bioRxiv preprint doi: https://doi.org/10.1101/009589; this version posted September 25, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

earthzine
FOSTERING EARTH OBSERVATION & GLOBAL AWARENESS
OES IEEE

Accepted manuscript

Applying geospatial semantic-array programming for a reproducible set of bioclimatic indices in Europe

Giovanni Caudullo [1]

[1] European Commission, Joint Research Centre, Institute for Environment and Sustainability, Forest Resources and Climate Unit, Via E. Fermi 2749, I-21027 Ispra, VA, Italy

This manuscript has been accepted
for publication in IEEE Earthzine.
IEEE Earthzine 2014 Vol. 7 Issue 2, (not yet published)+
2nd quarter theme. Geospatial Semantic Array Programming

Abstract

Bioclimate-driven regression analysis is a widely used approach for modelling ecological niches and zonation. Although the bioclimatic complexity of the European continent is high, a particular combination of twelve climatic and topographic covariates was recently found able to reliably reproduce the FAO ecological zoning for forest resources assessment at pan-European scale, generating the first fuzzy similarity map of FAO ecozones in Europe. The reproducible procedure followed to derive this collection of bioclimatic indices is now presented. It required an integration of data-transformation modules (D-TM) using both geospatial tools such as GIS software, and array-based mathematical implementation such as semantic array programming (SemAP). Base variables, intermediate and final covariates are described and semantically defined by providing the workflow of D-TMs and the mathematical formulation following the SemAP notation. Source layers to derive base variables were extracted by exclusively relying on global-scale public open geodata in order for the same set of bioclimatic covariates to be reproducible in any region worldwide. In particular, two freely available datasets were exploited for temperature and precipitation (WorldClim) and elevation (Global Multi-resolution Terrain Elevation Data). The working extent covers the whole European continent to the Urals with a resolution of 30 arc-second. The proposed set of bioclimatic covariates will be made available as open data in the European Forest Data Centre (EFDAC). The forthcoming complete set of D-TM codelets will enable the twelve covariates to be easily reproduced and expanded through free software.

 © 2014 IEEE

bioRxiv preprint doi: https://doi.org/10.1101/009589; this version posted September 25, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Accepted manuscript

Introduction

The integrated use of geospatial tools and regression analyses (from basic approaches such as distance weighting, polynomial interpolation and regression, up to more complex nonlinear regression techniques) is widely expanding for modelling habitat suitability, ecological niches and zonation starting from field observations - i.e. presence-only or presence/absence of species (*e.g.* [1][2][3]). The selection of predictor variables is an important factor for developing species distribution models [4]. In this work a set of twelve bioclimatic covariates are derived from publicly-available datasets and proposed for ecological niche modelling and zonation. This particular combination of climatic and topographic indices has been elaborated by de Rigo et al. [5] to reliably reproduce the FAO ecological zoning for forest resources assessment at pan-European scale, generating the first fuzzy similarity map of FAO ecozones in Europe. The original layer of FAO ecological zones was produced on the basis of bioclimatic variables and potential vegetation and was subject to expert-based refinement [6]. The final product is distributed as a crisp (i.e. non-fuzzy) vector file. The application of a model based on the relative distance similarity (RDS) technique allowed each ecological zone to be compared with climatically and geographically analogous grid-cells, reducing boundary artefacts and modelling a fuzzy raster map at 1 km^2. The analogous approach has been used in Ciscar et al. [7] performing the RDS method to estimate the impacts of habitat suitability change of a forest tree species over Europe for different future climate scenarios. The maximum suitability maps have been computed on the basis of field survey datasets of tree species presence/absence, modelled with a set of bioclimatic and geographic indicators. This proposed set of predictors has been selected on the basis of ecological and bio-geographical knowledge and adapted to the European climate. In this paper these covariates are presented with emphasis on their definitions, technical details and modelling procedures. The approach of reproducibility has been followed in all processes in order to allow these covariates to be easily re- used and/or modified for further ecological regression analyses of niche/zoning modelling.

Datasets

The datasets used for extracting base variables are freely available as GIS layers at global scale.

WorldClime v1.4: is a set of global climate layers with a spatial resolution of 30-arc-second, which consists of four climate variables: precipitation and mean, minimum, maximum temperature. Data were gathered from a variety of resources, covering as a current baseline the period from 1950 to 2000. Future scenarios developed with different models are also available. There are twelve layers for every variable, representing its 50-years monthly average values [8].

