PINE RIDGE REFORESTATION PROJECT

Lean Six Sigma Process Improvement Report

Richard E.W. Berl

May 6, 2018

Contents

Executive Summary

Problem Statement

Project partner PrintReleaf estimates a 47% survival rate for the 32,000 Rocky Mountain ponderosa pine (*Pinus ponderosa* var. *scopulorum*) seedlings planted by Trees, Water & People on the Pine Ridge Reservation in May 2017. This led to a reduction in payments from the originally estimated $2.35 per tree planted to $1.45. The cost of these seedlings to Trees, Water, & People was $27,989 and labor costs for the planting effort were $20,729, leading to a budget deficit of over $2,318.

Summary of Problem-Solving Approach

We formed a team consisting of members of the organizations that are involved with planning, implementing, and supporting the reforestation project, including Trees, Water & People, PrintReleaf, and Village Earth. The team collaborated in the use of the Lean Six Sigma DMAIC problem-solving methodology to define the problem, identify areas of improvement, and implement solutions for the current planting season. First, we used a variety of tools to map the process and understand all of the factors that could influence seedling survival. Next, I collected data on the present occurrence of Rocky Mountain ponderosa pine and on the climatic and environmental variables that our team identified as relevant. Using these data, I built habitat suitability models specific to this variety for present conditions and for two potential future climate scenarios projected out to 2050. This allowed us to determine the most influential factors in their survival and locate areas where planting will be most effective at encouraging future survival. Other improvements were recommended in the planting process to further increase the chances of seedling survival.

Major Findings and Recommendations

We find that major climatic shifts are expected across South Dakota and within the Pine Ridge Reservation, which will have significant impacts on the habitat suitability and survivability of newly planted Rocky Mountain ponderosa pine. To prepare for these predicted changes, I recommend planting in areas of currently high habitat suitability where projected future suitability is also relatively high. Analyses of correlations and the results of random forest modeling suggest that these are generally areas of high elevation, on cooler northeasterly slopes, in fine-grained, mildly acidic soil with high organic content.

In addition, to focus efforts on variables that we can reasonably control and in the interest of increasing survival as much as possible, I recommend implementing all of the solutions developed that do not depend directly on climatic conditions, including: comprehensive training of volunteers on planting methods, the use of auxin-fortified absorbent polymers during planting, and placement of flagging around planting sites to prevent incidental damage.

I make recommendations for a **Data Collection Plan** to be implemented in the current planting season, so that the impacts of these changes can be measured over time and strategies modified for future seasons. An **Experimental Design** is also presented, whereby the effects of seedling age, polymer usage, and scalping technique on seedling growth and survival can be tested systematically. These will allow for the long-term maintenance of these process changes and continued improvement.

Introduction

This report details the efforts of the project team to improve the performance of ongoing reforestation efforts on Pine Ridge Reservation in southwestern South Dakota, United States. The process is a collaboration between a number of parties, including nonprofit organizations Trees, Water & People and Village Earth, project sponsor PrintReleaf, the Oglala Lakota Tribe of Pine Ridge, and Lakota Solar Enterprises at the Red Cloud Renewable Energy Center.

The primary goal of this project, shared by all parties involved, is to enhance the survival of Rocky Mountain ponderosa pine seedlings (*Pinus ponderosa* var. *scopulorum*) planted annually by volunteers and TWP staff during the May planting season. Reducing the costs associated with planting is a secondary but relevant concern. The methods by which these goals were identified and solutions developed are detailed over the course of the report.

To address these issues, we applied a Lean Six Sigma statistical thinking approach, utilizing the DMAIC methodology to increase the efficiency of the process and to reduce failure rates and operational waste.

What is Lean Six Sigma?

The term **Six Sigma** refers to a business philosophy and set of tools for quality improvement developed by Bill Smith, Mikel Harry, and Robert Galvin at the Motorola Corporation in the mid-1980s. Sigma is the Greek character used to represent standard deviation, a measure of the variation in a set of data points. By reducing the variation in a process to the point where the failure specifications fall outside of six standard deviations, the number of defects are reduced to a rate of roughly 3.4 per million opportunities. Therefore, the goal of Six Sigma is mainly to reduce variation.

The concept of **Lean** comes from lean manufacturing practices, applied and combined with Six Sigma within the automobile industry at Toyota Motor Corporation in the 1980s and 1990s. The focus of lean, as suggested by the name, is to "trim the fat" from a process, eliminating wasted materials, effort, and time.

Lean Six Sigma integrates these two concepts into one approach that examines the entire course of the process, using data-driven methods to improve efficiency and quality and to reduce variation and waste in order to achieve measurable goals. The **DMAIC** method, or: **Define**, **Measure**, **Analyze**, **Improve**, and **Control**, breaks down the Lean Six Sigma approach into multiple discrete stages, each of which has its own goals, tools, and products. The report is structured around these stages, with brief discussions about the tools used and key findings prior to the presentation of the materials.

DEFINE: Improvement Opportunity

Tools

- **Project Charter**: Identifies the need, objective, support, and overall scope of the project. Includes the problem statement and metrics for success of the overall project.
- **5W2H**: Asks critical questions about the Why, Who, Where, When, What, How, and How Many aspects of the project. Provides information that helps to identify areas for improvement.
- **Thought Process Map**: Visual representation of the entire process and its interconnected concepts, structured around project goals.
- **CT Tree**: Links the “voice of the customer” to specific requirements needed to fulfill those needs that are “Critical To” quality or satisfaction.
- **SIPOC**: Documents the boundaries of the process by detailing the specific inputs and outputs and their requirements.

Summary

From these materials, we learned a great deal about the process and the opportunities for addressing the problem, which is that survival rates for seedlings planted in previous years were too low and the costs of volunteer labor too high to create an impactful and sustainable reforestation effort. The project represents significant environmental and cultural benefits for the people and lands involved, and all parties are interested and invested in its success, including Trees, Water & People, PrintReleaf, and the Oglala Lakota Tribe.

Each of the two objectives of the process—maximizing seedling survival and minimizing labor costs—were identified to have a number of important inputs, generally related to climate, ecology, and planting methods for survival and to training, experience, and living expenses for volunteers. Both are linked, in that the efforts of volunteers in planting seedlings contribute directly to survival and, therefore, to the effectiveness, efficiency, and value of their work in terms of surviving trees planted per hour of paid labor. Though it was found that there is little flexibility in reducing the costs of volunteers on-site, we can take steps to increase the value of their time and labor by improving training and increasing seedling survival.

In the following stages of the project, we explore ways to successfully address these two interrelated issues.

Project Charter

Problem Statement

Project partner PrintReleaf estimates a 47% survival rate for the 32,000 Rocky Mountain ponderosa pine (*Pinus ponderosa* var. *scopulorum*) seedlings planted by Trees, Water & People on the Pine Ridge Reservation in May 2017. This led to a reduction in payments from the originally estimated $2.35 per tree planted to $1.45. The cost of these seedlings to Trees, Water, & People was $27,989 and labor costs for the planting effort were $20,729, leading to a budget deficit of over $2,318.

Project Objective

Increase seedling survival rate from 47% to 75% (60% improvement) and minimize labor costs for the May 2018 planting season.

Project Team

Richard E.W. Berl, *Ph.D. Candidate*, Human Dimensions of Natural Resources, Colorado State University
Sebastian Africano, *Executive Director*, Trees, Water & People
Eriq Acosta, *National Director*, Trees, Water & People
Jordan Darragh, *CEO and Founder*, PrintReleaf
Dave Bartecchi, *Executive Director*, Village Earth

Primary Metrics

Survival rate = (Number seedlings planted / Number seedlings survive)

Labor cost per tree = [Hourly payment * (Number of hours planting + Number of hours training) / Number seedlings planted]

Critical to Satisfaction

Poor tree survival leads to less compensation, which means less money to purchase seedlings for the following year, which could lead to a lack of confidence from stakeholders. This could affect the overall success of this project and loss of potential future projects.

Defect Definition

Any seedling that fails to survive the first 4 months following planting.

Scope of Project

From the delivery of seedlings to the site to the survival assessment 3-4 months after planting.

5W2H

Why?

“Pine Ridge might as well be called Ridge.” Deforestation (due in part to wildfires) affects grazing, livelihoods, water retention, landslide reduction, and the health of the land. It is important to succeed in reforestation and land restoration efforts primarily for the social impacts. The benefits extend to current and future generations, connecting children in a significant and tangible way with place and engaging them in the positive aspects of their history, culture, and land.

Who?

Trees, Water & People coordinates the purchasing, planting, and volunteer efforts. PrintReleaf assesses survival with students from the University of Colorado Boulder and provides funding through third-party corporate donors. Village Earth provides GIS support. The Oglala Lakota Tribe of the Pine Ridge Reservation govern the land and, through Henry Red Cloud and the Red Cloud Renewable Energy Center, collaborates with the reforestation project. The Colorado State Forest Service is contracted to supply the seedlings. All parties involved hold an interest in maximizing tree survival.

Where?

Last year, 32,000 seedlings were planted in the areas of Lakeside and Evergreen in the Pine Ridge Reservation. This year, 28,000 seedlings will be planted on Reservation land and 5,000 seedlings will be planted in Bear Butte State Park, with 1,000-2,000 held for experimental plots. Within the process, issues could arise in the planning, planting, training, monitoring, and growth stages.

When?

Seedlings are planted annually in May between the Spring thaw and Summer drought periods. The project began in 2015 with 7,500 seedlings and continued in 2016 with 15,000 seedlings and 2017 with 32,000 seedlings. Problems with deforestation began as far back as the beginnings of colonization in the area and westward expansion by Europeans, and more recently have been exacerbated by government fire suppression policies which led to destructive wildfires.

What?

The defect is a low estimated rate of seedling survival, currently 47%. Seedling delivery, volunteer training and planting, and follow-up could contribute to this.

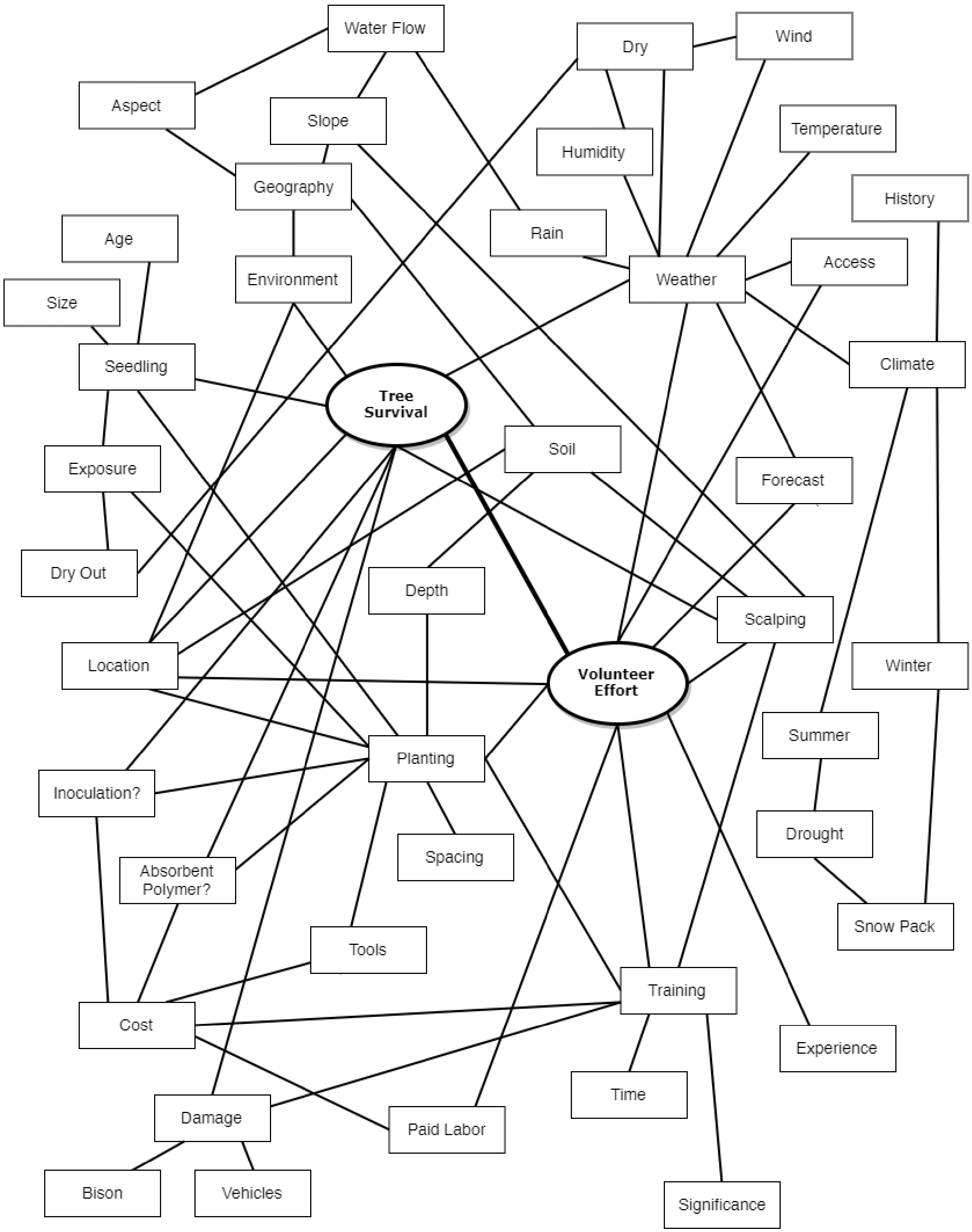
How?

The problem is identified during the follow-up assessment by PrintReleaf, in which a sample of trees are examined and survival rate determined. This occurs 3-4 months after planting.

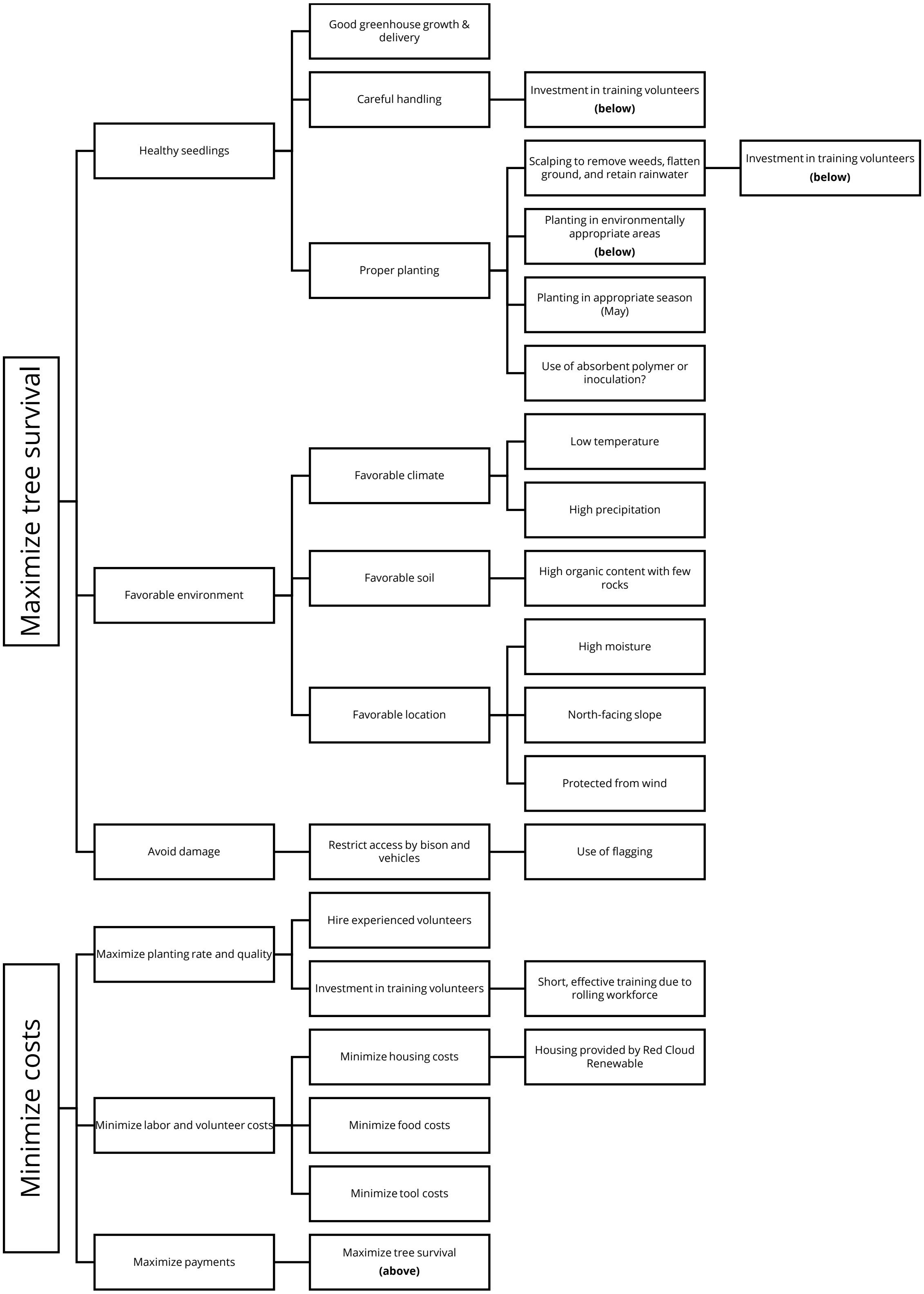
How many?

The estimated survival rate of last year's seedlings was 47%. Seedling cost was $27,989 and labor costs for the planting effort were $20,729. Additional unquantified costs were involved in purchasing tools and providing food for volunteers, as well as staff time investments. Costs are variable from year to year as the project has increased in size and scope.

Thought Process Map



CT Tree


Maximize tree survival
Healthy seedlings
Good greenhouse growth & delivery
Careful handling
Investment in training volunteers
(below)
Proper planting
Scalping to remove weeds, flatten ground, and retain rainwater
Investment in training volunteers
(below)
Planting in environmentally appropriate areas
(below)
Planting in appropriate season (May)
Use of absorbent polymer or inoculation?
Favorable environment
Favorable climate
Low temperature
High precipitation
Favorable soil
High organic content with few rocks
Favorable location
High moisture
North-facing slope
Protected from wind
Avoid damage
Restrict access by bison and vehicles
Use of flagging
Minimize costs
Maximize planting rate and quality
Hire experienced volunteers
Investment in training volunteers
Short, effective training due to rolling workforce
Minimize labor and volunteer costs
Minimize housing costs
Housing provided by Red Cloud Renewable
Minimize food costs
Minimize tool costs
Maximize payments
Maximize tree survival
(above)

SIPOC

Suppliers	Inputs	Process		Outputs	Customers	
		Input Requirements			Output Requirements	
PrintReleaf, Corporate Sponsors	Funding	Cover all costs		Higher payments	Contract survival estimates	TWP
				Continuation of project	Successful, future funding	TWP, Oglala Lakota Tribe
CO State Forest Nursery	Seedlings	Ready, healthy, appropriate variety	Prepare for Planting			
				Significance to community	Engagement with youth	TWP, Oglala Lakota Tribe
Oglala Lakota Tribe, Volunteers	Volunteer labor	Enough to plant all seedlings				
			⇩	Confidence of stakeholders	Continued support for success	TWP, PrintReleaf, Corporate Sponsors
Experienced Volunteers	Volunteer training	Trained in planting method				
			Plant Seedlings	Carbon sequestration	Biomass	Environment, Future Generations
TWP	Tools	Adequate for all volunteers, low cost				
			⇩	Environmental health	Soils, habitat, microclimates	Environment, Future Generations
TWP	Transportation	“				
Red Cloud Renewable	Lodging	“	Measure Survival			
				Wind buffering	Windbreaks	Environment, Oglala Lakota Tribe
TWP	Food	“				
TWP, Oglala Lakota Tribe	Planting sites	Best areas for survival				

MEASURE: Current State of the Process

Tools

- **Process Flow Map**: Walks through the process and helps to identify decision points and measurable inputs (below) and outputs (above).
- **Cause & Effect / Fishbone Diagram**: Also known as an Ishikawa diagram. Categorizes the potential root causes of failure in the process.
- **Run Test**: Visualizes and tests trends in data over time, as the data cluster in “runs” above, below, or about the median.
- **Data Collection Plan**: Specifies the data to be collected from the process, ensuring it is useful and meaningful.

Summary

Using these tools, we gather additional information about the process and find opportunities for measurement. Since little data is available from past planting efforts (coordinates for 100 surviving trees planted in close proximity), we benefit from looking at the decision points and risks of failure within the process and considering the data that would be helpful in addressing those issues, both at present and for the future.

Though the problem of low survival does not manifest until the end of the process, it is determined by a number of points upstream. The process revolves primarily around the planting effort, which itself is driven by the training and labor of volunteers. Both of these aspects, then, are critical to seedling survival.

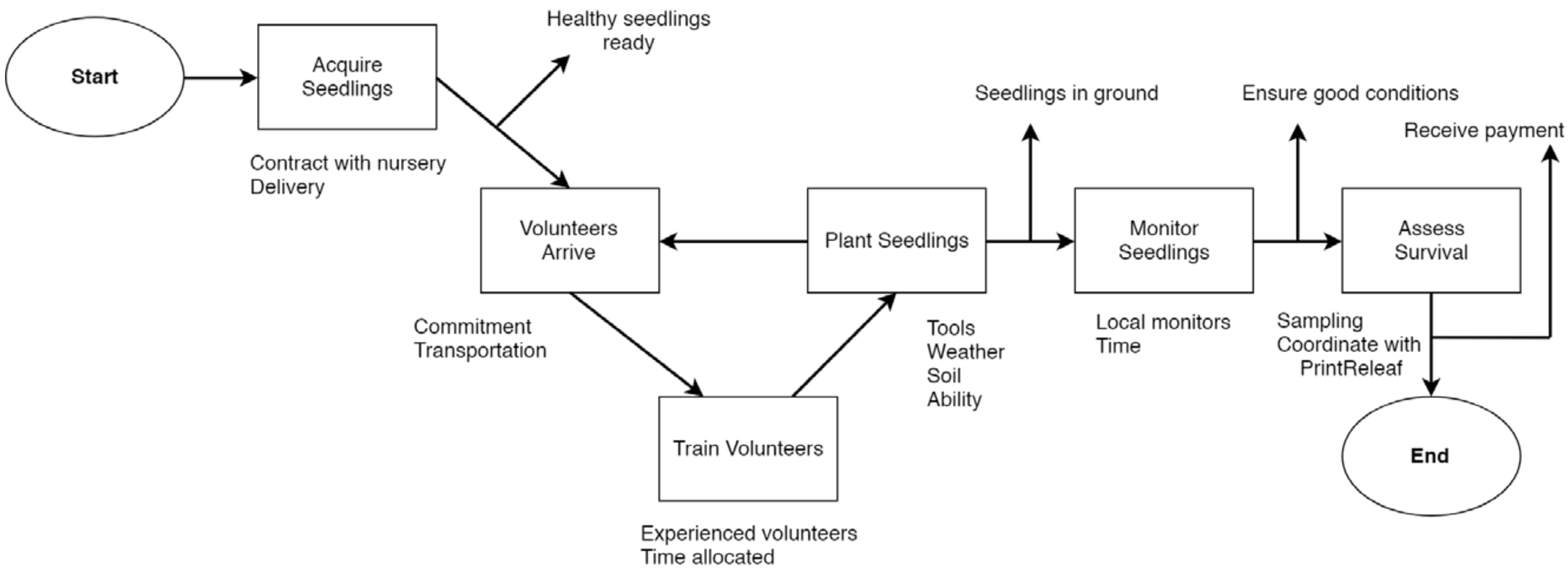
Given the lack of necessary data on seedling survival and the variables of interest—ones generated over the course of the Define and Measure phases—we took two courses of action. First, to ensure adequate data is available in the future, we developed a **Data Collection Plan** for the current season, balancing the need for data with a restriction to those variables that could be collected without causing significant burden during the critical planting stage. This will require data collection at the time of planting this season as well as at a later point for the assessment of survival.

Second, we sought data on relevant climatic and soil variables from publicly accessible data sources to address the issues to the best of our ability for the current planting season. I obtained relatively recent occurrence data for the specific Rocky Mountain variety of ponderosa pine being planted, collected between 2001 and 2012, from Maguire et al. 2018 (originally collected for Potter et al. 2013). The data set consists of 2,500 positive occurrence points and 5,528 randomly selected points where they do not occur. Having these data allows us to see where *P. p.* var. *scopulorum* is currently able to survive, within a range where interbreeding with other varieties is unlikely and the climate is roughly similar to that of South Dakota (see **Exploratory Data Analysis** in the Analyze phase). The Run Test performed here uses these data, and finds no evidence of systematic irregularities or trends.

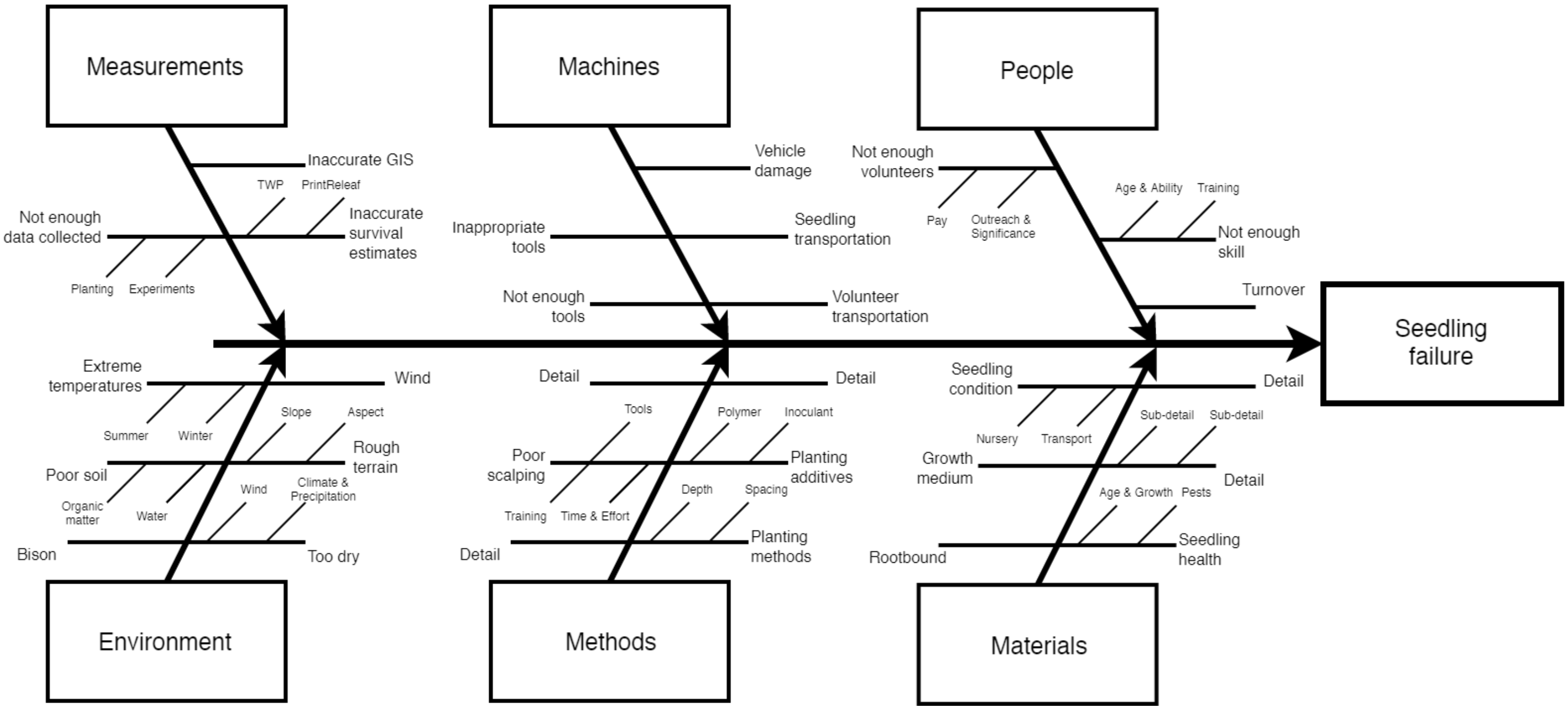
Along with the occurrence data, I obtained data on other variables from a variety of sources. Climate data was obtained from the AdaptWest project (2015; Wang et al. 2016) for all available “bioclimatic variables,” which are those most meaningful for the ecology and survival of living organisms (full list available in **Metadata** at the end of the report). Data on current conditions were derived from verified climate normals, or three-decade averages, from 1981-2010. These should be relevant for currently living ponderosa pine due to their moderate growth rate and long lifespan. Climate projections for future conditions at the year 2050 were also obtained, based on an ensemble (combination) of the most reliable climate models (CMIP5 AOGCMs) under two possible future emissions scenarios: low, controlled emissions for RCP4.5 (Thomson et al. 2011) and high, uncontrolled emissions for RCP8.5 (Riahi et al. 2011).

These climatic variables were supplemented with elevation, soils, wind, solar radiation, and burn area data from other sources. The digital elevation data used were sourced from Jarvis et al. (2008), which used NASA Shuttle Radar Topography Mission (“SRTM”) data and filled areas of missing data. Soils data were obtained from POLARIS (Chaney et al. 2016), wind and solar radiation data from WorldClim 2 (Fick & Hijmans 2017), and burn area data from the Landsat BAECV (Hawbaker et al. 2017). SRTM data were used for elevation instead of the more modern 3DEP national data files, as they were too large to deal with analytically and the increased resolution was not necessary for this project and would have been discarded during resampling. Soils data were restricted to the most biologically-relevant variables, and only for the first 5cm of soil since the composition of deeper layers is highly correlated with the first layer. Burn area data was restricted to 2001, the first year of species occurrence sampling, and was not used for prediction of future conditions as estimating the probability of future burn sites would be beyond the scope of this project. A list of all the variables obtained from these various data sources and details on their resolutions and interpolation are listed in the **Metadata** section at the end of this report.

