**Deep learning classification of cheatgrass invasion using biophysical and remote sensing data**

Kyle B. Larson\*, Aaron R. Tuor

*Pacific Northwest National Laboratory, Richland, Washington, United States*

*\*Corresponding author: Tel: +1 509 371 7207, E-mail address: kyle.larson@pnnl.gov*

Contents

[Keywords: 3](#_Toc533166979)

[Abstract [KL] 3](#_Toc533166980)

[1. Introduction [KL] 3](#_Toc533166981)

[2. Cheatgrass Invasion [KL] 4](#_Toc533166982)

[3. Datasets [KL] 5](#_Toc533166983)

[3.1. Field Observations [KL] 6](#_Toc533166984)

[3.2. Satellite Imagery [KL] 7](#_Toc533166985)

[3.3. Biophysical & Ancillary Data [KL] 8](#_Toc533166986)

[3.4. Variable Partitioning [KL] 9](#_Toc533166987)

[4. Modeling Approaches 9](#_Toc533166988)

[4.1. Logistic Regression 10](#_Toc533166989)

[4.2. Random Forest [AT] 10](#_Toc533166990)

[4.3. Deep Learning Models [AT] 10](#_Toc533166991)

[4.4. Parameter Optimization and Training 13](#_Toc533166992)

[5. Results 14](#_Toc533166996)

[5.1. Comparison of Model Types and Partitioning [AT/KL] 14](#_Toc533166997)

[5.2. Geographic Interpolation and Assessment [KL] 15](#_Toc533166999)

[6. Discussion 16](#_Toc533167001)

[7. Summary and Conclusions [KL/AT] 17](#_Toc533167002)

[8. Acknowledgements [KL/AT] 17](#_Toc533167003)

[9. References [KL] 17](#_Toc533167004)

[List of Tables 17](#_Toc533167005)

[List of Figures 18](#_Toc533167006)

# Keywords:

[TBD]

# Abstract [KL]

[TBD]

# Introduction

Machine learning methods have become increasingly popular for remote sensing classification due to their ability to model complex class signatures, accept a variety of input data, and often outperform traditionally used parametric methods. Some of the more commonly used machine learning methods in remote sensing include support vector machines (SVMs), Random Forests (RFs), single decision trees (DTs), boosted DTs, k-nearest neighbor (k-NN), and artificial neural networks (ANNs). While there has been considerable use and comparison of machine learning methods for remote sensing classification, there appears to be little consensus on whether one is more superior than the others. A number of excellent reviews of these machine learning methods in remote sensing exist, which we defer to for readers seeking more in-depth information and guidance on those specific methods. Here, we focus on the use and performance of a rapidly growing class of machine learning methods called “deep learning” for classification in remote sensing, with an emphasis on providing an applied perspective for the domains of remote sensing and ecological modeling. [Refs: Maxwell et al. 2018; Maxwell et al. 2015; Crisci et al. 2012; Lu & Weng 2007; Mountrakis et al. 2011; Rodriguez-Galiano et al. 2012; Rogan et al. 2008; Shao et al. 2012].

Deep learning methods are widely used in fields such as computer vision, speech recognition, natural language processing, audio recognition, and bioinformatics, and are emerging in their use in remote sensing. Their connotation as “deep” networks is derived from them generally containing many compositions of non-linear data transformation which enables them to learn multiple levels of representation (or abstraction) in a hierarchical manner. Their ability to produce highly discriminative feature representations from low-level input data using either supervised, unsupervised, or semi-supervised techniques makes them an attractive tool for many remote sensing tasks, including image preprocessing, classification, and target recognition. Also, the ability to derive complex relationships from a variety of data makes deep learning methods particularly useful for cross-cutting applications in remote sensing and other fields of research. [Refs: Chen et al. 2014; Kussul et al. 2017; Yu et al. 2017; Zhang et al. 2016; Zhao et al. 2015]

* General types of DL architectures (DNN, RNN, CNN) and examples of their use in remote sensing
  + Many variants of deep learning models have been developed although they are generally derived from several basic architectures including deep neural networks (DNNs), recurrent neural networks (RNNs), and convolutional neural networks (CNNs). Collectively, these architectures offer a high degree of modeling flexibility for assessing spectral-spatial relationships.
* DL use in ecological modeling
  + Very few examples of DL
  + Examples of other ML methods
  + Refs: Cutler et al. 2007; Franklin 1995
* Challenges to using DL
  + Overfitting
  + Compute resources and time
  + Not good for evaluating relative importance of predictor variables

Here we explore the use of two deep learning approaches (DNN and RNN) for conducting predictive mapping of an invasive exotic grass, cheatgrass (*Bromus tectorum*), over a very large and ecologically diverse region in the western United States. We also benchmark our neural network models to analogously structured linear logistic regression and Random Forest models to provide further context about their predictive ability. A comparative approach is taken with variable selection by using different suites of biophysical and spectral-spatial predictor variables to assess modeling approaches that combine ecological niche partitioning and spectral-spatial remote sensing. As part of this process, spectral data from two of NASA’s land observation satellites (Moderate Resolution Imaging Spectroradiometer, or MODIS, and Landsat-7 ETM+) are also compared and combined to evaluate tradeoffs of using single or multi-sensor datasets. We chose these sensors because they have complimentary characteristics regarding spatial resolution, spectral bandwidth, and acquisition frequency, and have been used in other studies to map cheatgrass.

# Cheatgrass Invasion

Introduced in the late 19th-century, cheatgrass is now found in every state in the contiguous United States. Nowhere has its invasion been more prevalent than in western states where it has become a dominant component in many shrubland and grassland ecosystems. It is now estimated to dominate at least 40,000-km2 in the states of Nevada and Utah alone. Cheatgrass invasion poses a variety of threats to ecosystem function, rangeland health, and human safety. A central thread in many of these threats is the increase in fine fuels associated with cheatgrass, which can lead to increased fire frequency and irreversible loss of native vegetation and wildlife habitat. Following a fire, cheatgrass is able to more effectively compete with native vegetation, giving rise to a feedback cycle of further invasion and fire. [Refs: Mack 1981; Pellant 1996; Bradley & Mustard 2005; USDA Plant Database]

