

Predicting the spectral information of future land cover using machine learning

Patil, Sopan; Gu, Yuting; Dias, A.; Steiglitz, Marc; Turk, Greg

International Journal of Remote Sensing

DOI:

[10.1080/01431161.2017.1343512](https://doi.org/10.1080/01431161.2017.1343512)

Published: 01/01/2017

Peer reviewed version

[Cyswllt i'r cyhoeddiad / Link to publication](#)

Dyfyniad o'r fersiwn a gyhoeddwyd / Citation for published version (APA):

Patil, S., Gu, Y., Dias, A., Steiglitz, M., & Turk, G. (2017). Predicting the spectral information of future land cover using machine learning. *International Journal of Remote Sensing*, 38, 5592-5607. <https://doi.org/10.1080/01431161.2017.1343512>

Hawliau Cyffredinol / General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Predicting the spectral information of future land cover using machine learning

Sopan D. Patil¹, Yuting Gu², Felipe S. A. Dias³, Marc Stieglitz³, Greg Turk²

¹ School of Environment, Natural Resources and Geography,
Bangor University,
Deiniol Road, Bangor, LL57 2UW, United Kingdom

² School of Interactive Computing,
Georgia Institute of Technology,
85 Fifth Street NW, Atlanta, GA 30308, United States of America

³ School of Civil and Environmental Engineering,
Georgia Institute of Technology,
790 Atlantic Drive, Atlanta, GA 30332, United States of America

Submission to: International Journal of Remote Sensing

Corresponding author: Sopan D. Patil (email: s.d.patil@bangor.ac.uk, Tel: +44 1248388294)

Funding information: This work was supported by the National Science Foundation under Award Number 1027870 (CDI-Type Small Resources Supercomputing: High Performance Computing in the Earth Sciences).

1 **Abstract**

2 Application of machine learning models to study land cover change is typically restricted to the
3 change detection of categorical, i.e., classified, land cover data. In this study, our aim is to
4 extend the utility of such models to predict the spectral band information of satellite images. A
5 Random Forests (RF) based machine learning model is trained using topographic and historical
6 climatic variables as inputs to predict the spectral band values of high-resolution satellite
7 imagery across two large sites in the western United States, New Mexico (10,570 km²) and
8 Washington (9,400 km²). The model output is used to obtain a true colour photorealistic image
9 and an image showing the Normalized Difference Vegetation Index (NDVI) values. We then
10 use the trained model to explore what the land cover might look like for a climate change
11 scenario during the 2061-2080 period. The RF model achieves high validation accuracy for both
12 sites during the training phase, with the coefficient of determination (R^2) = 0.79 for New Mexico
13 site and R^2 = 0.73 for Washington site. For the climate change scenario, prominent land cover
14 changes are characterized by an increase in the vegetation cover at the New Mexico site and a
15 decrease in the perennial snow cover at the Washington site. Our results suggest that direct
16 prediction of spectral band information is highly beneficial due to the ability it provides for
17 deriving ecologically relevant products, which can be used to analyse land cover change
18 scenarios from multiple perspectives.

19

20 **Keywords:** Land cover change; climate change; machine learning; Random forest; Landsat.

21 **1 Introduction**

22 Recent warming of the climate has led to large-scale changes in earth's land cover. Large
23 scale warming has resulted in a shift in the dominant vegetation species to higher latitudes and
24 higher elevations, which has been reported in many parts of the world [*Walther et al.*, 2002; *Root*
25 *et al.*, 2003; *Kelly and Goulden*, 2008; *Lenoir et al.*, 2008; *VanDerWal et al.*, 2013]. Throughout
26 the southwest US, woody species have been encroaching on grasslands [*Barger et al.*, 2011]. In
27 southwestern Wyoming, where precipitation has been trending down for the last thirty years,
28 sagebrush vegetation have been giving way to bare ground [*Homer et al.*, 2015]. In many
29 western states of the US, where seasonal snowmelt accounts for a large fraction of the annual
30 water supply, winter snow accumulation and perennial snow cover has been decreasing. *Mote*
31 [2003] has shown that from the mid to latter half of the twentieth century, winter snow
32 accumulation at several locations along the Cascades Mountain Range fell by more than 40%.
33 *Hall et al.* [2015] have reported that in north-western Wyoming the winter snowmelt is 16 ± 10
34 days earlier in 2000s compared to the period 1972 - 1999. At higher latitudes, where warming
35 has been significantly greater than the planetary average, there has been simultaneous shortening
36 of the snow season [*Groisman et al.*, 1994; *Stow et al.*, 2004] and lengthening of the vegetation
37 growing season [*Foster*, 1989; *Foster et al.*, 1992; *Stone et al.*, 2002]. These are just some of the
38 land cover changes that studies have documented within the last 100 years. However, while this
39 evidence of change provides a view to the future change, it nevertheless remains highly uncertain
40 what changes will occur in the global land cover over the next 100 years.

41 Despite high uncertainty, numerous studies have attempted to model the potential impact
42 of climate change on future land cover [*Pearson and Dawson*, 2003; *Sitch et al.*, 2003, 2008;
43 *Krinner et al.*, 2005; *Rogan et al.*, 2008]. We can broadly classify these modelling efforts into

44 those using physically based and statistically based models. Physically based models provide a
45 mechanistic framework in which mathematical representation of individual processes, such as
46 vegetation growth and decline, snow dynamics, and land-atmosphere exchanges of water and
47 carbon, can be coupled to simulate an integrated landscape response to climate forcing. For
48 instance, *Sitch et al.* [2003] developed the Lund-Potsdam-Jena (LPJ) Dynamic Global
49 Vegetation Model (DGVM) to simulate the response of terrestrial vegetation to climate forcing
50 and demonstrated its application globally at $0.5^\circ \times 0.5^\circ$ spatial resolution. *Campbell et al.* [2010]
51 used the Simultaneous Heat and Water (SHAW) model to simulate future changes in snowpack
52 and soil frost at the Hubbard Brook Experimental Forest in New Hampshire, USA with climate
53 forcing from three different General Circulation Models (GCMs). Physically based models have
54 the benefit that they can be used to infer the cause and effect of land cover change at the level of
55 individual physical processes [*Parker et al.*, 2003; *Pauleit et al.*, 2005; *Pitman et al.*, 2009].
56 However, these models suffer from the large number of simulations necessary to adequately
57 constrain parameter values, and therefore can be both time consuming and, in many instances,
58 beyond the available computing power for many researchers. As a result, physically based
59 simulations tend to make a compromise in their spatial resolution [*Brovkin et al.*, 2006; *Verburg*
60 *et al.*, 2011] or their areal extent [*Tague et al.*, 2009; *Abdelnour et al.*, 2011, 2013].

61 Statistically based land cover change models, on the other hand, operate on the premise
62 that a strong relationship exists between the geographical distribution of land cover and the
63 environmental and climate conditions and that these relationships can be empirically extracted
64 using statistical machine learning methods [*DeFries and Chan*, 2000; *Guisan and Zimmermann*,
65 2000; *McIver and Friedl*, 2002; *Brown de Colstoun et al.*, 2003; *Guisan et al.*, 2006; *Klein et al.*,
66 2012]. Machine learning refers to a broad set of computational techniques used for identifying

67 patterns in data and are usually applied where standard techniques such as regression analysis are
68 not applicable. Machine learning algorithms statistically learn patterns and rules based on
69 present correlations defined by a training set of data and provides a learned mapping between
70 predictor variables (or attributes) and a target variable [Witten *et al.*, 1993; Bishop, 2006]. Once
71 a model is developed through training, it can be used to predict the target variable in situations
72 where the predictor variables are known but the target variable is not [Mitchell, 1997]. Some of
73 the widely used machine learning techniques include Neural Networks (NN), Support Vector
74 Machines (SVM), Classification Trees (CT), Regression Trees (RT), Random Forests (RF),
75 Boosted Regression Trees (BRT), and Multivariate Adaptive Regression Splines [Vapnik, 1999;
76 Domingos, 2012; Alpaydin, 2014].

