Assessing the accuracy of a spatial model of habitat suitability for *Calypso bulbosa*

Michael Doorly

Assessing the accuracy of a spatial model of habitat suitability for *Calypso bulbosa*

Michael Doorly

Abstract

Calypso bulbosa is a rare and visually striking orchid that grows in older mesic to moist forests in the northern half of Sweden. *C. bulbosa* is red listed as a threatened species (Vulnerable, VU) with a reduction in numbers linked to modern forestry practices and exacerbated by the warming climate. The species is protected in Sweden and appears in two appendices of the EU's Species and Habitats Directive (EEA, 2016). To properly monitor the population and its response to a warming climate, an accurate model of species distribution is required. This data will be crucial in sustainable forestry management and conservation strategy development.

In 2022 a model was developed by SLU Artdatabanken *(English: Swedish Species Information Centre)* at the Swedish University of Agricultural Sciences. This model suggested a number of potential locations for *C. bulbosa* presence in the alpine region, suggesting that current estimates of a *C. bulbosa* population of around 1000 individuals in the region may not be entirely accurate. This thesis tested the 2022 model's predictive capability by visiting a number of areas in the Swedish alpine region with varying probability levels for *C. bulbosa* presence, with a special focus on "hotspots" that the model deemed to have a high likelihood of *C. bulbosa* occurrence. The result of these visits was then tested statistically against expected occurrence values from the 2022 model.

Over the course of the fieldwork, two new *C. bulbosa* populations were discovered in the alpine region, one of which being the largest in the region discovered in Sweden in the 21st century thus far. In total, 1190 individuals of *C. bulbosa* were observed in the alpine region over the course of the fieldwork, a figure which surpasses the currently estimated population for the entire region. Despite these successes, the model itself did not appear to have a good predictive power in the alpine region at the hectare scale, with an AUC=0.445 making the model's predictive capacity no better than random chance in this regard.

The presence/absence data collected from the period of fieldwork can now be used to potentially improve the model itself, thereby improving the accuracy of *C. bulbosa* population estimations in Sweden. Accurate modelling and population tracking are crucial tools for responsible ecological management and policymaking.

Keywords: Calypso bulbosa, norna, orchid, conservation, Sweden, alpine region, habitat suitability, distribution model, ecology, biodiversity

Table of contents

List of tables

List of figures

Abbreviations

AUC	Area Under Curve
CCF	Continuous Cover Forestry
ENFA	Ecological Niche Factor Analysis
GAM	Generalised Additive Model
GBIF	Global Biodiversity Information Facility
HSM	Habitat Suitability Model
IUCN	International Union for Conservation of Nature
Lu lpm	Lule lappmark: a Swedish flora province in Lappland, Norrbotten
Ly lpm	Lycksele lappmark: a Swedish flora province in Southern Lappland, Västerbotten
NR	Nature reserve
OR	Opportunistic reporting
Pi lpm	Pite lappmark: a Swedish flora province in Lappland, Norrbotten
SDM	Species Distribution Modelling
SLU	Swedish University of Agricultural Sciences
SP	Systematic Protocol
To lpm	Torne lappmark: a Swedish flora province in Lappland, Norrbotten
VU	Vulnerable red-list class, according to IUCN

1. Introduction

1.1 *Calypso bulbosa*: species ecology, distribution, and challenges

According to population estimates the majority of the red-listed calypso orchid (*Calypso bulbosa)* population in Sweden is located in the north of the country within the boreal region (EEA, 2016). The majority of observations reported via the Swedish species observation system 'Artportalen' (SLU, 2023), a service developed and operated by the SLU Swedish Species Information Centre (Swedish: SLU Artdatabanken) on behalf of the Swedish Environmental Protection Agency, are located in this area (figure 1). In 2019 the number of individuals estimated in the Swedish boreal region was 90 000, with around 1000 individuals estimated in the alpine region (Naturvårdsverket, 2020).
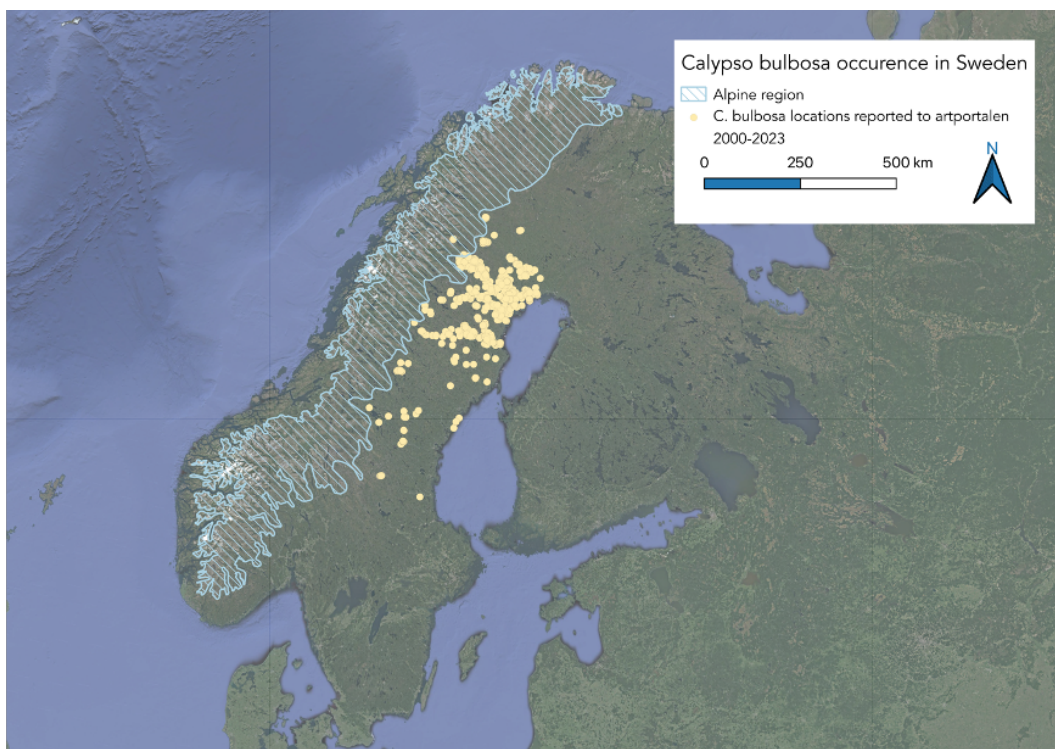
Figure 1: Map showing C. bulbosa locations reported into Artportalen between 2000-2023. Geodata for the alpine region created from a map of Nordic biogeographical regions ©2020 Naturvårdsverket. Background image ©2019 Google

In 2022 a species distribution model (SDM) developed by SLU Artdatabanken indicated that there may be a significant and hitherto undocumented population in the alpine region (Stephan et al., 2022). To evaluate the predictive power of the 2022 SDM, sites throughout the alpine region were visited and the predicted probability was statistically assessed against the actual presence of *C. bulbosa* individuals within each 1ha quadrant visited.

1.1.1 Description, distribution, and ecology

When developing an SDM, a thorough understanding of the target species' ecology is crucial to creating a model with good predictive power (Thuiller et al., 2023). *C. bulbosa* is a diploid (2n=24), perennial orchid in the sub-family Epidendroideae, distributed circumboreally and divided into four varieties: var. *bulbosa* in Eurasia, var. *speciosa* (Schelct.) Makino in Japan, var. *americana* (R.Br.) Luer in North America from the Rocky Mountains and eastward, and var. *occidentalis* (Holzinger) Calder & Taylor in the Pacific Northwest. The variety *Calypso bulbosa* var. *bulbosa* is found in the Northern parts of Sweden, Finland and Russia (Alexandersson and Ågren, 2000), primarily in mesic to moist, shady, and preferably calcium rich areas within older coniferous forests with an herbaceous plant understory (Mossberg and Stenberg, 2018). *C. bulbosa* is often found growing amidst mosses, typically red-stemmed feathermoss (*Pleurozium schreberi*) and/or mountain fern moss (*Hylocomium splendens*) as well as the shrubs lingonberry (*Vaccinium vitis-idaea*) and bilberry (*V. myrtillus*).

Calypso bulbosa prefers lower-lying areas on the forest floor with a high site quality and flowing groundwater. Other species often found in the same area and that are indicative of the higher site index typical for a *C. bulbosa* supporting biotope include wintergreen (*Pyrola spp.*), the woodland geranium (*Geranium sylvaticum*), *Paris quadrifolia*, and *Linnaea borealis* as well as less common species such as red baneberry (*Actaea rubra*) and orchids such as *Cypripedium calceolus* and creeping lady's-tresses (*Goodyera repens*).

Calypso bulbosa grows to a height of 10-20 cm, and in the Nordic region it is most easily observable between mid-May to June, when the striking flowers, typically purple/pink (occasionally white) in colour, make the otherwise unassuming orchid suddenly visible. The zygomorphic, hermaphroditic flower is accented with a labellum typically paler pink to white in colour and with a tonally darker spotting (figure 2). Other distinguishing features include the singular ovoid leaf located at the orchid's base as well as a faint aroma of vanilla. The leaf is attached to an underground ellipsoid-to-ovoid bulb-like corm that gives the plant it's Latin epithet of "*bulbosa*" and produces the orchids 5-8 cm long roots (Currah et al., 1988). *C. bulbosa* is frequently found in half-open areas of the forest and on slopes with a north to easterly direction. This openness leads to early snow coverage which creates a beneficial microclimate for overwintering of the leaf and corm. (ArtDatabanken, 2019).

Figure 2: C. bulbosa specimen (Photo: Michael Doorly, 2023)

Calypso bulbosa lacks a spur – in fact it produces no nectar or useable pollen, instead relying on colour, scent and anther-like hairs to deceive potential pollinators (Proctor and Harder, 1995) and the plant is primarily pollinated by bees such as *Bombus* queens of *B. hypnorum* (L.), *B. jonellus* (Kirby), *B. pratorum* (L.), and *B. pascuorum* (Scop) ssp. *sparreanus* (Alexandersson and Ågren, 1996). The percentage of fruit-setting individuals varies between studies, from 11-34% in var. *occidentalis* (Ackerman, 1981) to a relatively high value of 21-48% in var. *bulbosa* (Alexandersson and Ågren, 1996). Pollination is dependent on a large population of preferably "naïve" newly-emerged bumblebees that have yet to learn about the plant's lack of nectar. As the season develops, these insects learn that *C. bulbosa* is not a viable food source and prioritise other plants, leading to a reduction in the pollination rate. Cross-pollination is achieved through the deposition of pollinaria onto the scutellum of visiting bees, though this process is not guaranteed to result in a successful pollination, as observed in var. *occidentalis*, where the variability of both pollinator size as well as flower structure lead to an estimated 25% pollination success rate (Ackerman, 1981), with one flower pollinated for every four pollinaria deposited.

Like many other non-saprophytic terrestrial orchids, *C. bulbosa* is also reliant on the presence of endophytic fungi. This symbiosis allows for the utilisation of mycorrhizal roots, though it is not limited to one specific mycorrhizal fungus. In studies from Alberta, USA, multiple taxa of endophytic mycorrhizal fungi were observed throughout the roots, outer cortex and corm tissue of plant specimens (Currah et al., 1988). The presence of symbiotic fungi is also essential for the successful germination of *C. bulbosa* seeds (Smallwood and Trapnell, 2022). The combination of specific microclimate requirements as well as the necessary fungal

symbiosis for germination, deceptive pollination and a threatened, fragmented habitat all contribute to the calypso orchid's vulnerability in an era of rapidly changing environmental conditions.

1.1.2 Challenges and conservation

The calypso orchid is red listed as a threatened species (VU) in Sweden (Ahrné et al., 2020) based on Red List criteria outlined by the IUCN (IUCN, 2001). In the case of *C. bulbosa* the VU status is the result of an estimated population reduction of at least 30% over three generations (Ahrné et al., 2020).

The majority of reported calypso orchid locations in Sweden are found within the boreal region. The Swedish boreal *C. bulbosa* population has been experiencing a period of negative growth, even if the smaller alpine population is currently judged to be stable (ArtDatabanken, 2019). There are multiple explanatory factors likely contributing to this decline with many exacerbated by the interplay of the calypso orchid's complex life history and the rapidly changing environment that surrounds it at both macro and local levels.

Changes in land use since the industrial revolution have had a profound effect on global biodiversity resulting in many habitats being fragmented, disrupted or destroyed (Weiskopf et al., 2020). Habitat fragmentation favours plants adept at spreading and colonizing, creating challenges for rare plants with poor seed dispersal and a more niched ecology such as *C. bulbosa* (Maina and Howe, 2000). Developments in mechanised agriculture, with its focus on monocultures, fertilisation and history of herbicides and pesticides is but one of a complex array of factors negatively impacting biodiversity globally. In the north of Sweden even "green" hydroelectric energy has flooded landscapes and removed many sites that were particularly suitable for *C. bulbosa* growth (ArtDatabanken, 2019).

On a local level, modern forestry practices are replacing old-growth forest with high production monoculture forests of even aged Norway spruce (*Picea abies* (L.) Karst.) grown in short rotations, leading to a significant habitat loss for many species (Nordlind and Östlund, 2003). *C. bulbosa* is especially sensitive to clear-cut methods which are common practice in Swedish forestry (Law and Yeung, 1989). The calypso orchid typically does not survive the period of surface drying that follows a clear-cut felling. In addition, the presence of slash (leaves and branches that are usually left on site after a clear-cut) negatively impacts *C. bulbosa* survival in a site (ArtDatabanken, 2019). Even less-invasive forestry treatments such as continuous cover forestry (CCF) can prove problematic to *C. bulbosa*, as machinery used in felling (harvesters and forwarders) and site preparation (e.g. disc trenching) can lead to local extinction of the plant. Advice for foresters and landowners for conservation of *C. bulbosa* is to encourage a varied mix of trees, both old and similarly aged, retain soil moisture and natural variation of ground water levels. Trees and the living environment should be protected by leaving groups of trees, areas or zones whilst avoiding crude disturbance of soil or vegetation through clear cuts, soil preparation, driving damage, fertilisation or ditching (Berglund, 2015).

Local extinctions are compounded on a landscape level by *C. bulbosa*'s poor propensity for reproduction. The orchid's strategy of deceptive pollination is reliant on high insect populations, and low fruit setting in combination with poor seed dispersal can also limit gene flow, making the population less genetically diverse and reducing its ability to deal with changes in climate, diseases, and changes in pollinator populations (Alexandersson and Ågren, 2000).

