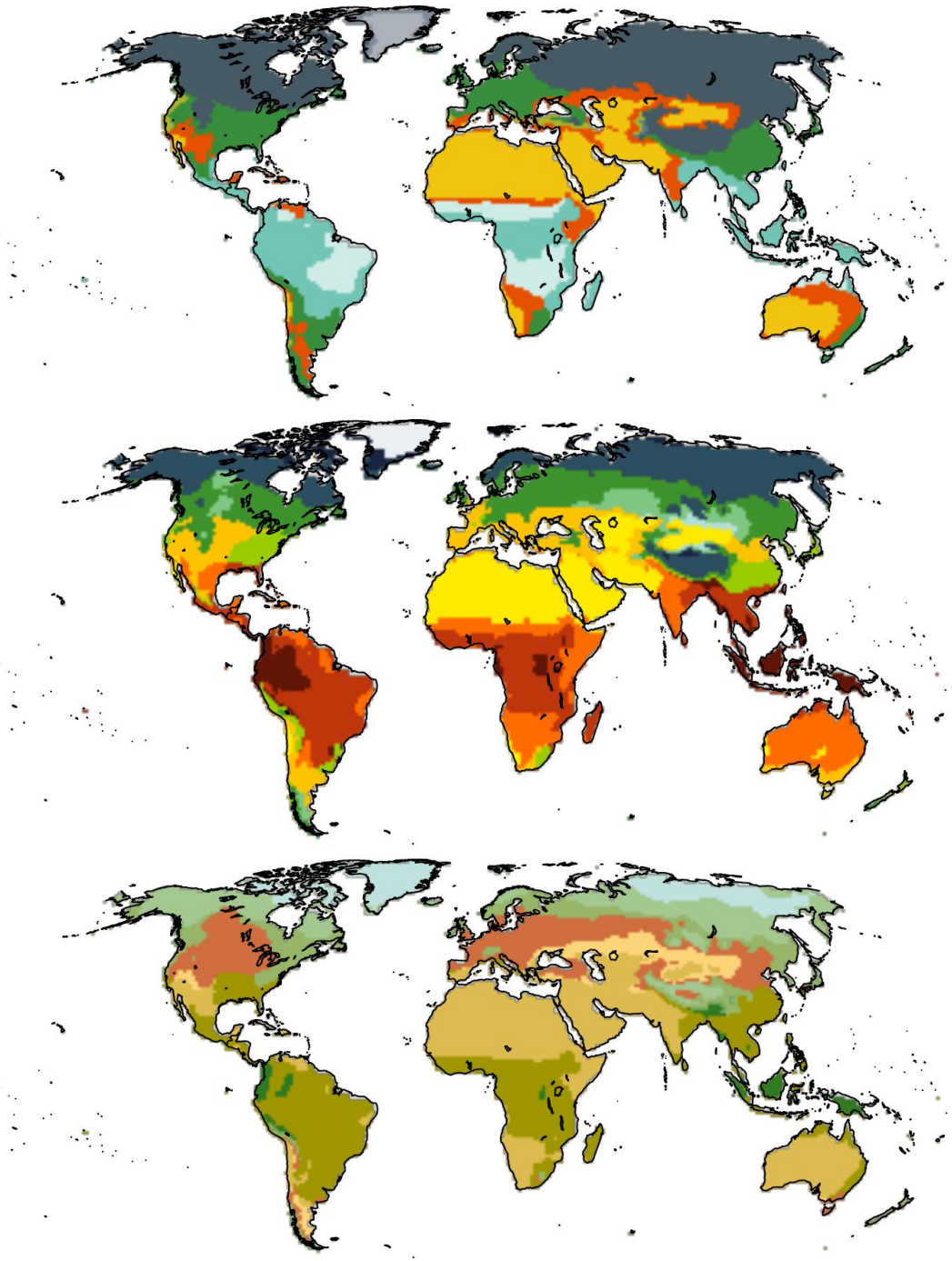
Applications of cluster analysis to Geospatial data

***Project**: A worldwide bioclimate classification scheme*

Proyectos en ingeniería de datos e inteligencia artificial



Prerequisites

- Theoretical
  - Climate classification fundamentals
  - Climate data manipulation
  - Clustering algorithms
- Technical
  - Linux / OSx (Native or virtual machine)
  - Miniconda / Anaconda
  - NetCDF / HDF5
  - CDO
  - Github

IOPscience Journals Books Publishing Support Login

ENVIRONMENTAL RESEARCH
CLIMATE

TOPICAL REVIEW • OPEN ACCESS

Climate classification systems for validating Earth System Models

Andrés Navarro* and Francisco J Tapiador

Published 2 August 2024 • © 2024 The Author(s). Published by IOP Publishing Ltd

Environmental Research: Climate, Volume 3, Number 4

Citation Andrés Navarro and Francisco J Tapiador 2024 *Environ. Res.: Climate* **3** 042001

https://doi.org/10.1088/2752-5295/ad6632

The challenge

You are given with a list of 300 cities around the world and you are asked to divide that list into categories based on the climates those cities experience, how would you do it?

Remember: climate vs weather

- Weather is what we experience over the course of hours, days, and weeks.
- Climate is the average of weather over years, decades, and longer

  "The distribution of weather states"

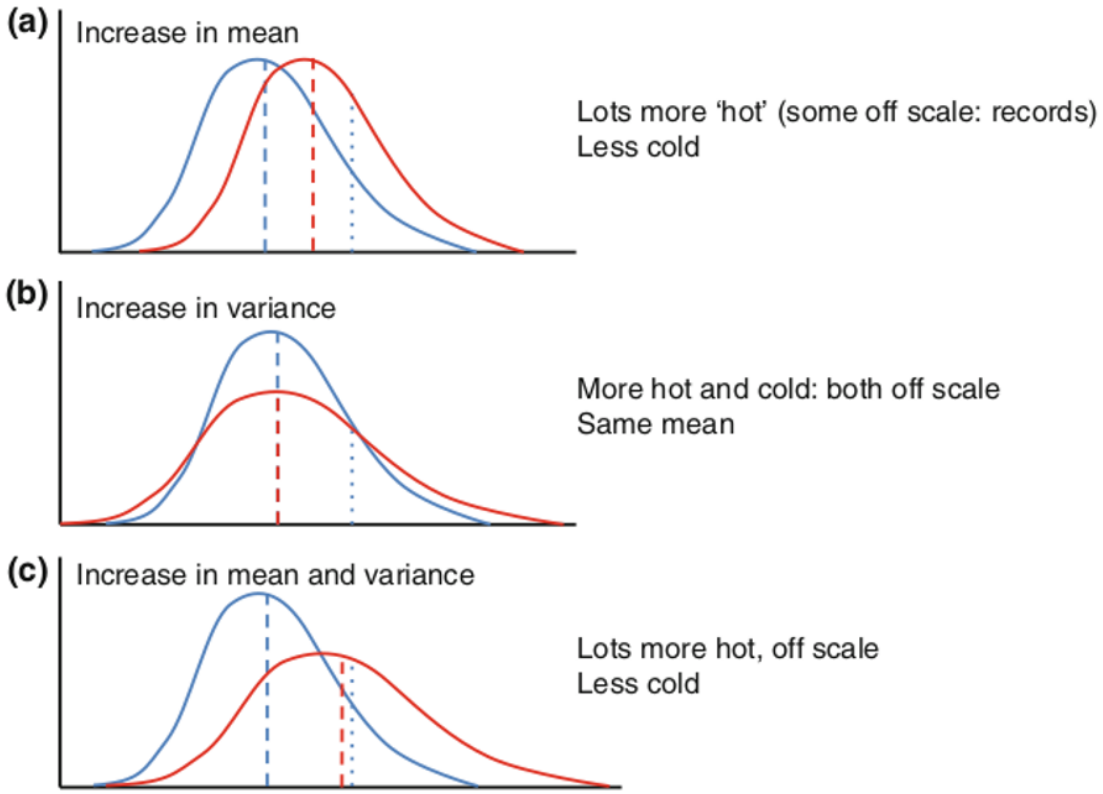


Fig. 1.3 Shifting probability distribution functions are illustrated in different ways going from the blue to red distribution. The thick lines are the distribution, the thin dashed lines are the mean of the distributions and the dotted lines are fixed points to illustrate probability. Shown is **a** increase in mean, **b** increase in variance (*width*), **c** increase in mean and variance

The challenge

What physical variables would you use?

A non-exhaustive list

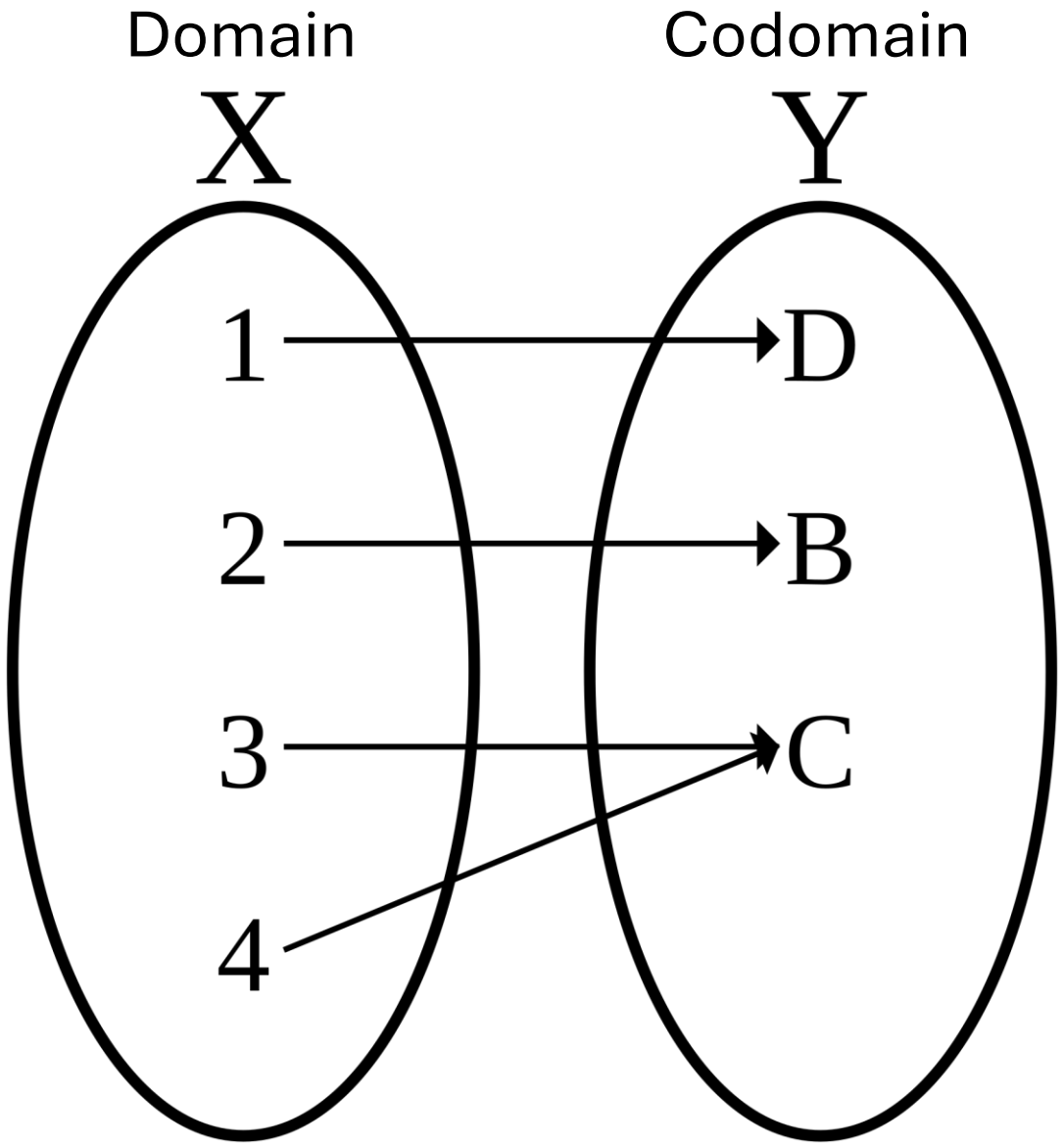
- Rainfall (annual total)
- Min. Rainfall (driest month)
- Max. Rainfall (wetter month)
- Temperature (mean)
- Max. Temperature (hottest month)
- Min. Temperature (coldest month)
- Soil Moisture
- Incoming Radiation
- Wind speed (avg)

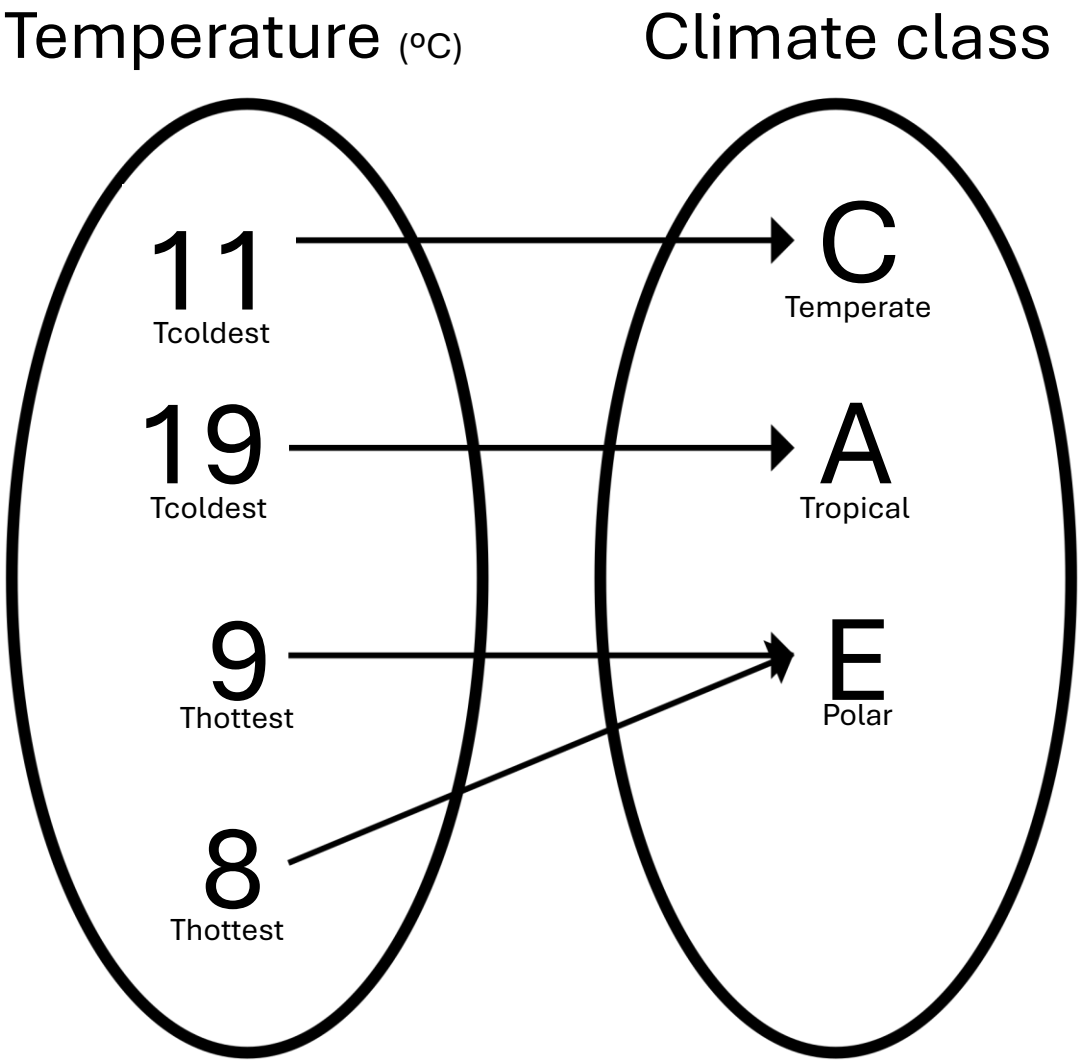
What is a CCS?

Climate classification schemes (CCS) divide the Earth’s surface into regions based on the similarity of climatic features.

What is a CCS?

CCS is a surjection from a set of physical magnitudes to a set of a few categories.


Domain
Codomain
X
Y
1
2
3
4
D
B
C



as in W. Köppen (1900)

Why are they important?

- Condense multivariate information into a single categorical index.
- Encode climatologies into a set of classes that are meaningful for environmental applications.
- Serve as a multidimensional index for climate model validation.

A good CCS must …

- Be based on well defined univocal rules.
- Minimize the differences between members of the same class.
- Maximize the differences between classes.