Global Multi-resolution Terrain Elevation Data 2010 (GMTED): is an enhanced elevation model developed by U.S. Geological Survey and the National Geospatial-Intelligence Agency at global scale. The product, available in different spatial resolutions up to 7.5-arc-second, is based on data derived from 11 raster-based elevation sources, and aggregated in six products: minimum, maximum, mean and median elevation, standard deviation of elevation, systematic subsample, breakline emphasis [9].

Methods

The process for modelling the final covariates required different data-transformation modules (D-TMs) integrating geospatial tools and array-based mathematical implementations such as semantic array programming – i.e. GeoSemAP [10].

bioRxiv preprint doi: https://doi.org/10.1101/009589; this version posted September 25, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

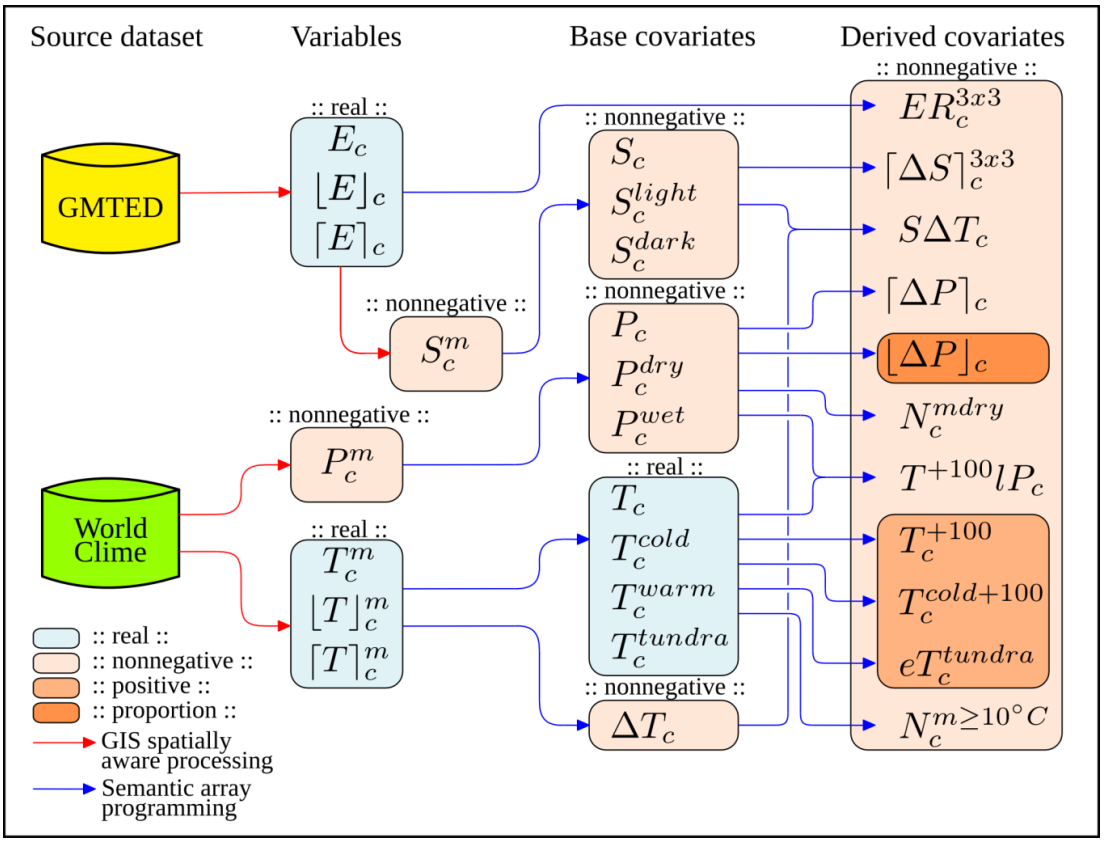

Figure 1. Process modelling workflow: the arrows represent the data transformation modules to derive the indices, which are grouped and coloured according to their semantics.

The workflow in Figure 1 shows D-TMs from source dataset to bioclimatic covariates. The colours of arrows define the transformation methods, while the colours of the grouping subsets of indices define the semantics of the outputs and intermediate layers. GIS technologies (red arrows) have been used principally for extracting data from source layers and preparing base covariates as a TIFF raster. In particular, Python scripting [11] with GDAL/OGR libraries [12][13] and Grass GIS v.7 libraries [14] have been used to operate with spatial layers. Data-transformation for converting raster maps into semantic array and SemAP procedure (blue arrows) have been based on GNU Octave/MATLAB programming language [15] with Mastrave library [16] and on Python language with SciPy library [17].