Increasing temperatures from global warming are predicted to affect interactions between heterotrophs and autotrophs, with disturbances to pollination and seed dispersal being prime examples of the way in which this disruption can manifest (Traill et al., 2010). Modelling on another deceptive orchid, the montane orchid (*Traunsteinera globosa*) suggests that the effect of climate change on pollinators will lead to a reduced orchid population (Kolanowska, 2021). An accurate model would help ensure that *C. bulbosa* population and dispersal were accurately monitored across the entirety of Sweden, including areas where less data is sampled. Habitat suitability & distribution modelling is one such strategy that can aid in monitoring populations and help shape policies that can hopefully protect this vulnerable species from further decline. Accurate predictive modelling can improve population number estimations, track population change over time and in response to environmental changes, as well as inform conservation and forestry strategies (Johnson and Gillingham, 2007, Puchnina, 2017, Smallwood and Trapnell, 2022).

1.2   Habitat suitability & distribution modelling

1.2.1 Basic concepts

The goal of distribution modelling is to use available data to understand why species are distributed the way they are, either spatially or temporally (Guisan et al., 2017). When developing any form of distribution model, a sound understanding of species ecology is essential in determining which of the available environmental factors are likely to be strongly correlated to the presence or absence of the species in question in a given area. SDMs are a useful tool for predicting species distribution in areas not yet surveyed, or for predicting how populations might change over time in response to different variables, such as climate change (Kolanowska, 2021). Models such as the 2022 SDM assessed in this paper (Stephan et al., 2022) can form the basis of a more accurate estimation of population and thereby aid with future conservation efforts (Maina and Howe, 2000).

Thanks to exponential developments in computing power, computer-based predictive modelling has been increasing in popularity since the early 1970s. SDMs typically function by quantifying the relationship between a species of interest and its environmental requirements (Guisan et al., 2017). Today environmental and spatially explicit biological data are readily available online, as

are powerful packages for statistical analysis that can be freely downloaded and implemented in R studio (R core team, 2023) such as Hmsc (Tikhonov et al., 2022) or biomod2 (Thuiller et al., 2023).

A concept essential to all species modelling is the ecological definition of a species niche, which can be divided into two main subcategories: fundamental and realised (Smith and Smith, 2015). The fundamental niche is defined as the set of environmental conditions under which a species can persist and is determined by abiotic physical variables such as pH and temperature. The realised niche is the portion of the fundamental niche that is ultimately exploited by a species, determined by biotic interactions such as competition from other species (Smith and Smith, 2015). Ultimately, an SDM will model the realised, not the fundamental niche of the species (Williams et al., 2009). A narrow species distribution may have less to do with a narrow fundamental niche, instead reflecting other ecological constraints such as biotic interactions or physical barriers.

As the majority of the data available for building these models is related to species presence/absence and environmental data, the fundamental niche has the most pragmatic value when in the modelling stage of SDM development, when making considerations on the types of spatial and environmental data that are to be used (Higgins et al., 2012), though biotic factors can be designed into a model through other inferences.

In order to build a model with good predictive power, explicit goals must be outlined and grounded in the early stages of model conception based on the type of data used (Guisan et al., 2017) including what types of environmental data are available, the explanatory power of this data, correlations between different explanatory variables and whether species presence and absence data are both available, or if pseudo absences must be generated.

The 2022 SDM for *C. bulbosa (Stephan et al., 2022*) includes many design choices that are intended to make it as robust as possible given the data, timescale and processing power available to the team behind it. What follows is a brief description of some of these considerations and their implications on the predictive power of the model.

1.2.2 Sampling

When estimating population numbers of a species, design-based surveys (also referred to as systematic protocols, or SPs) provide the least biased estimates, though reaching the required sampling intensity to give an accurate estimation is a challenge for rare species (Thomas C. Edwards et al., 2005). Comprehensively surveying the Swedish *C. bulbosa* population with an SP would be difficult given the rarity and geographic spread of the species. Instead, population estimates are extrapolated from reports from citizen scientists in combination with sound knowledge of species ecology (European Environment Agency, 2022).

The 2022 model is based on *C. bulbosa* presence data originating from citizen scientists reporting via Artportalen. This opportunistic reporting (OR) with no sampling design creates certain challenges when it is used as the foundation of a SDM, where quantity of data may come at a cost to quality (Henckel et al., 2020). One factor that can affect OR data is observer error when sampling rare plant populations. In the case of rare plant species *Physaria filiformis,* 36.5% of estimates of plant abundancy from observers were incorrect (i.e., they deviated from subsequent exhaustive counts) (Morrison and Young, 2016). Underestimation of abundance increased with shading and overlapping vegetation layers, which may be of particular issue for *C. bulbosa,* especially in its less visible vegetative stage (figure 8), as the species tends to grow amidst the herbaceous understory and is often shaded by larger trees and shrubs. Temporal issues may also affect *C. bulbosa* detection, as the species is most striking when its pink florescence is visible during the short period of mid-May to late June, meaning that OR early or late in the flowering season may miss less-visible vegetative individuals in the area (Johnson and Gillingham, 2007).

The nature of the locations frequented by citizen scientists creates further potential bias in SDMs based on OR data, with roadsides and more easily accessed areas typically being more frequently surveyed (Kadmon et al., 2004), leaving large swathes of more remote, inaccessible landscape unaccounted for. However, even though the frequency of plant OR data near roads is typically higher than it would be in a spatially randomised design distribution, the predictive maps produced, though less accurate, are typically not significantly so (Kadmon et al., 2004). A more pervasive issue of what areas are sampled may come from preferential sampling, i.e. areas not being sampled based on the observers judgement/assumptions that an area is not suitable (Edwards Jr. et al., 2006), a phenomenon that could explain the lack of OR data from the alpine region in the 2022 model.

Despite sampling biases, model predictions based on opportunistic reporting are usually comparable to data collected through systematic protocol when high-quality absences are inferred in the modelling, with OR models outperforming SP based species distribution models when modelling for particularly rare species (Henckel et al., 2020). OR data from citizen scientists is also consistent over time, as any under- or over-estimations based on area sampled, observer error, duration and/or time will typically remain constant (Devictor et al., 2010). As any estimation error from bias remains fairly constant in relation to actual population numbers, OR data such as the data used in the 2022 SDM is therefore statistically comparable over longer time periods, making it suitable for population monitoring.

1.2.3 Pseudo absences

Databases such as Artportalen or GBIF primarily contain "presence-only" datasets from opportunistic reporting. Opportunistic data collection contrasts with, for example, a typical stratified survey of bird life (Felton et al., 2016) where trained researchers are out in the field recording both the presence and absence data for a

target species in a given area. Creating an HSM from presence-only data is challenging as there is no design available (Thuiller et al., 2023) and the data collected through non-stratified sampling is often assumed to be biased, problematic and unsuitable for statistical analysis (Zaniewski et al., 2002). When using presence-only data, little is known about under-sampled areas, as is the case for *C. bulbosa* in the Swedish alpine region.

One commonly used strategy for creating an HSM from presence-only datasets is to generate pseudo-absences, which are equivalent to background data from the area of interest (Thuiller et al., 2023). Pseudo-absences provide information on the environment in the study region (Phillips et al., 2009) and can be sampled in multiple ways including random distribution, environmentally weighted distribution (Zaniewski et al., 2002) or even generated using frequently associated species (Phillips et al., 2009). Once generated, pseudo-absences can be utilised to extend the statistical power and range of an HSM beyond the original study region (Graham et al., 2004).

1.3 Spatial modelling of habitat suitability for *Calypso bulbosa* in the 2022 SDM

In 2022, *C. bulbosa* distribution in northern Sweden was modelled from a starting point based upon known observations reported into Artportalen, as well as observations of indicator species (both with and without the presence of *C. bulbosa*) in relation to different environmental conditions (Stephan et al., 2022). Pseudo-absences were generated from OR presence data of 70 associated plant species collected via artportalen. The 2022 model works on the assumption that if an observer reported one of these 70 plant species known to grow in a similar environment to *C. bulbosa* during the period of May 15th – July 1st, and had not observed the presence of *C. bulbosa,* then *C. bulbosa* could be assumed absent in this hectare, in a similar methodology to (Henckel et al., 2020).

Environmental conditions were characterised with the help of GIS mapping layers describing climate using detailed vegetation zones, primary vegetation types (with data from Nationella Marktäckedata), hydrology (with data from SLU's markfuktighetskarta), soil types, land gradient and soil acidity. In total 11 explanatory variables (out of a starting point with >100 variables) that could help predict the presence or absence of *C. bulbosa* were eventually selected based on their predictive power and lack of correlation with one another (table 1) (Stephan et al., 2022). In terms of the raw data used, while there is no agreed upon boundary in the literature on the minimum number of occurrences for generating
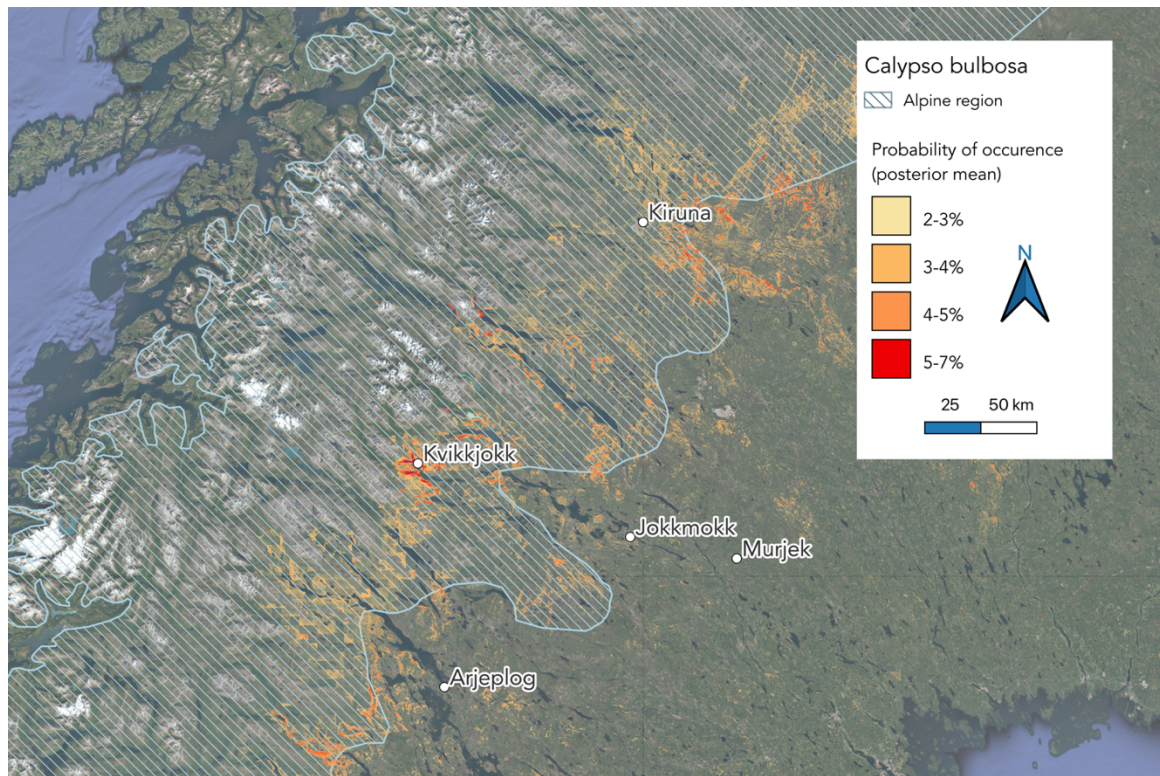
Figure 3: Heat map showing likelihood of C. bulbosa (Stephan et al., 2022). Geodata for the alpine region created from a map of the Nordic biogeographical regions ©2020 Naturvårdsverket. Background image ©2019 Google.

SDMs, the number of presences used in the 2022 model is well above minimum sufficient numbers from other studies (Williams et al., 2009) Full details of how the 2022 model was constructed can be found in *(Stephan et al., 2022)*. What follows is a summary of some key features of the model:

- The study area was divided into 1 ha quadrants
- Presence data was imported from Artportalen, marked as "present" if *C. bulbosa* reported in a quadrant
- Pseudo absences were inferred using 70 plant species frequently associated with *C. bulbosa*. *C. bulbosa* recorded as 'absent' if one of these 70 plants found, but not *C. bulbosa* in the period of May 15th – July 1st
- Environmental variables taken from quadrants with *C. bulbosa* presence
- 11 explanatory variables chosen (table 1) based on highest explanatory power and low correlation with one another
- Hierarchical modelling of species communities (Hmsc) (Tikhonov et al., 2022) in R (R core team, 2023) to make an SDM (Bayesian multivariate, hierarchical generalised linear mixed model)

- Spatial random effect for old and new quadrants – Gaussian predictive process (GPP) implemented to account for spatial structure utilising "knot locations"
- Result: a prediction of species presence for each quadrant in the study area (posterior mean, i.e. a revised/updated probability taking into account new information)

The model utilises Hmsc (Tikhonov et al., 2022), the model rated highest in a comparison of 33 SDMs when ranked for predictive performance (Norberg et al., 2019).

Table 1. The 11 variables used in the 2022 model (Stephan et al., 2022).

Variables
Mean height of 100 m^2 cell
Mean percent south of 100 m^2 cell (from aspect, if x larger than 180-> x=360-x)
pH in the top humus layer
Mean soil moisture in 100 m^2 cell
Mean annual precipitation
Mean annual temperature
Proportion of 100 m^2 cell with granite
Proportion of 100 m^2 cell with limestone
Proportion of 100 m^2 cell with silicate in soil
Proportion of 100 m^2 cell with organic soil
Proportion of 100 m^2 cell with spruce forest

The resulting map layer produced by the 2022 model showed the predicted likelihood of *C. bulbosa* presence on a hectare scale. In contrast to the previously known distribution (primarily clustered near the coast in Norrbotten and Västerbotten, figure 1) the model pointed to many potential locations in forests nearer the Scandes mountain range (figure 3), an area where there are currently very few observations reported. If the model is accurate, and this large, undocumented population exists in the alpine region, it would have implications for our understanding of *C. bulbosa* population numbers and future conservation strategies.

