Title

Deep learning to extract the meteorological by-catch of wildlife cameras

Running title

Meteorological by-catch of wildlife cameras

List of authors

Jamie Alison[1]

Stephanie Payne[2]

Jake M. Alexander[3]

Anne D. Bjorkman[4,5]

Vincent Ralph Clark[6]

Onalenna Gwate[6]

Maria Huntsaar[7,8]

Evelin Iseli[3]

Jonathan Lenoir[9]

Hjalte Mads Rosenstand Mann[1,10]

Sandy-Lynn Steenhuisen[2]

Toke Thomas Høye[1,10]

Institutional affiliations

[1]Department of Ecoscience, Aarhus University, Denmark

[2]Afromontane Research Unit and Department of Plant Sciences, University of the Free State, South Africa

[3]Institute of Integrative Biology, ETH Zurich, Switzerland

[4]Department of Biological and Environmental Sciences, University of Gothenburg, Sweden

[5]Gothenburg Global Biodiversity Centre, Gothenburg, Sweden

[6]Afromontane Research Unit and Department of Geography, University of the Free State, South Africa

[7]Arctic Biology Department, The University Centre in Svalbard (UNIS), Norway

[8]Department of Arctic and Marine Biology, The Arctic University of Norway (UiT), Norway

[9]UMR CNRS 7058 “Ecologie et Dynamique des Systèmes Anthropisés” (EDYSAN), Université de Picardie Jules Verne, Amiens, France

[10]Arctic Research Centre, Aarhus University, Denmark

Abstract

Microclimate – proximal climatic variation at scales of meters and minutes – can exacerbate or mitigate the impacts of climate change on biodiversity. However, most microclimate studies are temperature-centric, and do not consider meteorological factors such as sunshine, hail and snow. Meanwhile, remote cameras have become a primary tool to monitor wild plants and animals, even at micro-scales, with deep learning tools to rapidly convert images into ecological data. However, deep learning applications for wildlife imagery have focused exclusively on living subjects. Here, we identify an overlooked opportunity to extract latent, yet ecologically relevant, meteorological information.

We produce an annotated image dataset of micrometeorological conditions across 49 wildlife cameras in South Africa's Maloti-Drakensberg and the Swiss Alps. We train ensemble deep learning models to classify conditions as overcast, sunshine, hail or snow. Our best model achieves 91.7% accuracy on test cameras not seen during training. Furthermore, we show how effective accuracy is raised to 96% by disregarding 14.1% of classifications where ensemble member models did not reach a consensus. Unlike previous studies we test model performance in remote and novel locations, distinguishing overcast and sunny conditions in Svalbard, Norway with 79.3% accuracy (93.9% consensus accuracy).

Our model rapidly classifies sunshine, snow and hail in almost 2 million unlabeled images. Resulting micrometeorological data illustrated common seasonal patterns of summer hailstorms and autumn snowfalls across mountains in the northern and southern hemispheres. However, daily patterns of sunshine and shade diverged between sites, impacting daily temperature cycles. Crucially, we leverage micrometeorological data to demonstrate that (1) experimental warming using open-top chambers shortens early snow events in autumn, and (2) image-derived sunshine marginally outperforms sensor-derived temperature when predicting bumblebee foraging. These methods generate novel micrometeorological variables in synchrony with biological recordings, enabling new insights from an increasingly global network of wildlife cameras.

Keywords

alpine ecology, automated monitoring, bees, micrometeorology, proximal sensing, snow melt, time-lapse photography, *Trifolium pratense*

Introduction

Climate change is redistributing species in space and time, with profound impacts on ecosystem function and human wellbeing (Pecl et al., 2017). While biodiversity impacts of climate change are usually reported at coarse spatial resolutions across large spatial extents (Lenoir et al., 2020), species often respond to climate at much finer scales (Lembrechts et al., 2019; Maclean & Early, 2023; Potter et al., 2013). Furthermore, studies of the biodiversity impacts of climate change are traditionally temperature-centric (Antão et al., 2020). As the study of microclimate has expanded, this temperature-centricity has remained (Maclean et al., 2021), due in part to the availability of inexpensive and easy-to-use temperature loggers (Lembrechts et al., 2020). Still, non-temperature aspects of climate and meteorology have profound, fine-scale impacts on species and communities. Solar radiation (i.e., sunshine) underpins ambient temperature, but also directly impacts plant growth and animal behavior (Möhl et al., 2020; Roales et al., 2013; Valladares et al., 2016). Furthermore, snow cover severely constrains the onset of plants' growing and flowering seasons (Möhl et al., 2022), impacting plant-pollinator interactions (Gillespie et al., 2016; Gillespie & Cooper, 2022). Hail receives very little attention in ecological research, yet presents a direct physical threat that can devastate poorly-adapted plant species (Fernandes et al., 2011). As the frequency of extreme weather events increases (IPCC, 2023), it is vital that sensor networks capture not only temperature, but also a range of other fine-scale meteorological variables.

Meanwhile, novel technology is revolutionizing the monitoring of ecological communities, generating data with unprecedented temporal continuity and resolution (August et al., 2015; Besson et al., 2022). Wildlife cameras – *in situ* autonomous cameras that record wildlife – are particularly promising. Camera traps are an established method to monitor large animals (Burton et al., 2015), but wildlife cameras are now regularly deployed to study small mammals and birds (Ortmann & Johnson, 2021), insects and other arthropods (Høye et al., 2021), and plants (Katal et al., 2022). For vegetation, the "PhenoCam" approach has gained traction, capturing community-level characteristics such as plant productivity (Brown et al., 2016). However, the volume of images from wildlife cameras has proven difficult to manage, so algorithms are being developed to automatically extract ecological data (Høye et al., 2021; Tuia et al., 2022). Deep learning models are a prevalent family of algorithms used to detect, classify and track animals in images (Norouzzadeh et al., 2018). For plants, algorithms can be trained to flag phenological events such as the onset of budburst or flowering (Katal et al., 2022), or to detect and track individual flowers (Mann et al., 2022).

The consistency of image-based monitoring allows for incidental recording of non-target biota, known as "by-catch". For example, a camera trap network intended to study the Eurasian lynx in the Swiss Jura mountains has proven useful to explore habitat use by wild boar and roe deer (Wevers et al., 2021). However, it is increasingly recognized that wildlife cameras also capture non-target abiotic information, including meteorological conditions not easily captured with alternative sensors (Hofmeester et al., 2020). Furthermore, capturing both biotic and abiotic data with a single sensor ensures that they are measured simultaneously, at equivalent spatial and temporal scales. Several studies have manually extracted the presence, cover or depth of snow in ecological imagery (Lumbrazo et al., 2022). Furthermore, some studies automate the quantification of snow in PhenoCam images (Caparó Bellido & Rundquist, 2021; Julitta et al., 2014), or the classification of frost (Noh et al., 2021). However, the use of a single, efficient classifier to simultaneously detect sunshine and frozen precipitation in wildlife images has not been explored.

Working with wildlife imagery from matching experimental sites in mountains in the northern (Switzerland; CH) and southern (South Africa; ZA) hemispheres, we train ensemble deep learning models to detect micrometeorological conditions of sunshine, hail, and snow. Our objectives are as

follows: (1) Build a dataset and train deep learning classifiers to derive micrometeorological conditions in images; (2) evaluate different data-splitting and ensemble approaches to maximize model performance with out-of-sample and out-of-distribution test datasets; and (3) demonstrate the application of image-derived micrometeorological variables for ecological research. For this last objective, we classify conditions in almost 2 million unlabeled images in CH and ZA. Then we (i) use snow in ZA images to determine the effects of experimental warming on snowmelt, and (ii) use sunshine in CH images to explore the relative importance of sunshine and ambient temperature for bumblebees foraging at high elevations. Our approach efficiently attaches micrometeorological data to remote biodiversity observations. This is particularly useful for biodiversity monitoring in mountains, which are characterised by high levels of endemism, complex topography, and microclimatic heterogeneity (Spehn & Körner, 2005).

Methods

Study area

Imaging was carried out over 49 1×1m montane grassland plots across four replicated experimental field sites. In the Calanda massif in the Alps in Switzerland (CH), 24 plots were imaged from June 2021 to October 2021, of which eight were located at a low elevation site (46.869266° N 9.490098° E; 1,440m elevation) and 12 were located at a high elevation site (46.887824° N 9.489510° E; 2,000m). In the Sentinel region of the Maloti-Drakensberg in South Africa (ZA), 25 subplots were imaged from November 2021 to April 2022, of which five were located at a low elevation site (28.679532° S 28.894816° E; 2,200m) and 20 were located at a high elevation site (28.754951° S 28.866981° E; 3,060m). The grassland plots we imaged comprised native vegetation, but half of the plots additionally had small plant specimens, of species typical of lower elevations, introduced within them. Half of the high site plots in ZA were also exposed to open top chamber (OTC) warming of approximately 2°C.

Sensor deployment

A Wingscapes TimelapseCam Pro camera (with LED flash) was mounted on a steel frame over each plot, ca. 62cm from the ground, pointing towards the ground. Each camera captured an area of ca. 30×17cm. All cameras were 'continuous', recording images day and night at 5-minute intervals, except for 16 'focused' cameras in CH which recorded at 1-minute intervals between 12.00–15.00 and 01.00–03.00 (Alison et al., 2022). Each camera required eight AA lithium batteries and a 128GB SD card, replaced approximately every two months.