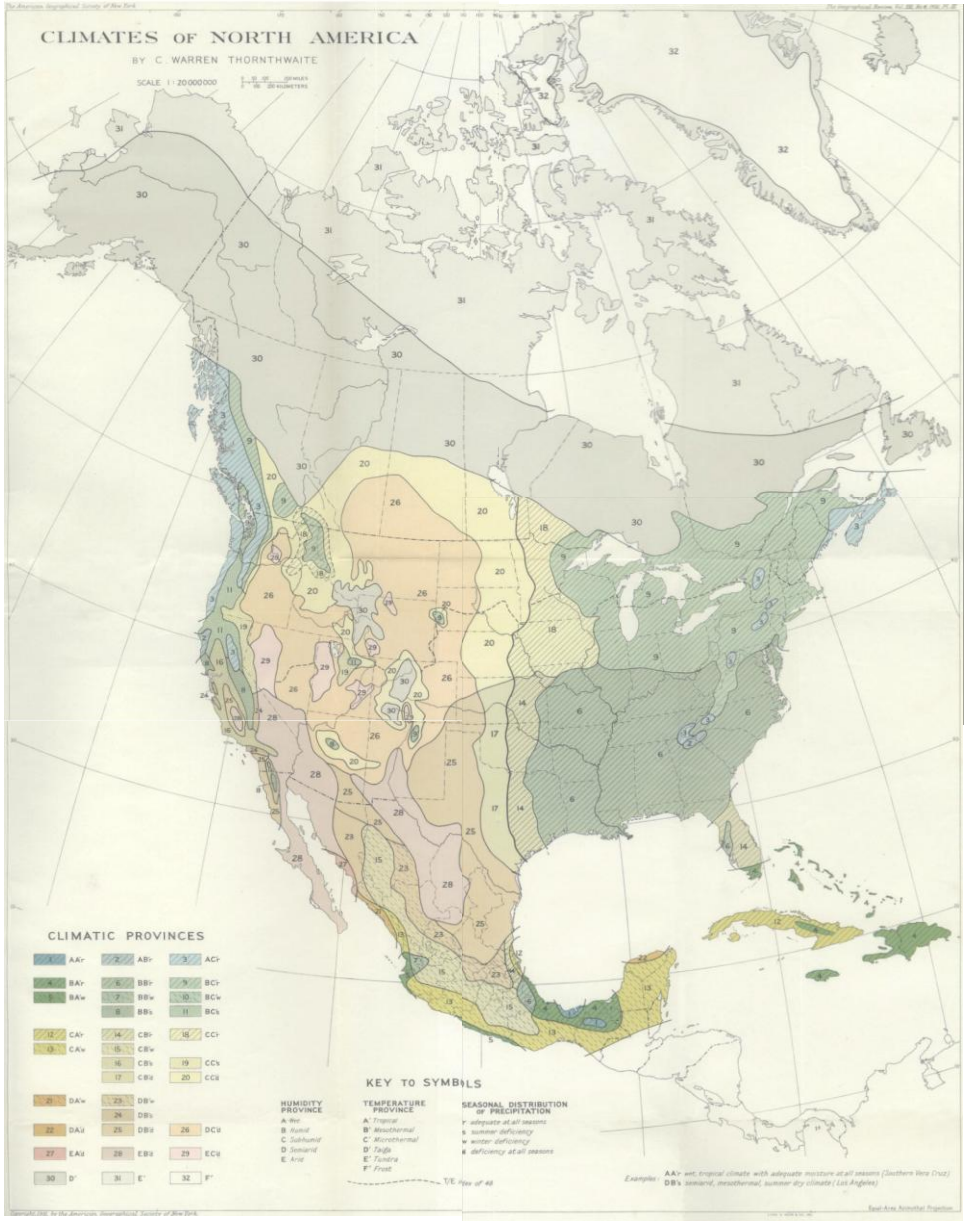
How many categories?

- The amount of variance allowed in each group depends on how many categories you select.
- The more categories you have, the more refined classification you get.
- The more categories you have, the less intuitive and clear the classification becomes.

Too complex for global applications

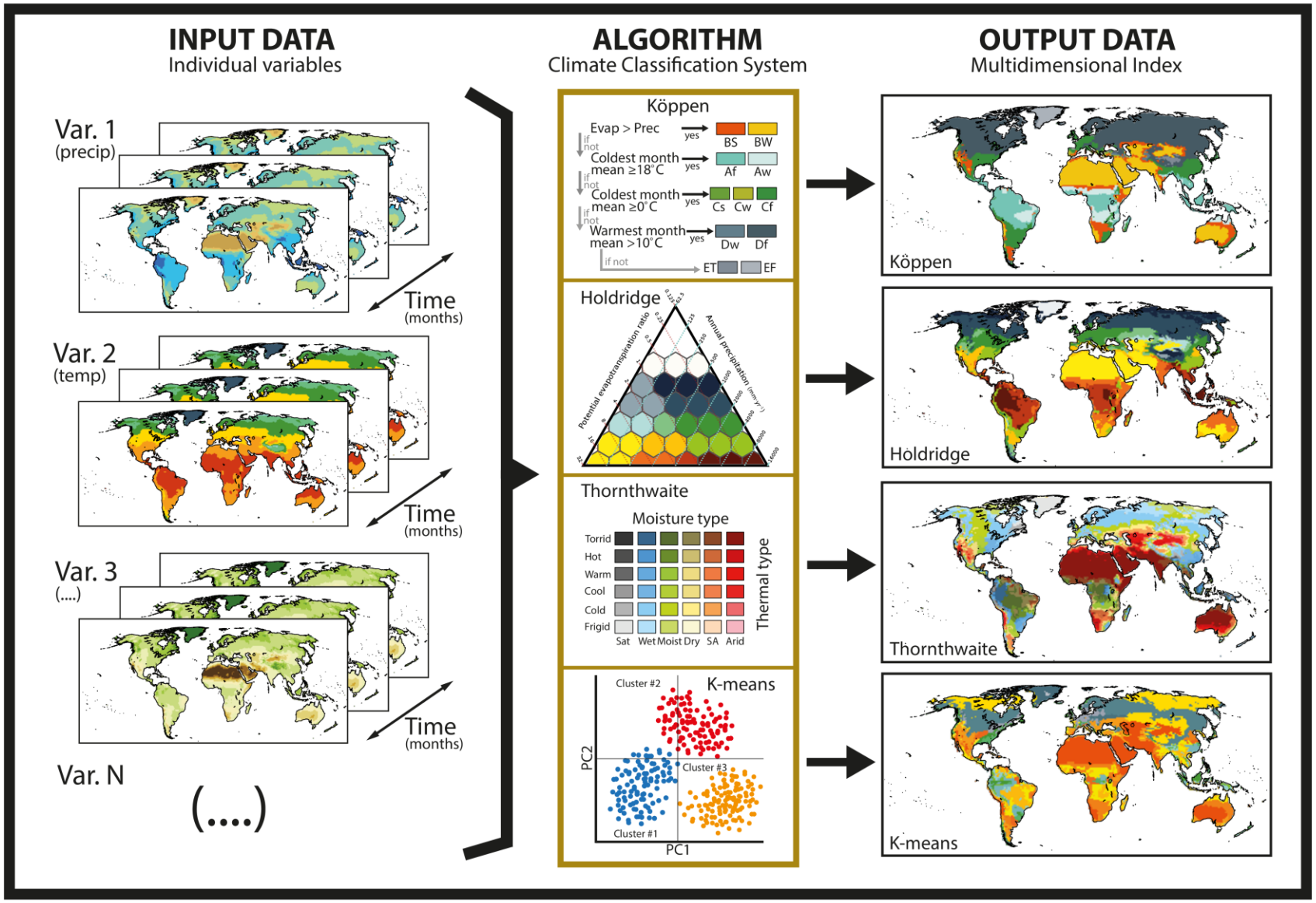
An example: the Thornthwaite's scheme

- 32 categories for 'NA only'
- More than 100 categories worldwide



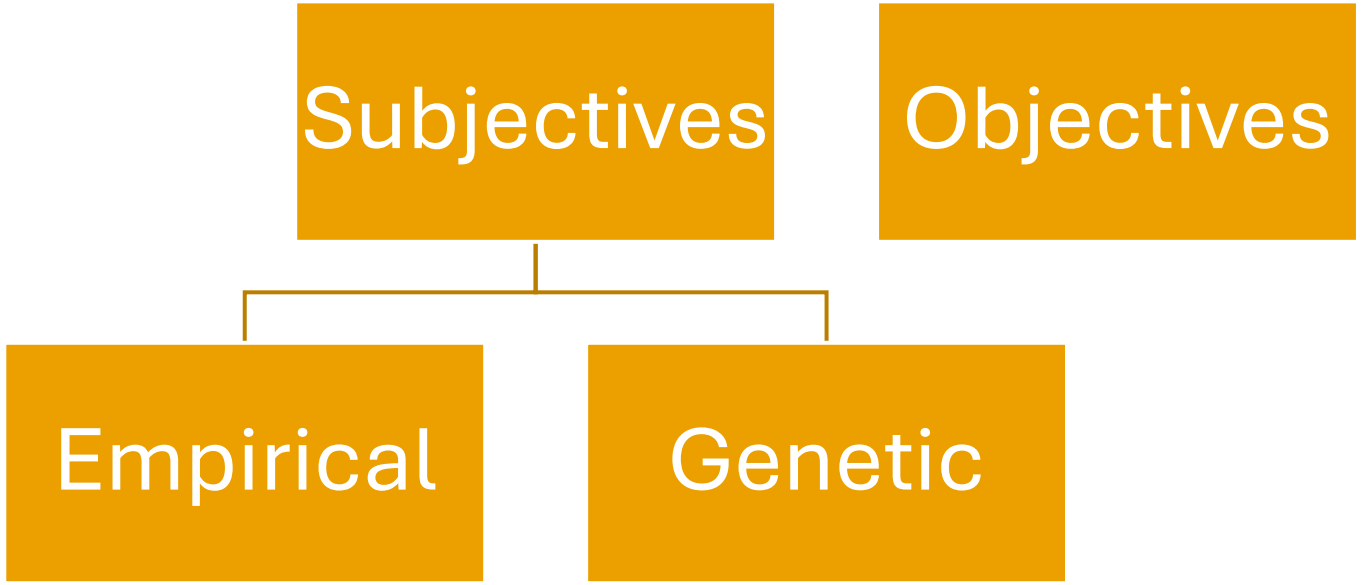
Thornthwaite, C.W. (1948). An approach toward a rational classification of climate. *Geographical Review*, **38** (1), 55-94. http://www.jstor.org/stable/210739

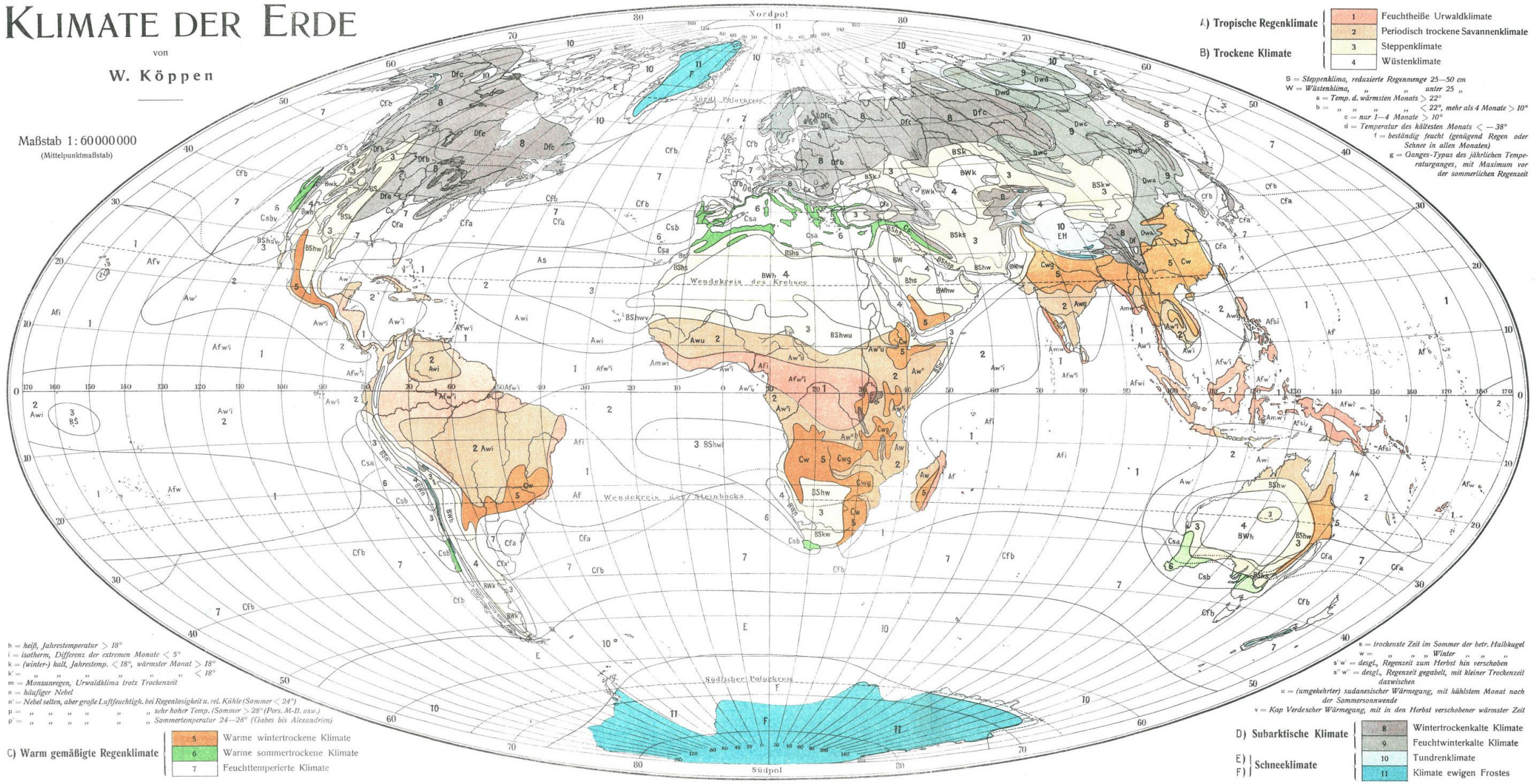
How does CCS work?



Navarro, A., et al. (2025) Uncertainty maps for model-based global climate classification systems. *Sci Data* **12**, 35. https://doi.org/10.1038/s41597-025-04387-0

Types of CCS





Types of CCS: Köppen's climate

Empirical

Klassifikation der Klimate nach Temperatur, Niederschlag und Jahreslauf.

Von Prof. Dr. *Wladimir Köppen*, Hamburg.

(Mit Karte, s. Tafel 10 und 8 Figuren, s. Tafel 11.)

I. Fortschritte seit 1901.

Es ist nicht schwer, das Klima eines Ortes in seine Bestandteile zu zerlegen und die Unterschiede im Klima verschiedener Orte nachzuweisen. Aber in der so geschaffenen Mannigfaltigkeit sich zurechtzufinden und die großen Züge darin zu erkennen, ist, namentlich für den Nichtfachmann, recht schwierig.

Wie der Botaniker sich nicht damit begnügen kann, möglichst viele Wurzeln, Blätter und Blüten zu beschreiben, sondern die Pflanze als Ganzes betrachten und die vielerlei Pflanzen in ein übersichtliches System bringen muß, so soll auch der Physiogeograph das Klima im Zusammenhang erfassen und das Verbindende und das Trennende zwischen den Klimaten durch eine großzügige Klassifikation leicht erkennbar machen. Sind auch seine Elemente nicht so greifbar wie Wurzel und Blatt, so sind dafür die physikalischen Vorgänge in ihnen einfacher und meßbarer.

Seit dem Erscheinen meines »Versuchs einer Klassifikation der Klimate, vorzugsweise nach ihren Beziehungen zur Pflanzenwelt« [1]) im Jahre 1901 sind mehrere Schriften erschienen, die das Problem mehr vom geographischen oder mehr vom biologischen Standpunkt behandeln [2]). Anwendungen meiner Klassifikation auf einzelne Gebiete sind mir nur bekannt geworden von H. Maurer auf Ostafrika und von Figurowski auf den Kaukasus — beides interessante und schwierige Ausnahmegebiete.

Durch die genannten Abhandlungen ist das Problem in lehrreicher Weise erweitert und näher beleuchtet. Ein genauer Vergleich der Einteilungen der in der Fußnote genannten Verfasser mit der meinigen ist allerdings deshalb nicht möglich, weil sie nur wenige klimatologische Zahlen angeben und, mit Ausnahme von Martonne, auch keine kartographische Darstellung der unterschiedenen Gebiete oder ausreichende Beschreibung ihrer Grenzen liefern.

Am meisten stimmt mit meiner Darstellung diejenige von Hettner überein, die in gewissem Sinne schon vor Erscheinen meiner Arbeit in einer Erdkarte der Pflanzendecke in Spamers Handatlas von ihm niedergelegt war, die mir unbekannt geblieben war (vgl. meinen »Versuch«, vorletzte Seite). Nur verfahren wir verschieden in der Darstellung, Hettner mehr deduktiv, ich mehr induktiv. Beides hat seine Vorteile. Vorsicht veranlaßt mich indessen, auch jetzt die induktive Feststellung der Tatsachen in den Vordergrund zu stellen; ist doch der innere Zusammenhang der atmosphärischen Zirkulation in manchen Hauptzügen noch keineswegs sichergestellt.

