Materials and Methods

Figure 1 provides an overview of the workflow. Using time-lapse images captured in the northern Japanese Alps, we quantified Sasa expansion within the study area and measured the snowmelt day of year (DOY). We quantified expansion by comparing vegetation classification maps derived from 2012 and 2021 images. Snowmelt DOY was measured from images spanning 2011–2021. Using the resulting Sasa distribution and its change, snowmelt DOY, and topographic features derived from a digital elevation model, we built HSMs to predict Sasa distribution and expansion hotspots. Finally, we input future snowmelt DOY projected from our observations into the HSMs to generate detailed forecasts of future Sasa invasion.

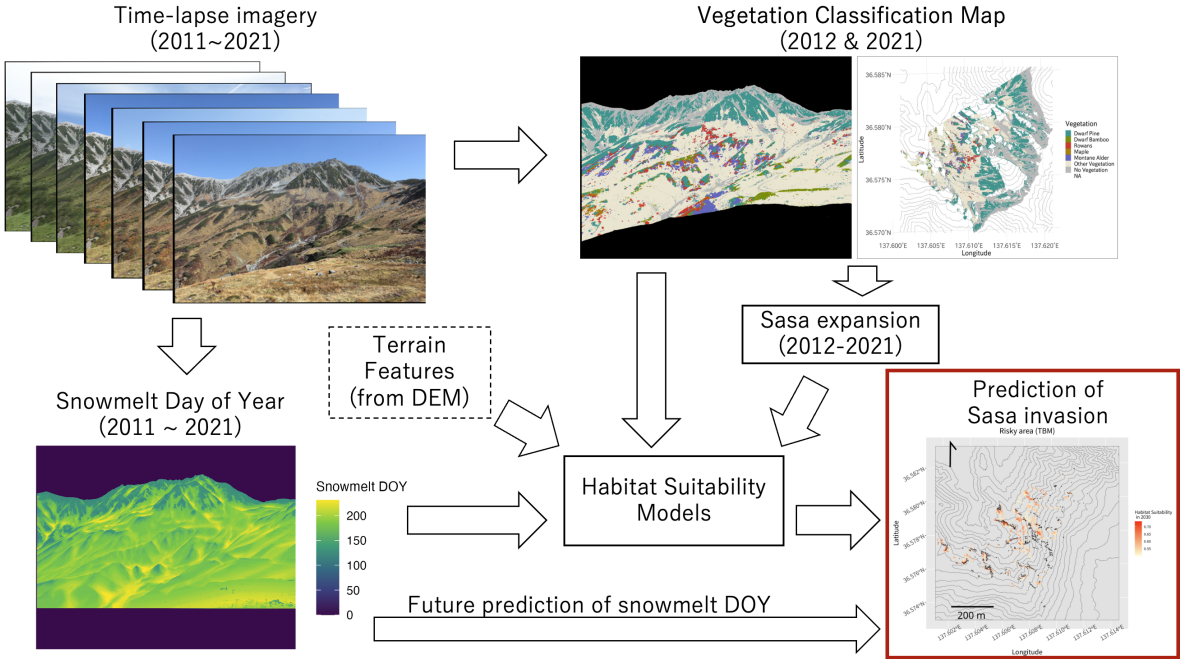

Figure 1: Overview of the proposed method.
Vegetation classification results in 2012 (left) and 2021 (right). The expansion was quantified by comparing the spatial extent of *Sasa* distribution between the two vegetation classification maps.

Image acquisition

To monitor the distribution of *Sasa*, we used images captured by a digital time-lapse camera (Canon EOS 5D Mark II) installed by the National Institute for Environmental Studies on a mountain lodge Murodo-sanso (about 2,450 m a.s.l.) in the northern Japanese Alps (Mt. Tateyama). Since 2010, the camera has recorded snowmelt and vegetation changes on

the western slope of Mt. Tateyama (2,350–3,015 m), taking photographs every hour from 6:00 to 19:00 between April and November. For details of installation and operation, see @Okamoto2024RSEC. In this study, we used images taken in September–October of 2012 and 2021.

