Modelling species distributions and environmental suitability highlights risk of plant invasions in western United States

Devin E. McMahon\textsuperscript{1,2} | Alexandra K. Urza\textsuperscript{1} | Jessi L. Brown\textsuperscript{3} | Conor Phelan\textsuperscript{4} | Jeanne C. Chambers\textsuperscript{1}

\textsuperscript{1}USDA Forest Service, Rocky Mountain Research Station, Reno, NV, USA
\textsuperscript{2}USDA Forest Service, Six Rivers National Forest, Eureka, CA, USA
\textsuperscript{3}Department of Biology, University of Nevada, Reno, NV, USA
\textsuperscript{4}Department of Natural Resources and Environmental Science, University of Nevada, Reno, NV, USA

Correspondence
Alexandra K. Urza, USDA Forest Service, Rocky Mountain Research Station, Reno, NV, USA.
Email: alexandra.urza@usda.gov

Funding information
US Joint Fire Science Program, Grant/Award Number: 19-2-02-11

Abstract

Aim: Non-native invasive plants impact ecosystems globally, and the distributions of many species are expanding. The current and potential distributions of many invaders have not been characterized at the broad scales needed for effective management. We modelled the distributions of 15 non-native invasive grass and forb species of concern in western North America to define their environmental niches and predict potential invasion risk.

Location: Arid and semi-arid western United States.

Methods: We modelled species distribution using presence/absence data from >148,000 vegetation survey plots to predict the probability of presence of each species based on associated climate, soil, topography and disturbance records. Boosted regression tree models were trained and tested within a buffer distance from presence observations of each species. Buffers were optimized by adding absence observations from increasingly large areas until prediction of presence was maximized, as defined by the area under the precision-recall curve (AUPRC). We evaluated model AUPRC and other metrics by cross-validation within training areas, then projected final models within focal EPA level III ecoregions.

Results: Focal species were present in 2.7% to 21% of plots within model training areas of 40 to 340 km radius. Models conservatively estimated potential species presence, with typically low false positive rates and moderate sensitivity. The most influential predictors of presence included minimum temperature (especially for grasses), climatic water deficit, precipitation seasonality and fire history. Cold desert ecoregions were impacted by the most species, with predicted range expansion into the prairie ecoregions and infilling for Warm Desert species.

Main conclusions: Invasive forb and grass species are likely to expand their ranges and continued increases in temperature, aridity and area burned will increase invasion risk. Monitoring species presence and absence and mapping known and potential ranges with a focus on presence detection, as in our methodology, will aid in identifying new invasions and prioritizing prevention and control.
1 | INTRODUCTION

Exotic plant invasions are damaging ecosystems around the world, changing ecosystem structure and function, reducing biodiversity, and harming social and economic systems (Mack et al., 2000; Pejchar & Mooney, 2009; Simberloff et al., 2013). Across the western United States, invasive plant species are compounding the ecosystem stresses of changing climate and fire regimes. Swaths of the sagebrush biome have converted from perennial- and shrub-dominated ecosystems to invasive annual plant dominance, severely impacting sagebrush-dependent species and ecosystem functions (Chambers et al., 2019; Coates et al., 2016; Germino et al., 2016). The annual grass Bromus tectorum has been extensively studied due to its ability to form monocultures, increase fire severity and thwart restoration (Balch et al., 2013; Bishop et al., 2019). However, much less is known about other non-native invaders in the region, which also have the capacity to impact ecological processes. Grasses including Bromus rubens, Schismus barbatus and Taeniatherum caput-medusae also alter fire behaviour, contributing to the development of grass-fire cycles that can lead to progressive loss of the native plant community (Fusco et al., 2019). Annual forbs such as Lactuca serriola, Tragopogon dubius, Sisymbrium altissimum and Salsola tragus can displace native vegetation, limit forage and habitat for animals, and facilitate annual grass establishment (Antill et al., 2012; Piemeisel, 1951; Prevéy et al., 2010).

Mitigating the impacts of plant invasions requires rapidly identifying and controlling local populations before they spread farther (Reaser et al., 2020), yet monitoring efforts may not detect non-native species until invasion is well underway. To anticipate future invasions, models of invasion risk based on the current distribution of non-native species have become increasingly common (Abella et al., 2012; Ibáñez et al., 2009). However, modelling suitable habitat for invading species poses unique challenges (Bradley, 2013). Newly arrived species have not yet explored the range of available environmental conditions and may have patchy and sparse distributions that reflect dispersal limitations (Jarnevich & Reynolds, 2011; Welk, 2004). Because species currently undergoing range expansion are not yet at climate equilibrium, presence data may not represent the full range of suitable habitat and absences may not necessarily represent unsuitable conditions (Uden et al., 2015). Describing habitat suitability thus requires a flexible modelling approach that uses broad-scale, comprehensive data on species occurrences and can distinguish informative and uninformative absences.

Multiple factors define an invasive species’ niche and determine whether it establishes and threatens native ecosystems once it arrives. Temperature restricts species’ ranges directly based on physiological thermal tolerances and interacts with precipitation to determine ecosystem water inputs (Allington et al., 2013; Bradford & Lauenroth, 2006). Precipitation seasonality influences the timing of available water and its storage across soil layers and thus has important effects on the distribution of plant species based on rooting habits (Lauenroth et al., 2014). At local scales, topography modifies climate conditions that promote or inhibit invasions (Chambers, Bradley, et al., 2014; Chambers, Miller, et al., 2014; Peeler & Smithwick, 2018). Soil properties influence availability of water and nutrients to invasive species (Brooks et al., 2016), and soil type and water availability are useful indicators of ecosystem resistance to invasion (Chambers et al., 2007; Roundy et al., 2018).

Disturbances can facilitate the arrival (Gill et al., 2018) or increase the abundance of non-native plant species (Jauni et al., 2015). Successful invaders often share traits that allow exploitation of habitat openings, such as short generation times, high fecundity and capacity for long-distance dispersal (Van Kleunen et al., 2010). Wildland fires suppress native species and create pulses of available resources, which can release invasive plants from competition (Gill et al., 2018; Hanna & Fulgham, 2015; Steers & Allen, 2010). Anthropogenic land use and infrastructure can also remove competing vegetation and increase resource availability, and the associated movement of equipment and animals provides a vector for the spread of seeds (Gelbard & Belnap, 2003; González-Moreno et al., 2014; Leu et al., 2008).

Our primary objective was to assess the current distributions and environmental suitability for non-native invasive plants and the relative risk of invasion across the arid and semi-arid western USA. We leveraged presence and absence observations from extensive, publicly available vegetation survey and monitoring data (BLM, 2020b; Pyke et al., 2011; USGS, 2012). We defined a model training region for each species using threshold distances from observed presences, in which absences were more likely to represent environments where species may have dispersed but did not establish and inclusion of absence data improved model performance (Barve et al., 2011). We projected the model predictions across ecoregions to identify areas at risk of future invasion.

2 | METHODS

2.1 | Study area

Our study focused on six-level II EPA ecoregions in the arid and semi-arid western USA: Cold and Warm Deserts, South-Central and West-Central Semi-Arid Prairies, Western Cordillera and Upper Gila Mountains (US Environmental Protection Agency, 2011). EPA ecoregions are based on abiotic and biotic factors at nested scales for consistent analysis across the continent (Omernik & Griffith, 2014). The level II ecoregions outline the arid and semi-arid western USA, and level III ecoregions within them.
represent units of soils, climate and vegetation patterns within which species may have similar invasion potential. Observations of invasive species presence data were obtained from more than 3.3 million km$^2$ in 28 level III ecoregions (Figure 1, Table 1). We included species presence data from outside the focal level II ecoregions to assess suitable habitat of invasive species and areas at risk of future invasion both within and adjacent to the focal area.