Die kurze Beschreibung, die Martonne von seinen 30 Klimaten gibt, ist gewiß in allem Wesentlichen richtig. Aber da er für jedes Klima mehrere Züge angibt, ohne zu sagen, welchen er dessen Begrenzung zugrunde legt, ist ein genauer Vergleich seiner Abteilungen mit meinen durch feste Merkmale umschriebenen schwierig.

Eine sehr wertvolle Bereicherung des Bildes der Klimate hat Pencks kurze Abhandlung gebracht, in der er die Frage ausschließlich vom Standpunkt des Verhaltens der wässerigen Niederschläge zum Erdboden behandelt. Auf diesem Gebiete waren ihm Woeikof und Hilgard vorangegangen [1]). Ich werde im folgenden suchen, auch diese Beziehungen zu berücksichtigen, soweit sie sich mit den von mir unterschiedenen Klimagruppen sicher verknüpfen lassen.

Die Klassifikation in Drudes Werk, die in dieser Zeitschrift kürzlich von Eckardt besprochen worden ist, bietet, bis auf die verstärkte Betonung des Periodischen, wenig für die Klimatologie Verwendbares.

Sie betrifft Pflanzen, nicht Klimate [2]). Neben Wärme und Feuchtigkeit führt sie auch das Licht ein durch Anfügung einer großen Pflanzengruppe, für deren Leben die Lichtperiode entscheidend sei. Für die Klimatologie würde die Einbeziehung des Lichts eine vermeidbare Komplikation bedeuten, da die Strahlung auf die meteorologischen Vorgänge ganz vorwiegend durch die Erwärmung wirkt. Bei der Pflanze ist dies natürlich ganz anders. Indessen ist es auch da vielleicht nicht so »unverzeihlich« (Drude, S. 147), wenn de Candolle u. a. für die geographischen Hauptzüge ohne das Licht auszukommen suchten; denn entscheidend ist doch auch hier das Moment, das im Minimum für den physiologischen Bedarf vorhanden ist. Und das ist, unter freiem Himmel, wohl nur selten das Licht, sondern in niederen Breiten das Wasser, in höheren die Wärme.

II. Trennende Merkmale für die Klassifikation.

Da wir es in der Klimatologie fast durchweg nur mit Quantitäten zu tun haben, so muß sie, um bestimmte Grenzen zu erhalten, zu Schwellen, und wenn diese nicht

[1]) Hettners Geographische Zeitschrift VI. Auch als Sonderabdruck erschienen (Leipzig 1901, B. G. Teubner). — [2]) E. de Martonne: Traité de géographie physique, Paris 1909, S. 206—25. — A. Penck: Versuch einer Klimaklassifikation auf physiogeographischer Grundlage. (Sitzber. Akad. Wiss. Berlin I, 1910, S. 236—46.) — A. Hettner: Die Klimate der Erde. (G. Z. 1911, S. 425, 545, 618, 675.) — O. Drude: Ökologie der Pflanzen, Braunschweig 1913, S. 149—62.

[1]) A. Woeikof: Flüsse und Seen als Produkte des Klimas. (Z. Ges. Erdk. Berlin 1885.) — E. W. Hilgard: Über den Einfluß des Klimas auf die Bildung und Zusammensetzung des Bodens. Heidelberg 1893. — Ders.: Soils ... in the humid and arid regions. Neuyork 1910 und London 1906. — [2]) Die 18 »Klimagruppen« des Buches sind nicht Gruppen von Klimaten, sondern klimatisch bedingte Pflanzengruppen.

Vegetation distribution based on weather patterns

- Based on PR and TS
- **5** first-order cats. (A, B, C, D, E)
- **3** second-order cats. (f, s, w)

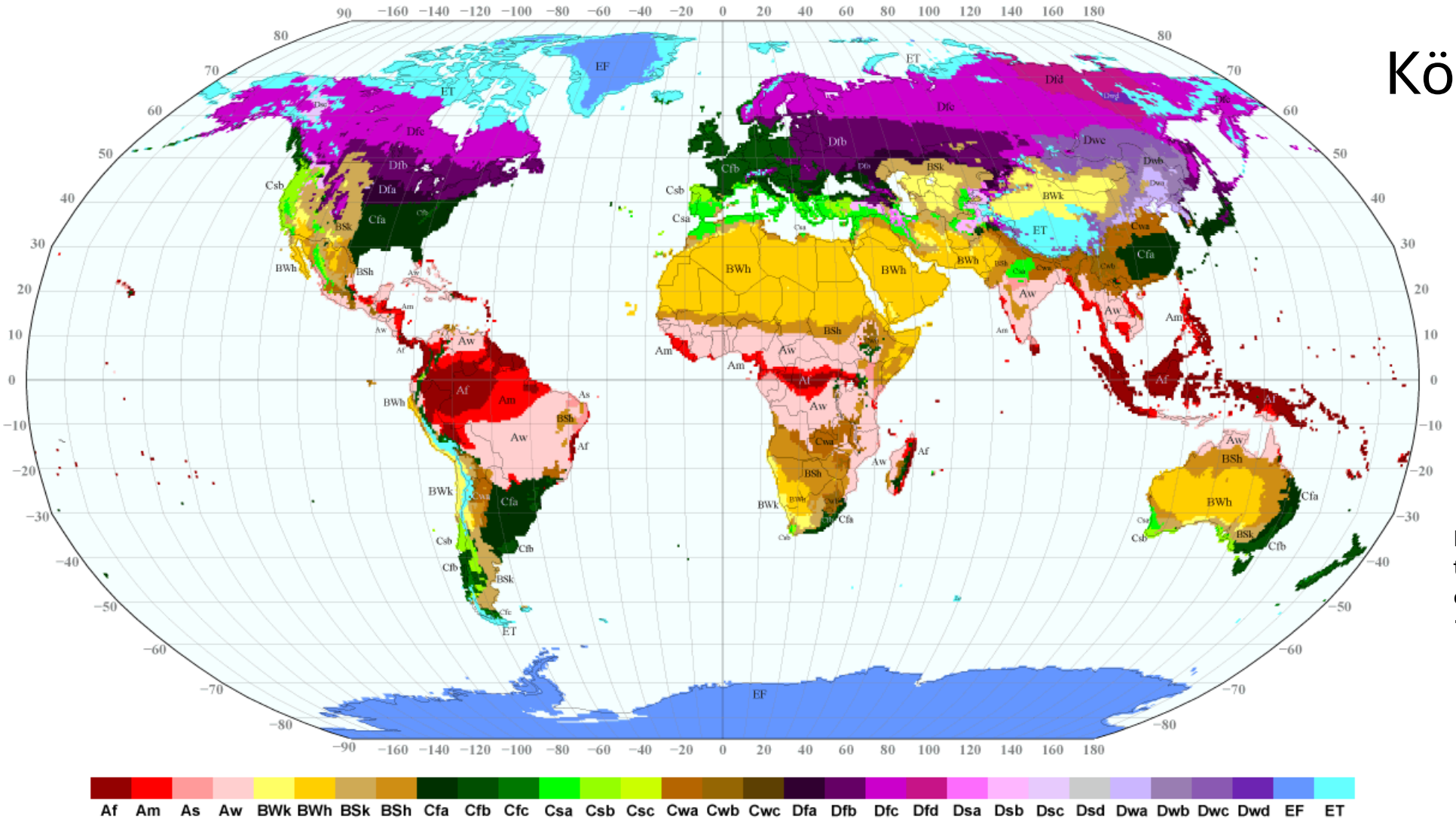
Köppen, W., 1918: Klassifikation der Klimate nach Temperatur, Niederschlag und Jahresablauf Petermanns Geogr. Mitt., 64, 193-203, 243-248.

Class	Type	Criteria
A (tropical)		T_min ≥ 18°C
	Af (tropical rainforest)	P_min ≥ 60 mm
	Aw (tropical savanna)	P_min < 60 mm
B (dry)		P_annual ≤ P_threshold
	BS (steppe)	P_annual ≥ P_threshold/2
	BW (desert)	P_annual < P_threshold/2
C (mesothermal)		T_min ≥ −3°C and T_min < 18°C
	Cs (warm climate dry summer)	P_wmax ≥ 3 × P_smin
	Cw (warm climate dry winter)	P_smax ≥ 10 × P_wmin
	Cf (humid temperate)	P_smax < 10 × P_wmin and P_wmax < 3 × P_smin
D (microthermal)		T_min < −3°C and T_max > 10°C
	Dw (cold climate dry winter)	P_smax ≥ 10 × P_wmin
	Df (cold climate moist winter)	P_smax < 10 × P_wmin
E (polar)		T_max < 10°C
	ET (tundra)	T_max ≥ 0°C and T_max < 10°C
	EF (permafrost)	T_max < 0°C

Note: More restrictive classes are calculated first in this paper.
Abbreviations: T_min, temperature coldest month; T_max, temperature hottest month; P_min, precipitation driest month; P_annual, annual precipitation; P_smin, minimum summer precipitation; P_smax, maximum summer precipitation; P_wmin, minimum winter precipitation; P_wmax, maximum winter precipitation; P_threshold, varies according to the following rules (if 70% of P_annual occurs in winter then P_threshold = 2 × T_avg; if 70% of P_annual occurs in summer then P_threshold = 2 × T_avg + 28; otherwise P_threshold = 2 × T_avg + 14).

Navarro, A., *et al.* (2022). Towards better characterization of global warming impacts in the environment through climate classifications with improved global models. *International Journal of Climatology*, **42**(10), 5197–5217. https://doi.org/10.1002/joc.7527

A popular scheme with many variants



Köppen-Geiger

Kottek, *et al.* (2006) World Map of the Köppen-Geiger climate classification updated. Meteorol. Z., **15**, 259-263.

A popular scheme with many variants

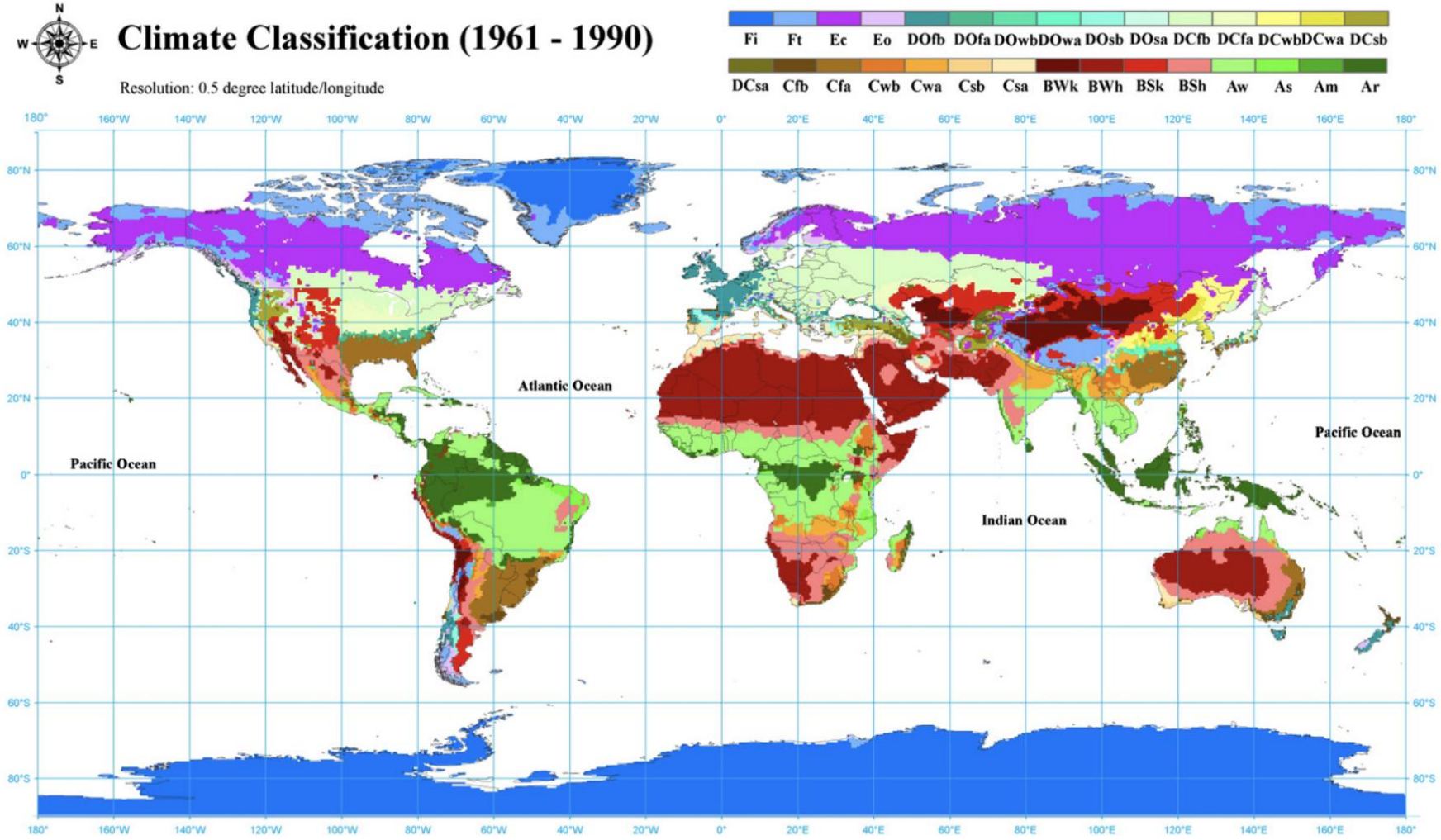
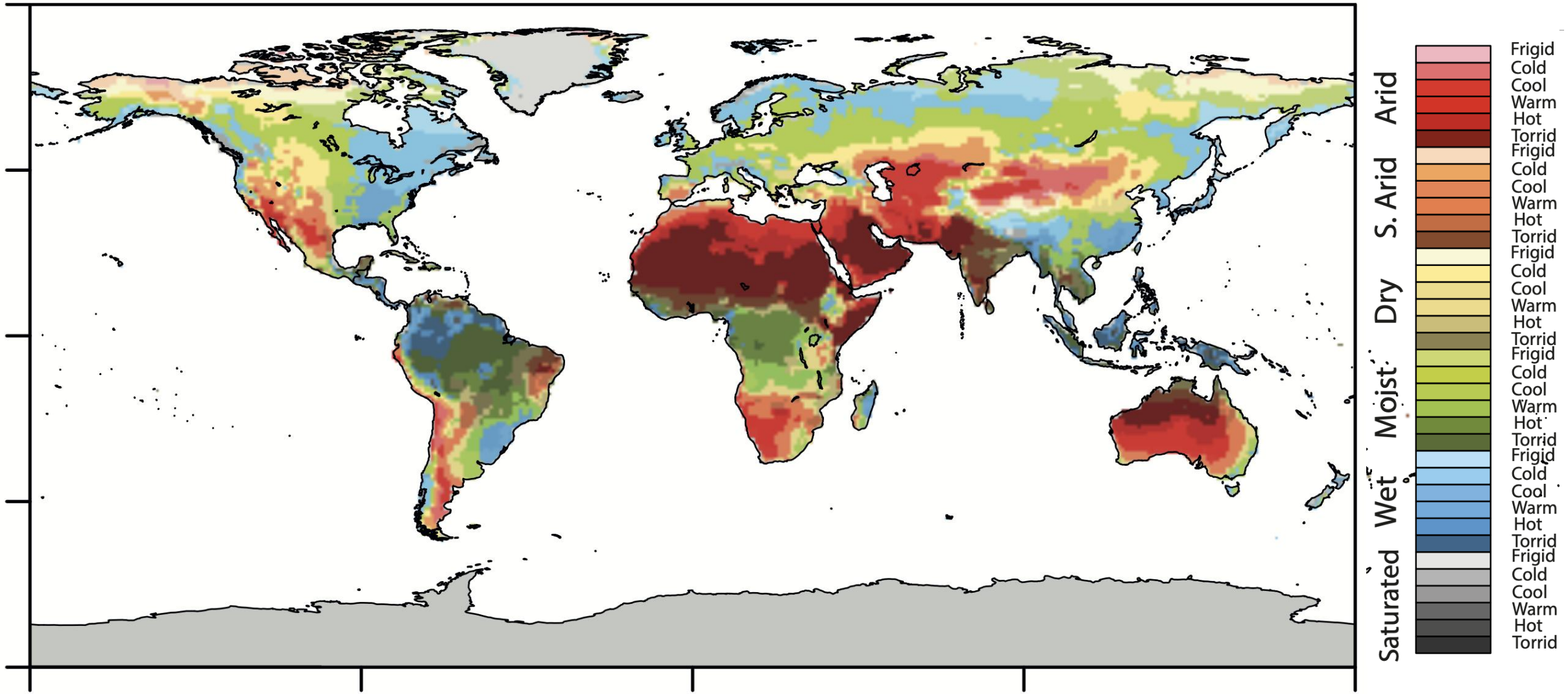


Fig. 3. Spatial distribution of the K–T climate types deduced from CPC climatology (1961–90).