77 Machine learning models have been widely used to predict the changes in land cover for
78 a given site or region. Rogan *et al.* [2008] compared three different machine learning models
79 (CT, Maximum Likelihood Classification, and NN) to detect changes in land cover classes
80 across two sites in California, USA between the years 1990 and 1996. Similar model
81 comparison was done by Schneider [2012] for land cover change detection in China across five
82 time periods between 1988 and 2009. Pearson *et al.* [2013] used a RF model to identify
83 relationships between 19 bioclimatic variables from the WorldClim database and eight tundra
84 vegetation types in the Arctic, and then used the trained model to predict future vegetation cover
85 classes for the climate change scenarios in the 2050s. Statistical machine learning models have
86 an advantage over the physically based models due to their significantly faster computational
87 speed and better predictive capacity [Im and Jensen, 2005; Rogan *et al.*, 2008]. Thus, they can
88 operate at both high spatial resolutions and over very large areas with much lower computational
89 overhead. However, one limitation of the machine learning models is that their application has

90 so far been restricted to the change detection/prediction of categorical (i.e., classified) land cover
91 information.

92 In this paper, our goal is to extend the utility of machine learning models to predict the
93 spectral band information of high-resolution satellite based land cover images (which is
94 continuous scale numerical data) for a future climate change scenario. The rationale for doing so
95 is two-fold. First, there is a body of evidence that strongly relates remote sensing proxies, such
96 as the Normalized Difference Vegetation Index (NDVI), to ecologically important processes
97 [*Roughgarden et al.*, 1991; *Kerr and Ostrovsky*, 2003; *Pettorelli et al.*, 2005], and their
98 prediction into the future will offer a quantitative understanding of ecological change.

99 Availability of spectral band information for a future scenario would be critical to derive such
100 proxy data. Second, as will be demonstrated, our methodology can be used to provide a
101 photorealistic view of land cover change, which from a conceptual vantage point, provides new
102 and intuitive insights to understand the implications of change. To conduct this research, we
103 have used the topographic and historical climate data (1950-2000) from two large sites in the
104 United States, one in the state of New Mexico and the other in the state of Washington, to train a
105 RF machine learning model. The model is trained to predict the spectral values from bands 1
106 (Blue), 2 (Green), 3 (Red), and 4 (Near Infrared) of a Landsat 7 image. Then, with the GCM
107 climate forecast data from the 2061-2080 period as input, we use the trained model to predict the
108 future band information and its derivative RGB and NDVI images. The data used include
109 Landsat 7 reflectance imagery, mean annual temperature and annual precipitation for the 1950-
110 2000 period, Digital Elevation Model data, and the future climate projections generated using the
111 Goddard Institute for Space Studies (GISS) GCM version E2 that are downscaled and bias
112 corrected to the current climate.

113

114 2 Study Area and Data

115 2.1 Study sites

116 Our New Mexico site (Figure 1) is located in the north-central region of New Mexico
117 state in the western US and includes the N-S flowing Rio Grande River, the Jimenez Mountains
118 on the west, and the Santa Fe National Forest on the east. Elevation ranges from 1573 to 3972
119 m. The annual mean temperature ranges from -1.3°C at higher altitudes to 12.7°C in the valleys.
120 Annual precipitation ranges from 250 mm in the valleys to 1000 mm in the uplands. Dominant
121 vegetation types include grasses near the river channel, shrubs in the lowlands and along the
122 mountain slopes, and evergreen vegetation in the uplands. Uplands also include grasses and a
123 small fraction of mixed forest. The soil type in this region consists mainly of Entisols,
124 Inceptisols and Alisols. Exposed rock formations are also present in areas surrounding the
125 mountain peaks [Wolock, 1997]. The total area covered is 10,570 km².

126 [Insert Figure 1 here]

127 Our Washington site (Figure 1) is located in the northwest part of Washington state and
128 includes the North Cascades National Park and part of the Mount Baker-Snoqualmie National
129 Forest. Elevation ranges from 70 m in the southwest to 3300 m in the northeast. The North
130 Cascade Range is oriented in a NW-SE direction and divides the region into distinct regimes;
131 cool and wet to the west of the range during winter and cold and dry to the east. Summers are
132 typically dry throughout the region. The predominant vegetation is evergreen forest, which
133 covers more than 60% of the site area. Major tree species include Western Hemlock, Pacific
134 Silver Fir, Subalpine Mountain Hemlock, Alpine, Subalpine Fir, and Douglas Fir [Crawford *et*
135 *al.*, 2009]. Other significant vegetation are shrubs, covering 14% of the territory, grasslands are

136 8% of the area, and deciduous forests are 1% of the area. The most distinctive feature of this
137 landscape is the arc of perennial snow that covers about 1.5% of the land area. Soils are
138 predominantly Andisols, Inceptisols, and exposed rock formation (rock outcrops) at higher
139 altitudes of the mountain range [Wolock, 1997]. Rock outcrops account for almost 8% of the
140 area. Annual mean temperature ranges from -4.9°C at higher altitudes to 10.5°C at lower
141 altitudes. Annual precipitation varies between 460 mm, east of the Cascade Mountains, to 2087
142 mm on the western side of the mountains [Hijmans *et al.*, 2005]. The total area covered is 9,400
143 km².

144 **2.2 Data**

145 Table 1 summarizes the spatial data used as model inputs at each of the two study sites.
146 We use the 32 day raw composite satellite images from Landsat 7, specifically seeking
147 information on the values of spectral bands 1, 2, 3, and 4, which correspond to the blue, green,
148 red, and near infrared colour channels, respectively. Both historical and future climate datasets
149 (items 4 and 5 in Table 1) are obtained from the WorldClim dataset [Hijmans *et al.*, 2005]. The
150 historical observed climate data by Hijmans *et al.* [2005] has used observed meteorological
151 station data (47544 stations for precipitation and 24542 stations for air temperature) from a
152 variety of sources, such as Global Historical Climatology Network (GHCN v2), World
153 Meteorological Organization's Climate Normals (WMO CLINO), and Food and Agricultural
154 Organization's Agroclimatic Database (FAOCLIM 2.0). These observed point data have been
155 interpolated over a 1 km global grid using the thin-plate smoothing spline algorithm. As shown
156 in Figure 1 of Hijmans *et al.* [2005], the meteorological station density is amongst the highest in
157 the continental United States. The downscaled 2061-2080 climate data from the GISS E2 model
158 output are for the Representative Concentration Pathway (RCP) 8.5 scenario [IPCC, 2013]. To

159 ensure fast computation as well as uniformity amongst the different datasets, we resample all the
160 above-mentioned data onto a common 150 m resolution grid.

161 [Insert Table 1 here]

162

163 **3 Methods**

164 **3.1 Machine learning model**

165 We use the RF model [Breiman, 2001], which is an ensemble-based machine learning
166 method, to predict the spectral band information of Landsat images. The spatial data used as
167 model inputs, i.e., the predictor variables, are elevation, aspect, slope, mean annual precipitation
168 and temperature. The model outputs, i.e., the target variables, are the spectral values from bands
169 1, 2, 3, and 4 of the Landsat image.

170 Each ensemble member in the RF model is a Decision Trees (DT) model, which is
171 essentially an inverted binary tree structure where splitting rules govern the flow of decisions.
172 The DT algorithm begins at the top node and proceeds down through internal nodes and
173 branches. There are two main types of DT models: (1) CT, which are used when the data type of
174 target variables is categorical, and (2) RT, which are used when the data type of target variables
175 is numerical. Since our target variables are the spectral band values in every pixel of the Landsat
176 data, we use RT as the base ensemble constituent of our RF model. Each node of RT is a binary
177 split that is conditional based on the value of a predictor variable. The particular form of RT that
178 we use here is Classification and Regression Tree (CART). CART builds a RT in a top-down
179 manner, first creating a root node and progressively splitting the data into two sub-trees. The
180 final output of RF model is the mean of the output from all individual RT models in the
181 ensemble.

182 A drawback of the DT models is their tendency to overfit the training dataset by building
183 very deep trees [Bramer, 2007]. This can lead to poor model performance when making
184 predictions outside the training dataset. RF models reduce the risk of overfitting in two main
185 ways. Firstly, given that the RF model structure is an ensemble of a large number of DT models,
186 its output is not overly dependent on that of any single DT model. Secondly, when creating the
187 training dataset for its ensemble member models, the RF model uses the bootstrap aggregating
188 method (also referred to as bagging) [Breiman, 1996]. In this method, the original training
189 dataset is sampled with replacement, thereby creating a sub-sampled dataset that has the same
190 length as the original training dataset. The use of bagging method ensures that: (1) each
191 individual DT model in the ensemble is trained with a slightly different dataset, and (2) part of
192 the original training dataset that is left out due to bagging can be used as the test dataset to
193 determine model performance (also known as the out-of-bag (OOB) score). Here, we use the
194 coefficient of determination (R^2) to measure the RF model's OOB score.

