Constructing Political Region Agglomerations for Effective Science Communication

James Rising

Ciriacy-Wantrup Postdoctoral Fellow, UC Berkeley

jrising@berkeley.edu, http://existencia.org/pro

Motivation

Often political regions form a better scale for analysis than regular grids, since they are used for collecting socioeconomic data and are more relevant to stakeholders and policy-makers. However, different political regions can enclose very different areas, populations, ranges of climatic variability. Also, large-scale analyses can struggle to balance resolution with computational feasibility. When **no available political division offers comparable and manageable regions**, it is necessary to combine regions to satisfy these concerns.

We provide a solution to this problem, by providing **a general system for agglomerating regions** to a larger scale, to approximately optimize arbitrary objectives. These regions provide an intermediate scale for modeling and greater comparability than unagglomerated regions.

Regions for the Global Climate Prospectus

The **Global Climate Prospectus (GCP)** is a project to estimate the costs of climate change from the empirical literature. The project aims to calculate impacts at a fairly high political resolution: more than country-level, but less than the 295 000 administrative regions in GADM 2.8. For this analysis, we attempt to develop **20 000 regions**.

We first allot region targets to each country, based on population density and climatic variability.

The population weighted target is $20000 P_i / \sum_i P_i$, for populations P_i. The climate weighted target is $20000 A_i V_i / \sum_i A_i V_i$ for areas A_i and $V_i = \frac{Var[T_i]}{E[Var[T]]} + \frac{Var[D_i]}{E[Var[D]]} + \frac{Var[P_i]}{E[Var[P]]} + \frac{Var[Q_i]}{E[Var[Q]]}$ where T_i is mean temperature, $E[T] = 8\,°C$, D_i is diurnal temperature range, $E[D] = 2.1\,°C$, P_i is precipitation in the wettest month, $E[P] = 250\,mm$, Q_i is precipitation in the driest month, and $E[Q] = 26\,mm$ (from Bioclim). The final target is the average of the population and climate weighted targets.

The target regions, relative to the available administrative levels, are shown below. For most countries, there is no available administrative division for our preferred resolution.