2.2 | Data sources

2.2.1 | Focal species

We identified invasive species of management concern using state weed classifications and expert opinion. We used a cut-off of 600 presence observations of these species in our database, resulting in 15 species: five annual grasses, one perennial grass and nine annual forbs (Table 2). Intentionally introduced species (e.g. perennial grasses for forage or erosion control) were excluded. We classified species by rooting type (deep or shallow) to select the depth to which soil predictor variables were integrated: 0–30 cm for shallow-rooted species and 0–100 cm for deep-rooted species. Nomenclature follows USDA Plants (USDA NRCS, 2020).

2.2.2 | Vegetation survey data

We compiled species presence/absence data from multiple regional-scale data sources. Most data came from recent releases (as of April 2020) of the Landfire Reference Database (USGS, 2012) and Bureau of Land Management Assessment, Inventory and Monitoring (AIM) database (BLM, 2020b). The Landfire database included 126,898 vegetation survey plots in the western USA, which were sampled between 1940 and 2006 (> 90% since 1983). Landfire data sources included the USDA Forest Service Forest Inventory and Analysis, US Geological Survey Gap Analysis, National Park Service Vegetation Inventory Program and other vegetation mapping and long-term monitoring efforts conducted by agencies and universities. The BLM AIM database included 27,264 survey plots sampled between 2006 and 2020 (> 90% between 2012 and 2019), including data from the Terrestrial AIM Database (TerrADat) and Landscape Monitoring Framework (LMF) databases (14,926 and 12,338 plots, respectively) (BLM, 2020a, 2020b). TerrADat and LMF data were collected using standardized monitoring protocols for grassland, shrubland and savanna ecosystems (Herrick et al., 2017). We supplemented these datasets with 826 plots from a Joint Fire Sciences Program study of post-fire reseeding treatment responses, the “Chronosequence Study” (Pyke et al., 2011). This study recorded species composition in paired plots within and outside fire perimeters for 88 fires occurring between 1987 and 2003 in the Central and Northern Basin and Range.

Prior to analysis, plot data were thinned to one plot per grid cell at the 3-arc-second resolution (~94 m) of the digital elevation model (described below). This minimized pseudoreplication while preserving the biologically meaningful topographic differences between cells of the elevation model. If plots from the AIM and Landfire databases fell within the same cell, we chose AIM plots, which were sampled more recently. If two Landfire plots fell within the same cell, we chose the
plot that contained the greatest number of focal species, excluding the near-ubiquitous Bromus tectorum. If plots were sampled more than once, we chose the most recent date for AIM and Landfire data and the date when more species were present for the Chronosequence data to provide a comprehensive and up-to-date set of species presence records. The final dataset included 148,404 plots.

2.2.3 | Environmental predictors of presence/absence

We compiled data on climate, topography, soils and factors related to disturbance and dispersal (Table 3). All rasterized data were resampled by bilinear interpolation to the resolution of the digital elevation model to create harmonized prediction surfaces. Climate data were calculated from the TerraClimate monthly products at 150-arc-second resolution (Abatzoglou et al., 2018). Soil data from the National Soil Survey were obtained from gNATSGO (Soil Survey Staff, 2019). Distance from roads, a proxy for human-mediated disturbance and dispersal, was calculated based on the TIGER roads database (US Census Bureau, 2019).

As another metric of disturbance, we included a variable indicating whether the plot had burned between 1984 and the time of sampling based on fire perimeters from the Monitoring Trends in Burn Severity and GeoMACs (Geospatial Multi-Agency Coordination Group, 2020; MTBS Project, 2017). A plot was considered to have burned if it intersected a fire from either dataset prior to the sampling year. Burn status of plots in the Chronosequence study was assigned based on the burned/unburned treatment categories. When projecting models across the landscape, we considered a cell as burned if it overlapped a fire footprint prior to 2019.

Annual species may not be detected if sampling occurs before emergence or after senescence resulting in inaccurate absence records. To incorporate phenology effects, we included day of year of plot sampling as a predictor. When predicting species presence, we set the day of sampling for each species to the day with highest probability of observed presence for that species.

To reduce collinearity in our initial list of 21 possible predictor variables, we removed annual mean and hottest-month temperature (retaining coldest-month temperature), monthly temperature variance (retaining precipitation coefficient of variation), total annual precipitation (retaining climatic water deficit) and sand content (retaining clay and silt content). Remaining variables had Spearman rank correlations of rho < 0.6 across all plot locations. We also removed those variables that never ranked in the ten most influential in preliminary model fitting and variable importance testing, which included topographic position index, soil calcium carbonate concentration, soil available water content and soil depth to restrictive layer. Final models included 13 predictor variables (Table 3).

2.3 | Species distribution modelling

We trained and evaluated predictive models of species presence, then used those models to predict invasion risk within model training areas and across ecoregions where the species was observed. We fit boosted regression trees using the “dismo” and “gbm” packages in R version 3.6.2 (Greenwell et al., 2019; Hijmans et al., 2017; R Core Team, 2019). This method allows for nonlinear relationships and interactions between multiple variables and species presence, is robust to missing data and correlations between variables and combines the strengths of multi-model averaging (fitting many simple regression trees) and boosting (focusing on prediction of hard-to-predict cases, rather than simple majority voting as in random forests) (Compton et al., 2012; Elith et al., 2008; Van Ewijk et al., 2014). We present our methods in ODMAP format (Zurell et al., 2020) in a supplement (Table S1).

### Table 1: Focal level III US EPA ecoregions for prediction of potential distributions of 15 invasive plants species in the USA. Numeric codes correspond to map in Figure 1

| Code | US EPA Level III Ecoregion | Level II Ecoregion                        |
|------|---------------------------|------------------------------------------|
| 4    | Cascades                  | Western Cordillera                       |
| 5    | Sierra Nevada             |                                          |
| 9    | Eastern Cascades Slopes and Foothills |            |
| 11   | Blue Mountains            |                                          |
| 15   | Northern Rockies          |                                          |
| 16   | Idaho Batholith           |                                          |
| 17   | Middle Rockies            |                                          |
| 19   | Wasatch and Uinta Mountains |                                      |
| 21   | Southern Rockies          |                                          |
| 41   | Canadian Rockies          |                                          |
| 77   | North Cascades            |                                          |
| 78   | Klamath Mountains/California High North Coast Range | |
| 10   | Columbia Plateau          | Cold Deserts                             |
| 12   | Snake River Plain         |                                          |
| 13   | Central Basin and Range   |                                          |
| 18   | Wyoming Basin             |                                          |
| 20   | Colorado Plateaus         |                                          |
| 22   | Arizona/New Mexico Plateau |                                      |
| 80   | Northern Basin and Range  |                                          |
| 14   | Mojave Basin and Range    | Warm Deserts                             |
| 24   | Chihuahuan Deserts        |                                          |
| 81   | Sonoran Basin and Range   |                                          |
| 25   | High Plains               | South-Central Semi-Arid Prairies         |
| 26   | Southwestern Tablelands   |                                          |
| 42   | Northwestern Glaciated Plains |                              |
| 43   | Northwestern Great Plains |                                          |
| 44   | Nebraska Sand Hills       |                                          |
| 23   | Arizona/New Mexico Mountains | Upper Gila Mountains                  |

We present our methods in ODMAP format (Zurell et al., 2020) in a supplement (Table S1).
2.3.1 Model building

To distinguish between informative and uninformative absences, we adapted methods from species distribution modelling based on presences and pseudoabsences (Itúrbide et al., 2015). We developed species-specific buffers from observed presences, beyond which including additional absence data did not improve the models’ ability to accurately predict species presence (Barga et al., 2018). These model training buffers approximated the area where species may have already explored; absences within buffers likely represented unsuitable, rather than unexplored, habitat.