The working spatial environment is established to cover Europe completely to the Urals (north 72°, south 28°, east 75°, west -25°). The adopted projection is WGS84 with a spatial resolution of 30-arc-second, operating with TIFF raster files with a dimension of approximately 63 million pixels as floating point values. Variables are defined in each spatial grid c and for each month m. In Table 1 both variables and covariates are described with relative symbols, descriptions and reference equations. The temperature T is expressed in degree Celsius + 100°, avoiding negative values. Temperature without symbols is defined as the average value of the month, while symbol ⌈ ⌉ defines the maximum of the month and the symbol ⌊ ⌋ the minimum. The precipitation P is in millimetres and values state the total sum of monthly precipitation. The elevation E without symbols is the average value inside the 1 km grid measured in metres. The solar irradiation S is expressed in watt-hour per square metre (Whm^{-2}). Quantities averaged with a spatial moving window of 3 by 3 km are denoted with ^{3x3}. The definition of variable average, maximum and minimum guarantees them to be sematically sortable ensuring ::nonnegative:: values.

Accepted manuscript

bioRxiv preprint doi: https://doi.org/10.1101/009589; this version posted September 25, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Table 1. List of the computed variables and covariates with relative descriptions and equations.

Variable code	Description of variable
E_c	Mean elevation
$\lfloor E \rfloor_c$	Minimum elevation
$\lceil E \rceil_c$	Maximum elevation
S_c^m	Total solar irradiation in the central day of the month m
P_c^m	Total precipitation of the month m
T_c^m	Average temperature of the month m
$\lfloor T \rfloor_c^m$	Minimum temperature of the month m
$\lceil T \rceil_c^m$	Maximum temperature of the month m

Base covariate	Description of the base covariate	Reference equation
S_c^{light}	Total potential solar irr. of the 6 lighter months	
S_c^{dark}	Total potential solar irr. of the 6 darker months	
P_c	Annual total precipitation	$\sum_m P_c^m$
P_c^{dry}	Total precipitation of the driest month	$\min_m P_c^m$
P_c^{wet}	Total precipitation of the wettest month	$\max_m P_c^m$
T_c	Annual average temperature	$\sum_m T_c^m$
T_c^{cold}	Average temperature of the coldest month	$\min_m T_c^m$
T_c^{warm}	Average temperature of the warmest month	$\max_m T_c^m$
T_c^{tundra}	Nordenskiöld index	$T_c^{\text{warm}} + 0.1 T_c^{\text{cold}} - 9$
ΔT_c	Mean of monthly temperature range	$1/12 \sum_m \lceil T \rceil_c^m - \lfloor T \rfloor_c^m$

	Derived covariates	Description of derived covariate	Reference equation
1.	$ER_c^{3\times3}$	$3\,km \times 3\,km$ average elevation range	$(\lceil E \rceil_c - \lfloor E \rfloor_c)^{3\times3}$
2.	S_c	Annual potential solar irradiation	$S_c^{\text{light}} + S_c^{\text{dark}}$
3.	$\lceil \Delta S \rceil_c^{3\times3}$	$3\,km \times 3\,km$ seasonal variation of pot. solar irr. (normalised by semestral minimum)	$\frac{S_c^{\text{light},3\times3} - S_c^{\text{dark},3\times3}}{S_c^{\text{dark},3\times3}}$
4.	$S\Delta T_c$	Potential irradiation by montly temp. range (simplified avg. monthly solar entransy flux)	$S_c \cdot \Delta T_c$
5.	$\lceil \Delta P \rceil_c$	Seasonal variation of monthly precipitation (normalised by minimum: driest month)	$\frac{P_c^{\text{wet}} - P_c^{\text{dry}}}{P_c^{\text{dry}}}$
6.	$\lfloor \Delta P \rfloor_c$	Seasonal variation of monthly precipitation (normalised by maximum: wettest month)	$\frac{P_c^{\text{wet}} - P_c^{\text{dry}}}{P_c^{\text{wet}}}$
7.	$N_c^{m\text{dry}}$	Number of dry months	$\sum_m [P_c^m \leq 2T_c^m]$
8.	$T^{+100} l P_c$	Shifted temperature per rain magnitude order	$T_c^{+100} / \log_{10}(P_c + 1)$
9.	T_c^{+100}	Shifted average annual temperature	$T_c + 100\,°C$
10.	$T_c^{\text{cold}+100}$	Shifted average temp. of the coldest month	$T_c^{\text{cold}} + 100\,°C$
11.	eT_c^{tundra}	Nordenskiöld exponential index	$\exp(T_c^{\text{tundra}})$
12.	$N_c^{m \geq 10\,°C}$	Number of months with $T_c \geq 10\,°C$	$\sum_m [T_c^m \geq 10]$