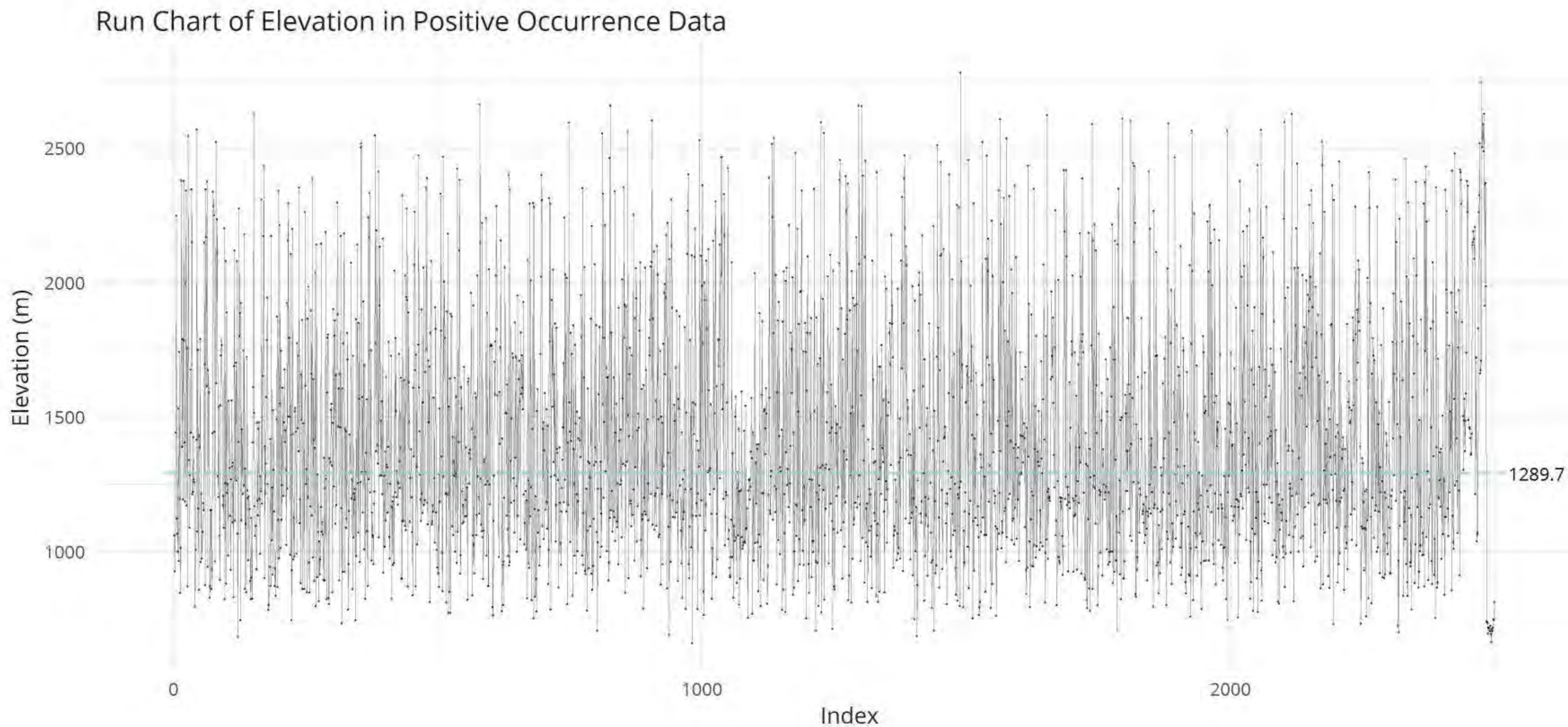
Process Flow Map



Cause & Effect / Fishbone Diagram



Run Test



Number of observations	2500
Number of runs	1237
Longest run	29
Number of crossings	1236

Wald–Wolfowitz test statistic	-0.5601
Wald–Wolfowitz p-value	0.5754
Expected maximum longest run	14
Expected minimum crossings	1208

Data Collection Plan

Research Questions

- What factors are most influential in encouraging the growth and survival of Rocky Mountain ponderosa pine (*Pinus ponderosa* var. *scopulorum*) in and around Pine Ridge Reservation in South Dakota?
- Are there any interactions between climatic or soil variables, planting conditions or methods, and the quality of volunteers?

Data Table

Data		**Operational Definition and Procedures**	
What	**Measure type/data type**	**How measured**	**Sampling notes**
ID number	Integer	Recorded on garden tape or tag attached to stem	Assigned at or before planting; Sequential, from 1 to sample size; Must be unique for each seedling
Date and time	Date-time (YYYY-MM-DD hh:mm:ss)	Clock or watch, synchronized between data recorders	Record during planting and again at assessment
Geographic coordinates	Continuous (decimal degrees)	Latitude and longitude found using a GPS unit or GPS-enabled cell phone or camera	Record during planting, or during assessment if data missing; Ensure accuracy is good, to at least ± 10m
Age	Discrete/Continuous (years)	Based on information from nursery, if available	Recorded before planting
Height*	Continuous (cm)	Ruler or meter stick from the cotyledon scar (near base of stem) to the base or tip of the terminal bud (or end of growing tip if no bud formed)	Record during planting and again at assessment
Stem diameter*	Continuous (mm)	Measured with a ruler or calipers just below the cotyledon scar (near base of stem)	Record during planting and again at assessment; Important to ensure that the calipers are perpendicular to the stem during measurement
Polymer use	Binary (yes/no) or Continuous (g or ml)	Visual assessment of presence, or measurement by mass or volume if amounts not consistent	Record during planting; Only applicable if used
Planting quality	Ordinal (1-5 scale: Very Poor, Poor, Acceptable, Good, Very Good)	How well proper scalping and planting procedure was followed: that the plot is flat or slightly angled in toward the hill, free of existing vegetation, planted at sufficient depth, stem vertical, roots not exposed, etc.	Record during planting and again at assessment (average the two values for analysis); May need to be assessed by someone other than the planter to avoid bias; Take brief notes on reasons behind assessment

Nearest neighbor distance	Continuous (m)	Meter stick or measuring tape to nearest other seedling or existing tree	Record during planting; Ensure no other seedlings are to be planted closer than the one being measured
Planter name	Categorical	Written	Record during planting
Planter age	Discrete/Continuous (years)	Written	Record during planting
Planter gender	Categorical (male/female/nonbinary)	Written	Record during planting
Planter training duration	Discrete/Continuous (hours)	Written	Record during planting; Number of hours of training undergone by the planter prior to planting; Change only if additional training received between plantings
Planter experience level	Ordinal (1-4 scale: None, Low, Moderate, High)	Written	Record during planting; Subjective, but levels could be regarded as: Little/no experience, Casual gardener or outdoor enthusiast, Serious gardener with some tree planting or landscaping or previous volunteer for this project, Professional landscaper, arborist, or botanist, or multiple years volunteering for this project
Survival	Binary (yes/no)	Visual assessment; Dead seedlings will be drooping or wilted with brown needles and no new growth of green needles or buds	Record during assessment; If unsure, use the "scratch test": remove a small strip of bark to check for living green tissue beneath, if it is brown and dry then it is dead, check several areas in case only one branch is dead
Notes	Qualitative	Written	Any additional observations or notes on conditions that could be relevant to seedling survival; Can be used in future to find additional potentially relevant variables

* Adapted from Haase 2008, see for additional details and other relevant measurements. See Menes & Mohammed 1995 for assistance with identifying the cotyledon scar on pine seedlings.

Ensuring Consistency, Stability, and Reliability

Since it is unlikely that we will be able to record data on every seedling that is planted, a sampling scheme will need to be used to ensure that samples are reasonably random and evenly distributed. This could mean sampling every seedling planted by a subset of individuals (e.g. 3 prespecified volunteers per weekend of planting) with the same level of ability and experience, or by tagging a subset of seedlings before planting and collecting data on each of them regardless of who plants them. The first design would attempt to control for skill level, while the second assumes it is sufficiently randomized. Managers would need to ensure other potential influences are either controlled for or sufficiently randomized as well; for example, planting day, time of day, location, and weather conditions.

If only a few individuals are recording data on all of their seedlings, these individuals should be trained to collect data consistently between themselves. If seedlings are randomly selected for data collection between different planters, it would likely be best to have one or two designated volunteers, trained to collect the data consistently, that circulate between planters or planting areas.

How Data Will Be Used

- Exploratory data analysis
- Coordinate with climate and soil data
- Identification of most significant contributors to survival
- Examine possible interactions between variables
- Check for geographic trends in survival
- Evaluate predictions of habitat suitability models
- Cross-verification of survival rate estimate
- Re-evaluate methods and continue to improve future survival rates

How Data Can Be Displayed

- Scatter plots
- Geospatial plots
- Control chart
- Pareto chart
- Run chart

ANALYZE: Examination and Findings

Tools

- **Exploratory Data Analysis**: Summarizes and compares the characteristics of a data set to provide insights on its structure and possible trends and relationships.
- **Hypothesis Testing**: Uses statistical tests to determine whether variation between groups of data is due to true differences or to natural variation in the process.
- **Habitat Suitability Modeling**: Predicts geographic locations where a particular species could occur, given relevant environmental variables.

Summary

In the Analyze phase, we take a deep dive into the data collected during the Measure phase and apply statistical tools and modeling toward our goal of improving seedling survival. We focus primarily on the data we were able to collect on occurrence, climate, and soils—rather than the potential variables detailed in the **Data Collection Plan**—as the additional collected data would facilitate and require a different set of analytical methods.

In exploratory analyses, we examine the correlations between the full set of variables, and specifically with the occurrence of Rocky Mountain ponderosa pine (variable `PA`). Descriptions of each variable are given in the **Metadata** at the end of the report. In terms of raw correlations, those most strongly linked with occurrence (absolute correlation above 0.3) are:

1. `slope`: Angle of slope
2. `TD`: Range in temperature between coldest and warmest months (negatively associated)
3. `MCMT`: Temperature of the coldest month
4. `bd_mean_0_5`: Soil particulate size (negatively associated)
5. `Tave_wt`: Winter temperature

From this, we could conclude that present trees tend to occur on steep slopes with little variation in annual temperature, winter temperatures that are relatively mild, and in topsoil that is free of large rocks. Associations with other variables support this picture.

In the following maps, I overlay the occurrence points on an elevation base layer for the Rocky Mountain Front and Northern Great Plains region where the sample was drawn, and for those points that fall within Pine Ridge. I then quantify the degree of climate change expected to occur in South Dakota and in the Pine Ridge Reservation between present conditions and the year 2050, with a value of 1 indicating an increase of 100% in that variable and -1 indicating a decrease of 100%. We do see more drastic changes in the RCP8.5 scenario than in the RCP4.5 scenario, especially in the Black Hills where we see an increase of 296% in the number of days above 18°C under RCP4.5 and 489% under RCP8.5, and also in the northeastern parts of the state. The effects are not as severe in Pine Ridge, but we do still see significant changes, such as: general warming indicated by the increased number of days above 18°C, drought indicated by decreases in summer precipitation (`PPT_sm`) and in the amount of snow (`PAS`), and greater extremes such as a decrease in winter temperatures.

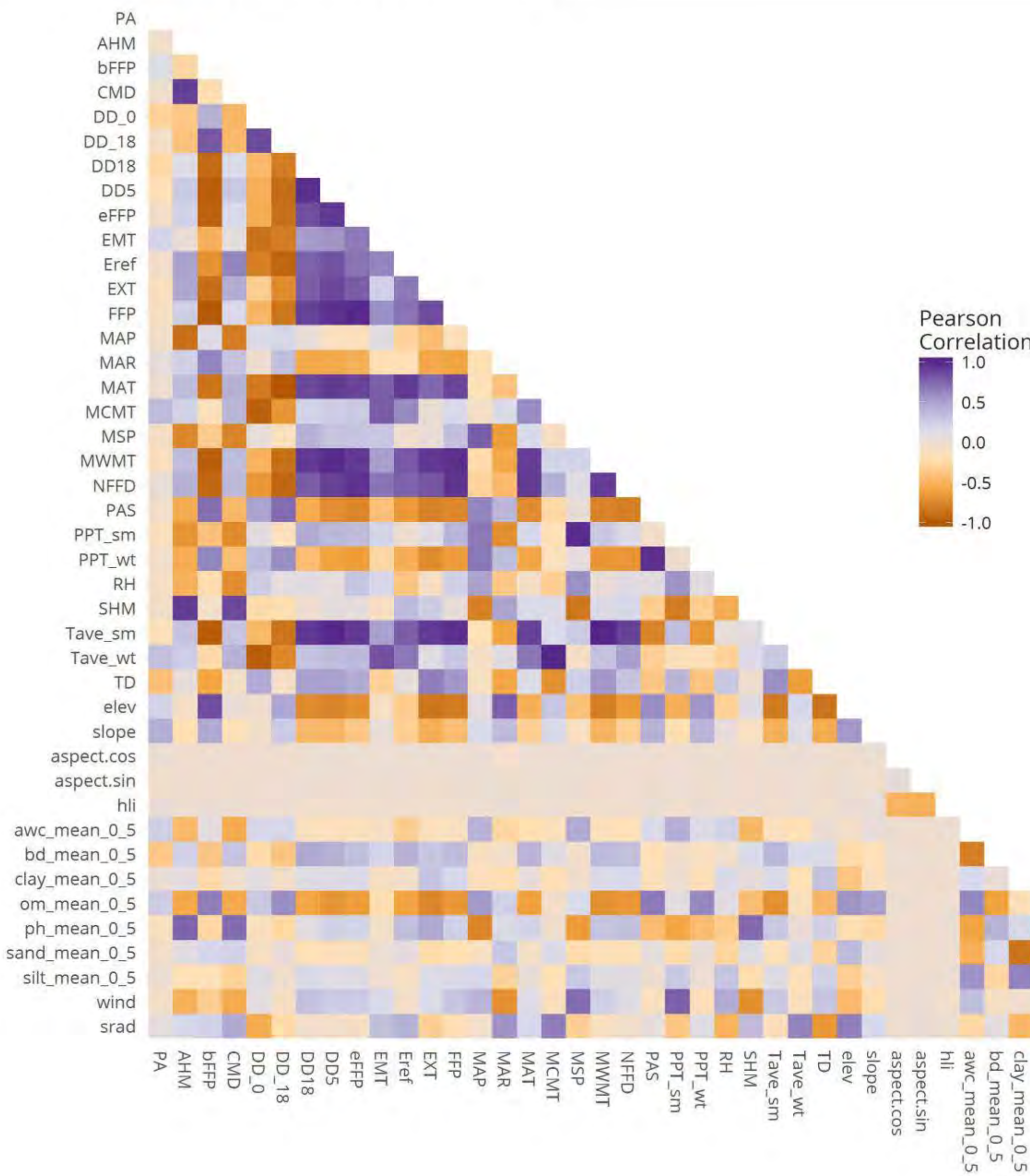
Since we do not have future projections for burn areas to predict occurrence through habitat suitability (see the **Habitat Suitability Modeling** section), I used a logistic regression to assess its statistical significance in predicting the present occurrence of Rocky Mountain ponderosa pine. I found that a quadratic model—incorporating a squared term to represent an increase and then a decrease in probability—fits the data better than a linear model of constantly increasing probability. This suggests that past burns that have run their course do encourage the growth of new *P. p.* var. *scopulorum*, up to a burn probability of around 0.734, but that more recent or high-intensity burns (indicated by higher burn probabilities near 1.0) likely destroyed any trees (and possibly seed banks) that were present prior to the burn. This fits with our knowledge of fire adaptation in ponderosa pine and their reliance on frequent, low-intensity fires for reproduction and recolonization, as well as the problems created by past fire suppression policies (Shepperd

& Battaglia 2002). Indeed, past large-scale wildfires are one of the main reasons why this reforestation project is necessary.

Finally, I use a tool from ecology known as **Habitat Suitability Modeling** to determine the areas in which Rocky Mountain ponderosa pine could occur, given present and future conditions. The methods used, the results obtained, and their implications are described in more detail prior to that section.

Exploratory Data Analysis

Correlation Matrix Heatmap



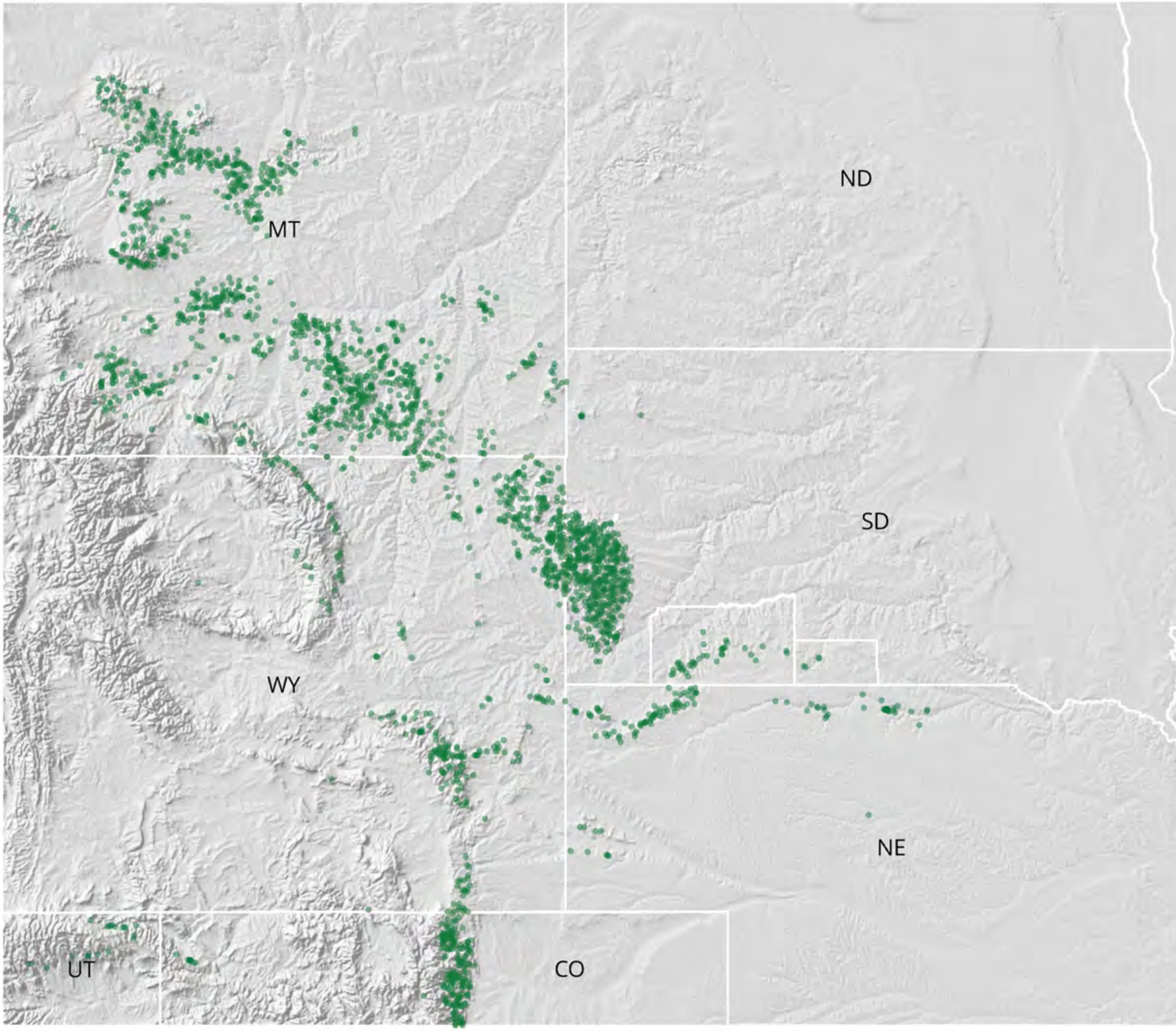
Correlations with Occurrence

Variable	Correlation
slope	0.4198
TD	-0.4132
MCMT	0.3614
bd_mean_0_5	-0.3513
Tave_wt	0.3306
DD_0	-0.2930
om_mean_0_5	0.2697
awc_mean_0_5	0.2468
DD18	-0.2460
elev	0.2364
EMT	0.2277
DD5	-0.1763
Tave_sm	-0.1691
PPT_sm	-0.1355
MWMT	-0.1354
bFFP	0.1353
ph_mean_0_5	-0.1298
EXT	-0.1228
FFP	-0.1075
MAR	0.0821
MSP	-0.0786

Variable	Correlation
RH	-0.0760
srad	0.0748
DD_18	-0.0670
eFFP	-0.0624
wind	-0.0563
SHM	-0.0556
sand_mean_0_5	-0.0531
MAP	-0.0525
NFFD	0.0471
PAS	-0.0392
hli	0.0375
PPT_wt	-0.0373
Eref	-0.0355
aspect.sin	-0.0284
AHM	-0.0222
CMD	-0.0219
aspect.cos	0.0210
clay_mean_0_5	0.0199
silt_mean_0_5	0.0103
MAT	0.0101

Rocky Mountain Front & Northern Great Plains

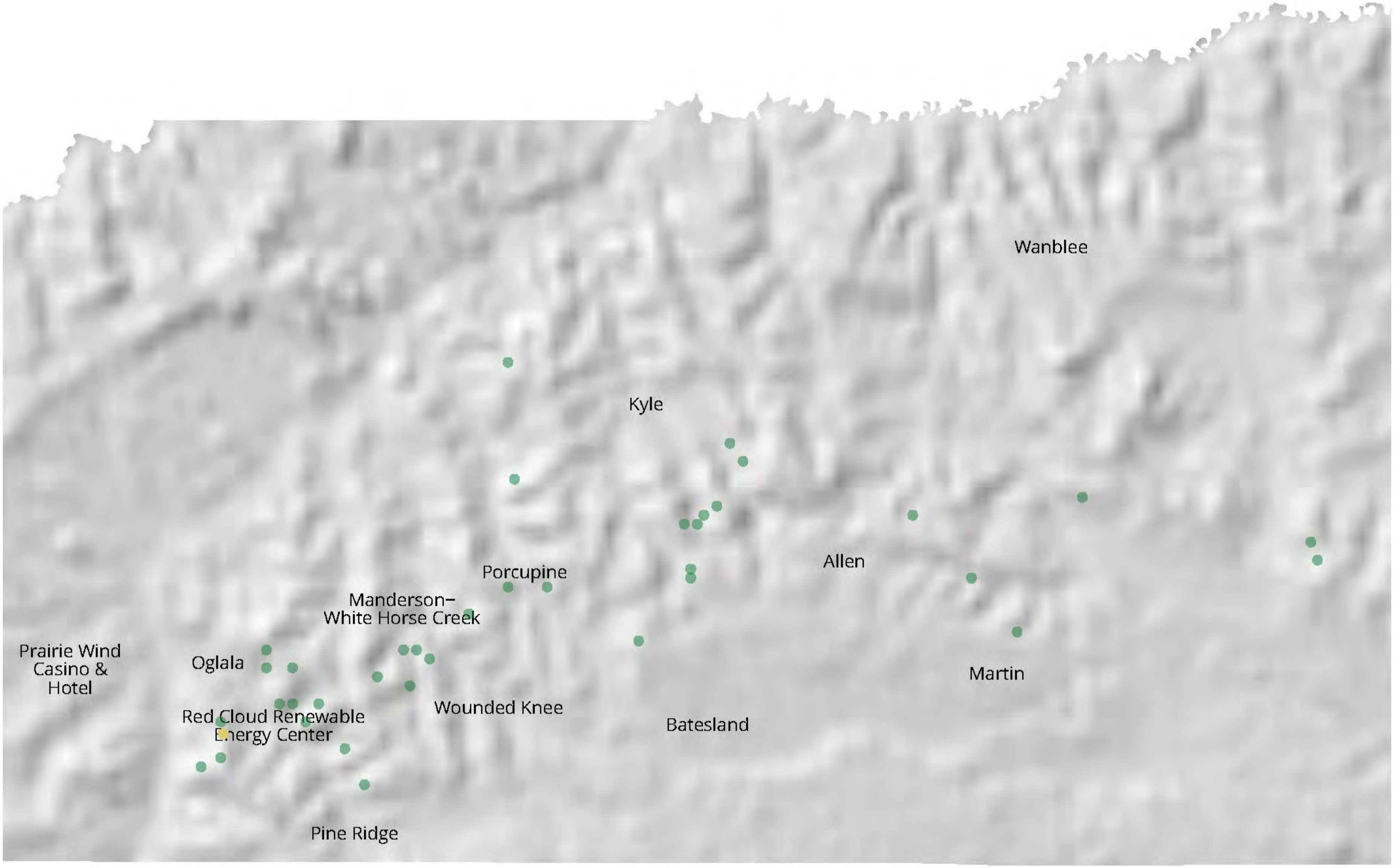
Present Occurrence of Pinus ponderosa var. scopulorum



Author: Richard E.W. Berl; Map Theme: Timo Grossenbacher; Font: Open Sans; Data Sources: Maguire et al. 2018, GTOPO30, TIGER/Line

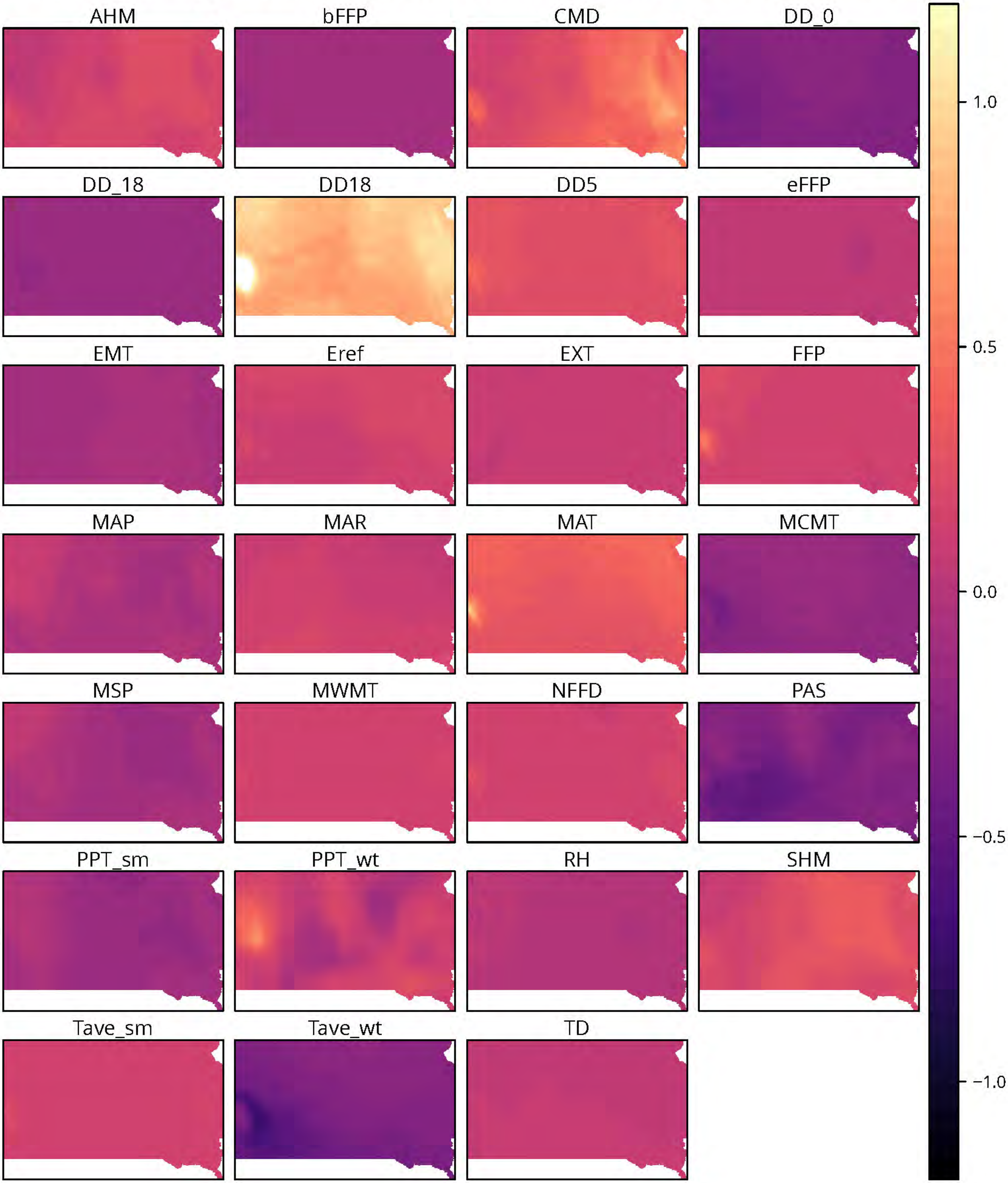
Pine Ridge Reservation and Trust Land

Present Occurrence of Pinus ponderosa var. scopulorum
(Gold Point Indicates 2017 Planting)

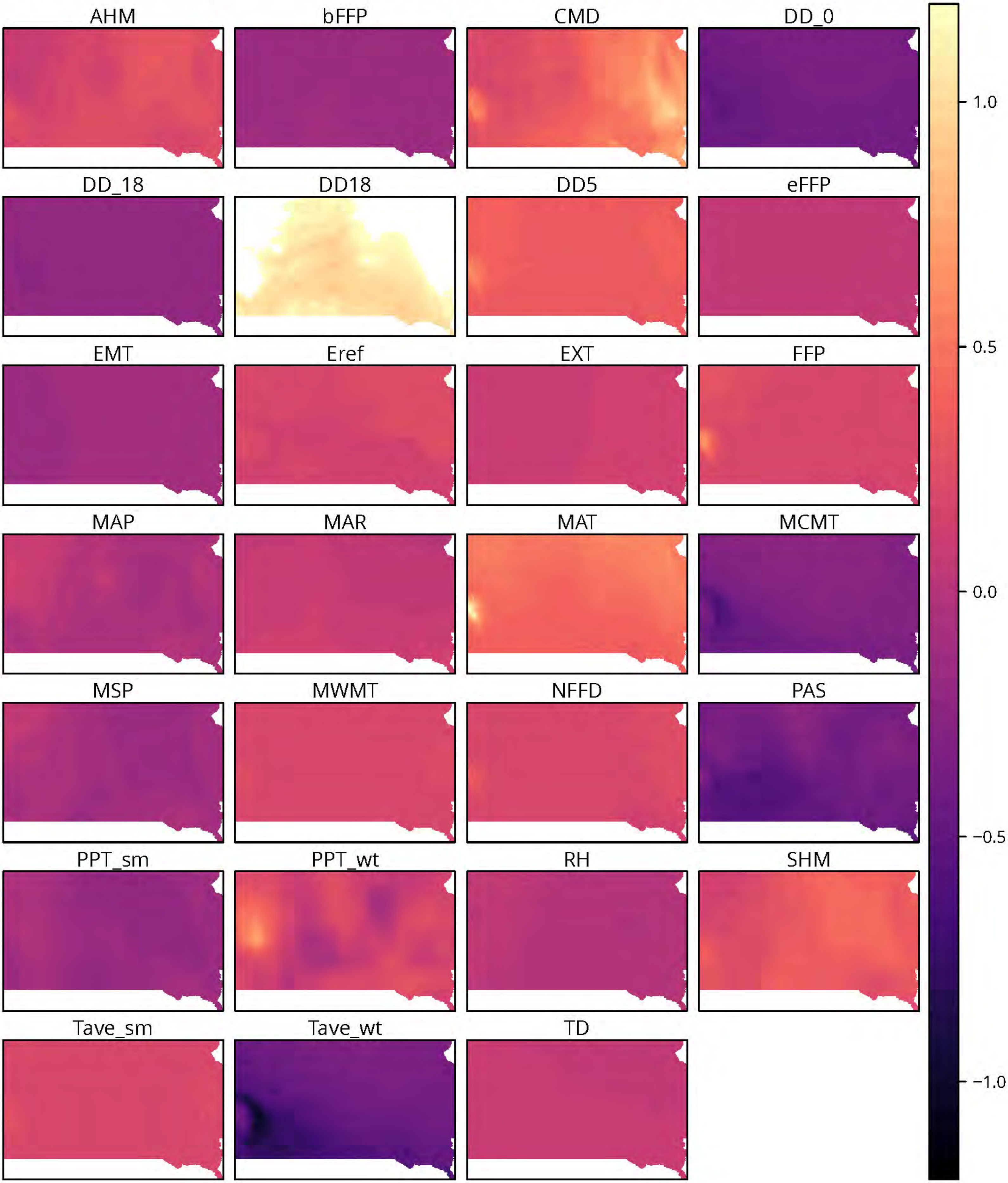


Author: Richard E.W. Berl; Map Theme: Timo Grossenbacher; Font: Open Sans; Data Sources: Maguire et al. 2018, GTOPO30, TIGER/Line

