Gorteria – Dutch Botanical Archives 40, 2018: 073–078
ISSN (online) 2542-8578

SHORT COMMUNICATION

An approach to calculate a Species Temperature Index for flora based on open data

L.B. Sparrius[1], G.G. van den Top[2], C.A.M. van Swaay[2]

Key words

Ascomycotina
Bryophytes
Climate change
GBIF
Open data
Plantae
R
Bootstrapping
STI
WorldClim

Abstract – To study the relation between changes in local plant communities and shifts in temperature, the temperature niches of the species need to be defined. We calculated the Species Temperature Index (STI) as a proxy for this niche. STI is defined as the mean annual temperature within the range of a species. In this paper, a method is described to calculate the STI for the European flora from open data provided by the Global Biodiversity Information Facility (GBIF) and WorldClim global climate data, for 7254 taxa of European vascular plants, bryophytes, algae, and ascomycetes including lichens. The algorithm accounts for incomplete and unbalanced species distribution data. The Community Temperature Index (CTI) is defined as the weighted mean of the STIs of a species assemblage. A 1 × 1 km CTI grid map for the Netherlands is presented as an example of the use of STIs.

Samenvatting – Om de relatie tussen de veranderingen in lokale plantengemeenschappen en verschuivingen in de temperatuur te bestuderen is het nodig om de temperatuurniche van de individuele soort te kwantificeren. Wij berekenden de Species Temperature Index (STI) als een proxy voor deze niche. STI wordt gedefinieerd als de gemiddelde jaartemperatuur binnen het verspreidingsgebied van een soort. In dit artikel beschrijven we een methode om de STI voor de Europese flora te berekenen op basis van open data voor 7254 taxa van Europese vaatplanten, mossen, algen en korstmossen. De methode is ontworpen om te corrigeren voor incomplete of ruimtelijk ongelijk verdeelde verspreidingsgegevens. De Community Temperature Index (CTI) wordt gedefinieerd als het gewogen gemiddelde van STI's voor gemeenschappen van soorten. Een 1 × 1 km CTI rasterkaart voor Nederland wordt gepresenteerd als voorbeeld van het gebruik van STI's.

Published on 30 October 2018

INTRODUCTION

Climate change is often regarded as one of the major factors affecting biodiversity (Mace et al. 2005; Thuiller et al. 2005b, c; Parmesan 2006). Many species are adapted to a certain temperature niche (e.g. Dullinger et al. 2009) and climatic change may act as a selective pressure causing a decline of species that cannot adapt well to this shift and an increase in species with high adaptation capabilities and dispersal capacity (Thuiller et al. 2005a; Devictor et al. 2008; Devictor et al. 2012).

In order to study the relation between changes in local plant communities and shifts in temperature, the temperature niche needs to be quantified. A previous study (Thuiller et al. 2005b) used range maps for a limited number of vascular plants, a proprietary dataset by Atlas Florae Europaeae. In this paper we describe the calculation of the Species Temperature Index (STI) for European vascular plants, bryophytes, algae, and ascomycetes including lichens based on open data. The increasing number of data published by the Global Biodiversity Information Facility (GBIF) allows researchers to include many more taxa in their study. However, species occurrence data in GBIF has significant data gaps. Therefore, we describe a bootstrapping method to minimize the effect of biased data.

The Species Temperature Index is defined as the long-term average temperature within a species range. The Community Temperature Index (CTI) is defined as the (weighted) average STI of a species assemblage (Devictor et al. 2008). The CTI

[1] FLORON, Toernooiveld 1, 6525 ED Nijmegen; e-mail: sparrius@floron.nl;
[2] De Vlinderstichting, Mennonietenweg 10, 6702 AD Wageningen;
e-mail: ggvdtop@gmail.com;
e-mail: chris.vanswaay@vlinderstichting.nl

corresponding author e-mail: sparrius@floron.nl

can be used for a trend-analysis if time-series are available, and for spatial analysis. The STI can be used as a quantitative replacement for the commonly used temperature indicator values (Ellenberg 1992).