195 One of the main controlling factors in RF model's performance is the number of its
196 ensemble members (i.e., individual RT). Typically, having too few ensemble members leads to a
197 poor OOB score, and increasing the number of ensemble members can improve the OOB score.
198 However, the improvement in model performance becomes marginal once a certain threshold of
199 ensemble members is crossed, and having too many ensemble members simply adds to the
200 computational cost without any performance gain. During preliminary tests of the RF model
201 with our datasets, we found that having more than 100 RT model ensembles provides virtually
202 no improvement in the OOB score. Therefore, for all the results presented in this paper, our RF
203 model consists of an ensemble of 100 RT models.

204 During the training phase of RF model, we use the historical climate data (see Table 1)

205 and topographic variables (elevation, slope, and aspect) as model inputs. The spectral band
 206 information from Landsat 7 images is used for comparison with model outputs to calibrate the
 207 RF model. In the prediction phase, the RF model uses the future climate data and the
 208 topographic variables as model inputs. We use the RF model in the scikit-learn machine learning
 209 package that is implemented in Python® programming language [Pedregosa *et al.*, 2011].

210 **3.2 Post-processing of the model outputs**

211 The output of RF model is the spectral band information of the Blue, Green, Red, and
 212 Near Infrared bands of the Landsat image. We use this output information to create two derived
 213 products: (1) a true colour photorealistic image consisting of the Red, Green and Blue (RGB)
 214 colour bands, and (2) an image showing the NDVI values of the study sites. NDVI value for
 215 each pixel is calculated using the following formula.

$$216 \quad \text{NDVI} = \frac{B_4 - B_3}{B_4 + B_3} \quad (1)$$

217 where, B_4 is the Near Infrared colour band and B_3 is the Red colour band of a Landsat 7 satellite
 218 image.

219 In addition to the OOB score obtained during the RF model's training phase (see Section
 220 3.1), we calculate two more error metrics to assess the model performance for the final trained
 221 images. For the photorealistic image, the error at each pixel is calculated as follows.

$$222 \quad E_{\text{RGB}} = \sqrt{(B_{1,\text{obs}} - B_{1,\text{pred}})^2 + (B_{2,\text{obs}} - B_{2,\text{pred}})^2 + (B_{3,\text{obs}} - B_{3,\text{pred}})^2} \quad (2)$$

223 where, B_1 is the Blue colour band, B_2 is the Green colour band, and obs and pred denote the
 224 observed and model predicted spectral band values, respectively. For the NDVI image, the error
 225 at each pixel is calculated as follows.

$$226 \quad E_{\text{NDVI}} = \text{NDVI}_{\text{obs}} - \text{NDVI}_{\text{pred}} \quad (3)$$

227

228 4 Results and Discussion

229 We first present the results from the RF model's training phase, which uses the
230 topographic and historical climate data to train the model for predicting the four spectral band
231 values (Blue, Green, Red, and Near Infrared) of the Landsat image. For the New Mexico site,
232 the OOB R^2 value for the prediction of four spectral band values is 0.79. For the Washington
233 site, the OOB R^2 value is 0.73. Figure 2 compares the original Landsat and the trained true
234 colour photorealistic images for both study sites. Images produced using the RF model are able
235 to capture almost all the major land cover features at both sites, and there is good visual
236 agreement with the original Landsat images.

237 [Insert Figure 2 here]

238 Figure 3 compares the NDVI values between the original and trained images at both
239 study sites. For the New Mexico site, the R^2 value between observed and simulated NDVI
240 values is 0.97. For the Washington site, $R^2 = 0.96$ between the observed and simulated NDVI
241 values. It is worth noting here that the R^2 values are much higher for NDVI because at each site
242 we compare all the pixels between the observed and simulated data, whereas for the raw spectral
243 band values, we only compare the pixels that were left out from training due to bagging.

244 [Insert Figure 3 here]

245 Figure 4 shows the RGB error between the original Landsat and the model generated
246 photorealistic images calculated at each pixel using Equation 2. The error across RGB band
247 values is lower at the New Mexico site, where there is no prominent geographical pattern for
248 high error values. Conversely, the Washington site has higher error across the RGB band values,
249 and the high error pixels are predominantly located in areas adjacent to the perennial snow cover.
250 Figure 5 shows the error between the original and model generated NDVI images calculated at

251 each pixel using Equation 3. Consistent with the RGB error shown in Figure 4, the NDVI error
252 values are lower at the New Mexico site compared to the Washington site.

253 [Insert Figure 4 here]

254 [Insert Figure 5 here]

255 Next, we focus on the prediction phase of the RF model, which uses the topographic and
256 future climate data (see Table 1) to predict the spectral band values for the RCP 8.5 climate
257 change scenario. Figure 6 compares the historical (trained) and the future (predicted) true colour
258 photorealistic images for both study sites. For the New Mexico site, the most prominent change
259 is the increase in vegetation cover within the forested areas on either side of the Rio Grande
260 river. For the Washington site, there is a substantial decrease in the perennial snow cover in the
261 vicinity of Mount Baker (top left of the image) as well as across other mountainous areas along
262 the Cascades Mountain Range. Many areas that appear as snow covered in the trained historical
263 image are replaced by bare ground in the future scenario image. Figure 7 shows the NDVI
264 images at both study sites for the historical (trained) and the future (predicted) scenarios. The
265 overall increase in vegetation cover at the New Mexico site is discernible from the NDVI
266 comparison. Interestingly, the reduction in perennial snow cover for the Washington site can be
267 perceived through the increase in NDVI values in the mountainous areas.

268 [Insert Figure 6 here]

269 [Insert Figure 7 here]

270 We have attempted to demonstrate that a machine learning model that is trained to predict
271 the spectral band information of satellite images can be highly useful for scenario-based
272 assessment of future land cover. Moreover, given the richness of information available from
273 spectral band values, it is possible to create several derived products to analyse (and visualize)

274 land cover response to climate change from multiple perspectives. In our view, this is a non-
275 trivial improvement from previous land cover change studies which had limited the application
276 of machine learning models to categorical land cover classification data [Rogan *et al.*, 2008;
277 Schneider, 2012; Pearson *et al.*, 2013]. It is worth mentioning here that the categorical land
278 cover classification data itself is a product that is derived from satellite image data, similar to the
279 photorealistic images and NDVI data shown in our study. Several methods, many of them based
280 on machine learning, exist to convert the satellite's spectral band information into land cover
281 classes [Friedl and Brodley, 1997; DeFries and Chan, 2000; Hansen *et al.*, 2000; Qian *et al.*,
282 2015]. We would also like to note that our focus on predicting only the first four spectral bands
283 of the Landsat 7 images was governed by our choice of derivative products, the NDVI and RGB
284 images (which require the use of first four bands only). Nonetheless, the techniques presented in
285 this study are applicable to predicting the information from any desired number of satellite
286 spectral bands, depending on the final product sought by the end user.

287 Our preference for choosing a RF machine learning model in this study was partly due to
288 the fact that its ensemble constituents are comprised of DT models, which offers a number of
289 attractive features over other statistical learning techniques. DT models are non-parametric and
290 therefore make no assumptions regarding the distribution of the data. They are structurally
291 explicit models and provide for a clear interpretation of the connections between the predictor
292 and target variables. Normalization of attribute distances is unnecessary in these models, and
293 their internal structure (essentially a cascading set of data splitting decisions) makes them much
294 more tolerant to redundancies in the information content among the input variables [Song and
295 Lu, 2015]. In addition, they tend to be computationally faster than other machine learning
296 techniques [Witten and Frank, 2005; Kotsiantis *et al.*, 2007; Rogan *et al.*, 2008; Schneider, 2012]

297 such as NN or BRT, and certainly faster than the physically based mechanistic models for a
298 similar resolution data and areal extent. Lastly, as we had mentioned in Section 3.1, the
299 ensemble averaging process in a RF model mitigates the drawbacks caused by the direct use of a
300 standalone DT model. Nonetheless, there are a few assumptions and limitations built into our
301 methodology. Firstly, our model requires long term climatic averages of precipitation and air
302 temperature as inputs. These were chosen because the development of natural vegetation cover
303 is a gradual process and would be a function of past climate over a long time period (in the order
304 of decades) [Dale, 1997; Kangur *et al.*, 2005; Soudzilovskaia *et al.*, 2013], especially for forested
305 areas which are abundant in both our study sites. Unfortunately, this makes the model unsuitable
306 for change detection at short time scales, and a time gap of several decades would be needed
307 between the training and prediction dataset to obtain meaningful change detection. Secondly,
308 our input data was resampled to a common grid resolution of 150 m prior to running the model,
309 which was done to limit the computational expenditure in the desktop runtime setting. Grid
310 resampling does bring another source of uncertainty to the model, but is unavoidable due to
311 different resolutions of our input datasets. Nonetheless, it would be possible to run our model at
312 finer spatial resolutions if additional computational resources are available to the user.