1.4   Goal and research questions

The goal of this study is to assess the predictive power of the 2022 model by visiting the alpine region and surveying sites highlighted as hotspots for *C. bulbosa* occurrence. Using the SDM, new populations in the alpine region may potentially be discovered (Williams et al., 2009). Regardless of whether new *C. bulbosa* presences are recorded or not, the observations from this study can nevertheless be used to improve the model, as the most effective way to remove

bias from correlation of site and survey effort to survey under-represented regions (Phillips et al., 2009).

In order to thoroughly assess the model, areas sampled should include a range of probabilities. Low probability areas that from ecological knowledge and experience seem like a potential *C. bulbosa* site will also be visited, thereby allowing both false positive and false negatives to be evaluated (Stephan et al., 2022).

Given the scarcity of information on the *C. bulbosa* population in the alpine region, any populations discovered in the region should be inventoried, and habitat conditions recorded for future research.

1.5 Limitations

The data collected in this study will be analysed against the probability values from the 2022 to assess its predictive power. Though it is strongly suggested that the presence/absence data collected from the period of fieldwork be used to augment and improve the original model, this is beyond the scope of this paper due to constraints on time and computing power.

Ideally, the model would be assessed by a large-scale survey of the alpine region, though this is impractical given the restraints on time, manpower and budget for this study. The fieldwork will thus be in the form of multiple small-scale surveys spread over a large area, based around particular areas of interest or 'hotspots' during the period of florescence for *C. bulbosa* (May to June).

Additionally, many of the 'hotspots' are located in remote areas, often on the less accessible side of a large body of water, or far away from any access road. In order to reduce the potential of bias from primarily sampling easily accessible sites near the roadside, more remote sites will be prioritised when feasible.

2. Method

Data was collected through fieldwork in the north of Sweden between the 1st and 20th of June, when *C. bulbosa* was in bloom and thus more easily observed. During this period the flower was in bloom and the surrounding vegetation had not yet fully developed, meaning that even the vegetative stage was relatively visible to the trained observer (figure 8).

The locations surveyed were selected based on the heat map produced in QGIS (figure 3) using data from the 2022 model (Stephan et al., 2022). A more explicit version of this map was produced (figure 4), where only areas with a probability over 3.5% were visible. These concentrated patches of high probability became the basis for selecting the areas of interest that would be visited via the routes surveyed in this paper.
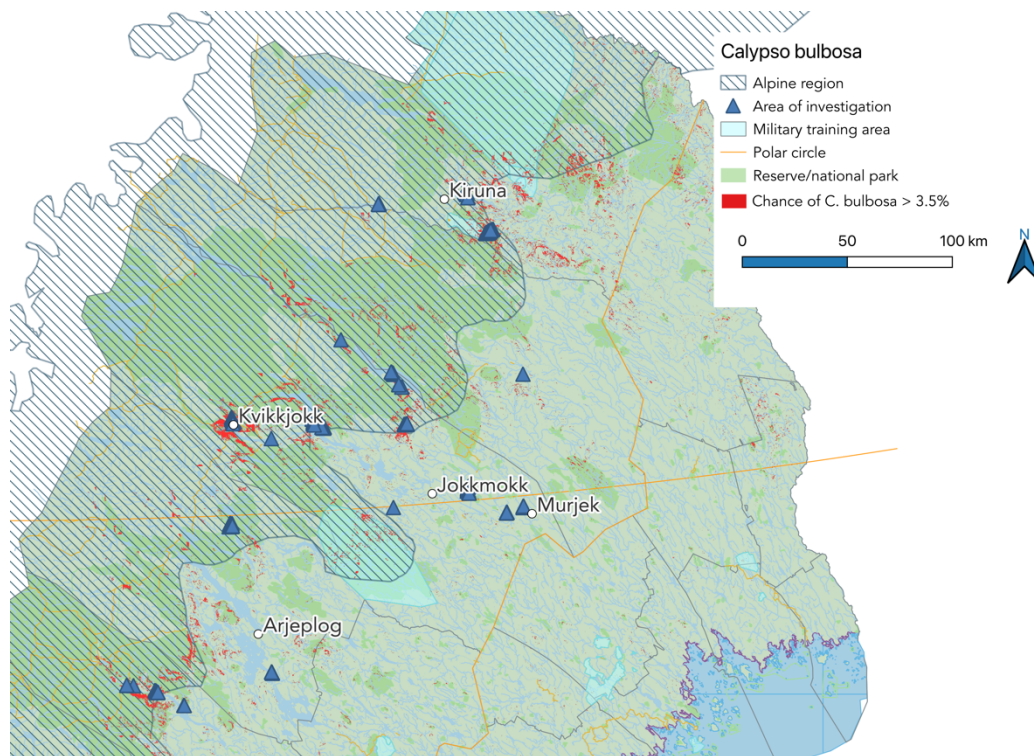
Figure 4: Map showing areas of interest selected using probability data from (Stephan et al., 2022). Geodata for the alpine region created from a map of Swedish biogeographical regions ©2020 Naturvårdsverket. Background map ©2020 Lantmäteriet.

Access was another important consideration, as many areas of high probability were too far away from a road to be accessible, especially in areas where the landscape was divided by a large body of water with road access on one side only. There were multiple areas of interest located in old growth forest southwest of Jokkmokk, as well as to the northeast of Kiruna (figure 3), but access to these areas was restricted to the public as they are military training grounds, meaning they could not be included in this study.

Table 2. List of sites identified as having a high probability of C. bulbosa presence (The highlighted sites are previously identified C. bulbosa locations)

Area of interest	County	Floral province	Biogeographical region
Liebben	Jokkmokk	Lu lpm	Alpine
Ågåsgielas	Jokkmokk	Lu lpm	Alpine
Serri nature reserve	Jokkmokk	Lu lpm	Boreal
Ballemvárásj, Tjåmotis	Jokkmokk	Lu lpm	Alpine
Stenudden, Arjeplog	Arjeplog	Pi lpm	Alpine
Sorsele, Arjeplog	Arjeplog	Ly lpm	Alpine
Ålloluokta	Jokkmokk	Lu lpm	Alpine
Guossanjárga	Jukkasjärvi	T lpm	Alpine
Aptasvare nature reserve	Jukkasjärvi	T lpm	Alpine
Akkar	Jukkasjärvi	T lpm	Alpine
Kvikkjokk	Jokkmokk	Lu lpm	Alpine
Siejdejávrásj, Tjåmotis	Jokkmokk	Lu lpm	Alpine

In total 12 areas of interest were identified (table 2) and preliminary routes of around 4 km in length were mapped out for each area in QGIS in such a way that the routes included a range of hectare values ranging from a low to high probability of *C. bulbosa* presence.

Each hectare square visited was surveyed for the presence/absence of *C. bulbosa* (with presence being recorded if the number of individuals ≥ 1) as well as a brief description of the forest type and the surrounding flora including the presence of companion species. Riparian areas with a low percentage change in the model but seemingly good likelihood of presence from an ecological standpoint were also prioritised for opportunistic spot-checks. The coordinates of the areas inventoried were logged by a GPS (Garmin Montana® 700) using the SWEREF 99 TM coordinate system. If *C. bulbosa* presence was noted, then the number of individuals in the quadrant was surveyed and recorded. The collected data was then analysed with a logistic regression in R (R core team, 2023) using the ROCR package (Sing et al., 2005).

In addition to the planned routes, time was given to visiting sites of particular interest, such as in Jaurekaska, a known *C. bulbosa* location that has been consistently reported into Artportalen over a number of years, situated far inside the northern alpine region. The purpose of these visits was to more thoroughly inventory these sites and their surroundings, to understand the challenges *C. bulbosa* faces in the region, and to provide greater insight into the type of habitat that can support *C. bulbosa* within the alpine region for future research. These sites are summarised in a separate section (3.3) and were not included in the statistical analysis.

3. Results

3.1 Summary of data

Over the period of field research for this study *C. bulbosa* was only present on one of the ten pre-planned alpine routes designed for statistical analysis. *C. bulbosa* was also observed at a spot-check on day one in Liebben.

The location at Liebben was chosen based on favourable ecological conditions including the gradient of the site and proximity to a nearby brook, despite the model showing a low chance of *C. bulbosa* presence (figure 26) and has been included in statistical analyses.

Table 3 shows a breakdown of the average percentages (probability of occurrence, posterior mean) across all hectare squares investigated in a given area of interest using data from the 2022 model (Stephan et al., 2022).

Table 3. Summary of locations visited, with average probability for hectare squares investigated both in the presence and absence of C. bulbosa.

Area visited	Average probability (%) of hectare squares investigated where *C. bulbosa was* **absent**	Average probability (%) of hectare squares investigated where *C. bulbosa was* **present**
Liebben	-	1,9
Ågåsgielas	3,00	3,03
Ballemvárásj, Tjåmotis	2,75	-
Stenudden, Arjeplog	2,61	-
Sorsele, Arjeplog	3,25	-
Ålloluokta	2,47	-
Guossanjárga	2,55	-
Aptasvare nature reserve	2,73	-
Akkar	2,43	-
Kvikkjokk	2,76	-
Siejdejávrásj, Tjåmotis	1,86	-

Data was also collected by Elina Larsson, who visited a number of sites across Jämtland in both the alpine and boreal regions (figure 5) with the same goal of assessing the 2022 model (Larsson, 2023). *C. bulbosa* was not observed in any of the alpine routes investigated during E. Larsson's fieldwork. The data from these two studies has been collated to provide a more complete picture of the alpine region and has been included in the statistical analysis performed in this paper.
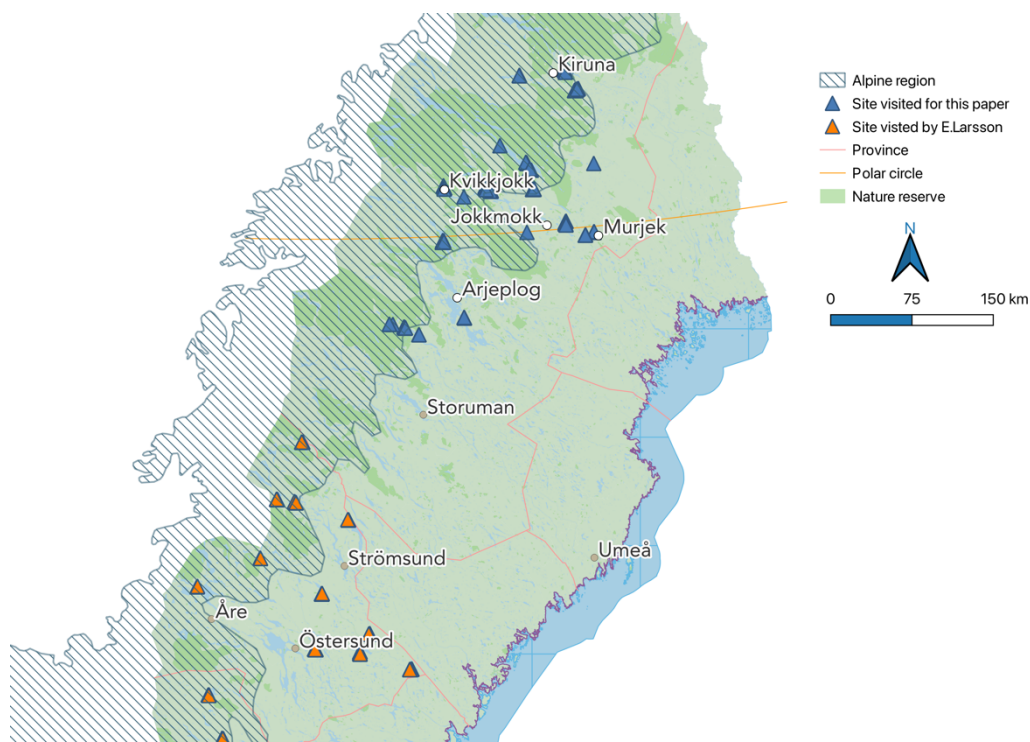
Figure 5: Map showing locations visited for the purposes of this paper, and those of (Larsson, 2023). Geodata for alpine region created from map of Swedish biogeographical regions ©2020 Naturvårdsverket. Background map ©2020 Lantmäteriet

3.2 Observations from field visits

The following are summaries of the sites visited between 04/06/23 and 20/06/23.

3.2.1 Liebben, 2,5 km NE of Tjåmotis, Jokkmokk, Lu lpm, 04/06/23

Whilst scouting for appropriate growing conditions near Heliga Fallet, NE of Tjåmotis, a new *C. bulbosa* location within the alpine region was discovered (pink polygon, figure 20). The area was not selected specifically based on probability values from the 2022 model, instead several locations in this general area were sampled as they were deemed interesting in regard to their topography and proximity to flowing water. This area of fairly sparse, low forest (figure 6)

consisting of downy birch (*Betula pubescens*), Norway spruce (*Picea abies)*, pine (*Pinus spp.)* and juniper (*Juniperus communis*) followed a small watercourse as it flowed downhill from NE to SW.

In total the approximately 200 m^2 area contained an estimated 1000-1500 *C. bulbosa* individuals, including a confirmed 560 fertile and 178 vegetative individuals (figure 6) recorded over a three-hour period. Other companion species noted in the area included *H. splendens*, *V. vitis-idaea*, crowberry (*Empetrum nigrum),* stiff clubmoss (*Lycopodium annotinum),* bilberry (*Vaccinium myrtillus*), and wood horsetail (*Equisetum sylvaticum*). None of the companion species indicated calcium rich conditions.

