermis*	0	0	0.5	0	0.5
Graminoid	Nodding brome	*Bromus porteri*	0	1	0	1	2
Forb	Harebell	*Campanula rotundifolia*	1	0	1	1	3
Forb	Hairy goldenaster	*Heterotheca villosa*	0	0	1	1	2
Forb	Canada thistle	*Cirsium arvense*	0	0	1	0	1
Forb	Flodman’s thistle	*Cirsium flodmanii*	0	0	1	1	2
Forb	Dandelion Hawsbeard	*Crepis runcinata*	1	1	1	1	4
Graminoid	Tufted hair grass	*Deschampsia cespitosa*	1	1	1	1	4
Graminoid	Prairie dropseed	*Sporobolus heterolepis*	1	1	1	1	4
Shrub	Wolf willow	*Elaeagnus commutata*	0	1	0	1	2
Graminoid	Marsh spikerush	*Eleocharis palustris*	1	0	0.5	1	2.5
Forb	Fleabane	*Erigeron* spp.	0	0	1	1	2
Shrub	Wild strawberry	*Fragaria* spp.	1	1	0	1	3
Forb	Great Blanketflower	*Gaillardia aristata*	0	0	1	1	2
Forb	Northern bedstraw	*Galium boreale*	0	1	0	1	2
Forb	Wild Licorice	*Glycyrrhiza lepidota*	1	1	0	1	3
Forb	Sunflower	*Helianthus spp.*	0	1	0	1	2

Graminoid	Porcupine grass	*Hesperostipa spartea*	0	0	1	1	2
Forb	Umbellate Hawkweed	*Hieracium umbellatum*	0	1	0	1	2
Forb	Yellow Stargrass	*Hypoxis hirsuta*	0	1	0	1	2
Forb	Tall Blue Lettuce	*Mulgedium pulchellum*	0	0	1	1	2
Forb	Meadow Blazingstar	*Liatris ligulistylis*	0	1	0	1	2
Forb	Wood lily	*Lilium philadelphicum*	0	0	1	1	2
Forb	Lobelia	*Lobelia* spp.	1	0	1	1	3
Forb	Sea milkwort	*Lysimachia maritima*	0	1	0	1	2
Forb	Black medick	*Medicago* spp.	0	0	1	0	1
Forb	White sweet-clover	*Melilotus* spp.	0	0	1	0	1
Graminoid	Mat muhly	*Muhlenbergia richardsonis*	0	0	0	1	1
Graminoid	Switchgrass	*Panicum virgatum*	0	1	0.5	1	2.5
Forb	Grass of parnassus	*Parnassia glauca*	1	1	0	1	3
Forb	Swamp lousewort	*Pedicularis lanceolata*	0	1	0	1	2
Forb	Prairie-clover	*Dalea* spp.	1	1	1	1	4
Graminoid	Common timothy	*Phleum pratense*	0	0	0	0	0
Graminoid	Grass	*Poa* spp.	0	0	0.5	0.5	1
Forb	Seneca Snakeroot	*Polygala senega*	0	1	0	1	2
Shrub	Silverweed	*Potentilla anserina*	1	0	0	1	2
Shrub	Shrubby cinquefoil	*Dasiphora fruticosa*	0	1	0	1	2
Forb	Seaside Crowfoot	*Halerpestes cymbalaria*	1	0	0	1	2
Forb	Upright Prairie Coneflower	*Ratibida columnifera*	0	0	1	1	2
Forb	Black-eyed Susan	*Rudbeckia hirta*	1	0	1	1	3
Graminoid	Little bluestem	*Schizachyrium scoparium*	1	1	1	1	4
Forb	Upland White Goldenrod	*Solidago ptarmicoides*	1	1	1	1	4
Forb	Goldenrod	*Solidago* spp.	1	0	0	1	2
Forb	Field Sow-thistle	*Sonchus arvensis*	1	0	0	0	1
Forb	Rush aster	*Symphotrychum boreal*	1	0	0	1	2

Forb	Many-flowered aster	*Symphyotrichum ericoides*	1	0	0	1	2
Forb	Smooth aster	*Symphyotrichum laeve*	0	0	0	1	1
Forb	Clover	*Trifolium* spp.	0	0	1	0	1
Forb	White camas	*Anticlea elegans*	0	1	1	1	3
Graminoid	Redtop	*Agrostis gigantea*	0	0	0	0	0
Graminoid	Golden sedge	*Carex aurea*	0	0	0.5	1	1.5
Shrub	Blue honeysuckle	*Lonicera caerulea*	0	0	0	1	1
Forb	Evening primrose	*Oenothera biennis*	0	0	1	0	1
Graminoid	River grass	*Scolochloa festucacea*	0	0	0	1	1
Forb	Blue-eyed grass	*Sisyrincium montanum*	0	0	0	1	1
Forb	Tall meadow rue	*Thalictrum dasycarpum*	0	0	0	1	1
Forb	Vetch	*Vicia* spp.	0	0	0	1	1