313 As we look forward, the method presented in our study offer both challenges and
314 opportunities. Firstly, our model presumes that the land cover change for the 2061-2080 period
315 is simply the application of learned rules from the historical period to the climate changed
316 environment. Many sites within our two study regions have experienced disturbance due to, for
317 example, grazing pressure and fires [Everett *et al.*, 2000; Floyd *et al.*, 2003; Allen, 2007].
318 However, to a large extent, this is mitigated by the fact that our land cover training is conducted
319 over regions that are much larger than the scale of a typical disturbance. Secondly, the predicted

320 land cover for 2061-2080 period does not indicate the velocity of land cover change in response
321 to changes in precipitation and air temperature [*Loarie et al.*, 2009]. Thus, our model does not
322 provide any mechanistic understanding of how the final predicted state of land cover will be
323 reached.

324 Within the limits of these challenges, the method presented here does provide a few
325 opportunities. Monthly Landsat images are available at the 16 and 32 day time frames going
326 back to 2002, and can provide ample raw data to explore how the seasonality of vegetation will
327 be altered in a future scenario. Ongoing improvements in the satellite sensor technology, such as
328 those in the recently launched Landsat 8 satellite [*Knight and Kvaran*, 2014; *Roy et al.*, 2014],
329 also have the potential to provide increasingly better quality input data to land cover change
330 models. The fast computational speed of the machine learning models permit the rendering of
331 future land cover over much larger areas than our study regions, possibly even covering the
332 entire continental USA. The five predictor variables we used were obtained from three primary
333 data sources: rainfall, air temperature, and elevation (slope and aspect are derivative products of
334 elevation), and were chosen based on what we judged to be important factors for predicting land
335 cover. Nonetheless, we cannot rule out the possibility that, at least in some regions, inclusion of
336 different types of predictor variables could improve the machine learning model's capability to
337 predict land cover. Therefore, there is opportunity to experiment with the predictor variables by
338 adding to or modifying the data sources.

339

340 **5 Conclusions**

341 In this paper, our goal was to extend the utility of machine learning based land cover
342 change models to predict the spectral band information of satellite based land cover images. We

343 used the topographic and historical climate data from two large sites in the United States to train
344 a RF machine learning model to predict the spectral values from bands 1 (Blue), 2 (Green), 3
345 (Red), and 4 (Near Infrared) of Landsat 7 image. We then used the trained model to explore
346 what the land cover might look like for a climate change scenario during the 2061-2080 period
347 through the two derived products. Our results showed that the RF model can accurately
348 reproduce the land cover properties for historical data and is able to provide realistic rendering of
349 future land cover for a climate change scenario. The two derived land cover products
350 (photorealistic RGB image and NDVI image) shown in our results demonstrate that the direct
351 prediction of spectral band information is helpful for deriving ecologically relevant products.
352 We consider this a major strength of our proposed approach because it enables the analysis of
353 land cover change from multiple perspectives.

354 What land cover change will occur over the next 100 years is highly uncertain. However,
355 presuming little is done to reduce the rate of CO₂ emissions, the global air temperatures for the
356 2081–2100 period are projected to be to 1.5 - 4.8 °C higher than for the 1986–2005 period
357 [IPCC, 2013]. This will almost certainly impact regional and global land cover [Krinner *et al.*,
358 2005; Beer *et al.*, 2007; Sitch *et al.*, 2008; Anav *et al.*, 2010; Hickler *et al.*, 2012]. We hope that
359 the method presented here makes a useful contribution towards understanding and predicting
360 these changes.

361

362 **Acknowledgements**

363 This work was supported by the National Science Foundation under Award Number 1027870
364 (CDI-Type Small Resources Supercomputing: High Performance Computing in the Earth
365 Sciences).

367 **References**

- 368 Abdelnour, A., M. Stieglitz, F. Pan, and R. McKane (2011), Catchment hydrological responses
 369 to forest harvest amount and spatial pattern, *Water Resour. Res.*, 47, W09521–W09521,
 370 doi:10.1029/2010WR010165.
- 371 Abdelnour, A., R. B. McKane, M. Stieglitz, F. Pan, and Y. Cheng (2013), Effects of harvest on
 372 carbon and nitrogen dynamics in a Pacific Northwest forest catchment, *Water Resour. Res.*,
 373 49(3), 1292–1313, doi:10.1029/2012WR012994.
- 374 Allen, C. D. (2007), Interactions Across Spatial Scales among Forest Dieback, Fire, and Erosion
 375 in Northern New Mexico Landscapes, *Ecosystems*, 10(5), 797–808, doi:10.1007/s10021-
 376 007-9057-4.
- 377 Alpaydin, E. (2014), *Introduction to machine learning*, MIT press.
- 378 Anav, A., F. D’Andrea, N. Viovy, and N. Vuichard (2010), A validation of heat and carbon
 379 fluxes from high-resolution land surface and regional models, *J. Geophys. Res.*
 380 *Biogeosciences*, 115(G4), G04016, doi:10.1029/2009JG001178.
- 381 Barger, N. N., S. R. Archer, J. L. Campbell, C. Huang, J. A. Morton, and A. K. Knapp (2011),
 382 Woody plant proliferation in North American drylands: A synthesis of impacts on
 383 ecosystem carbon balance, *J. Geophys. Res. Biogeosciences*, 116(G4), n/a-n/a,
 384 doi:10.1029/2010JG001506.
- 385 Beer, C., W. Lucht, D. Gerten, K. Thonicke, and C. Schmullius (2007), Effects of soil freezing
 386 and thawing on vegetation carbon density in Siberia: A modeling analysis with the Lund-
 387 Potsdam-Jena Dynamic Global Vegetation Model (LPJ-DGVM), *Global Biogeochem.*
 388 *Cycles*, 21(1), n/a-n/a, doi:10.1029/2006GB002760.
- 389 Bishop, C. M. (2006), *Pattern recognition and machine learning*, springer New York.
- 390 Bramer, M. (2007), *Principles of data mining*, Springer.
- 391 Breiman, L. (1996), Bagging predictors, *Mach. Learn.*, 24(2), 123–140,
 392 doi:10.1007/BF00058655.
- 393 Breiman, L. (2001), Random Forests, *Mach. Learn.*, 45(1), 5–32,
 394 doi:10.1023/A:1010933404324.
- 395 Brovkin, V., M. Claussen, E. Driesschaert, T. Fichefet, D. Kicklighter, M. F. Loutre, H. D.
 396 Matthews, N. Ramankutty, M. Schaeffer, and A. Sokolov (2006), Biogeophysical effects of
 397 historical land cover changes simulated by six Earth system models of intermediate
 398 complexity, *Clim. Dyn.*, 26(6), 587–600, doi:10.1007/s00382-005-0092-6.
- 399 Brown de Colstoun, E. C., M. H. Story, C. Thompson, K. Commisso, T. G. Smith, and J. R.
 400 Irons (2003), National Park vegetation mapping using multitemporal Landsat 7 data and a
 401 decision tree classifier, *Remote Sens. Environ.*, 85(3), 316–327,
 402 doi:http://dx.doi.org/10.1016/S0034-4257(03)00010-5.
- 403 Campbell, J. L., S. V Ollinger, G. N. Flerchinger, H. Wicklein, K. Hayhoe, and A. S. Bailey
 404 (2010), Past and projected future changes in snowpack and soil frost at the Hubbard Brook

405 Experimental Forest, New Hampshire, USA, *Hydrol. Process.*, 24(17), 2465–2480,
406 doi:10.1002/hyp.7666.

407 Crawford, R. C., C. B. Chappell, C. C. Thompson, and F. J. Rocchio (2009), *Vegetation*
408 *classification of Mount Rainier, North Cascades, and Olympic National Parks. Plant*
409 *association descriptions and identification keys*, Natural Resource Technical Report
410 NPS/NCCN/NRTR—2009/D-586. US Department of the Interior, National Park Service,
411 Natural Resource Program Centre, Fort Collins, CO, US.