Figure 6: Left: Image of surrounding habitat with two individuals of C. bulbosa in foreground. Note the openness of the forest, and the mixture of deciduous and coniferous tree species in this location. Right: C. bulbosa in situ amongst leaf litter Photos: Michael Doorly, 2023

3.2.2 Ågåsgielas towards Gárddevárjåhkå, Jokkmokk, Lu lpm, 05/06/23
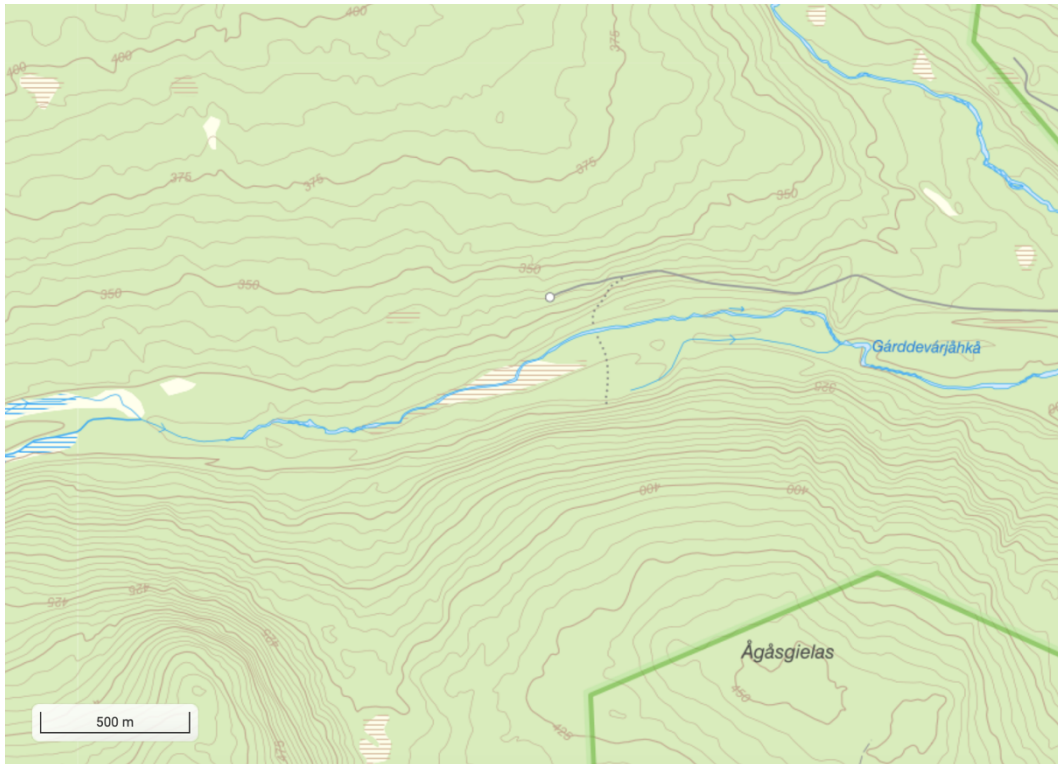
Figure 7: The area investigated, surrounding Gárddevárjåhkå in the centre of the valley. Note the steeper elevation of the slope south of the watercourse, where the majority of C. bulbosa were discovered. Map ©2020 Lantmäteriet

This route, located in the alpine region NW of Jokkmokk, was situated in a valley with a watercourse (Gárddevárjåhkå) flowing through its lowest point (Figure 7). 248 individuals of *C. bulbosa* (86 fertile, 52 budding, 110 vegetative) were observed along the approximately 4 km route through the valley, mainly clustered at the base of the steeper, southern slope. Water was observed flowing down this southern slope from a saturated area at the top of the elevation, down towards Gárddevárjåhkå in the centre of the valley.

Tree coverage was sparse, with low shrubs dominating amongst a mix of deciduous and coniferous trees. Many of the *C. bulbosa* sites were in close proximity to *J. communis,* with companion plants *H. splendens*, *V. vitis-idaea*, *E. nigrum, L. annotinum,* and *V. myrtillus* all present in the area. Additionally, *Linnaea borealis*, *Orthilia secunda*, *Pyrola minor*, *Antennaria dioica*, the red-listed *Actaea rubra* as well as orchids *Corallorhiza trifida* and *Goodyera repens* were all noted.

Figure 8: Left: Example of C. bulbosa in vegetative stage. Centre: Image showing habitat on Southern side of the valley, looking into the centre towards Gárddevárjåhkå. Right: C. bulbosa growing amidst the low shrub layer

3.2.3 Ballemvárásj, Tjåmotis, Jokkmokk, Lu lpm, 07/06/23
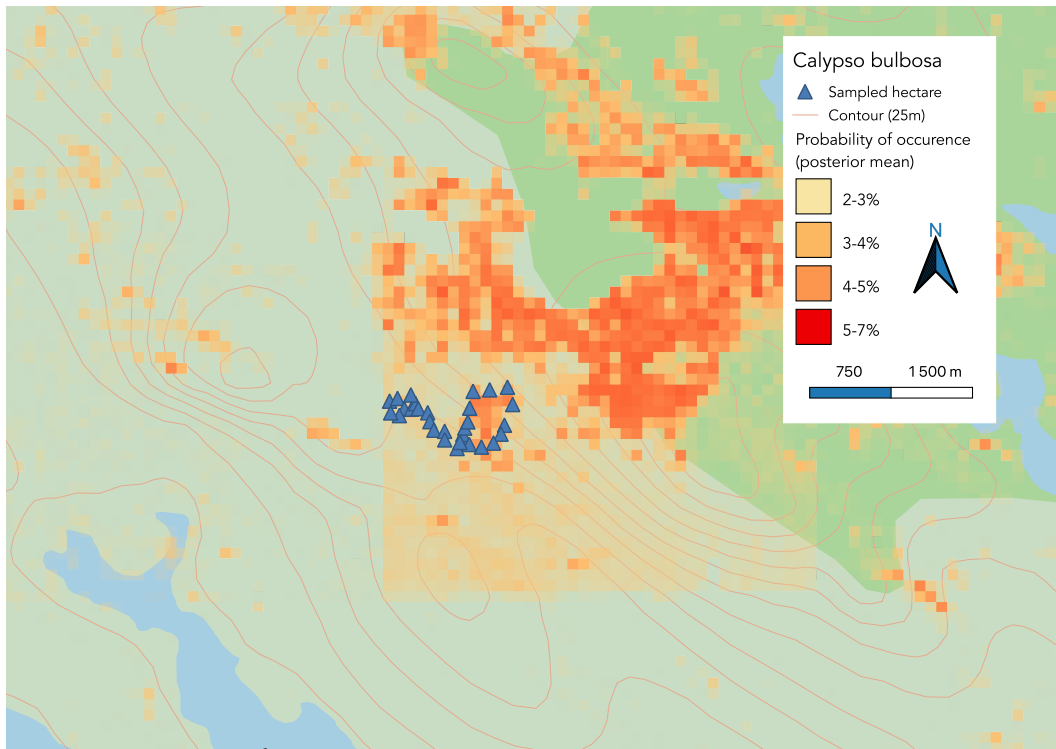
Figure 9: Heat map showing the likelihood of C. bulbosa using probability data from (Stephan et al., 2022). Background map ©2020 Lantmäteriet

The route was selected due to the model suggesting a high probability of calypso orchid occurrence in the area (figure 9). To sample multiple habitats the route began to the west of the valley, crossed over the watercourse and followed the valley to the SE until it continued up the valley wall and into the forested area to the east. The close proximity to the location visited at Liebben three days prior in combination with an apparently similar topography made this location seem like a

promising candidate. In reality the valley was steep and difficult to access, and the conditions were far more waterlogged and swamp-like than those witnessed at Liebben. Navigating this area was slow going, and ultimately the route was ended up being around 3 km. On the latter part of the route several moss dominated locations were potentially interesting, with companion species such as *J. communis, L. borealis* and *Pyrola minor* noted. No new *C. bulbosa* locations observed.

Figure 10: Left: view down into the valley from SW side. Centre: One of the more suitable habitats at the top of the NE slope. Right: One of 17 G. repens orchids found in the more suitable habitat on the NE side of the valley.

3.2.4 Stenudden, Arjeplog, Pi lpm, 10/06/23

The area around Stenudden is further south than the sites visited up until this point, but still within Sweden's alpine region. The area visited had a high probability for *C. bulbosa* presence (figure 11) and indeed at many points on the route the environment looked promising from the perspective of calypso orchid presence (figure 12), with running water, moss and shrub coverage, sparse forest and the presence of a number of companion species. Despite this, no new *C. bulbosa* locations were discovered.
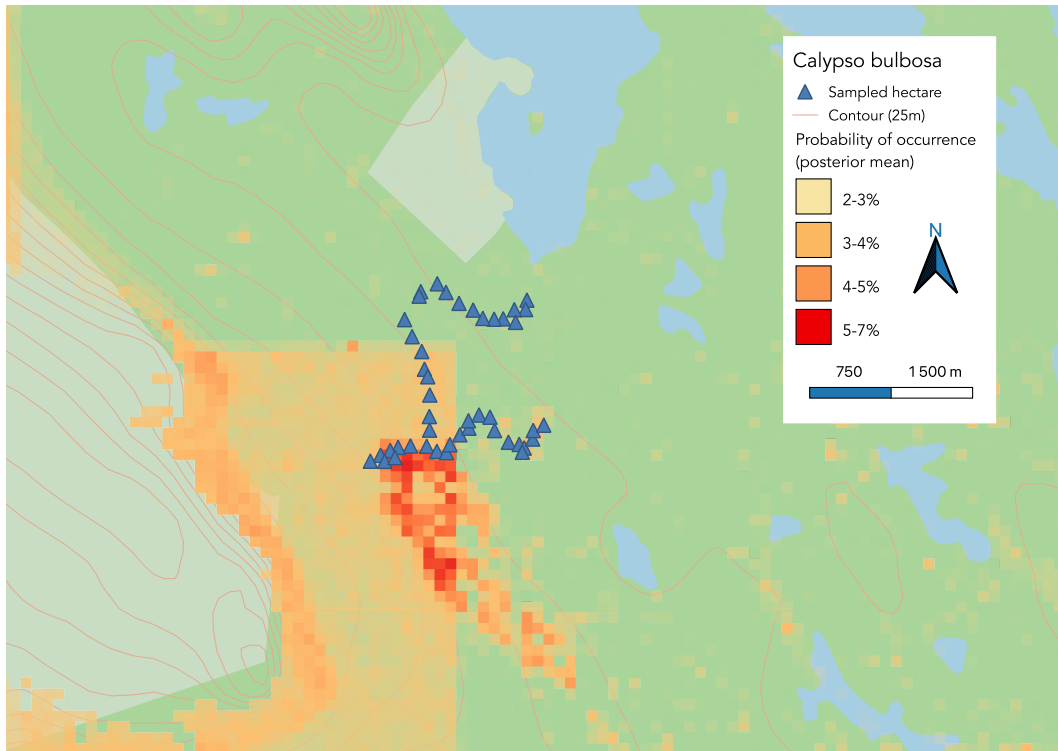
Figure 11: Heat map showing likelihood of C. bulbosa in the area surrounding Stenudden, using probability data from (Stephan et al., 2022). Background map ©2020 Lantmäteriet

Figure 12: Images of habitat around Stenudden. Left: Flowing water, thin forest and the presence of moss and J. communis Centre: Leaves of another orchid, Dactylorhiza maculata, emerging from the leaf litter. Right: more flowing water presents another potential calypso habitat along the route.

3.2.5 Sorsele, Ly lpm 11/06/23

The planned route was focused around Jillesnåle, an area to the east of figure 13 with a relatively high probability for *C. bulbosa* presence. Two additional areas were also sampled to the west of the of Jillesnåle: Kraddsele and Kyrkberget. These locations were chosen due to their topography and proximity to water.

The slopes around Jillesnåle featured a mixture of coniferous and mixed forest with a number of small streams running through it. Around these areas of increased hydrological activity the ground was often too waterlogged for *C. bulbosa,* indicated by the presence of *Sphagnum* spp., with the areas in between the waterlogging being too dry. Around Kraddsele and Kyrkberget the conditions seemed promising, at times strikingly similar to known *C. bulbosa* locations visited that morning: mesic with a moss and shrub dominated understory that included many companion species such as brittle bladder fern (*Cystopteris fragilis*), May lily (*Maianthemum bifolium*), *P. minor, G. sylvaticum, L. borealis, E. nigrum* and *P. quadrifolia.* No individuals of *C. bulbosa* were observed.

Figure 13: Heat map showing likelihood of C. bulbosa in the area surrounding Sorsele, using probability data from (Stephan et al., 2022). Background map ©2020 Lantmäteriet

3.2.6 Ålloluokta, Jokkmokk, Lu lpm, 12/06/23

This area was selected because of its position on a NE slope and favourable proximity to a large body of inland water (figure 14). Additional sites were sampled en route to Ålloluokta, as they were close to smaller streams running out into Stora Lulevatten. The landscape was dominated by mixed forest, often growing from a base of boulders. No individuals of *C. bulbosa* were observed.

Figure 14: Heat map showing likelihood of C. bulbosa in the area surrounding Ålloluokta, using probability data from (Stephan et al., 2022). Background map ©2020 Lantmäteriet

3.2.7 Guossanjárga, Jukkasjärvi, T lpm, 13/06/23

The area surveyed was dominated by mixed forest, with a shrub rich understory mainly consisting of *E. nigrum* and *V. vitis-idaea*. Small streams flowing through the area supported a greater diversity of plants including *Solidago spp.*, *L. borealis* and *C. fragilis* (figure 16). Additional hectares were surveyed to the south of the main route (figure 15) along the boundaries of a fast-flowing stream (figure 16). No individuals of *C. bulbosa* were observed.

Figure 15: Heat map showing likelihood of C. bulbosa in Guossanjárga, Jukkasjärvi using probability data from (Stephan et al., 2022). Background map ©2020 Lantmäteriet

Figure 16: Left: The area around a small stream slowing through Guossanjárga, Jukkasjärvi. Right: The faster flowing watercourse sampled to the southwest of the main route.

## 3.2.8 Aptasvare fjällurskogs nature reserve, Jukkasjärvi, T lpm, 15/06/23

Figure 17: Heat map showing likelihood of C. bulbosa in Aptasvare, Jukkasjärvi using probability data from (Stephan et al., 2022). Background map ©2020 Lantmäteriet

An area of mixed forest dominated by *B. pubescens, P. abies* with a relatively high probability of *C. bulbosa* occurrence in the model (figure 17), Aptasvare was for the most part too waterlogged to be a suitable *C. bulbosa* habitat. *Sphagnum spp.* presence indicated that water from the numerous small streams in the area had created a generally marshy environment. Despite this, several promising looking habitats around small streams in the area were observed, containing companion plants including *G. sylvaticum*, *Chamaenerion angustifolium*, *Phegopteris connectilis*, *Solidago spp.*, *J. communis* and the mosses *P. schreberi and H. splendens*. No individuals of *C. bulbosa* were observed.