South Dakota
Relative Change in Climatic Variables from Present Conditions to RCP4.5 in 2050



South Dakota
Relative Change in Climatic Variables from Present Conditions to RCP8.5 in 2050



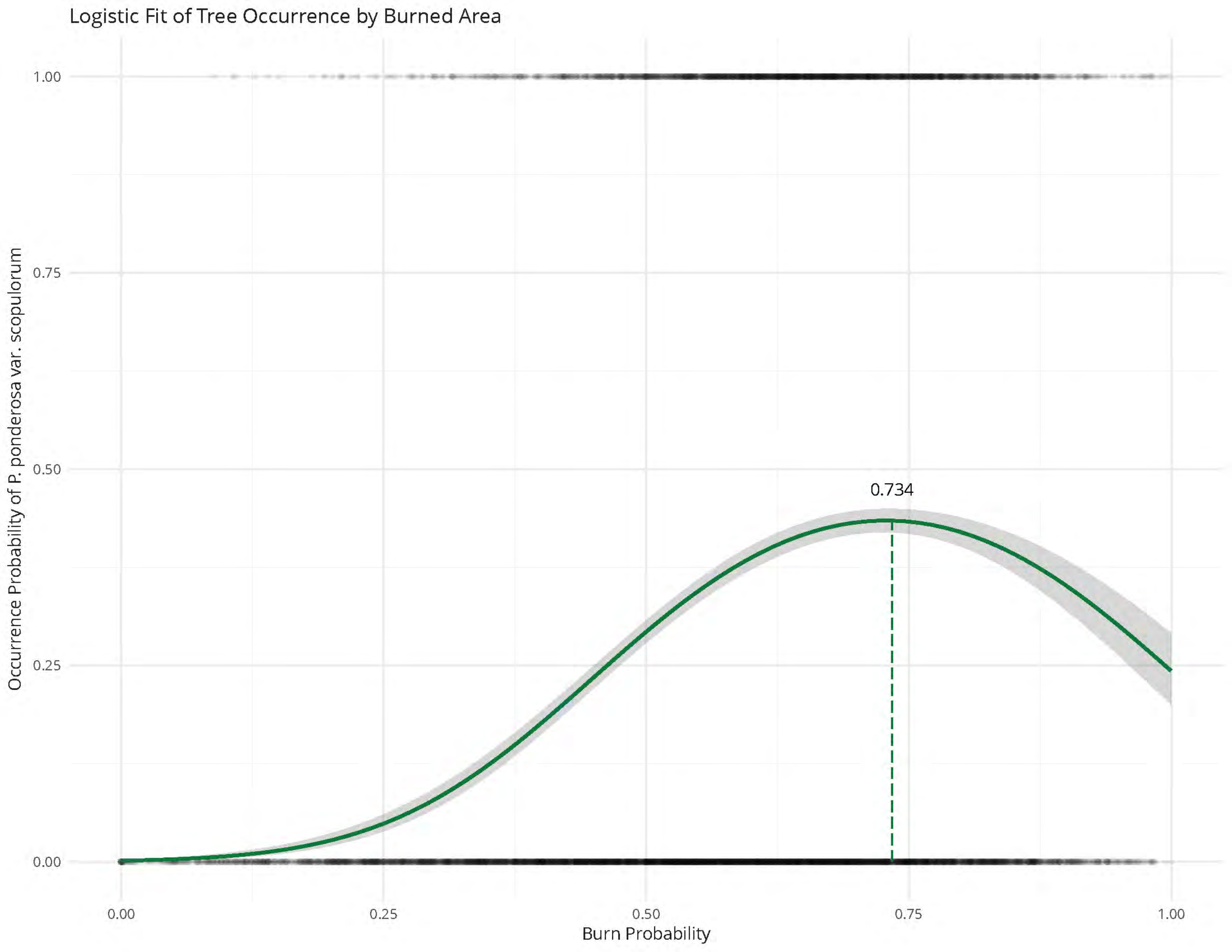
Pine Ridge Reservation and Trust Land
Relative Change in Climatic Variables from Present Conditions to RCP4.5 in 2050

Pine Ridge Reservation and Trust Land
Relative Change in Climatic Variables from Present Conditions to RCP8.5 in 2050

Hypothesis Testing

Logistic Regression of Occurrence on Burn Areas

```
Call:
glm(formula = PA ~ bp + I(bp^2), family = "binomial")

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.0679  -0.9898  -0.6423   1.2987   3.2030

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -6.5460     0.3195  -20.49   <2e-16 ***
bp           17.2529     1.0535   16.38   <2e-16 ***
I(bp^2)     -11.8443     0.8465  -13.99   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 9958.3  on 8027  degrees of freedom
Residual deviance: 9123.5  on 8025  degrees of freedom
AIC: 9129.5

Number of Fisher Scoring iterations: 5
```


Logistic Fit of Tree Occurrence by Burned Area
Occurrence Probability of P. ponderosa var. scopulorum
Burn Probability
0.734
0.00
0.25
0.50
0.75
1.00

Habitat Suitability Modeling

Overview

Habitat suitability modeling, one type of species distribution modeling (Elith & Leathwick 2009), is used in the discipline of ecology for a variety of applications related to the analysis and prediction of habitat niches across space and time, and is employed here to compare the conditions in which Rocky Mountain ponderosa pine currently occurs to conditions across South Dakota. We are interested specifically in this year's prospective planting locations around the communities of Oglala and Porcupine in Pine Ridge Reservation, and in Bear Butte State Park near Sturgis, South Dakota, and we gradually focus in on those three locations.

The first two steps in habitat suitability modeling are to obtain data on present occurrence for the species and on any ecologically relevant predictor variables (Elith & Leathwick 2009). These have been completed in the Measure phase of the project, so we proceed to the next task, which is to find the statistical model that best predicts the occurrence patterns seen in the data based upon the environmental variables we have available.

Model Building

To fit a model to our data, I employed a commonly-used algorithm from the field of machine learning called random forests. In essence, a random forest model creates a number of decision trees to decide how the data points should be classified (e.g. if X > some threshold, then classify it as A; otherwise, B), each based on a sample of the data and a subset of the variables. From this "forest" of decision trees, it creates a consensus of the trees that performed best at predicting the correct output, and uses these to estimate the parameters of our final model. For our purposes, the random forest method also has a number of advantages, including the ability to deal with a large number of variables, regardless of collinearity (when variables are confounded or too highly correlated to be useful), and its dual utility for classification or regression contexts.

In this way, I built a random forest model for the occurrence of Rocky Mountain ponderosa pine, using the `randomForest` package for the R statistical environment (R Core Team 2017; version 3.4.0). All environmental variables in the **Metadata** were included as predictors in the model, except where otherwise noted. I intentionally excluded elevation, slope, and aspect in an attempt to avoid survivorship biases in the occurrence data, since I considered that the first two variables could be more predictive of relative inaccessibility to logging and other human activities than of potential habitat suitability. I also felt that all three were adequately represented by proxy through a number of other variables that have clear relationships with terrain, such as heat load index (`hli`), temperature, precipitation, and soils. Supporting this view, it was found that including elevation, slope, and aspect actually increased the error and decreased the amount of variance explained, showing that these variables do not add anything to the model.

As noted previously, and as shown in the first occurrence map in the **Exploratory Data Analysis** section, the occurrence data used as the target for the model was restricted to an area of the Rocky Mountain Front and Northern Great Plains where the climate is roughly comparable to South Dakota and where we can be reasonably certain that there is no interbreeding with other ponderosa pine varieties.

Using a forest of 500 trees, a regression model fitting procedure was optimized to use 24 variables at each split, with a resulting mean of squared residuals of 0.0368 and 82.82% of the variance explained. Regression was used rather than classification due to the need to predict relative probabilities of occurrence across a landscape rather than binary presence-absence. The importance of each variable to the fitted model, represented by the percent increase in the mean standard error of predictions, is shown below.

Variable	% Inc. MSE
srad	56.03
om_mean_0_5	48.58
wind	41.73
bd_mean_0_5	41.26
hli	39.05
MCMT	38.86
MSP	37.71
clay_mean_0_5	36.61
silt_mean_0_5	33.65
EMT	33.23
bFFP	32.7
PPT_wt	32.64
ph_mean_0_5	28.59
NFFD	27.83
Eref	24.89
PPT_sm	24.73
MAP	24.4
MAR	23.83
awc_mean_0_5	22.82
sand_mean_0_5	21.56
eFFP	20.97

Variable	% Inc. MSE
AHM	20.86
DD_18	20.75
RH	20.69
EXT	19.72
PAS	19.66
CMD	19.64
DD5	19.57
MAT	19.46
TD	18.56
Tave_sm	17.34
FFP	16.55
DD18	16.2
Tave_wt	15.88
SHM	15.22
DD_0	14.94
MWMT	13.35
srad	56.03
om_mean_0_5	48.58
wind	41.73
bd_mean_0_5	41.26

We find that the most important variables differ substantially from those that were most highly correlated with occurrence in our exploratory analyses. Given the robustness of the random forest procedure compared to a simple correlation, these importance values should carry more weight. However, we get no indication of the direction of a relationship from a variable's importance in a random forest, since it takes combinations of variables into account simultaneously. Therefore, the interpretation of whether each variable is positively or negatively related to occurrence should be made cautiously (though correlations should provide some guidance, especially if relationships are strong).

From the importance values, we can conclude that solar radiation (`srad`) and heat load index—both of which are based on slope and aspect of a hill—are very significant predictors and that a northeast-facing hillside which is cooler and retains more water is likely to be most favorable to habitat suitability due to these factors. Wind is also significant and, judging by the slight negative correlation in the data, areas that are more sheltered from wind may have higher suitability (or, conversely, they may simply happen to occur in areas of high wind). Other climatic variables indicate relatively mild winter temperatures and low summer precipitation levels are predictive, though empirical studies show that greater summer precipitation is highly beneficial for seedling survival (Shepperd & Battaglia 2002) and so this may be the result of later survivorship bias. Soil variables suggest that high organic content and fine particulate size are important, as are the clay and silt content, though the directionality of the latter two variables is unclear.

Prediction under Present Conditions

The main goal of our analyses is not the explanation of current occurrence, but the prediction of habitat suitability elsewhere (Elith & Leathwick 2009), specifically within areas where planting efforts will be undertaken. To accomplish this, I used the model developed using random forests with occurrence data to predict the relative probability of occurrence across South Dakota under the estimated parameters. This gives us an indication of how well suited other environments are to the survival of Rocky Mountain ponderosa pine and can help guide planting efforts.

Maps of habitat suitability under present conditions are shown following these introductory and summary sections. The first plot for each location depicts only the probability of occurrence, for clarity and visibility, while the following plot includes roads, settlements, and waterways as points of reference. Note that, to aid in the differentiation of areas within each plot, color scales are based upon local maximum and minimum values, so color values are only comparable within, and not between, plots.

It is necessary to note that there is an area in east-central Pine Ridge, northeast of the town of Allen, that shows some erroneous patterns in the soils data. This is visible as a roughly square-shaped, cardinally-aligned border of higher predicted occurrence relative to the the surrounding area. After cross-checking with topography and USDA Web Soil Survey data, this appears to be due to an error in the POLARIS algorithm’s interpolation. Therefore, the predictions for this area in the maps of present and future conditions are not considered to be accurate and should not be used. There do not appear to be similar issues with the data in any other areas.

Prediction under Projected Future Conditions

In addition to present conditions, it is relevant for us to consider the survival of the seedlings that are being planted well into the future, given the desire for long-term impact and continuance of benefits. As detailed previously in the Measure phase, we have reliable consensus projections of future climate conditions in the 2050s under two emissions scenarios: RCP4.5, which takes into account some curtailing and regulation of global CO_2 emissions, and RCP8.5, which represents the continuation of the current rate of increase in CO_2 emissions levels absent any meaningful intervention. Of the two, the last several years of data have most closely followed the predictions of the RCP8.5 scenario and this trajectory seems unlikely to change given present political narratives, so I advise treating this as the most realistic projection. However, there is little real difference in the areas of interest based on either scenario. Plots of habitat suitability under these projected future conditions are shown following those of present conditions and use the same conventions.

Predictions for Previous Planting Locations

Presented below are the mean occurrence probabilities under present and future conditions for the 100 data points that we have for surviving seedlings of prior planting seasons. These points are all from a relatively small area south of Oglala Lake and include only seedlings that have survived, so are not a representative sample. The extracted values show that habitability is good for the area where these seedlings were planted under current conditions, but also show that these conditions are likely to become substantially less favorable in the future. Therefore, continued monitoring of these seedlings and those planted elsewhere and in future seasons would be worthwhile, to address issues as they arise and help ensure long-term survival.

	Present Conditions	*RCP4.5 in 2050*	*RCP8.5 in 2050*
Mean Probability of Occurrence	0.9513	0.0735	0.0701
Standard Deviation	0.0471	0.0198	0.0192

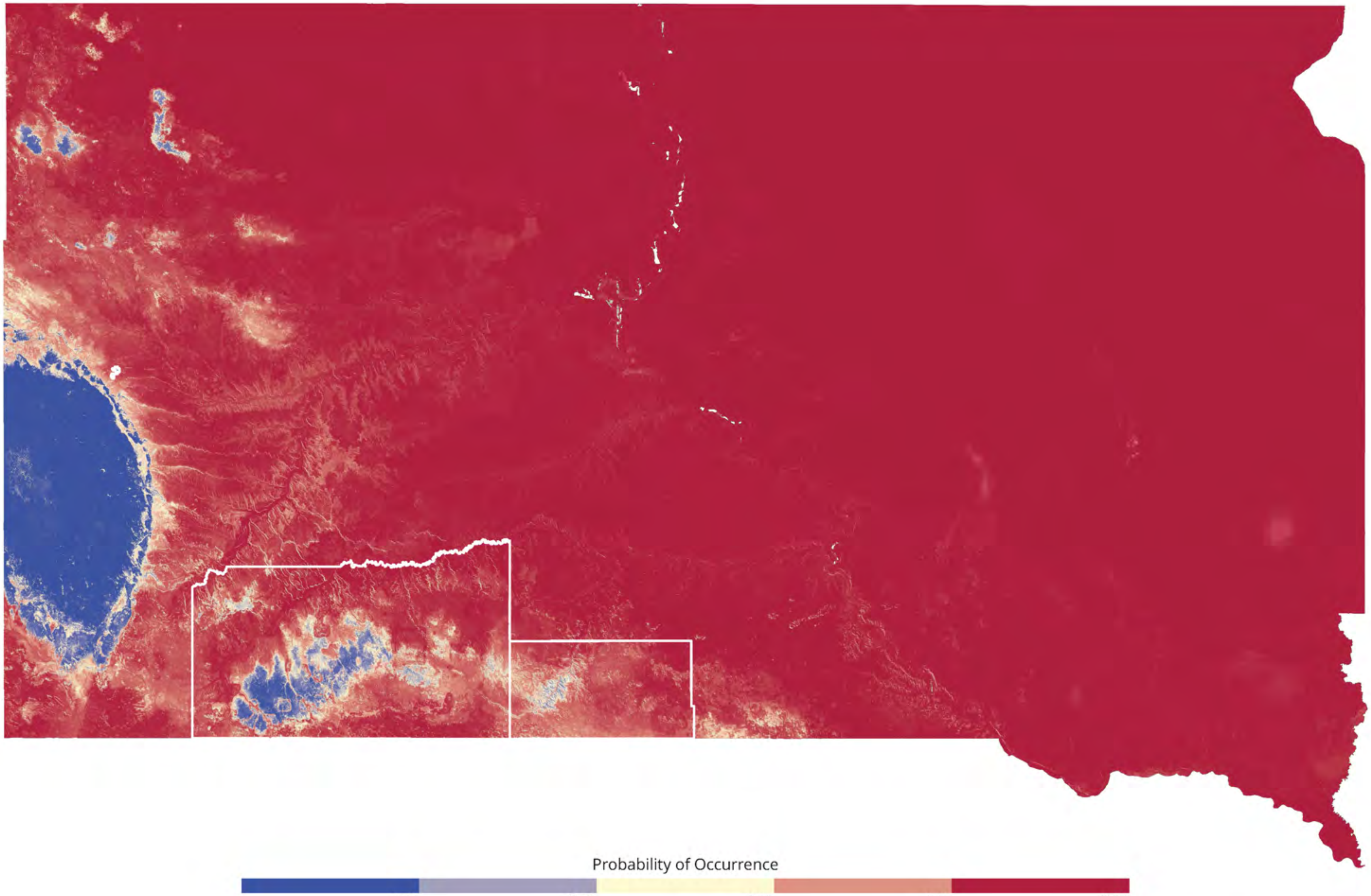
Summary

The findings of this habitat suitability modeling effort are among the most significant in this report. They illustrate that the habitat suitability and likelihood of survival for seedlings planted this year and in previous seasons decline dramatically across much of South Dakota and the Pine Ridge Reservation under projected future conditions. The few refugia in Pine Ridge where suitability remains highest appear to be in areas of high elevation on northeasterly slopes. This pattern mirrors the reduction in suitable range within the Black Hills, which has historically been one of the few areas in South Dakota where Rocky Mountain ponderosa pine have continued to thrive—likely in part due to their protected status within a National Forest, along with the favorable environmental conditions.

However, this is not to say that less suitable areas in Pine Ridge could not show successful planting efforts, or that the areas in which planting has been done previously will not continue to thrive. Rather, the awareness of present and future habitat suitability imparted by these analyses illustrate the importance of increasing the chances of seedling survival by any means possible—"stacking the deck in their favor," so to speak—rather than relying on areas that have been suitable historically. We will need to adapt our methods in the face of this information, as the trees will need to adapt to the changing climate.

South Dakota

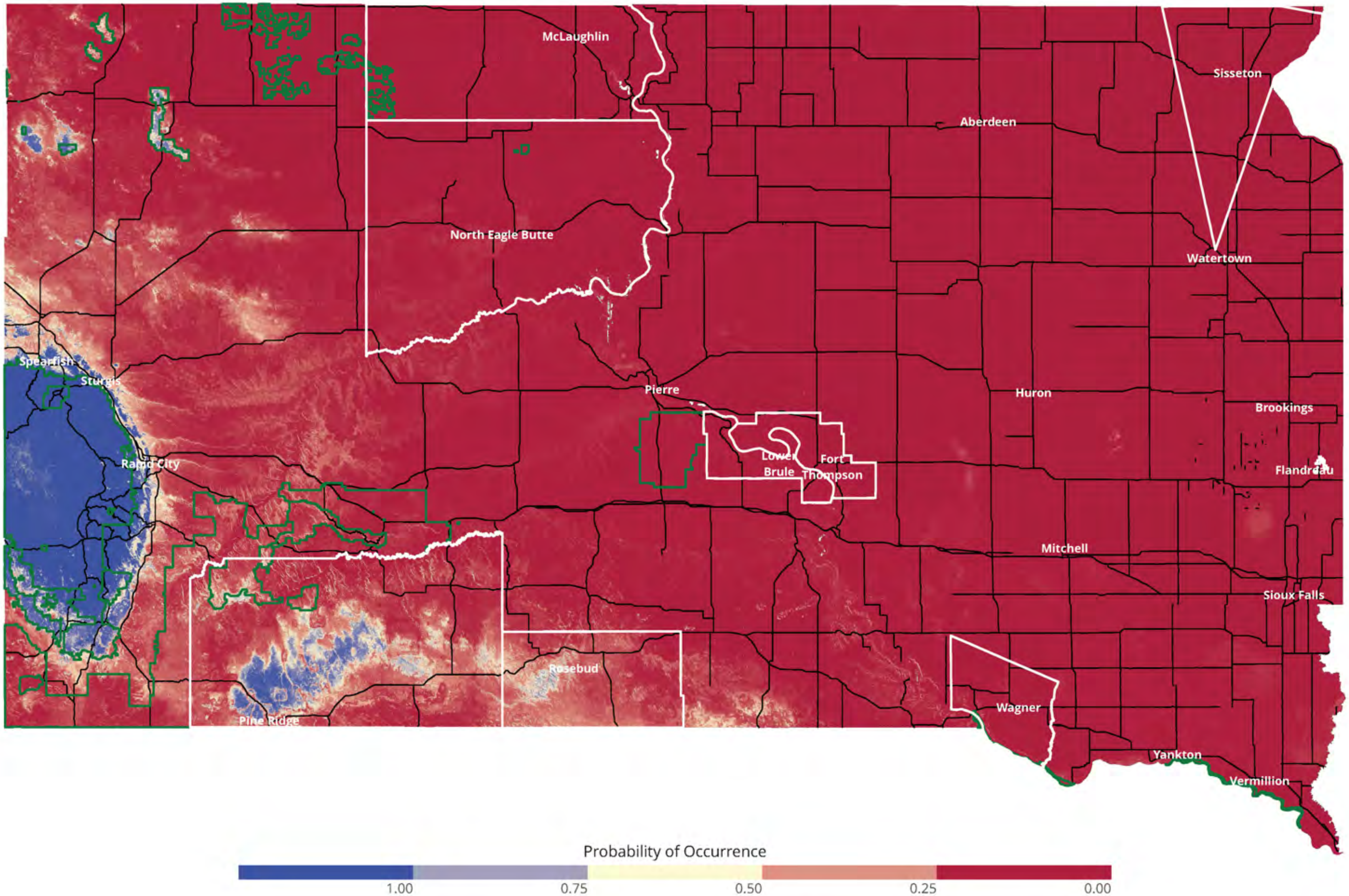
Random Forest Prediction of Pinus ponderosa var. scopulorum Habitat Suitability under Present Conditions



Author: Richard E.W. Berl; Map Theme: Timo Grossenbacher; Font: Open Sans; Data Sources: Maguire et al. 2018, AdaptWest (PRISM), SRTM, POLARIS, WorldClim2, TIGER/Line

South Dakota

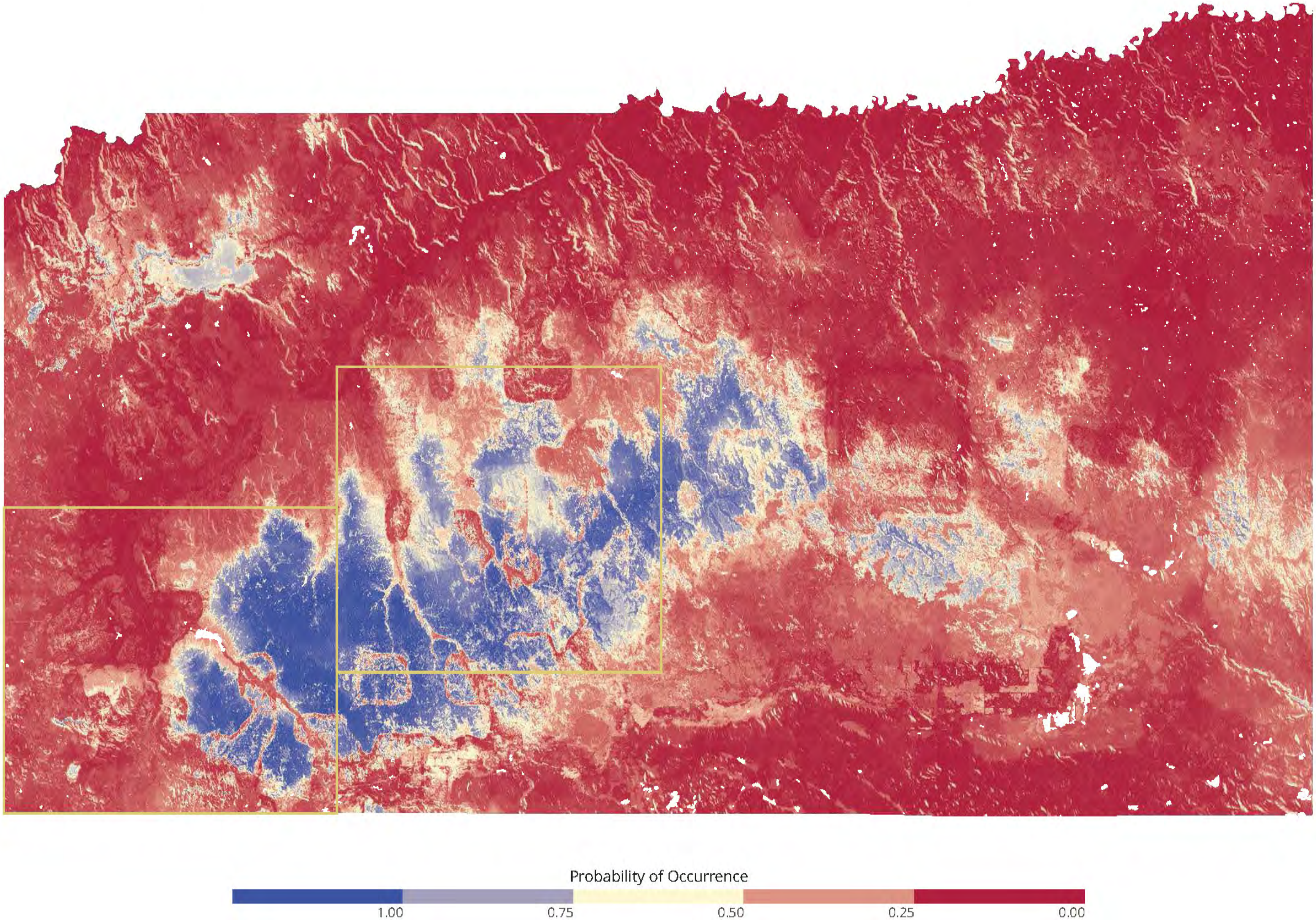
Random Forest Prediction of Pinus ponderosa var. scopulorum Habitat Suitability under Present Conditions



Author: Richard E.W. Berl; Map Theme: Timo Grossenbacher; Font: Open Sans; Data Sources: Maguire et al. 2018, AdaptWest (PRISM), SRTM, POLARIS, WorldClim2, TIGER/Line, NPS, USFS

Pine Ridge Reservation and Trust Land

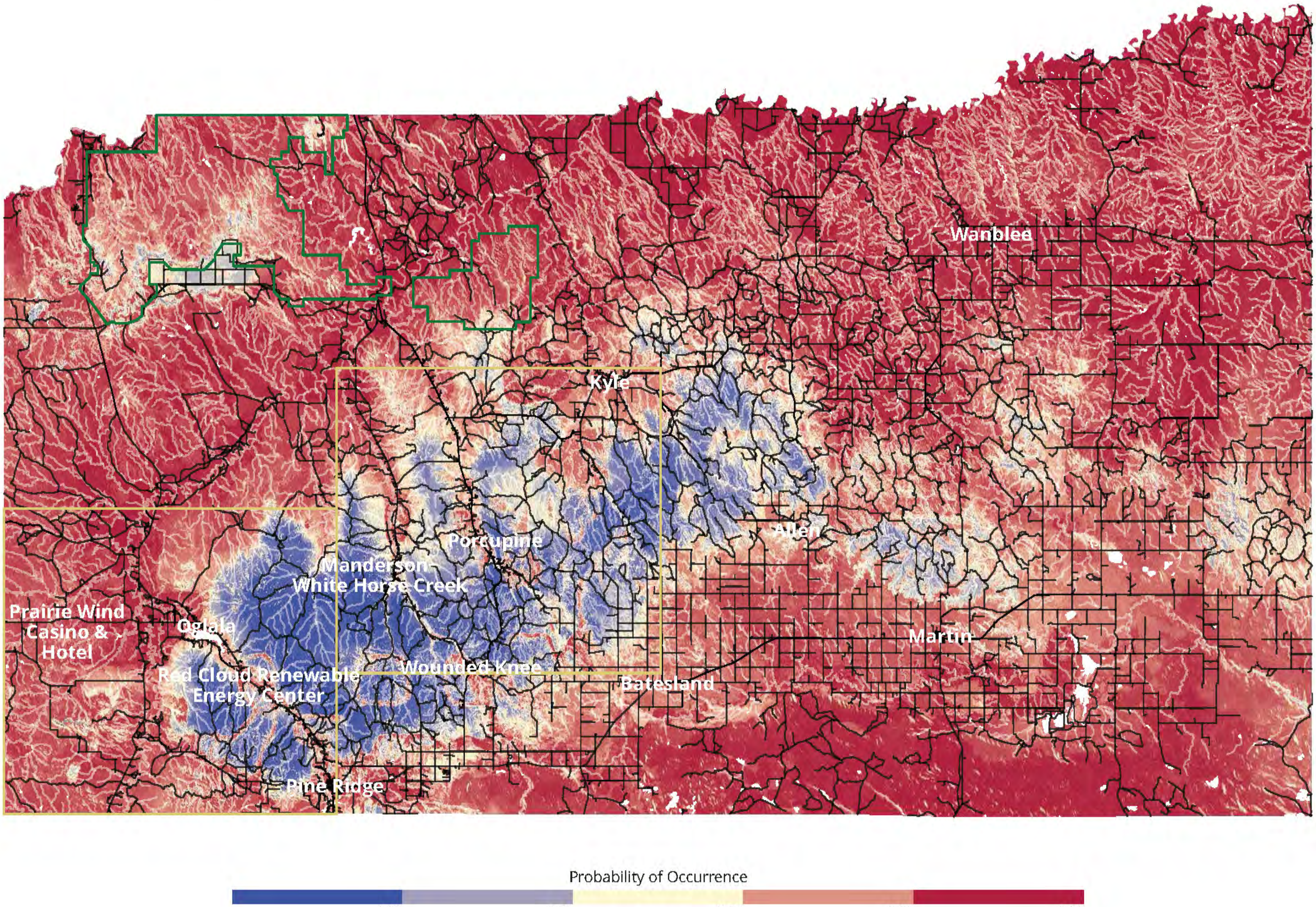
Random Forest Prediction of Pinus ponderosa var. scopulorum Habitat Suitability under Present Conditions