412 Dale, V. H. (1997), The Relationship Between Land- Use Change and Climate Change, *Ecol.*
413 *Appl.*, 7(3), 753–769.

414 DeFries, R. S., and J. C.-W. Chan (2000), Multiple Criteria for Evaluating Machine Learning
415 Algorithms for Land Cover Classification from Satellite Data, *Remote Sens. Environ.*,
416 74(3), 503–515, doi:http://dx.doi.org/10.1016/S0034-4257(00)00142-5.

417 Domingos, P. (2012), A Few Useful Things to Know About Machine Learning, *Commun. ACM*,
418 55(10), 78–87, doi:10.1145/2347736.2347755.

419 Everett, R. L., R. Schellhaas, D. Keenum, D. Spurbeck, and P. Ohlson (2000), Fire history in the
420 ponderosa pine/Douglas-fir forests on the east slope of the Washington Cascades, *For. Ecol.*
421 *Manage.*, 129(1–3), 207–225, doi:http://dx.doi.org/10.1016/S0378-1127(99)00168-1.

422 Floyd, M. L., T. L. Fleischner, D. Hanna, and P. Whitefield (2003), Effects of Historic Livestock
423 Grazing on Vegetation at Chaco Culture National Historic Park, New Mexico, *Conserv.*
424 *Biol.*, 17(6), 1703–1711, doi:10.1111/j.1523-1739.2003.00227.x.

425 Foster, J. L. (1989), The Significance of the Date of Snow Disappearance on the Arctic Tundra
426 as a Possible Indicator of Climate Change, *Arct. Alp. Res.*, 21(1), 60–70,
427 doi:10.2307/1551517.

428 Foster, J. L., J. W. Winchester, and E. G. Dutton (1992), The date of snow disappearance on the
429 Arctic tundra as determined from satellite, meteorological station and radiometric in situ
430 observations, *Geosci. Remote Sensing, IEEE Trans.*, 30(4), 793–798,
431 doi:10.1109/36.158874.

432 Friedl, M. A., and C. E. Brodley (1997), Decision tree classification of land cover from remotely
433 sensed data, *Remote Sens. Environ.*, 61(3), 399–409, doi:http://dx.doi.org/10.1016/S0034-
434 4257(97)00049-7.

435 Groisman, P. Y., T. R. Karl, and R. W. Knight (1994), Observed Impact of Snow Cover on the
436 Heat Balance and the Rise of Continental Spring Temperatures, *Sci.*, 263(5144), 198–200,
437 doi:10.1126/science.263.5144.198.

438 Guisan, A., and N. E. Zimmermann (2000), Predictive habitat distribution models in ecology,
439 *Ecol. Modell.*, 135(2–3), 147–186, doi:http://dx.doi.org/10.1016/S0304-3800(00)00354-9.

440 Guisan, A., A. Lehmann, S. Ferrier, M. Austin, J. M. C. C. Overton, R. Aspinall, and T. Hastie
441 (2006), Making better biogeographical predictions of species' distributions, *J. Appl. Ecol.*,
442 43(3), 386–392, doi:10.1111/j.1365-2664.2006.01164.x.

443 Hall, D. K., C. J. Crawford, N. E. DiGirolamo, G. A. Riggs, and J. L. Foster (2015), Detection of
444 earlier snowmelt in the Wind River Range, Wyoming, using Landsat imagery, 1972–2013,
445 *Remote Sens. Environ.*, 162, 45–54, doi:http://dx.doi.org/10.1016/j.rse.2015.01.032.

446 Hansen, M. C., R. S. Defries, J. R. G. Townshend, and R. Sohlberg (2000), Global land cover
447 classification at 1 km spatial resolution using a classification tree approach, *Int. J. Remote*
448 *Sens.*, 21(6–7), 1331–1364, doi:10.1080/014311600210209.

449 Hickler, T. et al. (2012), Projecting the future distribution of European potential natural
450 vegetation zones with a generalized, tree species-based dynamic vegetation model, *Glob.*
451 *Ecol. Biogeogr.*, 21(1), 50–63, doi:10.1111/j.1466-8238.2010.00613.x.

452 Hijmans, R. J., S. E. Cameron, J. L. Parra, P. G. Jones, and A. Jarvis (2005), Very high
453 resolution interpolated climate surfaces for global land areas, *Int. J. Climatol.*, 25(15),
454 1965–1978, doi:10.1002/joc.1276.

455 Homer, C. G., G. Xian, C. L. Aldridge, D. K. Meyer, T. R. Loveland, and M. S. O’Donnell
456 (2015), Forecasting sagebrush ecosystem components and greater sage-grouse habitat for
457 2050: Learning from past climate patterns and Landsat imagery to predict the future, *Ecol.*
458 *Indic.*, 55, 131–145, doi:http://dx.doi.org/10.1016/j.ecolind.2015.03.002.

459 Im, J., and J. R. Jensen (2005), A change detection model based on neighborhood correlation
460 image analysis and decision tree classification, *Remote Sens. Environ.*, 99(3), 326–340,
461 doi:http://dx.doi.org/10.1016/j.rse.2005.09.008.

462 IPCC (2013), *Climate Change 2013: The Physical Science Basis. Contribution of Working*
463 *Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate*
464 *Change*, Cambridge, United Kingdom and New York, NY, USA.

465 Kangur, A., H. Korjus, K. Jõgiste, and A. Kiviste (2005), A conceptual model of forest stand
466 development based on permanent sample-plot data in Estonia, *Scand. J. For. Res.*, 20(S6),
467 94–101.

468 Kelly, A. E., and M. L. Goulden (2008), Rapid shifts in plant distribution with recent climate
469 change, *Proc. Natl. Acad. Sci.*, 105(33), 11823–11826, doi:10.1073/pnas.0802891105.

470 Kerr, J. T., and M. Ostrovsky (2003), From space to species: ecological applications for remote
471 sensing, *Trends Ecol. Evol.*, 18(6), 299–305, doi:10.1016/S0169-5347(03)00071-5.

472 Klein, I., U. Gessner, and C. Kuenzer (2012), Regional land cover mapping and change detection
473 in Central Asia using MODIS time-series, *Appl. Geogr.*, 35(1–2), 219–234,
474 doi:http://dx.doi.org/10.1016/j.apgeog.2012.06.016.

475 Knight, J. E., and G. Kvaran (2014), Landsat-8 Operational Land Imager Design,
476 Characterization and Performance, *Remote Sens.*, 6(11), doi:10.3390/rs61110286.

477 Kotsiantis, S. B., I. Zaharakis, and P. Pintelas (2007), Supervised machine learning: A review of
478 classification techniques,

479 Krinner, G., N. Viovy, N. de Noblet-Ducoudré, J. Ogée, J. Polcher, P. Friedlingstein, P. Ciais, S.
480 Sitch, and I. C. Prentice (2005), A dynamic global vegetation model for studies of the
481 coupled atmosphere-biosphere system, *Global Biogeochem. Cycles*, 19(1), n/a-n/a,
482 doi:10.1029/2003GB002199.

483 Lenoir, J., J. C. Gégout, P. A. Marquet, P. de Ruffray, and H. Brisse (2008), A Significant
484 Upward Shift in Plant Species Optimum Elevation During the 20th Century, *Sci.*,
485 320(5884), 1768–1771, doi:10.1126/science.1156831.

486 Loarie, S. R., P. B. Duffy, H. Hamilton, G. P. Asner, C. B. Field, and D. D. Ackerly (2009), The

487 velocity of climate change, *Nature*, 462(7276), 1052–1055.

488 McIver, D. K., and M. A. Friedl (2002), Using prior probabilities in decision-tree classification
489 of remotely sensed data, *Remote Sens. Environ.*, 81(2–3), 253–261,
490 doi:http://dx.doi.org/10.1016/S0034-4257(02)00003-2.

491 Mitchell, T. M. (1997), *Machine learning*, McGraw Hill, Burr Ridge, IL.

492 Mote, P. W. (2003), Trends in snow water equivalent in the Pacific Northwest and their climatic
493 causes, *Geophys. Res. Lett.*, 30(12), n/a-n/a, doi:10.1029/2003GL017258.