3.2.9 Akkar, Jukkasjärvi, T lpm, 16/06/23

Figure 18: Heat map showing likelihood of C. bulbosa in Akkar, Jukkasjärvi using probability data from (Stephan et al., 2022). Background map ©2020 Lantmäteriet

An area with a low site index as indicated by the dry, mineral soil more common in the alpine region. Occasional traces of historical water flow were observed along the route, but at the time of visit the area was dominated by overwhelmingly dry, mountainous landscape with sparse forest consisting of *P. abies* and *B. pubescens*. Few companion plants beyond *V. vitis-idaea*, *V. myrtillus*, and *E. nigrum* with sporadic *J. communis*. Some smaller patches indicating higher site index with *Solidago spp.* and mosses. No individuals of *C. bulbosa* were observed in the area.
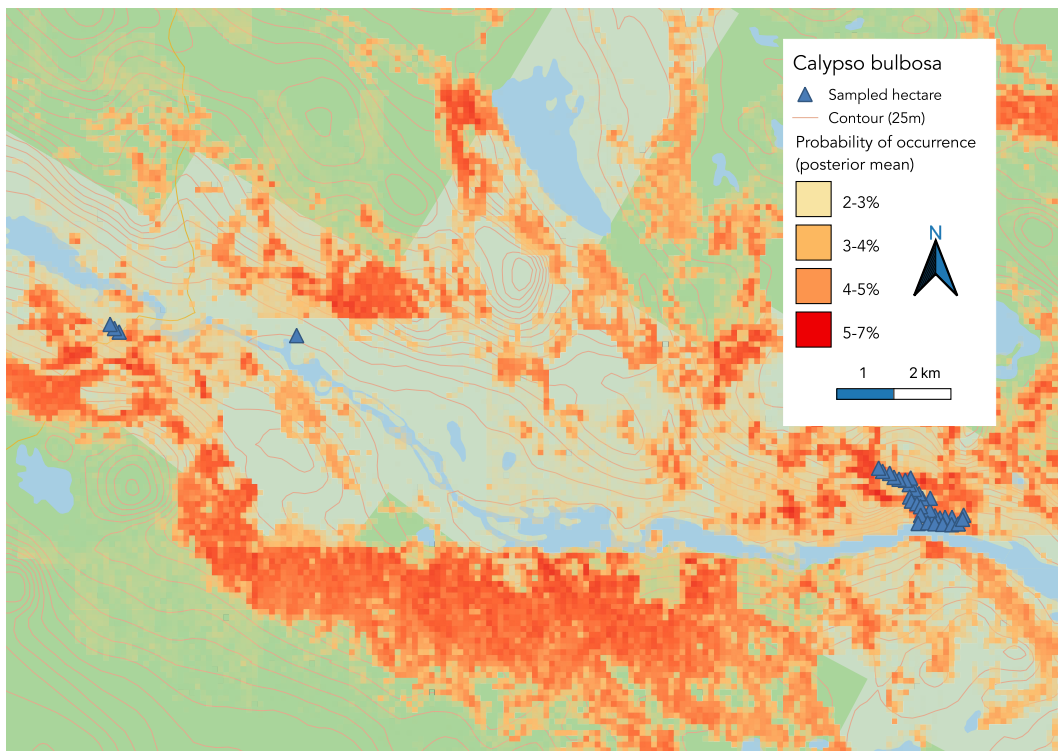
3.2.10 Kvikkjokk, Jokkmokk, Lu lpm, 19/06/23
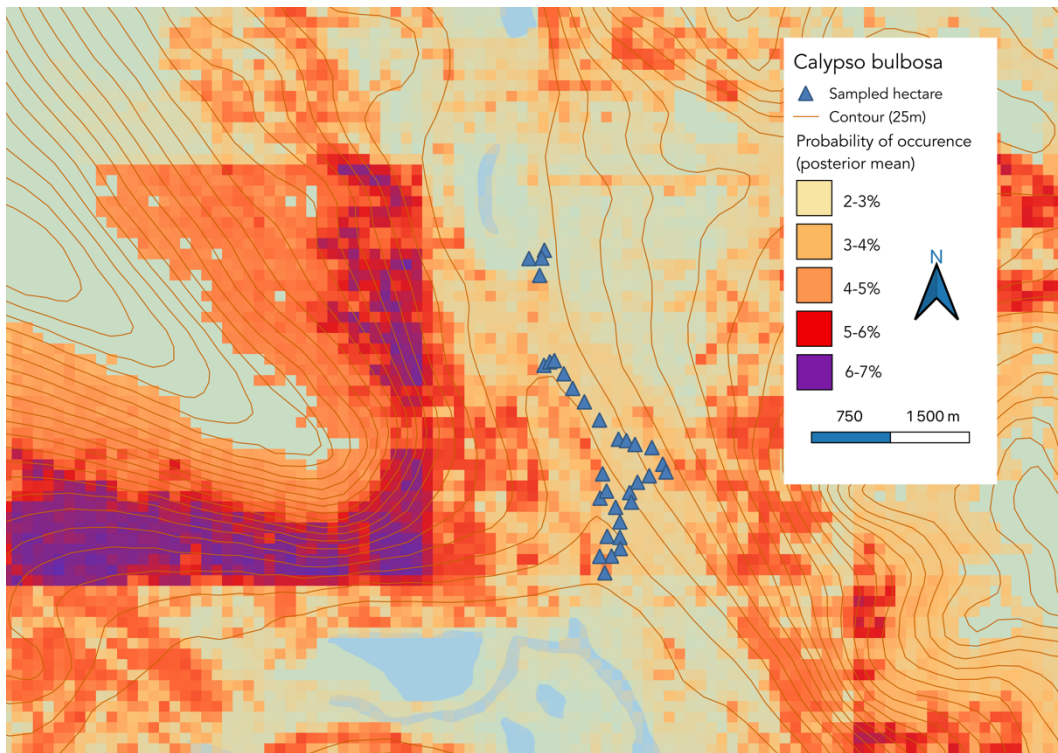
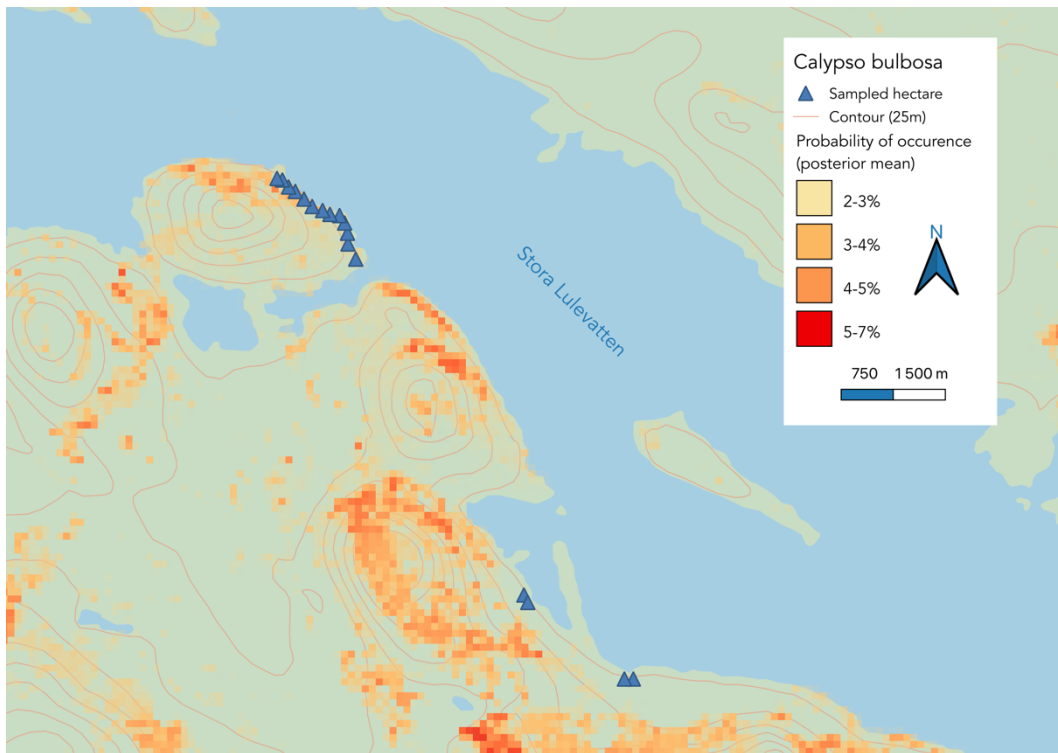*Figure 19: Heat map showing likelihood of C. bulbosa in Kvikkjokk using probability data from (Stephan et al., 2022). Background map ©2020 Lantmäteriet*

This area had some of the highest probability of *C. bulbosa* presence in the model (figure 19). The purple area representing the highest probability of occurrence was inaccessible over the fast-moving Gamájåhkå river. No individuals of *C. bulbosa* were observed, though another orchid, *Dactylorhiza maculata* subsp. *Maculata*, was seen in the area.

3.2.11 Tjåmotis, Siejdejávrásj, Jokkmokk, Lu lpm, 20/06/23

Based on the successful discoveries on day one, a further investigation of the area surrounding Tjåmotis was planned. Despite both the proximity to the site discovered on day one (figure 20: pink polygon, bottom left) and the similarity in environmental conditions, no further individuals of *C. bulbosa* were observed in the area, although another orchid, *Neottia cordata*, was recorded.
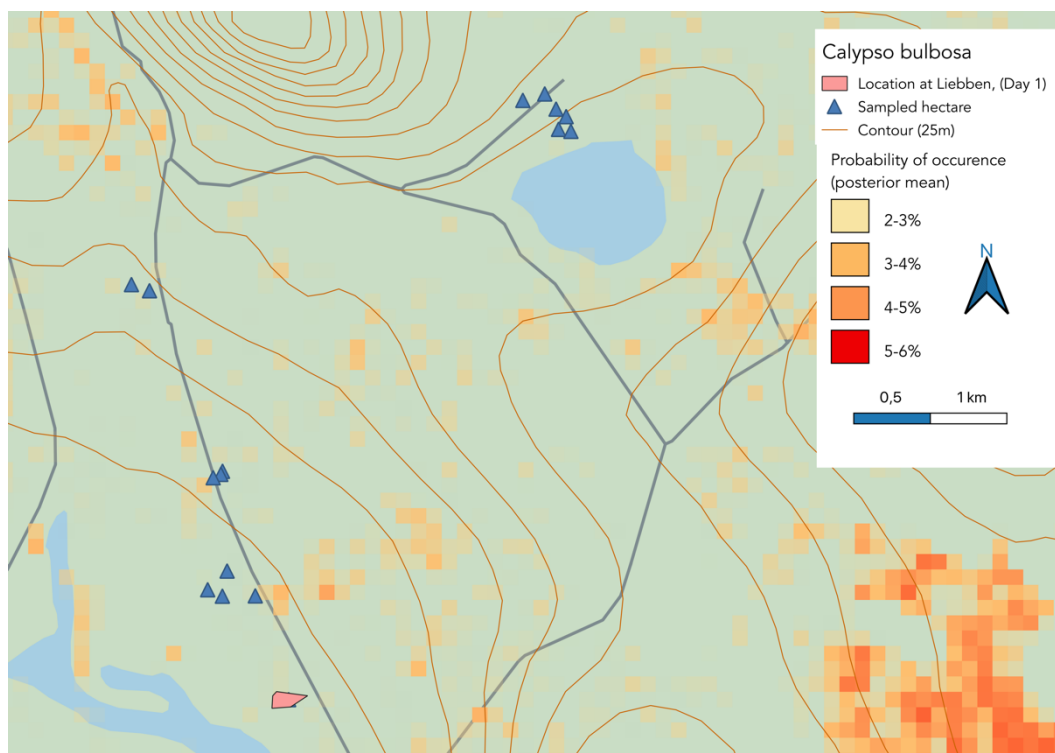
Figure 20: Heat map showing likelihood of C. bulbosa in the area around Heliga fallet, Tjåmotis. The pink polygon towards the south of the map represents the area where an estimated 1000-1500 individuals were found at Liebben, day 1. Probability data from (Stephan et al., 2022). Background map ©2020 Lantmäteriet

3.3 Other sites of interest

The following sites were visited to gather more information about *C. bulbosa* habitats and populations and are not included in the statistical analysis.

3.3.1 Serri NR, Jokkmokk, Lu lpm, 06/06/23

Serri is a nature reserve located to the southeast of Jokkmokk. The area is outside the alpine region but was deemed to be of interest due to a notable lack of *C. bulbosa* reports, which seemed unusual given the high number of suitable environments, lakes and watercourses within the park. Despite seemingly suitable locations and the presence of multiple companion species, no *C. bulbosa* locations were discovered. 30 individuals of the red-listed orchid *G. repens* were observed.

3.3.2 Murjek, Jokkmokk, Lu lpm, 09/06/23

Two previously recorded *C. bulbosa* locations were revisited around Murjek, including one recorded by biologist Mats Karström as part of a larger inventory of plant life in the area. *C. bulbosa* was not recorded in either location, despite the inclusion of a broad search area around both coordinates. At both sites visited signs of fairly recent clear cutting were noted in the surrounding area, which may

have altered site conditions to the detriment of the calypso orchid, resulting in a local extinction.