We set the buffer distance for each species based on fivefold cross-validation, evaluating a range of possible buffers between 40 and 340 km in 30-km increments. To minimize influence of spatial outliers on model training areas, we defined buffers based on mean distance to the second-through sixth-nearest presences, so that small clusters or isolated presence observations were not automatically included. For each buffer distance, we trained the model on a random sample of four-fifths of the plots within the buffer (training folds), stratified by presence and tested the model on the remaining fifth (test fold).

We then selected the buffer distance that optimized area under the precision-recall curve (AUPRC) across cross-validation folds. In an invasion-risk framework, it is more important to accurately predict species presences than absences (Jiménez-Valverde et al., 2011). AUPRC is better suited to assessing model performance on imbalanced datasets (e.g. with many more absences than presences) than the commonly used area under the receiver-operator curve (AUROC) (Lobo et al., 2008; Saito & Rehmsmeier, 2015). AUROC is high for models that correctly predict absences and fail to detect presences, because both types of errors are weighted equally, whereas AUPRC focuses on correct prediction of presences. Because AUPRC for a null model (random guessing) is equal to species prevalence, the proportion of plots where the species was observed (Carrington et al., 2020), we based our buffer selection on the difference between AUPRC and species prevalence within the buffer. Larger buffers generally decreased AUPRC, but also decreased prevalence. If the difference in the mean net AUPRC (AUPRC minus prevalence) between several buffers was less than 1% of the maximum net AUPRC, the largest of these buffers was selected to include the maximum number of training plots.

To increase model sensitivity and emphasize accurate prediction of presences, we trained all models on a subset of data from within the invaded-area buffer, such that at least 10% of observations in the training data represented species presences. In each cross-validation step, we selected training data from within the buffer by including all presence observations and adding randomly selected absences until the 10% prevalence threshold was reached. If species prevalence within the buffer already exceeded 10%, this step was skipped.

In all model-building steps, we fit 2000 boosted regression trees per model, with a tree complexity (maximum possible order of interactions between predictors) of 5 and a learning rate of 0.02. These

| Name                        | Code | Common name                  | Functional group | Rooting type | Number of presences |
|-----------------------------|------|------------------------------|------------------|--------------|---------------------|
| Bromus japonicus            | BRJA | Field brome/ Japanese brome  | Annual grass     | Shallow      | 640                 |
| Bromus rubens               | BRRU2| Red brome                    | Annual grass     | Shallow      | 4,014               |
| Bromus tectorum             | BRTE | Cheatgrass                   | Annual grass     | Shallow      | 31,156              |
| Schismus barbatus           | SCBA | Common Mediterranean grass   | Annual grass     | Shallow      | 1,354               |
| Taeniatherum caput-medusae  | TACA8| Medusahead                   | Annual grass     | Shallow      | 964                 |
| Poa bulbosa                 | POBU | Bulbous bluegrass            | Perennial grass  | Shallow      | 2011                |
| Ceratocephala testiculata   | CETE5| Curvedseed butterwort        | Annual forb      | Shallow      | 2,504               |
| Draba verna                 | DRVE2| Spring draba                 | Annual forb      | Shallow      | 1,216               |
| Erodium cicutarium          | ERCI6| Redstem storksbill           | Annual forb      | Shallow      | 3,723               |
| Halogeton glomeratus        | HAGL | Saltlover                    | Annual forb      | Deep         | 3,313               |
| Lactuca serriola            | LASE | Prickly lettuce              | Annual forb      | Deep         | 3,023               |
| Lepidium perfoliatum        | LEPE2| Clasping pepperweed          | Annual forb      | Deep         | 2,752               |
| Salsola tragus              | SATR12| Prickly Russian thistle      | Annual forb      | Deep         | 3,445               |
| Sisymbrium altissimum       | SIAL2| Tall tumbledmustard          | Annual forb      | Deep         | 4,639               |
| Tragopogon dubius           | TRDU | Yellow salsify               | Annual forb      | Deep         | 6,571               |
| Predictor | Short name | Range (median) within focal Level II ecoregions | Range (median) within sampled plots | Source | Definition |
|-----------|------------|-----------------------------------------------|----------------------------------|--------|------------|
| Climatic Water Deficit (annual) | CWD | 22–2,342 mm (806) | 55.5–2,331 mm (780) | Derived from TerraClimate monthly products at 1/24th degree (Abatzoglou et al., 2018) | Annual sum of monthly difference between potential and actual evapotranspiration. Monthly values are means of 1958–2018 values for each month and grid cell. |
| Summer precipitation | PPT summer | 4–315 mm (112) | 1–306 mm (73) | | Sum of June, July, August precipitation |
| Median temperature of coldest month | TMIN | −13.8–14.6°C (−2.3) | −12.5–14.6°C (−2.2) | | Monthly median temperatures calculated as mean of the monthly minimum and maximum temperature for each month |
| Coefficient of variation of monthly precipitation | PPTCV | 0.08–1.04 (0.56) | 0.08–1.10 (0.42) | | Standard deviation divided by mean |
| Correlation between monthly precipitation and median temperature | PTCOR | −1.00–1.00 (0.37) | −1.00–1.00 (−0.48) | | Pearson correlation coefficient between monthly median temperature and monthly precipitation |
| Topographic heat load index | HLI | 0.22–1.10 (0.81) | 0.52–1.01 (0.81) | Derived from 3-arc-sec (~94 m) gap-filled NASA Shuttle Radar Tomography Mission data (Jarvis et al., 2008) | Linear function of aspect (relative to SW), slope and latitude. Arc Gradient Metrics toolbox (Evans et al., 2014), following (McCune & Keon, 2002) |
| Topographic ruggedness index | TRI | 0.0–916 (3.5) | 0.0–135.4 (7.9) | | Mean of the absolute differences in elevation of a cell and its eight neighbours (Wilson et al., 2007) |
| Clay content | Clay | 0–100 cm: 0.0%–74.5% (17.6) | 0–100 cm: 0.0%–74.5% (15.8) | gNATSGO (Soil Survey Staff, 2019) | Calculated from the dominant soil series of the map unit covering each plot or grid cell. Integrated for horizons from 0–30 or 0–100 cm, depending on the species. |
| Silt content | Silt | 0–100 cm: 0.0%–86.4% (26.4) | 0–100 cm: 0.0%–83.2% (26.6) | | |
| Electrical conductivity | EC | 0–100 cm: 0.0–381 dS/m (0.60) | 0–100 cm: 0.0–381 dS/m (0.0) | | |

(Continues)
parameters were derived from preliminary models for each species using the gbm.step method to optimize the number of trees per model based on presence-stratified 10-fold cross-validation and minimization of model deviance, which resulted in between 1,700 and 2,400 trees.

2.3.2 | Model evaluation

For each species, we calculated AUPRC, prevalence and AUROC, as the mean values from the five model cross-validation steps within the final model training area. We also calculated model specificity, sensitivity, false positive rate and the Symmetric Extremal Dependence Index (SEDI, a model performance measure for predicting low-frequency events (Wunderlich et al., 2019)) at two thresholds of predicted probability of species presence for classifying a species as present at a point: 0.15 (low threshold for maximizing prediction of presence) and 0.5 (more conservative). Choice of a threshold for converting probabilities to expected presence/absence, or integration across possible thresholds as with AUPRC, depends on the data user's priorities. Variable importance was measured as the relative reduction in model mean square error resulting from data partitions based on each variable, averaged across all trees in the final model, as implemented in the "gbm" package (Friedman, 2001; Greenwell et al., 2019). We produced partial effects plots to visualize relationships between individual predictors and species presence (Hijmans et al., 2017).

2.3.3 | Prediction of invasion probability

After evaluating the models, we fit a single final model per species using all presence observations within the training area and enough absence observations from the area to maintain 10% prevalence. We projected model-predicted probability of occurrence at 3-arc-second resolution across the level III ecoregions where each species was observed, within the focal area.