494 Parker, D. C., S. M. Manson, M. A. Janssen, M. J. Hoffmann, and P. Deadman (2003), Multi-
495 Agent Systems for the Simulation of Land-Use and Land-Cover Change: A Review, *Ann.*
496 *Assoc. Am. Geogr.*, 93(2), 314–337, doi:10.1111/1467-8306.9302004.

497 Pauleit, S., R. Ennos, and Y. Golding (2005), Modeling the environmental impacts of urban land
498 use and land cover change—a study in Merseyside, UK, *Landsc. Urban Plan.*, 71(2–4),
499 295–310, doi:http://dx.doi.org/10.1016/j.landurbplan.2004.03.009.

500 Pearson, R. G., and T. P. Dawson (2003), Predicting the impacts of climate change on the
501 distribution of species: are bioclimate envelope models useful?, *Glob. Ecol. Biogeogr.*,
502 12(5), 361–371, doi:10.1046/j.1466-822X.2003.00042.x.

503 Pearson, R. G., S. J. Phillips, M. M. Loranty, P. S. A. Beck, T. Damoulas, S. J. Knight, and S. J.
504 Goetz (2013), Shifts in Arctic vegetation and associated feedbacks under climate change,
505 *Nat. Clim. Chang.*, 3(7), 673–677.

506 Pedregosa, F. et al. (2011), Scikit-learn: Machine Learning in {P}ython, *J. Mach. Learn. Res.*,
507 12, 2825–2830.

508 Pettorelli, N., J. O. Vik, A. Mysterud, J.-M. Gaillard, C. J. Tucker, and N. C. Stenseth (2005),
509 Using the satellite-derived NDVI to assess ecological responses to environmental change,
510 *Trends Ecol. Evol.*, 20(9), 503–510, doi:10.1016/j.tree.2005.05.011.

511 Pitman, A. J. et al. (2009), Uncertainties in climate responses to past land cover change: First
512 results from the LUCID intercomparison study, *Geophys. Res. Lett.*, 36(14), n/a-n/a,
513 doi:10.1029/2009GL039076.

514 Qian, Y., W. Zhou, J. Yan, W. Li, and L. Han (2015), Comparing Machine Learning Classifiers
515 for Object-Based Land Cover Classification Using Very High Resolution Imagery, *Remote*
516 *Sens.*, 7(1), doi:10.3390/rs70100153.

517 Rogan, J., J. Franklin, D. Stow, J. Miller, C. Woodcock, and D. Roberts (2008), Mapping land-
518 cover modifications over large areas: A comparison of machine learning algorithms, *Remote*
519 *Sens. Environ.*, 112(5), 2272–2283, doi:http://dx.doi.org/10.1016/j.rse.2007.10.004.

520 Root, T. L., J. T. Price, K. R. Hall, S. H. Schneider, C. Rosenzweig, and J. A. Pounds (2003),
521 Fingerprints of global warming on wild animals and plants, *Nature*, 421(6918), 57–60.

522 Roughgarden, J., S. W. Running, and P. A. Matson (1991), What Does Remote Sensing Do For
523 Ecology?, *Ecology*, 72(6), 1918–1922, doi:10.2307/1941546.

524 Roy, D. P. et al. (2014), Landsat-8: Science and product vision for terrestrial global change
525 research, *Remote Sens. Environ.*, 145, 154–172,
526 doi:http://dx.doi.org/10.1016/j.rse.2014.02.001.

527 Schneider, A. (2012), Monitoring land cover change in urban and peri-urban areas using dense

528 time stacks of Landsat satellite data and a data mining approach, *Remote Sens. Environ.*,
529 124, 689–704, doi:http://dx.doi.org/10.1016/j.rse.2012.06.006.

530 Sitch, S. et al. (2003), Evaluation of ecosystem dynamics, plant geography and terrestrial carbon
531 cycling in the LPJ dynamic global vegetation model, *Glob. Chang. Biol.*, 9(2), 161–185,
532 doi:10.1046/j.1365-2486.2003.00569.x.

533 Sitch, S. et al. (2008), Evaluation of the terrestrial carbon cycle, future plant geography and
534 climate-carbon cycle feedbacks using five Dynamic Global Vegetation Models (DGVMs),
535 *Glob. Chang. Biol.*, 14(9), 2015–2039, doi:10.1111/j.1365-2486.2008.01626.x.

536 Song, Y., and Y. Lu (2015), Decision tree methods: applications for classification and prediction,
537 *Shanghai Arch. Psychiatry*, 27(2), 130–135, doi:10.11919/j.issn.1002-0829.215044.

538 Soudzilovskaia, N. A., T. G. Elumeeva, V. G. Onipchenko, I. I. Shidakov, F. S. Salpagarova, A.
539 B. Khubiev, D. K. Tekeev, and J. H. C. Cornelissen (2013), Functional traits predict
540 relationship between plant abundance dynamic and long-term climate warming, *Proc. Natl.*
541 *Acad. Sci.*, 110(45), 18180–18184.

542 Stone, R. S., E. G. Dutton, J. M. Harris, and D. Longenecker (2002), Earlier spring snowmelt in
543 northern Alaska as an indicator of climate change, *J. Geophys. Res. Atmos.*, 107(D10), ACL
544 10-1-ACL 10-13, doi:10.1029/2000JD000286.

545 Stow, D. A. et al. (2004), Remote sensing of vegetation and land-cover change in Arctic Tundra
546 Ecosystems, *Remote Sens. Environ.*, 89(3), 281–308,
547 doi:http://dx.doi.org/10.1016/j.rse.2003.10.018.

548 Tague, C., L. Seaby, and A. Hope (2009), Modeling the eco-hydrologic response of a
549 Mediterranean type ecosystem to the combined impacts of projected climate change and
550 altered fire frequencies, *Clim. Change*, 93(1–2), 137–155, doi:10.1007/s10584-008-9497-7.

551 VanDerWal, J., H. T. Murphy, A. S. Kutt, G. C. Perkins, B. L. Bateman, J. J. Perry, and A. E.
552 Reside (2013), Focus on poleward shifts in species' distribution underestimates the
553 fingerprint of climate change, *Nat. Clim. Chang.*, 3(3), 239–243.

554 Vapnik, V. N. (1999), An overview of statistical learning theory, *Neural Networks, IEEE Trans.*,
555 10(5), 988–999, doi:10.1109/72.788640.

556 Verburg, P. H., K. Neumann, and L. Nol (2011), Challenges in using land use and land cover
557 data for global change studies, *Glob. Chang. Biol.*, 17(2), 974–989, doi:10.1111/j.1365-
558 2486.2010.02307.x.

559 Walther, G.-R., E. Post, P. Convey, A. Menzel, C. Parmesan, T. J. C. Beebee, J.-M. Fromentin,
560 O. Hoegh-Guldberg, and F. Bairlein (2002), Ecological responses to recent climate change,
561 *Nature*, 416(6879), 389–395.

562 Witten, I. H., and E. Frank (2005), *Data Mining: Practical machine learning tools and*
563 *techniques*, Morgan Kaufmann.

564 Witten, I. H., S. J. Cunningham, G. Holmes, R. McQueen, and L. Smith (1993), *Practical*
565 *machine learning and its application to problems in agriculture*, University of Waikato,
566 Department of Computer Science.