3.3.3 Jaurekaska, Jokkmokk, Lu lpm, 14/06/23

Figure 21: Left: The C. bulbosa supporting environment in Jaurekaska. Centre: A cluster of individuals growing in the mossy layer. Right: Specimens in-situ.
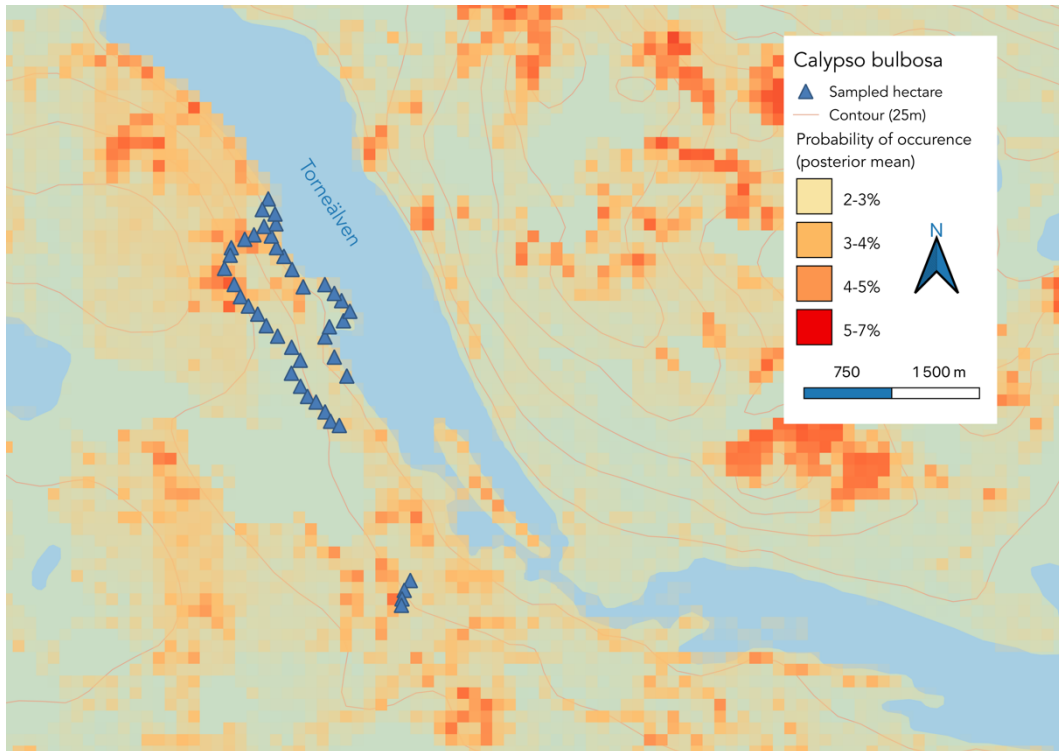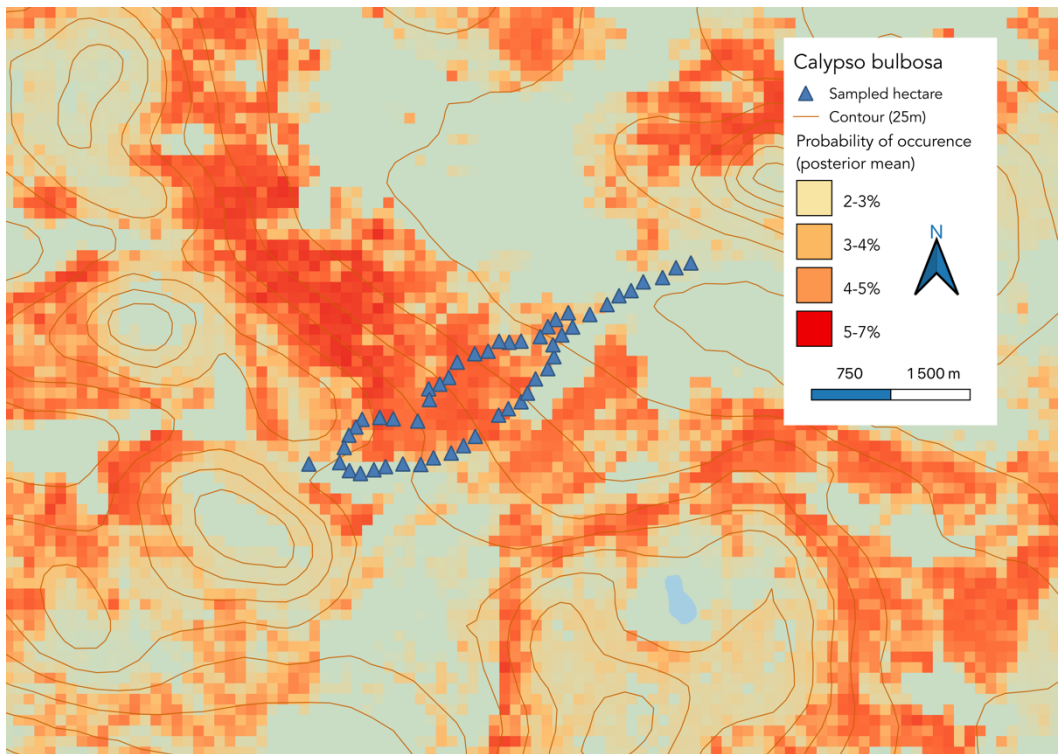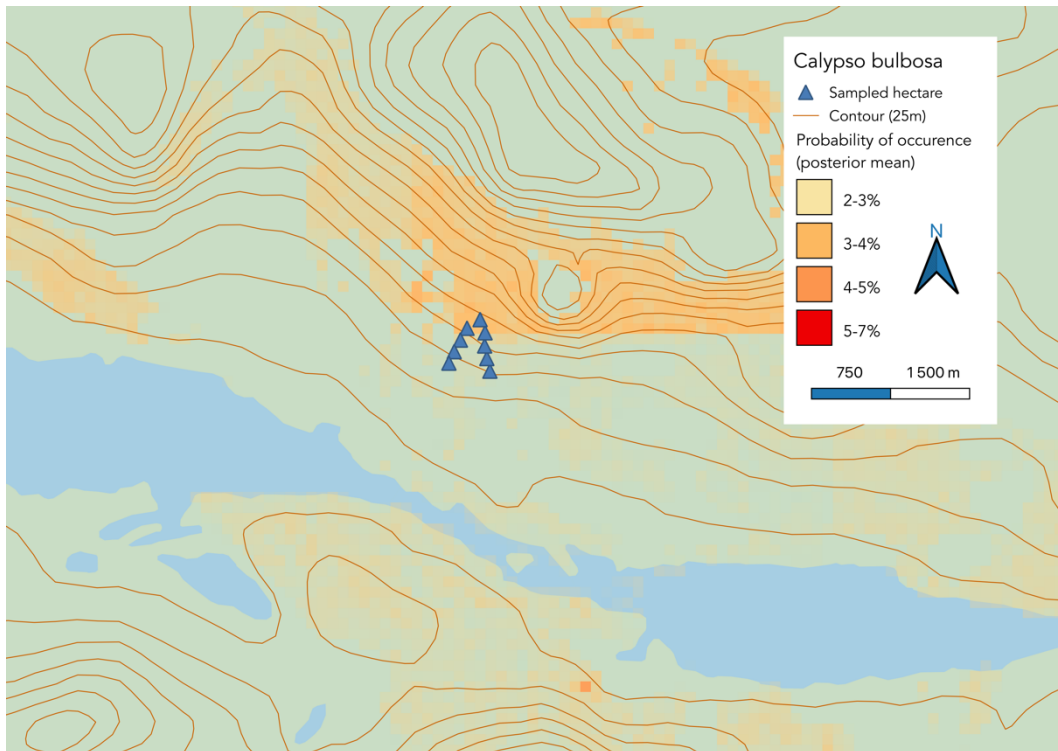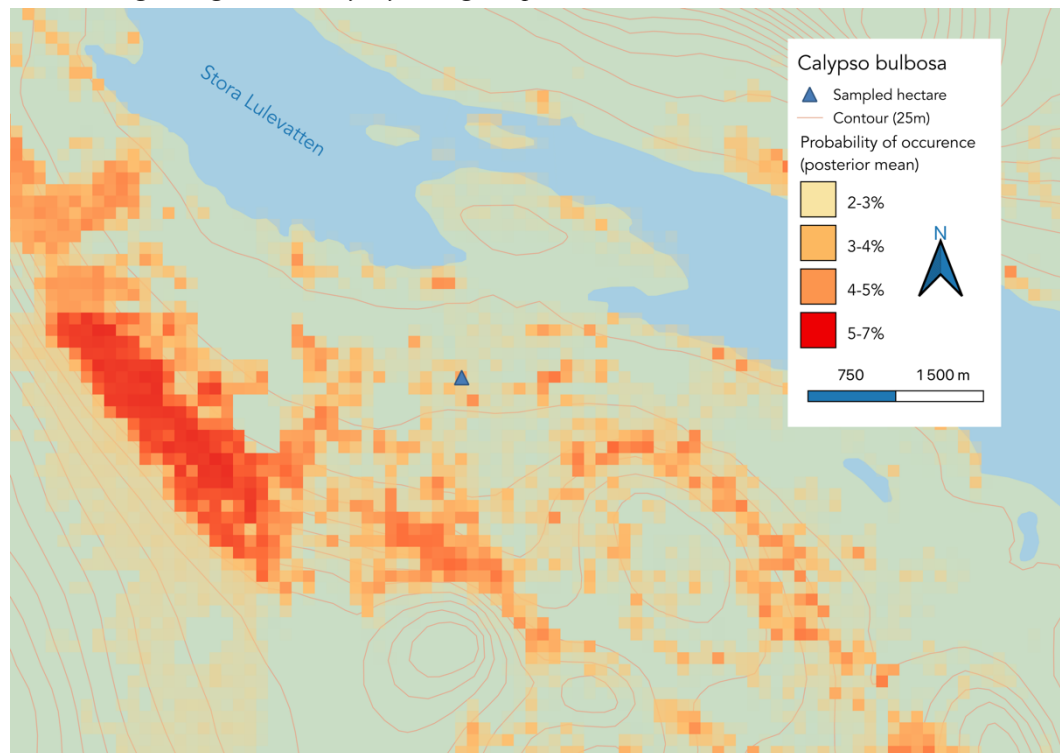

Figure 22: Heat map showing likelihood of C. bulbosa in area surrounding Jaurekaska using probability data from (Stephan et al., 2022). Background map ©2020 Lantmäteriet

Jaurekaska, a remote site accessible by boat from the northern shore of Stora Lulevatten, is a known *C. bulbosa* location with several years' worth of observations reported into Artportalen. It is also the most north-westerly *C. bulbosa* site reported in Sweden, situated well within the alpine region. The site was visited for

research purposes, primarily to observe the kinds of locations within the alpine region suitable for *C. bulbosa* growth. On the day of the field visit 204 individuals of *C. bulbosa* were inventoried, an increase from the far smaller number of plants previously recorded in the area. The plants observed at this site were comparatively modest in appearance compared to individuals observed further south.

Figure 23: The site flooded by melting snow on the 22 May 2023 (Photo: Lena Krey, 2023)

The site was located in a dip in the surrounding landscape, 380-382 m above sea level, with higher ground to the east and west and a slight slope downwards at the northern edge. Images recorded earlier in the year show that the area was covered in snow, which was then followed by a period of saturation caused by the melting snow layer (figure 23). At the time of visiting this flood water had long subsided, leaving a mesic, mixed and fairly open forest on rocky terrain with a moist humus layer varying from 1 to around 10 cm in depth. The ground was primarily covered by moss and small shrubs, with companion species recorded in the area including *P. schreberi, E. nigrum, V. myrtillus* and *L. borealis* (Figure 21, right).

3.4 Statistical analysis

Data from the period of field work was analysed in R (R core team, 2023) using a logistic regression to identify if there was a correlation between the observed presence/absence of *C. bulbosa* in the alpine region in comparison with the predicted presence/absence probability data from the 2022 SDM. The data analysed also includes data collected by Elina Larsson from her period of fieldwork in Jämtland (Larsson, 2023).

Every hectare quadrant located within the alpine region visited during the current period of fieldwork was plotted onto a map in QGIS, including a value for the presence or absence of *C. bulbosa* in that quadrant. The 2022 SDM prediction

values were then overlayed and extracted for each visited quadrant along with the observed presence/absence data using QGIS. This data was then imported into R, at which point a logistic regression was performed in order to calculate the power of the model to predict presence/absence based on predictions from the 2022 model data. An ROC curve was plotted using this data (Figure 24), and the AUC was then calculated for the dataset.
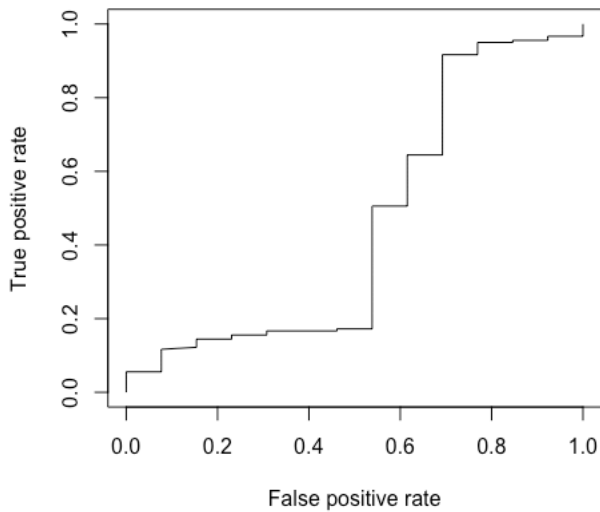
Figure 24: AUC visualised in R (R core team, 2023) using ROCR (Sing et al., 2005)

The AUC value of 0.455 represented in figure 24 implies that the model's predictions for the alpine region are no better than chance.

4. Discussion

The AUC value of 0,445 from the statistical analysis of the data collected in the fieldwork can be interpreted as being no better than random chance. An AUC of 0,7-0,8 is typically deemed to be acceptable, a value of 0,8-0,9 as excellent and >0,9 as outstanding. Prior to the assessment of the alpine region, the 2022 SDM showed an AUC=0,9 for all of Sweden when assessing the model's ability to predict the primarily boreal locations reported in to artportalen (Stephan et al., 2022). The lack of predictive power in alpine region contrasts to the significant discoveries of *C. bulbosa* in the alpine region made during the period of fieldwork, which will be discussed in detail below.

4.1 Alpine population of *C. bulbosa*

The site discovered on day one (Liebben, 2,5 km NE of Tjåmotis) is the largest *C. bulbosa* location discovered in the Swedish alpine region in the 21st century. This discovery casts serious doubt on current alpine population estimates of 1000 individuals (Naturvårdsverket, 2020) given that 738 individuals were observed in Liebben alone. Indeed, this figure of 738 is in itself likely to be an underestimation, with an estimated total of 1000 – 1500 individuals thought to be growing in this area.

During the period of fieldwork, a total of 1190 individuals were observed over three sites in the alpine region, a number equal to 119% of the entire alpine population at current estimations (Naturvårdsverket, 2020). Though the model may not have shown a significant predictive power for any given quadrant in the region, it does appear that the model's suggestion that the alpine region is more suitable for *C. bulbosa* occurrence than previously thought is correct.