Each indicator plant was scored on the following four criteria: (1) Highly abundant in excellent habitat, (2) An indicator of excellent habitat, (3) An adult nectar or larval food source, (4) Native to North America. The plant was awarded one point for each criteria it fulfilled, half a point if it partially fulfilled the criteria, and no point if it the criteria didn't apply. The points awarded for each plant were summed providing a final overall suitability score. Detailed indicator plant list and scores listed in Appendix 1 (from Rigney 2013; Dearborn *et al.* 2022).

Appendix 77. Detail of Site Sample Design.

Site evaluation consisted of employing either a detailed or rapid assessment protocol at each survey site. Detailed assessments were conducted when an initial walk-about indicated that a site had high native floristic diversity and was not severely disturbed by cattle or not agriculture cropland. Detailed and rapid assessments were modified between Common and Satellite due to the difference in pixel size. The intent was to establish a smaller quadrat square within the larger pixel to mitigate influence from the adjacent pixels when sampling for indicator plant species (Figure 6).

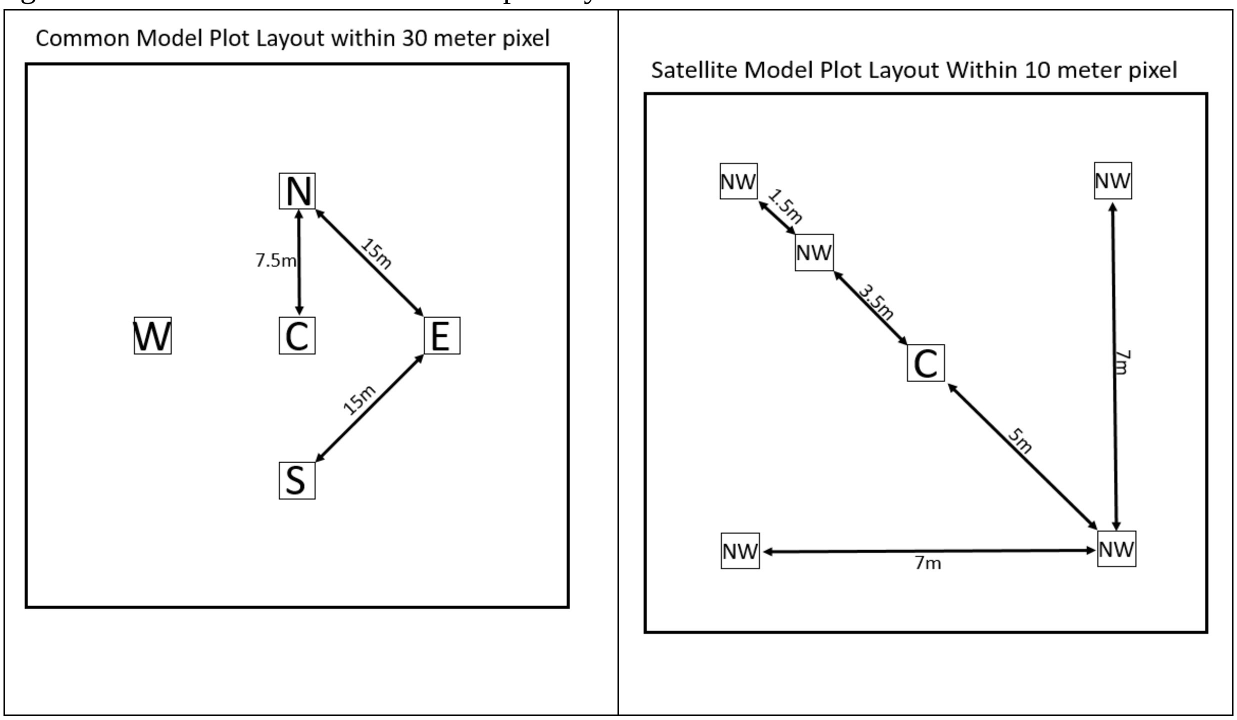
Common Model detailed methods consisted of navigating to the point site using GPS and marking the centroid location. Four other locations were established 7.5m from the centroid at the four cardinal directions. A 1x1m quadrat was then centered over each cardinal location and presence of indicator plant species was recorded. Finally, a five-minute timed walk-about was conducted within the established 15 x 15 m square recording any indicator plant species encountered.