Temperature and Precipitation variables (from WorldClime raster dataset) have "No Data" values covering water pixels, converted to "not a number" (NaN) values on semantic arrays. NaN creates a water mask which propagates on array computations up to derived covariates. The Tundra covariate is identified by the equation of Nordenskiold and Mecking [18]. The average temperature of coldest month and the number of dry months and months with a temperature higher than 10°C are defined by Food and Agriculture Organization of the United Nations [6] for the ecological zones. The combination of potential solar irradiation by temperature range is a simplification of the solar entransy flux [19][20]. Covariates counting the number of months were

Accepted manuscript

bioRxiv preprint doi: https://doi.org/10.1101/009589; this version posted September 25, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

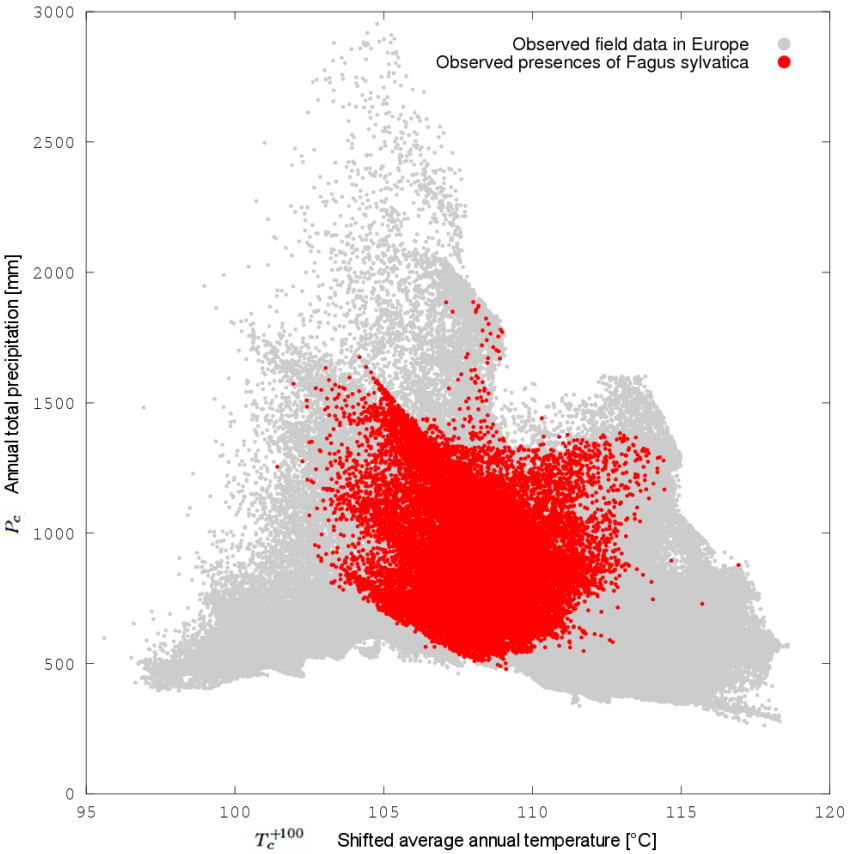

Figure 2. Ecological niche diagram of European beech (Fagus sylvatica) comparing the annual total precipitation and annual average temperature (shifted +100° C): the observed presence of the species (red dots) is highlighted over nearly 250,000 sample points (grey dots).

calculated tacking into account the temperature variation within the 1x1 km spatial grid due to difference in elevation adopting the lapse rate. Covariates derived from solar irradiation required heavy computations performed by Grass GIS 7 software. The potential solar irradiation was selected instead of the actual one due to its stability under different climate change scenarios. It was modelled starting from the average of the GMTED with a higher spatial resolution at 0.75 arc-second, gaining raster dimension of 273 million pixels. It is proved that the solar irradiation derived from a digital elevation model at 0.3 arc-second is less accurate compared with the aggregation to 0.3 arc-second of solar irradiation derived from a 0.75 arc-second. The average elevation was calculated using "r.sun" Grass module [21], combining the shadowing effects of "r.horizon" module derived every 5 degrees [22]. The atmospheric turbidity and albedo effects were not modelled, thus solar energy layer is a proxy of the potential one and essentially dependent on the aspect, latitude and orography-induced patterns of shadows. Twelve maps of solar irradiation were produced for the central days of each month. The monthly aggregation was computed by integrating data of grid maps using the trapezoidal rule. Thereafter the solar irradiation covariates were harmonized to 30-arc-second through a spatial mean aggregation.