MATERIALS AND METHODS

Study area

Data analysis in this study is restricted to the Western Palaearctic: Europe, Northern Africa and parts of the Middle East (Fig. 1).

Climate data

We used the WorldClim2 BIO1 dataset (Fick & Hijmans 2017), which contains the annual average temperature over the period 1970–2000 on a 30” grid (Fig. 1).

Species occurrence data

All occurrences of the taxonomical groups Plantae and Ascomycota were retrieved from GBIF (GBIF.org 2017). The period of observations ranges from 1970–2017, which was used as the climate dataset. This resulted in 70.3 million observations of 67023 taxa. The data shows a highly incomplete and biased distribution of the study area (Fig 1b), due to factors such as the lack of monitoring data in remote areas and the fact that not all available data has been published in GBIF.

Synonyms and accepted taxa

For vascular plants and bryophytes, the Taxonomic Name Resolution Service v4.0 (Boyle et al. 2013) with data from The Plant List (2015) was used to identify synonyms and acceptated taxa based on the “scientificname” field in the GBIF dataset. For fungi Mycobank (Robert et al. 2005) was used. Occurrence data of synonyms was merged with records of their accepted names. In the case of 27010 taxa, the name could not be found in The Plant List or Mycobank. The “speciesname” field in the GBIF dataset was then used, ignoring infraspecific taxa and accepted names according to the GBIF taxonomic backbone. This procedure was chosen because of a number of incorrect synonym relations were found in the GBIF taxonomic backbone, e.g. *Angelica sylvestris* L. being treated as a synonym of *Anthriscus sylvestris* (L.) Hoffm.

Of the accepted names, only taxa with more than 250 occurrences within the study area were selected to avoid errors due to a lack of data. Handling of large datasets was performed with the latest versions of EmEditor, MySQL, and QGIS.

STI calculation

In order to correct for the unbalanced data distribution (Fig. 1b), we used the following bootstrapping approach. At first we superimposed a 50 km UTM grid (European Environment Agency 2016) on the observation data and determined for each species which grid cells are occupied. This step not only makes the data more workable, but it also removes duplicate observation per grid cell. This is necessary because the STI is only based on presence/absence rather than abundance. Next we calculated the mean annual temperature for each 50 km UTM grid cell. The normal approach would be to perform a direct calculation of the mean temperature over the range of each species. Due to the nature of the data we decided to take a different approach.

A 250 km grid was superimposed on the 50 km UTM grid. Within each 250 km grid cell we randomly chose one 50 km grid cell that was occupied by the species of interest. This results in a more even distribution of 50 km UTM grid cells over the range of Europe and Northern Africa. These grid cells were used to calculate the mean annual temperature over the range of the species by repeating this process a 100 times and averaging the temperature estimation. The resulting STI is the average within the standard deviation of these 100 samples. These steps were performed in R. We published the R-script in a software repository (Sparrius et al. 2018). For comparison we also calculated the STI without the bootstrapping approach. In this case, the same GBIF data was projected on the 50 km grid cells and the mean temperature of all grid cells together was calculated.

We applied this STI dataset to the following case:

Plant community temperature index map of the Netherlands

The Plant Community Temperature Index (plant CTI) is here defined as the average STI of all vascular plants occurring in a km square, similar to Devictor (2008, 2012). This principle was applied to distribution data of vascular plants in the Netherlands (NDFF 2017). The CTI was calculated for each km square with more than 150 plant species, which we consider as sufficiently surveyed. No extrapolation to areas with less than 150 species was performed. We accepted that this selection left out a small number of grid cells located on the border of the country, grid cells that fall for a major part in sea or lakes, or are known to have a low plant diversity, such as extensive inland dunes, dry heathlands and pine forest.