Köppen-Trewartha

Feng, S. *et al.* (2014) Projected climate regime shift under future global warming from multi-model, multi-scenario CMIP5 simulations. Glob. Planet. Chang.,**112**, 41-52.

Empirical



Types of CCS: Feddema's scheme

A REVISED THORNTHWAITE-TYPE GLOBAL CLIMATE CLASSIFICATION

Johannes J. Feddema
Department of Geography
University of Kansas
Lawrence, Kansas 66045

Abstract: Although the Köppen climate classification is the most common climate classification in use today, the 1948 Thornthwaite classification is frequently cited as an improved climate classification system for its rational approach. However, the Thornthwaite classification is infrequently used because it tends to be too complex for use in everyday settings and world maps of the classification were never produced. This paper will present global maps of all four components of the 1948 Thornthwaite climate classification—a long-time wish of John "Russ" Mather, to whom this paper is dedicated. In addition, a revised Thornthwaite-type climate classification is presented with the intent of providing a more rational climate classification for everyday use in a classroom setting. This classification uses an amended version of the Thornthwaite moisture index, not only to delineate climatic moisture gradients but also to define a single seasonality index responsive to mean seasonal variation in both thermal and moisture conditions. Replacing the two cumbersome seasonality indicators in the original Thornthwaite classification with one variable greatly improves the utility of the classification. Results from this classification are compared to the Köppen and original Thornthwaite climate classification schemes. [Key words: climate classification, moisture index, W. Köppen, C. W. Thornthwaite, J. R. Mather.]

INTRODUCTION

In the last few decades, John "Russ" Mather frequently discussed the idea of producing a world map of the Thornthwaite climate classification. Thornthwaite (1948) only applied his final classification scheme to the United States, in part because he did not have access to sufficient climate data to do a global study. To remedy this problem, C. W. Thornthwaite Associates did much work to develop and improve water budget applications and to collect climatic data for the entire world (Field, 2005 [this issue]). However, those data have never been applied to the 1948 Thornthwaite classification on a global scale. Russ's wish was to publish the full classification; however, he was also aware that the full classification was too complex for everyday classroom use. Therefore, we also developed a modified, less complex classification scheme was also developed. Like Thornthwaite himself, the objective of this work is to present a more systematic approach to classifying global climates that is easy to interpret and can be easily conveyed to students in a classroom setting.

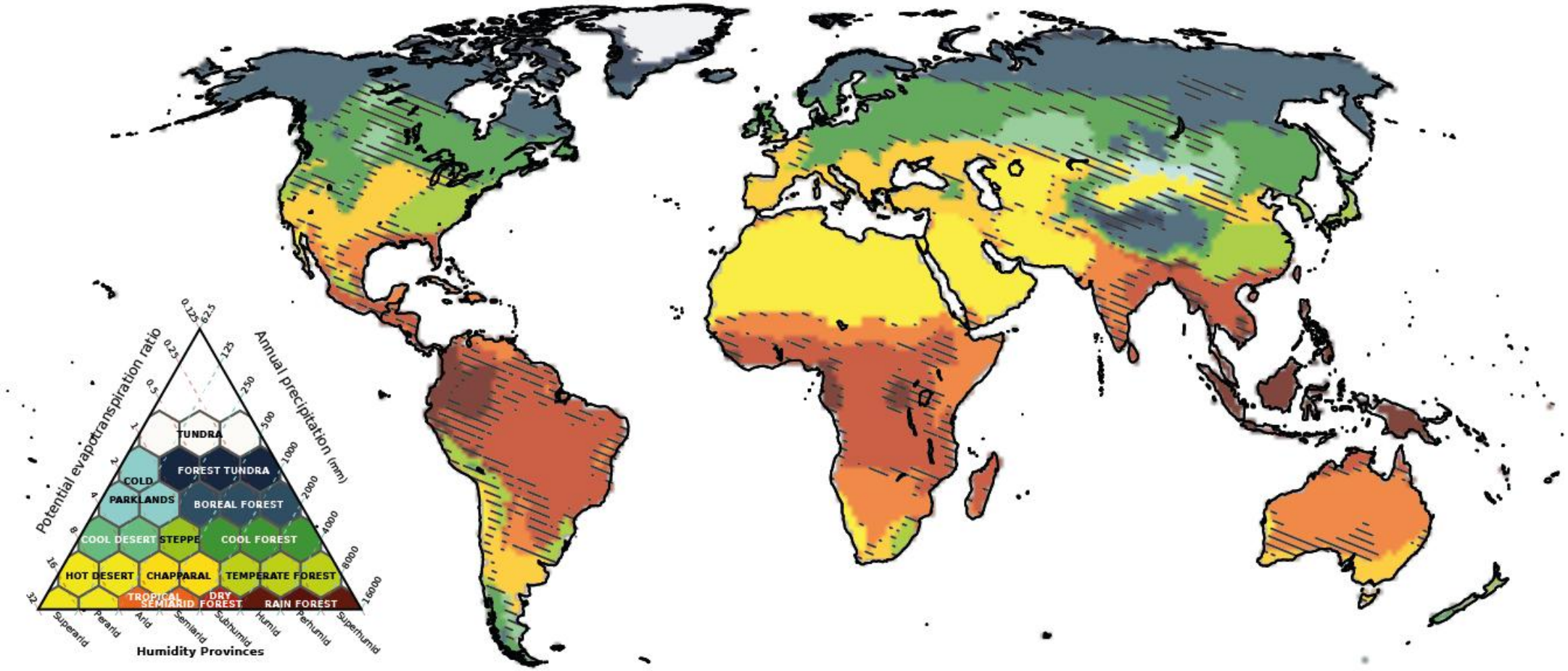
Although climate classification schemes have existed since the Ancient Greeks, modern climate classifications really began, and continue to this day, with the work of Wladimir Köppen (Köppen, 1900). The Köppen classification was a great achievement; first for identifying distinct climates, and second because it linked numerical climate statistics to vegetation distributions. The Köppen classification has been modified many times since its inception because of problems with the way

Physical Geography, 2005, **26**, 6, pp. 442–466.

- Based on amount of energy present and amount of available moisture
- Evapotranspiration= evaporation + plant transpiration
- Based on PR, TS and PET
- **6** Moisture types
- **6** Thermal types

Feddema, J., (2005) A revised Thornthwaite-type global climate classification. *Physical Geography* **26** , 442-466.

Empirical



Types of CCS: Holdridge's Life Zones

Determination of World Plant Formations From Simple Climatic Data

L. R. HOLDRIDGE

Department of Botany, University of Michigan, Ann Arbor

While attempting to understand relationships between the mountain vegetation of an area in Haiti and other vegetation units of the island and surrounding regions, the literature was searched unsuccessfully for a comprehensive system which presented formations or vegetation units on a relatively equal or comparable basis. To fulfill such a need, a chart (Fig. 1) was constructed which differentiates the vegetation of dry land areas of the world into 100 closely equivalent formations separated by temperature, precipitation, and evaporation lines of equal value.

Mean temperature values used approximate those of the growing season and are determined for specific sites by adding the average mean monthly temperatures greater than 0°C. and dividing by 12. Such values for lower elevations establish several regions north and south of the heat equator, as indicated on the left, and at higher elevations the belts listed at the right. Because of the difference in vegetation caused by the presence or absence of frost in the warm temperate region and the lower montane belt, a frost line is recognized which separates the low subtropical region and the subtropical belt formations as vegetation units usually present between the 24° isotherm and the limits of killing frosts.

The precipitation value used for a locality is the average mean annual precipitation in millimeters. The temperature and precipitation values for any site plotted logarithmically on the chart determine a point which falls within a hexagon representing a formation. When the point falls within one of the border triangles of a hexagon, the vegetation of that area will show a transitional character.

For altitudinal adjustment, approximate elevation in meters above sea level must be known to be certain of the region to which the formation belongs. All altitudinal belts will be found only in the tropics. In other regions, only the belts above the basal formations of the region on the chart will be encountered. Elevations of formation boundaries vary considerably, but the ranges of the belts approximate the following: nival—indefinite, alpine—500 m., subalpine—500 m., montane—1,000 m., and lower montane alone or with the subtropical if present—2,000 m. The tropical basal region varies from 0 to 1,000 m.; the warm temperate alone or with the low subtropical, 0–2,000 m.; and the basal formations of the other regions, from 0 m. to the maximum for the corresponding belt.

Thus, a station below 500 m. in elevation with means of

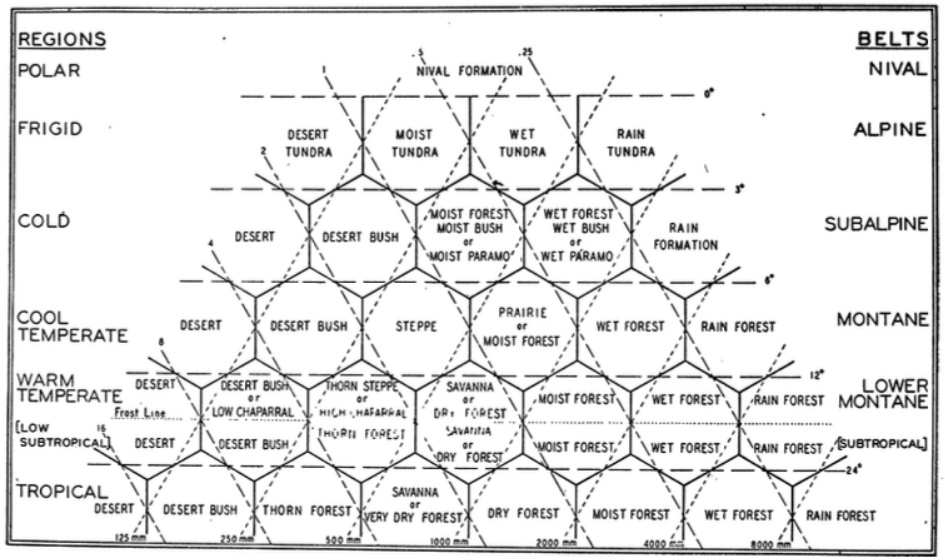


FIG. 1

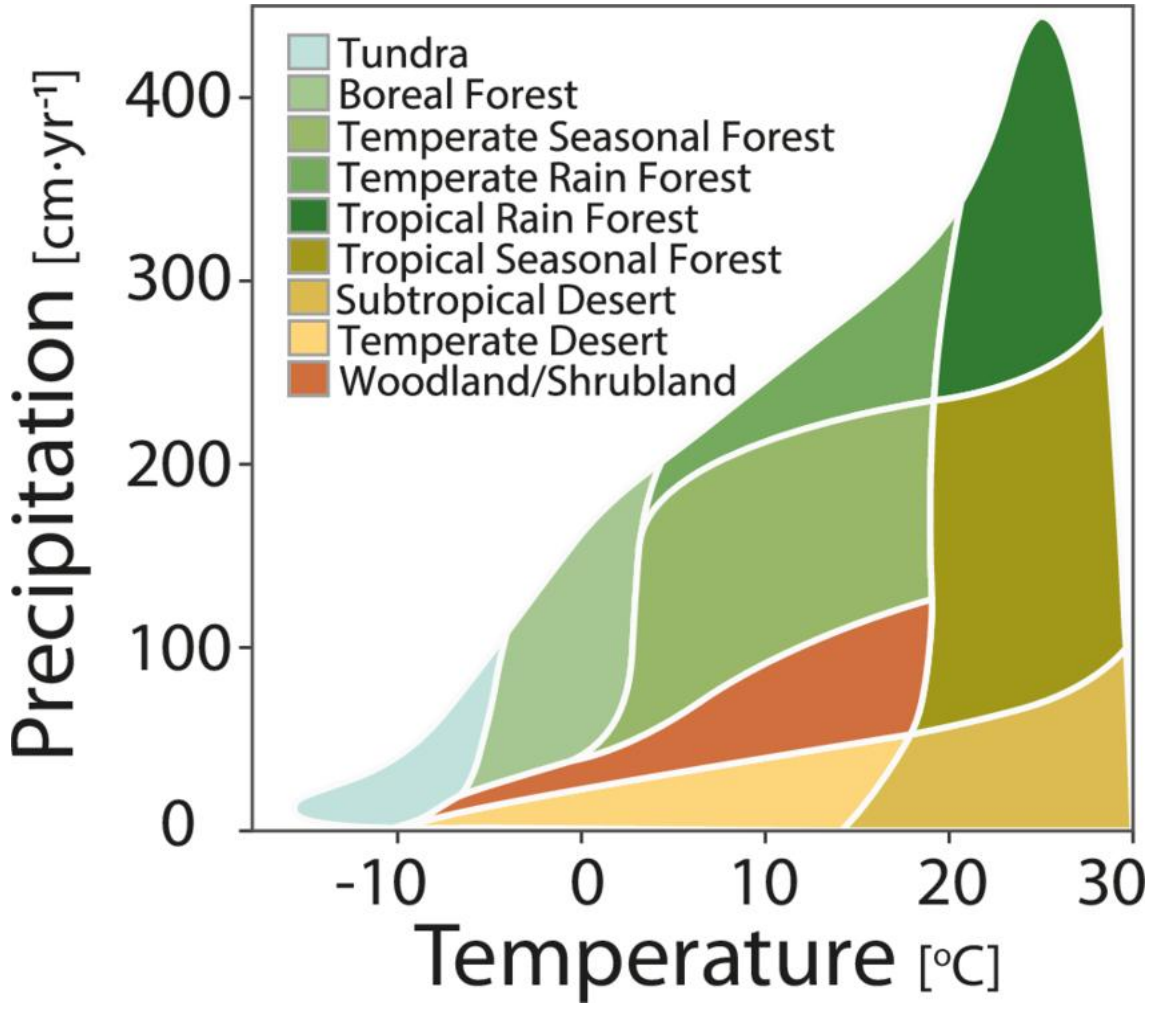
Thus, a station below 500 m. in elevation with means of

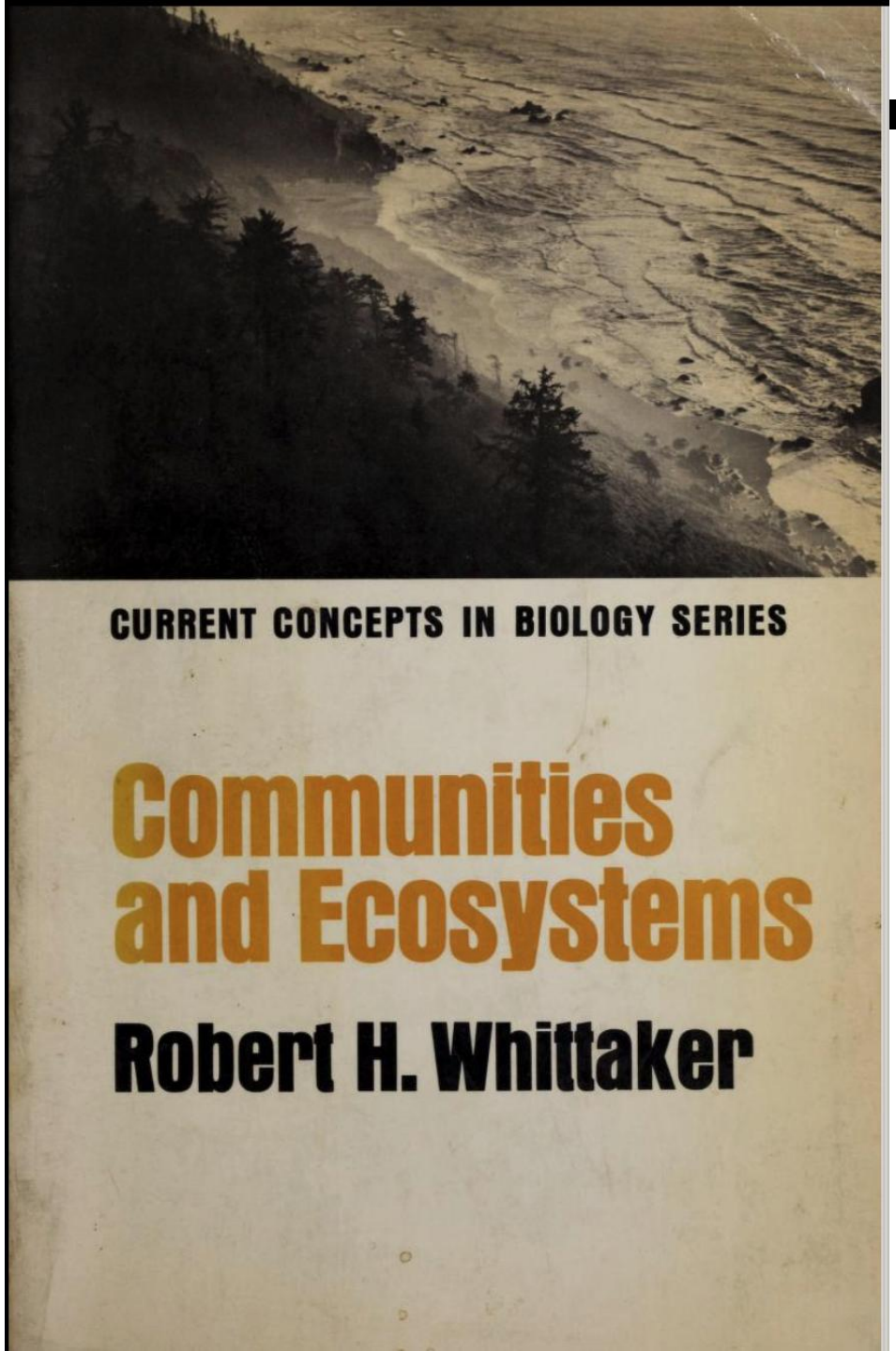
- Based on water and energy requirements
- Ecological units = Life Zones
- Based on PR, BIOTS and PET ratio
- **6** Biotemperature levels
- **8** Rainfall levels