Accepted manuscript

bioRxiv preprint doi: https://doi.org/10.1101/009589; this version posted September 25, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Accepted manuscript

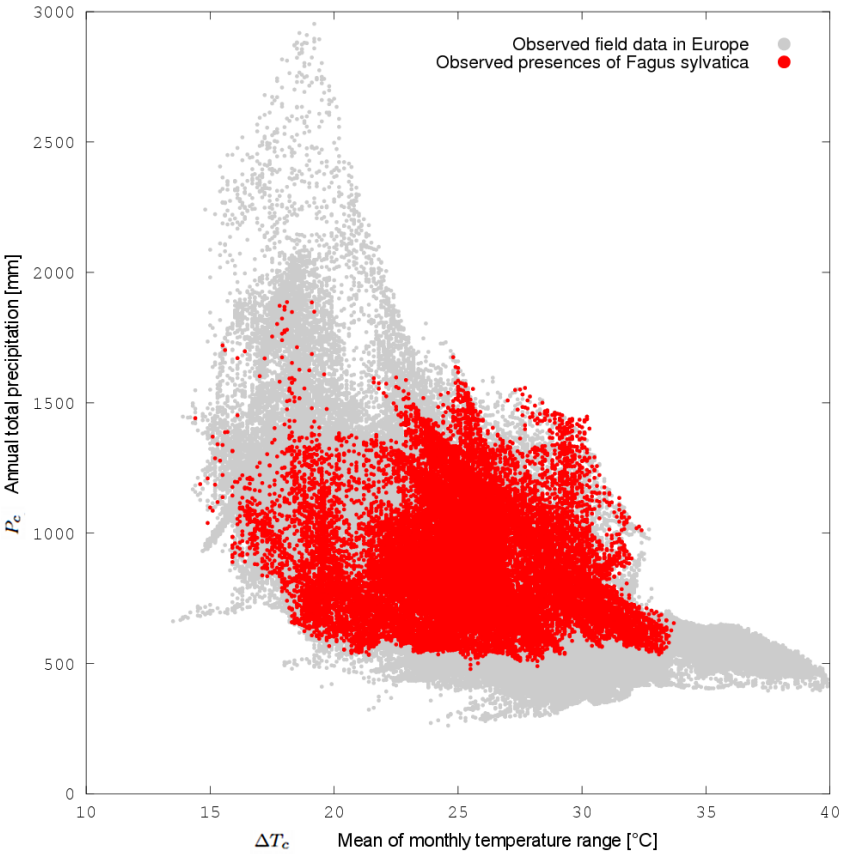

Figure 3. Ecological niche diagram of European beech (Fagus sylvatica) comparing the annual total precipitation and mean of monthly temperature range: the observed presence of the species (red dots) is highlighted over nearly 250,000 sample points (grey dots).

Results

All described covariates will be visualized and downloadable on the European Forest Data Centre (EFDAC at http://efdac.jrc.ec.europa.eu) of the Forest Information System for Europe (FISE). The modelling procedures have been performed with free software packages which allow a complete reproducibility of D-TMs. This set of twelve bioclimatic covariates, selected and elaborated for the first time to reproduce European forest ecological zones, is a combination of regression analysis predictors suitable for biological niche modelling related principally to forest resources at European scale – *e.g.* forest type and species suitability maps [23][24][25], forest pest outbreak predictions [26][27][28]. The same proposed set of predictors will be processed with the RDS regression method for the European Atlas of Forest Tree Species (FISE-EFDAC, in preparation), a publication where habitat suitability maps of main European tree species will be also presented. The codelets of described procedures were organized in modules corresponding to the D-TMs, enabling the complete set of covariate raster maps to be reproduced from different data sources and at different scales (technical details on de Rigo, Caudullo [29]). The forthcoming codelets will be also available on the EFDAC website.

In addition to geographic maps, covariates may be represented two by two as diagrams plotting values on the x and y axis within the study area. This family of charts can be used to analyse correlations between variables, and also to show ecological niche patterns when a certain species

bioRxiv preprint doi: https://doi.org/10.1101/009589; this version posted September 25, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Accepted manuscript