RESULTS

The STI was calculated for 7254 taxa (Sparrius et al. 2018: data repository) based on 66 million occurrences. The bootstrapping method resulted on average in c. 0.27 °C lower STIs (Fig. 2) than a non-bootstrapped approach. Although only taxa with a minimum of 250 observations were used, taxa occurring in a low number of observations (e.g. 1000) tend to have a higher standard deviation (Fig. 3). The CTI map of the vascular flora of the Netherlands (Fig. 4) shows a gradient from the warmer southern and western parts of the country to the colder northeast. The average temperature indicated by vascular plants is 8.7 °C.

DISCUSSION

Here we show a revised method for the calculation of the STI over a broad geographical range. This method is better at taking care of unbalanced observation data than the classical method without a bootstrapping procedure.

The Species Temperature Index (STI) has proven to be a valuable tool to study the effect of climate change on birds (Devictor et al. 2008) and butterflies (Devictor et al. 2012). Devictor et al. (2008) also show that it is a robust parameter in studying the effect on changes in the Community Temperature Index (CTI), even when the STI is replaced by its rank number or when the community data is restricted to presence/absence only. This makes the STI for plants, as documented in this paper, a valuable additional tool, which makes it possible to study the effects of climate change on taxa for which range maps are largely unavailable. The GBIF occurrence dataset shows a bias towards Western Europe, Mediterrean Europe, and Scandinavia, with lower data coverage in the boreal zone, mountainous areas in Eastern Europe, Russia, and Northern Africa. Bootstrapping led to lower STIs than the method without bootstrapping, as all 250 × 250 km regions were equally sampled. This also becomes visible in Fig 1b, with many white areas in cooler northeastern part of the study area.