3 | RESULTS

3.1 | Invasion patterns of grasses

3.1.1 | Cold desert grasses

Bromus japonicus was observed in a broad but patchy distribution in 16 of 28 analysed ecoregions. Its current distribution was characterized by cool and relatively moist areas (Figure 2). Within the relatively cool model training area (40 km buffer around presences), probability of occurrence was highest with warmer temperatures (TMIN > 0°C) and intermediate climatic water deficit (CWD 500–800 mm) (Figure S1a). Day of sampling was an important predictor of presence (Table 4), possibly because this species is hard to identify prior to flowering. Model precision and net AUPRC were lowest of the modelled species (Table 5). B. japonicus was projected to expand
its range in several ecoregions, filling in gaps between current pres-
ences in the Northwestern Great Plains and potentially expanding
throughout the semi-arid prairies (Figure 3).

The training area for *Bromus tectorum* included the entire western
USA; no vegetation survey plot was more than 340 km from the near-
est five presences and more than 20% of plots contained the species
(Figure 3). Although *B. tectorum* occurred across the range of environ-
mental conditions, it was most abundant in Cold Deserts (Figure 3).
Modelled probability of presence was higher with CWD > 500 mm,
summer precipitation < 100 mm, and in burned areas (Figure S1c),
although burn history was less important than climatic predictors
(Table 4). Although *B. tectorum* occurred across the range of environ-
mental conditions, it was most abundant in Cold Deserts (Figure 3).

*Taeniatherum caput-medusae* was observed in Cold Deserts, par-
ticularly those with clayey soils and winter-dominant precipitation
(Figure 2, Figure S2a). Its probability of presence was greatest for clay
contents from 20%-50%, CWD between 700 and 1,000 mm, and
minimum temperatures above −4°C (Figure S1e). Sampling day of year
was the most important predictor of presence (Table 4) with high-
est detection probabilities in early summer and late fall. This species
was predicted to expand within the Columbia Plateau and possibly
Snake River Plain, Northern Basin and Range, and Western Cordillera
ecoregions (Figure 3). Model performance was relatively high as was
its local prevalence (> 7% of plots within the training area) (Table 5).

### 3.1.2 Warm desert grasses

*Bromus rubens* was confined largely to the hot southwest margin
of the focal area. Absence data from up to 340 km of presences
improved model performance, defining a climatic niche with cold-
est-month temperatures below 0°C (generally > 5°C), low but
highly variable summer precipitation and annual CWD > 1,000 mm
(Figure 2). Models performed well with the highest values for net
AUPRC (0.64) and second-highest sensitivity and SEDI (Table 5). The
species was projected to fill in favourable areas of the Mojave Basin
and Range and Arizona/New Mexico Mountains (Figure 3).

Like *B. rubens*, *Schismus barbatus* showed a well-defined climatic
niche with TMIN > 5°C and CWD > 1,500 mm (Figure 2, Figure S1d).
Model predictions were robust with the highest sensitivity and SEDI
and high net AUPRC within the 340-km buffer (Table 5). Infilling was
predicted within the Mojave Basin and Range, and this species may
threaten ecosystems outside of the study area, particularly in south-
ern California (Figure 3).
3.2 | Invasion patterns of forbs

3.2.1 | Cold desert forbs

*Ceratocephala testiculata* occupied areas of the northwestern Cold Desert ecoregions with moderate climatic water deficits (CWD ~ 1,000 mm), precipitation distributed relatively evenly throughout the year (PPTCV < 0.5) and generally subfreezing coldest-month temperatures (TMIN < 0) (Figure 4). The species was predicted to expand in Cold Deserts within the 100-km training buffer and farther into eastern Cold Deserts (Figure 5). Its probability of presence was greatest in areas with moderate CWD and day of sampling had a strong influence on its detection (Table 6, Figure S1g). Model performance was intermediate (AUPRC of 0.37) (Table 5).

*Draba verna* was present in northwestern Cold Deserts and Blue Mountains, primarily in regions with winter-dominant precipitation and CWD < 1,000 mm annually (Figure 4, Figure 5). Like *C. testiculata*, it was predicted to occur most frequently for spring sampling dates and subfreezing minimum temperatures (Table 6, Figure S1h). This species was also associated with higher silt content. It was projected to expand substantially outside its 40-km model training buffer, primarily in the Middle Rockies (Figure 5).

The observed range of *Halogeton glomeratus* was centred in the Cold Deserts (Figure 5). It a halophyte with higher modelled probability in soils with nonzero electrical conductivity (i.e. saline soils) (Table 6). The species occurred in flat areas (low ruggedness) with low summer precipitation (<100 mm, but positive PTCOR) and high CWD (> 1,000 mm) and had a narrower range of topographic heat load index than most species (Figure 4, Figure S2b). This species was predicted to remain within the 160-km model training buffer filling in suitable habitat within the Central Basin and Range, Wyoming Basin and Colorado Plateaus (Figure 5). Models for this species had relatively high sensitivity and SEDI, consistent with the strength of salinity as a predictor (Table 5).

*Tragopogon dubius* had the most eastern range with extensive infilling predicted in the Northwestern Great Plains, Glaciated Plains, northwestern Cold Deserts and Columbia Plateau (Figure 5). Probability of presence was high in regions with minimum temperatures below freezing and CWD below 1,000 mm and either cold-season or warm-season precipitation (Table 6, Figure S1o). It was observed in areas with greater ruggedness than most species (Figure S2b).

3.2.2 | Warm desert forb

*Erodium cicutarium* inhabited a range similar to the warm-desert grasses (Figure 5) with modelled probability increasing at above-freezing temperatures (especially > 5°C), CWD > 500 mm (especially > 900 mm) and low summer precipitation, even where most precipitation falls in summer (Table 6, Figure S1l). It was also observed and predicted to expand in patches throughout the Cold Deserts. It was most commonly detected in early spring, particularly in areas with lower CWD, but could be detected throughout the year in areas with CWD > 1,200 mm (Figure 4).

### TABLE 4

Top five predictors of grass species presence based on final models including all available presence data. Relative influence is the improvement in model mean square error at all nonterminal nodes based on the target predictor averaged across all trees in the model (Friedman, 2001)

| Species               | Top 5 predictors of presence | Relative influence | Species               | Top 5 predictors of presence | Relative influence |
|-----------------------|------------------------------|--------------------|-----------------------|------------------------------|--------------------|
| *Bromus japonicus*    | TMIN                         | 13.66              | *Poa bulbosa*         | CWD                          | 17.78              |
|                       | Day                          | 13.13              |                       |                              | 15.30              |
|                       | CWD                          | 12.67              |                       |                              | 9.40               |
|                       | PPTCV                        | 11.90              |                       |                              | 9.07               |
|                       | PPT_summer                   | 10.01              |                       |                              | 8.89               |
| *Bromus rubens*       | TMIN                         | 44.62              | *Schismus barbatus*    | CWD                          | 46.49              |
|                       | PTCOR                        | 12.38              |                       |                              | 11.83              |
|                       | CWD                          | 10.99              |                       |                              | 9.79               |
|                       | PPT_summer                   | 6.62               |                       |                              | 7.42               |
|                       | PPTCV                        | 5.36               |                       |                              | 5.63               |
| *Bromus tectorum*     | CWD                          | 26.65              | *Taeniatherum caput-medusae* | Day                          | 16.71              |
|                       | TMIN                         | 20.33              |                       |                              | 15.01              |
|                       | PPT_summer                   | 15.77              |                       |                              | 13.71              |
|                       | PTCOR                        | 10.51              |                       |                              | 11.77              |
|                       | PPTCV                        | 6.94               |                       |                              | 7.76               |