567 Wolock, D. M. (1997), *STATSGO soil characteristics for the conterminous United States*, US
568 Geological Survey.

Tables

Table 1: Summary information of all the input data used for training the machine learning model

Attribute	Source	Resolution
Elevation	USGS National Elevation Dataset	30 m
Aspect	Calculated from Elevation data	30 m
Slope	Calculated from Elevation data	30 m
Historical mean annual temperature and precipitation	Worldclim – Normal 1950-2000 period [<i>Hijmans et al., 2005</i>]	1000 m
Future mean annual temperature and precipitation	Worldclim – Downscaled GISS E2 2061-2080 period [<i>Hijmans et al., 2005</i>]	1000 m
Landsat 7 reflectance imagery	For New Mexico: 16 October 1999 – 17 November 1999 For Washington: 12 July 2001 – 13 August 2001	30 m

Figures

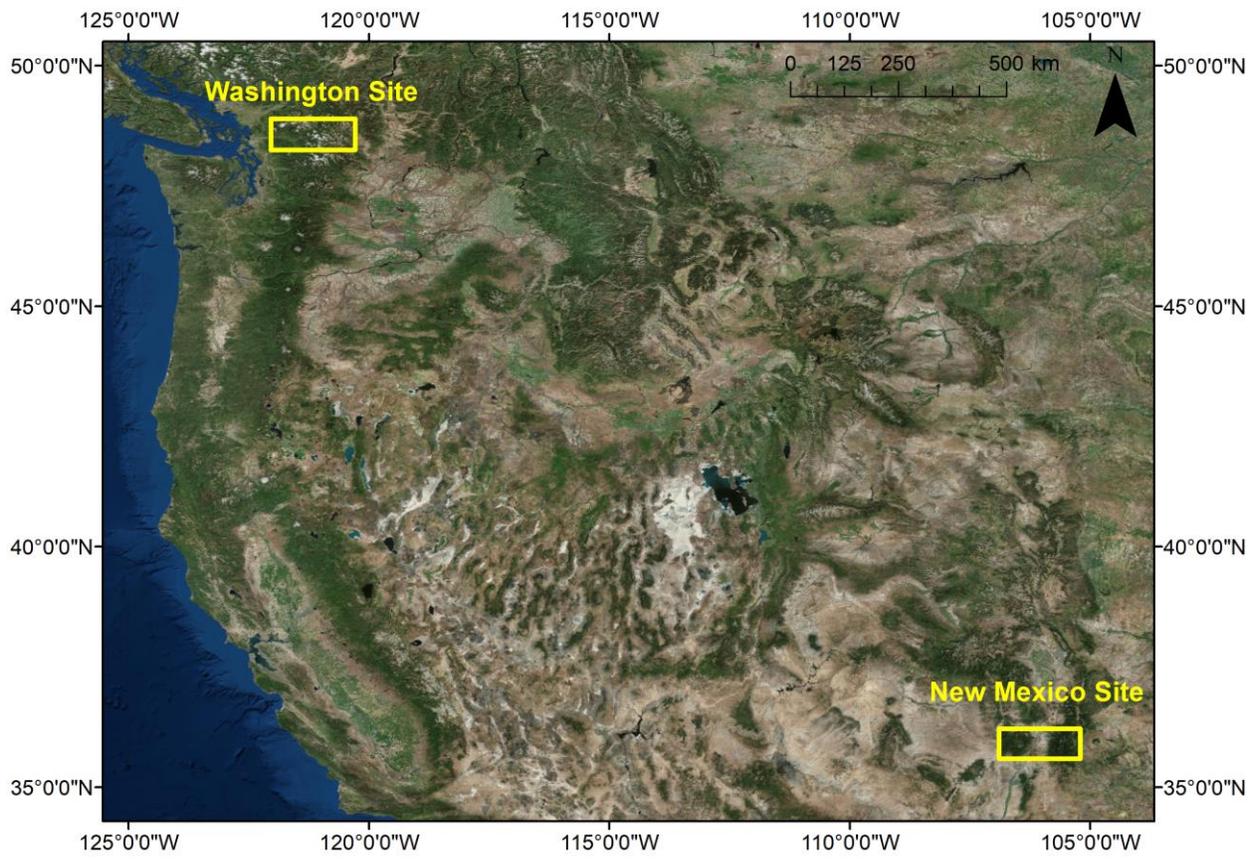


Figure 1: Location map of the two study sites.

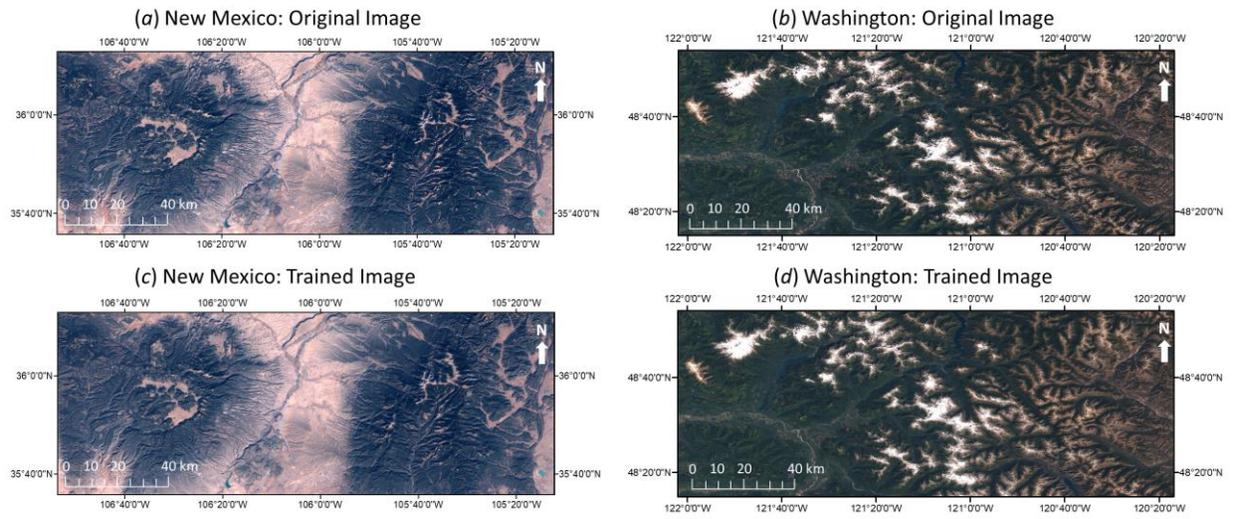


Figure 2: Comparison of the original Landsat 7 images and the RF model trained true colour photorealistic images for the two study sites.

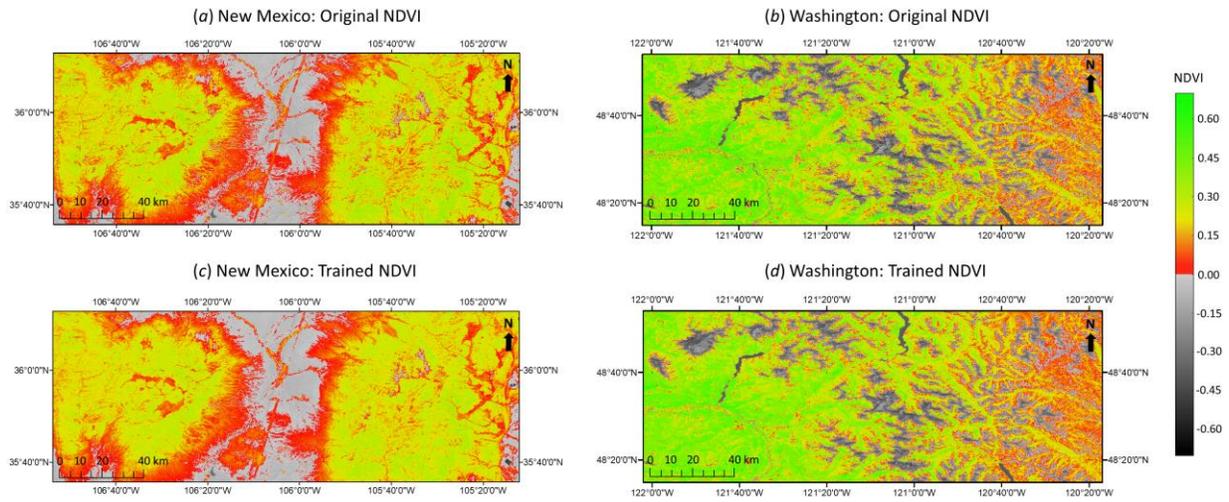


Figure 3: Comparison of the NDVI values between the original historical images (derived from Landsat 7 using Equation 1) and the RF model trained images for the two study sites.