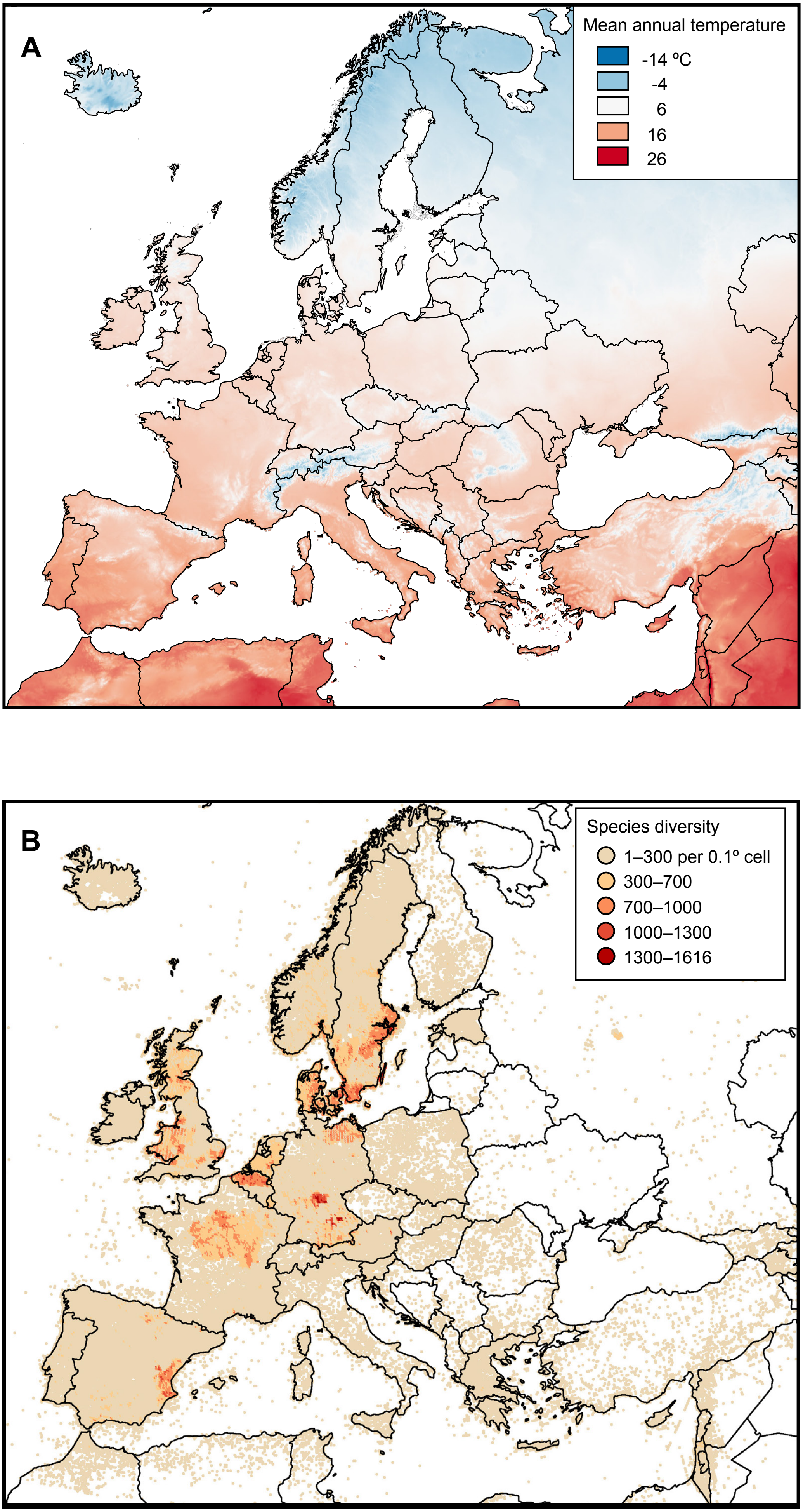

Fig. 1. Maps of (A) the average annual temperature in Europe and (B) the available data of plant diversity for Europe in GBIF (1970–2017).

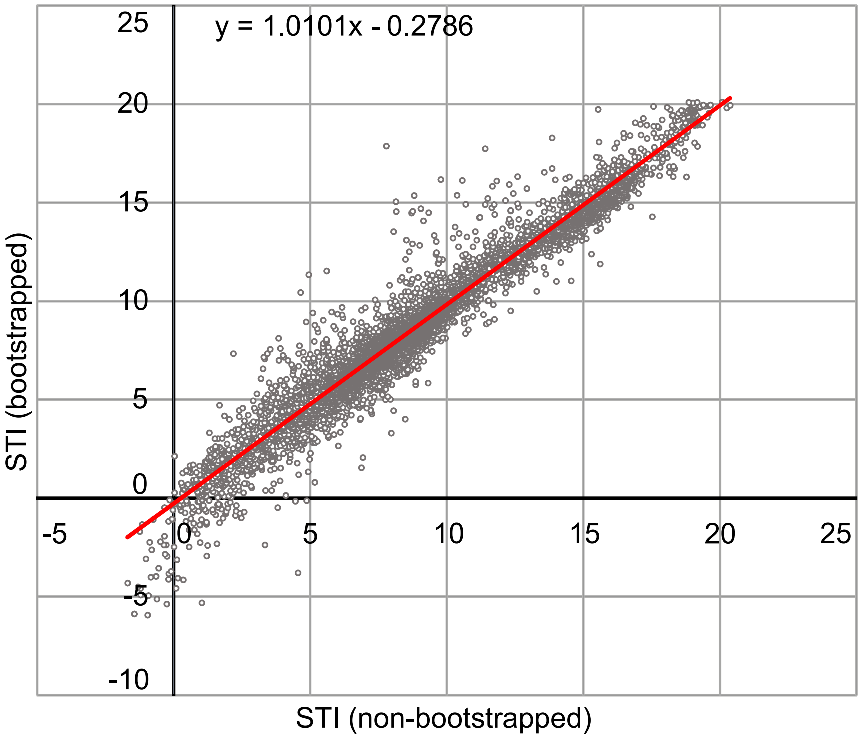

Fig. 2. Species Temperature Index (STI) calculated with all available GBIF data versus the bootstrapping approach. Trend line (linear regression) is shown in red together with its equation. Each data point represents a taxon (n = 7254).