Author: Richard E.W. Berl; Map Theme: Timo Grossenbacher; Font: Open Sans; Data Sources: Maguire et al. 2018, AdaptWest (PRISM), SRTM, POLARIS, WorldClim2, TIGER/Line

Pine Ridge Reservation and Trust Land

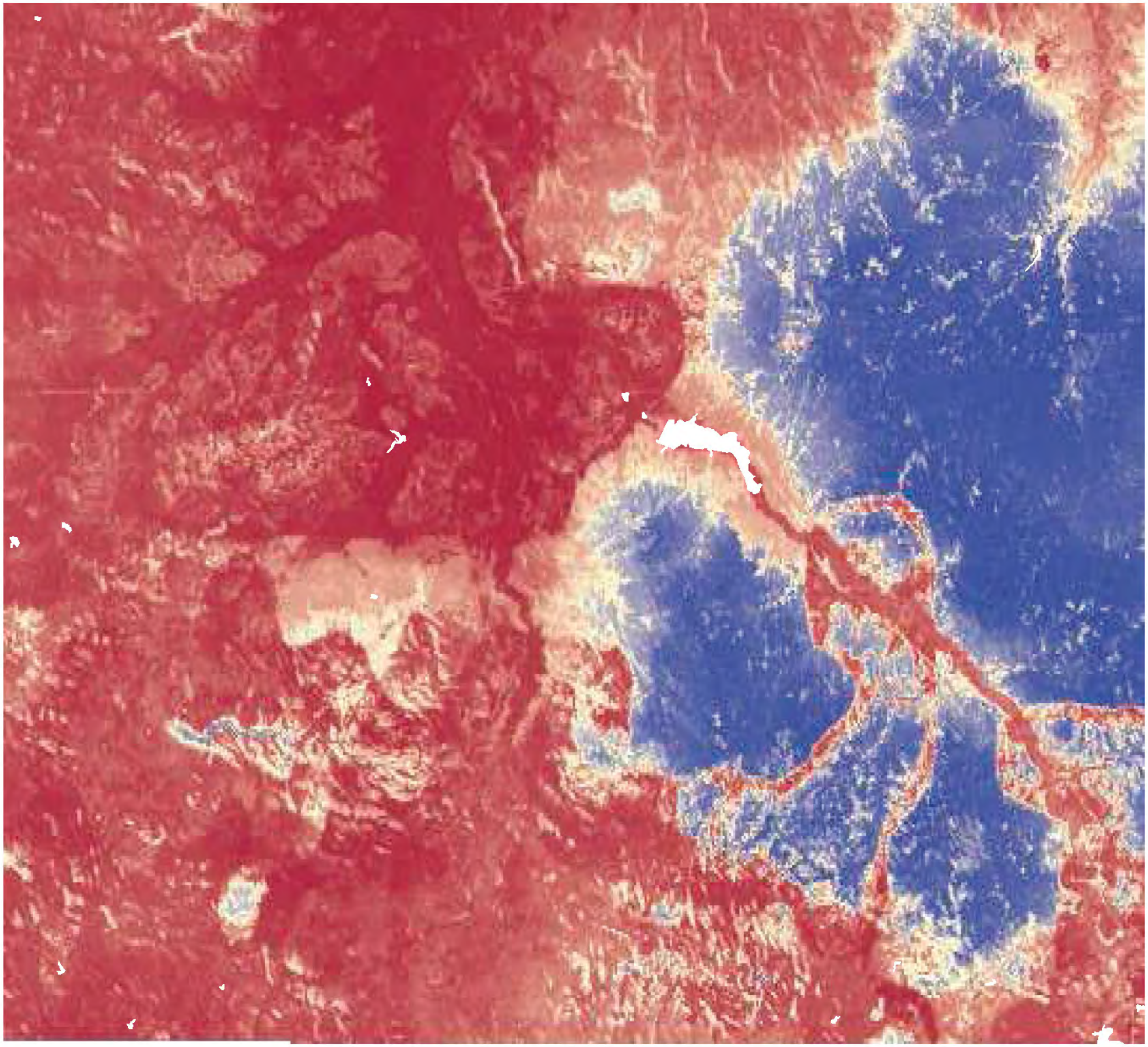
Random Forest Prediction of Pinus ponderosa var. scopulorum Habitat Suitability under Present Conditions



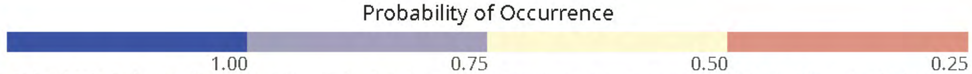
Author: Richard E.W. Berl; Map Theme: Timo Grossenbacher; Font: Open Sans; Data Sources: Maguire et al. 2018, AdaptWest (PRISM), SRTM, POLARIS, WorldClim2, TIGER/Line

Oglala Planting Area in Pine Ridge Reservation

Random Forest Prediction of Pinus ponderosa var. scopulorum Habitat Suitability under Present Conditions



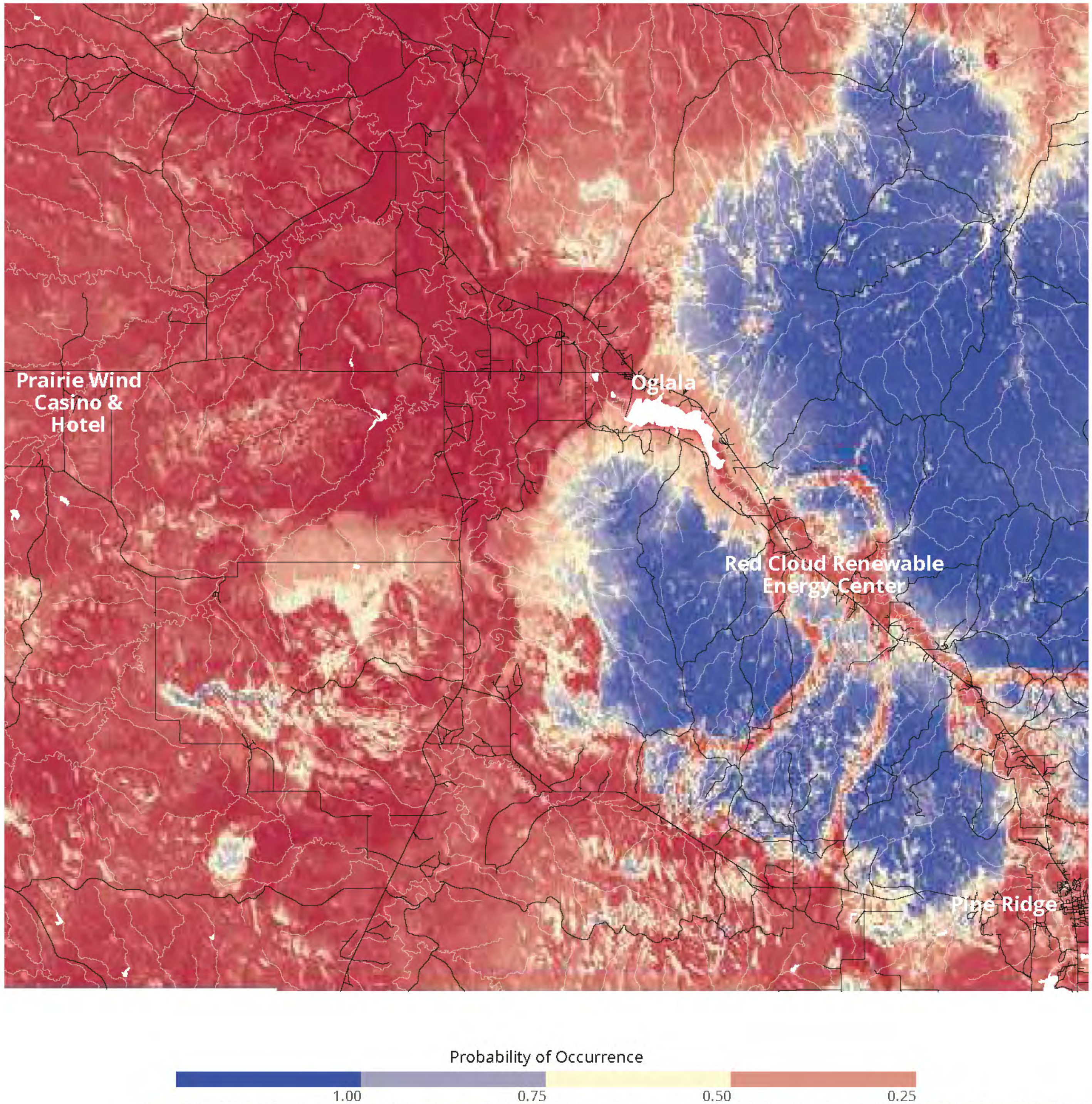
Probability of Occurrence



Author: Richard E.W. Berl; Map Theme: Timo Grossenbacher; Font: Open Sans; Data Sources: Maguire et al. 2018, AdaptWest (PRISM), SRTM, POLARIS, WorldClim2, TIGER/Line

Oglala Planting Area in Pine Ridge Reservation

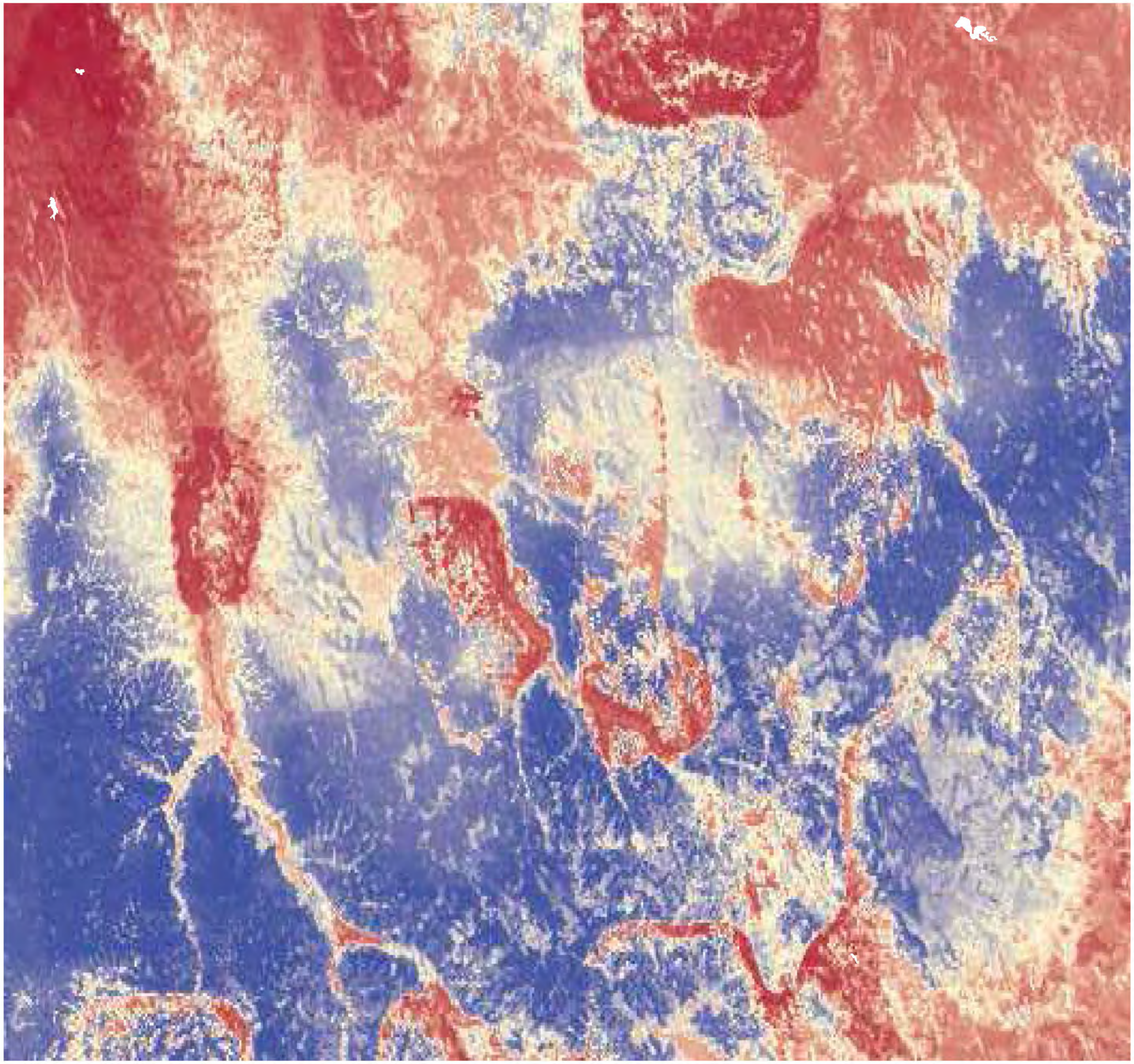
Random Forest Prediction of Pinus ponderosa var. scopulorum Habitat Suitability under Present Conditions



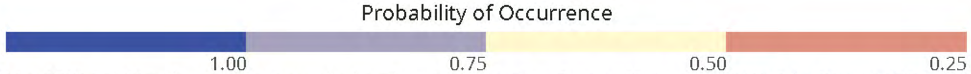
Author: Richard E.W. Berl; Map Theme: Timo Grossenbacher; Font: Open Sans; Data Sources: Maguire et al. 2018, AdaptWest (PRISM), SRTM, POLARIS, WorldClim2, TIGER/Line

Porcupine Planting Area in Pine Ridge Reservation

Random Forest Prediction of Pinus ponderosa var. scopulorum Habitat Suitability under Present Conditions



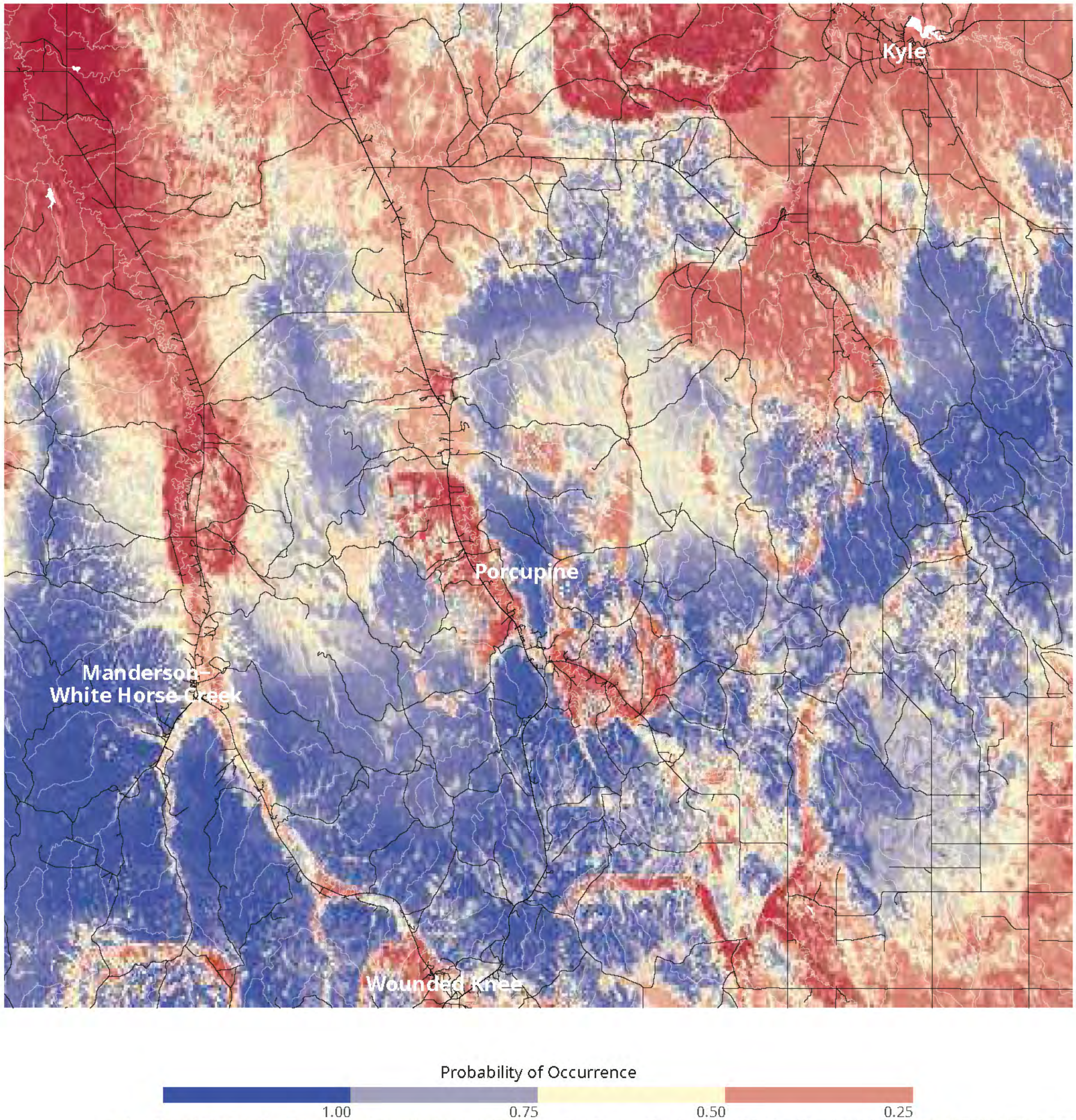
Probability of Occurrence



Author: Richard E.W. Berl; Map Theme: Timo Grossenbacher; Font: Open Sans; Data Sources: Maguire et al. 2018, AdaptWest (PRISM), SRTM, POLARIS, WorldClim2, TIGER/Line

Porcupine Planting Area in Pine Ridge Reservation

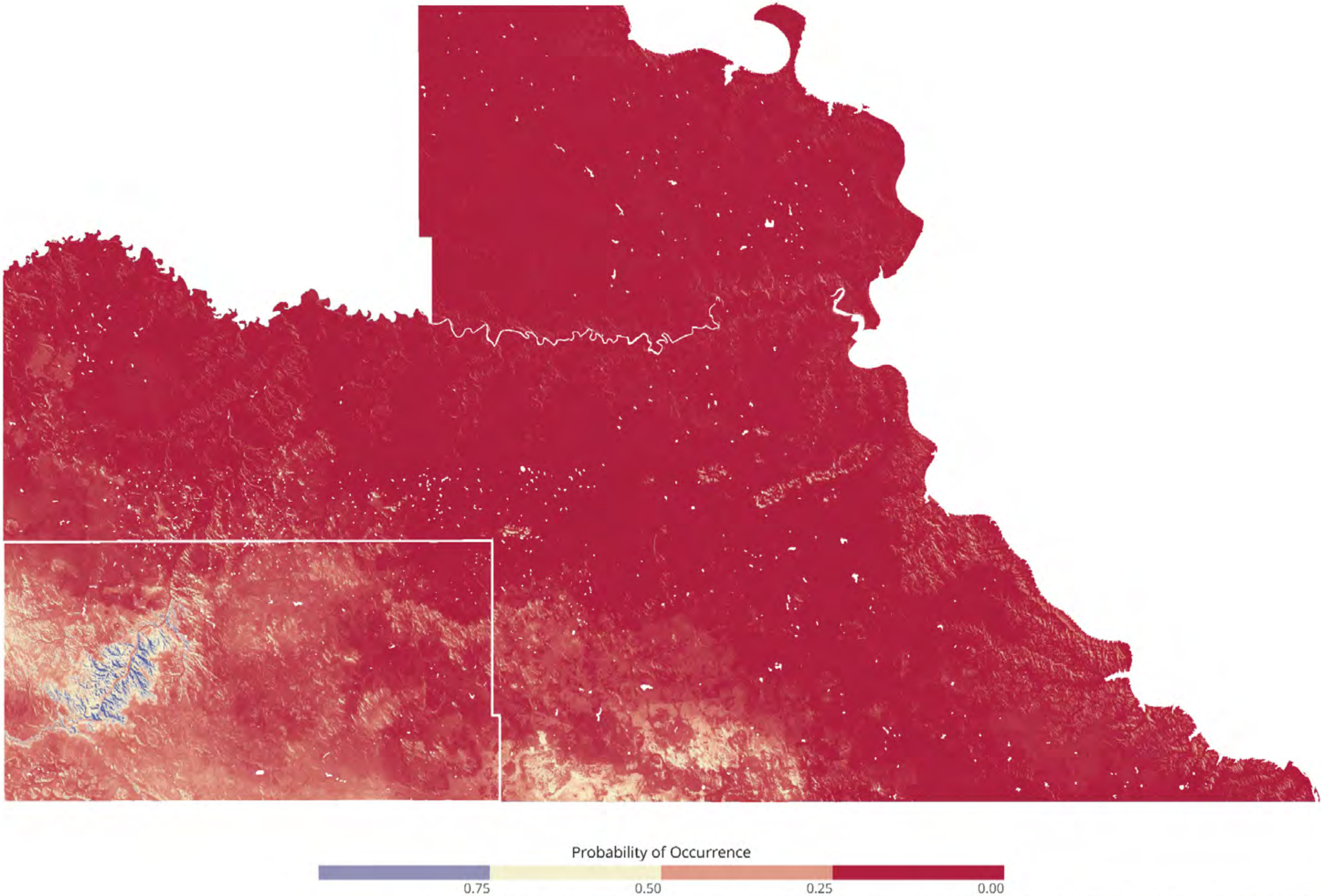
Random Forest Prediction of Pinus ponderosa var. scopulorum Habitat Suitability under Present Conditions



Author: Richard E.W. Berl; Map Theme: Timo Grossenbacher; Font: Open Sans; Data Sources: Maguire et al. 2018, AdaptWest (PRISM), SRTM, POLARIS, WorldClim2, TIGER/Line

Rosebud Reservation and Trust Land

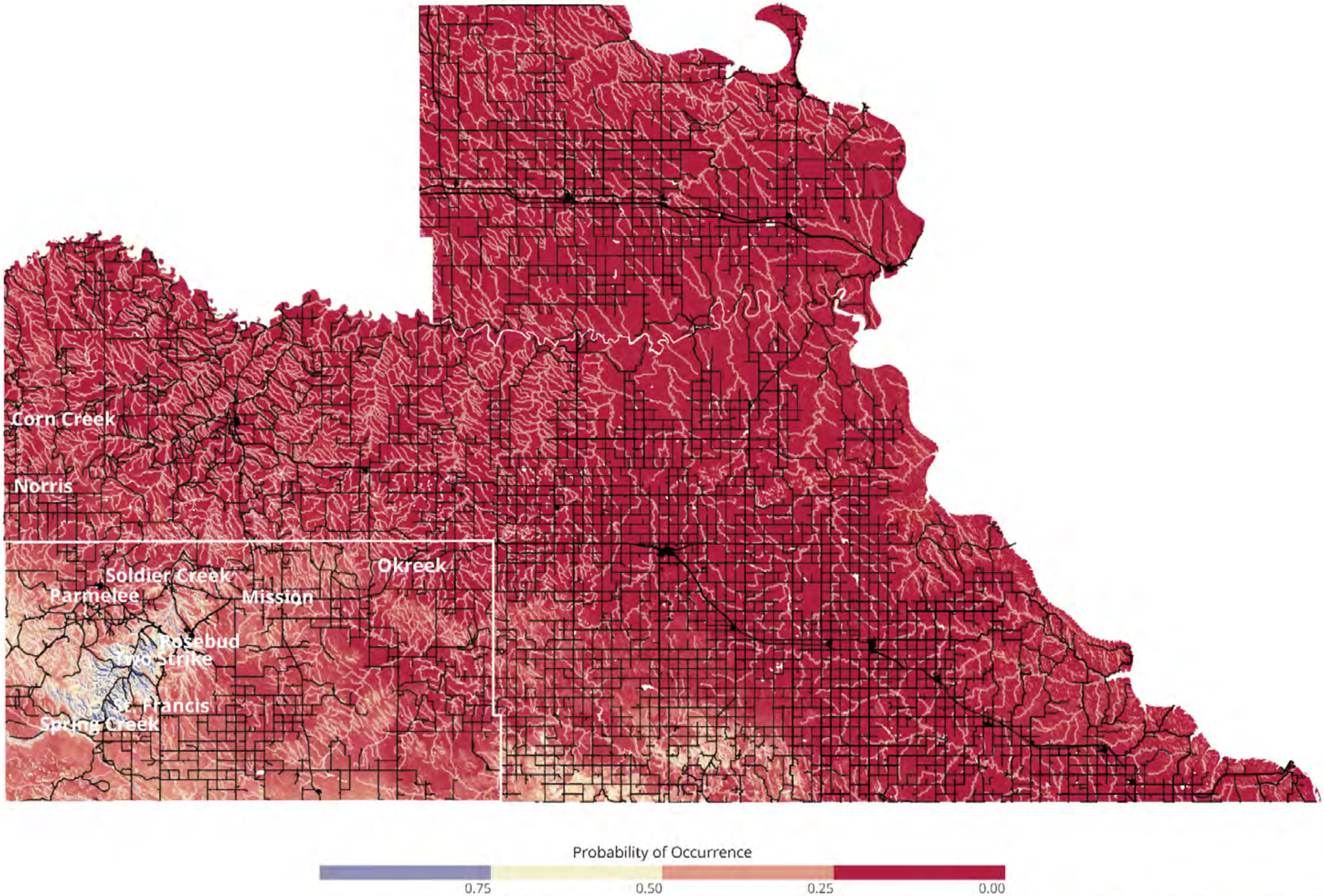
Random Forest Prediction of Pinus ponderosa var. scopulorum Habitat Suitability under Present Conditions



Author: Richard E.W. Berl; Map Theme: Timo Grossenbacher; Font: Open Sans; Data Sources: Maguire et al. 2018, AdaptWest (PRISM), SRTM, POLARIS, WorldClim2, TIGER/Line

Rosebud Reservation and Trust Land

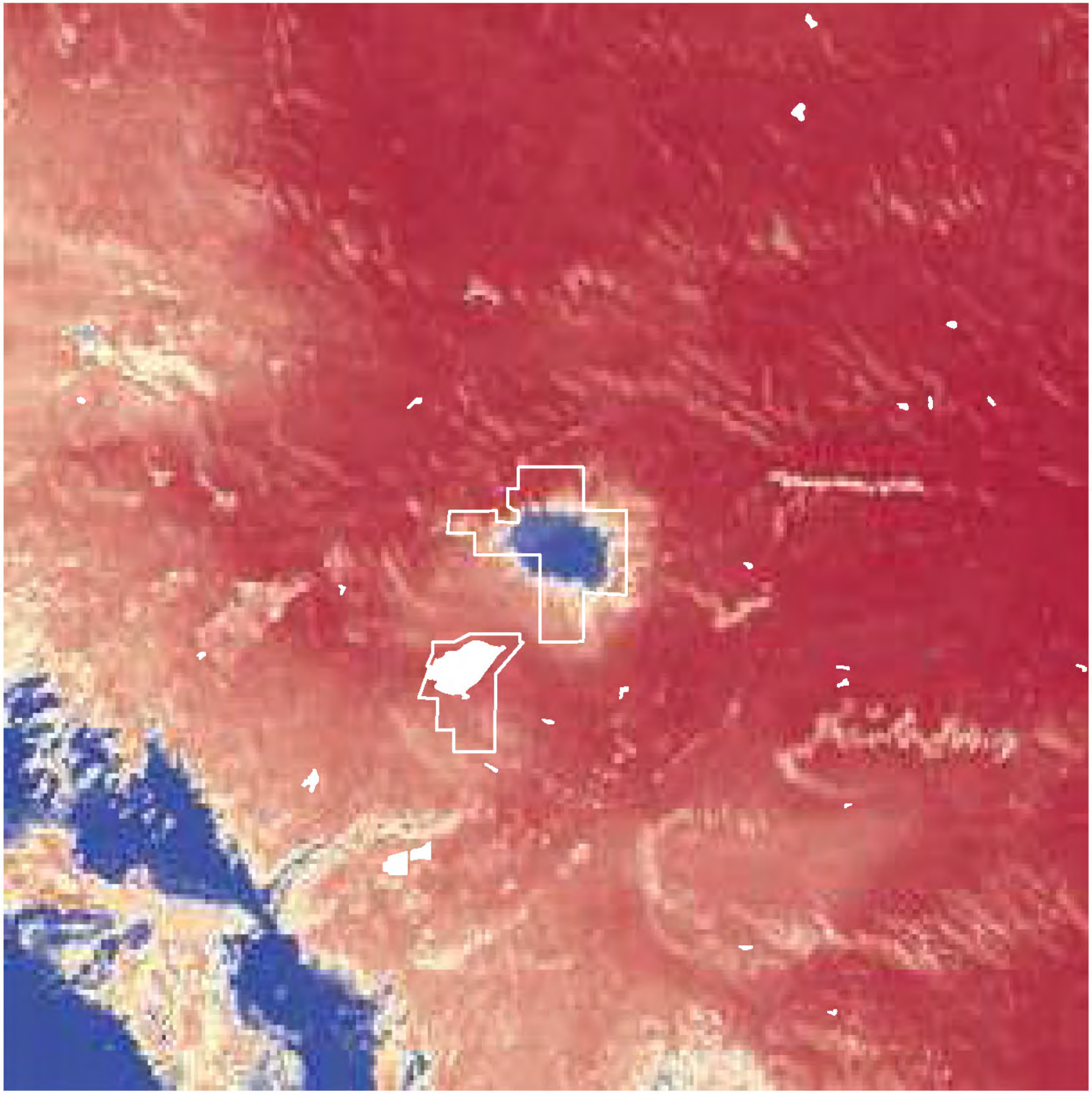
Random Forest Prediction of Pinus ponderosa var. scopulorum Habitat Suitability under Present Conditions



