Determination of climate zones in Conterminous United States for public health vulnerability assessment and modeling of infectious diseases

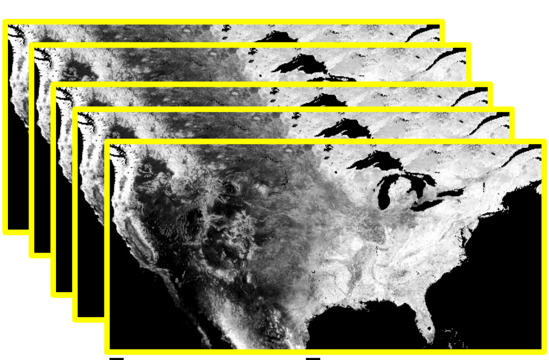
Alexander Liss, CFA, MBA, MSc, CFFA
Civil and Environmental Engineering
Center for Modeling Infectious Diseases
Tufts University, Medford, MA, USA

WHY?

The extreme weather events cause large number of adverse health outcomes. Heat waves and cold spells cause increase in mortality and hospitalizations for vulnerable population, hurricanes and tornado cause an increase in trauma related hospital admissions. It has been shown that there could be substantial difference in the rate of hospitalizations for similar events conditioned on the climate patterns at patient's locale. The climate differences play a major role in assessing vulnerability scores for different locations and efficiency of an early warning systems and public health interventions. Climate classification allows separating a large continuous territory into smaller areas with similar weather conditions.

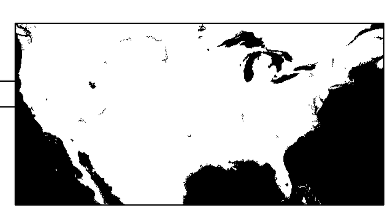
The goal of this work is to define, using remote sensing data, climate zones for the Cont. United States that would be useable for a healthcare vulnerability assessment and forecasting and modeling of infectious diseases and health impacts of extreme weather

HOW?



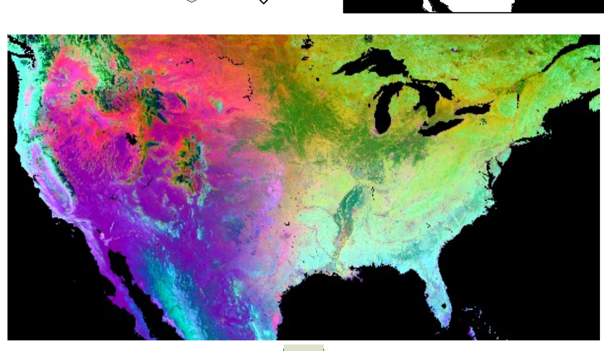
228 NDVI biweekly snapshots

MODerate-resolution Imaging Spectroradiometer (MODIS) on board NASA's Terra and Astra satellites produces worldwide NDVI snapshot every 16 days. The remote sensing data was downloaded from an FTP depository . The analysis used data set for the period from Jul 4, 2002 to March 13, 2012. There were 228 individual snapshots in this data set. Each NDVI snapshot has a worldwide coverage, with -180 to +180, -90 to +90 degrees extent, placed on 0.05x0.05 degree grid. For the climate analysis of the conterminous United States we clipped the worldwide NDVI snapshot to a bounding box of -125 to -65, 24 to 50 degrees



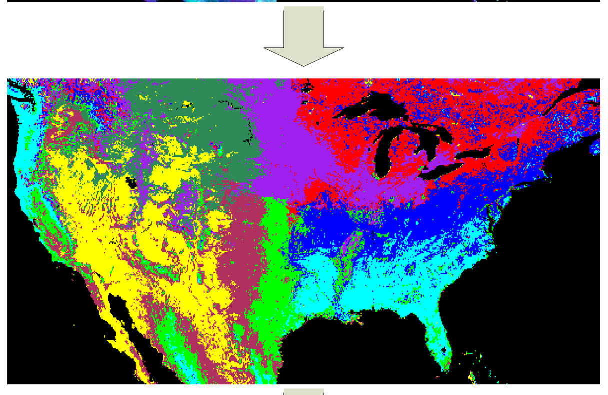
Water mask

The water reflectance pattern should produce very low values in NDVI data. However it often produced an uneven reflectance and the large water bodies needed to be masked in our analysis. The raster water mask was built based on a vector map data. The water mask is a snapshot in the same geographic coordinates as an original data, with cell value of 1 indicating a land mass and with a value of 0 indicating large bodies of water It was applied to the 223 NDVI snapshots prior to the analysis.