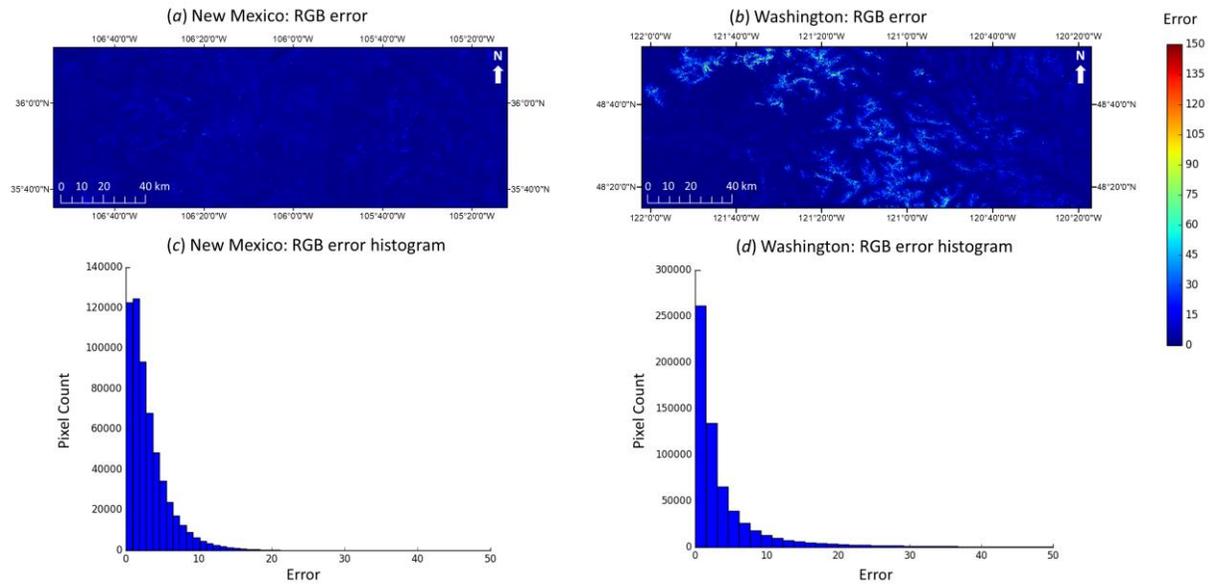


Figure 4: Spatial distribution of the error in RGB band values calculated for each pixel at (a) New Mexico site, and (b) Washington site. Also shown are the error histograms using the data from all the pixels at (c) New Mexico site, and (d) Washington site.

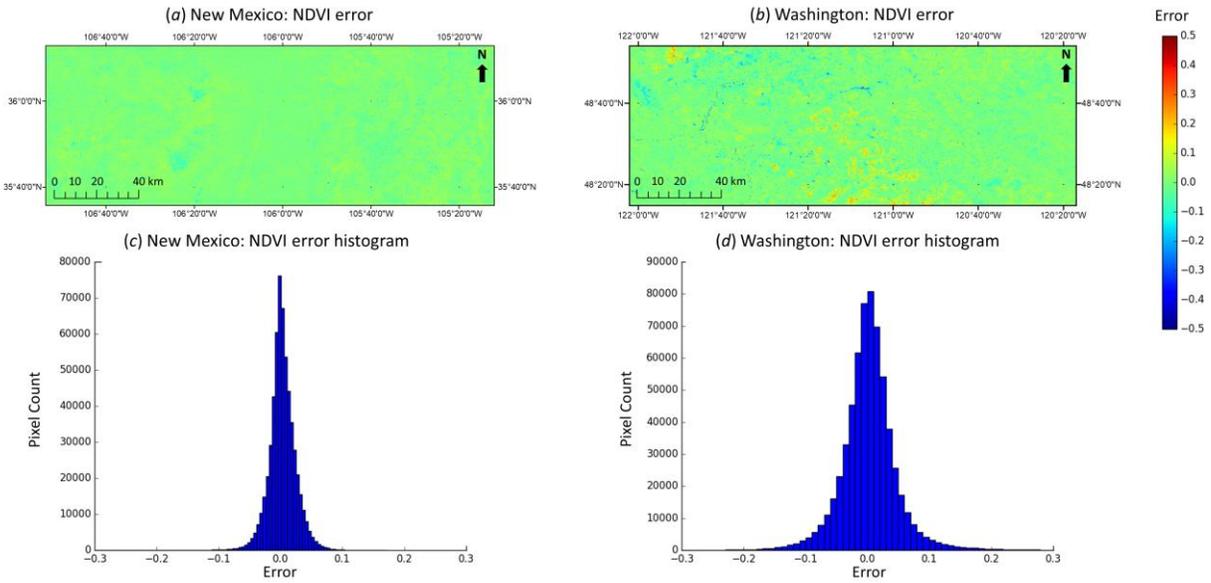


Figure 5: Spatial distribution of the error in NDVI values calculated for each pixel at (a) New Mexico site, and (b) Washington site. Also shown are the error histograms using the data from all the pixels at (c) New Mexico site, and (d) Washington site.

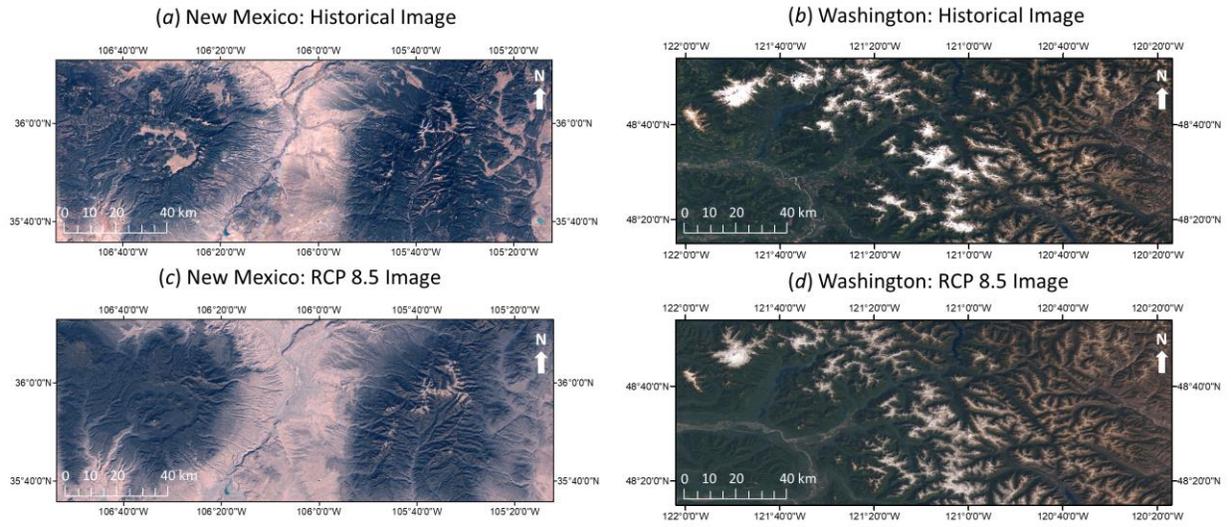


Figure 6: Comparison of the historical (RF model trained) and future (RF model predicted for RCP 8.5 scenario) true colour photorealistic images for the two study sites.

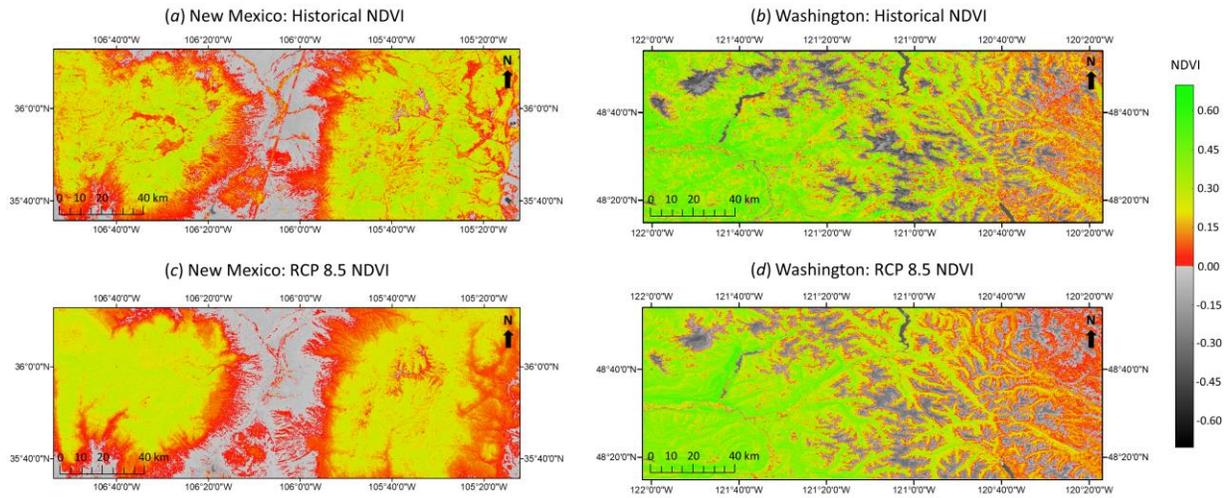


Figure 7: Comparison of the NDVI values between the historical (RF model trained) and future (RF model predicted for RCP 8.5 scenario) images for the two study sites.