The environmental conditions observed in the alpine locations with *C. bulbosa* presence in Jaurekaska, Liebben and Gárddevárjåhkå were similar in terms of:

- Forest type: Mesic, relatively open mixed forest consisting of coniferous trees (at times mixed with birch)
- Hydrology: located in a dip in the landscape or at the bottom of a slope where water could flow through but not stand still over a longer time period once snow had melted

- Humus layer present
- Low ground vegetation consisting primarily of a bottom layer of feather mosses and a field layer of herbs and dwarf shrubs, with no significant occurrence of calcium-indicating companion species

The majority of sites selected for survey were chosen based on information from the probability layer in QGIS, imported from the 2022 SDM. The site with the largest number of specimens recorded, however, was pinpointed with the help of a topographic map of the area, based on prior experience of ecologically suitable environments. This area did not appear particularly interesting in the model (figure 25). The model did, however, show that there were higher probability areas in the vicinity of the site that could indicate a potential population (figure 25), helping to focus the search. At this point in the model's development, it would seem that combining the models predictions with an ecological understanding of *C. bulbosa* may be a sound strategy for discovering new populations.
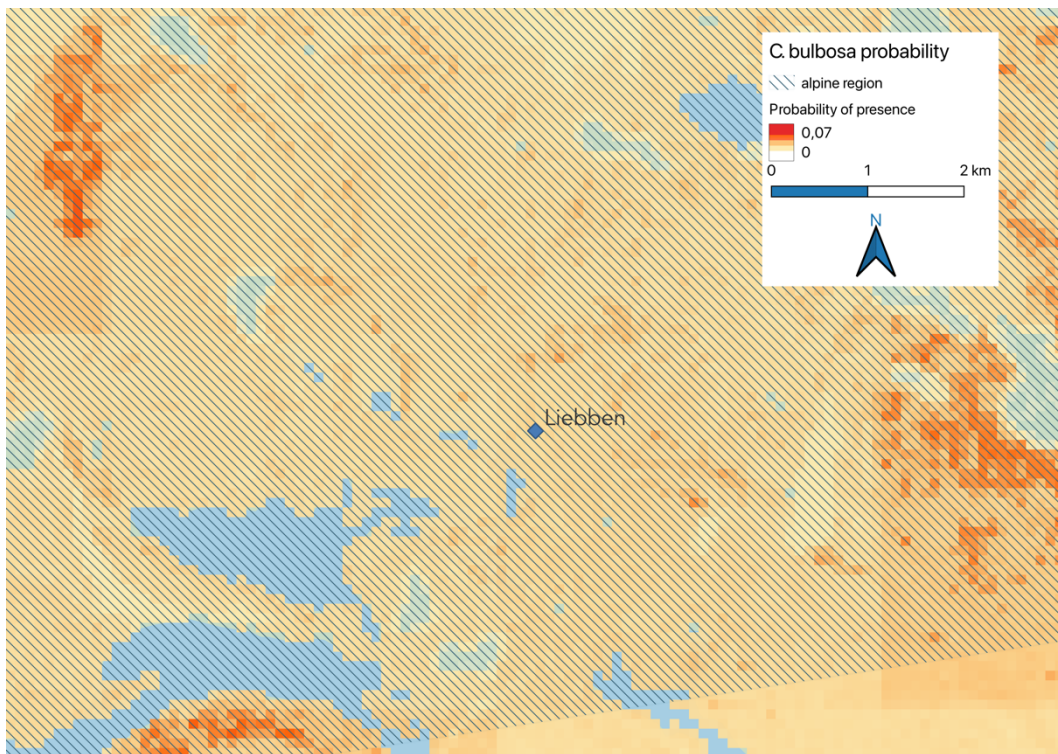
Figure 25: Probability data for area surrounding Liebben, image: Michael Doorly, 2023 with probability data from (Stephan et al., 2022). Background map ©2020 Lantmäteriet

Somewhat counterintuitively, ecological knowledge of *C. bulbosa* may also be a limiting factor in locating new populations. Prior knowledge of *C. bulbosa* distribution, for example that it is thought of as a primarily boreal plant, may have led to a phenomenon known as preferential sampling. Preferential sampling, the phenomenon in which areas may be under sampled based on an observer's judgement/presumptions (Edwards Jr. et al., 2006), could be a factor that has negatively affected surveying for the species in the alpine region historically, the

assumption being that as *C. bulbosa* is not commonly thought to grow in this area in any significant number, it is therefore not actively looked for. Given that the short window of appearance when the plant is at its most visible is situated early in the year when few hikers are visiting the region, it seems plausible the species is under-surveyed in the area, creating a self-fulfilling absence of data for a species that demonstrably can thrive within this region given the right location and conditions.

Even when individuals are observed and recorded, it can still be a challenge to produce accurate estimations from observer figures (Morrison and Young, 2016). Imperfect detection leads to omission of presences in presence-background data sets (Lahoz-Monfort et al., 2014) and can skew results. Additionally, *C. bulbosa* specimens documented in the alpine region appeared to be smaller and more modest in appearance compared to the specimens seen in the boreal region, which could explain the low number of individuals observed in Jaurekaska when compared to the more thorough inventory of the site undertaken in this study.

4.2 Challenges for the SDM in the alpine region

Even if the model's suggestion of *C. bulbosa* presence in the alpine region has now led to new and significant populations being discovered, the poor predictive power of the model for any given quadrant in this region is ultimately no better than random chance. The site discovered on day one in Liebben, for example, does not appear particular interesting in the model (figure 25), despite containing a significant *C. bulbosa* population.

In the original study AUC=0.90 (Stephan et al., 2022) meaning that the models predictive power explained observed variation in species occurrence well for the whole of Sweden. In this case, the majority of the presence/absence reports used for the ROC analysis were from within the boreal region, where *C. bulbosa* is at its most prevalent. What follows is an analysis of various factors that may have limited the model's capabilities for making accurate predictions in the alpine region, followed by suggestions for improvement.

Factors that may reduce the predictive power of an SDM, introduced in section 1.2, include the inherent limitations of modelling the full complexity of species niches (both fundamental and realised), missing explanatory values, bias originating from the type of data used, as well as limitations in the type of model used and the way in which it is implemented.

4.2.1 Uncertainty in the model

Uncertainty and bias from measurement- and systematic error, natural variation, model uncertainty, inherent randomness and subjective judgement can all interfere with the accuracy of a species distribution model (Regan et al., 2002). Whilst bias is a factor in any SDM based on OR data, it can be difficult to determine how and to what extent this bias has impacted results (Elith et al.,

2002), and whether in the case of the 2022 SDM this bias can meaningfully explain the poor performance of the model in the alpine region, when model performance for northern Sweden as a whole was otherwise deemed to be good.

Epistemic errors such as imperfect measurements can produce random variation in results (Elith et al., 2002). Measurement errors can originate from inaccurate GPS locations for reported presences in the OR data as well as from the GIS-data used for habitat modelling, as these interpolations of field-based measurements, satellite images and aerial photos can all potentially contain different degrees of measurement error. The heat map for probability from the 2022 SDM (figure 19), for example, shows a clear linear demarcation where percentage probability suddenly drops substantially as if following a grid line where data collection methods appear to have differed in some meaningful way for the model.

Systematic error can also effect results and create bias, e.g. oversampling close to roads (Phillips et al., 2009). Currently most *C. bulbosa* observations reported in to Artportalen are in places close to where people live and visit, in the boreal region, often near to the side of a road. Whilst maps produced from OR data may be slightly less accurate as a result of this tendency towards "roadside bias", they are usually not significantly so (Kadmon et al., 2004), with OR based SDMs often having even greater predictive power for rare species (Henckel et al., 2020).

The problem of roadside bias should theoretically be minimal as long as the data is consistently biased across all its applications (Phillips et al., 2009), however given the differences in population density and infrastructure further into the alpine region, this bias may not be so consistent. The relationship between the species and the road network is also complex, as the distribution of roads in Northern Sweden is frequently linked to the locations of bodies of water, rivers, lakes, damming and the bottom of slopes for practical reasons. Whilst these locations are often the most natural sites for road construction, they are also often sites with conditions favourable to *C. bulbosa*, complicating the distinction between cause and effect.

4.2.2 Explanatory factors – fundamental vs realised niche

The ecological niches of a species are key considerations when formulating any model. As discussed in section 1.2.1, predicting species distribution can be understood through three main conditions (Guisan et al., 2017):

1. The species potential for dispersal: the species must be able to reach the site
2. A species-suitable abiotic environment
3. A species-suitable biotic environment (interactions with other species)

The 11 explanatory variables selected for their predictive power in the 2022 SDM only paint a partial picture of the complex life history of *C. bulbosa.* They can be seen as indirectly representing a range of other crucial factors that are far harder to quantify and collect data on, such as biotic interspecies relationships. As previously discussed, an SDM will model the realised, not the fundamental niche

of the species (Williams et al., 2009). In the case of alpine vs boreal region, the assumptions made from the data the model is based on, collected almost entirely in the boreal region, may not be applicable to the plant in a different biogeographical region where it and many of the species it interacts with are existing at the extreme end of their tolerance to cold.

The quality of the information about a species in an unsampled (or in this case, less frequently sampled) location depends heavily on the quality of the data collected in the sampled locations (Tessarolo et al., 2021). This data quality is compromised when the unsampled area is dependent on data from sites that are spatially distant and environmentally dissimilar from it. Given, then, that the model is based upon observations that originate overwhelmingly from the boreal region, it seems plausible that these observations may not represent the observed niche of the species in the alpine region, and this could account, in part, for the low predictive power in this area. One such example could be that biotic species interactions may not be the same for *C. bulbosa* in the alpine and boreal regions. *C. bulbosa* is reliant on biotic interspecies reactions, for example bumblebee populations (Alexandersson and Ågren, 2000) as well as a reliance on mycorrhiza for increased root surface area and germination (Currah et al., 1988).

Neither of these biotic factors are directly considered in the explanatory variables from the 2022 SDM. Instead, the most tangible link to biodiversity originates from the 70 related plant species built into the model when creating pseudo-absences (Stephan et al., 2022). The process of generating pseudo-absences could be made more reliable by introducing more stringent criteria for the observers whose data qualifies for inclusion, as a less experienced observer is more likely to miss the presence of *C. bulbosa* in a given quadrant. To solve this issue, questionnaires could be sent to the observers who have been active in the relevant quadrants within the flowering period of *C. bulbosa,* as per (Henckel, et al. 2020). Additionally, using common species to develop models for rare ones has been shown to provide low predictive extrapolation which may increase the likelihood of prediction error in a model (Thomas C. Edwards et al., 2005).

In multiple sites visited over the period of fieldwork, conditions appeared to be suitable for *C. bulbosa*, though despite this the plant was ultimately absent in the majority of these areas. The explanation for this absence may be due, in part, to the climate being at the extreme cold end of species tolerance, though there are likely other factors involved. This raises questions of how much variation in the distribution of the species was stochastic in nature, or limited by dispersal effects in the alpine region, and if these effects are perhaps more prevalent in the alpine region where geographical conditions such as mountains and bare rock faces are more common. These conditions create larger geographical distances between potential sites which, combined with the limited dispersal capabilities of the plant (Alexandersson and Ågren, 1996), could mean that isolated populations have more difficulty dispersing to new sites to colonise (Guisan et al., 2017).

The quantification of niches powering the 2022 SDM could be missing other key explanatory values that may be more determinative for *C. bulbosa* distribution in

the alpine region in comparison to the boreal. From field visits it appears that snow melt (timing of when snow falls and how long it stays put before melting, run off and eventual water saturation) appears to be important, however this factor would be difficult to quantify directly.

4.3   Assessment of fieldwork

A stricter adherence to the principles of transect sampling (Ringvall et al., 2007) may have reduced the potential bias of the fieldwork caused by selective sampling of areas with better conditions for *C. bulbosa* growth. Additionally, more notes could have been made about frequently associated species, as these are important for the creation of pseudo-absences in future modelling based on companion species. However, the majority of these companion species, such as *V. vitis-idaea* and *L. borealis,* or the feather mosses *H. splendens* and *P. schreberi,* for example, are ubiquitous in the northern forests of Sweden, and their presence or absence in isolation gives little indication of *C. bulbosa* occurrence. The fieldwork could also have been improved by visiting more remote places, either by camping to allow longer hikes to more remote areas, or via access to a boat, as many hotspots were located on the far side of a river far away from any access road.

4.4   Improving the model

The original study used Hmsc (Tikhonov et al., 2022) in R (R core team, 2023) to produce the 2022 SDM. Other studies utilise different models for handling similar presence-only data with good results, for example GAM, ENFA (Zaniewski et al., 2002) or RF, which provided the best predictions in a 2009 comparison of SDMs (Williams et al., 2009) and ultimately lead to the discovery of 16 new rare plant populations.

A different approach, referred to as the "committee averaging method" (Gallien et al., 2012) is a concept in which several models are used (combining regressions, classification trees and machine learning) based on the premise that there is no perfect algorithm. Predicted probability maps that have been made by these models are then transformed into binary maps via a threshold, then averaged to create one final map. In this approach each model has therefore "voted" as part of the "committee" as to whether the species will be present/absent in a quadrant, resulting not in a probability of occurrence but instead a percentage of agreement between the different algorithms. The concept is similar to that of Bayesian Model Averaging (Wintle et al., 2003).

Ultimately, the Hmsc model gave good results in the initial study on Sweden as a whole (AUC=0.9), with the power of the model vindicated by a 2019 review of models which deemed that HMSC.3 provided the most accurate predictions, especially when predicting for rare species (Norberg et al., 2019). It is therefore not the suggestion of this paper that the model needs to be changed in any major

way, though results could be cross validated with GLM.5, SAM.1 and GLM.12 (Norberg et al., 2019).

High quality information visualisation is crucial for communicating with policy makers (Elith et al., 2002). With the current level of uncertainty in the model for the alpine region it is suggested that an "ignorance map" is created to communicate the areas in which predictions are weaker (Tessarolo et al., 2021), at least until more data has been collected from the alpine region and AUC for the area has been improved. These maps are based on 'biogeographical ignorance', which incorporates measures of the quality of distributional data as well as its longevity (time elapsed since data was collected) and distance between locations with and without presence/absence information (Tessarolo et al., 2021).

4.5 Areas for further study

Accurate SDM modelling in collaboration with OR data from citizen scientists and volunteer organisations such as the Swedish organisation of 'Flora Guardians' (*Swedish: 'Floraväktarna'*) is crucial to monitor the population size of *C. bulbosa* as the impact of a changing climate becomes ever clearer. Climate change and land use have been shown to reduce accessibility to suitable habitat, especially in rare and poor dispersal species (De Kort et al., 2020).

The relationship between orchids and climate change is complex, with other orchid species shifting northwards as edaphic conditions change (Smallwood and Trapnell, 2022), which raises the question of whether the Swedish *C. bulbosa* population could be moving northwards, expanding into the alpine region. This could explain the historic lack of detection in the region, though confirming the theory would take years of data collection over a large area.