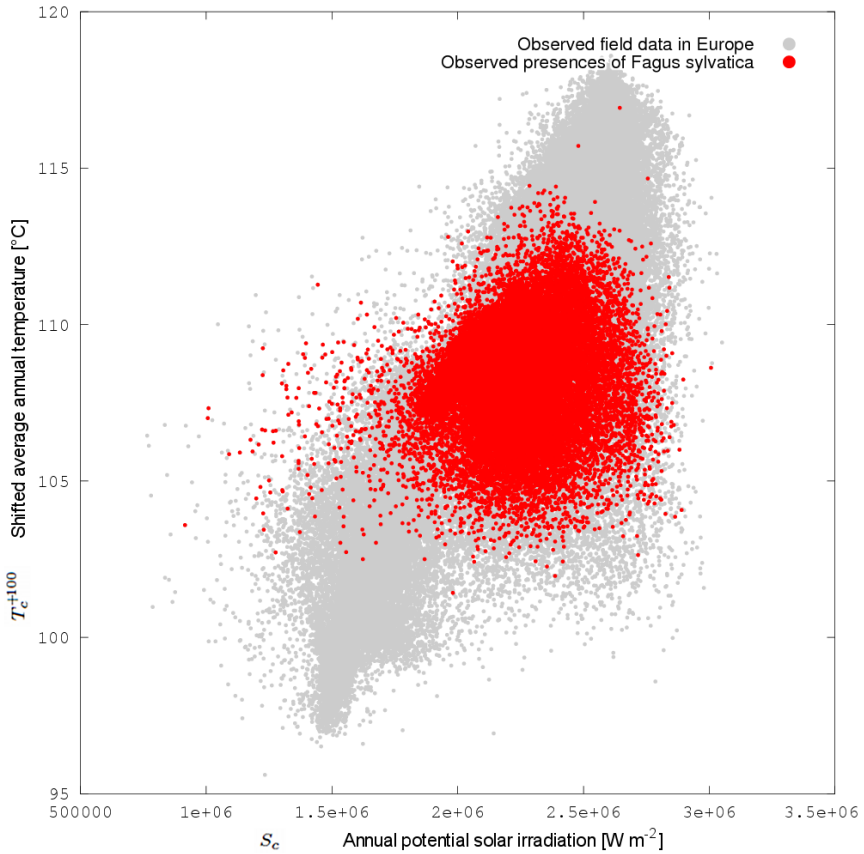

Figure 4. Ecological niche diagram of European beech (Fagus sylvatica) comparing annual average temperature (shifted +100° C) and the annual potential solar irradiation: the observed presence of the species (red dots) is highlighted over nearly 250,000 sample points (grey dots).

presence is highlighted. To exemplify this possible usage, the realised ecological niche of European beech (*Fagus sylvatica*) has been computed using the EFDAC-FISE datasets, part of them deriving from a harmonisation effort of the European National Forest Inventories. The data refer to observed presences of the tree species. In Figure 2 nearly 250,000 sample points are represented for annual Precipitation and annual average Temperature (grey dots). The presence of European beech is large (red dots) and centred in the middle of the observation cloud, confirming its common and widespread distribution in temperate forests [30]. Figure 3 shows the relation between Precipitation and the annual range in Temperature. Here it is more evident that the current presence of European beech is limited by a minimum precipitation of about 500 mm and a maximum of 2,000 mm. Annual range in Temperature gives a measure of the continentality of the climate. Beech is clearly absent above a range of 33-34°C in Temperature, confirming that this tree has a realised niche characterized by more oceanic climates [30]. Solar irradiation and annual average Temperature covariates show a positive correlation (Figure 4), since ground temperature depends principally on solar energy. The observation cloud shows a linear growing trend even if wider in the middle. On the other hand beech presence does not present crisp boundaries on the ecological niche modelled with these two covariates. Isolated samples are still present on the cloud edges, in particular where solar irradiation has lower values, without evident thresholds.

bioRxiv preprint doi: https://doi.org/10.1101/009589; this version posted September 25, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Author Bio

Giovanni Caudullo holds a degree in Forestry and Environmental Science from the "Università degli Studi di Padova". He has since worked for 6 years as a freelance environmental and GIS consultant for private companies and public authorities. Now he is working at the European Commission's Joint Research Centre in the Forest Resources and Climate Unit as senior GIS analyst, operating with spatial and temporal indicators from local to Europe-wide scale and developing procedures for automatic data processing.

References

[1] R. Store and J. Jokimäki, "A GIS-based multi-scale approach to habitat suitability modeling," *Ecological Modelling*, vol. 169, no. 1, pp. 1-15, 2003. doi: 10.1016/S0304-3800(03)00203-5

[2] L. Brotons, W. Thuiller, M. B. Araújo, and A. H. Hirzel, "Presence-absence versus presence-only modelling methods for predicting bird habitat suitability," *Ecography*, vol. 27, no. 4, pp. 437-448, 2004. doi: 10.1111/j.0906-7590.2004.03764.x

[3] A. H. Hirzel and G. Le Lay, "Habitat suitability modelling and niche theory," *Journal of Applied Ecology*, vol. 45, no. 5, pp. 1372-1381, 2008. doi: 10.1111/j.1365-2664.2008.01524.x

[4] M. B. Araújo and A. Guisan, "Five (or so) challenges for species distribution modelling," *Journal of Biogeography*, vol. 33, no. 10, pp. 1677-1688, 2006. doi: 10.1111/j.1365-2699.2006.01584.x