All cameras were equipped with on-board temperature sensors, recording temperature at ca. 62cm above ground with every photograph (either at 5- or 1-minute intervals depending on the camera). At the high elevation sites, we also deployed TMS4 microclimate loggers to record temperature every 15 minutes at -8 cm, 0 cm and 15 cm above ground. We deployed loggers in blocks of six, with three blocks at the CH high site and one block at the ZA high site. Each block included two loggers in vegetated plots with OTC warming, two in vegetated plots without warming, and one each in bare soil plots with and without OTC warming. Loggers were not targeted to the same plots as cameras, though some of the same plots were sampled by coincidence.

Image labelling

We sorted a total of 8,953 images into four classes based on micrometeorological conditions: overcast, sunshine, hail or snow (Fig. 1; further details in Appendix S1). First, to generate training data representing the entire recording period, we labelled a systematic sample of 6,205 images. We sampled one image per hour in both ZA and CH, cycling through the continuous cameras. In CH, we included an additional one image per hour cycling through all focused cameras (five hours per day). The systematic sample yielded 1,320 sunshine, 110 snow, 39 hail, and 4,736 overcast images. Second, to generate adequate training data and improve detection of snow and hail, we took a strategic sample of 2,748 images that was informed by the systematic sample. The strategic sampling protocol had three tiers, such that the rarest meteorological events were sampled more intensively (i.e. oversampled; Table S1). The strategic sample yielded an additional set of 266 sunshine, 923 snow, 803 hail and 756 overcast images. CH images were labelled by JA, and ZA images by SP, using VGG Image Annotator (VIA v2.0.11; https://www.robots.ox.ac.uk/~vgg/software/via/).

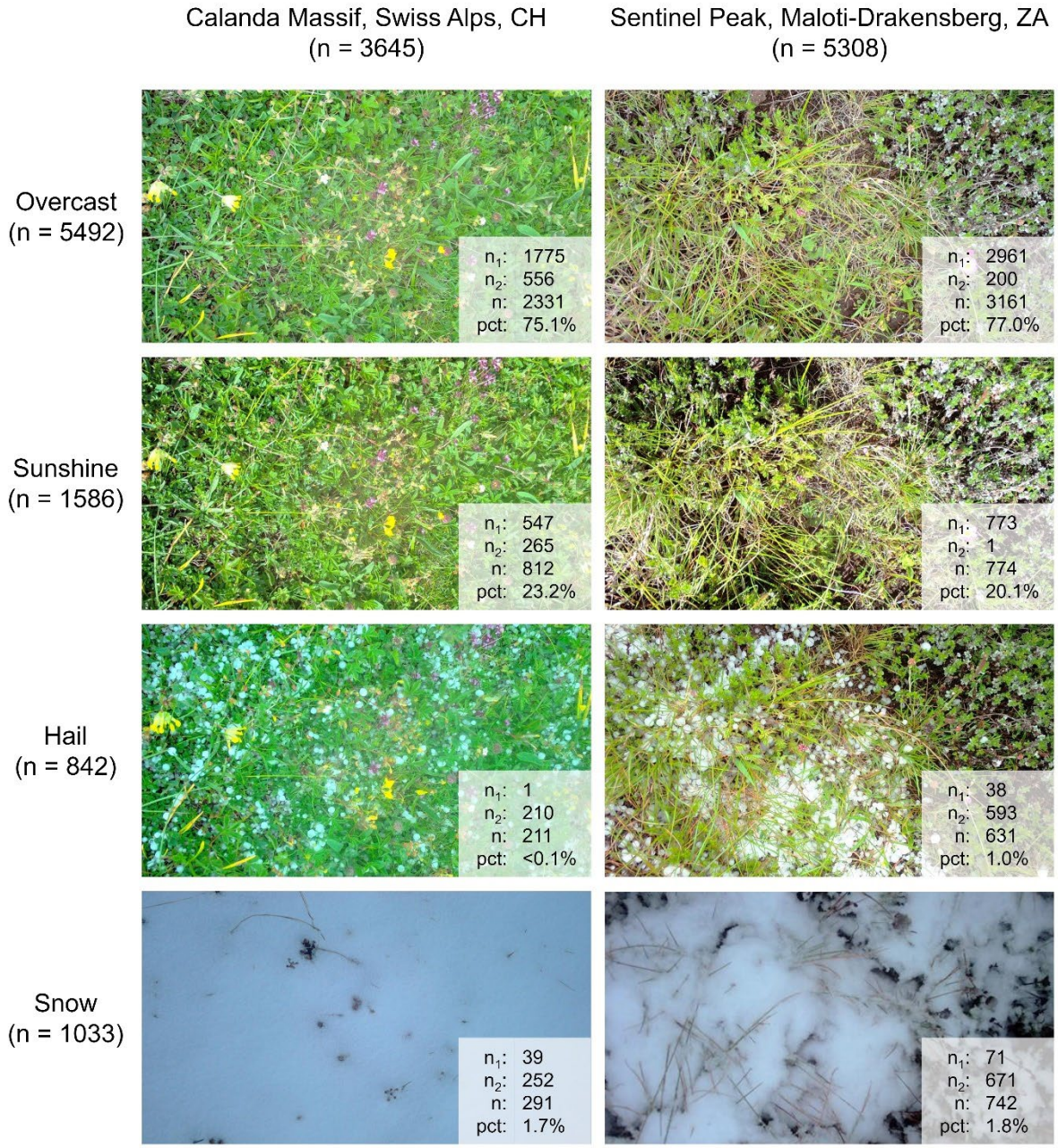

Figure 1. Examples of the four classes of micrometeorological conditions in two representative high elevation plots in Switzerland (CH) and South Africa (ZA). Total numbers of labelled images (n) are shown across classes and regions (including high and low elevation sites). The breakdown of labels across the systematic sample (n_1) and the strategic sample (n_2) is also shown, as is the representation of each condition within the systematic sample for a given region (pct).

Model training

We trained convolutional neural networks (CNNs) to classify conditions in images as overcast, sunshine, hail or snow using the *Keras* python library (Chollet, 2015). Specifically, we fine-tuned the MobileNet network (Howard et al., 2017) pretrained on ImageNet (Russakovsky et al., 2015), representing a lightweight and efficient family of CNNs that has been shown to perform well for

image classification in ecology (Besson et al., 2022; Schneider et al., 2022). To adapt the model to predict just four classes, we removed the top layer of the network and added a custom softmax activation layer with a flattened input. Training images were rescaled to 224×224px and put through an augmentation pipeline to reduce overfitting. Augmentation involved random vertical or horizontal flipping and up to 45° random rotation in any direction. Following preliminary tests, we trained the entire network in two stages using the Adam optimizer and a batch size of 128. First, we trained for five epochs specifying a learning rate of 0.001, to bring models rapidly towards a solution. Second, we trained the entire network for up to 200 epochs, specifying a learning rate of 0.0001. The change in learning rate allowed models to reach an optimal solution while minimizing spurious fluctuations in validation loss. Appendix S2 gives a primer on CNN parameters and concepts.

Model validation

We used model validation with early stopping to minimize overfitting to the training data. We stopped training if validation performance did not improve for 30 epochs and saved the model from the epoch with the best validation performance (concepts explained in Appendix S2). To account for variance related to data splitting for validation, we used cross-validation. Specifically, we carried out a 5-fold cross-validation in which each of five mutually exclusive data folds is iteratively treated as the validation dataset (e.g. Fig. S1, inner validation loop). Additionally, we compared two data splitting methods: “Cis” and “Trans” (Beery et al., 2018, 2020). Cis validation involves random splitting of images from all cameras. Trans validation splits cameras rather than images, such that images from one camera are always constrained to a single fold (Fig. S1). With the Cis validation method, the effective size of the training set is larger, because more camera positions are seen during training. However, the Trans validation method ensures that images from the same camera are not represented in both training and validation data. For both Cis and Trans validation, folds were stratified with respect to the four study sites (Wieczorek et al., 2022).

Model aggregation

We produced ensembles of sets of five “member models”, representing the five iterations of each 5-fold validation loop, in a procedure called cross-validation aggregation (crogging; Barrow & Crone, 2013; see Appendix S3 for further details). The crogging procedure produces ensemble models in which each observation contributes to both model training (in all but one member model) and model validation by early stopping (in one member model). Averaging across models which use distinct data folds for validation can improve generalization performance and model stability (Barrow & Crone, 2016; Krogh & Vedelsby, 1994). Ensemble models were produced under both Cis and Trans validation methods. For each 5-fold validation loop, five Cis member models and five Trans member models were aggregated into ensembles by unweighted averaging of output softmax probabilities.