The standard deviation increased with sample size, probably because common species have a wider distribution area. However, a low sample size led more frequently to higher standard deviations. This could be the case for taxa with disjunct distribution, but high standard deviations are more probably caused by data gaps near the center of the distribution area, or in the case of poorly known taxa with a wide distribution.

Map

Coastal areas, cities and river valleys show a higher CTI. For coastal areas, this can be explained by the more temperate climate near a large water body (Stefan et al. 1998). The temperature in cities is usually higher than the surrounding areas due to heating, drought and presence of stone surfaces that easily heat up in the sun. Western European city centers and can be up to 2 °C warmer than the surrounding landscape (Santamouris 2007). River valleys are under influence of species dispersal from upstream areas (Boedeltje et al. 2004; Johansson et al. 1996). In the Netherlands, upstream areas of the Rivers Meuse and Rhine extend towards the south. This phenomenon is supported by the fact that small rivers, with a limited catchment area, do not show up as warm spots on the map. Warm spots in the river valleys are especially visible in areas with extensive sand extraction followed by development of new natural areas with opportunities for pioneer vegetation (Rossenaar et al. 2006).

The average temperature indicated by vascular plants is 8.7 °C. Meteorological data shows that this was a commonly occurring annual average before 1980 (KNMI 2018). The average temperature in the country between 2007 and 2017 was 10.6 °C. This suggests that the climate in the Netherlands could offer favorable conditions for many Southern European taxa.

Potential applications

STIs can be used as a more precise replacement for temperature indicator values (Ellenberg 1992) being a continuous variable instead of an ordinal scale. STIs can also be used to predict the effects of climate change on the flora (Bertrand et al. 2011; Thuiller 2005b), although niche modelling remains the best solution to predict shifts in species distributions (e.g. Thuiller 2009). For example, species with a high dispersal capacity and an STI equal to or slightly higher than the average temperature within a certain area may be expected there in future. The STI may also be of use for risk assessments for invasive species (Thuiller 2005c).

Data quality

As demonstrated in Fig. 3, the accuracy of STIs calculated with the bootstrap method may improve in future when more data becomes available. Alternatively, instead of GBIF data, range maps of e.g. Atlas Florae Europaeae may be used as demonstrated by Thuiller et al. (2005a), however these maps are not publicly available and are cover only 20% of the vascular flora.

Primary data in GBIF is not free of errors either. A small percentage of records may have errors in identification or geolocation. However, this would not dramatically affect the STI when the sample size is large enough. A minimum sample size of 250 observations was chosen in this study.

With large standard deviations, caused by large distribution areas, individual STIs must be used with care, but can be useful in studies where many STIs are combined or related to other variables.