While cheatgrass is considered ubiquitous throughout much of the western U.S., detailed spatial information about its presence and abundance are still lacking for much of its range within this region. Previous efforts to map cheatgrass have focused on core areas of invasion such as the Great Basin and Snake River Plain. These efforts help paint a clearer picture of cheatgrass invasion in the western U.S., but their disparate nature – due to differing extents, methods, data, and focal periods – inhibits their reconciliation to inform range-wide management decisions. Conversely, mapping cheatgrass (or other vegetation) at a range-wide scale is also not straightforward or without its challenges; most notably the need for a greater volume of training and input predictor data to adequately model the phenomenon across many diverse ecological gradients. [Refs: ]

A variety of methods to map cheatgrass are described in the literature, ranging from imagery-driven methods that focus on spectral signatures or phenological indicators of cheatgrass to ecological niche partitioning approaches aimed at mapping biophysical and climatic conditions of where cheatgrass is known to occur. Much attention has been given to deriving phenological indicators of cheatgrass presence from overhead imagery (typically using spectral indices such as the Normalized Difference Vegetation Index, or NDVI) because its life cycle differs from many of the native plant species in its western range. Cheatgrass is a winter annual that may begin growth in the late fall and senesce in late spring, whereas many native plants begin growing in mid to late spring and continue growth through summer under favorable precipitation conditions. Thus, cheatgrass presence, and abundance to some degree, can be identified indirectly by comparing pixel-chronologies of NDVI, particularly in years when winter or early spring precipitation is above normal [Refs: Rice et al. 1992; Loik 2007; Bradley 2009;].

However, phenological differences between cheatgrass and non-target vegetation may be subtle and difficult to detect in most years under normal conditions. The phenological and spectral separability of cheatgrass from other vegetation is also strongly dependent on sub-pixel vegetation composition and structure. Hence, some remote sensing studies of cheatgrass have elected to use sensor platforms with high frequency of return and coarse spatial resolution (e.g., MODIS, AVHRR, VIIRS) to better capture within-season variation of cheatgrass growth, whereas other studies have chosen platforms with higher spatial resolution but less-frequent return cycles (e.g., Landsat-7, Landsat-8) to improve sub-pixel ratios of cheatgrass and non-target surfaces. The weight given to these tradeoffs is usually dependent on the study objective, as well as amount of available data, time, and resources, and results in selection of one or the other type of remote sensing platform. Here, we attempt to illustrate how deep learning can help overcome some of the difficulty in this decision process and enable more effective use of multiple platforms that offer complimentary advantages for mapping cheatgrass.

# Datasets

Three categories of data were used in this study: field observations of cheatgrass cover, time series of MODIS and Landsat-7 data, and spatial data of various biophysical parameters. Field data were acquired from a previous study (unpublished) of cheatgrass in the historic range of sage-grouse. The areal extent of the historic range of sage-grouse covers more than 1.25-e6 km2 spanning much of the western U.S and portions of southern Alberta and British Columbia, Canada. Data for this study was limited to the U.S. portion of the sage-grouse range, which covers approximately 1.17-e6 km2 (Figure 1). This area is ecologically diverse, spanning 23 Level-3 ecoregions (Omernik and Griffith 2014).

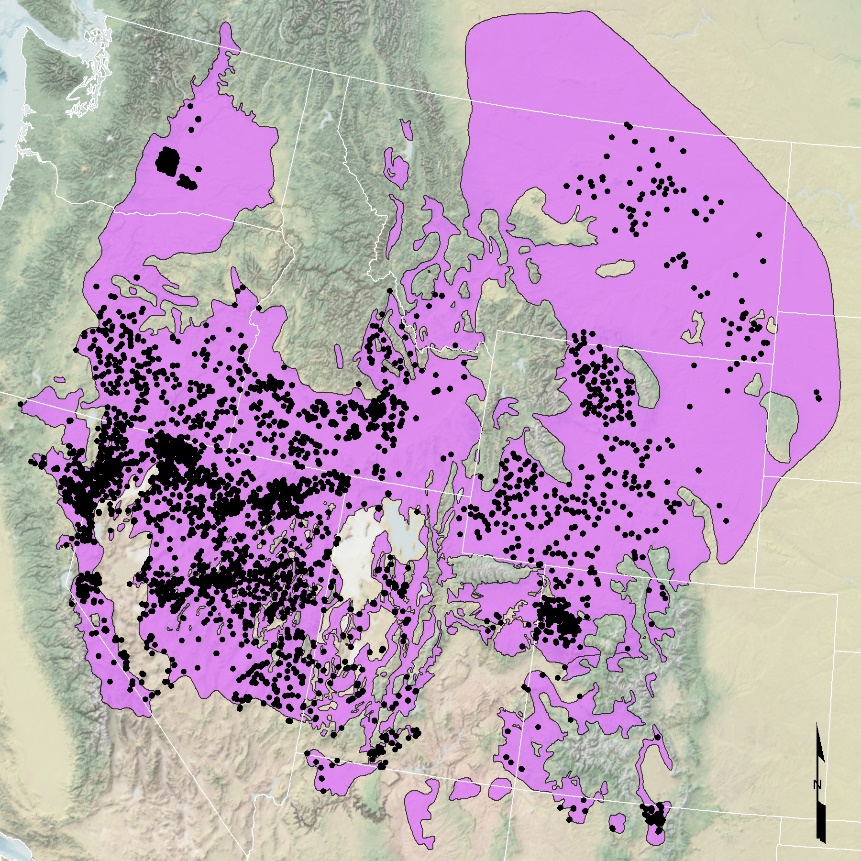


Figure 1. Location of field samples and extent of the historic range of sage-grouse that served as the geographic domain for this study.

### Field Observations

Field observations were solicited from multiple internal and external sources that collected vegetation measurements for unrelated studies between 2001 and 2014. Of these, 6602 observations were deemed useful for our study after extensive review for geographic accuracy and overlap with our study area, completeness, and rigor of collection methods. The majority of observations were collected in 2011 and 2012. All field data were collected along transects ranging from 25-m to 100-m in length using point intercept or standardized frame techniques. A strong break was observed in the distribution of cheatgrass canopy cover values at approximately 2 percent canopy cover (Figure 2) which coincidentally represented approximately equal percentages of the data (48% and 52%); thus, we chose to map two classes of cheatgrass above and below this natural break (i.e., Class 1 = 0-2 percent cover, Class 2 = >2 percent cover). By selecting these classes we also avoid the assumption of true absence for sites where cheatgrass was not observed.