Model testing and deployment

We aimed for models that would transfer to (1) novel camera positions within our sites and (2) novel sites with a similar recording protocol. To test transferability to novel positions within our sites, each 5-fold validation loop was nested within an outer 6-fold test loop (e.g. Fig. S1). In other words, we used a nested cross-validation or “double-cross” to obtain a robust and unbiased measure of model generalization (Cawley & Talbot, 2010). We randomly split 49 cameras into six folds that were stratified across the four sites (Wieczorek et al., 2022). During each test loop iteration, one of six folds was withheld as a test dataset. We compared model accuracy and macro-F1 (the mean of class

F1 scores) following Eq. 1-4 where: *TP* = true positives; *TN* = true negatives; *FP* = false positives; *FN* = false negatives; *Pr* = class precision; and *Re* = class recall. Class F1 scores represent class-level performance, and macro-F1 represents overall model performance while accounting for class-imbalance.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (1)$$

$$Pr = \frac{TP}{TP + FP} \quad (2)$$

$$Re = \frac{TP}{TP + FN} \quad (3)$$

$$Class\ F1 = 2\frac{Pr \times Re}{Pr + Re} \quad (4)$$

To test model transferability to novel sites, we trained three production models by incorporating the test data for training and validation. Specifically, we trained a six-fold Trans ensemble and a six-fold Cis ensemble. Novel site performance was then tested using images from Svalbard, Norway. In a separate study of pollination of *Silene acaulis*, six plots in Bjørndalen (78.21660° N 15.33280° E; 40m elevation) were imaged at 1-minute intervals in June and July 2020 using the same imaging methods. However, some images were cropped slightly to zoom in on *S. acaulis* during annotation. The first 11 images of every hour of every day were labelled for sunshine (n = 957) and overcast (n = 6222) conditions by MH.

Finally, we deployed the Trans ensemble model to predict micrometeorological conditions across the full set of 1,934,044 images taken across all four sites in CH and ZA. We validated our time-series of micrometeorological conditions against temperature data from TMS4 loggers (Wild et al., 2019). We then analyzed micrometeorological conditions to explore (1) effects of OTC warmed and unwarmed treatments on snow cover in ZA, and (2) effects of sensor-derived temperature and image-derived sunshine on *Bombus* visitation to *Trifolium pratense* in CH (using data previously published by Alison et al., 2022). For (2), we calculated the daytime degree days recorded by on-board temperature sensors (sum of mean daily temperatures above 0°C) and sunshine days (sum of mean daily proportional sunshine) during flowering of each inflorescence. We compared AIC of linear models predicting ln(number of visits) of each inflorescence. To further visualize the niche preferences of *Bombus* bees, we calculated mean temperature and proportional sunshine within an hour either side of each *Bombus* foraging event. This was overlaid on the distribution of temperature and sunshine during daytime hours in which 32 *T. pratense* inflorescences were flowering.

Results & Discussion

Wildlife cameras capture details about species' abiotic environments, and not just their behaviors, life cycles, and interactions (Hofmeester et al., 2020). We find that conditions such as sunshine, snow and hail can be readily and automatically extracted from wildlife imagery. Furthermore, we show how such micrometeorological data are complementary to sensor-derived temperature data.

Model performance

Ensemble models were highly transferable to novel camera positions, consistently achieving higher accuracy and macro-F1 than member models (Table 1). Ensemble models also outperformed full models, which used all data for training and none for validation, by around 1%. Beyond performance benefits, member models within a given ensemble disagree on 13-14% of predictions (Table 1) and this disagreement forms a useful measure of uncertainty; by omitting non-consensus predictions as uncertain, we raised the effective accuracy of our most accurate ensemble to 96%. Differences between models trained using the Cis and Trans validation methods were negligible, while cross-validation revealed considerable variation when models of the same type were trained and tested on different folds of the data (Table 1, see Appendix S3 for further discussion on validation approaches and ensembling). Our models classified overcast and sunshine conditions more successfully than snow or hail (Fig. 2, Fig. S2), perhaps reflecting the oversampling of snow and hail events. Our ensemble model based on a Trans validation method misclassified 29% of hail images as overcast, 12% of sunshine as overcast, and 3% of overcast as sunshine (Fig. 2).

The real-world utility of deep learning models hinges on generalization performance, including transferability to novel sites and conditions. When classifying images of *S. acaulis* across six cameras from a totally independent site in Svalbard, Norway, ensemble models using Cis and Trans validation methods distinguished sunshine from overcast conditions with 79.3% and 72.8% accuracy, respectively (model F1: 70.9% and 65.0%). Furthermore, omitting non-consensus predictions as uncertain, we raised the effective accuracy to 94% and 88.4%, respectively. This predictive performance is impressive given that these out-of-distribution images came from a site >1,000km away, at >1,000m lower elevation, with different height and width compared to training images. For detection of sunshine in images, our models show clear potential to generalize to novel sites with a similar recording protocol.

Table 1. Mean (±SD) performance of full models (n=6), ensemble models (n=6) and member models (n=5×6=30) in classification of micrometeorological conditions in out-of-sample test images. Performance metrics include macro-F1, accuracy, and consensus accuracy for ensemble models. Ensemble models are compared to member models under two validation data split methods: "Cis" and "Trans" (see Fig. S1). The consensus accuracy of ensemble models is the accuracy of predictions on which all five member models agree. Rate of consensus is the percentage of test images for which all five member models agree.

Validation split method	Model type	Macro-F1 (%)	Accuracy (%)	Consensus accuracy (%)	Rate of consensus (%)
Cis	Ensemble	**89.25** (2.02)	91.63 (1.96)	95.89 (1.69)	87.13 (1.92)
	Member	88.02 (1.59)	90.72 (1.67)		
Trans	Ensemble	88.82 (3.33)	**91.70** (2.02)	**96.00** (1.55)	85.90 (3.33)
	Member	87.22 (3.02)	90.41 (2.13)		
None	Full	88.00 (2.61)	90.94 (2.21)		

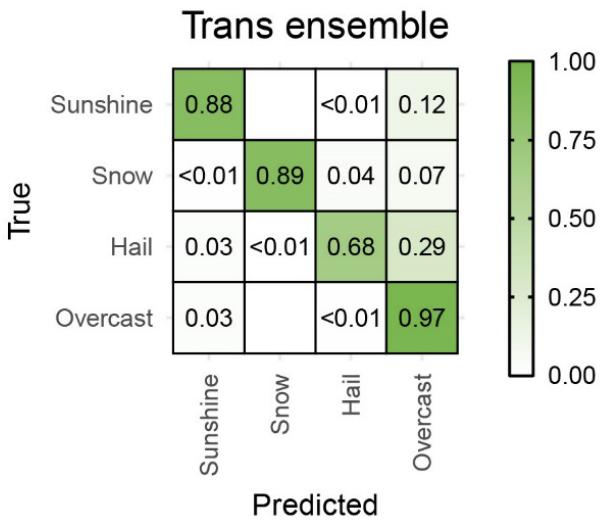

Figure 2. Confusion matrix of predicted micrometeorological conditions across 8,953 images. Six ensemble deep learning models were used to classify images in six distinct folds of hold-out test cameras in a cross-validation process. Data splitting for validation of these models was done using the Trans validation method (see text and Fig. S1 for explanation).

Extensive prediction of micrometeorological data

We predicted micrometeorological conditions in almost two million images across CH and ZA sites. A consensus prediction emerged for 87.7% of images, and these predictions were averaged across cameras over time to generate seasonal (Fig. 3) and diel (Fig. 4) micrometeorological profiles. Seasonal profiles revealed a common seasonal pattern of summer hailstorms and autumn snow across mountains in the northern and southern hemispheres (Fig. 3). As expected, during the day there was a striking match between image-derived sunshine and sensor-derived temperatures (Fig. 3). However, nighttime temperatures also appeared warmer following periods of high daytime cloud cover (i.e., low sunshine before nightfall; Fig. 3), as expected under longwave cloud forcing (Ramanathan et al., 1989). Furthermore, diel sunshine profiles revealed that the ZA high elevation site was characterized by morning sunshine and afternoon shade, with subtle impacts on the diel temperature profile (Fig. 4). Such insights would be difficult to obtain without using the meteorological by-catch of wildlife cameras. Crucially, our models extracted the meteorological by-catch of images very rapidly. On a laptop computer, a member model classified two million images in around 20 hours on an Intel Xeon E5-2697A v4 processor with 8 CPUs @~2.6GHz. Much faster times would be expected if using a GPU.

Effects of experimental warming on snowmelt

We found clear impacts of OTCs on the prevalence and duration of snow cover (Fig. 5). During the first (and longest) recorded snowfall, the snow melted completely after around 2.5 days in warmed plots. This contrasted with unwarmed plots, where snow persisted for up to four days (Fig. 5). Beyond temporal variation, we also captured fine-scale spatial variation in retention of snow cover – especially among warmed plots (Fig. 5, left). Sensor-derived temperatures showed evidence of warming within OTCs, especially during prolonged sunshine, which probably contributed to advanced snowmelt. However, snow was often less prevalent across warmed plots even immediately after snowfall (Fig. 5). This suggests that OTCs also intercept a fraction of falling snow, which could delay snowpack formation.

Few studies have reported effects of OTCs on snow cover in autumn. Several studies have reported effects on snow depth in winter and snowmelt in spring, although observations are often infrequent or anecdotal (Bokhorst et al., 2013; Wipf & Rixen, 2010). Heavy snows are known to accumulate in OTCs over winter, increasing snow depth and insulating plants and soil (Hollister et al., 2023; Rixen et al., 2022); the resulting winter warming of OTCs often exceeds their spring and summer warming effects (Aerts et al., 2004; Bjorkman et al., 2015). Such snow accumulation leads to unpredictable effects of OTCs on snowmelt; it may remain unchanged, be delayed by a week, or be advanced by several weeks (Aerts et al., 2004; Marion et al., 1997; Wipf & Rixen, 2010). Here we show that in early autumn, experimental warming can reduce the duration of snow cover – although this may partly result from interception of snowfall by OTCs. Furthermore, we demonstrate how to efficiently quantify persistence of snow at high spatiotemporal resolutions. This is a vital contribution given the impacts of snow cover on plant and soil ecosystem processes (Möhl et al., 2022), and the prevalence of OTCs in climate change research (Bjorkman et al., 2015; Hollister et al., 2023; Rixen et al., 2022).