Detailed assessment protocol for the Satellite Model sites consisted of marking the centroid location and marking four other locations 5m from the centroid location to the cardinal directions (northwest, southwest, southeast, southwest) making a 7 x 7 m square. An additional sixth location was marked at 3.5 m from the centroid along the northwest axis. Indicator plant presence-absence was evaluated at the two northwest points and the centroid point using a 1m x 1m quadrat. A five-minute timed walk-about survey was then done within the 5 x 5 m area.

Rapid assessment for the Satellite Model involved sampling the northwest point 3.5 m from the centroid and then the five-minute timed walk-about survey in the 7 x 7 m square. Rapid assessment for the Common Model entailed sampling for indicator plants at the north quadrat point and then the five-minute walk-about within the 15 x15 m square.

Survey forms were designed in ESRI Survey123 (ESRI (Environmental Systems Research Institute) 2018) to collect site information, and digital maps were created for navigation and to keep track of survey status for each site and overall progress. Photographs were captured of each site at the south corner looking north, east, and west. Aerial photographs were recorded using a DJI Phantom 4 UAV deployed at each site, flight conditions permitting, to photograph the site at 30 m, 50 m, and 100 m altitude.

Figure 6. Common and Satellite Model plot layout



Appendix 8. Common Model and Satellite Model run output

Table 12. Common Model environmental variable relative contribution to the model.

Variable	Percent Contribution
Landuse	34.4
Soil mode of deposition	20.7
SRTM DEM	17.1
Soil drainage	13.7
Distance to deciduous forest	7.5
Soil Ph	4.9
Closest forest stand area	1.7

Table 13. Satellite Model environmental variable relative contribution to the model.

Variable	Percent Contribution
SRTM DEM	44.7
NDVI 8a	24.5
Tasseled cap wetness	20.0
Tasseled cap greenness	9.8
ARVI 8a	1.0