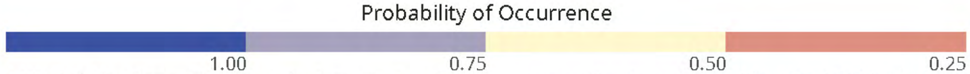
Author: Richard E.W. Berl; Map Theme: Timo Grossenbacher; Font: Open Sans; Data Sources: Maguire et al. 2018, AdaptWest (PRISM), SRTM, POLARIS, WorldClim2, TIGER/Line

Bear Butte

Random Forest Prediction of Pinus ponderosa var. scopulorum Habitat Suitability under Present Conditions



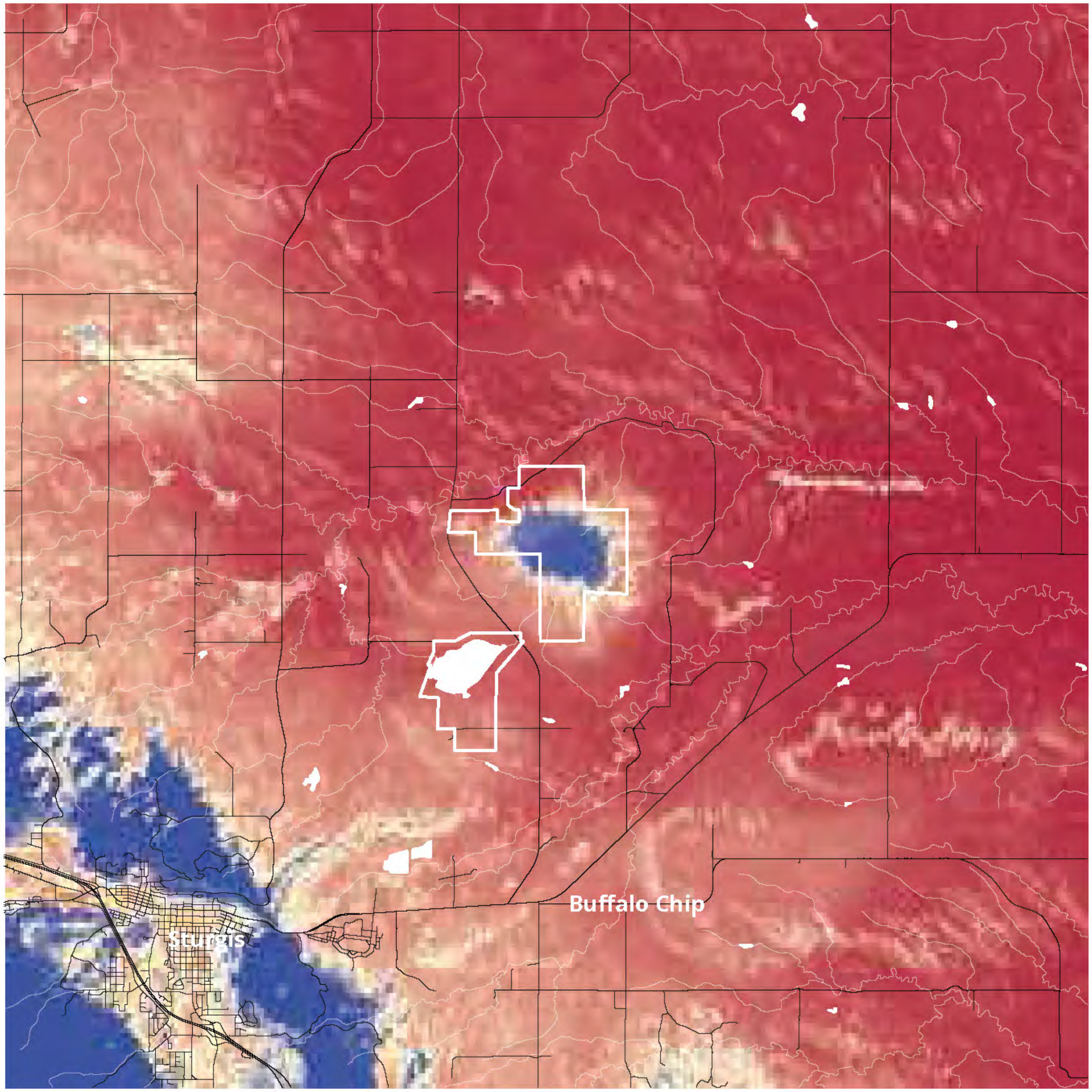
Probability of Occurrence



Author: Richard E.W. Berl; Map Theme: Timo Grossenbacher; Font: Open Sans; Data Sources: Maguire et al. 2018, AdaptWest (PRISM), SRTM, POLARIS, WorldClim2, TIGER/Line

Bear Butte

Random Forest Prediction of Pinus ponderosa var. scopulorum Habitat Suitability under Present Conditions



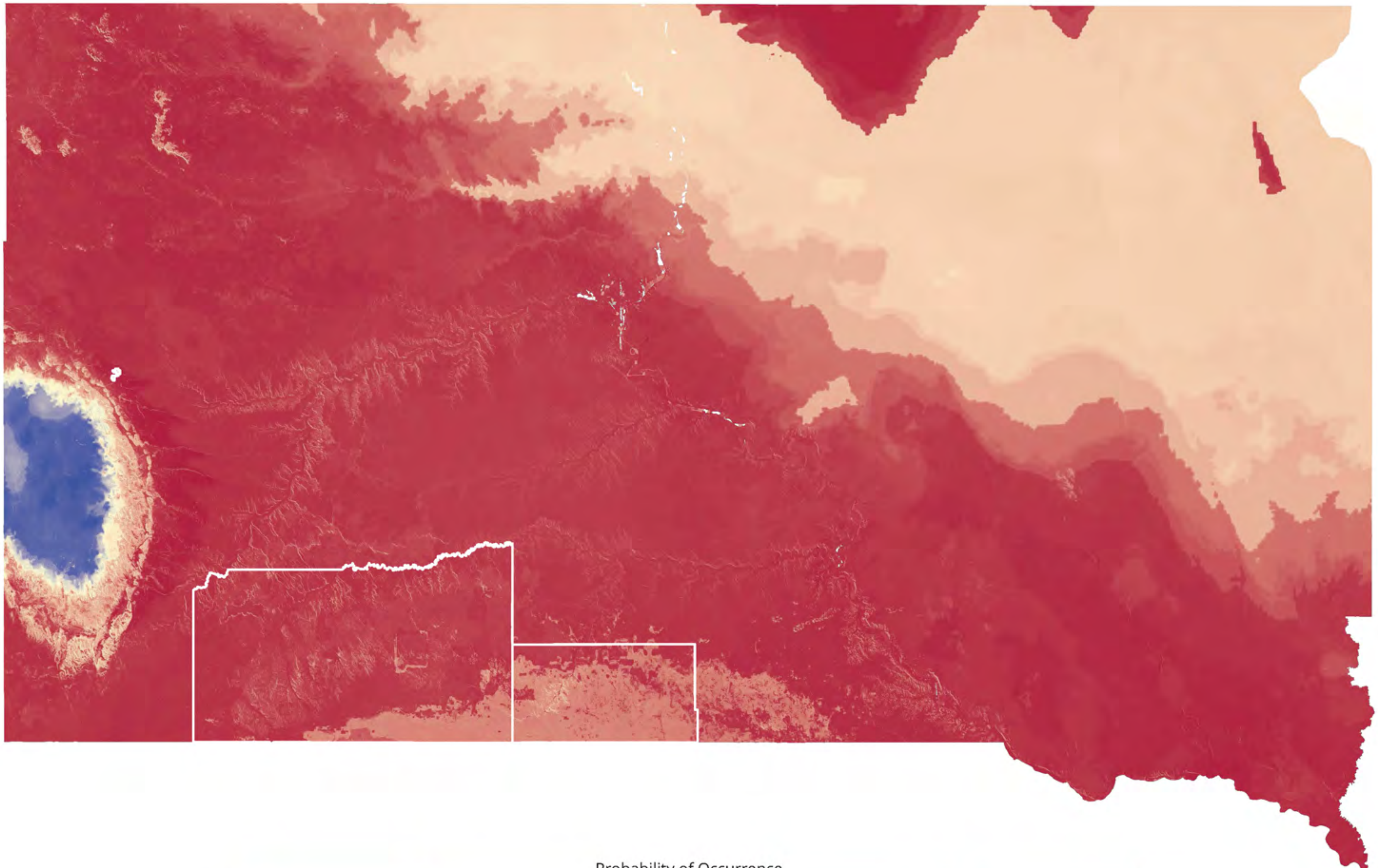
Probability of Occurrence

1.00 0.75 0.50 0.25

Author: Richard E.W. Berl; Map Theme: Timo Grossenbacher; Font: Open Sans; Data Sources: Maguire et al. 2018, AdaptWest (PRISM), SRTM, POLARIS, WorldClim2, TIGER/Line

South Dakota

Random Forest Prediction of Pinus ponderosa var. scopulorum Habitat Suitability in 2050 under RCP4.5 Climate Scenario



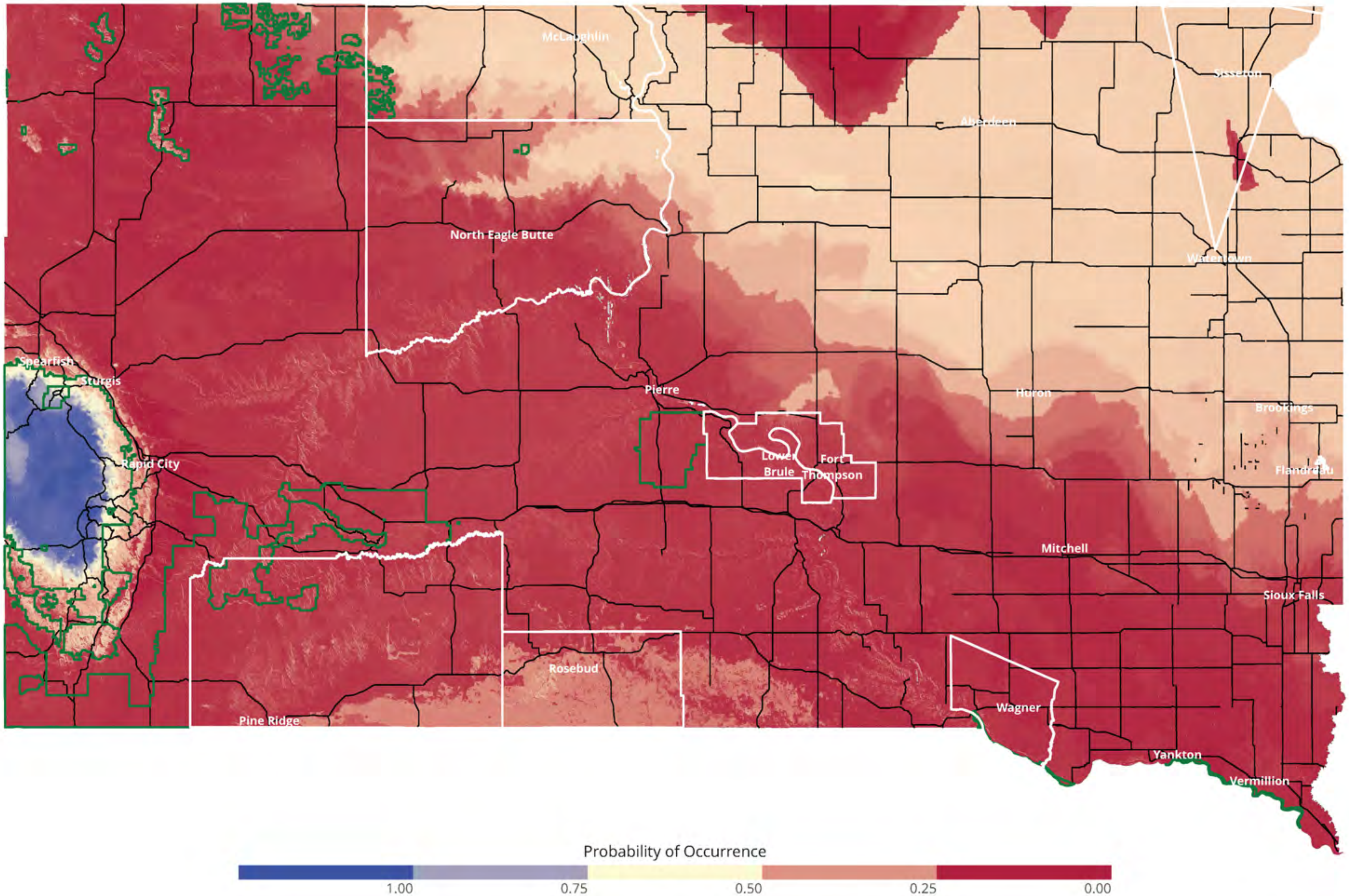
Probability of Occurrence

1.00 0.75 0.50 0.25 0.00

Author: Richard E.W. Berl; Map Theme: Timo Grossenbacher; Font: Open Sans; Data Sources: Maguire et al. 2018, AdaptWest, SRTM, POLARIS, WorldClim2, TIGER/Line

South Dakota

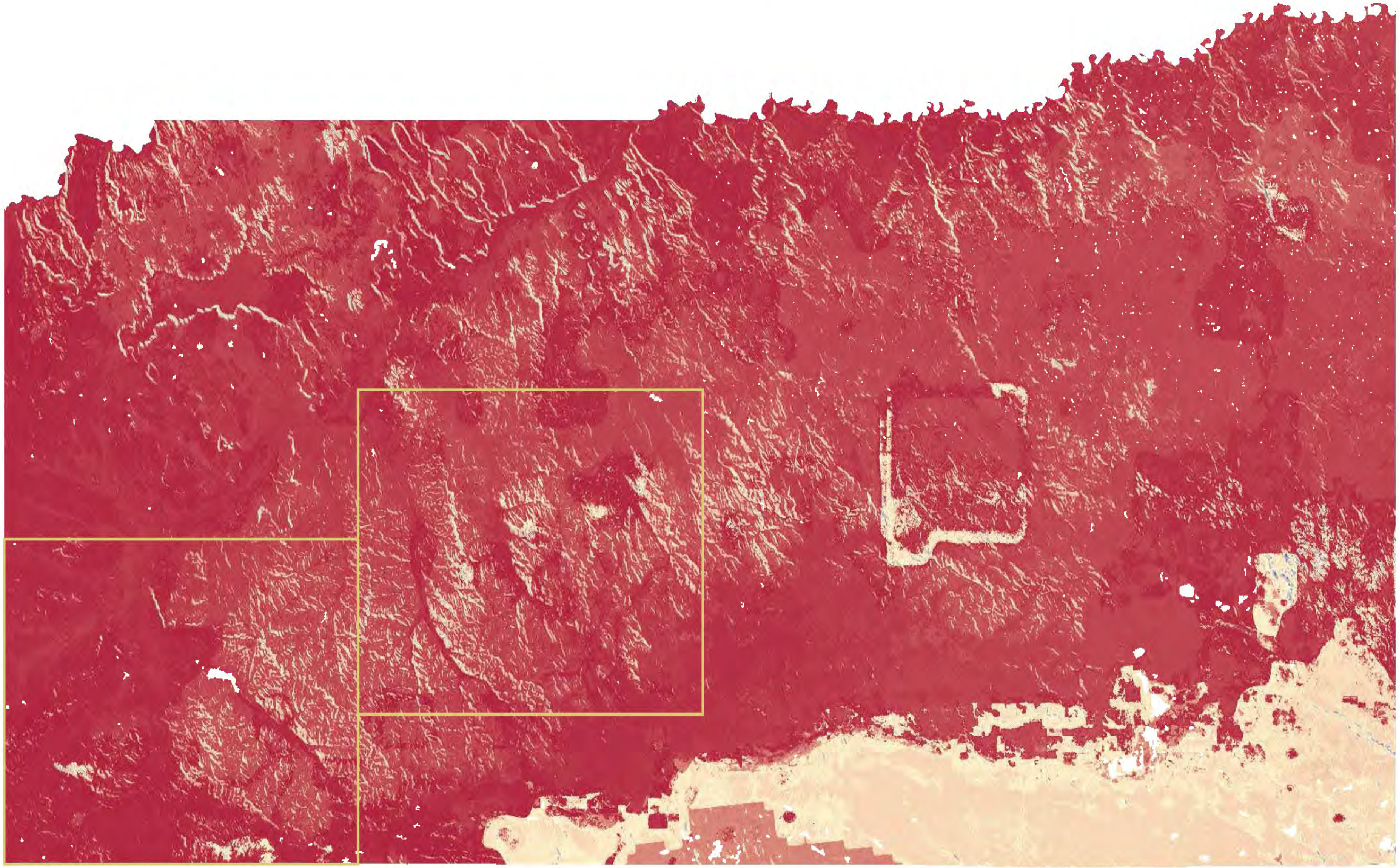
Random Forest Prediction of Pinus ponderosa var. scopulorum Habitat Suitability in 2050 under RCP4.5 Climate Scenario

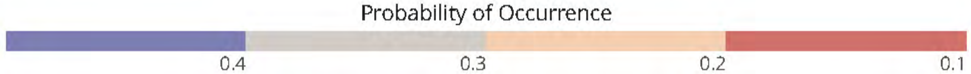


Author: Richard E.W. Berl; Map Theme: Timo Grossenbacher; Font: Open Sans; Data Sources: Maguire et al. 2018, AdaptWest, SRTM, POLARIS, WorldClim2, TIGER/Line, NPS, USFS

Pine Ridge Reservation and Trust Land

Random Forest Prediction of Pinus ponderosa var. scopulorum Habitat Suitability in 2050 under RCP4.5 Climate Scenario

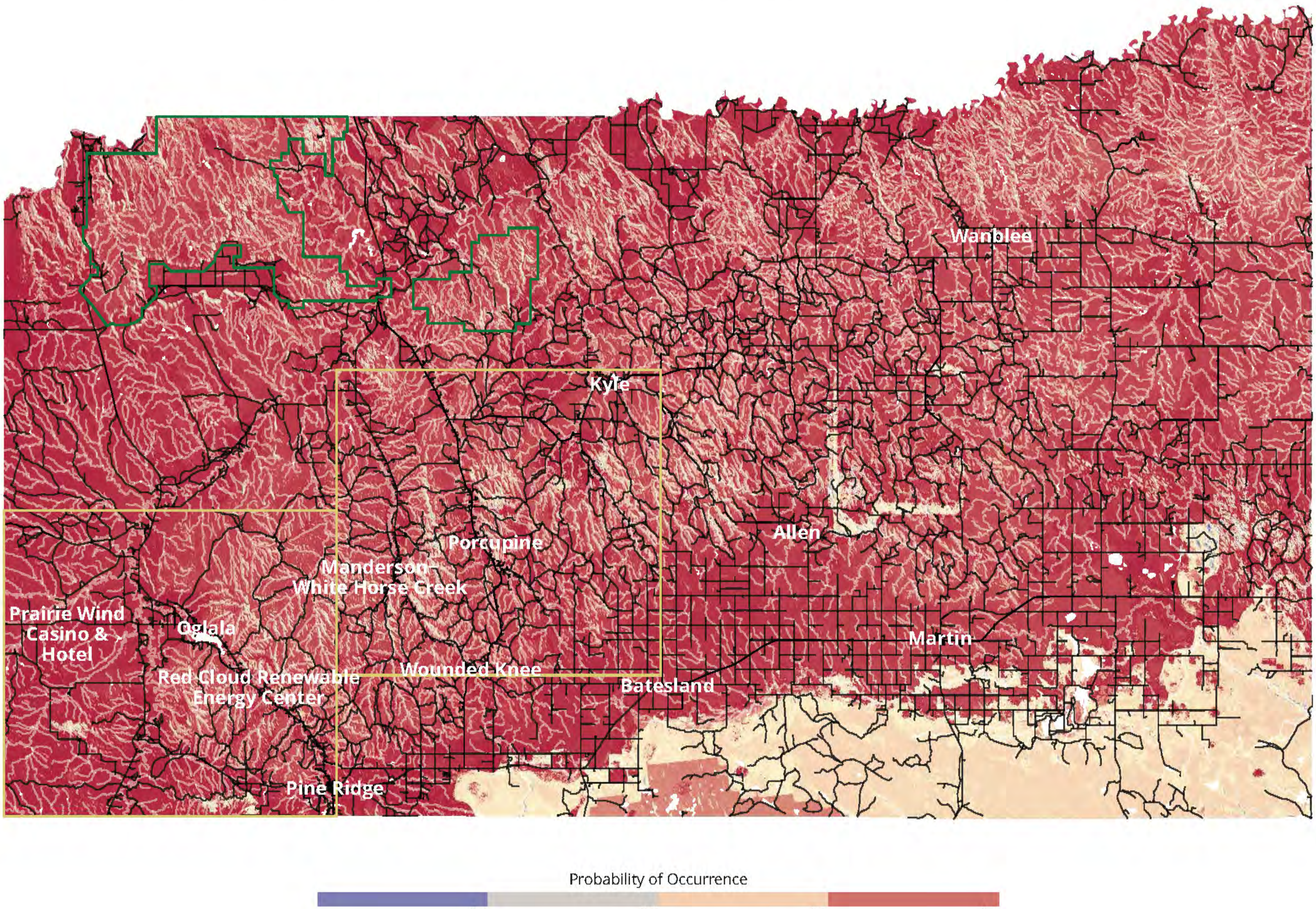




Author: Richard E.W. Berl; Map Theme: Timo Grossenbacher; Font: Open Sans; Data Sources: Maguire et al. 2018, AdaptWest, SRTM, POLARIS, WorldClim2, TIGER/Line

Pine Ridge Reservation and Trust Land

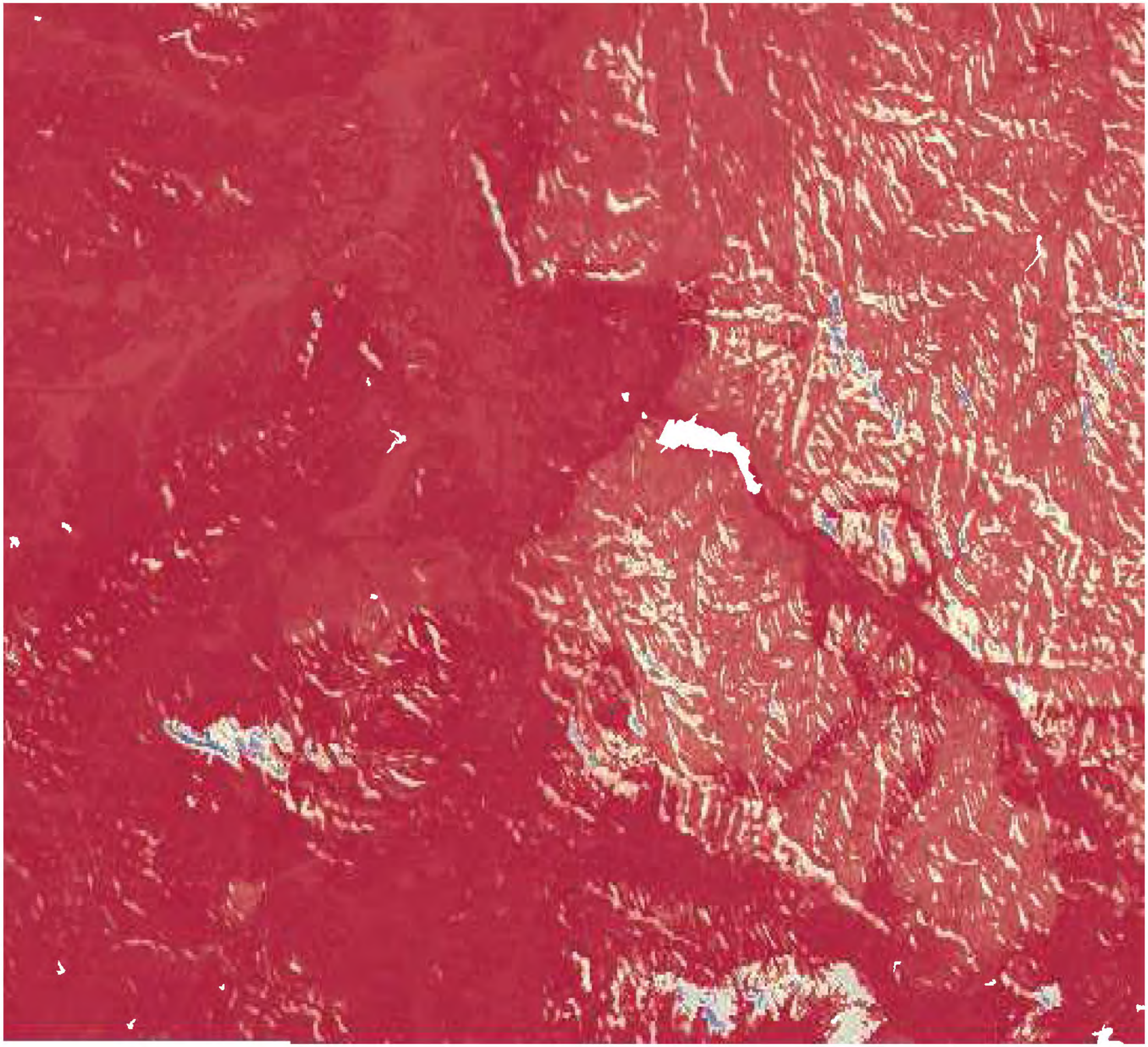
Random Forest Prediction of Pinus ponderosa var. scopulorum Habitat Suitability in 2050 under RCP4.5 Climate Scenario



Author: Richard E.W. Berl; Map Theme: Timo Grossenbacher; Font: Open Sans; Data Sources: Maguire et al. 2018, AdaptWest, SRTM, POLARIS, WorldClim2, TIGER/Line

Oglala Planting Area in Pine Ridge Reservation

Random Forest Prediction of Pinus ponderosa var. scopulorum Habitat Suitability under RCP4.5 Climate Scenario

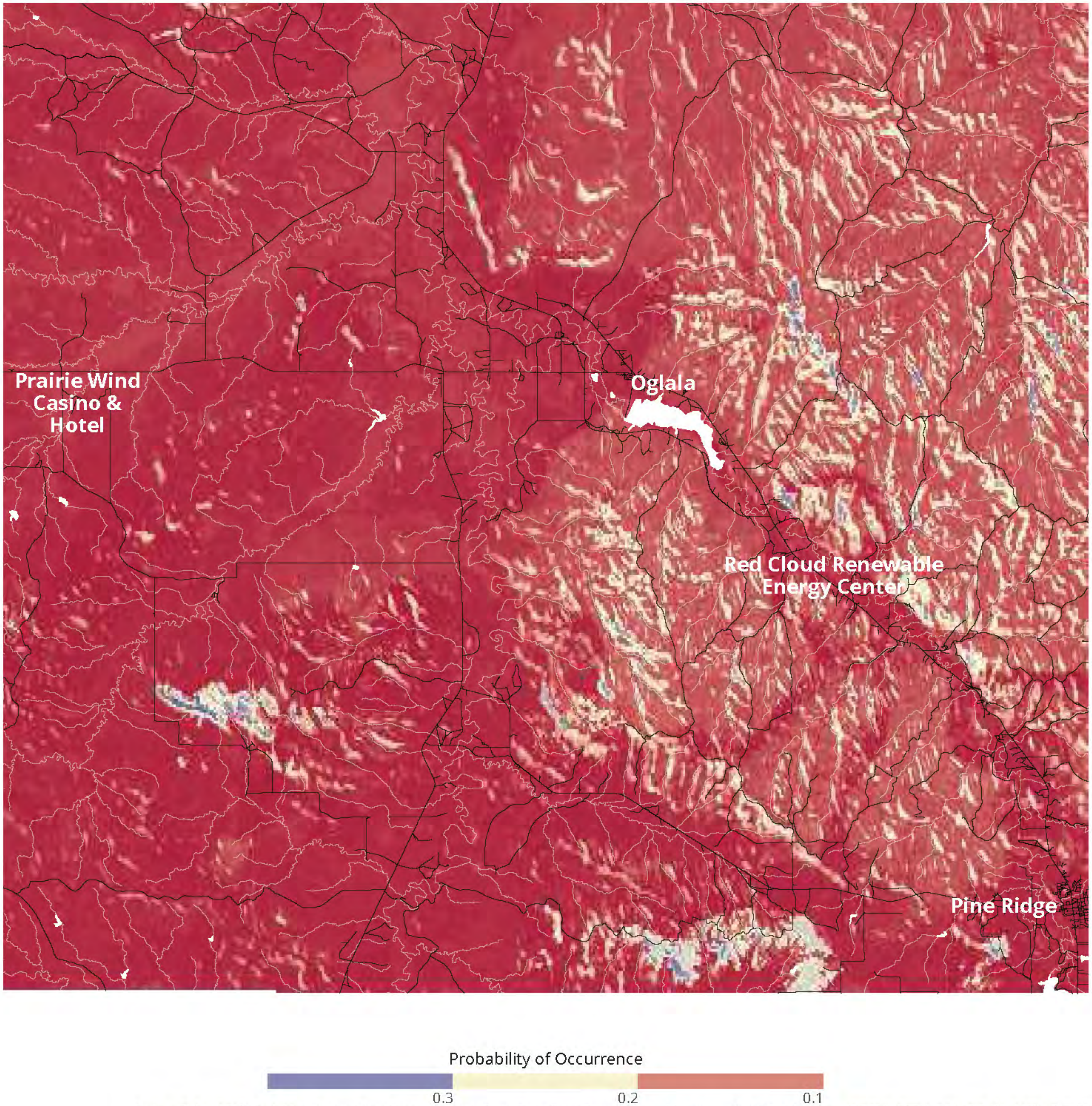


Probability of Occurrence

Author: Richard E.W. Berl; Map Theme: Timo Grossenbacher; Font: Open Sans; Data Sources: Maguire et al. 2018, AdaptWest (PRISM), SRTM, POLARIS, WorldClim2, TIGER/Line