12 Principal components

The NDVI data is a 223-dimensional data set. The data set is redundant and heavily correlated across time and space. The principal component analysis is a mathematical orthogonal transformation that creates a set of variables from an original data set that is linearly independent. The first principal component contains the largest amount of information with ach of the following components has diminishing marginal information. This causes the majority of useful information to concentrate in the first few components.

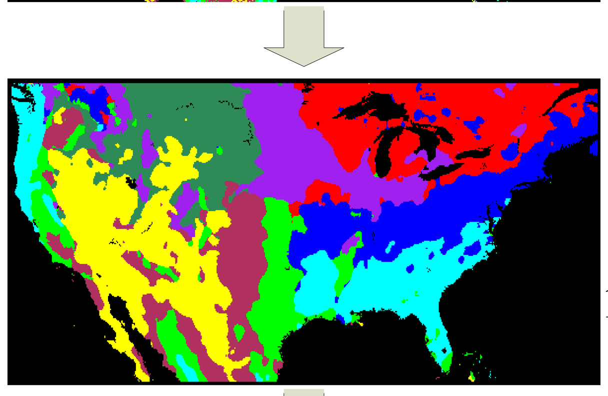


K-Means classification

8 zones

The k-means unsupervised classification algorithm groups data points based on the level of their "similarity", using Euclidean inter- and intra-cluster distances.

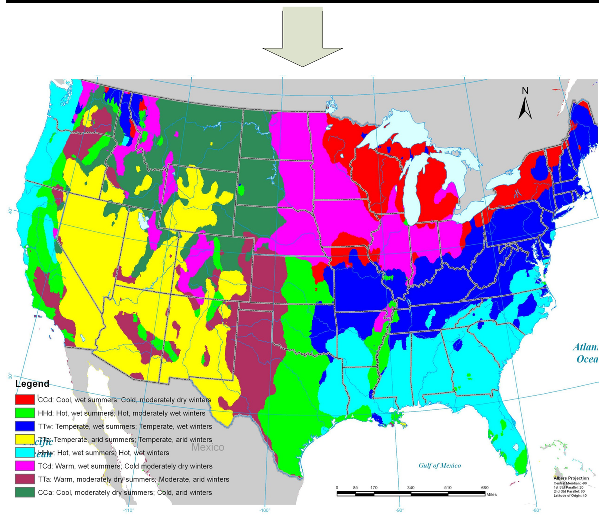
The number of clusters decided based on the value of a Calinski-Harabasz cluster validity index.



Majority Analysis

15x15 convolution kernel

The convolution with a "majority" 15x15 kernel applied to the output of the k-means algorithm. The "majority" kernel assigns value to each point in the data set equal to the value of the majority points surrounding the point, excluding those that do not have class values (ex: points over water). This transformation makes inter-cluster edges less jagged, more smoothly defined, and it removes smaller clusters within large areas.