Note: Clay = per cent clay, CWD = climatic water deficit, Day = sampling day of year, EC = electrical conductivity, PPT_summer = summer (June, July, August) precipitation, PPTCV = coefficient of variation of monthly precipitation, PTCOR = correlation between monthly temperature and precipitation, Silt = per cent silt, TMIN = temperature of coldest month (°C).
TABLE 5 Metrics of predictive ability for models trained and tested within buffered invaded areas. Metrics are means from fivefold cross-validation on all training data within the invaded areas. Sensitivity, false positive rate, precision and the Symmetric Extremal Dependence Index (SEDI) are each calculated at two thresholds for classifying a species as present. The low threshold, considering a species as present if the predicted probability of presence exceeds 0.15, maximizes prediction of presences, while a threshold of 0.5 is more conservative.

| Species | Buffer (km) | AUROC | Prevalence | Net AUPRC  | Probability threshold for labelling species as present | Training plots (n) |
|---------|-------------|-------|------------|------------|-------------------------------------------------------|--------------------|
|         |             |       |            |            | 0.15 0.50 0.15 0.50 0.15 0.50 0.15 0.50 0.15 0.50 |                    |
| BRJA    | 40          | 0.885 | 0.028      | 0.228      | 0.671 0.320 0.100 0.020 0.162 0.313 0.742 0.574 4,111 |
| BRRU2   | 340         | 0.974 | 0.044      | 0.637      | 0.927 0.728 0.066 0.023 0.395 0.592 0.948 0.874 18,200 |
| BRTE    | 340         | 0.919 | 0.210      | 0.554      | 0.915 0.599 0.251 0.054 0.492 0.748 0.822 0.742 29,681 |
| SCBA    | 340         | 0.983 | 0.110      | 0.620      | 0.939 0.779 0.057 0.023 0.350 0.539 0.958 0.902 8,549  |
| TACAY   | 40          | 0.945 | 0.075      | 0.581      | 0.810 0.316 0.080 0.017 0.452 0.707 0.869 0.751 2,433  |
| POBU    | 40          | 0.895 | 0.114      | 0.589      | 0.709 0.342 0.117 0.017 0.292 0.575 0.757 0.608 6,058  |
| CTE5    | 100         | 0.913 | 0.040      | 0.370      | 0.759 0.390 0.107 0.017 0.228 0.483 0.807 0.652 12,535 |
| DRIVE2  | 40          | 0.934 | 0.040      | 0.463      | 0.784 0.503 0.097 0.017 0.255 0.533 0.837 0.742 5,865  |
| ERCI6   | 340         | 0.941 | 0.027      | 0.434      | 0.790 0.505 0.086 0.017 0.201 0.454 0.850 0.745 27,973 |
| HAGL    | 160         | 0.93  | 0.037      | 0.336      | 0.815 0.404 0.120 0.021 0.209 0.425 0.840 0.650 17,728 |
| LASE    | 40          | 0.885 | 0.054      | 0.332      | 0.696 0.288 0.120 0.015 0.245 0.527 0.741 0.567 10,746 |
| LEPE2   | 40          | 0.848 | 0.114      | 0.341      | 0.634 0.181 0.130 0.010 0.239 0.547 0.665 0.471 8,377  |
| SATR12  | 40          | 0.848 | 0.051      | 0.231      | 0.635 0.169 0.142 0.011 0.193 0.446 0.659 0.445 12,983 |
| SIAL2   | 280         | 0.93  | 0.032      | 0.434      | 0.776 0.508 0.089 0.018 0.221 0.477 0.837 0.743 29,407 |
| TRDU    | 340         | 0.925 | 0.045      | 0.396      | 0.789 0.424 0.107 0.020 0.257 0.498 0.831 0.671 29,427 |

\(^a\text{AUROC} = \text{Area Under Receiver-Operator Curve.}\)
\(^b\text{Prevalence} = \text{true presences / all observations.}\)
\(^c\text{AUPRC} = \text{Area Under Precision-Recall Curve; Net AUPRC (AUPRC minus prevalence) represents increase in AUPRC above baseline of random chance.}\)
\(^d\text{Sensitivity} = \text{true presences / (true presences + false absences).}\)
\(^e\text{False-positive rate} = \text{false presences / (true presences + false presences).}\)
\(^f\text{Precision} = \text{true presences / (true presences + false presences).}\)
\(^g\text{SEDI} = \ln(\text{false pos rate}) − \ln(\text{sensitivity}) − \ln(1 − \text{false pos rate}) + \ln(1 − \text{sensitivity})\) / \ln(\text{false pos rate}) + \ln(\text{sensitivity}) + (1 − \text{false pos rate}) + \ln(1 − \text{sensitivity}).\)

3.2.3 | Disturbance-driven and widespread forbs

\textit{Lactuca serriola} was widely distributed throughout the western USA, though it was most common in northern latitudes. It was associated with moderately dry climates (CWD 500–1000 mm/year) with winter-dominant precipitation (PTCOR < 0) (Figure 4). Few presences were observed in the High Plains or Arizona/New Mexico Mountains, but it was predicted to expand its distribution widely in these regions and the Columbia Plateau (Figure 5). It was frequently observed in previously burned areas and discrete patches of predicted expansion corresponded to fire footprints (Figure 5).

\textit{Sisymbrium altissimum} was patchily observed in burned areas throughout the Cold Deserts (Figure 5). Burn status contributed to over 40% of regression tree nodes, whereas summer precipitation (greatest probability of presence with < 100 mm) and coldest-month temperatures (greatest probability from 0 to −10°C) each contributed about 11% (Figure S1n). It was observed and predicted to expand in eastern Cold Desert and prairie ecoregions within its 340-km modelled buffer (Figure 5). As with \textit{L. serriola}, this species did not have a clearly defined climate niche and model performance on training data was intermediate (Table 5).

\textit{Lepidium perfoliatum} was most common in the Cold Deserts but showed a wide and patchy distribution across the range of climate variables (Figure 5). Presence probability was highest in areas with summer precipitation < 50 mm and CWD > 600 mm (Table 6, Figure S1l). Model performance was relatively poor for this species, consistent with a generalist habitat and seasonal detection constraints (Table 5). With this caveat, expansion was projected for the Northwestern Great Plains and Mojave Basin and Range and infilling projected for the Northern and Central Basin and Range and Snake River Plain (Figure 5).

As with \textit{L. perfoliatum}, \textit{Salsola tragus} had a wide and patchy distribution, primarily in the Cold Deserts, and was similarly difficult to predict. Observations were associated with flat terrain in basins (narrow range of HLI and ruggedness) (Figure S1b).
Predicted probability of presence was highest in areas with moderately high CWD (typically > 1,000 mm), warm-season precipitation (PTCOR > 0) and nonzero electrical conductivity (Table 6, Figure S2m). In addition to infilling in the western Central Basin and Range, the model predicted extensive expansion into the Southwestern Tablelands, where warm-season precipitation dominates (Figure 5).

### 3.3 Model performance on test data

Overall, the models discriminated between presences and absences in data withheld from the training set (Table 5). Values of AUROC were generally high (0.85 or greater for all species). The more relevant AUPRC metric was usually an order of magnitude higher than its baseline of species prevalence in the test dataset, ranging from 0.228 for *B. japonicus* (present in only 2.8% of plots within the 40-km invaded-area buffer) to 0.637 for *B. rubens* (present in 4.4% of plots within a 340-km buffer and occupying a well-defined temperature niche).

Even when trained within buffers and with presence-weighted training data, models were more likely to miss observed presences than to predict presences in areas where none were observed. With a threshold of 50% probability for classifying a species as present in a grid cell, models predicted less than 20% of *S. tragus* and *L. perfoliatum* presences in cross-validation. At the 50% presence probability threshold, the false positive rate was less than 2.5% for all species except *B. tectorum*. Therefore, the models provided a conservative estimate of which areas might contain invasive species.