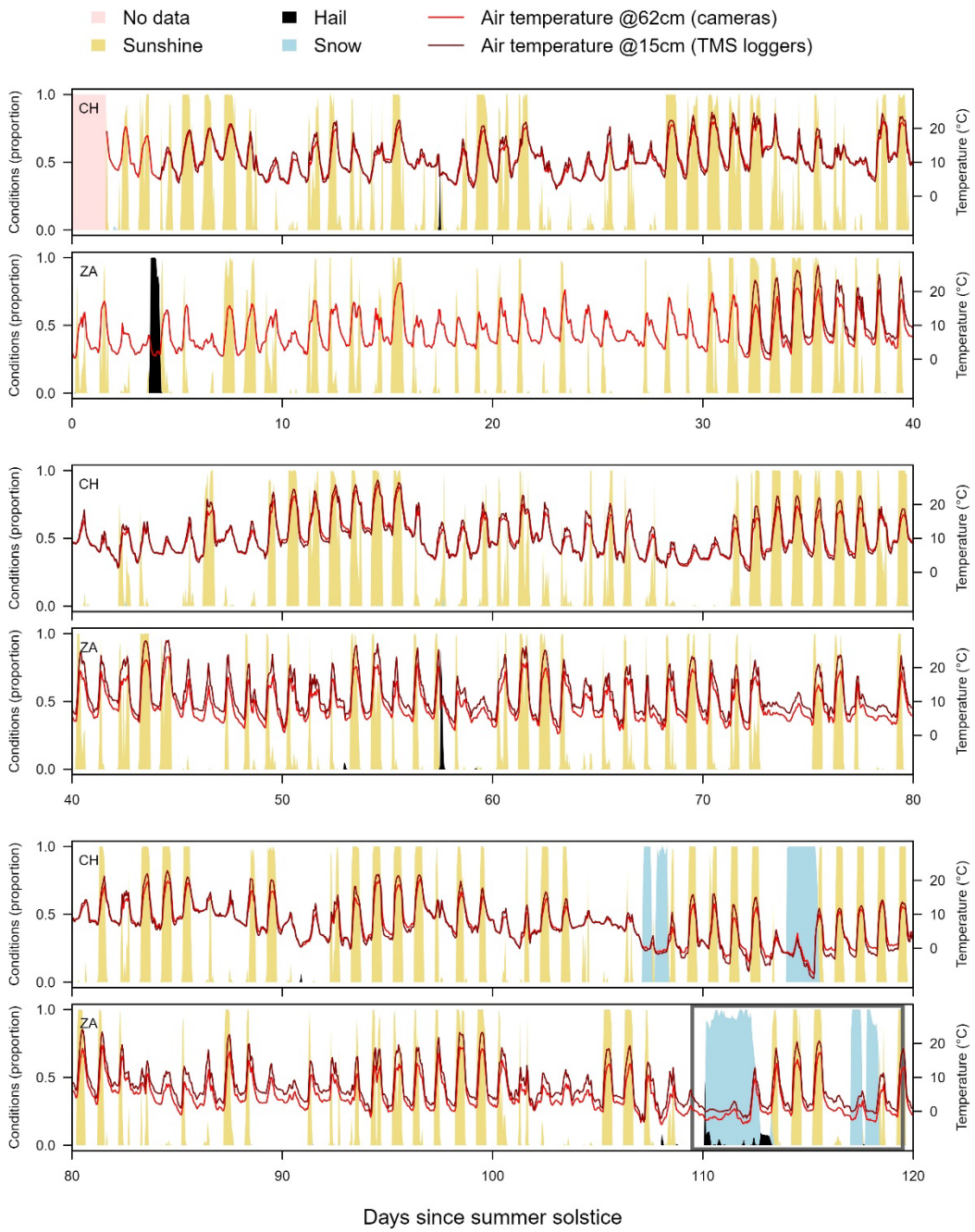

Figure 3. Sunshine (yellow), hail (black) and snow (blue) detected in images throughout summer solstice (top pair), early autumn (middle pair) and late autumn (bottom pair) across two mountain sites. Distinct time series from the northern hemisphere (top of each pair) and the southern hemisphere (bottom of each pair) are aligned based on summer solstice dates of 21st June 2021 in Switzerland (CH) and 21st December in South Africa (ZA). Conditions were classified in all images from continuous cameras at the two high-elevation sites, using a deep learning ensemble trained using the full set of 8,953 labelled images. Data are displayed with a resolution of 1.2 hours (20 time slices per day). We also present mean air temperatures from on-board sensors on cameras, measuring air temperature every 5 minutes at 62cm above ground (red lines), and TMS4 loggers, measuring air temperature every 15 minutes at 15cm above ground (dark red lines). Data in the grey box (bottom-right) are more closely explored in Figure 5.

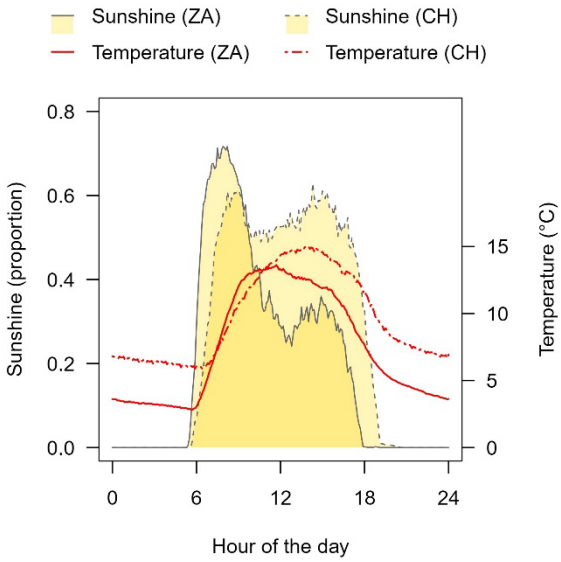

Figure 4. Average diel profiles of image-derived sunshine (yellow shading and grey lines) across high elevation sites in Switzerland (CH, dashed lines) and South Africa (ZA, solid lines) for 120 days following summer solstice. We also present mean air temperatures from on-board sensors on cameras, measuring air temperature every 5 minutes at 62cm above ground (red lines). Data are displayed with a resolution of 0.1 hours (240 time slices per day). The ZA site is characterized by early morning sunshine and afternoon shade. Sunshine was classified using a deep learning ensemble trained using the full set of 8,953 labelled images.

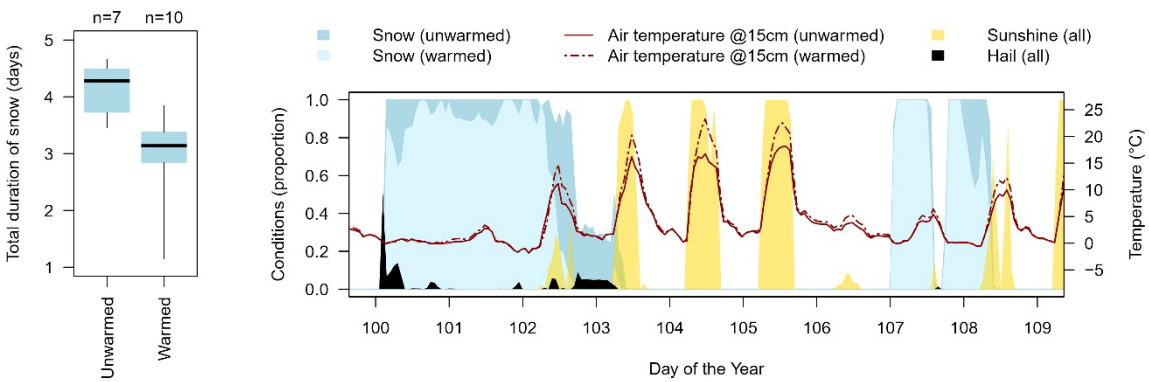

Figure 5. (left) Duration of snow, derived from images, across unwarmed plots and open-top chamber (OTC) warmed plots at the high elevation site in South Africa. Boxes capture the median and interquartile range, while whiskers capture the range of the data. Three cameras failed before the snows and are excluded here. **(right)** Proportion of snow over time across unwarmed (dark blue) and warmed (light blue) plots. Snow events were often shortened or dampened within OTCs. Furthermore, during periods of sunshine (yellow), temperature spikes recorded by TMS4 loggers were more intense in warmed plots (dashed red line) than unwarmed plots (solid red line). Small proportions of snow images were misclassified as hail (black), and this was often due to fog. Weather conditions were classified using a deep learning ensemble trained using the full set of 8,953 labelled images.