One major advantage of time-lapse cameras is their ability to capture daily changes in snow cover. Because earlier snowmelt is considered a primary driver of *Sasa* expansion (@Kudo2011EcoEvo), high-frequency and high-resolution data on snowmelt timing are crucial for understanding and predicting *Sasa* expansion. We therefore also used photographs taken from April to August between 2011 and 2021 to calculate the snowmelt DOY for each pixel. Snowmelt detection followed a workflow based on Otsu's binarization method (@otsu_binarization) as described in @IdeOguma2013EcolInfom. The extracted snowmelt DOYs were later used in the HSM.

Automated production of vegetation classification maps

Vegetation classification maps were produced using the method developed by @Okamoto2024RSEC, consisting of three steps: image alignment, vegetation classification, and automated georectification.

- **Image alignment**: All images from 2012 and 2021 were aligned to enable precise comparisons. Following @Okamoto2024RSEC, we used feature-based automatic alignment.

- **Vegetation classification**: Temporal changes in leaf color during autumn were analyzed using a recurrent neural network (RNN), classifying each pixel into seven categories: Dwarf Pine (*Pinus pumila*), Dwarf Bamboo (*Sasa* spp.), Rowans (*Sorbus sambucifolia*, *S. matsumurana*), Maple (*Acer tschonoskii*), Montane Alder (*Alnus viridis* subsp. *maximowiczii*), Other Vegetation (alpine shrubs and herbaceous plants), and No Vegetation.

- **Automated georectification**: Using the method proposed in @Okamoto2024RSEC, we automatically extracted ground control points (GCPs) from existing orthorectified aerial photographs and a digital elevation model (DEM), estimated camera parameters (orientation, focal length, and lens distortion), and transformed the images into georeferenced data at 1 m resolution. The aerial photographs and DEM used for georectification were identical to those in @Okamoto2024RSEC.

Quantification of *Sasa* expansion

Vegetation classification maps from 2012 and 2021 were overlaid to detect changes in *Sasa* distribution over nine years (Figure 2). The area occupied by *Sasa* in each year was calculated, and expansion was quantified by comparing the two classification maps. Because encroachment

and growth of shrubs can render *Sasa* undetectable in time-lapse imagery (i.e., *Sasa* persisting under shrub canopies), we defined the primary expansion metric as the area of transitions from non-*Sasa* to *Sasa* (i.e., expansion area). Areas of decrease (from *Sasa* to non-*Sasa*) were recorded as reference information but not used as the main metric to avoid conservative underestimation of expansion.

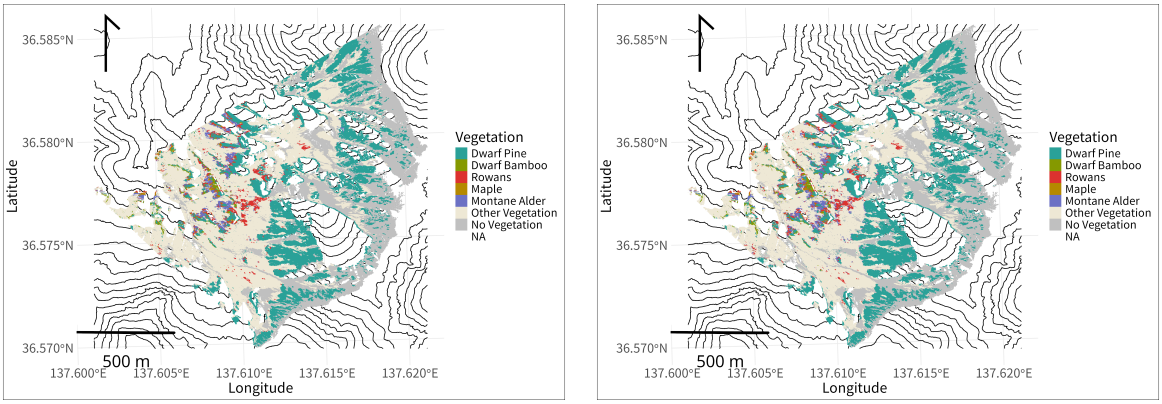

Figure 2: Vegetation classification results in 2012 and 2021

Habitat suitability modeling

To predict future *Sasa* expansion, we constructed HSMs to identify areas likely to become suitable under changes in snowmelt timing. *Sasa* forms complex rhizome systems and large clonal networks (@Suyama2000MolEcol). While it generally expands via rhizome elongation, seed dispersal occurs after mass flowering and has also been reported during partial flowering events (@Miyazaki2009JPlantRes; @Mizuki2014PLOS; @Kudo2011EcoEvo). Thus, expansion can potentially occur even at sites distant from current distributions.