Model projections beyond training-area buffers generally stayed within the range of environmental conditions covered by the training data (Figure S3) with a few exceptions. For *B. japonicus* and *D. verna*, survey points did not cover the coldest parts of the ecoregions, potentially affecting model predictions for *D. verna*, which was associated with low temperatures. Ranges of CWD in invaded ecoregions extended beyond those of sample plots for *P. bulbosa* and *B. japonicus*. Plots with *T. caput-medusae* did not exhibit the full range of intra-annual precipitation variation; this species was associated with both high and low values of PPTCV (Figure S1e).
Our distribution maps provide conservative predictions of invasion potential, differentiating between buffered areas around observed occurrences where models were trained and currently unoccupied areas that may be suitable for invasion. We compiled nearly 150,000 presence/absence observations to overcome the challenges of predicting invasive species distributions from patchy and expanding ranges, and systematically selected the most informative absence data to model the environmental niches of 15 common invasive plants, including relatively unstudied species.

To our knowledge, this is the first attempt to map regional invasion potential for these species at a resolution that can inform landscape-scale planning. For the high-profile Bromus species, our results build on prior studies of invasion potential by spatially mapping invasion probability across the region (e.g. Brooks et al., 2016; Shinneman & Baker, 2009). For other species, prior studies were smaller scale (e.g. Brandt & Rickard, 1994; Morris et al., 2013) or mapped outside of the western USA (Clements et al., 1999; Francis et al., 2012). Our maps supplement national compilations of distribution data, such as the Early Detection and Distribution Mapping System, by identifying drivers of species distribution and projecting invasion probability beyond point observations and county scales (EDDMapS, 2020; USDA NRCS, 2020).

Our model outputs suggested sources of ecosystem vulnerability to invasion under current and future conditions and were consistent with known species traits. Predicted probability of presence tended to follow ecoregional boundaries, indicating that models reflected established thresholds in environmental conditions. As expected, climate variables were the best predictors of most species’ distributions. Some specialists were predicted by soil characteristics, while some generalists were most responsive to disturbance. These predictions can be incorporated into an adaptive management framework to anticipate, detect and respond to the spread of invasive plant species.

Climate change may promote spread of many non-native invasive grasses and forbs

Annual invaders are highly sensitive to temperature effects in the climatically and topographically diverse western USA (Brooks et al., 2016). Minimum temperature was among the top five predictors of all but two species and was among the top two predictors for B. japonicus, B. tectorum, B. rubens, D. verna, E. cicutarium and T. dubius. Warmer temperatures are causing species distribution shifts poleward and upward in elevation around the globe (Chen...
et al., 2011; Parmesan & Yohe, 2003). The effect of minimum temperatures on the focal species’ future distributions will depend on their physiological tolerances. For example, *B. rubens* occurs in areas with higher minimum temperatures than *B. tectorum*, has lower freezing tolerance (Bykova & Sage, 2012), and is less broadly distributed in warmer, southern ecoregions (Salo, 2005). *Bromus tectorum’s* distribution is projected to expand into areas that are becoming warmer, but to contract in the hotter and drier edge of its range (Bradley et al., 2016). *Bromus rubens* is projected to expand into those areas becoming less suitable for *B. tectorum* as temperatures warm.

Temperature interacts with precipitation to influence climatic water deficit (CWD), which was among the top five predictors of all but one species. Many focal species were associated with relatively

**FIGURE 5** Predicted habitat suitability for forb species. Probability of presence is shown in a range from yellow to red within the model training area for each species, and from blue to red outside the training area, within the Level III ecoregions where the species has been observed. Small grey points represent survey plots where each species was present. Albers equal-area conic projection. CETE5 = *Ceratocephala testiculata*, DRVE2 = *Draba verna*, ERCI6 = *Erodium cicutarium*, HAGL = *Halogeton glomeratus*, LASE = *Lactuca serriola*, LEPE2 = *Lepidium perfoliatum*, SATR12 = *Salsolatragus*, SIAL2 = *Sisymbrium altissimum*, TRDU = *Tragopogon dubius*
high CWD, which is predicted to increase as greater evapotranspiration from higher temperatures more than offsets potential increases in precipitation (Cook et al., 2004; Jeong et al., 2014). Many annual or short-lived invasive plant species in these environments have highly plastic responses to resource availability and can rapidly grow and reproduce when soil water is available and temperatures are conducive (Hulbert, 1955; James et al., 2011). Regional increases in CWD may promote expansion of species that are currently associated with areas of high CWD, but differences will likely exist among species. Growth and reproduction of shallow-rooted species, including the focal grasses, occurs prior to drying of shallow soils in summer across much of the study area (Ryel et al., 2010). For deeper-rooted species, like many of the forbs, growth and reproduction usually extends into the summer (Prevéy et al., 2010). Decreases in precipitation and recharge of deeper soil water may disadvantage both deeper-rooted natives and invaders in areas with low or decreasing summer precipitation (Wilcox et al., 2012). Ecosystem conversion to shallow-rooted annual grasses and forbs and increased uptake from shallow soil layers may have a similar effect.

The abundance and timing of precipitation varies across the western USA and drives species distributions (Robertson et al., 2009). In general, the timing of precipitation ranges from winter-dominated precipitation in the northwest to monsoonal patterns of summer precipitation in the southeast. At local scales, precipitation abundance changes with elevation and affects the availability of summer soil moisture. In our study, modelled species distributions reflected both the abundance and seasonality of precipitation. Species associated with high summer precipitation, such as *B. japonicus* and *T. dubius*, occurred primarily in northern or higher elevation areas with low TMIN and CWD. Species associated with low summer precipitation—*B. rubens*, *L. perfoliatum*, *S. altissimum* and *S. barbatus*—were in southern or lower elevation areas with high TMINs and CWDs.

The correlation between precipitation and temperature (PTCOR), which indicates the seasonality of precipitation (winter-dominant for

### TABLE 6 Top five predictors of forb species presence based on final models including all available presence data. Relative influence is the improvement in model mean square error at all nonterminal nodes based on the target predictor averaged across all trees in the model (Friedman, 2001)

| Species                      | Top 5 predictors of presence | Relative influence | Species                      | Top 5 predictors of presence | Relative influence |
|------------------------------|------------------------------|--------------------|------------------------------|------------------------------|--------------------|
| *Ceratocephala testiculata*  | Day                          | 21.14              | *Lepidium perfoliatum*       | PPT_summer                   | 16.43              |
|                              | CWD                          | 12.32              |                              | Day                          | 11.56              |
|                              | PTCOR                        | 9.40               |                              | CWD                          | 10.12              |
|                              | TMIN                         | 8.71               |                              | TMIN                         | 8.24               |
|                              | Burn_since_84                | 7.37               |                              | TR              | 7.46               |
| *Draba verna*                | Day                          | 17.55              | *Salsola tragus*             | CWD                          | 11.37              |
|                              | TMIN                         | 11.18              |                              | PTCOR                       | 10.89              |
|                              | Silt                         | 10.50              |                              | EC                          | 10.73              |
|                              | Clay                         | 10.26              |                              | Day                          | 10.24              |
|                              | PPT_summer                   | 9.36               |                              | TMIN                         | 8.79               |
| *Erodium cicutarium*         | TMIN                         | 21.43              | *Sisymbrium altissimum*      | Burn_since_84                | 40.58              |
|                              | CWD                          | 19.54              |                              | PPT_summer                   | 11.46              |
|                              | Day                          | 12.53              |                              | TMIN                         | 11.38              |
|                              | PPT_summer                   | 9.47               |                              | CWD                          | 7.51               |
|                              | PTCOR                        | 8.35               |                              | PPTCV                       | 6.75               |
| *Halogegeton glomeratus*     | EC                           | 23.86              | *Tragopogon dubius*          | PTCOR                       | 15.78              |
|                              | PPT_summer                   | 13.05              |                              | TMIN                         | 14.51              |
|                              | CWD                          | 11.84              |                              | CWD                          | 14.15              |
|                              | PTCOR                        | 10.60              |                              | PPTCV                       | 11.27              |
|                              | PPTCV                        | 7.86               |                              | Burn_since_84                | 11.21              |
| *Lactuca serriola*           | Burn_since_84                | 13.20              |                              |                              |                    |
|                              | PPTCV                        | 12.88              |                              |                              |                    |
|                              | CWD                          | 12.30              |                              |                              |                    |
|                              | Day                          | 10.66              |                              |                              |                    |
|                              | PTCOR                        | 7.49               |                              |                              |                    |

Note: Clay = per cent clay, CWD = climatic water deficit, Day = sampling day of year, EC = electrical conductivity, PPT_summer = summer (June, July, August) precipitation, PPTCV = coefficient of variation of monthly precipitation, PTCOR = correlation between monthly temperature and precipitation, Silt = per cent silt, TMIN = temperature of coldest month (°C).
negative values, summer-dominant for positive and throughout the year for values near zero), was an important predictor of the distribution of many species. Several forb species, including *H. glomeratus*, *S. tragus* and *T. dubius*, were found across a range of PTCOR and were predicted to expand in areas where precipitation coincided with warm temperatures. The grass species *P. bulbosa* and *B. tectorum* were predicted to occur in locations with winter-dominated precipitation. Climate changes that result in shifts in the seasonality of precipitation (Polley et al., 2013) may substantially alter the distribution of suitable habitat for these species.