Although field observations were not collected in the same manner or period of time as is preferred for vegetation mapping studies, such an undertaking was not possible in this study due to the vast extent of the study area and limited resources. Combining field observations was appropriate for our objectives because it was important to maximize coverage of the potentially invaded portion of our study area. An assumption of this approach is that there was little to no change in cheatgrass occurrence at the field locations across time. Available literature and anecdotal evidence indicates that cheatgrass will persist for many years after invasion unless there is significant intervention to eliminate it; therefore, we are confident that our combined field data contains few false-positives due to disappearance of cheatgrass. We are also confident that proportion of field locations that were observed to have no cheatgrass but have since been invaded is extremely low because cheatgrass has been well-established throughout the study area since the mid-20th century and anecdotal evidence indicates that further invasion in the 17 years represented by the data used in this study has occurred primarily at local scales. In both cases, potential bias from these assumptions should be mitigated by our approach of including time-series satellite data and time-averaged biophysical data in our analyses.

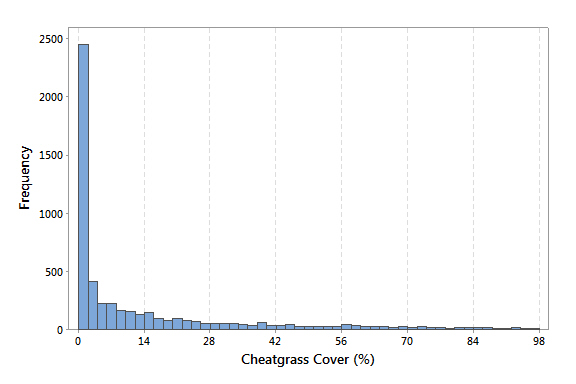


Figure 2. Distribution of cheatgrass cover at field sample locations.

### Satellite Imagery

Time series of both annual- and seasonal-composite MODIS (Terra) and Landsat-7 data were used in this study. Both sets of imagery were composited on a pixel-wise basis to reduce residual cloud and aerosol contamination and reflect the time of peak vegetation vigor as determined by maximum NDVI within the composite period (Didan 2015; Roy et al. 2010). These “maximum greenness” composites are valuable for identifying cheatgrass because its life cycle differs from most native vegetation. We used seasonal composite data for the Northern Hemisphere spring and summer periods (defined here as March-May and June-August, respectively) which correspond to distinguishing periods in the life cycle of cheatgrass of peak productivity in the spring and senescence in the summer.

The time series of satellite data spanned the entire period that field data was collected (2001-2014), except for Landsat-7 the data source we used offered data for 2003-2012 (Roy et al. 2010). We initially used MODIS 16-day composite data (MOD13Q1 MODIS Terra 16-day Vegetation Index product; Didan 2015) for the same time series, but later included data for 2001-2002 and 2015-2016 because it was found to improve results. MODIS 16-day composite data was obtained through Google EarthEngine, which we used to create annual and seasonal maximum NDVI composites and to down-sample the spatial resolution of MODIS to that of Landsat-7 (~30 meters). Spectral bands for each sensor that were used in this study are described in Table 1. For each MODIS and Landsat-7 composite time series we also derived a ΔNDVI grid for each year that represents the pixel-wise difference between NDVI and the long-term median NDVI based on all data in the time series. This metric has been found in previous investigations (unpublished) to be a useful indicator of cheatgrass infestation as such areas tend to exhibit greater inter-annual phenologic variability than native vegetation.

Table 1. MODIS and Landsat-7 bands included in this study.

|  |  |
| --- | --- |
| **Sensor** | **Band Name** |
| MODIS Terra | NDVI, EVI, Blue, Red, Near-infrared, Mid-infrared, Delta-NDVI\*, Median NDVI\* |
| Landsat-7 | Blue, Green, Red, Near-infrared, Shortwave infrared 1, Low-gain thermal, High-gain thermal, Shortwave Infrared 2, Delta-NDVI\*, Median NDVI\* |
| \* Derived band not included with original data. | |

### Biophysical & Ancillary Data

Six types of biophysical spatial data relevant to describing the ecological niche of cheatgrass at a regional scale were included in the study: soil moisture-temperature regimes, existing vegetation types, elevation, potential relative radiation (PRR), growing degree days (GDD), and climatic conditions (Table 2). Two ancillary datasets depicting ecoregions and geographic location (gridded latitude and longitude) were also incorporated to allow for more complex, multi-scale interactions among biophysical variables that likely occur throughout the study area.

The two categorical biophysical variables – soil moisture-temperature regimes and existing vegetation type – were selected because they are useful indicators of biophysical conditions that affect plant community dynamics. Soil moisture-temperature regimes also have been found to influence ecosystem resilience and resistance to invasive annual grasses (citations). We used available soils data derived from national soils databases (SSURGO and STATSGO2) by citation, and derived accompanying data from the same databases for remaining portions of our study area. Vegetation data was derived from the national-scale LANDFIRE Existing Vegetation Type (EVT) dataset, which we generalized into broader plant community associations that were more appropriate for the scale of our analysis and geographic representation of our field observations.

Continuous biophysical variables included PRR, GDD, elevation, and climatic conditions. PRR is a unitless index of available solar radiation for photosynthetic activity that is useful for landscape-scale vegetation analysis (Pierce et al. 2005, others). Similarly, GDD is an index that represents the relative period of time when temperatures are suitable for plant growth (citation). We calculated PRR as described by Pierce et al. (2005) for the general growing season of cheatgrass across the study area. GDD was calculated using 1-km gridded daily temperature data as:

(1)

where *TMAX* and *TMIN*  are the maximum and minimum daily temperature and *TBASE* is the minimum temperature for growth (assumed to be 0°C for cheatgrass), and if then . In both cases, PRR and GDD were calculated for the approximate growing season of cheatgrass throughout the study area (October to June).

The largest category of biophysical data in this study was a suite of 4-km gridded climatic datasets depicting monthly and annual 30-year norms for minimum and maximum temperature, and precipitation (Table 2). From these data we also derived five seasonal climatic datasets that correspond to important periods during the growth season of cheatgrass: cumulative winter (December-February) precipitation, cumulative spring (April-May) precipitation, cumulative summer (July-August) precipitation, winter (November-February) minimum and maximum temperature. Seasonal groupings were determined based on expert knowledge and exploratory analysis of cheatgrass occurrence from field data and climatic variables.