Considering these dispersal characteristics, we developed two complementary models. The **Topography-Based Model (TBM)** incorporates only topographic variables such as elevation, slope, and snowmelt DOY, and represents potential expansion areas assuming that *Sasa* can disperse via seeds. In contrast, the **Topography–Distance Model (TDM)** additionally includes the distance from the 2012 *Sasa* distribution, thereby accounting for the limited rhizome-based spread observed in most situations. By comparing the two models, we can distinguish between the broader potential niche of *Sasa* under seed dispersal and the more restricted realized expansion expected from clonal growth, providing a more realistic range of future scenarios.

Explanatory variables

Both models used elevation, slope, aspect, roughness, Terrain Ruggedness Index (TRI), Topographic Position Index (TPI), and snowmelt DOY as explanatory variables. These features were calculated from the 5 m DEM provided by the Geospatial Information Authority of Japan and resampled to 1 m resolution to match the vegetation maps.

Snowmelt DOYs were calculated from time-lapse imagery. Although annual fluctuations occurred, snowmelt timing advanced over the past decade. To predict how *Sasa* distribution would change if this trend continues, we fitted linear regressions to the per-pixel snowmelt DOYs in every year from 2011 to 2021 and used the predicted 2021 values as explanatory variables. The mean regression coefficient was −0.86, indicating an advance of 8.6 days over 10 years. We then predicted 2030 snowmelt DOYs for each pixel using the same regressions for future projections. In the TDM, distance from the 2012 *Sasa* distribution was included as an additional variable.

Target variable

The TBM was trained to predict the *Sasa* distribution in 2021. Because many *Sasa* patches present in 2012 remained present in 2021, pixels at 0 m distance from the 2012 distribution were excluded from the TDM. Accordingly, the TDM was designed to predict newly colonized locations between 2012 and 2021.

Data sampling and splitting

Data sampling

Because the *Sasa* distribution is relatively limited compared to other vegetation types (Figure 2), the dataset is imbalanced. Then presence data were retained at 1 m resolution, while absence data were resampled to 5 m resolution. Areas above 2,560 m, where *Sasa* is absent, were excluded.

Initial splitting

To evaluate model performance, we split the data into 80% training and 20% testing sets. Given strong spatial autocorrelation in the explanatory variables, we used spatial block splitting with the R packages `spatialsample` [@spatialsample] and `tidysdm` [@tidysdm] to mitigate overfitting (Figure 3, left).