Oglala Planting Area in Pine Ridge Reservation

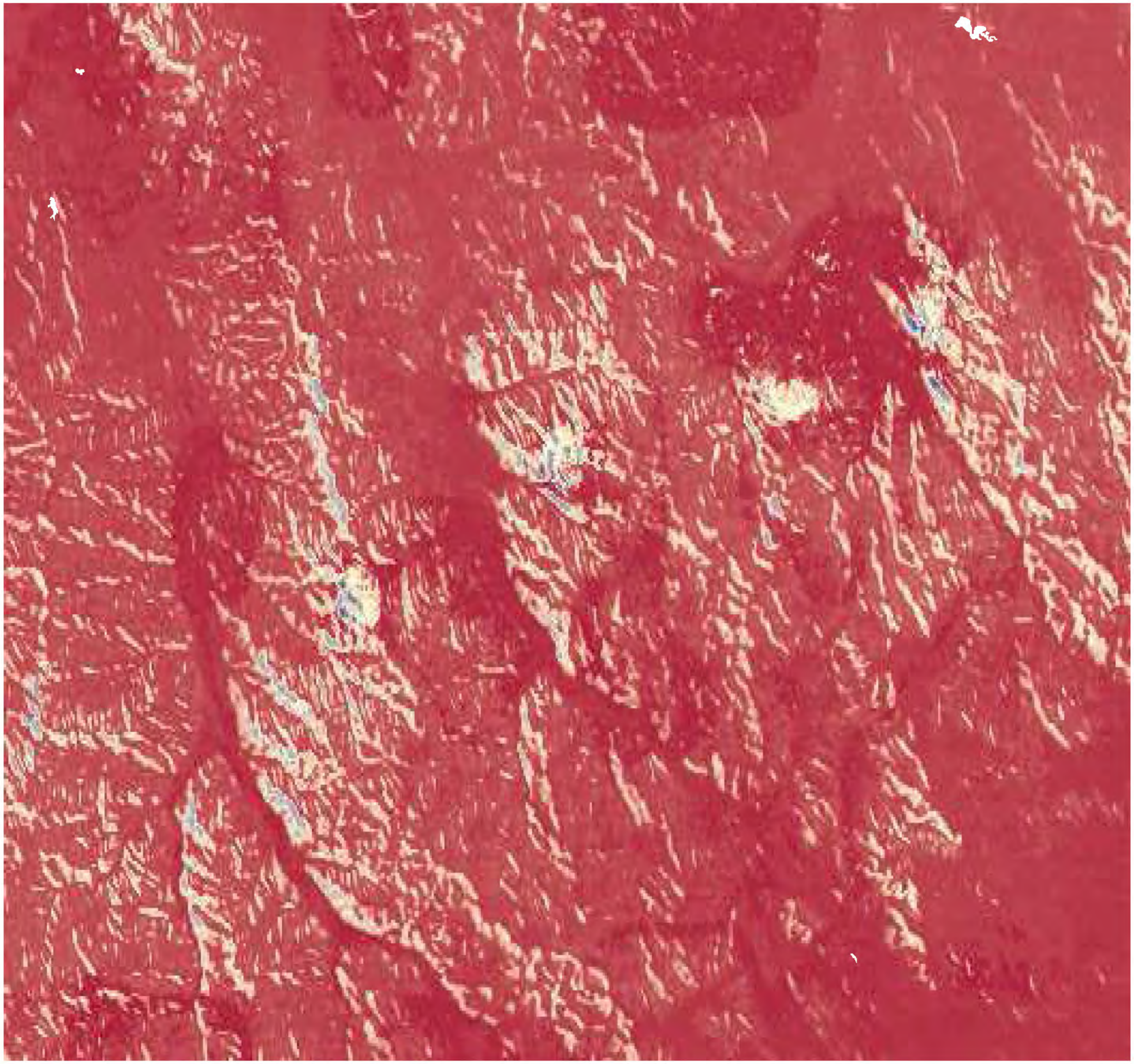
Random Forest Prediction of Pinus ponderosa var. scopulorum Habitat Suitability under RCP4.5 Climate Scenario

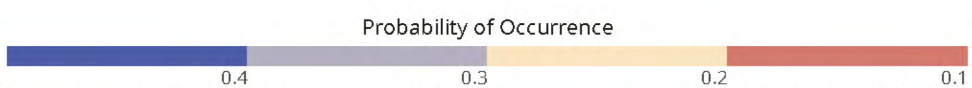


Author: Richard E.W. Berl; Map Theme: Timo Grossenbacher; Font: Open Sans; Data Sources: Maguire et al. 2018, AdaptWest (PRISM), SRTM, POLARIS, WorldClim2, TIGER/Line

Porcupine Planting Area in Pine Ridge Reservation

Random Forest Prediction of Pinus ponderosa var. scopulorum Habitat Suitability under RCP4.5 Climate Scenario

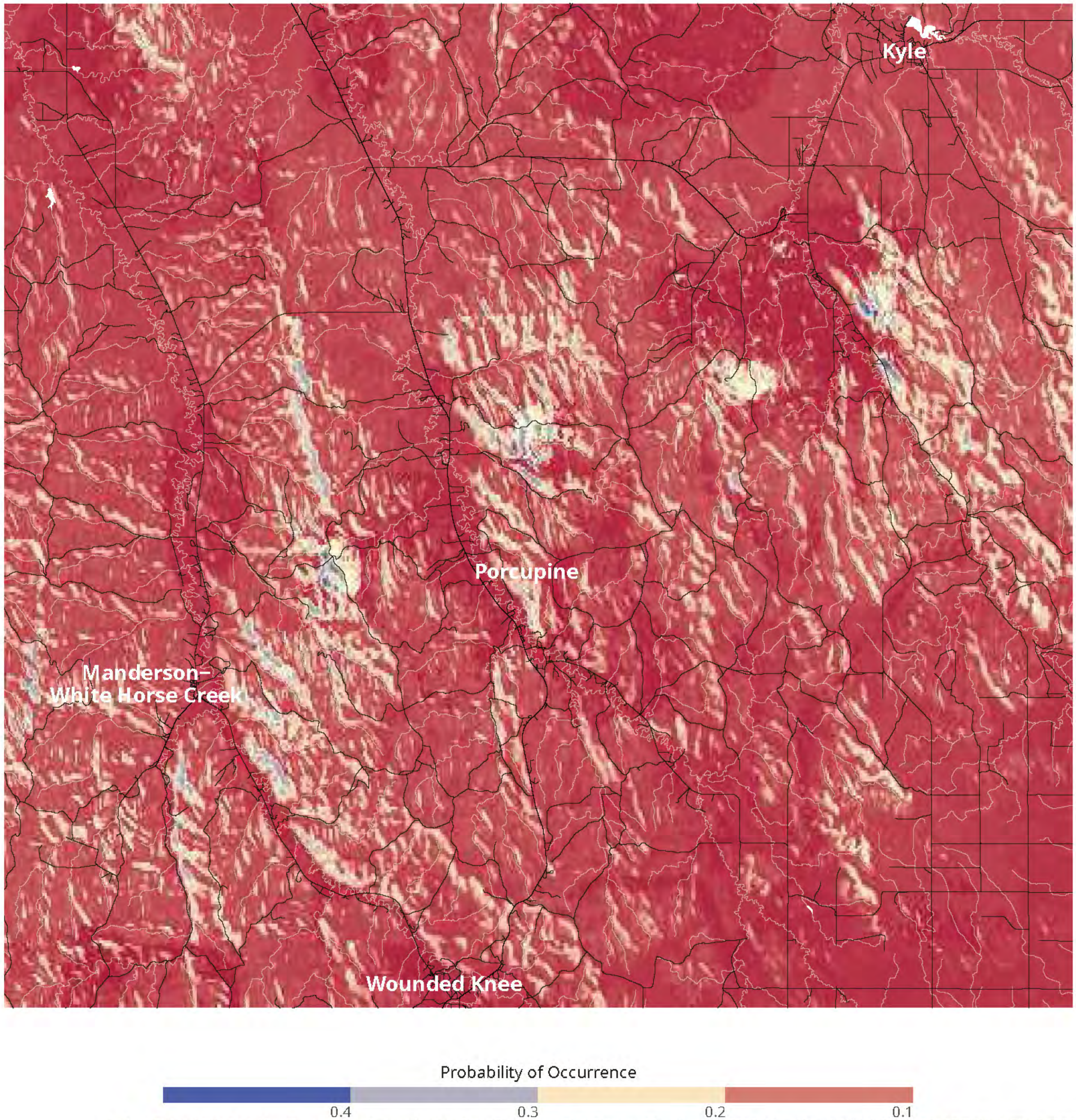




Author: Richard E.W. Berl; Map Theme: Timo Grossenbacher; Font: Open Sans; Data Sources: Maguire et al. 2018, AdaptWest (PRISM), SRTM, POLARIS, WorldClim2, TIGER/Line

Porcupine Planting Area in Pine Ridge Reservation

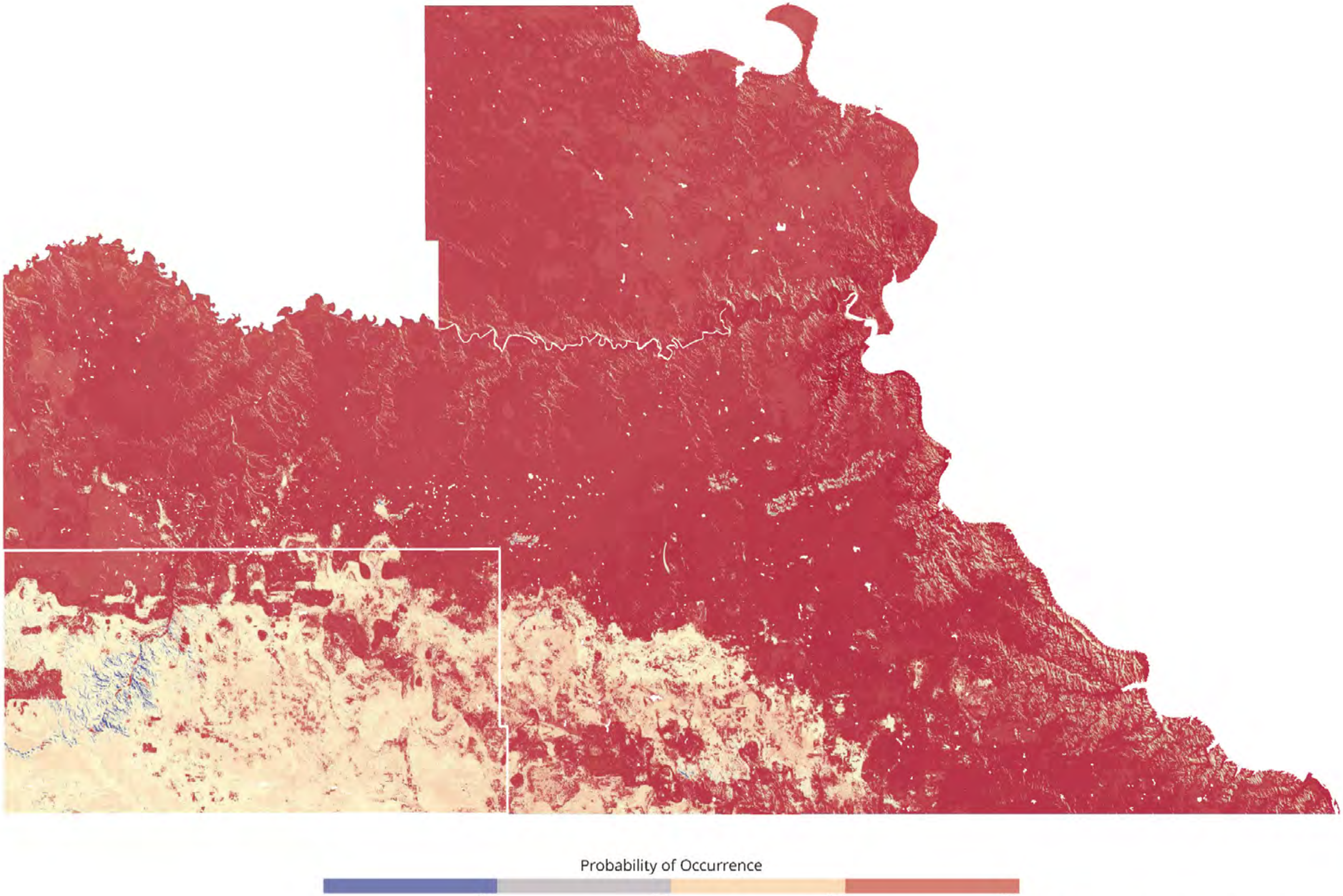
Random Forest Prediction of Pinus ponderosa var. scopulorum Habitat Suitability under RCP4.5 Climate Scenario



Author: Richard E.W. Berl; Map Theme: Timo Grossenbacher; Font: Open Sans; Data Sources: Maguire et al. 2018, AdaptWest (PRISM), SRTM, POLARIS, WorldClim2, TIGER/Line

Rosebud Reservation and Trust Land

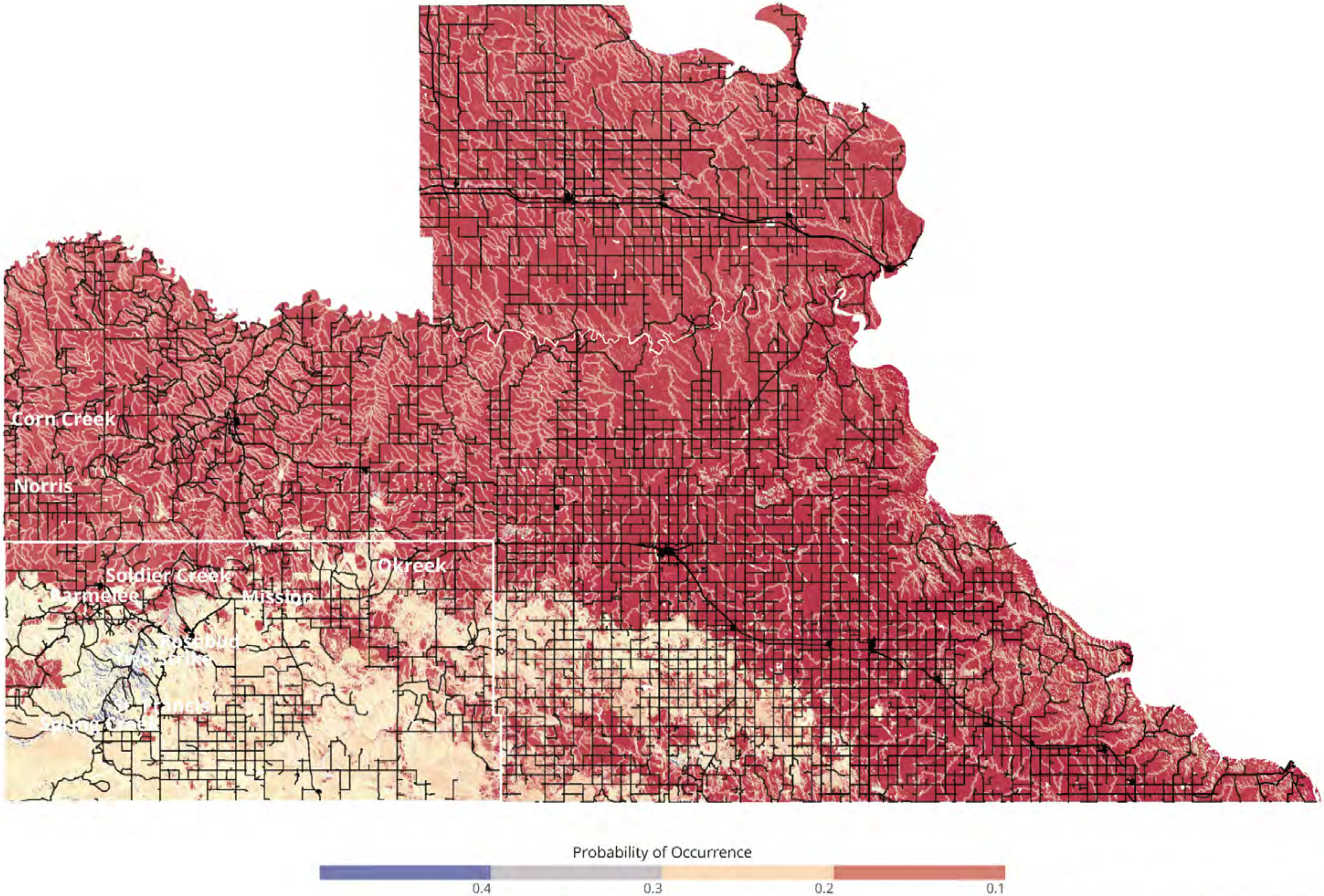
Random Forest Prediction of Pinus ponderosa var. scopulorum Habitat Suitability in 2050 under RCP4.5 Climate Scenario



Author: Richard E.W. Berl; Map Theme: Timo Grossenbacher; Font: Open Sans; Data Sources: Maguire et al. 2018, AdaptWest, SRTM, POLARIS, WorldClim2, TIGER/Line

Rosebud Reservation and Trust Land

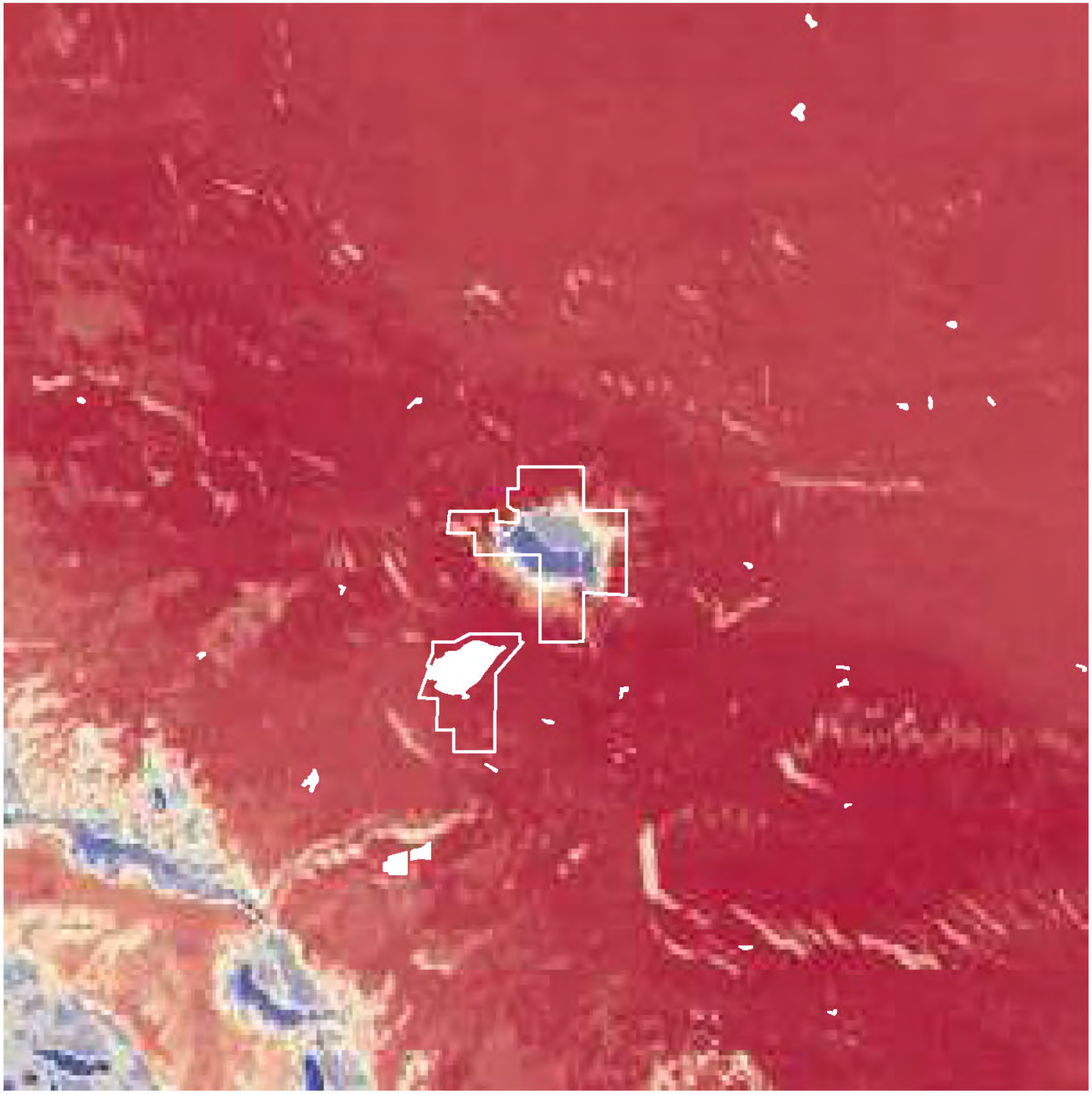
Random Forest Prediction of Pinus ponderosa var. scopulorum Habitat Suitability in 2050 under RCP4.5 Climate Scenario

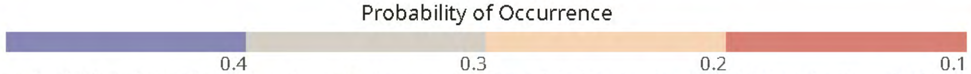


Author: Richard E.W. Berl; Map Theme: Timo Grossenbacher; Font: Open Sans; Data Sources: Maguire et al. 2018, AdaptWest, SRTM, POLARIS, WorldClim2, TIGER/Line

Bear Butte

Random Forest Prediction of Pinus ponderosa var. scopulorum Habitat Suitability in 2050 under RCP4.5 Climate Scenario

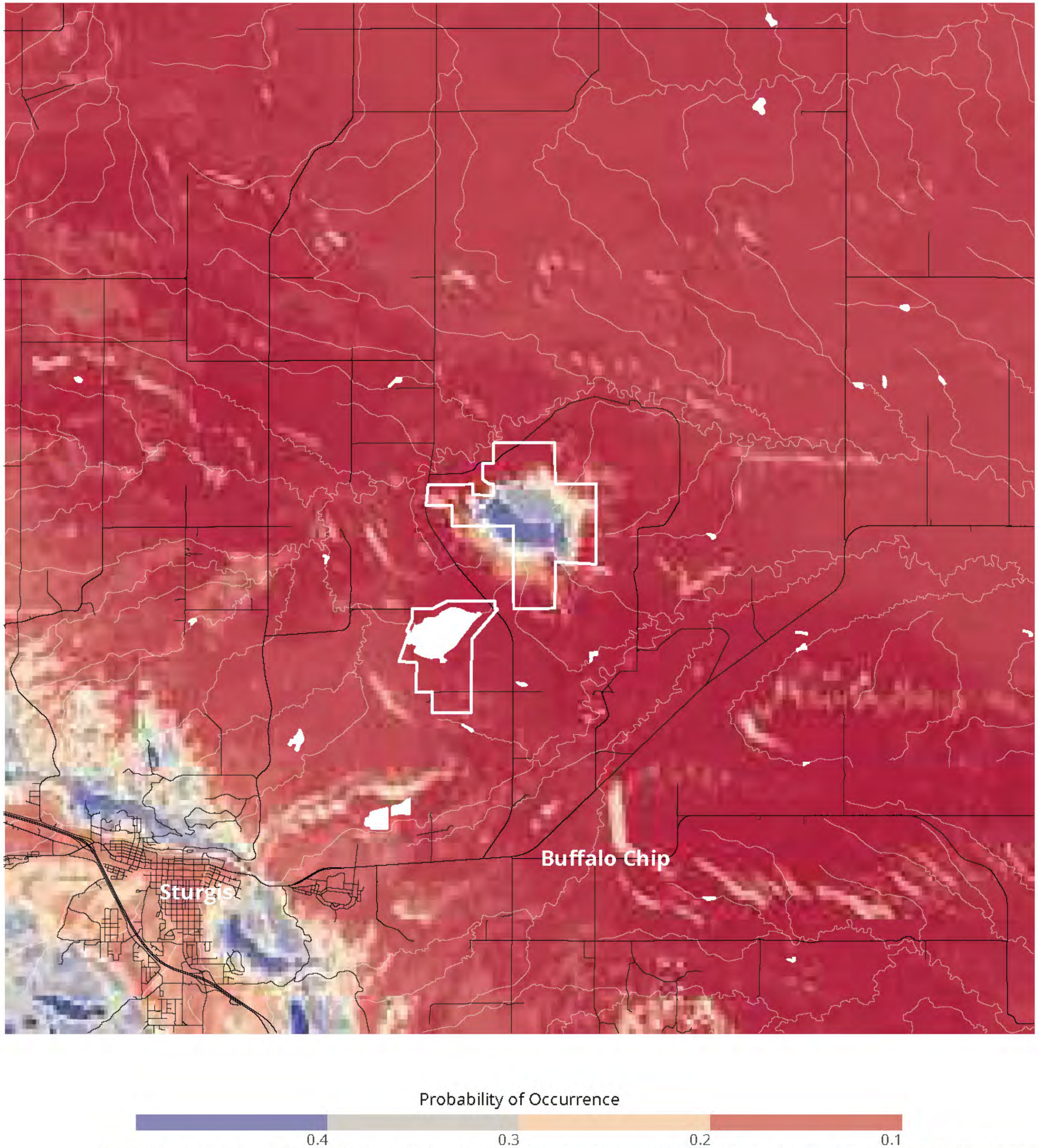




Author: Richard E.W. Berl; Map Theme: Timo Grossenbacher; Font: Open Sans; Data Sources: Maguire et al. 2018, AdaptWest, SRTM, POLARIS, WorldClim2, TIGER/Line

Bear Butte

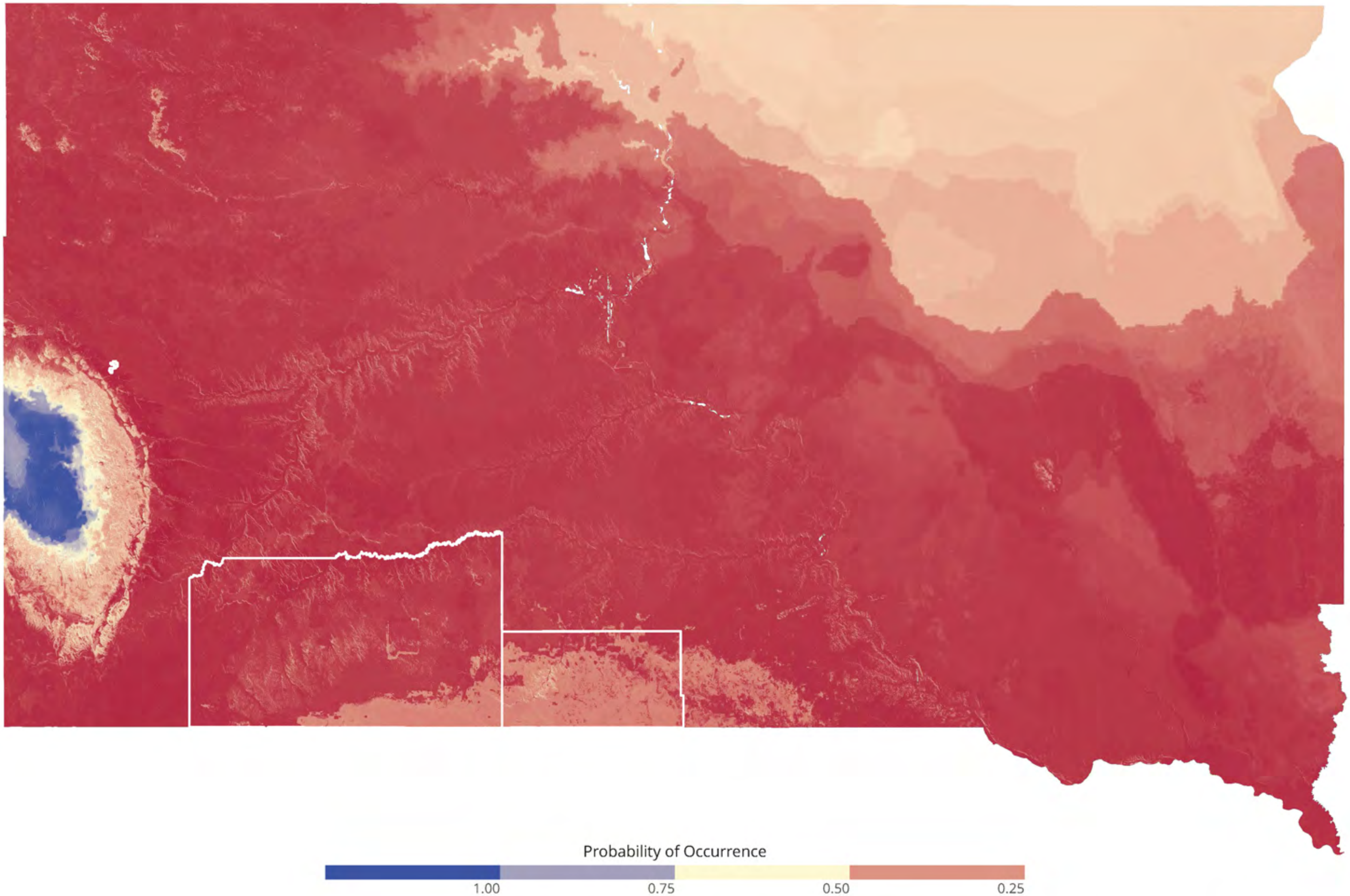
Random Forest Prediction of Pinus ponderosa var. scopulorum Habitat Suitability in 2050 under RCP4.5 Climate Scenario



Author: Richard E.W. Berl; Map Theme: Timo Grossenbacher; Font: Open Sans; Data Sources: Maguire et al. 2018, AdaptWest, SRTM, POLARIS, WorldClim2, TIGER/Line