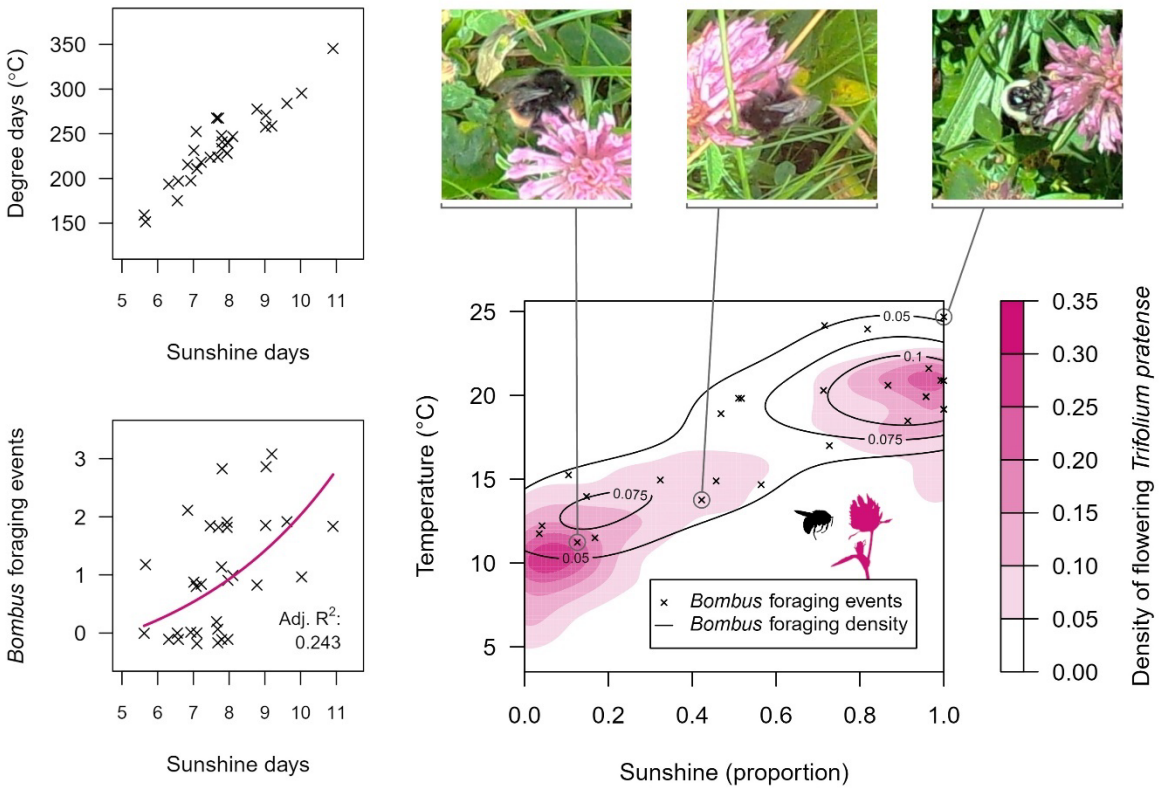

Figure 6. (top left) The cumulative daytime temperatures experienced by *Trifolium pratense* inflorescences, derived from temperature sensors, are highly positively correlated with sunshine exposure, derived from images. **(bottom left)** Sunshine exposure of an inflorescence increases the number of recorded *Bombus* foraging events. A linear model effectively predicts the natural log of the number of *Bombus* foraging events (pink line). **(right)** Sunshine and temperature preferences of foraging *Bombus* bees during late summer days at a high elevation site in Switzerland. The plot shows the temporal niche overlap between flowering *Trifolium pratense* (pink) and their *Bombus* bee pollinators (black crosses and lines) along axes of image-derived sunshine and sensor-derived temperature. The density of both flowering *T. pratense* and *Bombus* visits are bimodal with respect to sunshine and temperature, but *Bombus* bees are more likely to visit during warm and sunny periods. Image crops corresponding to three *Bombus* foraging events are shown above the plot.

Effects of sunshine on bumblebee foraging

Sunshine is a vital factor affecting the activity of flower visiting insects, especially in alpine environments (Bergman et al., 1996). Our data demonstrate how sunshine drives fluctuations in ambient temperature at our site (Fig. 3), but also the cumulative daytime temperatures experienced by individual *T. pratense* inflorescences (Fig. 6, top left). We also observe a strong effect of sunshine exposure on the number of *Bombus* foraging visits to each inflorescence (Fig. 6, bottom left), probably partly mediated by ambient temperatures. However, we found that sunshine days marginally outperformed degree days when predicting the number of *Bombus* visits to an inflorescence (ΔAIC = -0.98). As such, we suggest that sunshine also has direct effects on *Bombus* foraging activity. For example, solar radiation can directly raise thoracic temperatures of bees, relieving a major constraint on flight behavior (Corbet et al., 1993).

We overlaid the density of *Bombus* foraging events with the density of flowering *T. pratense* on a two-dimensional surface of sunshine and temperature. Even though *Bombus* visits are contingent on the presence of flowering *T. pratense*, we found clear evidence of microclimatic and micrometeorological mismatch. *Trifolium pratense* inflorescences were most available at ~10°C with little or no sunshine, with a secondary peak at ~21°C during constant sunshine (Fig. 6, right). However, *Bombus* foraging events were completely absent below 10°C, and disproportionately prevalent during warm temperatures and intermediate sunshine. Camera-based monitoring, with automated extraction of meteorological by-catch, allows us to quantify the constraints and preferences of species at unprecedented spatial and temporal resolutions.

Applications and future development

Micrometeorological variables extracted from images are highly complementary to those recorded by affordable microclimatic sensors. Unlike temperature and moisture, solar radiation is expensive and difficult to measure at high temporal resolution (Roales et al., 2013). Furthermore, cameras are perhaps the only in-field sensors that can record the prevalence of snow (Caparó Bellido & Rundquist, 2021) and hail at fine spatiotemporal resolutions. Crucially, the impacts of sunshine, snow, and hail on the abundance and phenology of wild species are significant, but rarely explored and poorly understood (Fernandes et al., 2011; Möhl et al., 2020, 2022; Valladares et al., 2016). This is particularly concerning given that the frequency of extreme meteorological events is expected to increase under climate change (IPCC, 2023).

Climate change impacts organisms at fine spatial and temporal scales (Potter et al., 2013), with complex effects on the frequency and intensity of species interactions (Blois et al., 2013). Furthermore, changes in phenology can cause mismatches in timing between interacting species, with potentially devastating demographic impacts (Miller-Rushing et al., 2010). Leveraging the micrometeorological by-catch of wildlife cameras can develop our understanding of microclimate ecology along the following lines. Firstly, cameras can capture variation in micrometeorology and organismal activity at very small spatial scales. This creates opportunities to study fine-scale microclimatic refugia, such as areas protected from sunshine or snow, which may be vital for species persistence in extreme environments (von Oppen et al., 2022). Secondly, cameras record continuously at high temporal resolution, allowing the analysis of phenological mismatches not only across days of the year, but hours of the day (Alison et al., 2022). Finally, cameras generate biological and meteorological data that are tethered in both space and time. This is a particularly useful property to explore behavioral adaptations to micrometeorological conditions.

Previous studies have automated the extraction of snow-covered regions of phenocam images (Caparó Bellido & Rundquist, 2021; Julitta et al., 2014). Our approach of classifying entire images is simpler, and thus broader in application. Our model rapidly detects images with snow, but also those with sunshine or hail. Unlike previous models, ours is shown to perform well on images from novel camera placements and even a distant site in Svalbard. Above all, we present a dataset and methods to train future models that will be even more transferable. Future work should focus on the assembly of larger training datasets, representing a greater diversity of backgrounds, conditions, and wildlife camera domains. There is potential for deep learning models to extract hydrological information from images, for example rainfall or even rare cryogenic processes (Grab et al., 2021). Furthermore, models could be trained to extract meteorological by-catch from phenocams for vegetation (Brown et al., 2016) or camera traps for large animals (Hofmeester et al., 2020). A particularly exciting prospect would be to extract latent meteorological information from existing wildlife camera network datasets, containing millions of labelled organisms (Norouzzadeh et al.,

2018). This could provide new insights into the occurrence and phenology of a huge variety of animal and plant species worldwide.

Acknowledgements

This research was funded through the 2019–2020 BiodivERsA joint call for research proposals, under the BiodivClim ERA-Net COFUND programme, with the funding organizations Innovation Fund Denmark (grant no. 0156-00022B), the Department of Science and Innovation Republic of South Africa, the Research Council of Norway, the Swiss National Science Foundation (grant no. 20BD21_193809), the Swedish Research Council for Environment, Agricultural Sciences and Spatial Planning and the German Research Foundation.