### Variable Selection

We selected four subsets of predictor variables that represent two generic approaches to mapping cheatgrass: one approach founded on ecological niche modeling, and the other founded on combining ecological-spectral modeling approaches. The purpose of using these subsets was twofold: to evaluate potential gains and losses in classification accuracy by combining ecological and spectral data; and to evaluate gains and losses in classification accuracy by combining data from multiple sensors. For the purpose of defining variable subsets, let be the complete set of predictor variables for location in the study area, and be a function that selects a defined subset of variables from . The four subsets of variables tested are further described in Table 3.

Table 2. Summary of biophysical and spectral predictor variables used in the study. Categorical and continuous variables are denoted as *c* and *x*, respectively.

|  |  |  |
| --- | --- | --- |
| **Variable** | **Dataset** | **Source** |
| *c1* | Soil temperature and moisture regime classes | Chambers et al. 2014; SSURGO; STATSGO |
| *c2* | Generalized vegetation cover type | LANDFIRE |
| *c3* | EPA Level III Ecoregions | U.S. EPA 2013 |
| *x1:2* | 30-m gridded latitude & longitude | Google EarthEngine |
| *x3* | Elevation above mean sea level | USGS 2018 |
| *x4* | Potential Relative Radiation | USGS 2018; Pierce et al. 2005 |
| *x5* | Median winter precipitation | PRISM Climate Group |
| *x6* | Median growing degree days | Thornton et al. 2014 |
| *x7:50* | 30-year normal climatic conditions | PRISM Climate Group |
| *x51:323* | Landsat-7 bands 1-10 (annual, spring, summer) | Roy et al. 2010 |
| *x324:664* | MODIS bands 1-8 (annual, spring, summer) | Didan 2015 |

Table 3. Subsets of input predictor variables used in all models.

|  |  |  |
| --- | --- | --- |
| **Subset Function** | ***Selected Variables*** | ***Description*** |
|  | *x1:50* | Biophysical and ancillary variables |
|  | *x1:50,326:664* | Biophysical, ancillary, and MODIS variables |
|  | *x1:325* | Biophysical, ancillary, and Landsat variables |
|  | *x1:664* | All variables (Biophysical, ancillary MODIS, Landsat) |

# Modeling Approaches

We assessed four machine learning methods for predictive mapping of cheatgrass occurrence: Random Forest (RF), Logistic Regression (LR), Deep Neural Networks (DNN), and Recurrent Neural Networks (RNN). The classification objective for all models was to map two classes of cheatgrass based the evident separation observed in the field data of 0-2 percent and >2 percent canopy cover. All continuous predictor variables were standardized by subtracting their respective mean and dividing by the standard deviation. The LR, DNN, and RNN models were implemented in Python using the Tensorflow library (1.2.1) and RF models were implemented in Python using the scikit-learn library (0.18.2).

### Categorical Variables

The input to our machine learning models contains three values from categorical variables, soil temperature and moisture regime (, generalized vegetation cover type and EPA Level III ecoregions (); (Table 2). Because the native inputs to our machine learning algorithms consist of real valued vectors we employ two simple strategies for incorporating vector representations of these predictive variables: one-hot vectors for RF, and embedding vectors for LR, DNN, and RNN. For a categorical variable with possible classes, for each class, we define a one-hot vector where:

|  |  |  |
| --- | --- | --- |
|  |  | (2) |

We define as the one hot vectors associated with the categorical values for a location.

Embedding vectors are a parametrized vector representation of categorical values. The values of embedding vectors are learned jointly with the other parameters of the LR, DNN, and RNN models. We define three embedding matrices for the categorical variables, , , and where the second dimension corresponds to the number of classes for a given categorical variable and is the size of the embedding vector. We define as the embedding vectors associated with the categorical values for a location.

### Random Forest

We selected RF for comparison to DNN and RNN methods because it has been shown to perform well for predictive vegetation mapping (citations), be resilient to overfitting (citations), and provide competitive results compared to deep learning models in low resource data regimes (citations). RF classification is an ensemble method where a set of decision trees is trained on subsets of the data and a classification decision is obtained from a vote among the collection of decision tree predictions. Trees are fit using variants of standard decision tree algorithms such as CART, but randomness is introduced in two ways: a different subset of the training data is randomly sampled when building each tree, and random sets of features are used for making decisions at each node of the tree.

We experimented with various parameter configurations for RF to improve model performance. Key parameters we manipulated (and their final values) were as follows: sampling method (with and without replacement), criterion for splitting nodes (GINI index), maximum number of features (square root of the total number of features), minimum number of samples in a leaf node ({1, 2, 4}), minimum number of samples in a split node ({2, 5, 10}), maximum depth of a tree ([10, 110]), and the number of decision trees in the forest (10-200).

### Logistic Regression

A common approach to machine learning classification is to use expert knowledge to craft discriminative features for a linear model from a set of raw predictor variables. One example of this type of feature is the NDVI index derived from spectral bands. Discovering highly discriminative features in this manner is likely to take substantive time and effort; the space of possible discriminative features is beyond exponential according to the dimension of the original input.

A principal concern of this work is to assess the capability of deep learning methods to automatically generate useful features from high dimensional geospatial and biophysical predictor variables. Proven effective, these methods have the potential to completely sidestep the feature engineering stage necessary to create a highly predictive linear model. To provide perspective on the value of deep learning for deriving highly predictive feature representations for ecological modeling, we explore as a baseline linear LR models based directly on the standardized geospatial and biophysical predictor variables.

With *n*  as the number of continuous variables in the data subset and the size of an embedding vector, the input for a given sample to our LR model is a vector composed of standardized real values of the continuous predictor variables and vector representations of the categorical variables described in Section (4.1).

(3)

is one of the subset functions defined in Table 3. The output of the LR model is a vectorcontaining a predicted probability distribution over the two classes of cheatgrass canopy cover for a location.

(4)

Where the matrix and vector are learned parameters and *softmax* is defined as the elementwise vector valued function:

(5)

We fit the parameters of the logistic regression model by minimizing the cross-entropy between the ground truth and predictions with the ADAM (cite) optimization algorithm[[1]](#footnote-2). The only hyper-parameter that we tuned is the weight on L2 regularization.