South Dakota

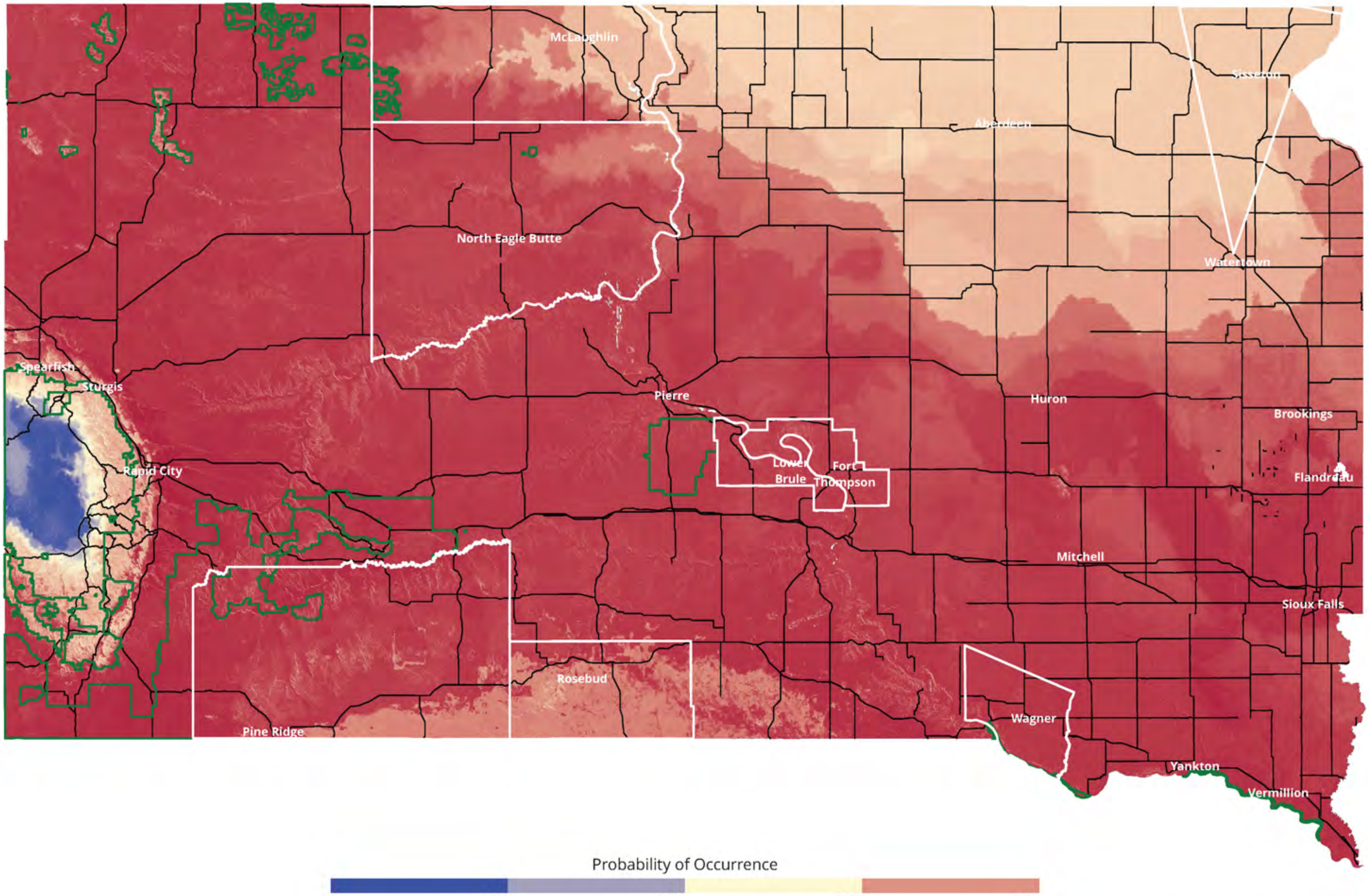
Random Forest Prediction of Pinus ponderosa var. scopulorum Habitat Suitability in 2050 under RCP8.5 Climate Scenario



Author: Richard E.W. Berl; Map Theme: Timo Grossenbacher; Font: Open Sans; Data Sources: Maguire et al. 2018, AdaptWest, SRTM, POLARIS, WorldClim2, TIGER/Line

South Dakota

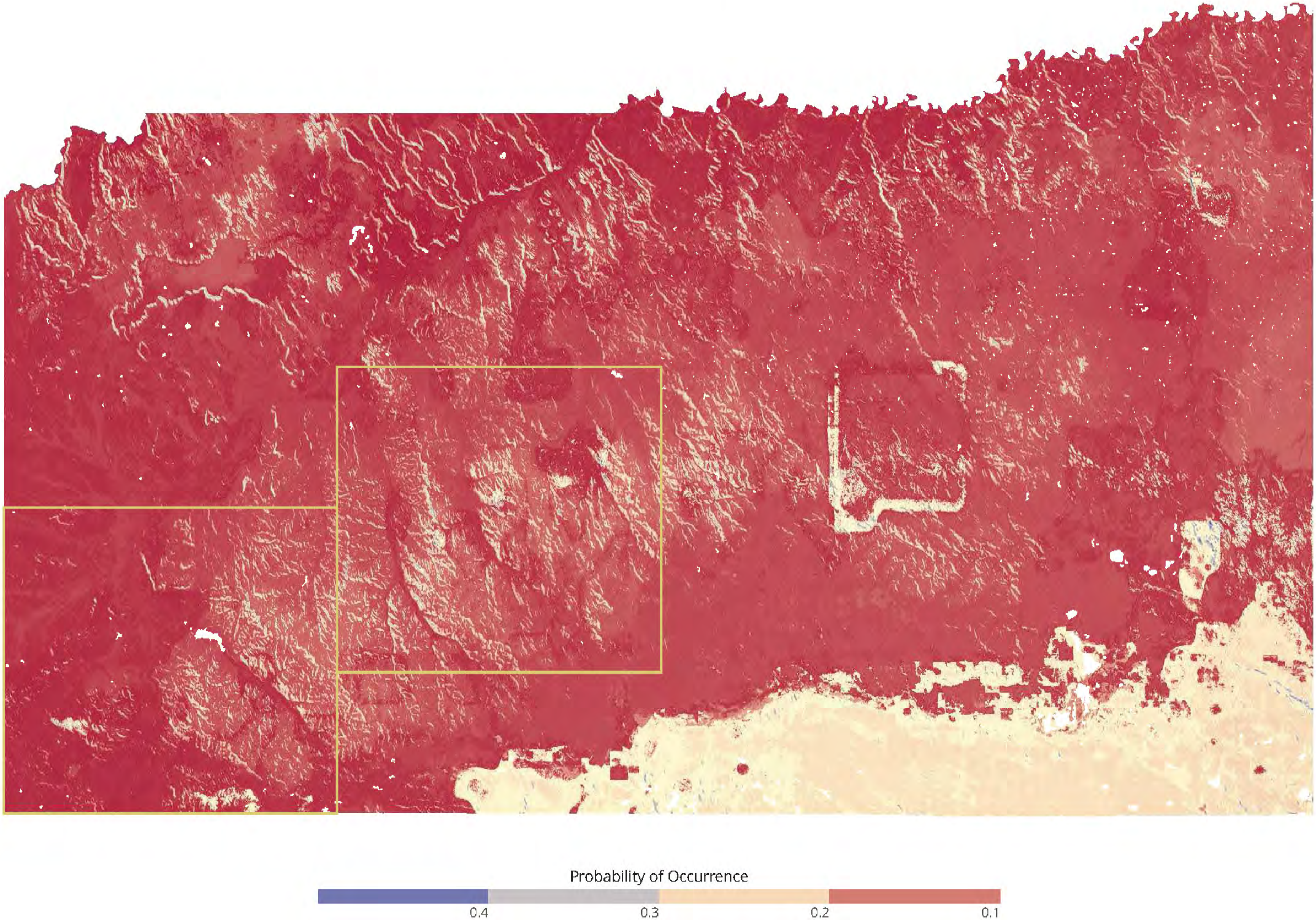
Random Forest Prediction of Pinus ponderosa var. scopulorum Habitat Suitability in 2050 under RCP8.5 Climate Scenario



Author: Richard E.W. Berl; Map Theme: Timo Grossenbacher; Font: Open Sans; Data Sources: Maguire et al. 2018, AdaptWest, SRTM, POLARIS, WorldClim2, TIGER/Line, NPS, USFS

Pine Ridge Reservation and Trust Land

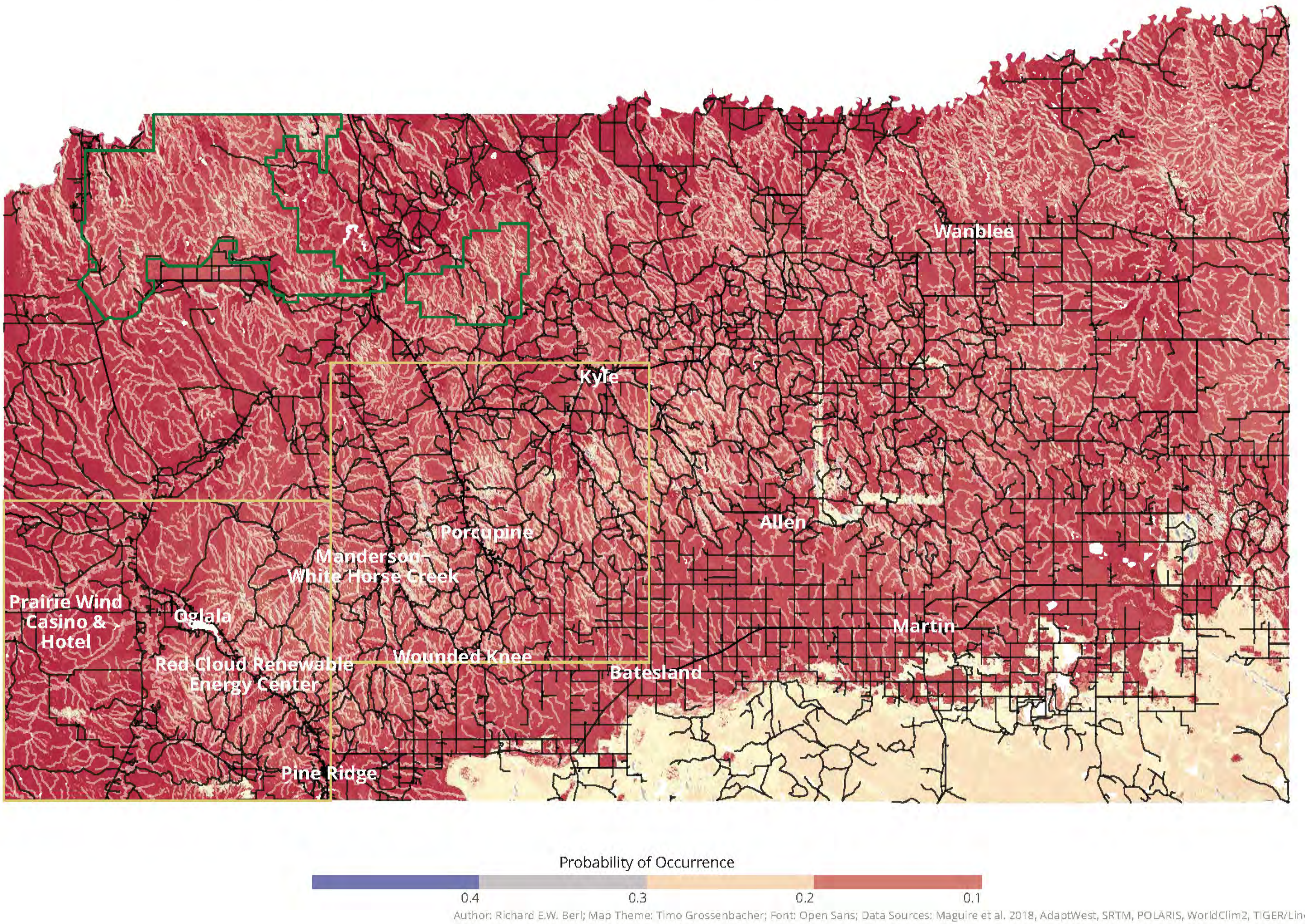
Random Forest Prediction of Pinus ponderosa var. scopulorum Habitat Suitability in 2050 under RCP8.5 Climate Scenario



Author: Richard E.W. Berl; Map Theme: Timo Grossenbacher; Font: Open Sans; Data Sources: Maguire et al. 2018, AdaptWest, SRTM, POLARIS, WorldClim2, TIGER/Line

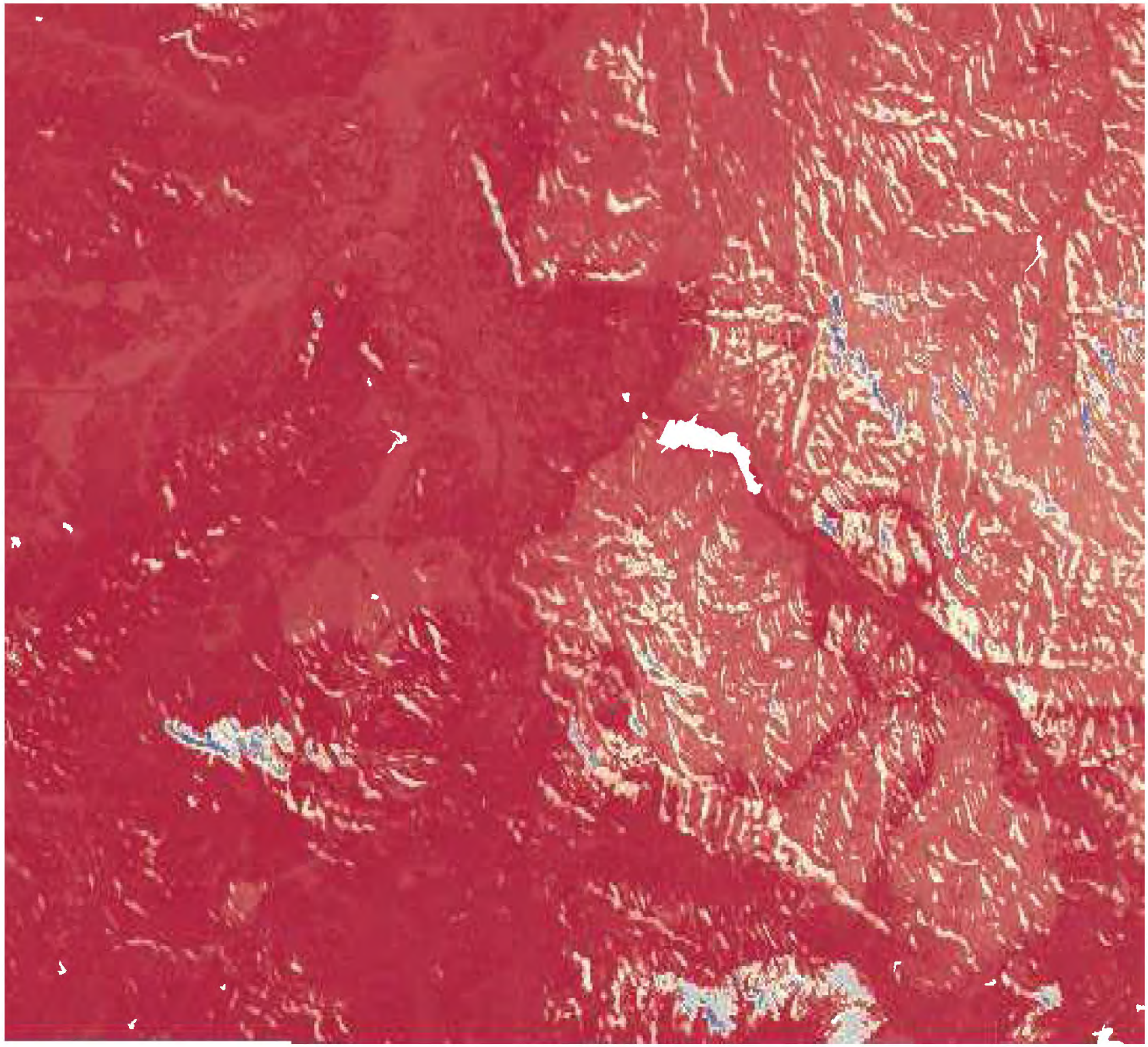
Pine Ridge Reservation and Trust Land

Random Forest Prediction of Pinus ponderosa var. scopulorum Habitat Suitability in 2050 under RCP8.5 Climate Scenario



Oglala Planting Area in Pine Ridge Reservation

Random Forest Prediction of Pinus ponderosa var. scopulorum Habitat Suitability under RCP8.5 Climate Scenario



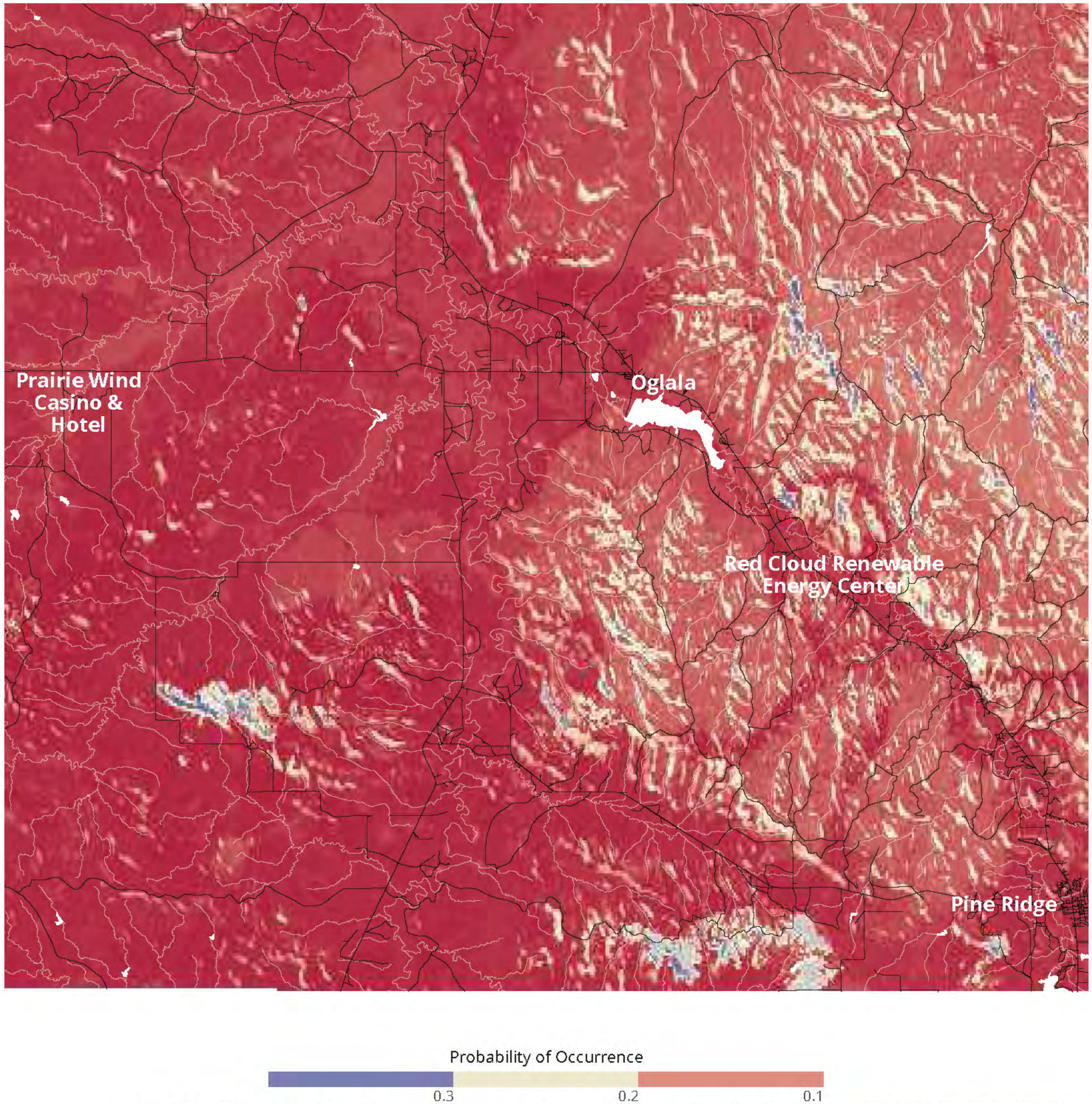
Probability of Occurrence

0.3 0.2 0.1

Author: Richard E.W. Berl; Map Theme: Timo Grossenbacher; Font: Open Sans; Data Sources: Maguire et al. 2018, AdaptWest (PRISM), SRTM, POLARIS, WorldClim2, TIGER/Line

Oglala Planting Area in Pine Ridge Reservation

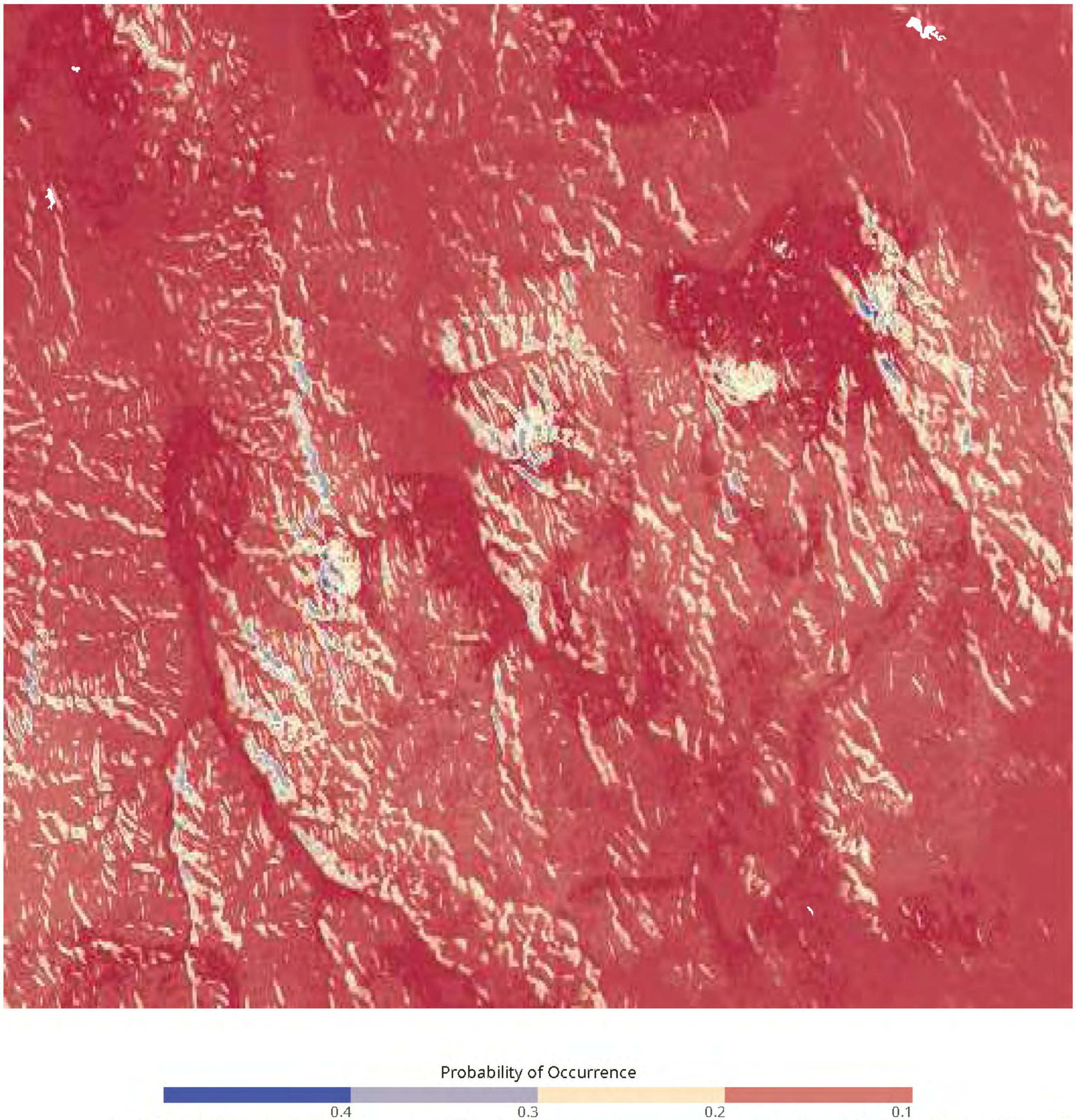
Random Forest Prediction of Pinus ponderosa var. scopulorum Habitat Suitability under RCP8.5 Climate Scenario



Author: Richard E.W. Berl; Map Theme: Timo Grossenbacher; Font: Open Sans; Data Sources: Maguire et al. 2018, AdaptWest (PRISM), SRTM, POLARIS, WorldClim2, TIGER/Line

Porcupine Planting Area in Pine Ridge Reservation

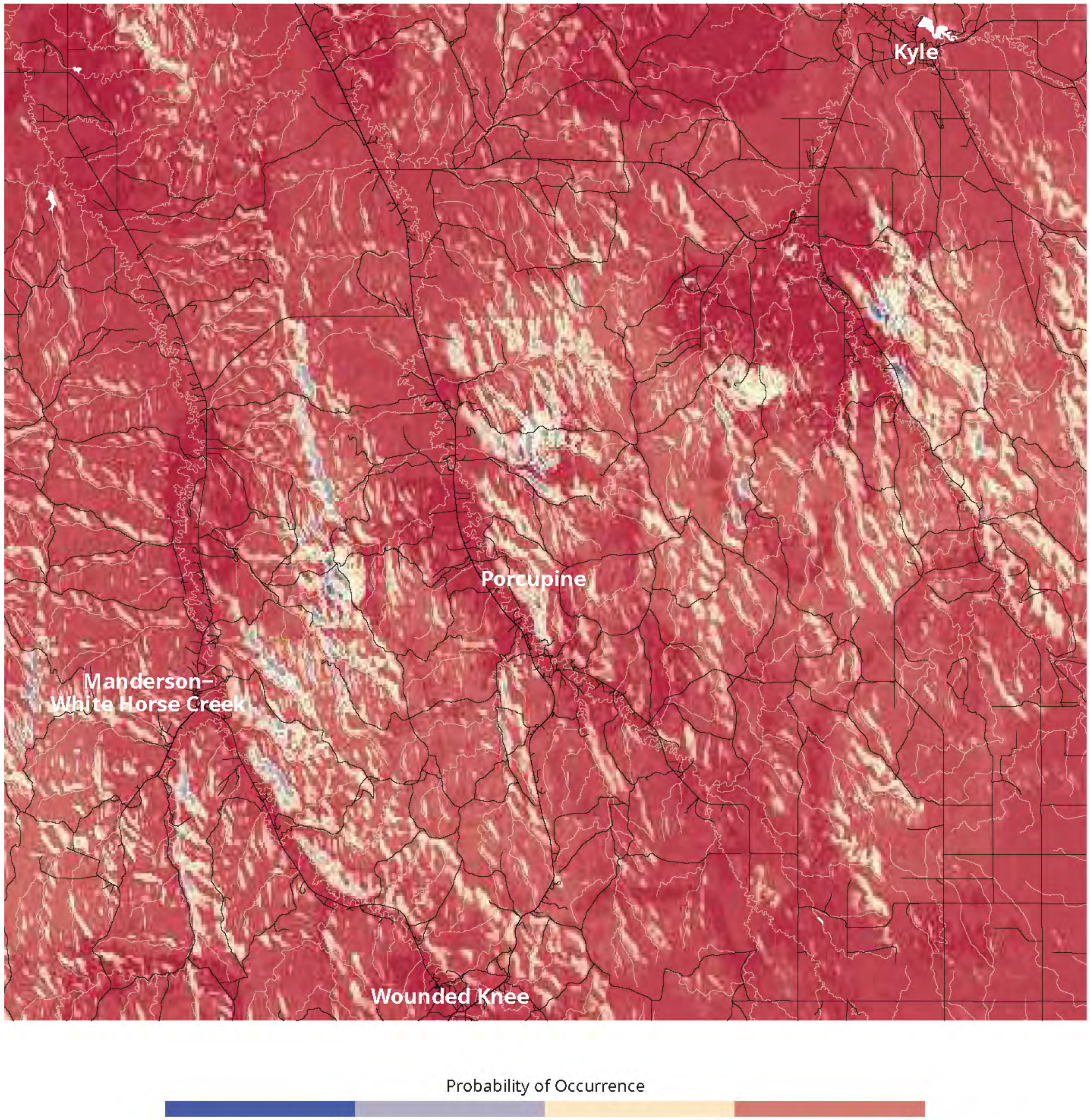
Random Forest Prediction of Pinus ponderosa var. scopulorum Habitat Suitability under RCP8.5 Climate Scenario



Author: Richard E.W. Berl; Map Theme: Timo Grossenbacher; Font: Open Sans; Data Sources: Maguire et al. 2018, AdaptWest (PRISM), SRTM, POLARIS, WorldClim2, TIGER/Line

Porcupine Planting Area in Pine Ridge Reservation

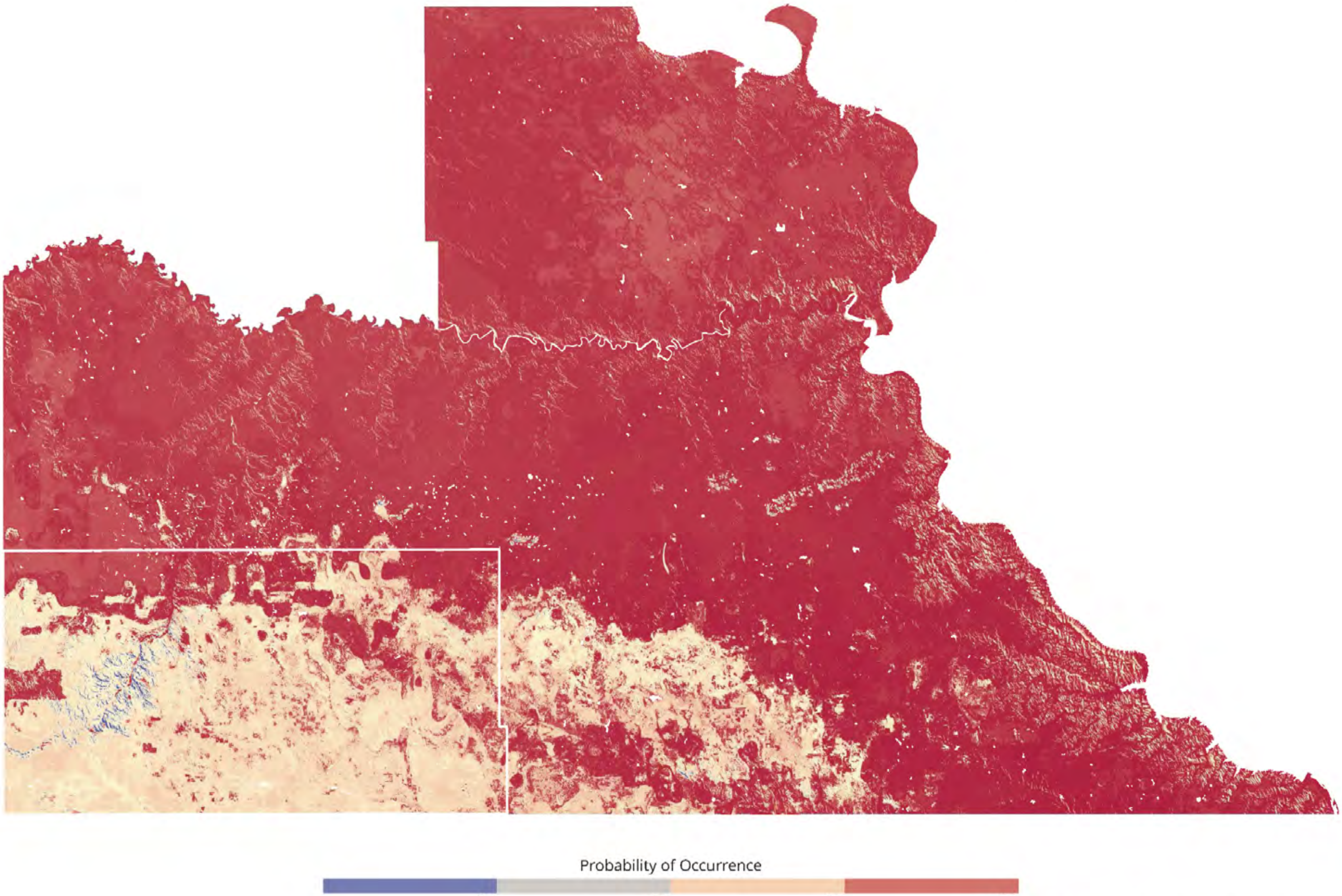
Random Forest Prediction of Pinus ponderosa var. scopulorum Habitat Suitability under RCP8.5 Climate Scenario