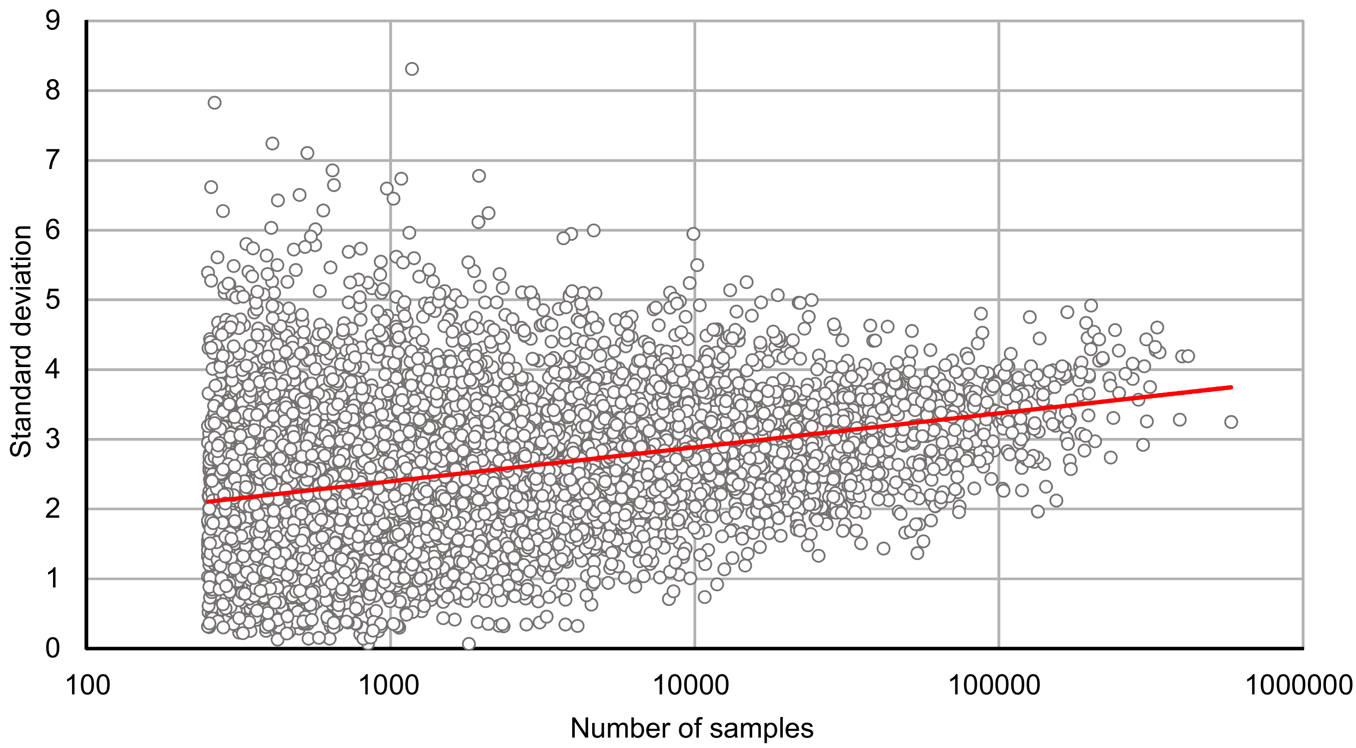

Fig. 3. The relation between the sample size (number of observations) and standard deviation of a species in the Species Temperature Index (STI) calculations. Trend line (linear regression) is shown in red.

CONCLUSIONS

The Species Temperature Index for the European flora can be calculated from open data sources. As more and more data becomes available, it is likely that STIs can be calculated more accurately and more species can be included in the calculation.

Acknowledgements – The work presented in this paper has been made possible by a grant of WWF Netherlands.

REFERENCES

Bertrand R, Lenoir J, Piedallu C, Riofrío-Dillon G, De Ruffray P, Vidal C, Pierrat J-C, Gégout JC. 2011. Changes in plant community composition lag behind climate warming in lowland forests. Nature 479: 517–520.

Boedeltje G, Bakker JP, ten Brinke A, van Groenendael JM, Soesbergen M. 2004. Dispersal phenology of hydrochorous plants in relation to discharge, seed release time and buoyancy of seeds: the flood pulse concept supported. J. Ecol. 92: 786–796; doi: 10.1111/j.0022-0477.2004.00906.x

Boyle B, Hopkins N, Lu Z, Raygoza Garay JA, Mozzherin D, Rees T, Matasci N, Narro ML, Piel WH, Mckay SJ, Lowry S, Freeland C, Peet RK, Enquist BJ. 2013. The taxonomic name resolution service: an online tool for automated standardization of plant names. BMC Bioinformatics 14:16. doi: 10.1186/1471-2105-14-16.

Devictor V, Julliard R, Couvet D., Jiguet F. 2008. Birds are tracking climate warming, but not fast enough. Proc. Roy. Soc. Biol. Sci. Ser. B 275: 2743–2748.

Devictor V, Van Swaay C, Brereton T, Brotons L, Chamberlain D , Heliölä J, Herrando S, Julliard R, Kuussaari M, Lindström Å, Reif J, Roy DB , Schweiger O, Settele J, Stefanescu C, Van Strien A, Van Turnhout C, Vermouzek Z, WallisDeVries M, Wynhoff I, Jiguet F. 2012. Differences in the climatic depts of birds and butterflies at a continental scale. Nature Climate Change 2: 121–124.