Figure 7. Common Model Jackknife test for training and test data

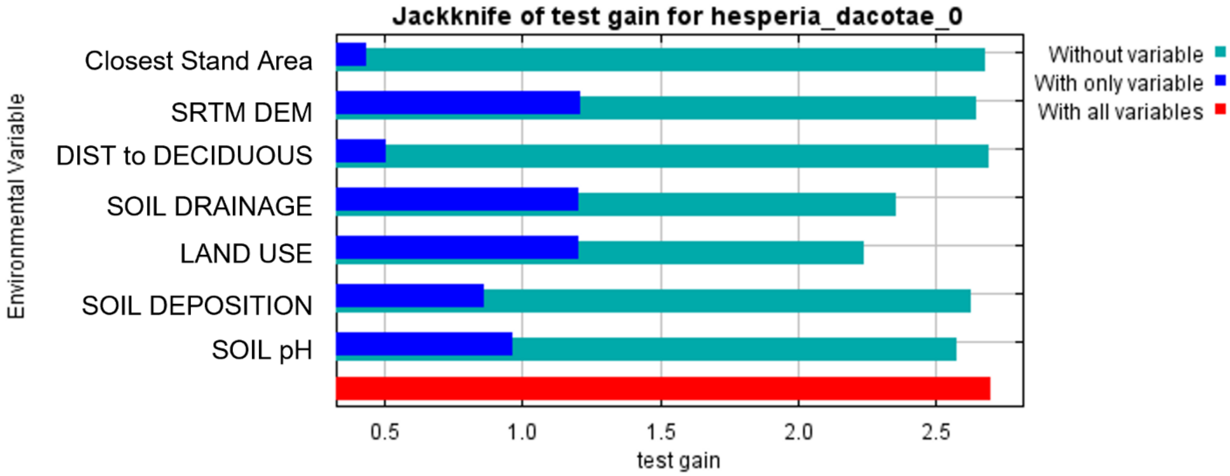


Figure 8. Satellite Model Jackknife test for training and test data

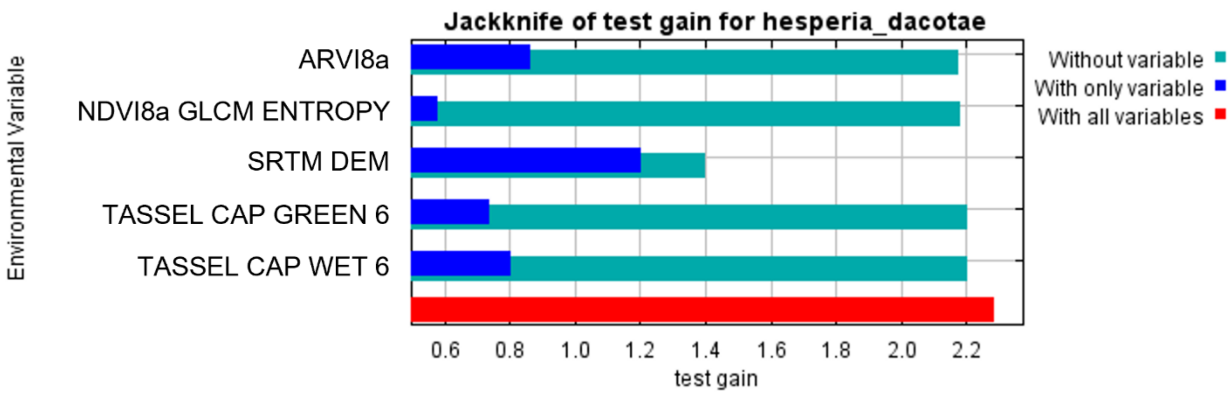
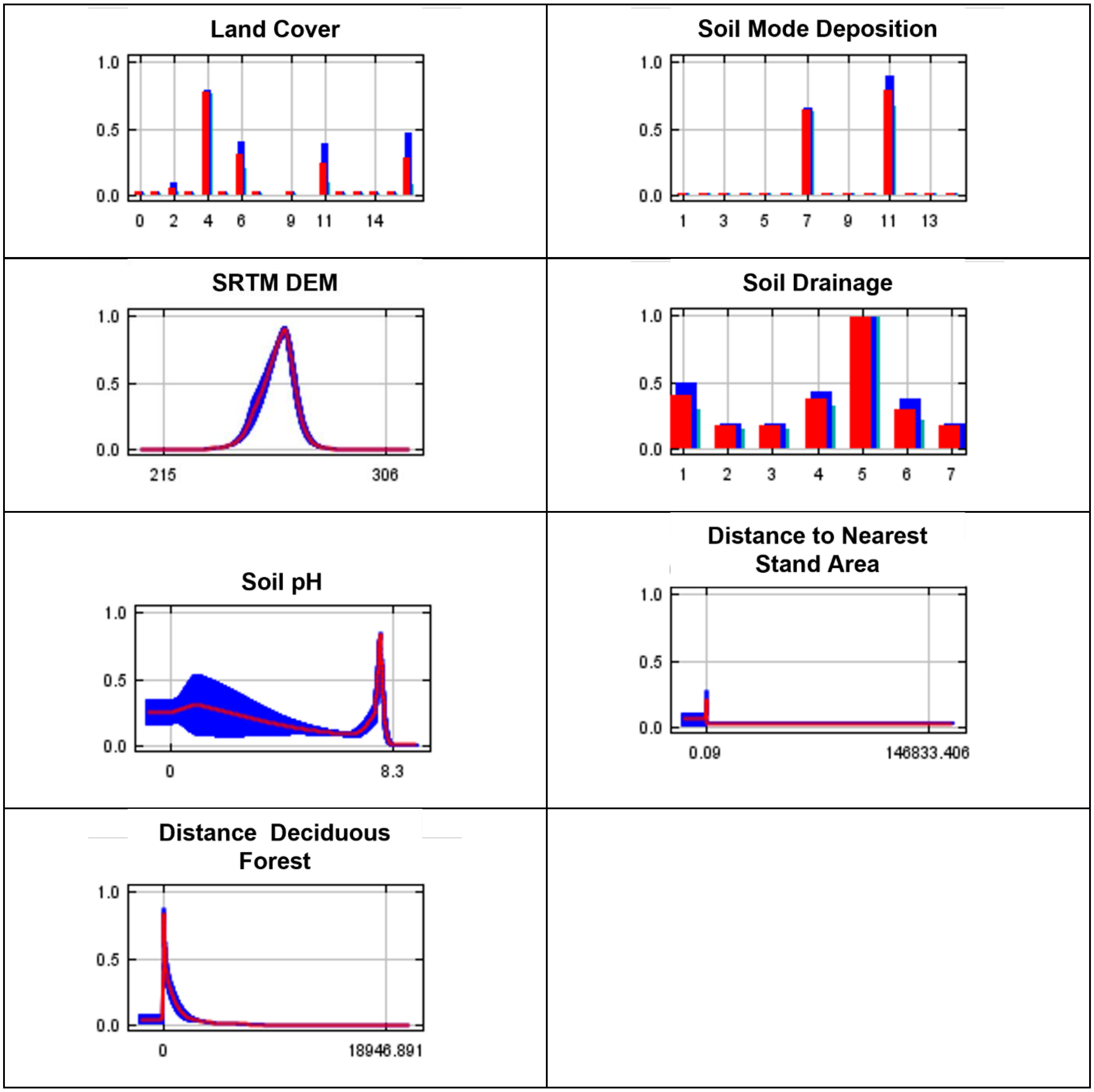
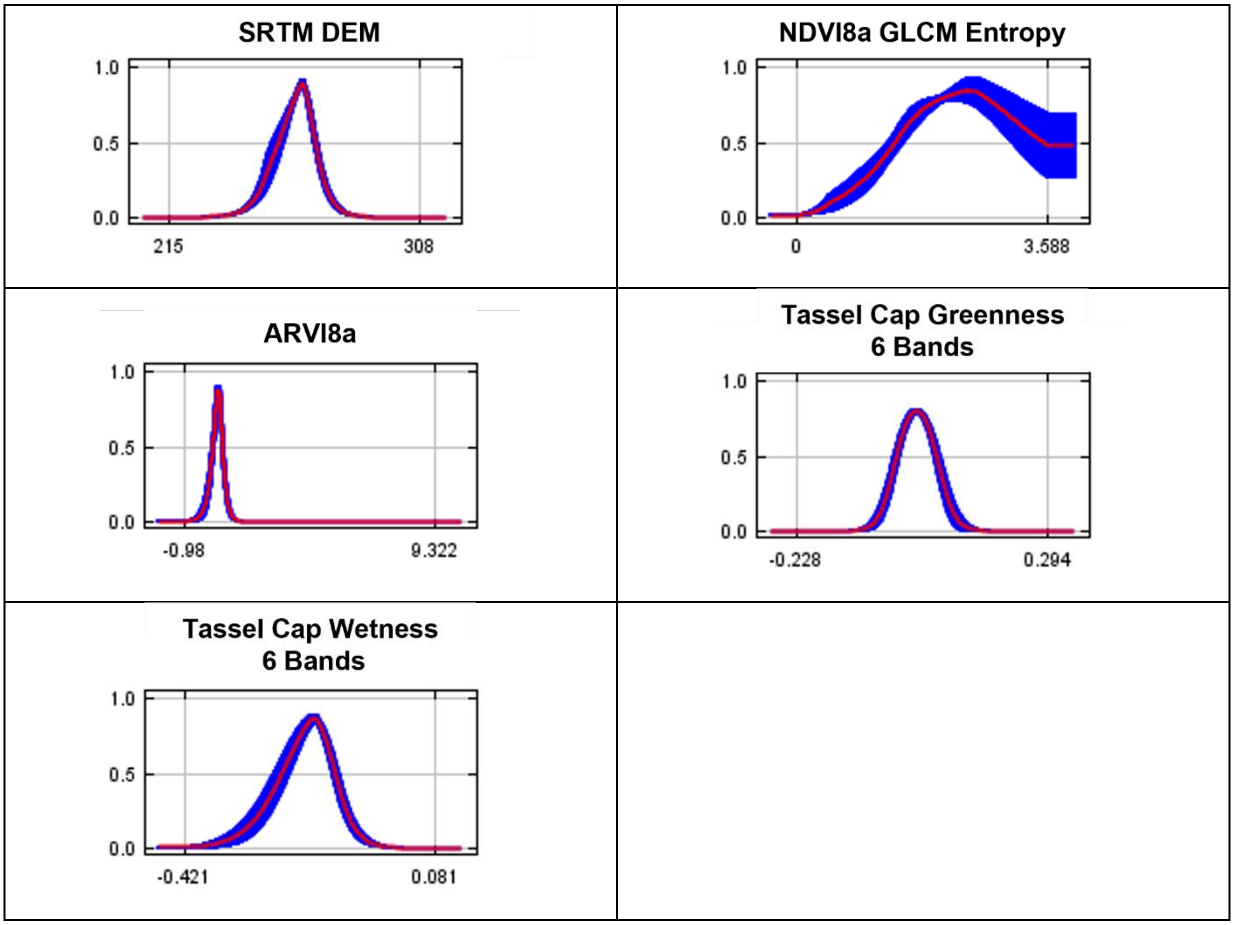


Figure 9. Common Model environmental variable response curves



Curves show predicted suitability for Dakota skipper when each variable is modelled separately (red line and bar is the mean of 25 model runs, blue is +/- one standard deviation). X-axis is environmental variable value range; y-axis is model predictive change.

Figure 10. Satellite Model environmental variable response curves



Curves show predicted suitability for Dakota skipper when each variable is modelled separately (red line and bar is the mean of 25 model runs, blue is +/- one standard deviation). X-axis is environmental variable value range; y-axis is model predictive change.