Author: Richard E.W. Berl; Map Theme: Timo Grossenbacher; Font: Open Sans; Data Sources: Maguire et al. 2018, AdaptWest (PRISM), SRTM, POLARIS, WorldClim2, TIGER/Line

Rosebud Reservation and Trust Land

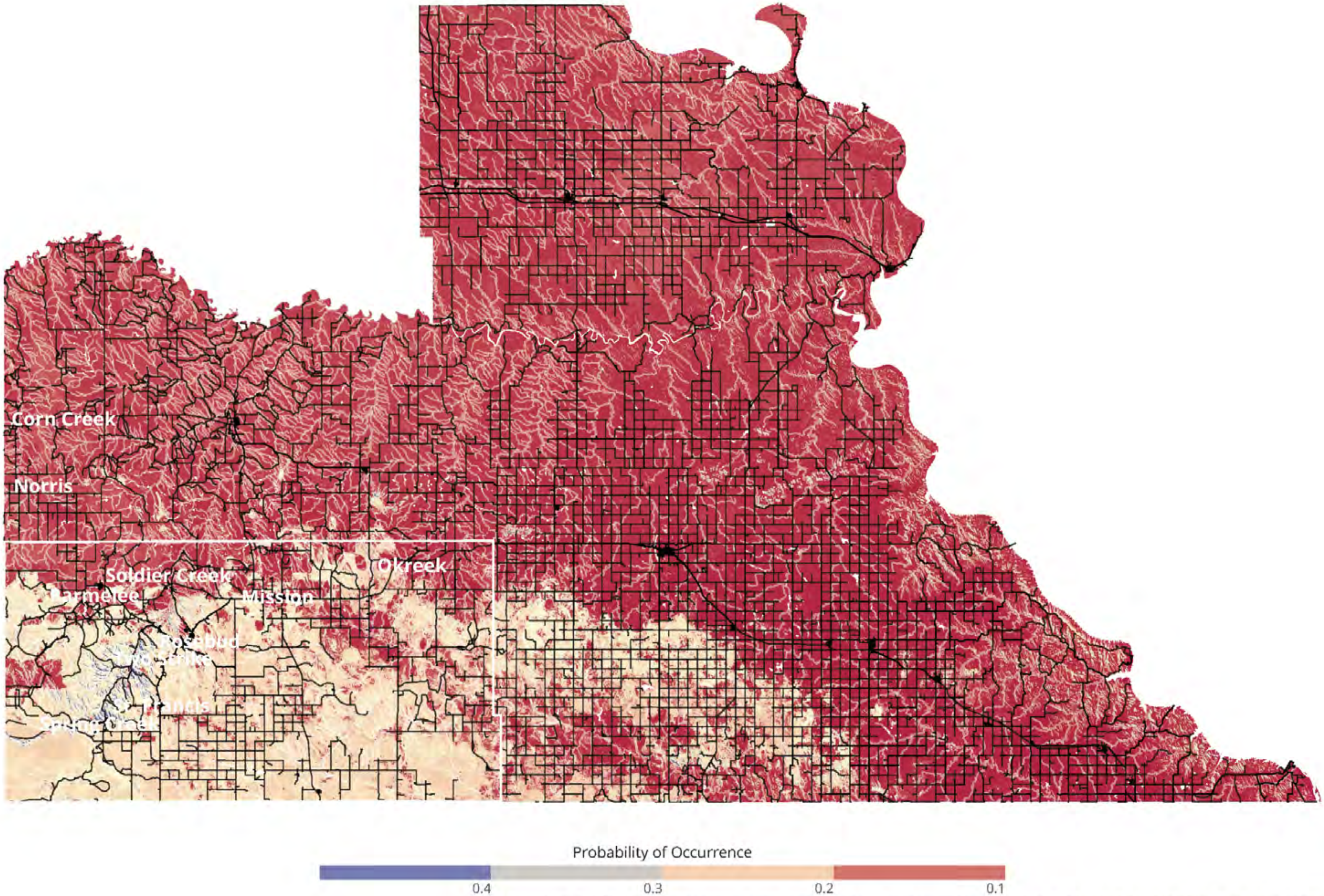
Random Forest Prediction of Pinus ponderosa var. scopulorum Habitat Suitability in 2050 under RCP8.5 Climate Scenario



Author: Richard E.W. Berl; Map Theme: Timo Grossenbacher; Font: Open Sans; Data Sources: Maguire et al. 2018, AdaptWest, SRTM, POLARIS, WorldClim2, TIGER/Line

Rosebud Reservation and Trust Land

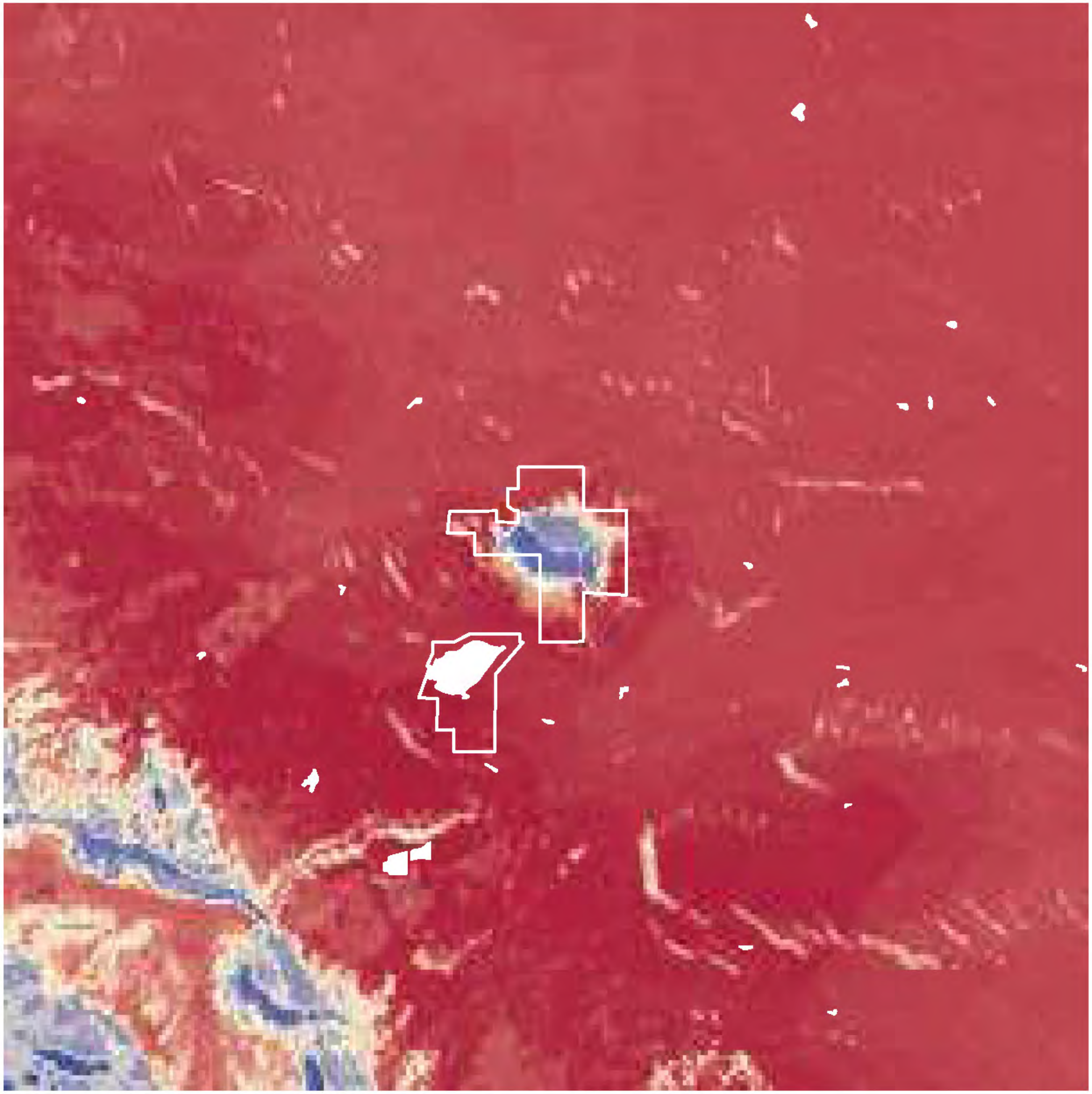
Random Forest Prediction of Pinus ponderosa var. scopulorum Habitat Suitability in 2050 under RCP8.5 Climate Scenario



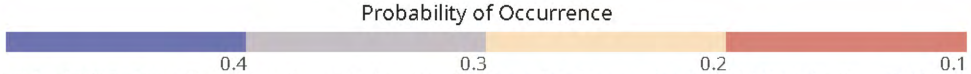
Author: Richard E.W. Berl; Map Theme: Timo Grossenbacher; Font: Open Sans; Data Sources: Maguire et al. 2018, AdaptWest, SRTM, POLARIS, WorldClim2, TIGER/Line

Bear Butte

Random Forest Prediction of Pinus ponderosa var. scopulorum Habitat Suitability in 2050 under RCP8.5 Climate Scenario



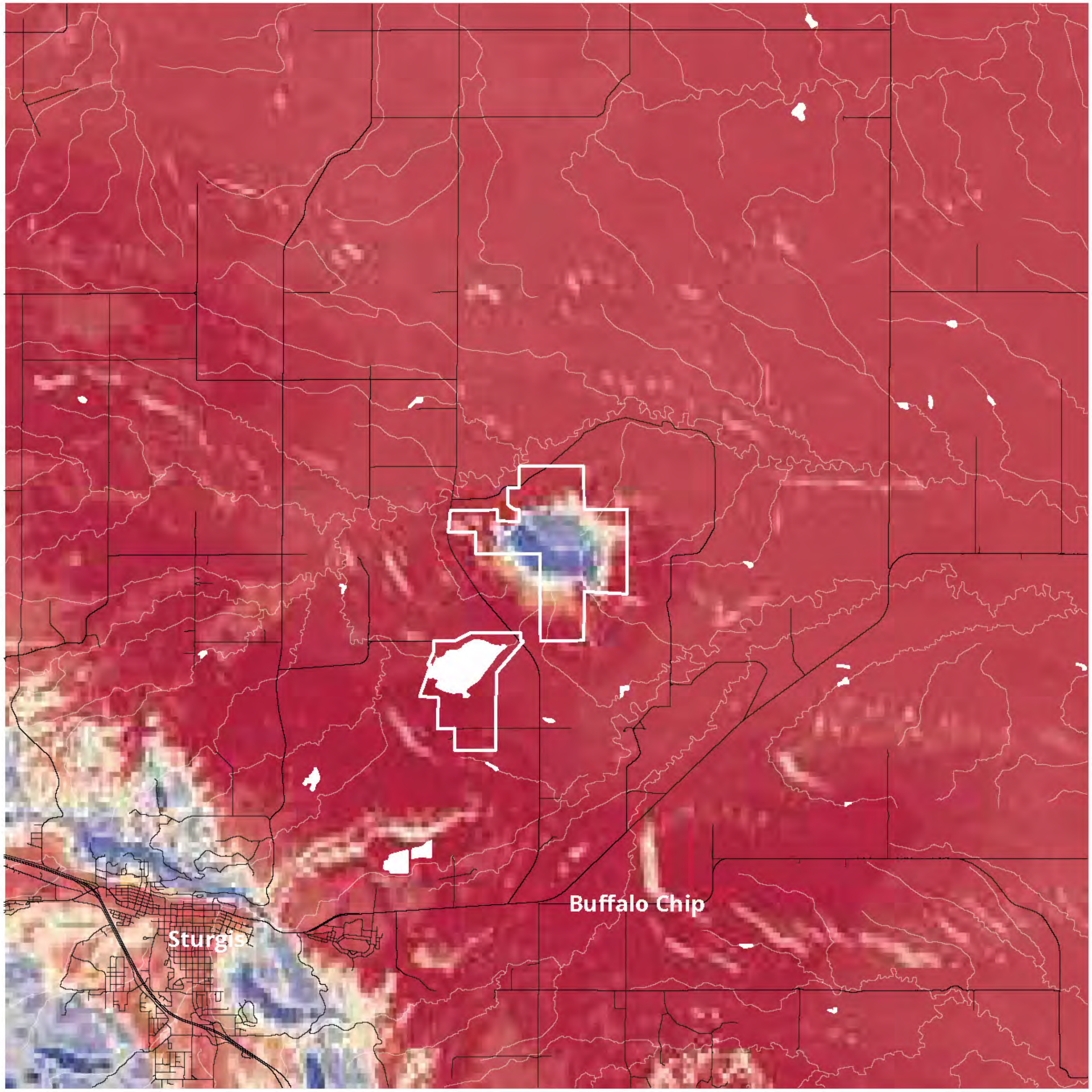
Probability of Occurrence



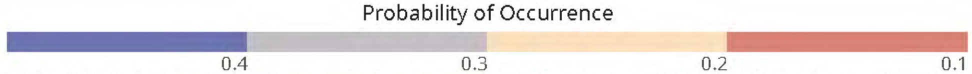
Author: Richard E.W. Berl; Map Theme: Timo Grossenbacher; Font: Open Sans; Data Sources: Maguire et al. 2018, AdaptWest, SRTM, POLARIS, WorldClim2, TIGER/Line

Bear Butte

Random Forest Prediction of Pinus ponderosa var. scopulorum Habitat Suitability in 2050 under RCP8.5 Climate Scenario



Probability of Occurrence



Author: Richard E.W. Berl; Map Theme: Timo Grossenbacher; Font: Open Sans; Data Sources: Maguire et al. 2018, AdaptWest, SRTM, POLARIS, WorldClim2, TIGER/Line

IMPROVE: Recommendations

Tools

- **Experimental Design**: Alters the inputs to the process in a structured way to study the effects of changing multiple inputs simultaneously.

Summary

The recommendations for improvement in this project are drawn primarily from the results of the tools we have used. Discussion and research have suggested that the use of mycorrhizal inoculant and auxin-fortified hydrophilic gel during planting can enhance seedling survival by roughly 18% and 15%, respectively (Shepperd & Battaglia 2002), and these arose in our discussions. We anticipate that a greater amount of time devoted to training in planting methods should also yield significant improvements in survival. Specifically, previous defects were identified in methods for scalping, which is used to remove other vegetation and provide a flat surface for planting and water retention. The publication by Sheppard & Battaglia (2002) can serve as a general reference for additional best practices during planting of seedlings (pp. 89-90), including taking care to avoid desiccation. Damage after planting was identified as an issue due to the difficulty in distinguishing planting areas, and can be addressed by the use of brightly-colored flagging.

In discussion, none of these proposed improvements appear to be cost-prohibitive (and inoculation is already being done in the nursery, prior to our stage of the process), so the recommendation at this time is to implement all of them. Additionally, they all contribute toward the primary goal of encouraging seedling survival, as we did not find any viable options for reducing the cost of volunteer labor, only for increasing its efficiency and value. If the situation changed and there was no longer time or funds to implement all of our proposed solutions, the proper tool from Lean Six Sigma would be a **Solution Prioritization Matrix**, which would assist in best meeting project goals with the resources available.

To aid in future planting efforts and to measure more precisely the relative effects of polymer use, scalping quality, and seedling age and size on survival in the field, we have drafted an experimental design to use with the additional seedlings that remain available after planting. The results of this experiment will provide valuable information and allow a better quantification of the costs and benefits associated with these factors.

Finally, through the **Habitat Suitability Modeling** conducted in the Analyze section of the report, we identified the environmental factors that we expect to contribute most to present and future survival of Rocky Mountain ponderosa pine. Using our model, I plotted the projected suitability across the planting areas at a resolution of 90 meters. With these resulting maps, we now have the best available recommendations for maximizing the relative survival within our planting areas.

Experimental Design

Problem Definition

We suspect that the use of auxin-fortified hydrophilic polymer, the quality of scalping of the planting site, and the age of the seedling may have effects on seedling survival. This experiment is designed to quantify the relative effects of these variables to enable future assessment of the cost-to-benefit ratio of polymer use and time spent on training.

Response Variables

Survival*: Binary (yes/no)

Height*: Continuous (cm)

Stem diameter*: Continuous (mm)

**See Data Collection Plan for notes on variables and measurement.*

Input Variables

Polymer use: Binary (yes/no)

Seedling age: Discrete (1 year, 2 years)

Scalping quality: Ordinal (none, poor, good)

Experimental Strategy

The size of the experiment will be contingent upon the exact number of seedlings available (ensuring an equal number of 1-year-old and 2-year-old seedlings), the maximum size of the experimental plot, and the effort required. Sheppard & Battaglia (2012) advise a minimum spacing of 12 feet by 12 feet (3.7 meters by 3.7 meters) for commercial planting, which should be adequate for this experimental context as well and requires a $(12^2)n$ feet or $(3.7^2)n$ m area for planting, where n is the number of seedlings to be planted.

Using a randomized mixed-level full factorial design (below) that allows for full testing of main effects and interactions between all three factors, 12 seedlings are required per replication, 6 of each age level. This means that for a single replication, a 1,728 ft^2

(164.28 m^2) area would be required for experimental plots. For each additional replication, 12 additional seedlings and plots are required. Plantings should follow the order of rows in the design (below) and should be re-randomized for each replication.

As an example, if 500 seedlings are available, then up to 41 replications can be conducted, using 492 seedlings and an area of 70,848 ft^2 (6,735.48 m^2). If 1,000 seedlings are available, up to 83 replications can be conducted, using 996 seedlings and an area of 143,424 ft^2 (13,635.24 m^2). This scenario would likely be beyond the limits of time, effort, and space available for experimental setup.

If a fractional factorial design was desired, in the interest of reducing the number of plantings required at the expense of a slight decrease in explanatory power, then the scalping quality variable would need to be reduced to two levels (none/good or poor/good) for ease of determining a feasible design. This requires 8 plantings for a full factorial design, or 4 plantings for a 2^{3-1} fractional factorial design in which three-way interactions may not be estimable (which is acceptable under most conditions). Either of these designs would allow more replications for the same number of seedlings used and a more efficient use of time, effort, and space. Therefore, if it is acceptable to eliminate the ability to distinguish between three levels of scalping quality, then the fractional factorial design is preferred. As in the description of the full factorial design, plantings should be situated in 12-foot by 12-foot plots and the order of rows within each replication randomized.

Full Factorial Design (Mixed-Level)

	polymer	age	scalping
1	0	1	2
2	0	2	2
3	0	1	0
4	0	2	1
5	1	2	2
6	1	1	1
7	1	1	0
8	1	1	2
9	1	2	1
10	1	2	0
11	0	2	0
12	0	1	1

Full Factorial Design (2-Level)

	polymer	age	scalping
1	1	1	1
2	0	2	1
3	0	1	0
4	1	2	0
5	0	1	1
6	0	2	0
7	1	2	1
8	1	1	0

Fractional Factorial Design (2^{3-1})

	polymer	age	scalping
1	0	2	0
2	1	2	1
3	0	1	1
4	1	1	0

Experimental Considerations

In conducting the experiment, it will be necessary to ensure that all other potential variables aside from the three input variables being manipulated are either controlled for or sufficiently randomized. For example, polymer amount will need to be consistent and aspects of the experimental plots will need to be as similar as possible in terms of hill slope and aspect and the amount and frequency of watering (if any). As a field experiment, a number of factors (e.g. weather, animal browsing, disease) will be beyond the control of the experiment and may create noise in the data—necessitating as many replications as possible—but will aid in the ecological validity of the experiment and the applicability of results to the reforestation effort.

Response data would need to be collected on a similar timescale as the post-planting assessment for the **Data Collection Plan**. Subsequent analysis of results can use multinomial logistic regression or other related tests to assess the effects of each input variable.

CONTROL: Monitoring for Success

Summary

Though there was not sufficient time to undergo full consultation and development of process controls before the conclusion of this project, I advise that the project team take steps following this year’s planting season to ensure that the changes implemented are maintained over successive seasons. This could be done in a formal way by developing a **Control Plan**, which would solidify the responsibilities for monitoring changes and maintaining improvements. This would include reviewing the **Process Flow Map** and other materials included in this report and revising them as needed on at least an annual basis.

Once sufficient data are collected using the proposed **Data Collection Plan**, a **Process Capability Analysis** can be conducted to quantify the mean and variation in survival over the course of the process. Following this, the Lean Six Sigma tool of **Statistical Process Control** could be used to stabilize the mean and reduce variation to achieve the goal of six standard deviations within process limits.

The overall goal of the Control phase is to monitor and maintain the changes recommended by this report and to continue the commitment to process improvement into the future. As the objective of the project is to increase seedling survival over the short and long terms, monitoring and maintaining the recommended improvements year after year will be vital to this effort.

Project Summary and Conclusions

The Lean Six Sigma project detailed in this report began with a problem: reforestation efforts with Rocky Mountain ponderosa pine in Pine Ridge Reservation were producing survival rates below that needed to sustainably continue the process. Further, this rate of failure had other potential negative impacts on external perceptions and reputation, relationships with project partners, community livelihoods and culture, and ecological health. The survival rate of last year's seedlings was estimated at 47% and, through the steps taken and changes advised toward improving the quality of the process, we intend to raise the rate of survival to 75%, an improvement of 60%.

If the increase in survival predicted from the use of auxin-fortified hydrophilic polymer is consistent with previous studies (Shepperd & Battaglia 2002), then we should expect to see an improvement of roughly 15% from that change alone, which leaves an additional 13% gain to be attained by other means.

We believe we can accomplish the remainder of this goal through: a) targeted planting based on current and projected future climate conditions; b) additional training to improve planting skill and quality, particularly in scalping methods and including a focus on the cultural and ecological significance of the project; c) and placement of brightly-colored flagging around planting sites to prevent incidental damage after planting. These three changes should attain the stated goal and, importantly, provide a foundation for continued improvement into the future.

The continued success of this project will require adherence to the **Data Collection Plan** so that data are available for more targeted future analyses on the factors driving the success or failure of individual seedlings, and on the execution of the structured field experiment detailed in the **Experimental Design,** to estimate more precisely the effects of the factors under our control. Following these, a focus on control and maintenance of the gains achieved will be critical to the sustainability of the project and maximizing its benefits.

Acknowledgements

First and foremost, I would like to thank my project partners, especially Sebastian Africano and Eriq Acosta of Trees, Water & People for their willingness to include me on this project and for their time, trust, participation, and encouragement over the course of its development.

This report was generated using materials, guidelines, and templates shared by Dr. Richard D. Schleusener of iSquared, Inc., and supplementary materials from the website of Dr. Pat Hammett of the University of Michigan. I am grateful to Dr. Schleusener for his detailed, practical instruction and thoughtful guidance in Lean Six Sigma methods and philosophy over the course of this project.

Cover image is courtesy of Trees, Water & People and PrintReleaf.

Fonts used in this document are Merriweather by Eben Sorkin (Headings), Open Sans by Steve Matteson (Body), and Inconsolata by Raph Levien (Code). All are open source and freely available from Google Fonts.

Color scales for habitat suitability maps are from a technical note by Paul Tol of SRON and are optimized for color-blind viewers.

Any opinions expressed in this report are my own and do not necessarily represent those of Colorado State University, Trees, Water & People, or any other entity.

References

AdaptWest Project. (2015). Gridded current and projected climate data for North America at 1km resolution, interpolated using the ClimateNA v5.10 software (T. Wang et al., 2015). Available at https://adaptwest.databasin.org.

Chaney, N. W., Wood, E. F., McBratney, A. B., Hempel, J. W., Nauman, T. W., Brungard, C. W., & Odgers, N. P. (2016). POLARIS: A 30-meter probabilistic soil series map of the contiguous United States. *Geoderma*, *274*, 54-67.

Elith, J., & Leathwick, J. R. (2009). Species distribution models: Ecological explanation and prediction across space and time. *Annual Review of Ecology, Evolution, and Systematics*, *40*, 677-697.

Fick, S.E., & R.J. Hijmans. (2017). WorldClim 2: New 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*, *37*(12), 4302-4315.

Haase, D. L. (2008). Understanding forest seedling quality: Measurements and interpretation. *Tree Planters' Notes*, *52*(2), 24-30.

Hawbaker, T.J., Vanderhoof, M.K., Beal, Y.G., Takacs, J.D., Schmidt, G.L., Falgout, J.T., Williams, B., Brunner, N.M., Caldwell, M.K., Picotte, J.J., Howard, S.M., Stitt, S., & Dwyer, J.L. (2017). Landsat Burned Area Essential Climate Variable products for the conterminous United States (1984 – 2015; ver. 1.1, Sep. 2017): U.S. Geological Survey data release. Available at https://doi.org/10.5066/F73B5X76.

Jarvis, A., Reuter, H.I., Nelson, A., & Guevara, E. (2008). Hole-filled SRTM for the globe Version 4. Available from the CGIAR-CSI SRTM 90m Database at http://srtm.csi.cgiar.org.

Maguire, K. C., Shinneman, D. J., Potter, K. M., & Hipkins, V. D. (2018). Intraspecific niche models for ponderosa pine (*Pinus ponderosa*) suggest potential variability in population-level response to climate change. *Systematic Biology*, *67*.

Menes, P. A., & Mohammed, G. H. (1995). Identifying the root collar on forest tree seedlings. *The Forestry Chronicle*, *71*(3), 304-310.

Potter, K. M., Hipkins, V. D., Mahalovich, M. F., & Means, R. E. (2013). Mitochondrial DNA haplotype distribution patterns in *Pinus ponderosa* (Pinaceae): Range-wide evolutionary history and implications for conservation. *American Journal of Botany*, *100*(8), 1562-1579.

R Core Team. (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available at https://www.R-project.org.

Riahi, K., Rao, S., Krey, V., Cho, C., Chirkov, V., Fischer, G., Kindermann, G., Nakicenovic, N., & Rafaj, P. (2011). RCP 8.5—A scenario of comparatively high greenhouse gas emissions. *Climatic Change*, *109*(1-2), 33.

Sanford, T., Frumhoff, P. C., Luers, A., & Gulledge, J. (2014). The climate policy narrative for a dangerously warming world. *Nature Climate Change*, *4*(3), 164.

Shepperd, W. D., & Battaglia, M. A. (2002). *Ecology, silviculture, and management of Black Hills ponderosa pine*. Fort Collins, CO: US Department of Agriculture, Forest Service, Rocky Mountain Research Station.

Thomson, A. M., Calvin, K. V., Smith, S. J., Kyle, G. P., Volke, A., Patel, P., Delgado-Arias, P., Bond-Lamberty, B., Wise, M. A., Clarke, L. E., & Edmonds, J. A. (2011). RCP4. 5: a pathway for stabilization of radiative forcing by 2100. *Climatic Change*, *109*(1-2), 77.

Wang, T., Hamann, A., Spittlehouse, D., & Carroll, C. (2016). Locally downscaled and spatially customizable climate data for historical and future periods for North America. *PLOS ONE*, *11*(6): e0156720.

Metadata

AdaptWest

(30 arc-second ~ 1km, resampled to 3 arc-second using bilinear interpolation)

MAT: mean annual temperature (°C)

MWMT: mean temperature of the warmest month (°C)

MCMT: mean temperature of the coldest month (°C)

TD: difference between MCMT and MWMT, as a measure of continentality (°C)

MAP: mean annual precipitation (mm)

MSP: mean summer (May to Sep) precipitation (mm)

AHM: annual heat moisture index, calculated as (MAT+10)/(MAP/1000)

SHM: summer heat moisture index, calculated as MWMT/(MSP/1000)

DD_0: degree-days below 0°C (chilling degree days)

DD5: degree-days above 5°C (growing degree days)

DD_18: degree-days below 18°C (days)

DD18: degree-days above 18°C (days)

NFFD: the number of frost-free days (days)

bFFP: the julian date on which the frost-free period begins

eFFP: the julian date on which the frost-free period ends

FFP: frost-free period

PAS: precipitation as snow (mm)

EMT: extreme minimum temperature over 30 years

EXT: extreme maximum temperature over 30 years

Eref: Hargreave's reference evaporation

CMD: Hargreave's climatic moisture index

MAR: mean annual solar radiation (MJ m-2 d-1) (excludes areas south of US)

RH: mean annual relative humidity (%)

Tave_wt: winter (Dec to Feb) mean temperature (°C)

Tave_sm: summer (Jun to Aug) mean temperature (°C)

PPT_wt: winter (Dec to Feb) precipitation (mm)

PPT_sm: summer (Jun to Aug) precipitation (mm)

SRTM

(3 arc-second ~ 90m)

elev: void-filled digital elevation (m) [Note: Only used to generate Heat Load Index, below]

slope: grade of elevation change (radians) [Note: Only used to generate Heat Load Index, below]

aspect.cos: cosine of the direction of slope face, where 0 radians represents north (N/A) [Note: Only used to generate Heat Load Index, below]

aspect.sin: sine of the direction of slope face, where 0 radians represents north (N/A) [Note: Only used to generate Heat Load Index, below]

hli: McCune & Keon (2002) Heat Load Index, ranging from 0 (low) to 1 (high)

POLARIS

(3 arc-second ~ 90m)

awc_mean_0_5: mean available water content in 0-5 cm depth from surface (m3/m3)

bd_mean_0_5: mean bulk density in 0-5 cm depth from surface (g/cm3)

om_mean_0_5: mean organic matter in 0-5 cm depth from surface (%)

ph_mean_0_5: mean soil pH in H2O in 0-5 cm depth from surface (N/A)

silt_mean_0_5: mean silt percentage in 0-5 cm depth from surface (%)

sand_mean_0_5: mean sand percentage in 0-5 cm depth from surface (%)

clay_mean_0_5: mean clay percentage in 0-5 cm depth from surface (%)

WorldClim 2

(30 arc-second ~ 1km, resampled to 3 arc-second using bilinear interpolation)

wind: mean annual wind speed (m/s)

srad: mean annual solar radiation (m/s)

BAECV

(1 arc-second ~ 30m, not resampled)

bp: burn probability (%) [Note: Only used in logistic regression]