References

Aerts, R., Cornelissen, J. H. C., Dorrepaal, E., Van Logtestijn, R. S. P., & Callaghan, T. V. (2004). Effects of experimentally imposed climate scenarios on flowering phenology and flower production of subarctic bog species. *Global Change Biology*, *10*(9), 1599–1609. https://doi.org/10.1111/j.1365-2486.2004.00815.x

Alison, J., Alexander, J. M., Zeugin, N. D., Dupont, Y. L., Iseli, E., Mann, H. M. R., & Høye, T. T. (2022). Moths complement bumblebee pollination of red clover: A case for day-and-night insect surveillance. *Biology Letters*, *20220187*, 1–5. https://doi.org/10.1098/rsbl.2022.0187

Antão, L. H., Bates, A. E., Blowes, S. A., Waldock, C., Supp, S. R., Magurran, A. E., Dornelas, M., & Schipper, A. M. (2020). Temperature-related biodiversity change across temperate marine and terrestrial systems. *Nature Ecology & Evolution*, *4*(7), Article 7. https://doi.org/10.1038/s41559-020-1185-7

August, T., Harvey, M., Lightfoot, P., Kilbey, D., Papadopoulos, T., & Jepson, P. (2015). Emerging technologies for biological recording. *Biological Journal of the Linnean Society*, *115*(3), 731–749. https://doi.org/10.1111/bij.12534

Barrow, D. K., & Crone, S. F. (2013). Crogging (cross-validation aggregation) for forecasting—A novel algorithm of neural network ensembles on time series subsamples. *Proceedings of the International Joint Conference on Neural Networks*. https://doi.org/10.1109/IJCNN.2013.6706740

Barrow, D. K., & Crone, S. F. (2016). Cross-validation aggregation for combining autoregressive neural network forecasts. *International Journal of Forecasting*, *32*(4), 1120–1137. https://doi.org/10.1016/j.ijforecast.2015.12.011

Beery, S., Liu, Y., Morris, D., Piavis, J., Kapoor, A., Meister, M., Joshi, N., & Perona, P. (2020). Synthetic examples improve generalization for rare classes. *Proceedings - 2020 IEEE Winter Conference on Applications of Computer Vision, WACV 2020*, 852–862. https://doi.org/10.1109/WACV45572.2020.9093570

Beery, S., Van Horn, G., & Perona, P. (2018). Recognition in Terra Incognita. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *11220 LNCS*, 472–489. https://doi.org/10.1007/978-3-030-01270-0_28

Beleites, C., & Salzer, R. (2008). Assessing and improving the stability of chemometric models in small sample size situations. *Analytical and Bioanalytical Chemistry*, *390*(5), 1261–1271. https://doi.org/10.1007/s00216-007-1818-6

Bergman, P., Molau, U., & Holmgren, B. (1996). Micrometeorological impacts on insect activity and plant reproductive success in an alpine environment, Swedish Lapland. *Arctic and Alpine Research*, *28*(2), 196–202. https://doi.org/10.2307/1551760

Besson, M., Alison, J., Bjerge, K., Gorochowski, T. E., Høye, T. T., Jucker, T., Mann, H. M. R., & Clements, C. F. (2022). Towards the fully automated monitoring of ecological communities. *Ecology Letters*, *25*(12), 2753–2775. https://doi.org/10.1111/ele.14123

Bjorkman, A. D., Elmendorf, S. C., Beamish, A. L., Vellend, M., & Henry, G. H. R. (2015). Contrasting effects of warming and increased snowfall on Arctic tundra plant phenology over the past two decades. *Global Change Biology*, *21*(12), 4651–4661. https://doi.org/10.1111/gcb.13051

Blois, J. L., Zarnetske, P. L., Fitzpatrick, M. C., & Finnegan, S. (2013). Climate Change and the Past, Present, and Future of Biotic Interactions. *Science*, *341*(6145), 499–504. https://doi.org/10.1126/science.1237184

Bokhorst, S., Huiskes, A., Aerts, R., Convey, P., Cooper, E. J., Dalen, L., Erschbamer, B., Gudmundsson, J., Hofgaard, A., Hollister, R. D., Johnstone, J., Jónsdóttir, I. S., Lebouvier, M., Van de Vijver, B., Wahren, C. H., & Dorrepaal, E. (2013). Variable temperature effects of Open Top Chambers at polar and alpine sites explained by irradiance and snow depth. *Global Change Biology*, *19*(1), 64–74. https://doi.org/10.1111/gcb.12028

Brown, T. B., Hultine, K. R., Steltzer, H., Denny, E. G., Denslow, M. W., Granados, J., Henderson, S., Moore, D., Nagai, S., SanClements, M., Sánchez-Azofeifa, A., Sonnentag, O., Tazik, D., & Richardson, A. D. (2016). Using phenocams to monitor our changing Earth: Toward a global phenocam network. *Frontiers in Ecology and the Environment*, *14*(2), 84–93. https://doi.org/10.1002/fee.1222

Burton, A. C., Neilson, E., Moreira, D., Ladle, A., Steenweg, R., Fisher, J. T., Bayne, E., & Boutin, S. (2015). Wildlife camera trapping: A review and recommendations for linking surveys to ecological processes. *Journal of Applied Ecology*, *52*(3), 675–685. https://doi.org/10.1111/1365-2664.12432

Caparó Bellido, A., & Rundquist, B. C. (2021). Semi-Automatic Fractional Snow Cover Monitoring from Near-Surface Remote Sensing in Grassland. *Remote Sensing*, *13*(11), Article 11. https://doi.org/10.3390/rs13112045

Cawley, G. C., & Talbot, N. L. C. (2010). *On Over-fitting in Model Selection and Subsequent Selection Bias in Performance Evaluation* (Journal of Machine Learning Research, Vol. 11, pp. 2079–2107).

Chollet, F. (2015). *Keras* [Computer software]. https://keras.io

Corbet, S. A., Fussell, M., Ake, R., Fraser, A., Gunson, C., Savage, A., & Smith, K. (1993). Temperature and the pollinating activity of social bees. *Ecological Entomology*, *18*(1), 17–30. https://doi.org/10.1111/j.1365-2311.1993.tb01075.x

Fernandes, G. W., Oki, Y., Sanchez-Azofeifa, A., Faccion, G., & Amaro-Arruda, H. C. (2011). Hail impact on leaves and endophytes of the endemic threatened Coccoloba cereifera (Polygonaceae). *Plant Ecology*, *212*(10), 1687–1697. https://doi.org/10.1007/s11258-011-9941-z

Gillespie, M. A. K., Baggesen, N., & Cooper, E. J. (2016). High Arctic flowering phenology and plant–pollinator interactions in response to delayed snow melt and simulated warming. *Environmental Research Letters*, *11*(11), 115006. https://doi.org/10.1088/1748-9326/11/11/115006

Gillespie, M. A. K., & Cooper, E. J. (2022). The seasonal dynamics of a High Arctic plant–visitor network: Temporal observations and responses to delayed snow melt. *Arctic Science*, *8*(3), 786–803. https://doi.org/10.1139/as-2020-0056

Grab, S., Knight, J., Mol, L., Botha, T., Carbutt, C., & Woodborne, S. (2021). Periglacial landforms in the high Drakensberg, Southern Africa: Morphogenetic associations with rock weathering rinds and shrub growth patterns. *Geografiska Annaler: Series A, Physical Geography*, *103*(3), 199–222. https://doi.org/10.1080/04353676.2020.1856625

Hofmeester, T. R., Young, S., Juthberg, S., Singh, N. J., Widemo, F., Andrén, H., Linnell, J. D. C., & Cromsigt, J. P. G. M. (2020). Using by-catch data from wildlife surveys to quantify climatic parameters and timing of phenology for plants and animals using camera traps. *Remote Sensing in Ecology and Conservation*, *6*(2), 129–140. https://doi.org/10.1002/rse2.136

Hollister, R. D., Elphinstone, C., Henry, G. H. R., Bjorkman, A. D., & Klanderud, K. (2023). A review of open top chamber (OTC) performance across the ITEX Network. *Arctic Science*, *9*(2), 1–14.

Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., & Adam, H. (2017). MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. *arXiv*. http://arxiv.org/abs/1704.04861

Høye, T. T., Ärje, J., Bjerge, K., Hansen, O. L. P., Iosifidis, A., Leese, F., Mann, H. M. R., Meissner, K., Melvad, C., & Raitoharju, J. (2021). Deep learning and computer vision will transform entomology. *Proceedings of the National Academy of Sciences of the United States of America*, *118*(2), 1–10. https://doi.org/10.1073/PNAS.2002545117

Intergovernmental Panel on Climate Change (Ed.). (2023). Weather and Climate Extreme Events in a Changing Climate. In *Climate Change 2021 – The Physical Science Basis: Working Group I Contribution to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* (pp. 1513–1766). Cambridge University Press. https://doi.org/10.1017/9781009157896.013

Julitta, T., Cremonese, E., Migliavacca, M., Colombo, R., Galvagno, M., Siniscalco, C., Rossini, M., Fava, F., Cogliati, S., Morra di Cella, U., & Menzel, A. (2014). Using digital camera images to analyse snowmelt and phenology of a subalpine grassland. *Agricultural and Forest Meteorology*, *198–199*, 116–125. https://doi.org/10.1016/j.agrformet.2014.08.007

Katal, N., Rzanny, M., Mäder, P., & Wäldchen, J. (2022). Deep Learning in Plant Phenological Research: A Systematic Literature Review. *Frontiers in Plant Science*, *13*. https://www.frontiersin.org/articles/10.3389/fpls.2022.805738

Krogh, A., & Vedelsby, J. (1994). Neural Network Ensembles, Cross Validation, and Active Learning. *Advances in Neural Information Processing Systems*, *7*, 231–238.