Holdridge, L.R., (1947) Determination of World Plant Formations From Simple Climatic Data. *Science* **105**, 367-368. https://doi.org/10.1126/science.105.2727.367

Types of CCS

Whittaker's biomes



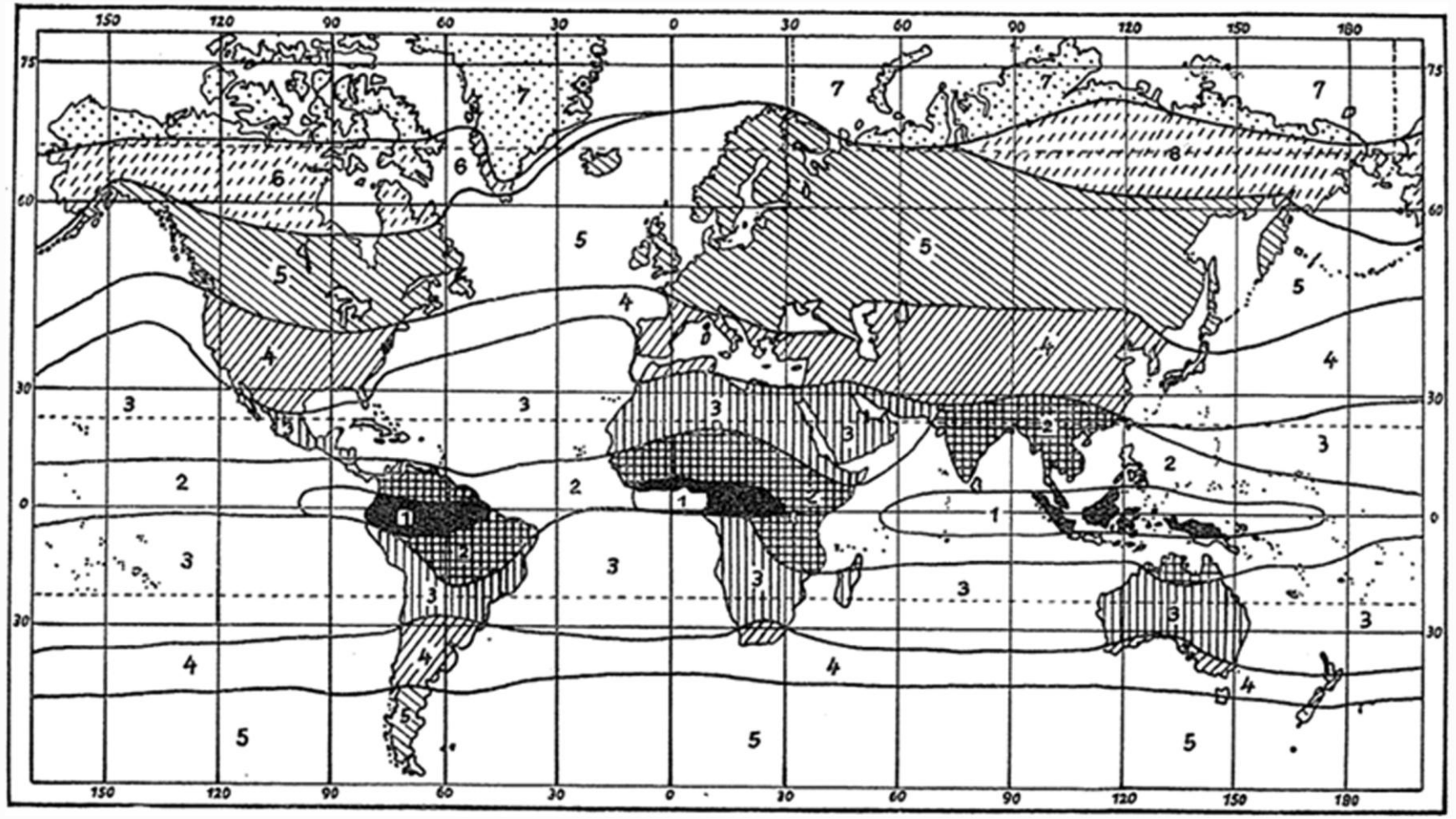


- Biome = ecorregion with specific climate, vegetation and animal life
- Based on PR, TS
- **9** major types

Whittaker RH (1975) *Communities and Ecosystems. 2nd edition.* Macmillan, New York.

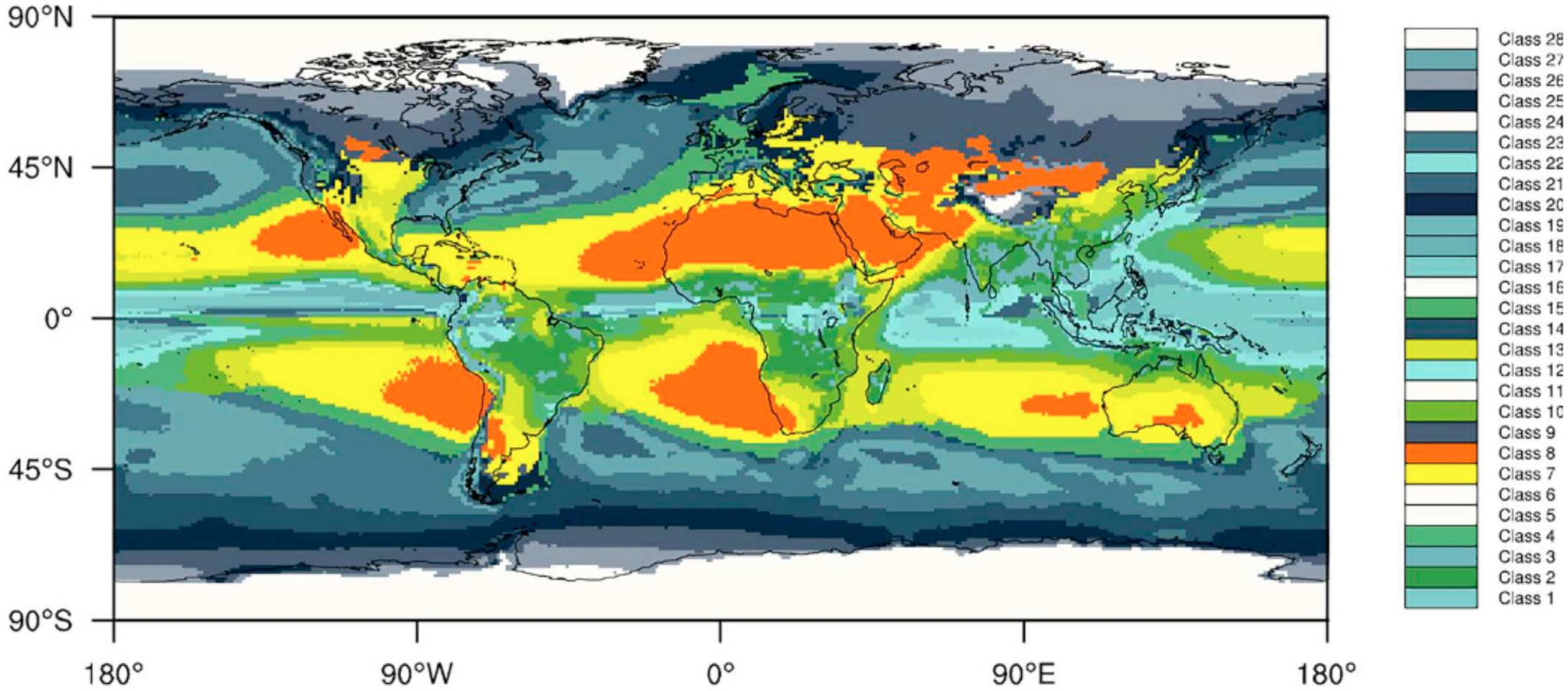
Types of CCS
Alisov

Alisov's scheme aims to understand global climatic features based on **air masses**.



Alisov's seven climatic zones (Alisov 1954). 1: Equatorial zone, 2: Subequatorial zone, 3: Tropical zone, 4: Subtropical zone, 5: Polar zone, 6: Subarctic zone, 7: Arctic/Antarctic zone

Cluster Analysis



Atmospheric Research 216 (2019) 26–36

Contents lists available at ScienceDirect

Atmospheric Research

journal homepage: www.elsevier.com/locate/atmosres

Consensus in climate classifications for present climate and global warming scenarios

Francisco J. Tapiador*, Raúl Moreno, Andrés Navarro

University of Castilla-La Mancha (UCLM), Institute of Environmental Sciences (ICAM), Faculty of Environmental Sciences and Biochemistry, Earth and Space Sciences Group, Avda. Carlos III s/n, 45071 Toledo, Spain

ARTICLE INFO

Keywords:
Climate classification
Global Climate Models
Köppen classification
Köppen-Trewartha classification
K-means clustering

ABSTRACT

Climate classifications of climate models' outputs have been used to assess environmental changes but systematic analyses of the differences between models, scenarios and classification methods are scarce. Here, the results of applying the most commonly used climate classifications to the outputs of 47 Global Climate Models (GCM) of different physical parameterizations and varied grid size are presented. The extent and intensity of changes for present climate, three different Representative Pathways Scenarios (RCP26, RCP45 and RCP85) and three increasingly-fine classification methods show that there is a consensus between models, and that climate classifications are indeed useful tools to translate physical climatology variables into environmental changes. The main conclusions are that climate classifications can indeed be used to gauge model performance at several grid sizes and that the classification method does not decisively affects the potential global changes in future climates under increasing greenhouse gas emissions. The analyses also reveal that there are several uncertainties that are not attributable to model grid size or to limitations in the reference datasets but more likely to deficiencies in the physics of the models.

1. Introduction

Classifications of atmospheric variables are useful tools in atmospheric research to unveil patterns, filter observational data and identify relationships buried in model outputs. Thus, they have been used in the analysis of urbanization effects (Lin et al., 2018); evaluation of precipitation (Ramos, 2001; Serra et al., 2014; Miró et al., 2017; Wen et al., 2017; Kim et al., 2017; Sharifi et al., 2018), and temperature (Peña-Angulo et al., 2016) products; modelling extreme weather and climatic events (Chu and Zhao, 2011; Tramblay and Hertig, 2018); radiation (Rozwadowska, 2004; Vindel et al., 2015); convection (Dimitrova et al., 2009; Aran et al., 2011; Lack and Fox, 2012); and deriving climatologies of processes such as fog (Cereceda et al., 2008) and tornadoes (Giaiotti et al., 2007).

Climate classifications in particular can be characterized as techniques to perform a dimensional reduction of physical variables (precipitation, temperature, evapotranspiration, etc.) into one of two index-classes that can be more readily related to the biota. They have been found useful for a variety of topics including Arctic research (Wang and Overland, 2004); studies of ecosystem impacts (Roderfeld et al., 2008); biome distribution (Leemans et al., 1996) and biodiversity analyses (Garcia et al., 2014); hydrological cycle studies (Manabe and Holloway, 1975); to compare vegetation distribution (Monserud and Leemans, 1992); analyze precipitation metrics (Tang and Hossain, 2012); analyzing vegetation changes in the future (Jiang et al., 2013); provide input to global models (Prentice, 1990) and to visualize climate change (Jylhä et al., 2010) to name but a few.

Climate classifications such as Köppen's date back to pioneer studies in the late 19th century, and were not intended to explain environmental factors in terms of climate variables. Rather, Köppen's specific aim was the other way around: his intention was to be able to define climates based on the observed distribution of vegetation at the time (Köppen, 1900). In synthesis, his method defines a set of temperature and precipitation thresholds to derive a tripartite climate classification, which a 3-letter key identifying the major climate, the precipitation cycles, and the summer temperature. Part of its success (Köppen's method is taught in Geography 101 everywhere) relies on its simplicity. Other index-based classifications such as Budyko's recently discussed by Caracciolo et al. (2018), or the one proposed by Thornthwaite (1948) are less popular.

The inception of coupled and dynamic climate models has allowed to turn upside down Köppen's approach. While his intention was help to define climates using vegetation distribution because of scarce and unreliable meteorological data, data availability is no longer an issue

* Corresponding author.
E-mail address: Francisco.Tapiador@uclm.es (F.J. Tapiador).

https://doi.org/10.1016/j.atmosres.2018.09.017
Received 14 July 2018; Received in revised form 17 September 2018; Accepted 18 September 2018
Available online 24 September 2018

- Multivariate analysis (PR, TS, SLP, PET...)
- Dimension reduction + cluster analysis
- Several approaches
  - K-means
  - K-medoids
  - Ward linkage
- Determining the optimal number of clusters.
- Physical interpretation requires a background in climate science.

Tapiador, F.J. Moreno, R. Navarro, A., (2019) Climate classifications for present and global warming scenarios. *Atmospheric Research*. **216** , 26-36.
https://doi.org/10.1016/j.atmosres.2018.09.017

Dimensionality reduction

The process of reducing the number of random variables or attributes or features under consideration.

Principal components analysis

[A.K.A. Karhunen-Loeve]

PCA searches for k d-dimensional ortonormal vectors that can best be used to represent the data, where $k \leq d$.

Principal components analysis

[A.K.A. Karhunen-Loeve]

PCA *“combines”* the essence of attributes by creating an alternative, smaller set of variables.

What is clustering?

The process of partitioning a set of data objects into subsets *(clusters)*

K-means

K-Means is an iterative algorithm that assigns *K* clusters to a dataset where each cluster has a center that is the average of all the points situated in it, always referred to as the centroid.

Algorithm: *k*-means. The *k*-means algorithm for partitioning, where each cluster's center is represented by the mean value of the objects in the cluster.

Input:

- *k*: the number of clusters,
- *D*: a data set containing *n* objects.

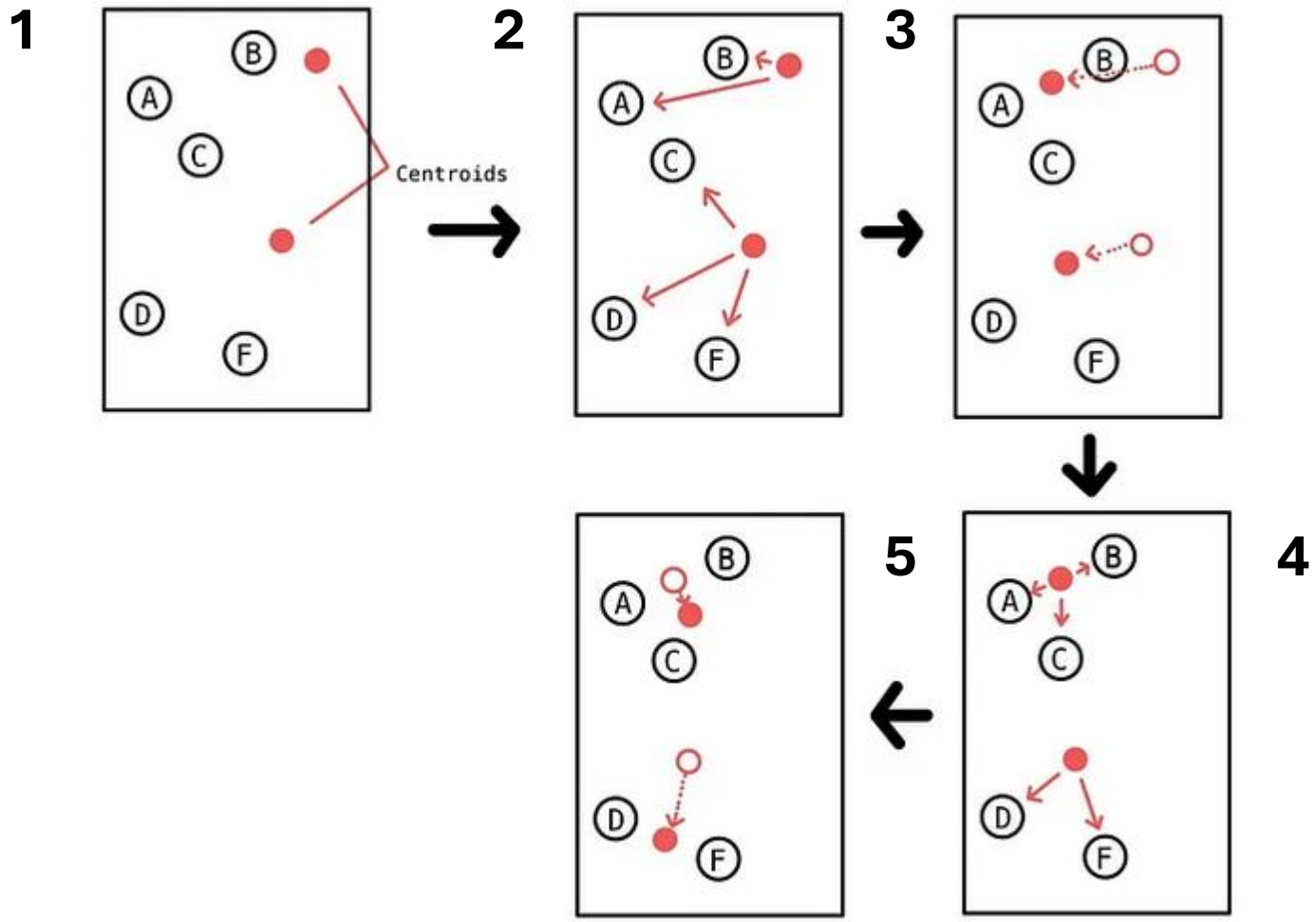
Output: A set of *k* clusters.