Climate change poses its own challenges to modelling, as the situation on the ground is changing fast making it difficult for models to keep pace (De Kort et al., 2020). Regardless of whether the population is increasing in the alpine region or simply that we are now discovering new growing sites, *C. bulbosa* population in Sweden is generally declining (Naturvårdsverket, 2020). Restoration of ecological qualities needs to include production forests and not just protected areas. The focus needs to be on a diverse landscape with a mix of large trees, dead trees, deciduous trees, natural processes such as fire, and the encouragement of natural structures and processes (Nordlind and Östlund, 2003).

5. Conclusions

The conclusions from this study suggest that, to improve the predictive power of the SDM in the alpine region, several options are available:

- More data should be collected from surveying sites in the alpine region, perhaps in collaboration with volunteer organisations such as the Flora Guardians, to create a more complete picture
- A separate model for the alpine region based entirely on presence/absence data from the region could be developed until sufficient data has been collected to reduce some of the bias from the original model
- The addition of new explanatory variables such as distance to nearest known location to account for dispersal challenges in the alpine region may improve predictive power
- Implementation of a filter for forest age to prioritise data collection in old growth areas favoured by *C. bulbosa*
- Hmsc.3 model results for the separate alpine region modelling could be cross validated with GLM.5, SAM.1 and GLM.12 (Norberg et al., 2019)
- Creation of an 'ignorance map' that visualises areas with higher and lower uncertainty in the model's predictive power, increasing its utility for researchers and policy makers (Tessarolo et al., 2021)

As suggested in the original study, data from field visits such as those undertaken for this paper should be used to improve the otherwise presence-only model (Williams et al., 2009), perhaps culminating in an iterative SDM where the model could be updated with each addition of new population and absence data from field work. The predictive power of the SDM should improve with the addition of more presence/absence data.

Despite its lack of accuracy on the hectare scale, in this study alone the 2022 SDM helped discover new *C. bulbosa* locations in the alpine region, as well as giving just cause to re-evaluate the population estimation figures for the area. An updated and more accurate SDM for the Swedish *C. bulbosa* population could help to better monitor population decline in the boreal region, track the effects of climate change on the population both in the alpine region and nationwide, as well as hopefully encouraging greater conservation efforts for this charming, vulnerable orchid.

References

ACKERMAN, J. D. 1981. Pollination biology of *Calypso bulbosa* var. *occidentalis* (Orchidaceae)*:* A food-deception system. *Madroño,* 28**,** 101-110.

AHRNÉ, K., BJELKE, U., NORDSTRÖM, S., OTTOSSON, E., SANDSTRÖM, J., SUNDBERG, S. & EIDE, W. 2020. *Tillstånd och trender för arter och deras livsmiljöer : rödlistade arter i Sverige 2020,* Uppsala, SLU Artdatabanken, Sveriges lantbruksuniversitet.

ALEXANDERSSON, R. & ÅGREN, J. 1996. Population size, pollinator visitation and fruit production in the deceptive orchid Calypso bulbosa. *Oecologia***,** 533-540.

ALEXANDERSSON, R. & ÅGREN, J. 2000. Genetic structure in the nonrewarding, bumblebee-pollinated orchid Calypso bulbosa. *Heredity***,** 401-409.

ARTDATABANKEN. 2019. *Artfakta* [Online]. SLU Artdatabanken. Available: https://artfakta.se/naturvard/taxon/calypso-bulbosa-233 [Accessed 04.05.2023].

BERGLUND, H. 2015. Hänsyn till arter i samband med skogsbruk. *In:* ARTDATABANKEN (ed.). Artfakta: SLU.

CURRAH, R. S., HAMBLETON, S. & SMRECIU, A. 1988. Mycorrhizae and mycorrhizal fungi of Calypso bulbosa. *Americal Journal of Botany,* 75**,** 739-752.

DE KORT, H., BAGUETTE, M., LENOIR, J. & STEVENS, V. M. 2020. Toward reliable habitat suitability and accessibility models in an era of multiple environmental stressors. *Ecology and Evolution,* 10**,** 10937-10952.

DEVICTOR, V., WHITTAKER, R. J. & BELTRAME, C. 2010. Beyond scarcity: citizen science programmes as useful tools for conservation biogeography. *Diversity and Distributions,* 16**,** 354-362.

EDWARDS JR., T. C., CUTLER, D. R., ZIMMERMAN, N. E., GEISER, L. & MOISEN, G. G. 2006. Effects of sample survey design on the accuracy of classification tree models in species distribution models. *Ecological Modelling,* 199**,** 132-141.

EEA 2016. Calypso bulbosa: Report under the Article 17 of the Habitats Directive Period 2007 - 2012. European Environment Agency.

ELITH, J., BURGMAN, M. A. & REGAN, H. M. 2002. Mapping epistemic uncertainties and vague concepts in predictions of species distribution. *Ecological modelling,* 157**,** 313-329.

EUROPEAN ENVIRONMENT AGENCY, E. T. C. O. B. D. 2022. Reporting format referred to in Article 17 of Directive 92/43/EEC (Habitats Directive). Available:

https://cdr.eionet.europa.eu/help/habitats_art17/Reporting2025/Art.17%20 report%20format%202019-2024.pdf/.

FELTON, A., P.O., H., LINDBLADH, M., NYBERG, T., FELTON, A. M., HOLMSTRÖM, E., WALLIN, I., LÖF, M. & BRUNET, J. 2016. The biodiversity contribution of wood plantations: Contrasting the bird communities of Sweden's protected and production oak forests. *Forest Ecology and Management,* 365.

GALLIEN, L., DOUZET, R., PRATTE, S., ZIMMERMAN, N. E. & THUILLER, W. 2012. Invasive species distribution models – how violating the equilibrium assumption can create new insights. *Global Ecology and Biogeography,* 21**,** 1126-1136.

GRAHAM, C. H., FERRIER, S., HUETTMAN, F., MORITZ, C. & PETERSON, A. T. 2004. New developments in museum-based informatics and applications in biodiversity analysis. *Trends in Ecology and Evolution,* 19**,** 497-503.

GUISAN, A., THUILLER, W. & ZIMMERMAN, N. E. 2017. *Habitat Suitability and Distribution Models with applications in R,* United Kingdom, Cambridge University Press.

HENCKEL, L., BRADTER, U., JÖNSSON, M., ISAAC, N. J. B. & SNÄLL, T. 2020. Assessing the usefulness of citizen science data for habitat suitability modelling: Opportunistic reporting versus sampling based on a systematic protocol. *Diversity and Distributions,* 26**,** 1276–1290.

HIGGINS, S. I., O'HARA, R. B. & RÖMERMANN, C. 2012. A niche for biology in species distribution models. *Journal of Biogeography,* 39**,** 2091-2095.

IUCN 2001. *IUCN Red List Categories and Criteria Version 3.1,* Gland, Switzerland and Cambridge, UK, IUCN - The World Conservation Union.

JOHNSON, C. J. & GILLINGHAM, M. P. 2007. Sensitivity of species-distribution models to error, bias, and model design: An application to resource selection functions for woodland caribou. *Ecological modelling,* 213**,** 143-155.

KADMON, R., FARBER, O. & AVINOAM, D. 2004. Effect of roadside bias on the accuracy of predictive maps produced by bioclimatic models. *Ecological Applications,* 14**,** 401-413.

KOLANOWSKA, M. 2021. The future of a montane orchid species and the impact of climate change on the distribution of its pollinators and magnet species. *Global Ecology and Conservation,* 32**,** e01939.

LAHOZ-MONFORT, J. J., GURUTZETA, G. A. & WINTLE, B. A. 2014. Imperfect detection impacts the performance of species distribution models. *Global Ecology and Biogeography,* 23**,** 504-515.

LARSSON, E. 2023. *Inventerarens underbara resa genom Jämtland - träffsäkerheten i artutbredningsmodell för norna (Calypso bulbosa) testad med fokus på Jämtlands län.* Bachelor thesis, Uppsala University.

LAW, S. K. & YEUNG, E. C. 1989. Embryology of *Calypso Bulbosa*. I. Ovule development. *American Journal of Botany,* 76**,** 1668-1674.

MAINA, G. G. & HOWE, H. F. 2000. Inherent Rarity in Community Restoration. *Conservation Biology,* 14**,** 1335-1340.

MORRISON, L. W. & YOUNG, C. C. 2016. Observer error in sampling a rare plant population. *Plant Ecology & Diversity,* 9**,** 289-297.

MOSSBERG, B. & STENBERG, L. 2018. *Nordens flora*, Bonnier Fakta.

NATURVÅRDSVERKET 2020. Sveriges arter och naturtyper i EU:s art- och habitatdirektiv. Resultat från rapportering 2019 till EU av bevarandestatus 2013–2018.

NORBERG, A., et al. 2019. A comprehensive evaluation of predictive performance of 33 species distribution models at species and community levels. *Ecological Monographs,* 89.

NORDLIND, E. & ÖSTLUND, L. 2003. Retrospective comparative analysis as a tool for ecological restoration: a case study in a Swedish boreal forest. *Forestry,* 76**,** 243-251.

PHILLIPS, S. J., DUDÍK, M., ELITH, J., GRAHAM, C. H., LEHMANN, A., LEATHWICK, J. & FERRIER, S. 2009. Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. *Ecological applications,* 19**,** 181-197.

PROCTOR, H. C. & HARDER, L. D. 1995. Effect of pollination success on floral longevity in the orchid *Calypso bulbosa* (orchidaceae). *Americal Journal of Botany,* 82**,** 1131-1136.

PUCHNINA, L. V. 2017. Status of *Calypso bulbosa* and *Cypredium calceolus* (Orchidaceae) populations in the Pinega state nature reserve. *Nature Conservation Research,* 2**,** 125-150.

R CORE TEAM 2023. R: A language and Environment for statsitcal computing. Vienna, Austria: R Foundation for statistical computing.

REGAN, H. M., COLYCAN, M. & BURGMAN, M. A. 2002. A Taxonomy and treatment of uncertainty for ecology and conservation biology. *Ecological applications,* 12**,** 618-628.

RINGVALL, A., SNÄLL, T., EKSTRÖM, M. & STÅHL, G. 2007. Unrestricted guided transect sampling for surveying sparse species. *Canadian Journal of Forest Research,* 37**,** 2575 - 2586.

SING, T., SANDER, O., BEERENWINKEL, N. & LENGAUER, T. 2005. ROCR: visualizing classifier performance in R. *Bioinformatics,* 21**,** 7881.

SLU. 2023. *Artportalen* [Online]. Available: www.artportalen.se [Accessed].

SMALLWOOD, P. A. & TRAPNELL, D. W. 2022. Species distribution modeling reveals recent shifts in suitable habitat for six North American Cypripedium spp. (Orchidaceae). *Diversity,* 14**,** 694.

SMITH, T. M. & SMITH, R. L. 2015. *Elements of Ecology,* England, Pearson.

STEPHAN, J., WIKBERG, S. & SUNDBERG, S. 2022. Spatial modelling of habitat suitability for *Calypso bulbosa*, a protected plant species in boreal moist forest. *SLU Artdatabanken,* SLU.dha.2021.5.2-87.

TESSAROLO, G., LADLE, R. J., LOBO, J. M., RANGEL, T. F. & HORTAL, J. 2021. Using maps of biogeographical ignorance to reveal the uncertainty in distributional data hidden in species distribution models. *Ecography,* 44**,** 1743-1755.

THOMAS C. EDWARDS, J., CUTLER, D. R., ZIMMERMANN, N. E., GEISER, L. & ALEGRIA, J. 2005. Model-based stratifications for enhancing the detection of rare ecological events. *Ecology,* 86**,** 1081-1090.

THUILLER, W., GEORGES, D., GUEGUEN, M., ENGLER, R., BREINER, F., LAFOURCADE, B. & PATIN, R. 2023. biomod2: Ensemble platform for species distribution modeling. R package version 4. 2-4 ed.

TIKHONOV, G., OVASKAINEN, O., OKSANEN, J., DE JONGE, M., OPEDAL, O. & DALLAS, T. 2022. Hmsc: Hierarchical model of species communities. R package version 3.0-13 ed.

TRAILL, L. W., LIM, M. L. M., SODHI, N. S. & BRADSHAW, C. J. A. 2010. Mechanisms driving change: altered species interactions and ecosystem function through global warming. *Journal of Animal Ecology,* 79**,** 937-947.

WEISKOPF, S. R., et al. 2020. Climate change effects on biodiversity, ecosystems, ecosystem services, and natural resource management in the United States. *Science of the Total Environment,* 733.

WILLIAMS, J. N., CHANGWAN, S., THORNE, J., NELSON, J. K., ERWIN, S., O'BRIEN, J. M. & SCHWARTZ, M. W. 2009. Using species distribution models to predict new occurrences for rare plants. *Diversity and Distributions,* 15**,** 565-576.

WINTLE, B. A., MCCARTHY, M. A., VOLINSKY, C. T. & KAVANAGH, R. P. 2003. The use of Bayesian model averaging to better represent uncertainty in ecological models. *Conservation Biology,* 17**,** 1579-1590.

ZANIEWSKI, A. E., LEHMANN, A. & OVERTON, J. M. 2002. Predicting species spatial distributions using presence-only data: a case study of native New Zealand ferns. *Ecological modelling,* 157**,** 261-280.

Acknowledgements

A huge thank you to Zebbe for his infectious enthusiasm for botany, as well as his generosity in sharing both his time and extensive knowledge with me on this fascinating project, and to my examiner Joachim Strengbom for taking the time to assess this thesis. Thank you also to Lena for rowing me across Store Lulevatten on a windy day to show me a unique Norna habitat on Jaurekaska. Thanks too to Kurt Kjær from the university of Copenhagen for lending me a GPS (and mosquito spray) at the last minute, to Artportalen and all the team involved in developing the 2022 SDM that formed the basis for this project and to all the citizen scientists out there for collecting and sharing their data.

Publishing and archiving

Approved students' theses at SLU are published electronically. As a student, you have the copyright to your own work and need to approve the electronic publishing. If you check the box for **YES**, the full text (pdf file) and metadata will be visible and searchable online. If you check the box for **NO**, only the metadata and the abstract will be visible and searchable online. Nevertheless, when the document is uploaded it will still be archived as a digital file. If you are more than one author, the checked box will be applied to all authors. You will find a link to SLU's publishing agreement here:

- https://libanswers.slu.se/en/faq/228318.

☒ YES, I/we hereby give permission to publish the present thesis in accordance with the SLU agreement regarding the transfer of the right to publish a work.

☐ NO, I/we do not give permission to publish the present work. The work will still be archived and its metadata and abstract will be visible and searchable.