### Deep Learning Models

The deep learning models used in this study can be viewed as a composition of successive “hidden” non-linear transformations of features that are mapped to a penultimate discriminative latent feature representation of the original input. This representation is then used as input to a linear LR model for prediction of cheatgrass classes. As a universal function approximator a DNN has the capability to derive highly discriminative features that represent complex relationships between input variables in a high dimensional space. RNN architectures are a powerful extension to DNNs which provide recurrent connections between nodes. These connections allow them to effectively model temporal dynamics. We include time-series satellite input data in both model types but treat it as a flat vector in the DNN models and as sequences of vectors in the hierarchical RNN models. This approach allowed us to establish parity between the models to compare their performance.

#### Deep Neural Network

The input to the DNN model is the same as defined for the LR model. The architecture of the DNN model (Figure 3) consists of *L* hidden layers *h* which are recursively defined as:

(6)

where the RELU (Rectified Linear Unit; citation) operation is defined as the elementwise vector valued operation:

(7)

The output of the DNN is then directed to the aforementioned linear LR classifier. Dropout normalization is also used to avoid overfitting and reduce generalization error (Srivastava et al. 2014) and batch normalization is used to get better conditioned gradient updates (Ioffe & Szegedy 2015). Together, these techniques help to stabilize learning during training.

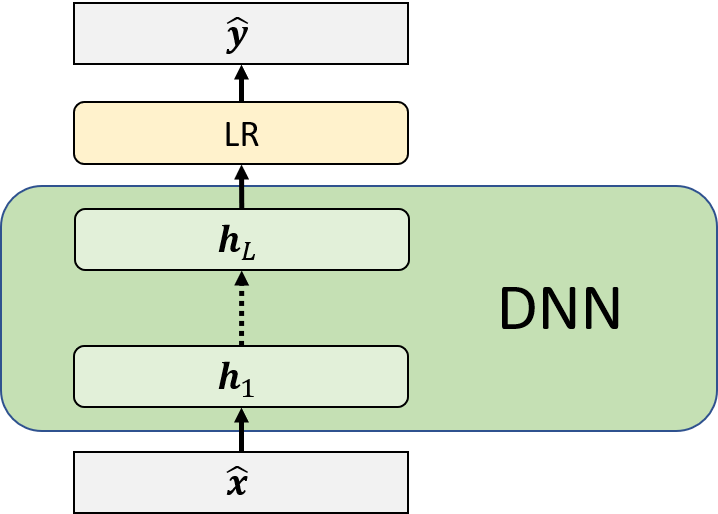


Figure 3. Computational graph depiction of a Deep Neural network as a feature extractor for a logistic regression model. The input vector is fed to the DNN model. The output from the DNN model, **,**  is used as a set of derived features for the location which are classified by the LR classifier.

#### Hierarchical Recurrent Neural Network

As an extension to the DNN, we employed a hierarchical composition of bidirectional RNNs (HRNN) with Long Short Term Memory (LSTM) to predict cheatgrass occurrence. Bi-directional RNNs have been proven effective for modeling difficult time series and operate by processing sequence data in both directions, allowing output nodes to get signals from both previous and future time steps (Schuster & Paliwal 1997). Use of LSTM in an RNN can reduce training difficulties and also improve the RNN’s ability to model long-term dependencies (Hochreiter & Schmidhuber 1997). LSTM’s capability to track discriminative values over arbitrary time intervals is especially useful if there are response lags of unknown duration between dependent events in a time series. Such is the case in this study as the timing and magnitude of cheatgrass growth (and subsequently its detectable presence in time-series satellite data) may be accelerated or lagged depending on climatic conditions across and within years.

We define two bidirectional LSTM networks for the MODIS and Landsat-7 time-series datasets respectively as and . The LSTMs provide condensed vector representations of the respective satellite datasets given a sequence of vectors of annual and seasonal composite data for a given pixel. Let be the vector of annual (*ann*) composite pixel values of the *t*-th year for the spectral bands of the LandSat imagery. Similarly defined are and for the spring (*spr*) and summer (*sum*) composite image products. The same notation is used for the MODIS data; thus, the *t*-th vector ( corresponds to year) in the Landsat-7 (10 spectral bands) and MODIS (8 spectral bands) time series are defined as:

(8)

(9)

The LSTM vectors are concatenated with the categorical embeddings and the continuous biophysical variables [] and used as input to a DNN as described in (Eq. 3,4) but with as:

(10)

The parameters of the HRNN and DNN models are fit with the same optimization algorithm, objective function, and regularization strategies as described for the LR model. The HRNN model is depicted in Figure 4.

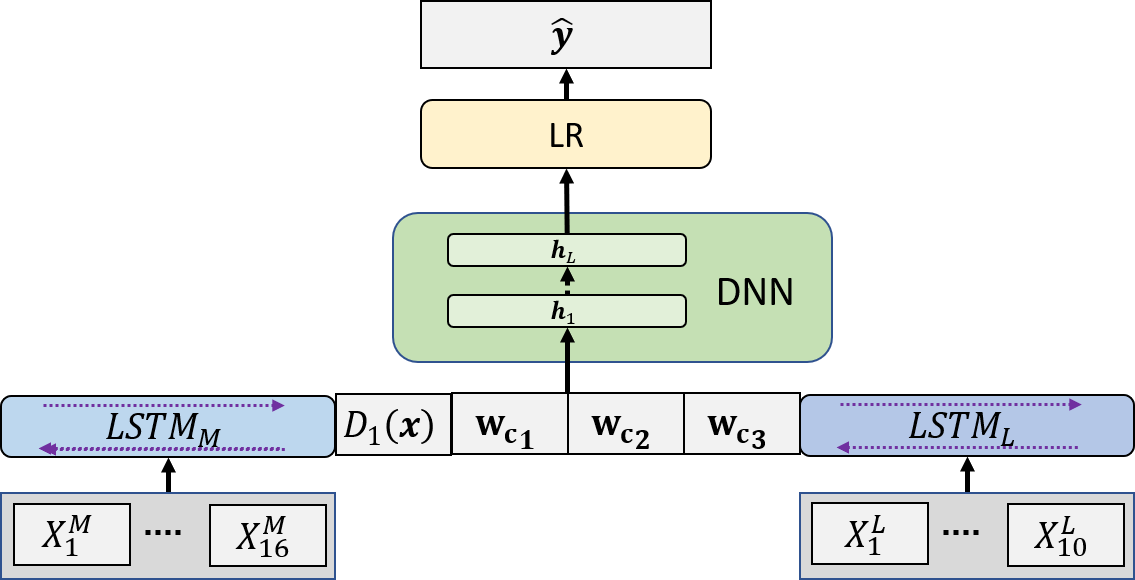


Figure 4. Computational graph depiction of the HRNN model. The MODIS and LandSat branches of the HRNN architecture extract features from the imagery sequences that are used in conjunction with biophysical variables as input to a DNN. The DNN then further refines features as input to the linear LR model.