### 4.1.2 Soil properties will influence expansion of specialist invaders

Soil salinity (EC) was a strong predictor of the range of *H. glomeratus*. This species appears to alter soil chemistry by concentrating salts and changing microbial communities, which may further decrease site suitability for other invaders (Duda et al., 2003). *Halogoton glomeratus* may benefit from climate warming in some areas as higher temperatures facilitate its germination (Khan et al., 2001). *Salsola tragus* is salt tolerant and was associated with higher salinity, but invades diverse environments (Beckie & Francis, 2009). Climatic factors were important in determining the species’ ranges and will interact with soil properties to influence future invasions.

### 4.1.3 Disturbance may rapidly expand potential habitat for invasion

Fires are becoming larger and more frequent across the arid and semi-arid West (Abatzoglou & Kolden, 2013; Dennison et al., 2014; Fusco et al., 2019) and are rapidly changing the map of habitat suitability for post-fire invaders. Overlap with a fire footprint was the top predictor for *S. altissimum* and *L. serriola* and was within the top 5 for *C. testiculata* and *T. dubius*. These species have seeds that are dispersed by wind or adhesion to animals, and the strength of fire as a predictor of their presence reflects both their capacity for widespread dispersal and establishment following disturbances (Brandt & Rickard, 1994). Projected increases in burned area may therefore translate into large increases in these species’ ranges.

Distance to roads was not highly predictive of any species, likely because more than 87% of sample plots occurred within 1,000 m of a road and no species was confined to roadside margins. In Wyoming, *H. glomeratus*, *C. testiculata* and *B. tectorum* were abundant > 1,000 m from human infrastructure, indicating non-human-mediated dispersal between disturbed areas (Manier et al., 2014). Disturbances such as livestock grazing or development of oil, gas and solar infrastructure also influence the distributions of these species. Expansion of invasive species mediated by disturbance and climate change can feedback to fire regimes, further altering invasion potentials (Liu & Wimberly, 2016; Parks et al., 2016).

### 4.2 Data availability and species characteristics affect model outcomes

For many species, timing of sampling contributed meaningfully to the models, suggesting that imperfect detection due to phenology artificially constrained modelled suitable habitats (Jarnevich et al., 2015). Sampling date may also have acted as a proxy for climatic parameters not fully captured in the other predictors. While we expect field crews accurately identified common invasive species during the growing season, pre-emergence or post-senescence sampling of short-lived forbs like *C. testiculata* and *D. verna* and misidentification of species like *T. caput-medusae* may have resulted in false absence records. This indicates that monitoring of these species will need to be timed to maximize detection and more accurately model their distributions.

Prevalence and habitat specificity also affected model performance: the distribution of species with high prevalence in a limited subset of environmental conditions, such as *S. barbatus*, was easiest to accurately model. Model performance was poorer for species that were widely dispersed across the study region and range of environmental conditions, particularly *L. perfoliatum* and *S. tragus*. These species may be generalists for which presences are defined more by rare long-range dispersal events than by environmental conditions (Beckie & Francis, 2009). In contrast, although *B. japonicus* was represented by the fewest presence points, the final model had higher sensitivity and SEDI than those for *L. perfoliatum* and *S. tragus*, suggesting that this species has a better-defined environmental niche and modelling ability was limited by available habitat and sample size.

### 4.3 Species distribution models can help manage emerging invasion risks

Anticipating future distributions of non-native invasive species aids managers in preventing introductions and eradicating new populations. Maps summarizing areas with environmental suitability for invaders can be useful tools for anticipating species’ invasions and controlling their spread (Jiménez-Valverde et al., 2011). Our novel approach illustrates a range of certainty about potential invasion locations, including known presences, buffer areas where the species may have explored, and likely unexplored areas with environmental suitability. Managers of large landscapes can use this information to design monitoring and control strategies based on the distance-to-presence thresholds identified in our models. For example, intensive surveys and control efforts can target areas with high probability of species occurrence within buffer areas, while more speculative high-probability areas outside of the buffer can be opportunistically monitored when funding and personnel are available.

Both species distributions and areas at risk will change as the influences of climate change and natural and anthropogenic disturbance increase. Models of species’ distributions should be revised as part of an adaptive management framework that incorporates
smaller scales. Predicted areas of expansion can be targeted to search for new populations and models can incorporate new records of occurrence and understanding of species’ ecology (Uden et al., 2015).

The environmental conditions associated with each species that are identified in the models can be used to help target species-specific monitoring and control efforts at finer management scales. The information provided by the models can be supplemented with local knowledge of soil characteristics, microclimates and disturbance history. Our approach provides the needed distribution maps and information on environmental suitability of common invaders to help increase their detection and prevent their expansion across the arid and semi-arid western USA.

ACKNOWLEDGMENTS
This work was supported by a grant from the US Joint Fire Science Program (JFSP Project ID: 19-2-02-11). We thank John Bradford, Brice Hanberry, Francis Kilkenny and Shannon Still for their feedback on the study design.

PEER REVIEW
The peer review history for this article is available at https://publon ns.com/publon/10.1111/ddi.13232.

DATA AVAILABILITY STATEMENT
Raster files of species probability maps that were generated for this study, analysis R code and associated metadata are available on the U.S. Forest Service Research Data Archive: (https://doi.org/10.2737/ RDS-2020-0078).