[5] D. de Rigo, J. I. Barredo, L. Busetto, G. Caudullo, and J. San-Miguel-Ayanz, "Continental-scale living forest biomass and carbon stock: a robust fuzzy ensemble of ipcc tier 1 maps for Europe," *IFIP Advances in Information and Communication Technology*, vol. 413, pp. 271-284, 2013. doi: 10.1007/978-3-642-41151-9_26

[6] Food and Agriculture Organization of the United Nations, *Global ecological Zones for FAO forest reporting: 2010 update*, ser. Forest Resources Assessment Working Paper. FAO, Forestry Department, 2012, vol. 179. [Online]. Available: http://www.fao.org/docrep/017/ap861e/ap861e00.pdf

[7] J.-C. Ciscar, L. Feyen, A. Soria, C. Lavalle, F. Raes, M. Perry, F. Nemry, H. Demirel, M. Rozsai, A. Dosio, M. Donatelli, A. K. Srivastava, D. Fumagalli, S. Niemeyer, S. Shrestha, P. Ciaian, M. Himics, B. Van Doorslaer, S. Barrios, N. Ibáñez, G. Forzieri, R. Rojas, A. Bianchi, P. Dowling, A. Camia, G. Libertà, J. San-Miguel-Ayanz, D. de Rigo, G. Caudullo, J. I. Barredo, D. Paci, J. Pycroft, B. Saveyn, D. Van Regemorter, T. Revesz, T. Vandyck, Z. Vrontisi, C. Baranzelli, I. Vandecasteele, F. Batista e Silva, and D. Ibarreta, "Climate Impacts in Europe - The JRC PESETA II project," ser. *EUR – Scientific and Technical Research*, J. C. Ciscar, Ed. Publications Office of the European Union, 2014, vol. 26586, 155 pp. doi: 10.2791/7409

[8] R. J. Hijmans, S. E. Cameron, J. L. Parra, P. G. Jones, and A. Jarvis, "Very high resolution interpolated climate surfaces for global land areas," *International Journal of Climatology*, vol. 25, no. 15, pp. 1965-1978, 2005. doi: 10.1002/joc.1276

[9] J. J. Danielson and D. B. Gesch, "Global multi-resolution terrain elevation data 2010 (GMTED2010)," *Open-File Report*, no. 2011–1073, 2011. [Online]. Available: http://pubs.usgs.gov/of/2011/1073/pdf/of2011-1073.pdf

Accepted manuscript

bioRxiv preprint doi: https://doi.org/10.1101/009589; this version posted September 25, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Accepted manuscript

[10] D. de Rigo, P. Corti, G. Caudullo, D. McInerney, M. Di Leo, and J. San-Miguel-Ayanz, "Toward open science at the European scale: Geospatial Semantic Array Programming for integrated environmental modelling," *Geophysical Research Abstracts*, vol. 15, pp. 13245+, 2013. doi: 10.6084/m9.figshare.155703

[11] G. van Rossum and F. L. Jr. Drake, *The Python Library Reference - Release 2.7.8*. Python Software Foundation, 2014. [Online]. Available: https://docs.python.org/2.7/download.html

[12] F. Warmerdam, "The geospatial data abstraction library," in *Open Source Approaches in Spatial Data Handling*, ser. Advances in Geographic Information Science, G. B. Hall and M. G. Leahy, Eds. Springer Berlin Heidelberg, 2008, vol. 2, pp. 87-104. doi: 10.1007/978-3-540-74831-1_5

[13] GDAL Development Team, *GDAL - Geospatial Data Abstraction Library: Version 1.10.1*, Open Source Geospatial Foundation, 2014. [Online]. Available: http://gdal.osgeo.org

[14] M. Neteler, M. H. Bowman, M. Landa, and M. Metz, "GRASS GIS: A multi-purpose open source GIS," *Environmental Modelling & Software*, vol. 31, pp. 124-130, May 2012. doi: 10.1016/j.envsoft.2011.11.014

[15] J. W. Eaton, "GNU octave and reproducible research," *Journal of Process Control*, vol. 22, no. 8, pp. 1433-1438, 2012. doi: 10.1016/j.jprocont.2012.04.006

[16] D. de Rigo, "Semantic array programming for environmental modelling: Application of the Mastrave library," in *International Environmental Modelling and Software Society (iEMSs) 2012 International Congress on Environmental Modelling and Software. Managing Resources of a Limited Planet: Pathways and Visions under Uncertainty, Sixth Biennial Meeting*, R. Seppelt, A. A. Voinov, S. Lange, and D. Bankamp, Eds., 2012, pp. 1167-1176. [Online]. Available: http://www.iemss.org/iemss2012/proceedings/D3_1_0715_deRigo.pdf

[17] E. Jones, E. Oliphant, P. Peterson, et al., *SciPy: Open Source Scientific Tools for Python*, 2001-. [Online]. Available: http://www.scipy.org

[18] O. Nordenskiold and L. Mecking, *The geography of the polar regions*. Amer. Geogr. SOC. Spec. Publ. 8, 1928, 359 pp.