### Parameter Optimization and Training

We experimented with various configurations of the DNN and HRNN as well as with different data sampling strategies to identify best-performing models for each approach. Final model configurations were chosen by assessing model performance for 200 random configurations of hyperparameters. Key hyperparameters (and their ranges) that we randomly manipulated were as follows: number of nodes per hidden layer (32-512 incremented by powers of 2), dropout rate (0-0.5), learn rate (0.001-0.03), and mini-batch size (16-128 incremented by powers of 2).

We also evaluated several stratified random sampling strategies for splitting data into training, validation, and test sets because our field observations were obtained from other efforts spanning different years and portions of our study area. We did so by assessing the variance of model performance across random permutations of each sampling method to determine which method provided more stable modeling outcomes. The most consistent model performance was obtained using 5-fold cross-validation where random splits (80% training, 20% validation) were stratified using equal joint distributions of ecoregion and generalized land cover classification.

# Results

Here we describe the performance of the four model types (LR, RF, DNN, HRNN) and four variable subsets observed during model development and validation, followed by results of post-modelling efforts to generate a predictive map of cheatgrass occurrence based on the best combination of model and variable partition. We used common metrics for assessing the performance of binary classifiers including unbalanced accuracy, precision, recall, *F1* score (harmonic mean of precision and recall), and receiver operating characteristic (ROC) curves. Standard one-side and two-side t-tests (citation) were also used to assess the statistical significance (α = 0.05) of performance gains among variable partitions and model types.

### Comparison of Model Types and Partitioning

All models exhibited similar likelihoods of misclassifying cheatgrass as present (greater than 2% canopy cover) when it was not (Type I error) or vice versa (Type II error), indicating that our selection of classes and use of unbalanced accuracies to compare models were appropriate. Classification accuracy among all model types and variable subset tested ranged from 76.7% for LR with the lowest dimension subset representing biophysical and ancillary data (*D1*), to 83.2% using DNN with the highest dimension subset containing biophysical, ancillary, MODIS, and Landsat-7 data (*D4*) (Table 4). All three machine learning methods (RF, DNN, RNN) outperformed LR benchmark models across all subsets, although the magnitude of this difference became less noticeable when satellite data was included. Machine learning methods were competitive amongst each other, exhibiting similar accuracy, balance between precision and recall (as exhibited by *F1* scores; Table 4), and ROCs (Figure 5).

Table 4. Summary of classification accuracy and *F1* scores across model types and variable partitions.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Partition | *D1* | | *D2* | | *D3* | | *D4* | |
| Model | *Acc.(%)* | *F1* | *Acc.(%)* | *F1* | *Acc.(%)* | *F1* | *Acc.(%)* | *F1* |
| LR | 76.72 | 0.768 | 79.62 | 0.808 | 81.19 | 0.822 | 81.44 | 0.822 |
| RF | 80.17 | 0.811 | 81.68 | 0.824 | 82.50 | 0.833 | 82.80 | 0.835 |
| DNN | 79.60 | 0.808 | 82.00 | 0.823 | 82.46 | 0.832 | 83.20 | 0.839 |
| RNN | -- | -- | 82.05 | 0.828 | 82.72 | 0.835 | 83.18 | 0.840 |

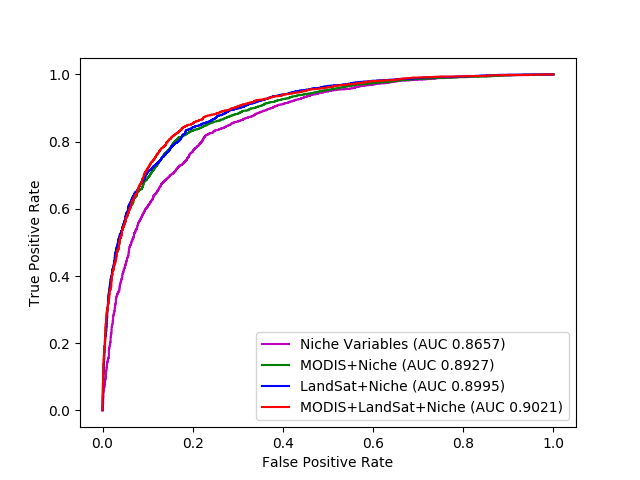


Figure 5. Receiver Operating Characteristic (ROC) curves of LR, RF, DNN and RNN models with *D4* partition (left), and ROC curves of all partitions for DNN (right).

Classification accuracy of all four model types significantly improved with the addition of MODIS (*D2*; p < 0.05) or Landsat-7 (*D3*; p < 0.01) compared to using only biophysical and ancillary data (*D1*) (Figure 5). The use of Landsat-7 appeared to improve accuracy for all model types compared to MODIS, although this difference was not statistically significant for the neural networks suggesting there was no obvious advantage of using one sensor versus the other in our modeling framework. All model types achieved their best classification results when both sensors were used (*D4*).

Because neural networks require training and the amount of training is at the discretion of the analyst, it is useful to compare trends in their performance during training and validation to better understand their ability to generalize a prediction. Inspection of cross-entropy loss during training and validation, which provides a measure of how closely the predicted probability diverges from the actual label, further illustrates that including MODIS and/or Landsat-7 data with biophysical and ancillary data (i.e., partitions *D2*, *D3*, *D4*) improves the neural networks’ (RNN not shown) ability to predict cheatgrass occurrence compared to using only biophysical data (Figure 6). The use of Landsat-7 appeared to have a better regularizing effect on cross-entropy loss compared to MODIS, allowing it minimize overfitting and converge toward a solution more quickly. These trends may be related to obvious differences between the two sensors that affect signal noise and specificity (e.g., spectral bandwidth, spatiotemporal resolution) as well as differences in the size of the networks (i.e., number of variables used). The variable partition with MODIS (D2) included X more variables than the partition with Landsat-7 (D3), which could increase chances of overfitting. However, we considered this an inherent aspect of the comparison because we were interested in including all of their respective information and assessing model performance.