Lembrechts, J. J., Aalto, J., Ashcroft, M. B., De Frenne, P., Kopecký, M., Lenoir, J., Luoto, M., Maclean, I. M. D., Roupsard, O., Fuentes-Lillo, E., García, R. A., Pellissier, L., Pitteloud, C., Alatalo, J. M., Smith, S. W., Björk, R. G., Muffler, L., Ratier Backes, A., Cesarz, S., … Nijs, I. (2020). SoilTemp: A global database of near-surface temperature. *Global Change Biology*, *26*(11), 6616–6629. https://doi.org/10.1111/gcb.15123

Lembrechts, J. J., Lenoir, J., Roth, N., Hattab, T., Milbau, A., Haider, S., Pellissier, L., Pauchard, A., Ratier Backes, A., Dimarco, R. D., Nuñez, M. A., Aalto, J., & Nijs, I. (2019). Comparing temperature data sources for use in species distribution models: From in-situ logging to remote sensing. *Global Ecology and Biogeography*, *28*(11), 1578–1596. https://doi.org/10.1111/geb.12974

Lenoir, J., Bertrand, R., Comte, L., Bourgeaud, L., Hattab, T., Murienne, J., & Grenouillet, G. (2020). Species better track climate warming in the oceans than on land. *Nature Ecology & Evolution*, *4*(8), Article 8. https://doi.org/10.1038/s41559-020-1198-2

Lumbrazo, C., Bennett, A., Currier, W. R., Nijssen, B., & Lundquist, J. (2022). Evaluating Multiple Canopy-Snow Unloading Parameterizations in SUMMA With Time-Lapse Photography Characterized by Citizen Scientists. *Water Resources Research*, *58*(6), e2021WR030852. https://doi.org/10.1029/2021WR030852

Maclean, I. M. D., Duffy, J. P., Haesen, S., Govaert, S., De Frenne, P., Vanneste, T., Lenoir, J., Lembrechts, J. J., Rhodes, M. W., & Van Meerbeek, K. (2021). On the measurement of microclimate. *Methods in Ecology and Evolution*, *12*(8), 1397–1410. https://doi.org/10.1111/2041-210X.13627

Maclean, I. M. D., & Early, R. (2023). Macroclimate data overestimate range shifts of plants in response to climate change. *Nature Climate Change*, *13*(5), Article 5. https://doi.org/10.1038/s41558-023-01650-3

Mann, H. M. R., Iosifidis, A., Jepsen, J. U., Welker, J. M., Loonen, M. J. J. E., & Høye, T. T. (2022). Automatic flower detection and phenology monitoring using time-lapse cameras and deep learning. *Remote Sensing in Ecology and Conservation*, *Early view*, 1–13. https://doi.org/10.1002/rse2.275

Marion, G. M., Henry, G. H. R., Freckman, D. W., Johnstone, J., Jones, G., Jones, M. H., Lévesque, E., Molau, U., Mølgaard, P., Parsons, A. N., Svoboda, J., & Virginia, R. A. (1997). Open-top

designs for manipulating field temperature in high-latitude ecosystems. *Global Change Biology*, *3*(SUPPL. 1), 20–32. https://doi.org/10.1111/j.1365-2486.1997.gcb136.x

Miller-Rushing, A. J., Høye, T. T., Inouye, D. W., & Post, E. (2010). The effects of phenological mismatches on demography. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *365*(1555), 3177–3186. https://doi.org/10.1098/rstb.2010.0148

Möhl, P., Hiltbrunner, E., & Körner, C. (2020). Halving sunlight reveals no carbon limitation of aboveground biomass production in alpine grassland. *Global Change Biology*, *26*(3), 1857–1872. https://doi.org/10.1111/gcb.14949

Möhl, P., von Büren, R. S., & Hiltbrunner, E. (2022). Growth of alpine grassland will start and stop earlier under climate warming. *Nature Communications*, *13*(1), Article 1. https://doi.org/10.1038/s41467-022-35194-5

Noh, I., Doh, H.-W., Kim, S.-O., Kim, S.-H., Shin, S., & Lee, S.-J. (2021). Machine Learning-Based Hourly Frost-Prediction System Optimized for Orchards Using Automatic Weather Station and Digital Camera Image Data. *Atmosphere*, *12*(7), Article 7. https://doi.org/10.3390/atmos12070846

Norouzzadeh, M. S., Nguyen, A., Kosmala, M., Swanson, A., Palmer, M. S., Packer, C., & Clune, J. (2018). Automatically identifying, counting, and describing wild animals in camera-trap images with deep learning. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(25), E5716–E5725. https://doi.org/10.1073/pnas.1719367115

Ortmann, C. R., & Johnson, S. D. (2021). How reliable are motion-triggered camera traps for detecting small mammals and birds in ecological studies? *Journal of Zoology*, *313*(3), 202–207. https://doi.org/10.1111/jzo.12849

Pecl, G. T., Araújo, M. B., Bell, J. D., Blanchard, J., Bonebrake, T. C., Chen, I.-C., Clark, T. D., Colwell, R. K., Danielsen, F., Evengård, B., Falconi, L., Ferrier, S., Frusher, S., Garcia, R. A., Griffis, R. B., Hobday, A. J., Janion-Scheepers, C., Jarzyna, M. A., Jennings, S., … Williams, S. E. (2017). Biodiversity redistribution under climate change: Impacts on ecosystems and human well-being. *Science*, *355*(6332), eaai9214. https://doi.org/10.1126/science.aai9214

Potter, K. A., Arthur Woods, H., & Pincebourde, S. (2013). Microclimatic challenges in global change biology. *Global Change Biology*, *19*(10), 2932–2939. https://doi.org/10.1111/gcb.12257

Ramanathan, A. V., Cess, R. D., Harrison, E. F., Minnis, P., Barkstrom, B. R., Ahmad, E., Hartmann, D., Science, S., Series, N., Jan, N., Ramanathan, V., Cess, R. D., Harrison, E. F., Minnis, P., Barikstrom, B. R., Ahmad, E., & Hartmann, D. (1989). Cloud-Radiative Forcing and Climate: Results from the Earth Radiation Budget Experiment. *Science*, *243*(4887), 57–63.

Rixen, C., Høye, T. T., Macek, P., Aerts, R., Alatalo, J. M., Jill, T., Arnold, P. A., Barrio, I. C., Bjerke, J. W., Björkman, M. P., Blok, D., Blume-werry, G., Boike, J., Bokhorst, S., Carbognani, M., Christiansen, C. T., Convey, P., Cooper, E. J., Cornelissen, J. H. C., … Zong, S. (2022). Winters are changing: Snow effects on Arctic and alpine tundra ecosystems. *Arctic Science*, *8*(February), 572–608.

Roales, J., Durán, J., Bechtold, H. A., Groffman, P. M., & Rosi-Marshall, E. J. (2013). High Resolution Measurement of Light in Terrestrial Ecosystems Using Photodegrading Dyes. *PLOS ONE*, *8*(9), e75715. https://doi.org/10.1371/journal.pone.0075715

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A. C., & Fei-Fei, L. (2015). ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision*, *115*(3), 211–252. https://doi.org/10.1007/s11263-015-0816-y

Schneider, S., Taylor, G. W., Kremer, S. C., Burgess, P., McGroarty, J., Mitsui, K., Zhuang, A., deWaard, J. R., & Fryxell, J. M. (2022). Bulk arthropod abundance, biomass and diversity estimation using deep learning for computer vision. *Methods in Ecology and Evolution*, *13*(2), 346–357. https://doi.org/10.1111/2041-210X.13769

Spehn, E. M., & Körner, C. (2005). A Global Assessment of Mountain Biodiversity and its Function. In U. M. Huber, H. K. M. Bugmann, & M. A. Reasoner (Eds.), *Global Change and Mountain*

Regions: An Overview of Current Knowledge (pp. 393–400). Springer Netherlands. https://doi.org/10.1007/1-4020-3508-X_39

Tuia, D., Kellenberger, B., Beery, S., Costelloe, B. R., Zuffi, S., Risse, B., Mathis, A., Mathis, M. W., van Langevelde, F., Burghardt, T., Kays, R., Klinck, H., Wikelski, M., Couzin, I. D., van Horn, G., Crofoot, M. C., Stewart, C. V., & Berger-Wolf, T. (2022). Perspectives in machine learning for wildlife conservation. *Nature Communications*, *13*(1), 1–15. https://doi.org/10.1038/s41467-022-27980-y

Valladares, F., Laanisto, L., Niinemets, Ü., & Zavala, M. A. (2016). Shedding light on shade: Ecological perspectives of understorey plant life. *Plant Ecology & Diversity*, *9*(3), 237–251. https://doi.org/10.1080/17550874.2016.1210262

von Oppen, J., Assmann, J. J., Bjorkman, A. D., Treier, U. A., Elberling, B., Nabe-Nielsen, J., & Normand, S. (2022). Cross-scale regulation of seasonal microclimate by vegetation and snow in the Arctic tundra. *Global Change Biology*, *28*(24), 7296–7312. https://doi.org/10.1111/gcb.16426

Wevers, J., Beenaerts, N., Casaer, J., Zimmermann, F., Artois, T., & Fattebert, J. (2021). Modelling species distribution from camera trap by-catch using a scale-optimized occupancy approach. *Remote Sensing in Ecology and Conservation*, *7*(3), 534–549. https://doi.org/10.1002/rse2.207

Wieczorek, J., Guerin, C., & McMahon, T. (2022). K-fold cross-validation for complex sample surveys. *Stat*, *11*(1), e454. https://doi.org/10.1002/sta4.454

Wild, J., Kopecký, M., Macek, M., Šanda, M., Jankovec, J., & Haase, T. (2019). Climate at ecologically relevant scales: A new temperature and soil moisture logger for long-term microclimate measurement. *Agricultural and Forest Meteorology*, *268*, 40–47. https://doi.org/10.1016/j.agrformet.2018.12.018

Wipf, S., & Rixen, C. (2010). A review of snow manipulation experiments in Arctic and alpine tundra ecosystems. *Polar Research*, *29*(1), 95–109. https://doi.org/10.1111/j.1751-8369.2010.00153.x

Table S1. Rare meteorological events were sorted into three tiers based on duration and representation in the systematic subset. This determines the sampling protocol for strategic sampling and allows for more intensive sampling of the rarest events.