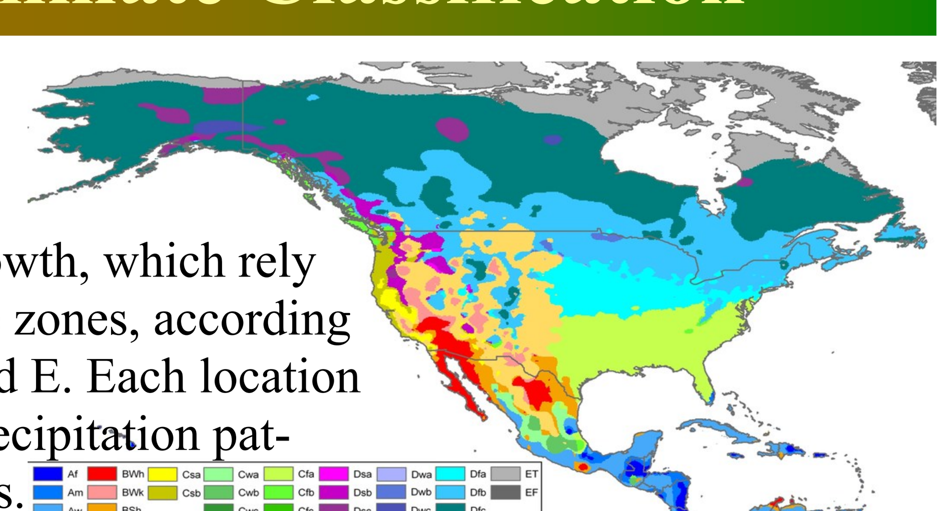


Climate and healthcare statistics per zone

Using Spatial Analyst and Zonal Analysis every zip code is assigned to a majority class. The zip code is a unit of analysis in the healthcare data set. The person at risk is defined as an elderly person, age 65 and older. The membership of a person at risk in a particular climate zone is defined by his/her residential address. The person is considered to belong to the same climate class as the zip code where he/she resides. Using zip code allocation computed an average number of elderly people (65 y.o. or older) living in each climate zone for 16 years between 1991 and 2006 and determined a number of hospitalizations due to hypothermia, an environmentally induced condition.

Original Köppen-Geiger Climate Classification

Today the climate classification of Köppen and Geiger is assigned to the classical climatology and is called a generic climate classification. This is a classification that identifies climates in similarity to their effects on plant growth, which rely mainly on aridity and warmth. There are five main climate zones, according to the original Köppen Classification: zones A, B, C, D and E. Each location is assigned to a specific zone based on temperature and precipitation patterns. Each major zone is subdivided to several sub regions.