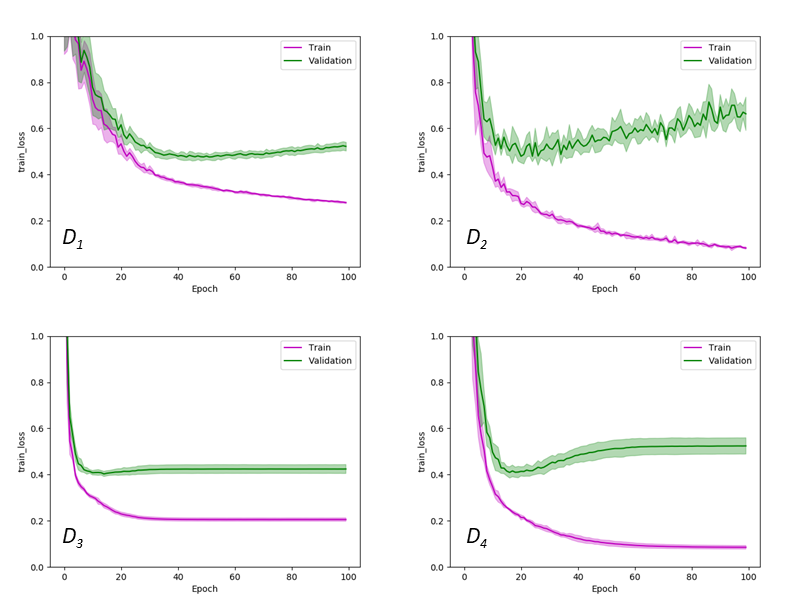


Figure 6. Cross-entropy loss during DNN training and validation for D1, D2, D3, and D4 variable partitions.

### Geographic Interpolation and Assessment

We selected our best performing modeling configuration (DNN with variable partition *D4*) to conduct predictive mapping of cheatgrass occurrence and to summarize biogeographic patterns of cheatgrass invasion throughout the historic range of sage-grouse. Mapping was performed by combining all input predictor datasets into a three-dimensional matrix and recursively applying each *k*-fold trained model to the vector of predictor values at each pixel location to produce a spatial grid of the probability (*p*) of >2 percent cheatgrass occurrence. The resultant distribution of probabilities was strongly bimodal so we used a simple decision rule of *p* ≥ 0.5 to produce thematic maps of cheatgrass occurrence for each *k*-fold outcome.

The k-fold models can be considered complimentary to one another because each was trained with different subsets of the same set of field observations; therefore, we used an ensemble approach to combine them based on their pixel-wise consensus. This method results in representative abstractions of all k-fold models that can be interpreted as ranging from least (1 out of k) to most conservative (k out of k). We determined the best consensus by evaluating balance between precision and recall (as determined by *F1* score) based on mapped predictions at all field observation locations. This was done strictly to determine the best geographic interpolation of the k-fold models, and we refer to the aforementioned results in 5.1 as the appropriate measures of model performance. We lacked enough field observations to have sufficiently sized (and geographically dispersed) independent sample sets for model development, validation, and map assessment, so we prioritized doing so for the first two tasks.

Accuracy and balance between precision and recall of the least (1 of 5) and most (5 of 5) conservative consensus maps (Table 5) was similar to that measured for DNN model validation (Table 4). Classification precision and recall were most evenly balanced at the mid-consensus mapping level (3 of 5), so we selected this level for subsequent evaluation and interpretation of cheatgrass invasion (Figure 7). Map accuracy was also maximized at this level and exceeded that from model validation by 7.6 percent.

Table 5. Summary of post-modeling consensus mapping accuracy, precision, recall, and *F1* scores.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Consensus | Accuracy | Precision | Recall | F-score |
| 1 of 5 | 0.842 | 0.792 | 0.947 | 0.863 |
| 2 of 5 | 0.893 | 0.878 | 0.924 | 0.900 |
| 3 of 5 | 0.908 | 0.923 | 0.899 | 0.911 |
| 4 of 5 | 0.891 | 0.946 | 0.840 | 0.890 |
| 5 of 5 | 0.832 | 0.969 | 0.701 | 0.814 |

Figure 7. Map of predicted cheatgrass occurrence based on best k-fold spatial consensus of DNN-D4 model combination.

# Discussion

We benchmarked the efficacy of two fundamental deep learning techniques (DNN and RNN) to their respective output classification function (linear LR) and to a more mature machine learning method (RF) for the purpose of predictive mapping of cheatgrass invasion in the Western U.S. Four suites of predictor variables were also compared to assess different data fusion strategies for modeling cheatgrass occurrence. The three machine learning methods tested (DNN, RNN, RF) provided more accurate predictions of cheatgrass occurrence compared to LR, although this result was expected given the presumed advantage these models have due to training. RF was found to be closely competitive with both deep learning techniques, but the latter produced slighter better accuracies with variable partitions that combined biophysical and spectra-spatial data. [other examples from literature]

RNN provided similar predictive ability as its DNN counterpart despite being deliberately designed to leverage time-series information. One hypothesis for this is that the temporal frequency and composite nature of satellite data (annual- and seasonal-composites) and climatic data (30-year annual and monthly norms) used in this study may be too coarse to obtain better performance with RNN. Higher frequency time-series data may be beneficial to our RNN models because cheatgrass growth can be temporally dynamic both within and among growing seasons. However, doing so can also exacerbate key sources of error or missing data such as increased prevalence of cloud and aerosols contamination in imagery, and greater spatiotemporal variability of interpolated climate data. Cursory examination of higher frequency MODIS and Landsat-7 data suggested these factors would be problematic in this study due to the large geographic scale and multi-year time period.

Our choice of predictor variables was an important aspect of this study. With many traditional statistical approaches, variables are often selected through a quantitative process of elimination designed to reduce noise and correlation among variables. Machine learning algorithms effectively perform this step by adjusting the weight each variable has in the outcome, making them relatively robust to broad variable inclusion. A disadvantage, however, is that resulting determinations of each variable’s weight are not easy to derive and latent interpretation of the model is nearly impossible. Ablative analysis can be used in this regard to make qualitative comparisons of different model configurations to gauge their relative effect on model accuracy. We performed ablative analysis of different configurations of predictor variables and found that predictions of cheatgrass occurrence improved for all four model classes by combining biophysical and spectral-spatial variable partitions. Presumably, increasing the number and complexity of predictor variables would eventually have a limiting effect on accuracy gains, which we did observe with the variable partitions we assessed. Nonetheless, there was a clear benefit to predicting cheatgrass occurrence by fusing multi-modal biophysical and remote sensing data. We also found that model performance improved with inclusion of both MODIS Terra and Landsat-7 satellite data compared to the use of just one. This result was observed for all model types suggesting these sensors do provide complementary indicators of cheatgrass occurrence. Reasonable arguments could be made for choosing other sensors for this study, such as sensors that enhance temporal frequency of acquisitions to better capture phenological phenomena of cheatgrass and non-target vegetation. Here, we demonstrate one possible combination and reiterate the importance of choosing sensors based on the perceived advantages and disadvantages they offer relative to the application, regardless of model selection.