Event duration	Representation	Strategic sampling protocol	Applies to:
> 1 day	< 100 images	Extract one image per-camera per-hour during the two non-consecutive days with highest prevalence of the given condition	Snow at ZA high site Snow at CH high site Frost at CH high site
< 1 day	< 100 images	Extract one image per-camera per-hour during three non-consecutive 12h windows with highest prevalence of the given condition	Hail at ZA high site
< 2 hours	< 10 images	All images across all cameras in the 2h window with the highest prevalence of the given condition.	Hail at CH high site

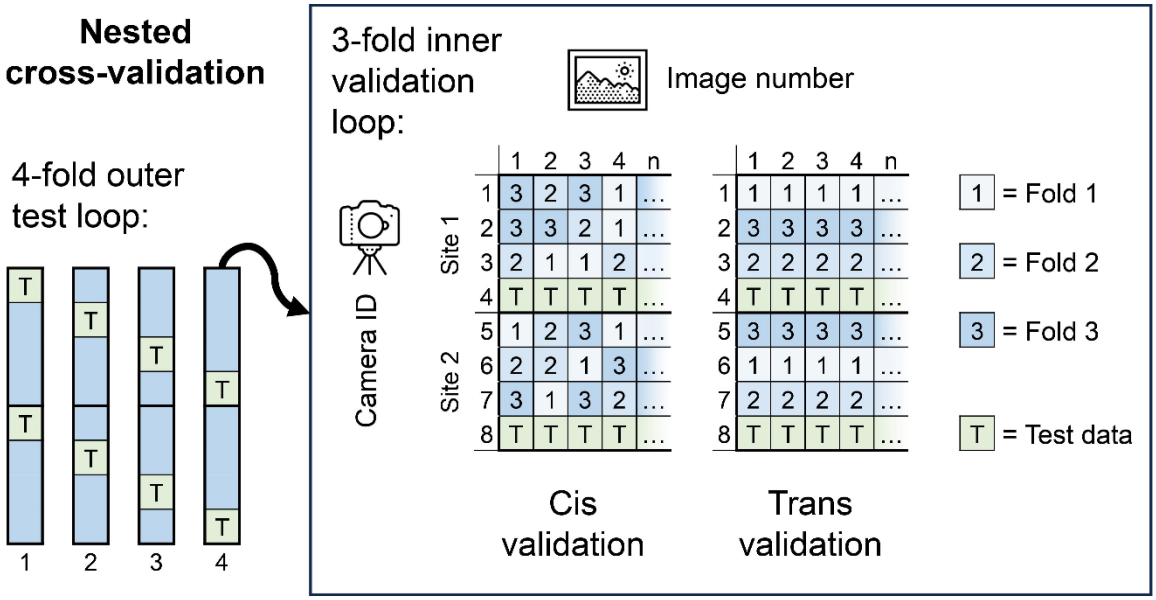

Figure S1. A simplified example of the nested cross-validation procedure – note, our study uses inner and outer validation loops with five and six folds respectively. Here, the outer test loop determines which of *k*=4 folds of cameras will be held back as out-of-sample test data. For each iteration of the outer test loop, an inner validation loop determines which of *k*-1=3 folds of images are used for early stopping. The delineation of folds in the inner validation loop is done using one of two methods: "Cis" or "Trans". Cis validation involves naïve random splitting of images from all cameras. Trans validation splits based on cameras rather than images, as in the outer test loop. During Trans validation, a given camera never contributes to both model training and validation. Data splitting is always stratified with respect to sampling site.

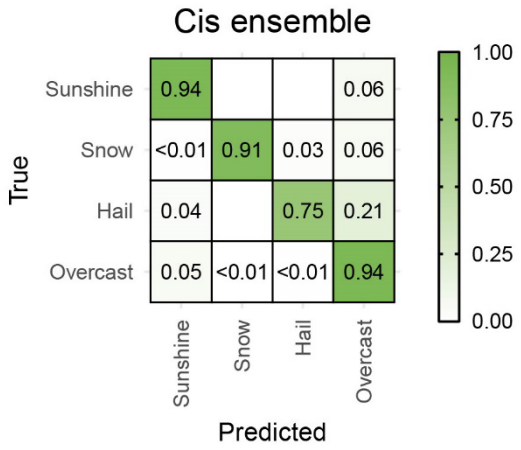

Figure S2. Confusion matrix of average predictions of an ensemble model following Cis validation (left) and Trans validation (right). Six Cis ensembles and six Trans ensembles were used to classify images from distinct hold-out test cameras.

Appendix S1. Further details on image labelling. We initially labelled evidence of sunshine, hail, snow and frost. Sunshine was evidenced by clear shadows cast by vegetation; hail by pellets of ice and ensuing slush; snow by snowflakes and snow cover, as well as ensuing slush; and frost by ice crystals forming on vegetation. Frost differed between the two regions; in ZA, heavy white frost erupted from the vegetation to form snowy crystals. This strongly resembled, and was often accompanied by, precipitated snow. In CH, frost was light and difficult to separate from dew. As such, frost was included in the snow class in ZA but the overcast class in CH. Overcast images did not adhere to a strict meteorological definition of overcast, being defined by the absence of sunshine, snow or hail. They comprised images of dry or wet vegetation, day and night, with occasional fog or light frost. A few images containing both sunshine and snow were included in the snow class.

Appendix S2. A primer on CNN parameters and concepts. During training of a CNN, training images are fed to an algorithm iteratively in computationally manageable **batches** of a given number of images. Model parameters are updated after each iteration, based on a given **optimizer** and a **learning rate**, to minimize a **loss function** (effectively a measure of how wrong the model's predictions are for a given batch of images). An **epoch** is a full cycle of iterations that spans the entire training dataset. The batch size, the optimizer, the learning rate, and the number of epochs are hyperparameters. **Overfitting** means an increase in model performance with respect to the training dataset (measured as the training loss), but a decrease in model performance with respect to independent observations in a validation dataset (measured as the validation loss). **Early stopping** means finding the epoch when training should be stopped to maximize model performance with respect to a validation dataset.

Appendix S3. Further details on Trans validation and crogging ensembles. Several obstacles hinder deep learning approaches to extract biotic or abiotic data from images. Chief among these is model transferability to novel camera deployments or locations, which often goes untested. This issue is exacerbated by redundancy or shared information between images from the same time-series; training datasets with tens of thousands of images may represent <100 individual deployments (Bjerge et al., 2023). To prevent overfitting to specific image backgrounds, studies must deploy many cameras across distinct locations. Furthermore, the presence of shared information between training, validation and test datasets can cause a phenomenon known as "information leakage",

leading to underestimation of prediction error (Wieczorek et al., 2022). This can be avoided by splitting images appropriately between training, validation, and test datasets (Beery et al., 2018, 2020). In this study, we compared two data splitting methods for validation in our study: “Cis” and “Trans” (Fig. S1). The Trans validation approach withholds entire deployments for an especially rigorous validation process. Few studies applying deep learning to wildlife camera imagery have used a Trans data splitting process for model validation, though some have used it for model testing (Beery et al., 2018, 2020).

A major limitation of the Trans validation approach is that it greatly reduces the effective size of the training dataset, especially if the number of deployments is far smaller than the number of images. In the field of forecasting, an approach called cross-validation aggregation (crogging) has helped to overcome this barrier by allowing distinct time-series to contribute to both training and validation datasets – but not in the same model (Barrow & Crone, 2013). Hitherto untested for wildlife camera applications, crogging is an ensemble modelling approach that combines the benefits of cross-validation and model aggregation. Crogging can outperform other ensemble methods such as bootstrap aggregation (bagging; Barrow & Crone, 2013), and has proven useful where there is non-independence in the data – for example when classifying multiple brain tissue images per patient (Beleites & Salzer, 2008) or forecasting from autocorrelated time-series (Barrow & Crone, 2016). However, there is the caveat that the number of folds used during crogging affects the proportion of data used for training vs. improving out-of-sample performance of each member model (Barrow & Crone, 2016; Krogh & Vedelsby, 1994).

We compared crogging ensembles using the Cis validation approach, which permits images from the same camera to be used in both training and validation sets, against those using the Trans validation approach, which does not (Beery et al., 2018, 2020). We found that performance of ensemble models from the Trans validation approach was comparable to that of the Cis validation approach, even though member models from the Trans validation approach saw 20% fewer cameras during training. Furthermore, our crogging ensemble models tended to outperform full models, which were trained on all available data for up to 29 epochs, retaining the model from the epoch with the lowest training loss. In any case, the average difference between member and ensemble models was generally greater when using the Trans validation approach (Table 1). Our results highlight that the combination of crogging and Trans validation approaches demands further exploration in deep learning applications to images from wildlife cameras.