Dullinger S, Kleinbauer I, Peterseil J, Smolik M, Essl F. 2009. Niche based distribution modelling of an invasive alien plant: effects of population status, propagule pressure and invasion history. Biol. Invas. 11: 2401–2414; doi: 10.1007/s10530-009-9424-5.

Ellenberg H, Weber HE , Düll R, Wirth V, Werner W, Paulißen D. 1992. Zeigerwerte von pflanzen in Mitteleuropa. Verlag Erich Goltze KG, Göttingen.

European Environment Agency. 2016. EEA reference grids. See http://eea.europa.eu/data-and-maps/data/eea-reference-grids.

Fick SE, Hijmans RJ. 2017. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. International Journal of climatology (in press). doi: 10.1002/joc.5086

GBIF.org. 20th March 2017. GBIF occurrence download http://doi.org/10.15468/dl.qw9lpv

Johansson ME, Nilsson C, Nilsson E. 1996. Do rivers function as corridors for plant dispersal? J. Veg. Sci. 7: 593–598; doi: 10.2307/3236309.

KNMI. 2018. Weerstatistieken, accessed through https://weerstatistieken.nl/

Mace G, Masundire H, Baillie JEM. 2005. Biodiversity. In: Hassan R, Scholes R, Ash N (eds), Ecosystems and human well-being: Current state and trends: findings of the condition and trends working group: 77–122. Island Press, Washington.

NDFF. 2017. Vascular plant distribution data download http://www.verspreidingsatlas.nl/.

Parmesan C. 2006. Ecological and evolutionary responses to recent climate change. Annual Rev. Ecol. Evol. Syst. 37: 637–669.

Robert V, Stegehuis G, Stalpers J. 2005. The MycoBank engine and related databases. See: http://www.mycobank.org (accessed April 2017).

Rossenaar AJGA, Odé B, Beringen R. 2006. Natuurontwikkeling en flora langs de grote rivieren. Levende Natuur 107: 237–241.

Santamouris M. 2007. Heat island research in Europe: the state of the art. Advances in Building Energy Research 1: 123–150.

Sparrius LB, van Strien AJ. 2014. Het berekenen van jaarlijkse trends van planten op basis van verspreidingsgegevens. Gorteria 37: 31–40.

Sparrius LB, van der Top G, van Swaay CAM. 2018. R script for species temperature index with open data (version 1.0). Zenodo. See http://doi.org/10.5281/zenodo.1155850.

Stefan HG, Fang X, Hondzo M. 1998. Simulated climate change effects on year-round water temperatures in temperate zone lakes. Climatic Change 40: 547–576.

The Plant List. 2013. Version 1.1. Published on the Internet; http://www.theplantlist.org/ [Accessed: 1 June 2017].

Thuiller W, Lavorel S, Araújo MB. 2005a. Niche properties and geographical extent as predictors of species sensitivity to climate change.