Obtaining sufficient numbers of field observations for model development, validation, and map assessment was a concern in this study, especially due to the large extent and ecological diversity of the study area. Generally, deep learning methods require very large labeled datasets. Fulfilling this requirement may be difficult for vegetation mapping efforts, especially if new field campaigns are required. This is less of a concern for studies that do not require new field data and can assimilate like-data from multiple sources, as was done in this study. We assimilated more than 24,000 field observations from various sources, but approximately 25% were deemed useful based on their geographic accuracy and overlap with our study area, completeness, and rigor of collection methods. Consequently, we were not able to obtain an independent set of observations for assessing the accuracy of mapped predictions, although we were able to thoroughly assess the accuracy of underlying models using *k*-fold cross-validation.

A caution to using k-fold cross-validation with deep learning or other neural network approaches is that it is not possible to create an ensemble model because each k-fold model contains potentially different hidden representations. This is problematic for predictive mapping applications because it is generally not useful to produce multiple maps of the same phenomenon. One solution is to spatially composite mapped predictions to create an ensemble representation. We used a straightforward approach by deriving the pixel-wise consensus among k-fold model predictions and evaluating which consensus level provided the best balance between classification precision and recall. Although we were not able to assess the accuracy of consensus maps with an independent set of observations, we believe the method provides an appropriate and useful means to depict cheatgrass occurrence because it utilizes ground-truth observations to identify an optimal balance in the spatial agreement among k-fold predictions. Additionally, the full range of consensus values can be visualized to identify areas where model uncertainty is greater. Many of the field observations used in this study were obtained from ongoing long-term environmental monitoring efforts; thus, we anticipate additional observations will be available in the future to reassess our mapped prediction of cheatgrass occurrence.

[Discussion of mapping results. Describe cheatgrass invasion across study area: core areas, most affected ecoregions, etc. How does this compare to what’s described in the literature?]

Results of consensus mapping suggest that the method provides a straightforward and representative spatial solution to the k-fold models.

# Summary and Conclusions

Modeling approaches that combine elements of ecological niche partitioning and remote sensing are attractive for mapping plant species or communities because it can be reasonably assumed that ecological linkages between target species/communities, their biophysical environment, and spatial-spectral signatures are discernable when appropriate information and data fusion techniques are used. In practice, these approaches can be difficult to implement due to uncertainties about the scope of predictor variables or most appropriate modeling framework to use, ability to satisfy underlying assumptions, challenges associated with assimilating and analyzing large multi-modal datasets, and limited availability of ground-truth data. Machine learning techniques, including a newer class of “deep learning” algorithms, are gaining attention in the remote sensing field because they are well-suited for complex modeling problems and have become increasingly accessible to the broader scientific community through open-source software.

The deep learning techniques proved to be effective for this purpose and show initial promise for data fusion in remote sensing applications. Likewise, RF was reconfirmed as a robust machine learning technique and in addition shown to be competitive with DNN and RNN models with identical predictor variable sets.

We were able to achieve strong predictive capability for mapping cheatgrass occurrence across a very large, ecologically diverse region.

# Acknowledgements

* Field data sources
* This work was funded by the Deep Learning for Scientific Discovery Laboratory Directed Research and Development investment at the Pacific Northwest National Laboratory. Pacific Northwest National Laboratory is operated for the U.S. Department of Energy by Battelle under Contract DE- AC05-76RL01830.

# References [KL]

## List of Tables

[Table 1. MODIS and Landsat-7 bands included in this study. 8](#_Toc533163901)

[Table 2. Summary of biophysical and spectral predictor variables used in the study. Categorical and continuous variables are denoted as *c* and *x*, respectively. 9](#_Toc533163902)

[Table 3. Partitions of input predictor variables used in all models. 9](#_Toc533163903)

[Table 4. Summary of classification accuracy and F1 scores across model types and variable partitions. 14](#_Toc533163904)

[Table 5. Summary of post-modeling consensus mapping accuracy, precision, recall, and *F1* scores. 15](#_Toc533163905)

## List of Figures

[Figure 1. Location of field samples and extent of the historic range of sage-grouse that served as the geographic domain for this study. 6](#_Toc533169850)

[Figure 2. Distribution of cheatgrass cover at field sample locations. 7](#_Toc533169851)

[Figure 3. Architecture of Deep Neural Network for Classification. 12](#_Toc533169852)

[Figure 4. Architecture of Recurrent Neural Network for Classification. 13](#_Toc533169853)

[Figure 5. Receiver Operating Characteristic (ROC) curves of LR, RF, DNN and RNN models with D4 partition (left), and ROC curves of all partitions for DNN (right). 14](#_Toc533169854)

[Figure 6. Cross-entropy loss during DNN training and validation for D1, D2, D3, and D4 variable partitions. 15](#_Toc533169855)

[Figure 7. Map of predicted cheatgrass occurrence based on best k-fold spatial consensus of DNN-D4 model combination. 16](#_Toc533169856)

Didan, K. (2015). MOD13Q1 MODIS/Terra Vegetation Indices 16-Day L3 Global 250 m SIN Grid V006. In N.E.L. DAAC (Ed.)

Omernik, J.M., & Griffith, G.E. (2014). Ecoregions of the Conterminous United States: Evolution of a Hierarchical Spatial Framework. *Environmental Management, 54*, 1249-1266

Roy, D.P., Ju, J., Kline, K., Scaramuzza, P.L., Kovalskyy, V., Hansen, M., Loveland, T.R., Vermote, E., & Zhang, C. (2010). Web-enabled Landsat Data (WELD): Landsat ETM+ composited mosaics of the conterminous United States. *Remote Sensing of Environment, 114*, 35-49

1. ADAM is a variant on stochastic gradient descent that adjusts the model parameters adaptively according to estimates of the second order moment of the gradient. [↑](#footnote-ref-2)