Method:

(1) arbitrarily choose *k* objects from *D* as the initial cluster centers;
(2) **repeat**
(3) (re)assign each object to the cluster to which the object is the most similar, based on the mean value of the objects in the cluster;
(4) update the cluster means, that is, calculate the mean value of the objects for each cluster;
(5) **until** no change;

The *k*-means partitioning algorithm.

K-means



K-means

Advantages

- Simplicity
- Efficency
- Speed

Disadvantages

- Sensitivity to outliers
- Shape assumption
- Initial centroids

K-medoids

Partitioning Around Medoids (PAM), is similar to the K-Means clustering method but requires the use of medians for the formation of subgroups. A medoid is a centroid that best represents the objects in a defined cluster.

Algorithm: *k*-medoids. PAM, a *k*-medoids algorithm for partitioning based on medoid or central objects.

Input:

- *k*: the number of clusters,
- *D*: a data set containing *n* objects.

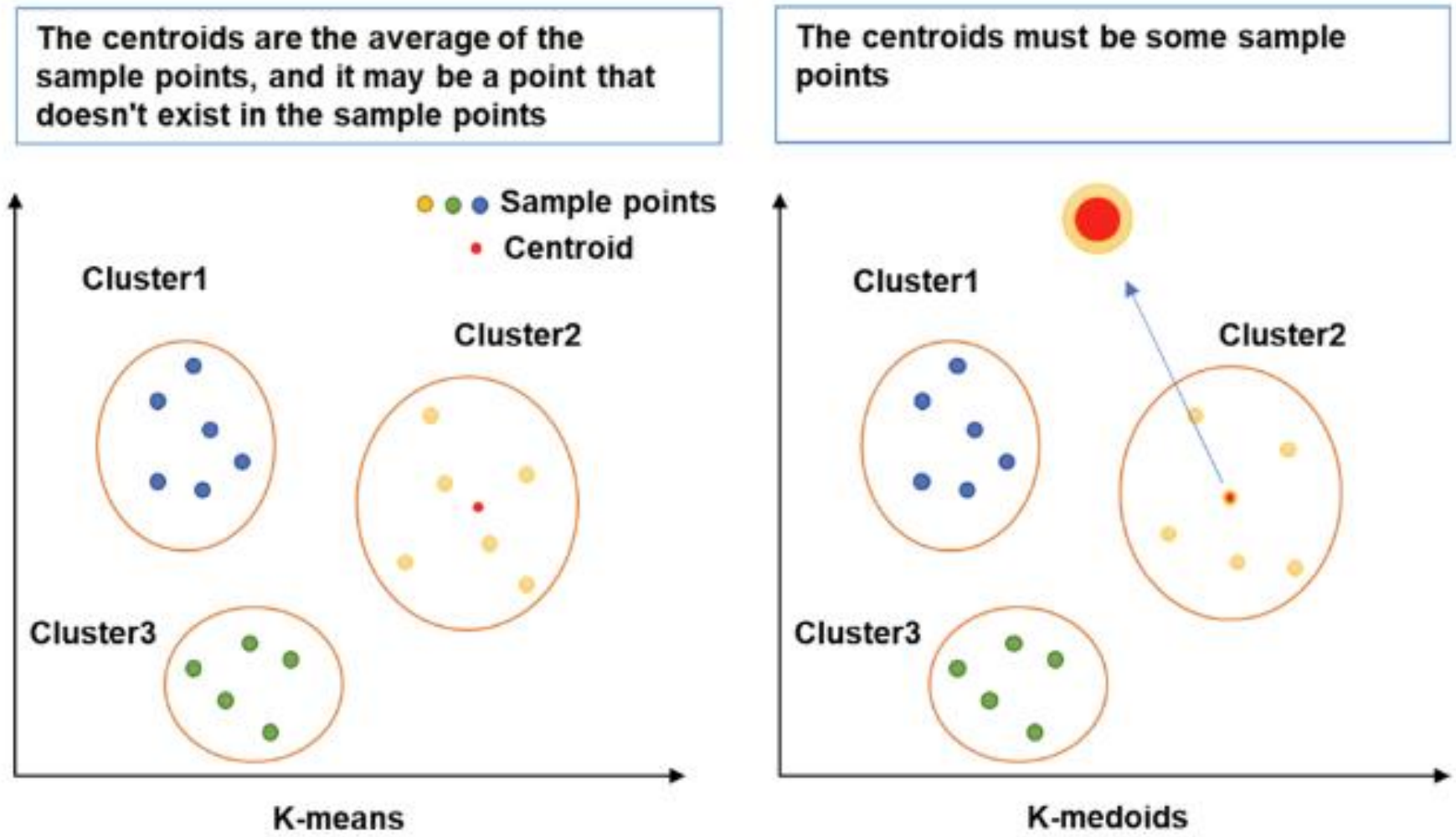
Output: A set of *k* clusters.

Method:

(1) arbitrarily choose *k* objects in *D* as the initial representative objects or seeds;
(2) **repeat**
(3) assign each remaining object to the cluster with the nearest representative object;
(4) randomly select a nonrepresentative object, o_{random};
(5) compute the total cost, *S*, of swapping representative object, o_j, with o_{random};
(6) **if** $S < 0$ **then** swap o_j with o_{random} to form the new set of *k* representative objects;
(7) **until** no change;

PAM, a *k*-medoids partitioning algorithm.

K-medoids


The centroids are the average of the sample points, and it may be a point that doesn't exist in the sample points
Sample points
Centroid
Cluster1
Cluster2
Cluster3
K-means
The centroids must be some sample points
Cluster1
Cluster2
Cluster3
K-medoids

K-medoids

Advantages

- Robustness to outliers
- Flexibility in distance metrics

Disadvantages

- Computationally intensive
- Complexity

Aspect	K -Means Clustering	K-Medoids Clustering
Representation of Clusters	K-Means Clustering uses the mean of points (centroid) to represent a cluster.	It uses the most centrally located point (medoid) to represent a cluster.
Sensitivity to Outliers	Highly sensitive to outliers.	More robust to outliers.
Distance Metrics	K-Means primarily uses Euclidean distance.	Whereas it can use any distance metric.
Computational Efficiency	K-Means is generally faster and more efficient	It is slower due to the need to calculate all pairwise distances within clusters.
Cluster Shape Assumption	It assumes spherical clusters.	It does not make strong assumptions about cluster shapes.

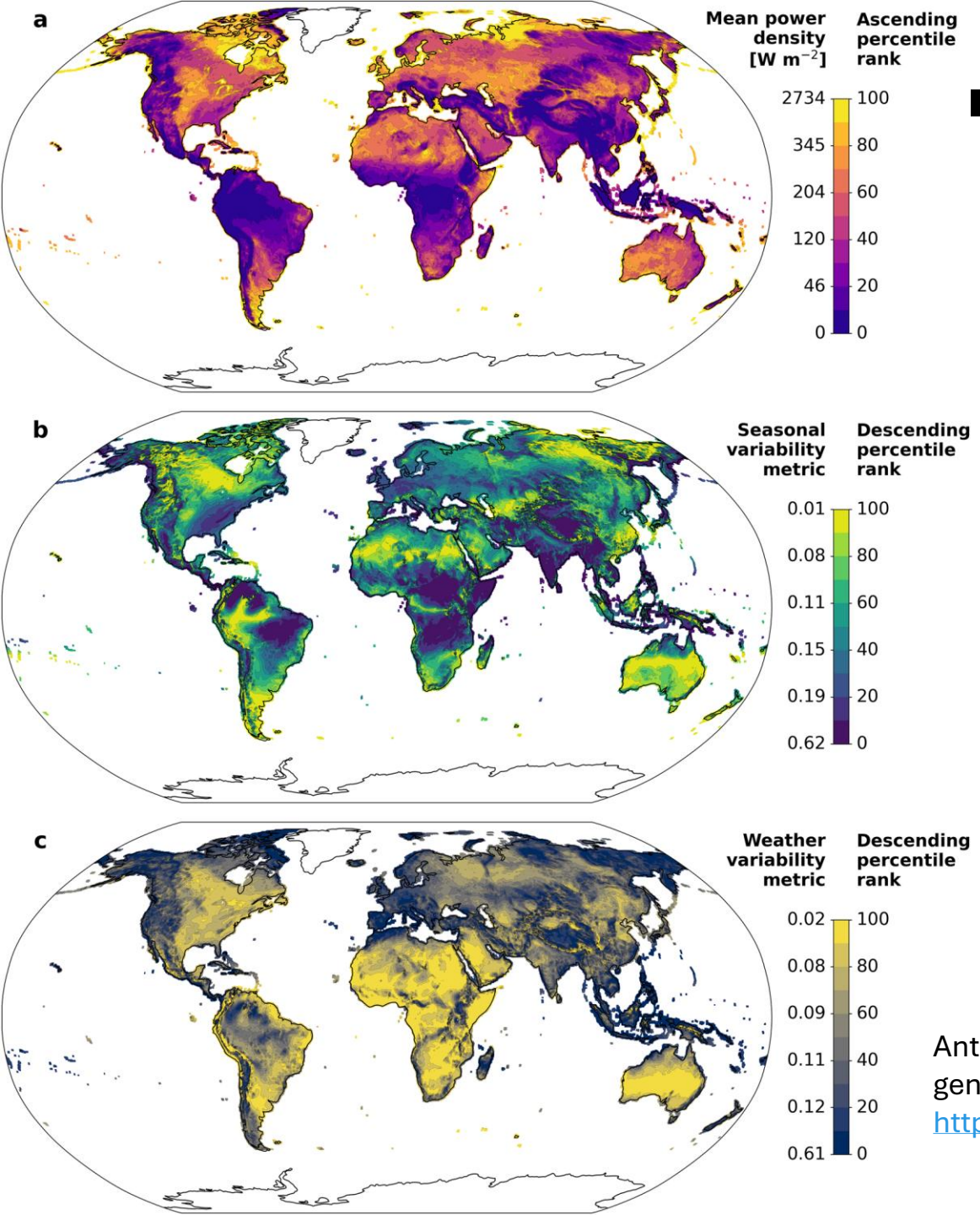
Source: www.geeksforgeeks.org

What variables are relevant?

The criteria used in the classification of climates depend on the use of the classification.

What variables are relevant?

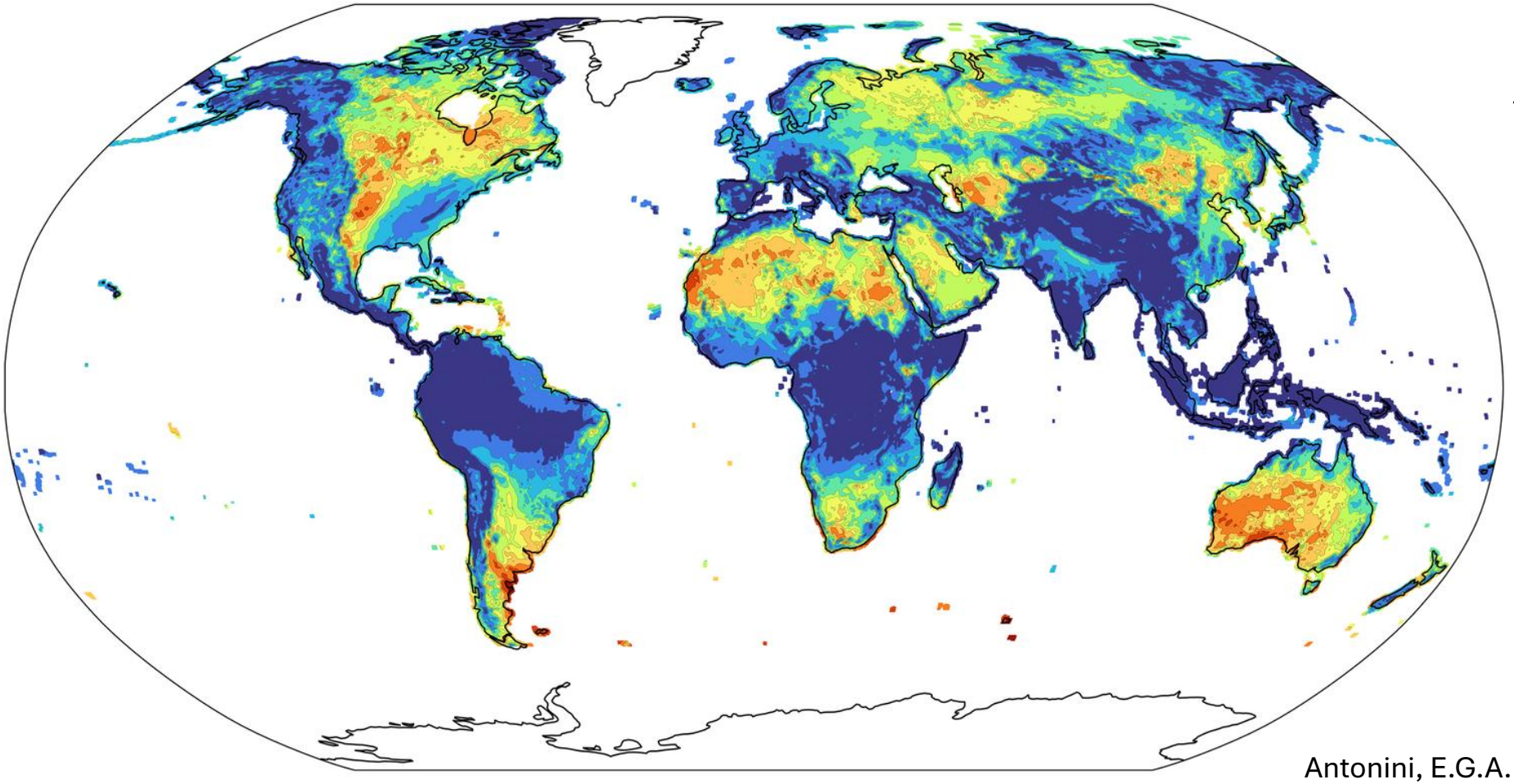
If you were conducting your climatic classification as an employee of a giant energy conglomerate that was investigating the feasibility of installing wind turbines around the world to generate wind power, the most logical atmospheric variables to include might be ….



3 metrics in a single index

High mean power densities, with low seasonal and weather variabilities, would tend to make a location more attractive for wind generation. Color maps in all panels are such that lighter colors indicate better quality of wind resources.