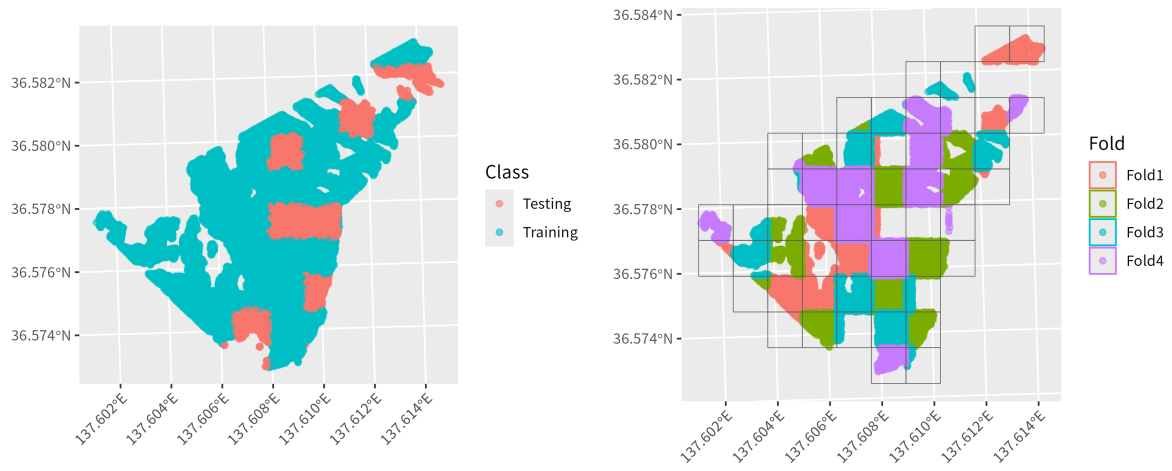

Figure 3: Visualizations of spatial sampling

Cross-validation

We performed four-fold cross-validation for hyperparameter tuning and ensemble construction, applying the same spatial block splitting (Figure 3, right).

Model preparation

Model training and evaluation

We applied four classifiers—Gradient Boosted Trees (GBT; @Friedman2001GBM), Maximum Entropy (MaxEnt; @Philips2006MaxEnt), Random Forest (RF; @Breiman2001RandForest), and Generalized Additive Model (GAM). For all but GAM, hyperparameters were tuned via grid search with up to 18 trials. Performance was evaluated with True Skill Statistics (TSS; @Allouche2006TSS; see Equation 1), ranging from 0 to 1, with higher values indicating better accuracy. (TP), (FP), and (FN) stand for the numbers of true positives, false positives, and false negatives, respectively.

$$
\begin{aligned}
recall &= \frac{TP}{TP + FN} \\
specificity &= \frac{TN}{TN + FP} \\
TSS &= recall + specificity - 1
\end{aligned}
\tag{1}
$$

Model ensembling

We built ensemble models using the `blend_predictions` function in the R package `stacks` (@RCranStack), which estimates blending coefficients via LASSO to maximize cross-validated performance. The ensemble TSS was subsequently measured on the test dataset.

Variable importance computation

For both TBM and TDM, variable importance was assessed using permutation loss, defined as the decrease in TSS when each variable is randomly permuted, thereby quantifying its contribution to predictive accuracy.

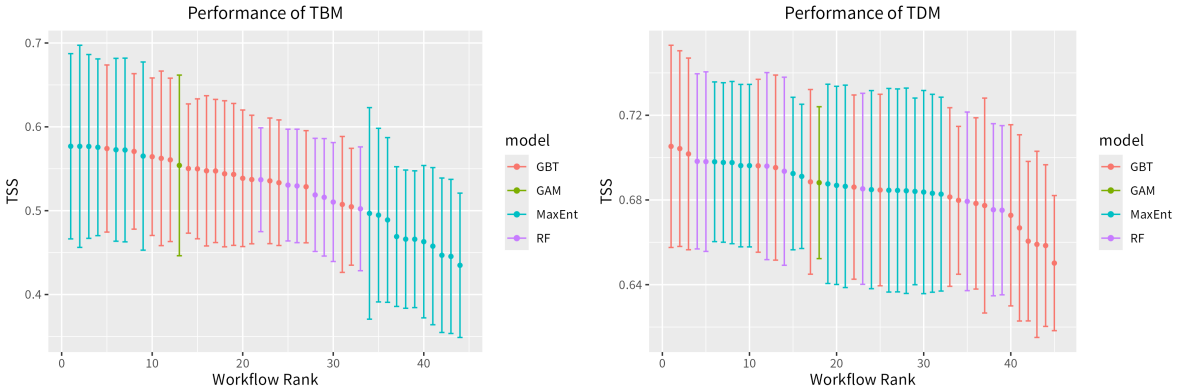

Figure 4: Performance metrics of Habitat Suitability Models (HSMs) obtained through hyper-parameter tuning during cross-validation.

Future prediction and definition of risky areas

Using the ensemble model of TBM and TDM, we predicted *Sasa* distribution for 2030 based on the estimated snowmelt DOYs (see Section). In this framework, the prediction of TBM is interpreted as representing areas that could become suitable if *Sasa* disperses by seed. Under such conditions, *Sasa* is more likely to invade grasslands, which in our classification scheme correspond to "Other Vegetation." Therefore, for conservation applications, we defined "risky areas" as cells classified as "Other Vegetation" in 2021 for which the TBM-predicted HS in 2030 exceeds 0.5, and extracted their spatial distribution.