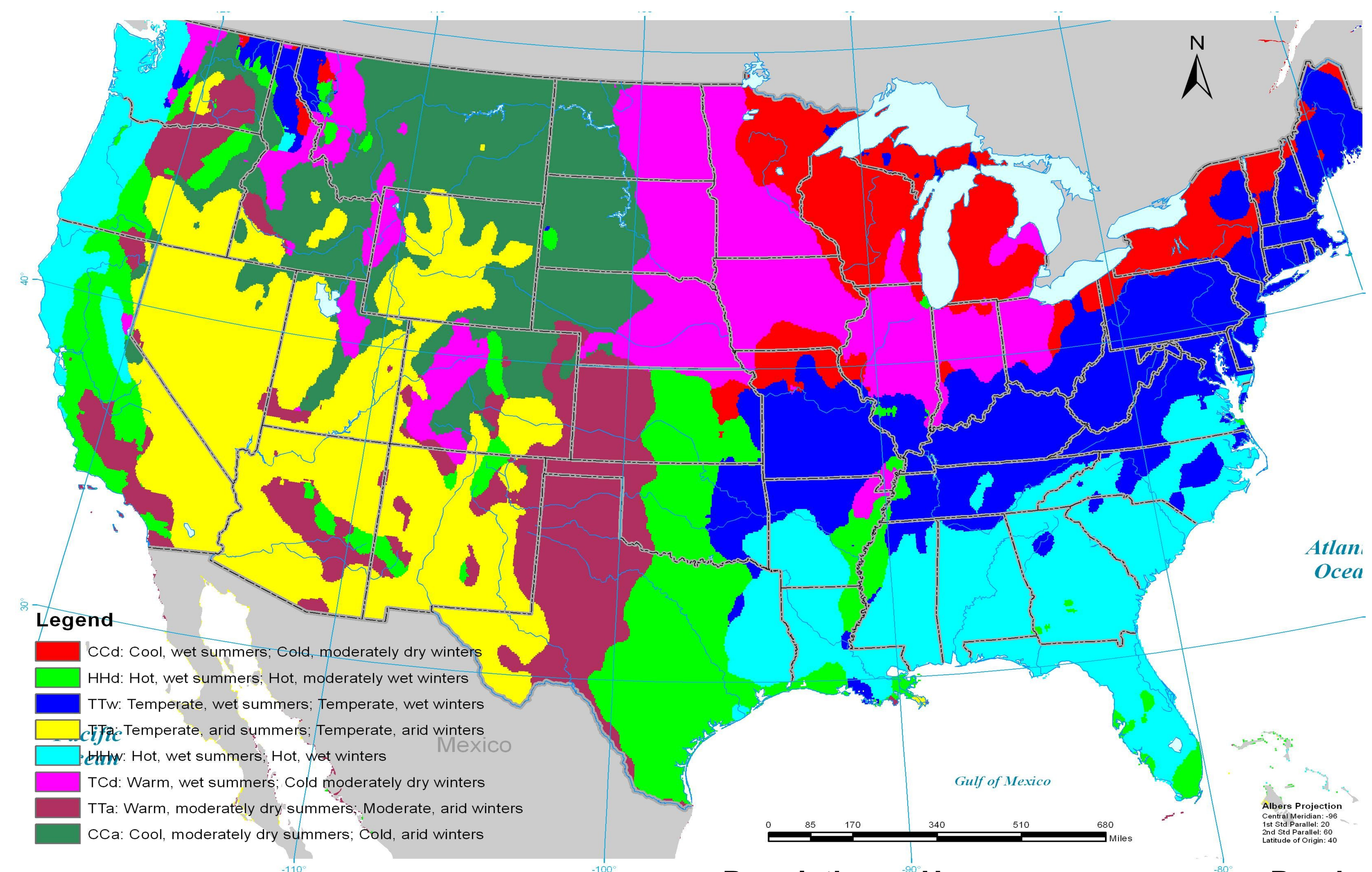


NDVI: Normalized Difference Vegetation Index

NDVI= (NIR-Red)/(NIR + Red)

The original Köppen-Geiger climate classification relied on proxy measurements, such as ambient temperature and aridity to identify climate similarity relative to plant's growth patterns. Recently with the application of remote sensors we can directly record properties of vegetation cover without the need of proxies. The plant leaves reflect strongly in Near Infrared (NIR) part of the spectrum (wavelengths of 700 to 1000 nanometers), while strongly absorbing in the Red spectrum (400 to 700 nanometers). The NDVI is a ratio between the difference and the sums of reflectance in these bands. The greater values of the index indicate stronger vegetation cover in the region. MODIS Terra and Astra satellites provide worldwide coverage. The NASA produces twice a month a worldwide NDVI map from 2002 to present with 0.05 degree resolution.