Antonini, E.G.A. *et al.*(2024) Identification of reliable locations for wind power generation through a global analysis of wind droughts. *Commun Earth Environ* **5**, 103. https://doi.org/10.1038/s43247-024-01260-7



Areas with abundant and reliable wind power

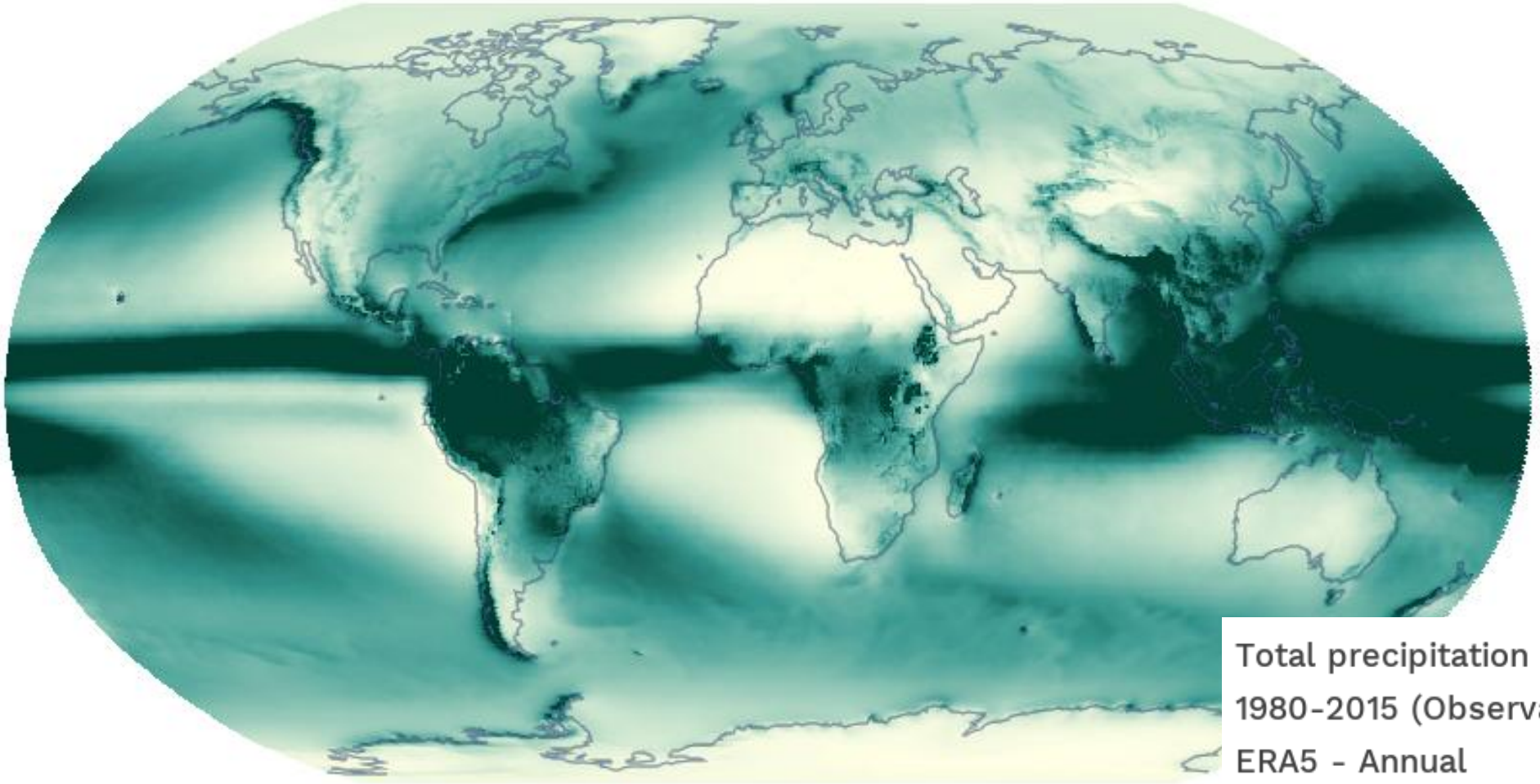
Minimum percentile rank across power density, seasonal variability, and weather variability

Areas that this metric identifies as having good wind resources are shown in orange and red colors.

Antonini, E.G.A. *et al.*(2024) Identification of reliable locations for wind power generation through a global analysis of wind droughts. *Commun Earth Environ* **5**, 103. https://doi.org/10.1038/s43247-024-01260-7

Key variables in CCS

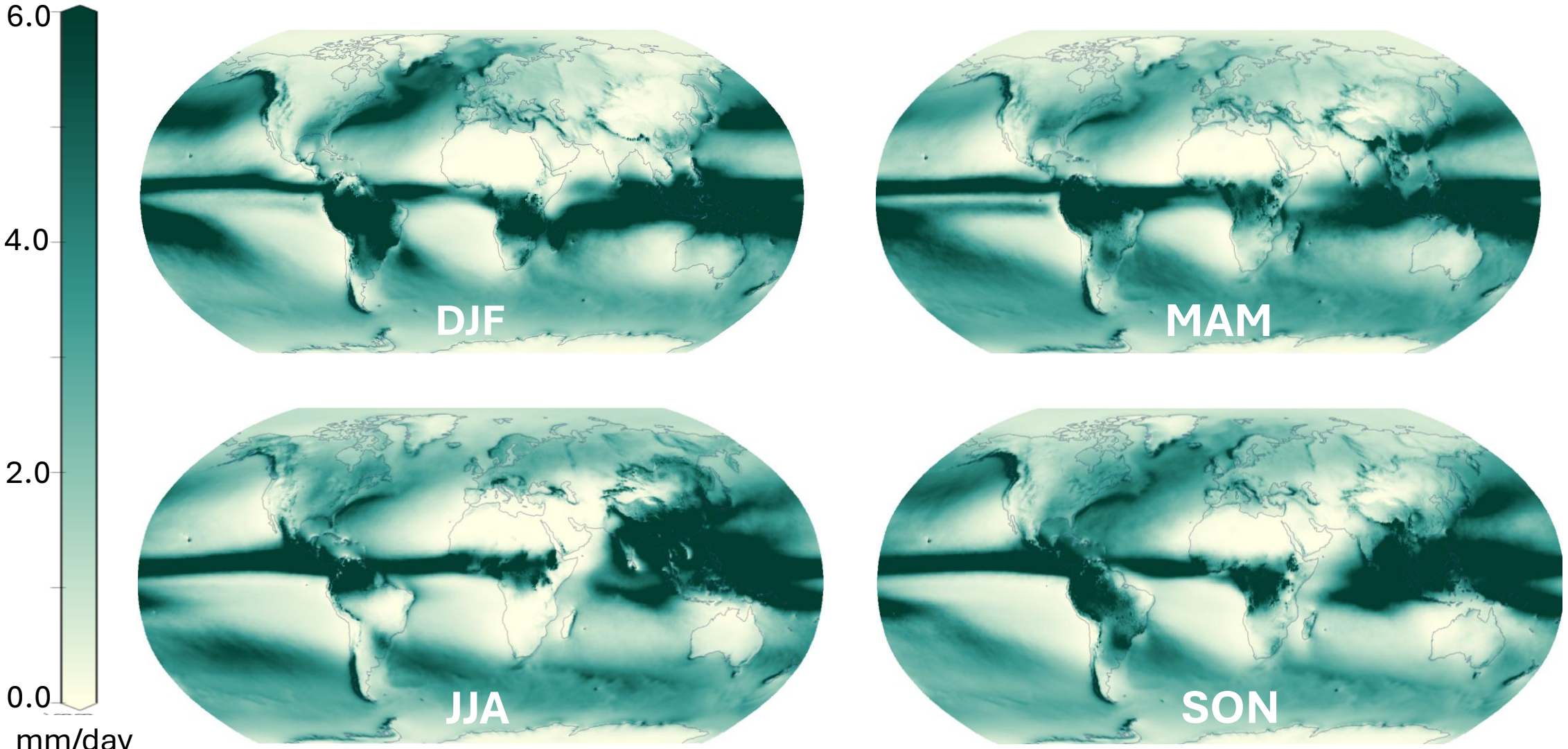
PR



Total precipitation (PR) - (mm/day)
1980-2015 (Observations)
ERA5 - Annual

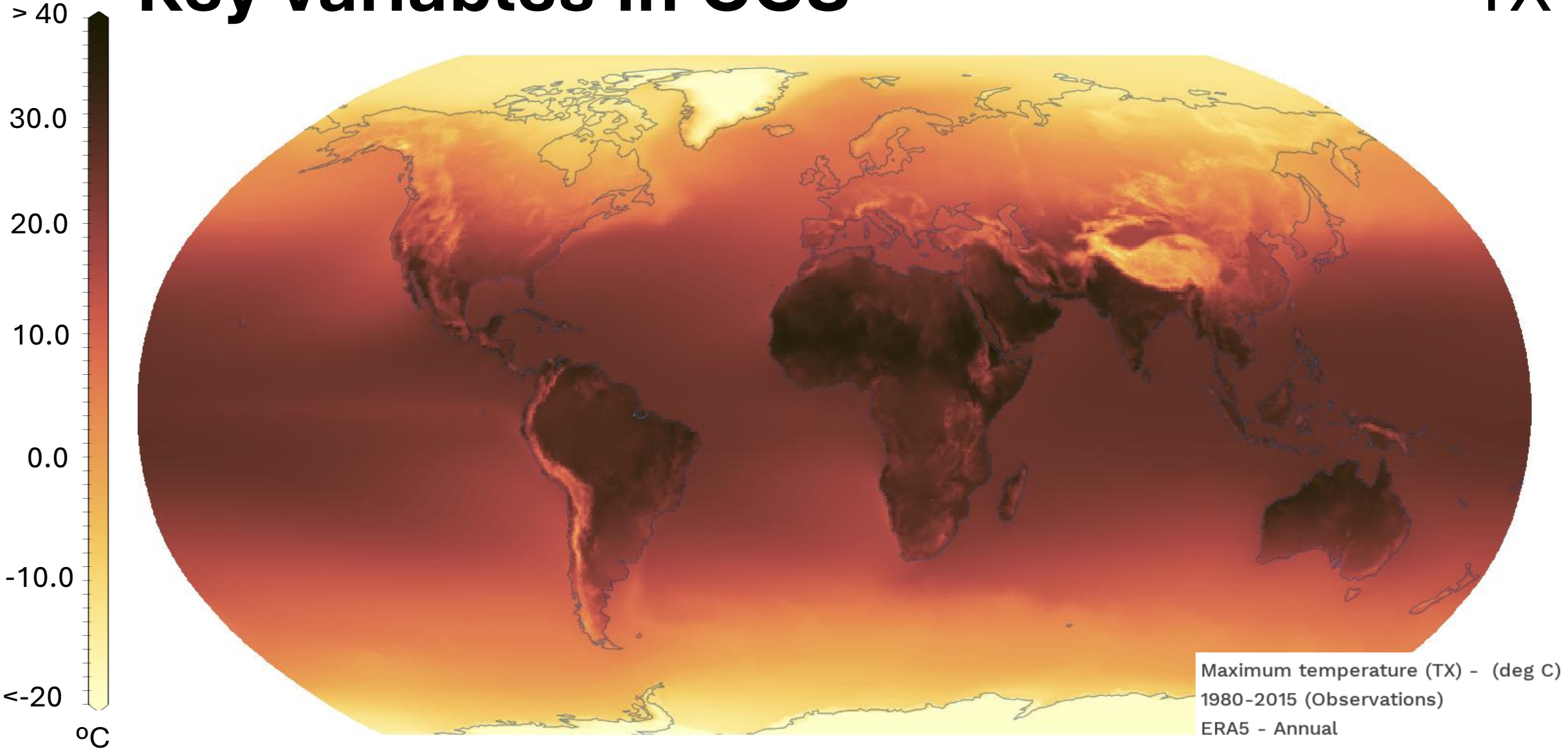
Source: https://interactive-atlas.ipcc.ch