[19] Z.-Y. Guo, H.-Y. Zhu, and X.-G. Liang, "Entransy—A physical quantity describing heat transfer ability," *International Journal of Heat and Mass Transfer*, vol. 50, no. 13-14, pp. 2545-2556, 2007. doi: 10.1016/j.ijheatmasstransfer.2006.11.034

[20] M. Xu, "The thermodynamic basis of entransy and entransy dissipation," *Energy*, vol. 36, no. 7, pp. 4272-4277, 2011. doi: 10.1016/j.energy.2011.04.016

[21] J. Hofierka, M. Suri, and T. Huld, "GRASS GIS manual: r.sun," in GRASS Development Team, *GRASS GIS 6.4.3svn Reference Manual*. Open Source Geospatial Foundation, 2007. [Online]. Available: http://grass.osgeo.org/grass64/manuals/r.sun.html

[22] T. Huld, T. Cebecauer, J. Hofierka, M. Suri, "GRASS GIS manual: r.horizon," in GRASS Development Team, *GRASS GIS 6.4.3svn Reference Manual*. Open Source Geospatial Foundation, 2007. [Online]. Available: http://grass.osgeo.org/grass64/manuals/r.horizon.html

[23] G. E. Rehfeldt, C. C. Ying, D. L. Spittlehouse, and D. A. Hamilton, "Genetic responses to climate in *Pinus contorta*: niche breadth, climate change, and reforestation," *Ecological Monographs*, vol. 69, no. 3, pp. 375-407, Aug. 1999. doi: 10.1890/0012-9615(1999)069%5B0375:grtcip%5D2.0.co;2

bioRxiv preprint doi: https://doi.org/10.1101/009589; this version posted September 25, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

[24] F. Rupprecht, J. Oldeland, and M. Finckh, "Modelling potential distribution of the threatened tree species *Juniperus oxycedrus*: how to evaluate the predictions of different modelling approaches?" *Journal of Vegetation Science*, vol. 22, no. 4, pp. 647-659, 2011. doi: 10.1111/j.1654-1103.2011.01269.x

[25] E. Conlisk, D. Lawson, A. D. Syphard, J. Franklin, L. Flint, A. Flint, and H. M. Regan, "The roles of dispersal, fecundity, and predation in the population persistence of an oak (*Quercus engelmannii*) under global change," *PLoS ONE*, vol. 7, no. 5, pp. e36 391+, 2012. doi: 10.1371/journal.pone.0036391

[26] B. H. Aukema, A. L. Carroll, Y. Zheng, J. Zhu, K. F. Raffa, R. Dan Moore, K. Stahl, and S. W. Taylor, "Movement of outbreak populations of mountain pine beetle: influences of spatiotemporal patterns and climate," *Ecography*, vol. 31, no. 3, pp. 348-358, 2008. doi: 10.1111/j.0906-7590.2007.05453.x

[27] S. Sobek-Swant, D. A. Kluza, K. Cuddington, D. B. Lyons, "Potential distribution of emerald ash borer: What can we learn from ecological niche models using Maxent and GARP?," *Forest Ecology and Management*, vol. 281, pp. 23-31, 2012. doi: 10.1016/j.foreco.2012.06.017

[28] D. de Rigo, G. Caudullo, L. Busetto, and J. San-Miguel-Ayanz, "Supporting EFSA assessment of the EU environmental suitability for exotic forestry pests: Final report," *EFSA Supporting Publications*, vol. 2014, no. EN-434, 61 pp., 2014. [Online]. Available: http://www.efsa.europa.eu/en/supporting/doc/434e.pdf

[29] D. de Rigo and G. Caudullo, "A reproducible extensible set of bioclimatic indices based on Geospatial Semantic Array Programming: dataset and application in Europe," (in preparation).

[30] J. R. Packham, P. A. Thomas, M. D. Atkinson, and T. Degen, "Biological Flora of the British Isles: *Fagus sylvatica*," *Journal of Ecology*, vol. 100, no. 6, pp. 1557-1608, 2012. doi: 10.1111/j.1365-2745.2012.02017.x

Accepted manuscript