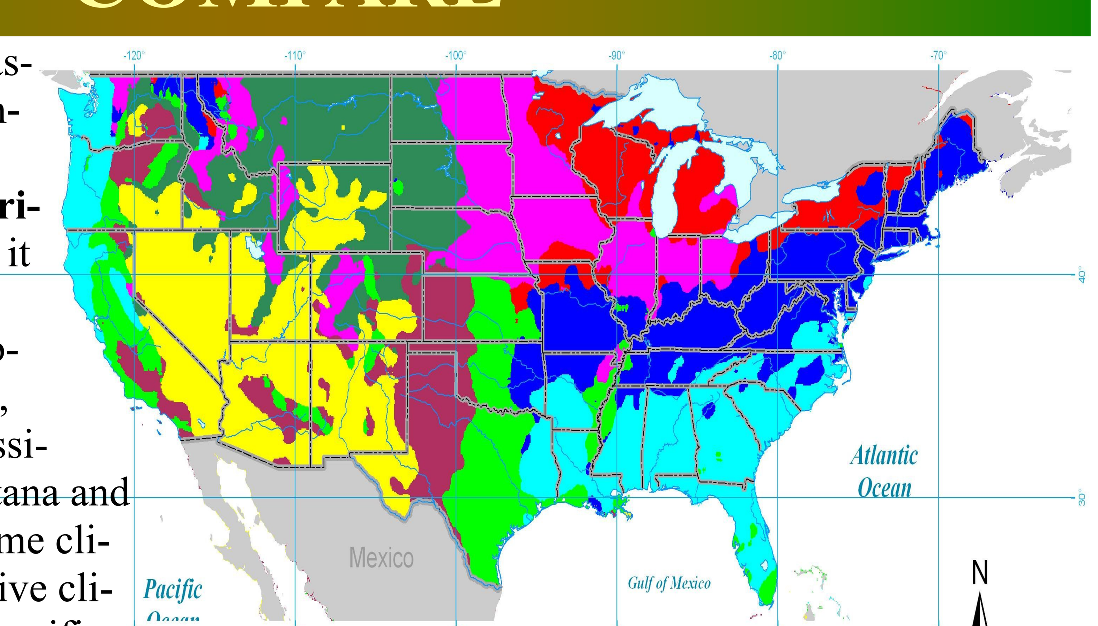
RESULTS

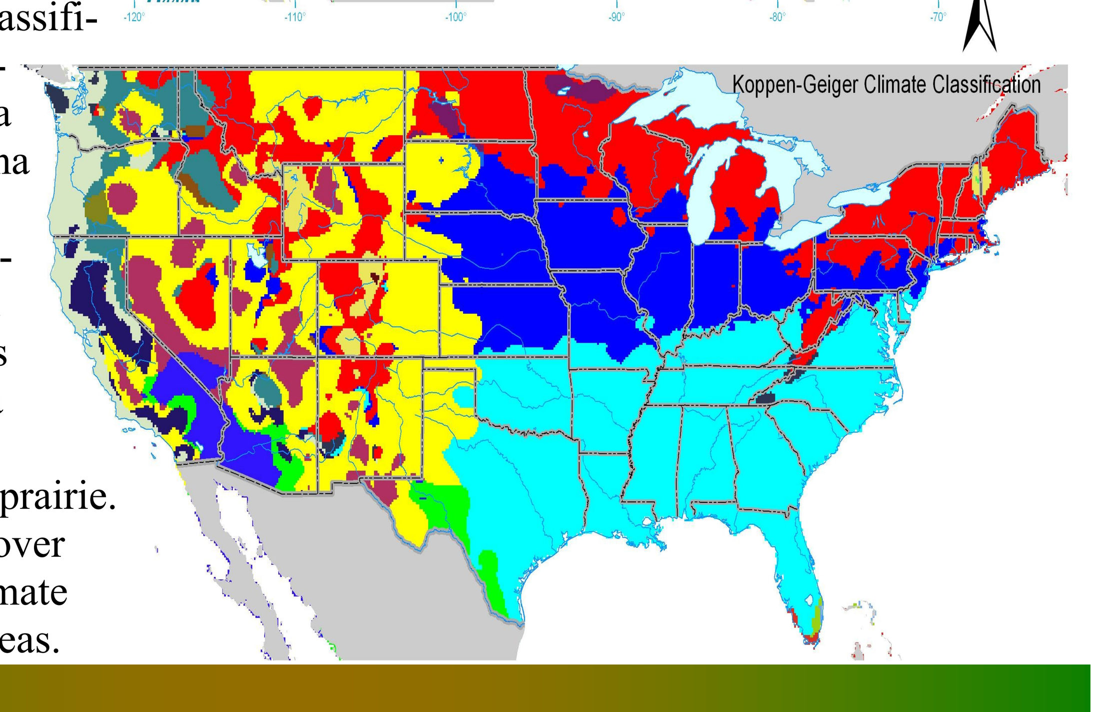


Zone Description	Population, at risk	Hosp, 65+ y.o.	Rate per 10K	Temp,	Precip, mm
1. Cool wet summers; Cold moderately dry winters	3.3MM (10.4%)	7,509	22.8	7.5 (5.9;9.1)	845
2. Hot wet summers; Hot moderately wet winters	5.1MM (16.2%)	10,376	20.2	16.1 (13;19)	737
3. Temperate wet summers; Temperate wet winters	10.9MM (34.5%)	24,803	22.7	12.1 (10;14)	1,088
4. Temperate arid summers; Temperate arid winters	1.2MM (3.9%)	1,790	14.6	11.1 (7;15)	253
5. Hot, wet summers; Hot wet winters	5.3MM (16.9%)	15,109	28.3	16.6 (14;19)	1,254
6. Warm, wet summers; Cold moderately dry winters	3.4MM (10.6%)	8,156	24.3	8.8 (6;11)	699
7. Warm moderately dry summers; Moderate arid winters	2.0MM (6.2%)	2,976	15.2	12.8 (10;16)	4328.
8. Cool moderately dry summers; Cold arid winters	0.4MM (1.4%)	1,240	27.9	6.8 (5; 8)	358