Seasonality



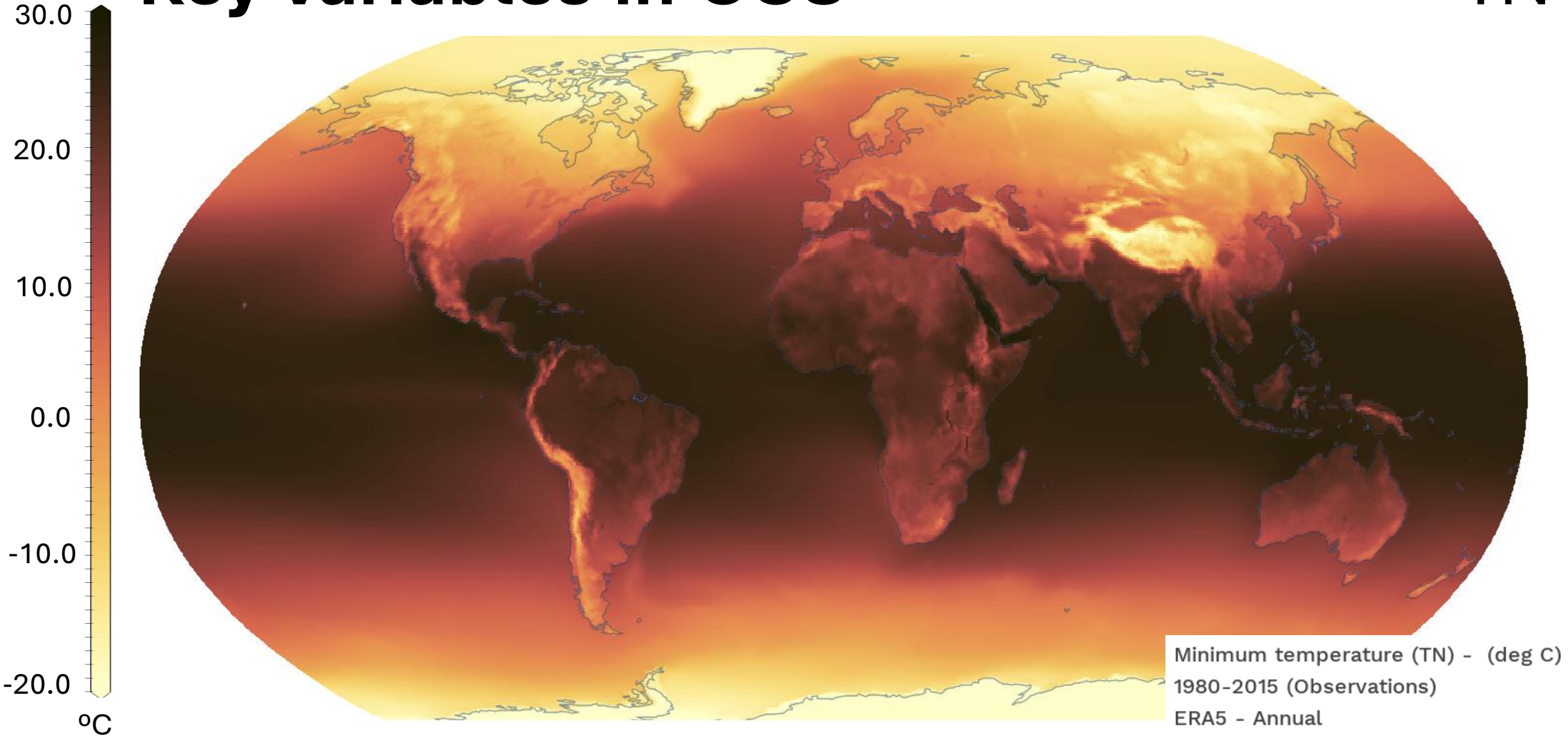
Key variables in CCS

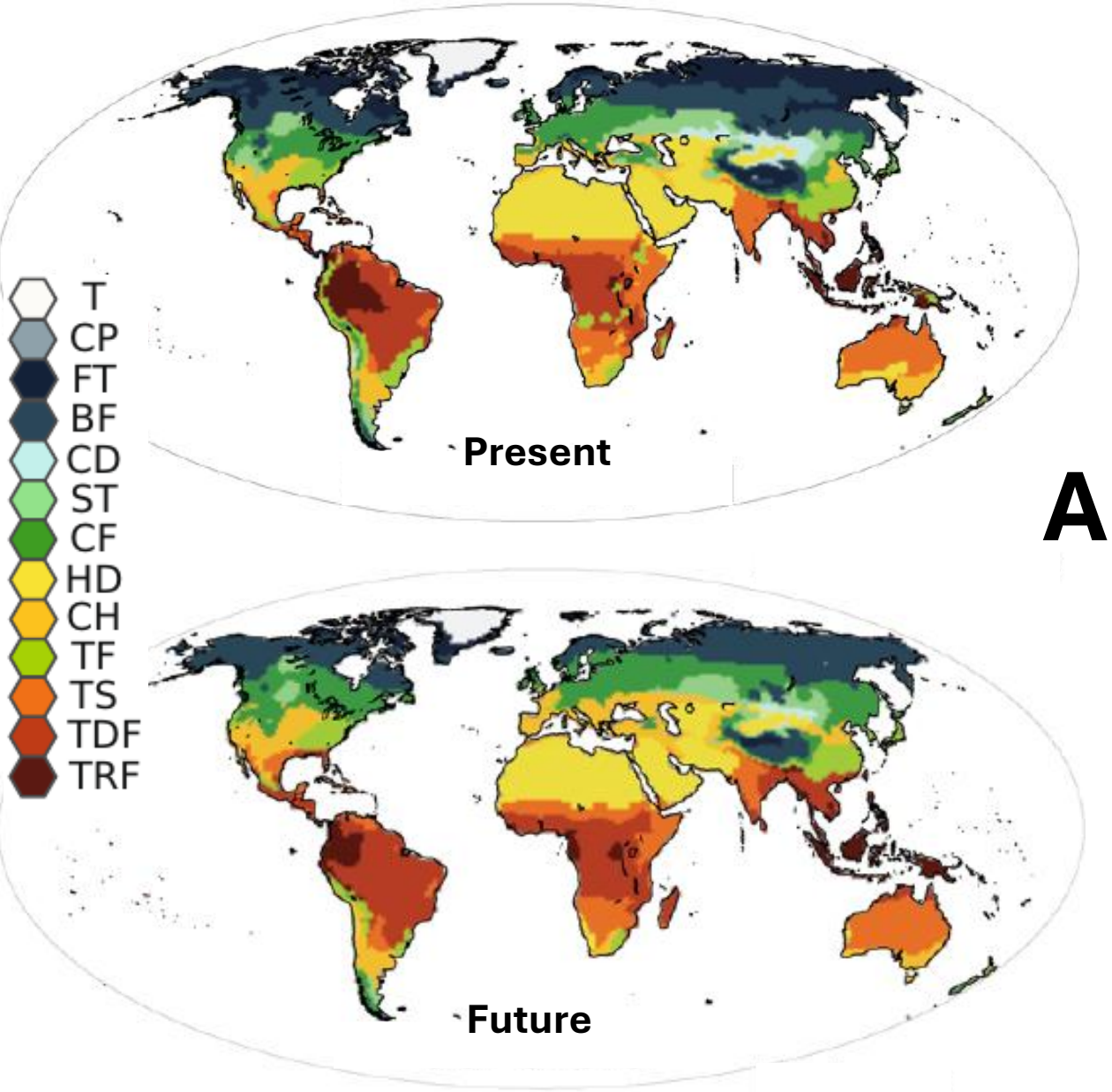
TX



Key variables in CCS

TN



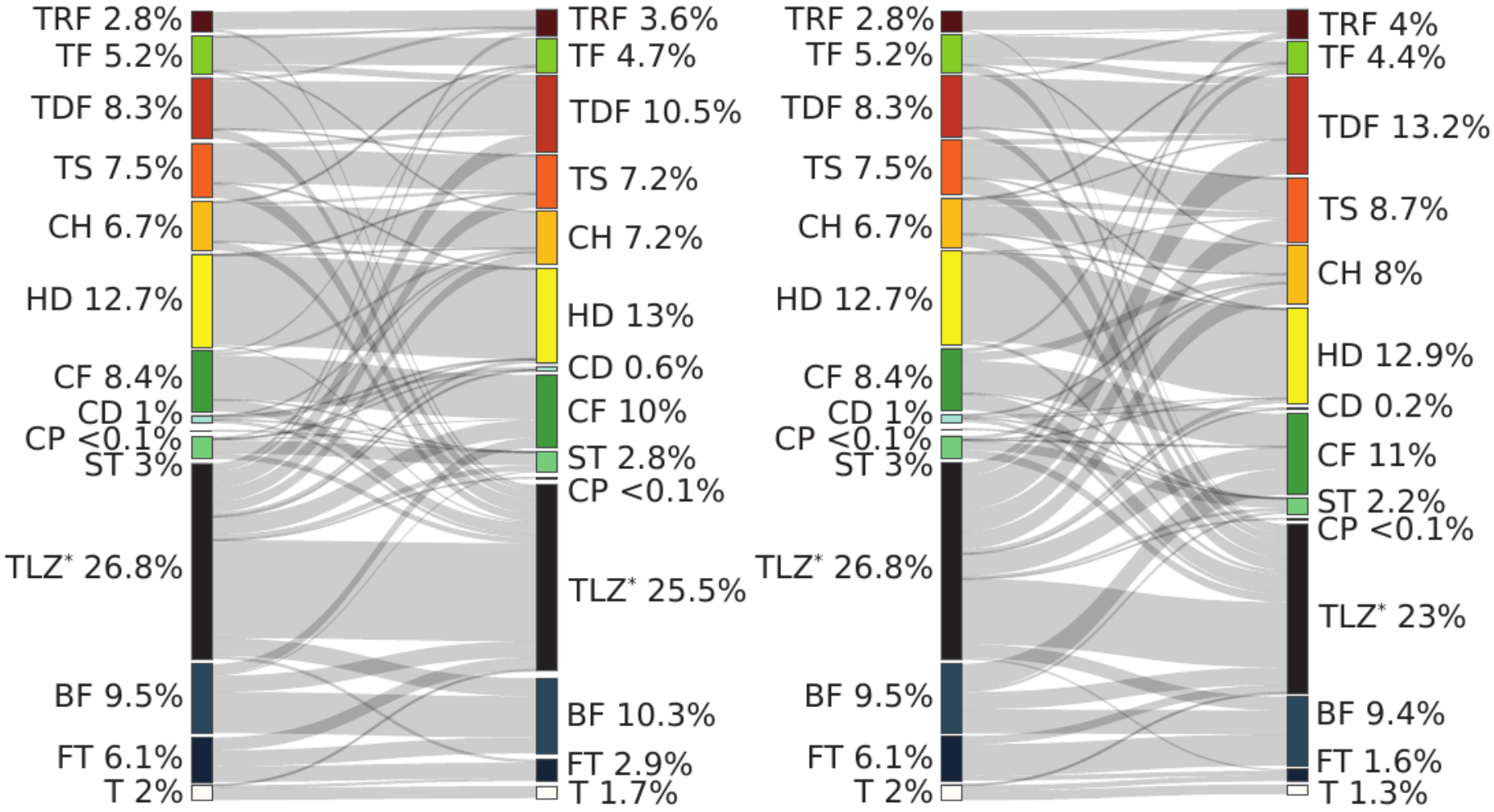


A Changing Climate...

Navarro *et al.*(2024) Uncertainties in measuring precipitation hinders precise evaluation of loss of diversity in biomes and ecotones. *npj climate and atmospheric science* **7**, 35. https://doi.org/10.1038/s41612-024-00581-w

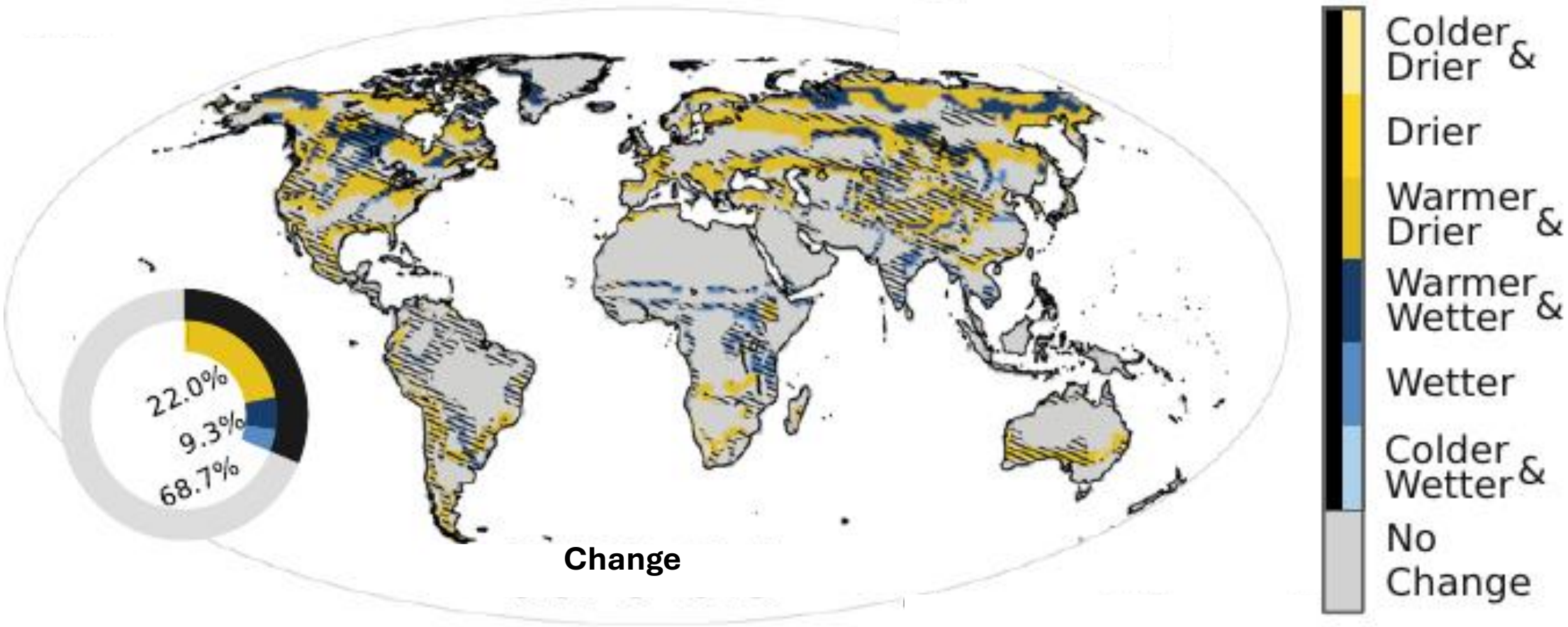
A Changing Climate…

Life zone area changes from present climate (1980–2014) to future climate (2015–2100). Flow diagrams for the low-emission scenario SSP1-2.6 (Left) and high-emisión scenario SSP5-8.5 (Right). The twenty-seven ecotones were aggregated into one category (TLZ*) for a better visualization.



Navarro. *et al.*(2024) Uncertainties in measuring precipitation hinders precise evaluation of loss of diversity in biomes and ecotones. *npj climate and atmospheric science* **7**, 35. https://doi.org/10.1038/s41612-024-00581-w

A Changing Climate...

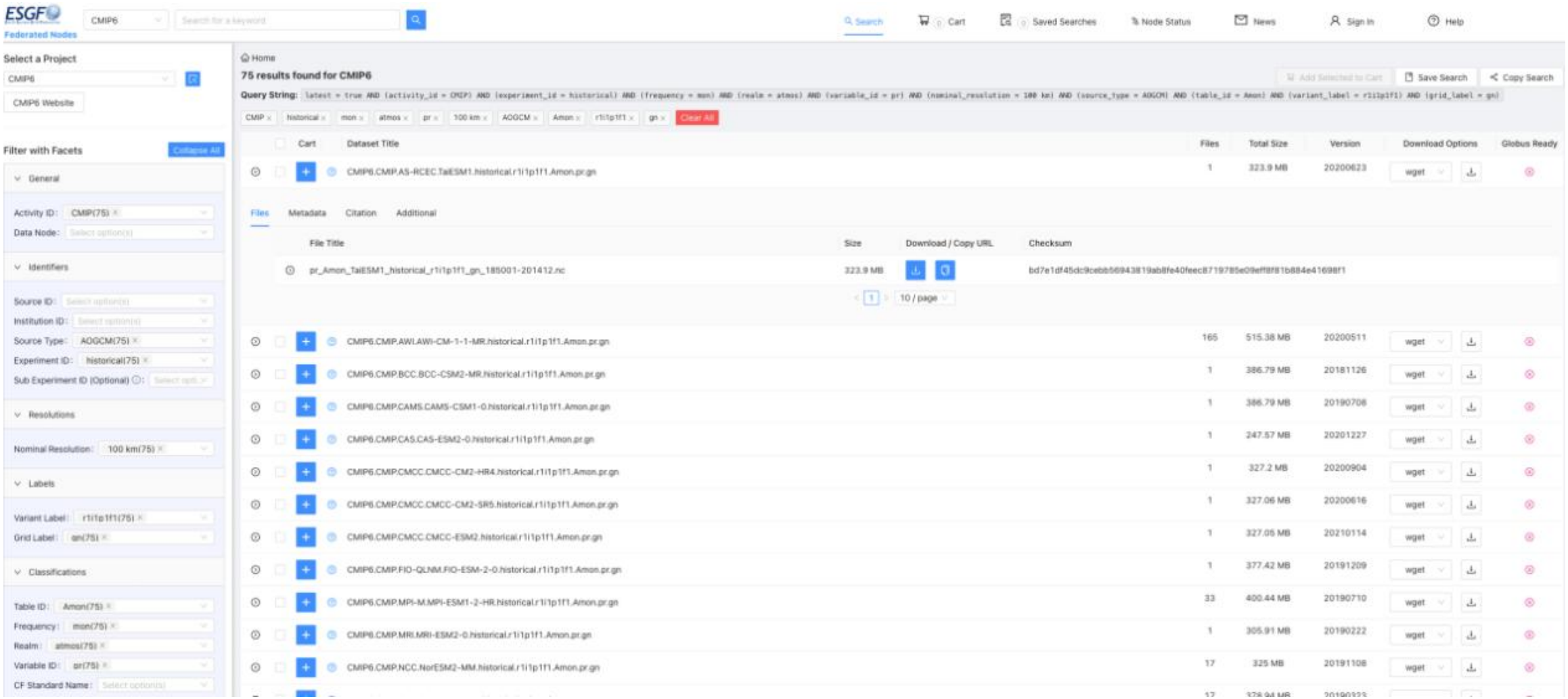


Navarro *et al.* (2024) Uncertainties in measuring precipitation hinders precise evaluation of loss of diversity in biomes and ecotones. *npj climate and atmospheric science* **7**, 35. https://doi.org/10.1038/s41612-024-00581-w

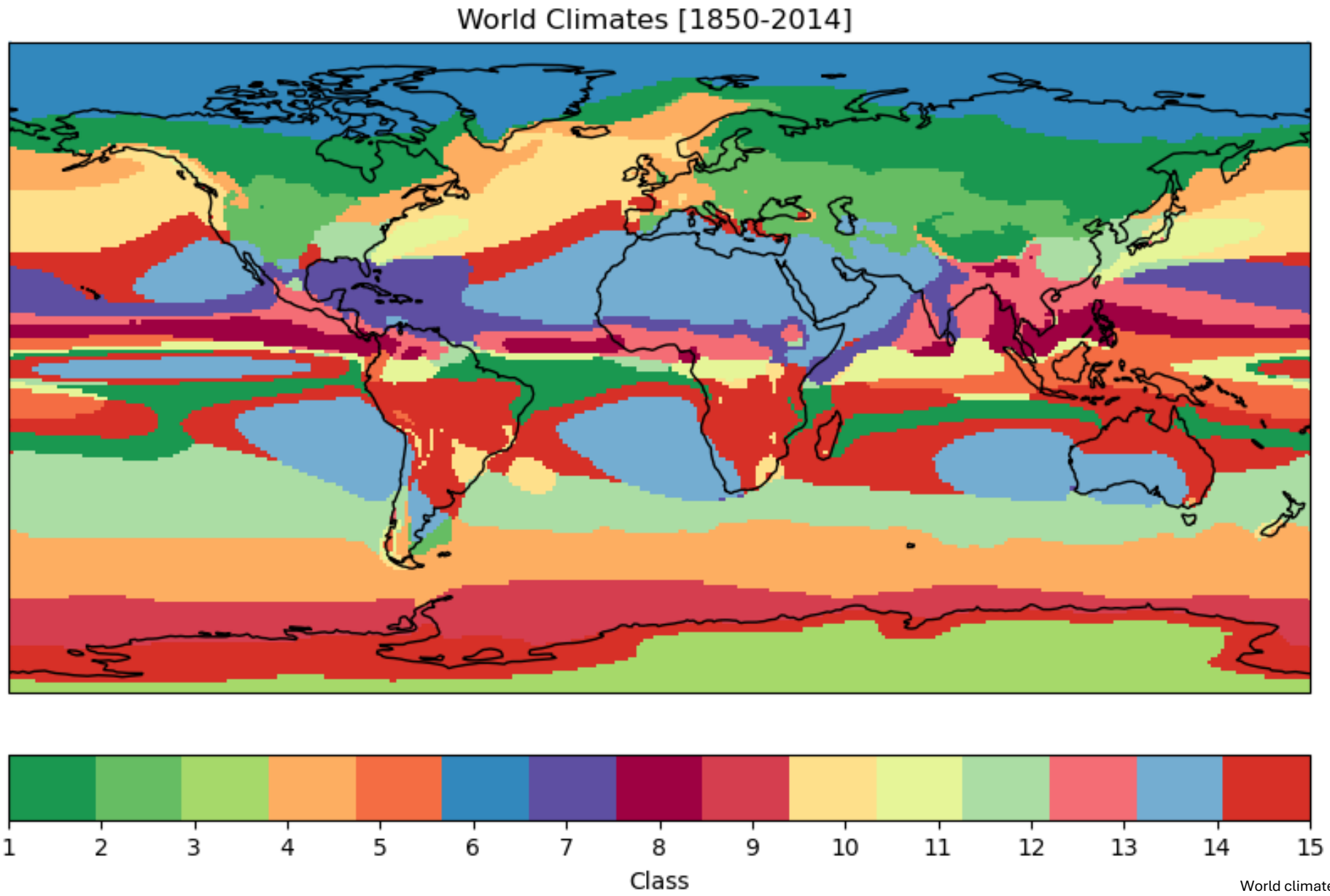
Assignment: Climate Classification using K-Means Clustering on CMIP6 Ensemble Data

Assignment

1. Get *pr & ts* data from CMIP6
2. Pre-process the data
3. Dimensionality reduction with PCA
4. K-means cluster analysis
5. Analysis and report



https://esgf-metagrid.cloud.dkrz.de/search



World climate distribution [15 categories]