ORCID
Alexandra K. Urza https://orcid.org/0000-0001-9795-6735

REFERENCES
Abatzoglou, J. T., Dobrowski, S. Z., Parks, S. A., & Hegewisch, K. C. (2018). TerraClimate, a high-resolution global dataset of monthly climate and climatic water balance from 1958–2015. Scientific Data, 5, 1–12. https://doi.org/10.1038/sdata.2017.191
Abatzoglou, J. T., & Kolden, C. A. (2013). Relationships between climate and macroscale area burned in the western United States. International Journal of Wildland Fire, 22(7), 1003–1020. https://doi. org/10.1071/WF13019
Abella, S. R., Embrey, T. M., Schmid, S. M., & Prengaman, K. A. (2012). Biophysical correlates with the distribution of the invasive annual red brome (Bromus rubens) on a Mojave Desert landscape. Invasive Plant Science and Management, 5(1), 47–56. https://doi.org/10.1614/ ipsm-d-11-00030.1
Allington, G. R. H., Koons, D. N., Morgan Ernest, S. K., Schutzenhofer, M. R., & Valone, T. J. (2013). Niche opportunities and invasion dynamics in a desert annual community. Ecology Letters, 16(2), 158–166. https://doi.org/10.1111/ele.12023
Antill, T. M., Naeth, M. A., Bork, E. W., & Westhaver, A. L. (2012). Russian thistle (Salsola tragus L.) control on bighorn sheep winter ranges in Jasper National Park. Natural Areas Journal, 32(4), 391–397. https://doi.org/10.3375/043.032.0407
Balch, J. K., Braden, B. A., D’Antonio, C. M., & Gómez-Dans, J. (2013). Introduced annual grass increases regional fire activity across the arid western USA (1980–2009). Global Change Biology, 19(1), 173–183. https://doi.org/10.1111/gcb.12046
Barga, S. C., Dilts, T. E., & Leger, E. A. (2018). Contrasting climate niches among co-occurring subdominant forbs of the sagebrush steppe. Diversity and Distributions, 24(9), 1291–1307. https://doi.org/10.1111/ddi.12764
Barve, N., Barve, V., Jiménez-Valverde, A., Lira-Noriega, A., Maher, S. P., Peterson, A. T., Soberón, J., & Villalobos, F. (2011). The crucial role of the accessible area in ecological niche modeling and species distribution modeling. Ecological Modelling, 222(11), 1810–1819. https://doi.org/10.1016/j.ecolmodel.2011.02.011
Beckie, H. J., & Francis, A. (2009). The Biology of Canadian Weeds. 65. Salsola tragus L. (updated). Canadian Journal of Plant Science, 89(4), 775–789. https://doi.org/10.4141/CJPS08181
Bishop, T. B., Munson, S., Gill, R. A., Belnap, J., Petersen, S. L., & St. Clair, S. B. (2019). Spatiotemporal patterns of cheatgrass invasion in Colorado Plateau National Parks. Landscape Ecology, 34(4), 925–941. https://doi.org/10.1007/s10180-019-00817-8
BLM (2020a). BLM AIM LMF point features. U.S. Department of Interior Bureau of Land Management (BLM).
BLM (2020b). Terrestrial AIM Data 2.0. U.S. Department of Interior Bureau of Land Management (BLM).
Bradford, J. B., & Lauenroth, W. K. (2006). Controls over invasion of Bromus tectorum: The importance of climate, soil disturbance and seed availability. Journal of Vegetation Science, 17(6), 693.
Bradley, B. A. (2013). Distribution models of invasive plants over-estimate potential impact. Biological Invasions, 15(7), 1417–1429. https://doi. org/10.1007/s10530-012-0380-0
Bradley, B. A., Curtis, C. A., & Chambers, J. C. (2016). Bromus response to climate and projected changes with climate change. In M. J. Germino, J. C. Chambers, & C. S. Brown (Eds.), Exotic brome-grasses in arid and semiarid ecosystems of the western US (pp. 257–274). Springer.
Brandt, C. A., & Rickard, W. H. (1994). Alien taxa in the North American shrub-steppe four decades after cessation of livestock grazing and cultivation agriculture. Biological Conservation, 68(2), 95–105. https://doi.org/10.1016/0006-3207(94)90339-5
Brooks, M. L., Brown, C. S., Chambers, J. C., D’Antonio, C. M., Keeley, J. E., & Belnap, J. (2016). Exotic annual bromus invasions: Comparisons among species and ecoregions in the Western United States. In M. J. Germino (Ed.), Exotic brome-grasses in arid and semiarid ecosystems of the Western US (pp. 11–59). Springer International Publishing Switzerland. https://doi.org/10.1007/978-3-319-24930-8_2
Bykova, O., & Sage, R. F. (2012). Winter cold tolerance and the geographic range separation of Bromus tectorum and Bromus rubens, two severe invasive species in North America. Global Change Biology, 18(12), 3654–3663. https://doi.org/10.1111/gcb.12003
Carrington, A. M., Fieguth, P. W., Qazi, H., Holzinger, A., Chen, H. H., Mayr, F., & Manuel, D. G. (2020). A new concordant partial AUC and partial c statistic for imbalanced data in the evaluation of machine learning algorithms. BMC Medical Informatics and Decision Making, 6, 1–12. https://doi.org/10.1186/s12911-019-0104-6
Chambers, J. C., Allen, C. R., & Cushman, S. A. (2019). Operationalizing ecological resilience concepts for managing species and ecosystems at risk. Frontiers in Ecology and Evolution, 7, 1–27. https://doi.org/10.3389/fevo.2019.00241
Chambers, J. C., Bradley, B. A., Brown, C. S., D’Antonio, C., Germino, M. J., Grace, J. B., Hardegree, S. P., Miller, R. F., & Pyke, D. A. (2014). Resilience to stress and disturbance, and resistance to Bromus tectorum L. invasion in cold desert shrublands of western North America. Ecosystems, 17(2), 360–375. https://doi.org/10.1007/s10021-013-9725-5
Chambers, J. C., Miller, R. F., Board, D. I., Pyke, D. A., Roundy, B. A., Grace, J. B., Schupp, E. W., & Tausch, R. J. (2014). Resilience and resistance of sagebrush ecosystems: implications for state and transition models
derived from LiDAR and high spatial resolution imagery. Remote Sensing of Environment, 150, 120-131. https://doi.org/10.1016/j.rse.2014.04.026

Van Kleunen, M., Weber, E., & Fischer, M. (2010). A meta-analysis of trait differences between invasive and non-invasive plant species. Ecology Letters, 13(2), 235-245. https://doi.org/10.1111/j.1461-0248.2009.01418.x

Welk, E. (2004). Constraints in range predictions of invasive plant species due to non-equilibrium distribution patterns: Purple loosestrife (Lythrum salicaria) in North America. Ecological Modelling, 179(4), 551-567. https://doi.org/10.1016/j.ecolmodel.2004.04.020

Wilcox, B. P., Turnbull, L., Young, M. H., Williams, C. J., Ravi, S., Seyfried, M. S., Bowling, D. R., Scott, R. L., Germino, M. J., Caldwell, T. G., & Wainwright, J. (2012). Invasion of shrublands by exotic grasses: Ecohydrological consequences in cold versus warm deserts. Ecohydrology, 5(2), 160-173. https://doi.org/10.1002/eco.247

Wilson, M. F. J., O'Connell, B., Brown, C., Guinan, J. C., & Grehan, A. J. (2007). Multiscale terrain analysis of multibeam bathymetry data for habitat mapping on the continental slope. Marine Geodesy, 30(1-2), 3-35. https://doi.org/10.1080/01490410701295962

Wunderlich, R. F., Lin, Y. P., Anthony, J., & Petway, J. R. (2019). Two alternative evaluation metrics to replace the true skill statistic in the assessment of species distribution models. Nature Conservation, 35, 97-116. https://doi.org/10.3897/natureconservation.35.33918

Zurell, D., Franklin, J., König, C., Bouchet, P. J., Dormann, C. F., Elith, J., Fandos, G., Feng, X., Guiller-aroita, G., Guisan, A., Lahoz-Monfort, J. J., Leitão, P. J., Park, D. S., Peterson, A. T., Rapacciuolo, G., Schmatz, D. R., Schröder, B., Serra-Diaz, J. M., Thuiller, W., ... Merow, C. (2020). A standard protocol for reporting species distribution models. Ecography, 43(9), 1261-1277. https://doi.org/10.1111/ecog.04960

BIOSKETCH
The authors are research scientists with the USDA Forest Service Rocky Mountain Research Station and the University of Nevada, Reno. Their research uses a collaborative, interdisciplinary approach to increase understanding of dryland ecosystems and to develop approaches for maintaining and restoring their integrity.

Author contributions: A.U. and J.C. conceived of the ideas; J.B., C.P, D.M. and A.U. collected and processed the data; D.M. analysed the data; and D.M., A.U. and J.C. led the writing.

SUPPORTING INFORMATION
Additional supporting information may be found online in the Supporting Information section.

How to cite this article: McMahon DE, Urza AK, Brown JL, Phelan C, Chambers JC. Modelling species distributions and environmental suitability highlights risk of plant invasions in western United States. Divers Distrib. 2021;27:710–728. https://doi.org/10.1111/ddi.13232