Climate data shows well defined separation between climate zones based on seasonal characteristics, as well as annual temperature and precipitation parameters. The largest population is in zone 3, 10.9 million persons at risk. The smallest population belongs to zone 8, 400 thousand people at risk. The highest hospitalization rate due to hypothermia is in zones five, 28.3 and three – 27.9 hospitalizations per 10,000 persons at risk. The lowest rate is in zone 4, 14.6 and in zone seven, 15.2 hospitalizations per 10,000 persons at risk. This data shows that relatively warm, south eastern region has highest rate of hypothermia hospitalizations per person at risk, whereas, a dry south-western region has the lowest rate.

COMPARE





Comparison of an objective classification with existing Köppen-Geiger climate zone divisions shows **a high degree of similarity between the two**. However, it also shows **more intuitive climate divisions** produced by objective clustering. For example, an original Köppen-Geiger classification assigns Northern Montana and Western New Mexico to the same climate zone. This is not an intuitive climate division. The objective classification, on the other hand, intuitively places Northern Montana and majority of Nevada, Arizona and New Mexico into separate climatic regions. However Mississippi river delta, an irrigated agricultural land is classified as a separate climate area due to a difference in crop types on the irrigated land and surrounding prairie. This difference in vegetation cover does not necessarily reflect climate differences in these adjacent areas.

Acknowledgments

I would like to express my deepest gratitude to all those wonderful people who gave me the possibility to complete this work.
I am deeply indebted to my adviser, **Prof. Elena Naumova** whose help, stimulating discussions, valuable suggestions and constant encouragement helped me to complete this work and to make it much better.
I want to thank **Magaly Koch** for her thoughtful presentation of a difficult material, willingness to help and valuable suggestions that tremendously improved this work.
I would like to thank **Jonathan Gale, Patrick Florance, Barbara Parmenter**
, and **Tufts GIS Lab Assistants** for their knowledge and willingness to help with difficult technical issues.
While many people helped me to improve this work, the errors and omissions are mine alone.

References

Calinski, R.B. and J. Harabasz, *A Dendrite Method for Cluster Analysis.* Comm. in Statistics, 1974. **3**
Fovell, R.G. and M.Y.C. Fovell, *Climate zones of the conterminous United States defined using cluster analysis.* Journal of Climate, 1993. **6**(11)
Fovell, R.G., *Consensus clustering of U.S. temperature and precipitation data.* Journal of Climate, 1997. **10**(6)
Mahlstein, I. and R. Knutti, *Regional climate change patterns identified by cluster analysis.* Climate Dynamics, 2010. **35**(4)
Köppen, W., E. Volken, and S. Brönnimann, *The thermal zones of the Earth according to the duration of hot, moderate and cold periods and to the impact of heat on the organic world.* Meteorologische Zeitschrift, 2011. **20**(3)
O'Neill, M.S. and K.L. Ebi, *Temperature extremes and health: Impacts of climate variability and change in the United States.* Journal of Occupational and Environmental Medicine, 2009. **51**(1).
Rubel, F. and M. Kottek, *Comments on: The thermal zones of the Earth'by Wladimir Köppen (1884).* Meteorologische Zeitschrift, 2011. **20**(3)
Thornthwaite, C.W., *An approach toward a rational classification of climate.* Geogr. Review, 1948. **38**
Wilks, D.S., *Cluster Analysis*, in *Statistical Methods in the atmospheric sciences,* 2011, Academic Press: USA
Tools and Data: ENVI 4.7; ESRI ArcMap 10.0; MATLAB 7.13.0.564, MS SQL Server 2010, CMS Dataset (1991-2006), MODIS 13C v.5.0