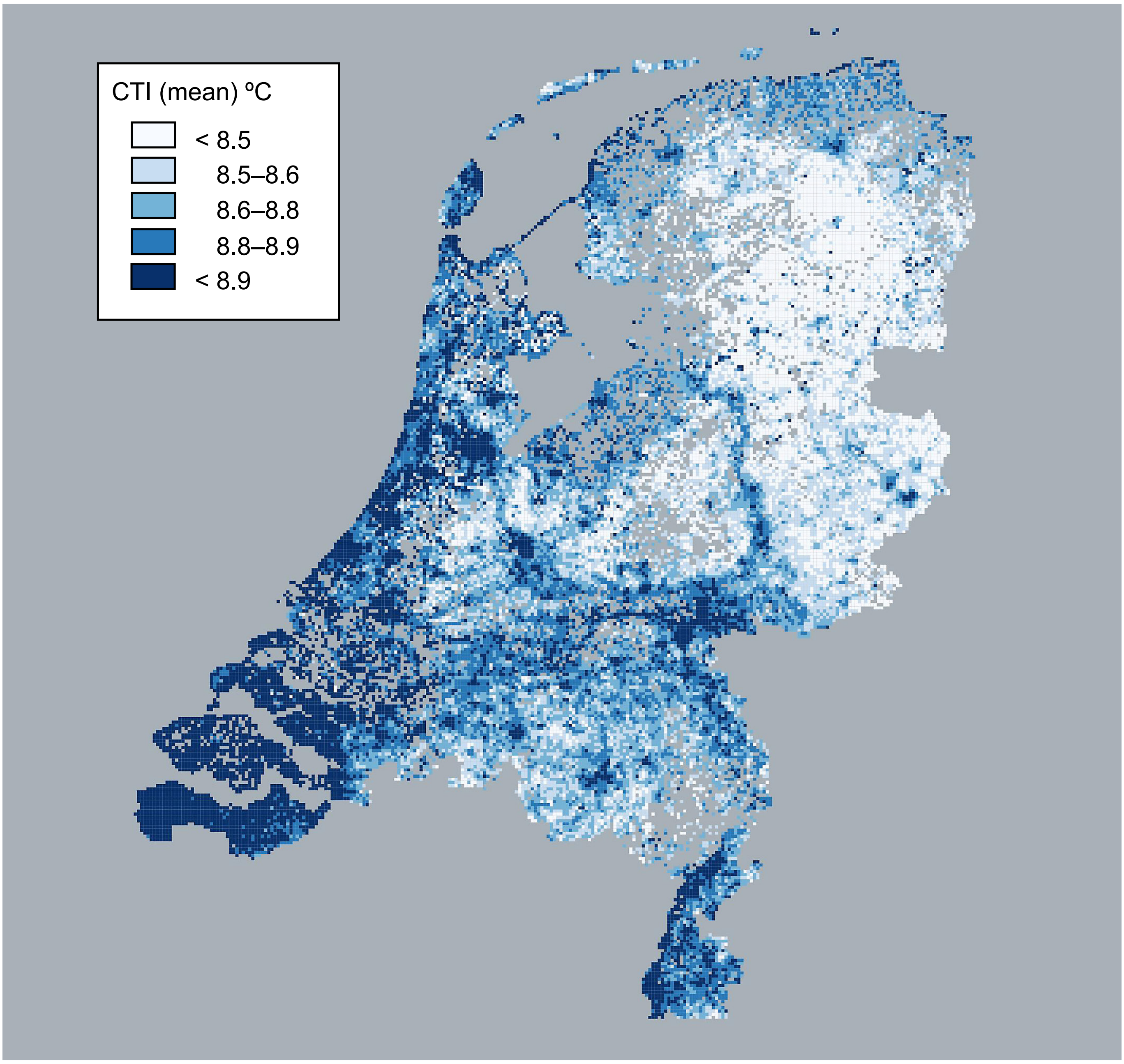

Fig. 4. Map of the Netherlands with the Community Temperature Index (CTI) of plants expressed as the mean Species Temperature Index (STI) for all plants found in km squares.

Thuiller W, Lavorel S, Araújo MB, Sykes MT, Prentice IC. 2005b. Climate change threats to plant diversity in Europe. PNAS 102: 8245–8250; doi: 10.1073/pnas.0409902102.

Thuiller W, Richardson DM, Pyšek P, Midgley GF, Hughes GO, Rouget M. 2005c. Niche-based modelling as a tool for predicting the risk of alien plant invasions at a global scale. Global Change Biology 11: 2234–2250; doi: 10.1111/j.1365-2486.2005.001018.x

Thuiller W, Lafourcade B, Engler R, Araujo MB. 2009. BIOMOD – A platform for ensemble forecasting of species distributions. Ecography 32: 369–373.

van Strien AJ, van Swaay CAM, Termaat T. 2013. Opportunistic citizen science data of animal species produce reliable estimates of distribution trends if analysed with occupancy models. J. Appl. Ecol. 50: 1450–1458; http://dx.doi.org/10.1111/1365-2664.12158.