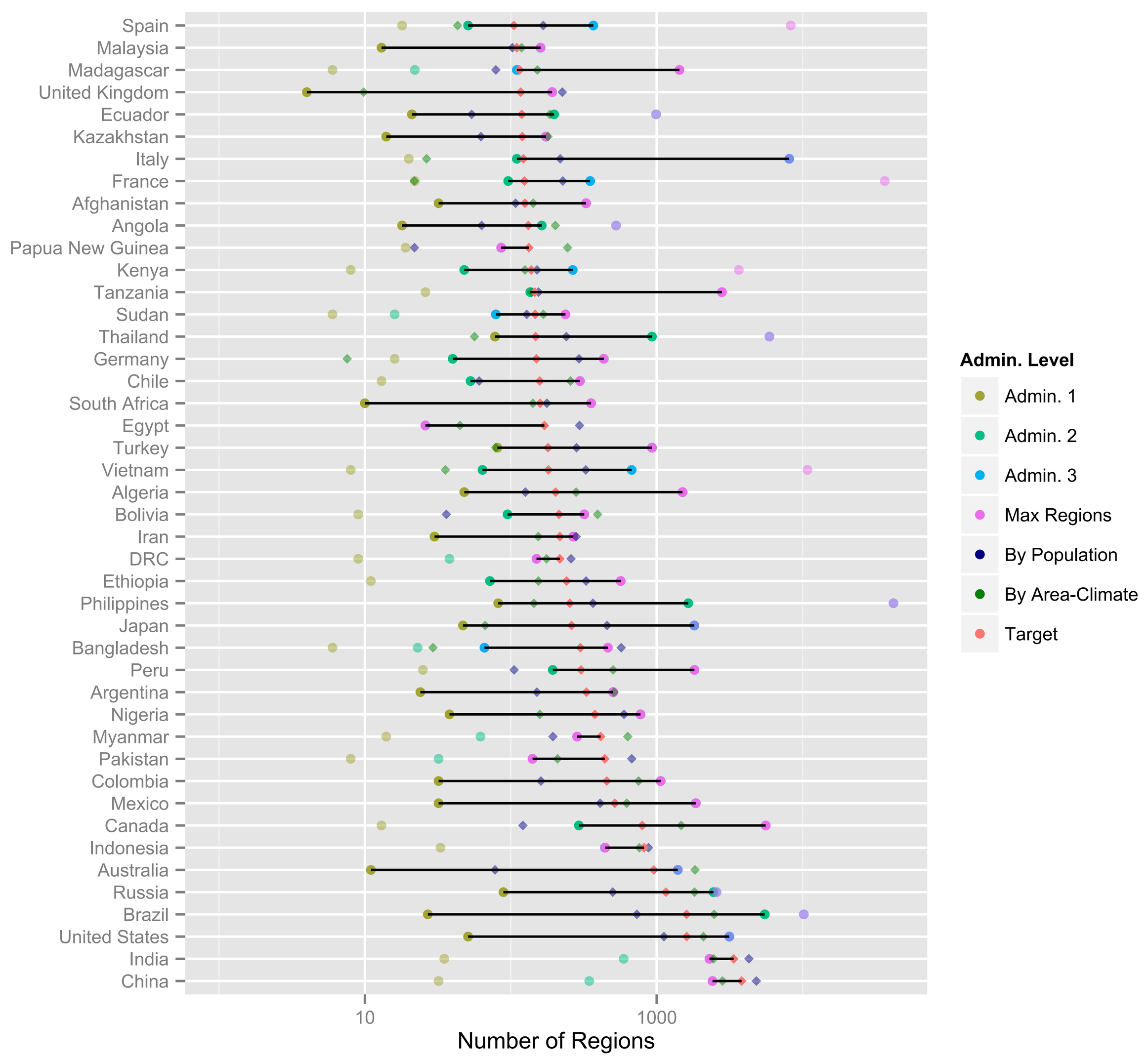


Figure 1: Countries with over 1000 target regions ♦, based on their population ♦ and climatic ♦ targets. The range between administrative region counts above and below these targets are shown in black.

When is agglomeration needed?

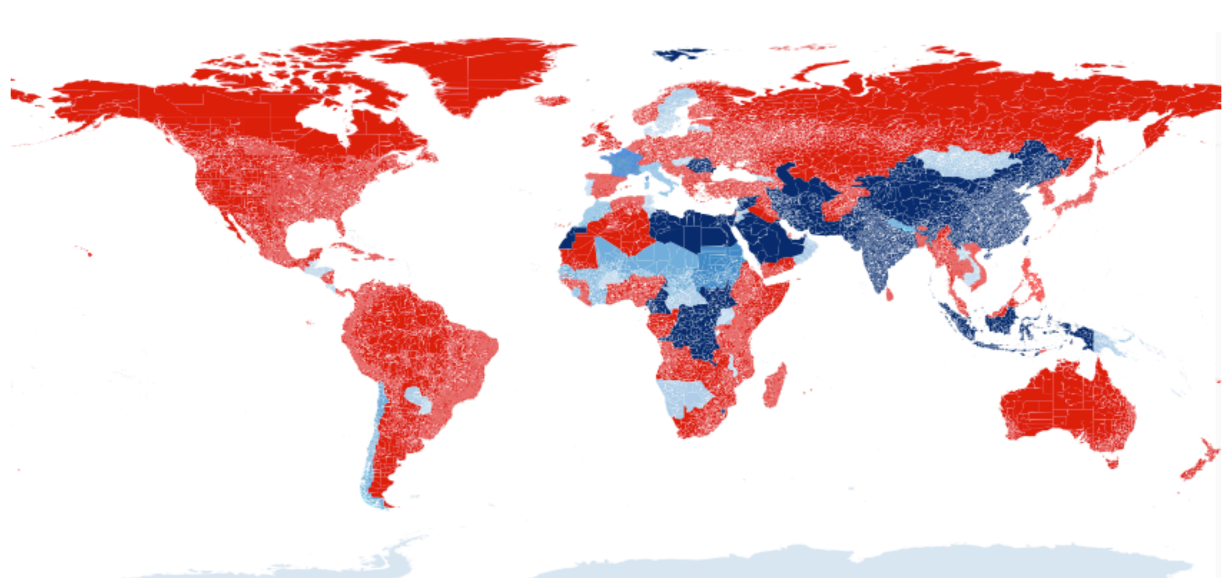


Figure 2: Let C be a country's target region count. Countries in dark blue have $C \geq$ Max Regions; lightest blue have $C \leq 1$; other shades of blue have an administrative region with $C/2 \leq N \leq C$. All others (red) need agglomeration.

Redrawing the United States

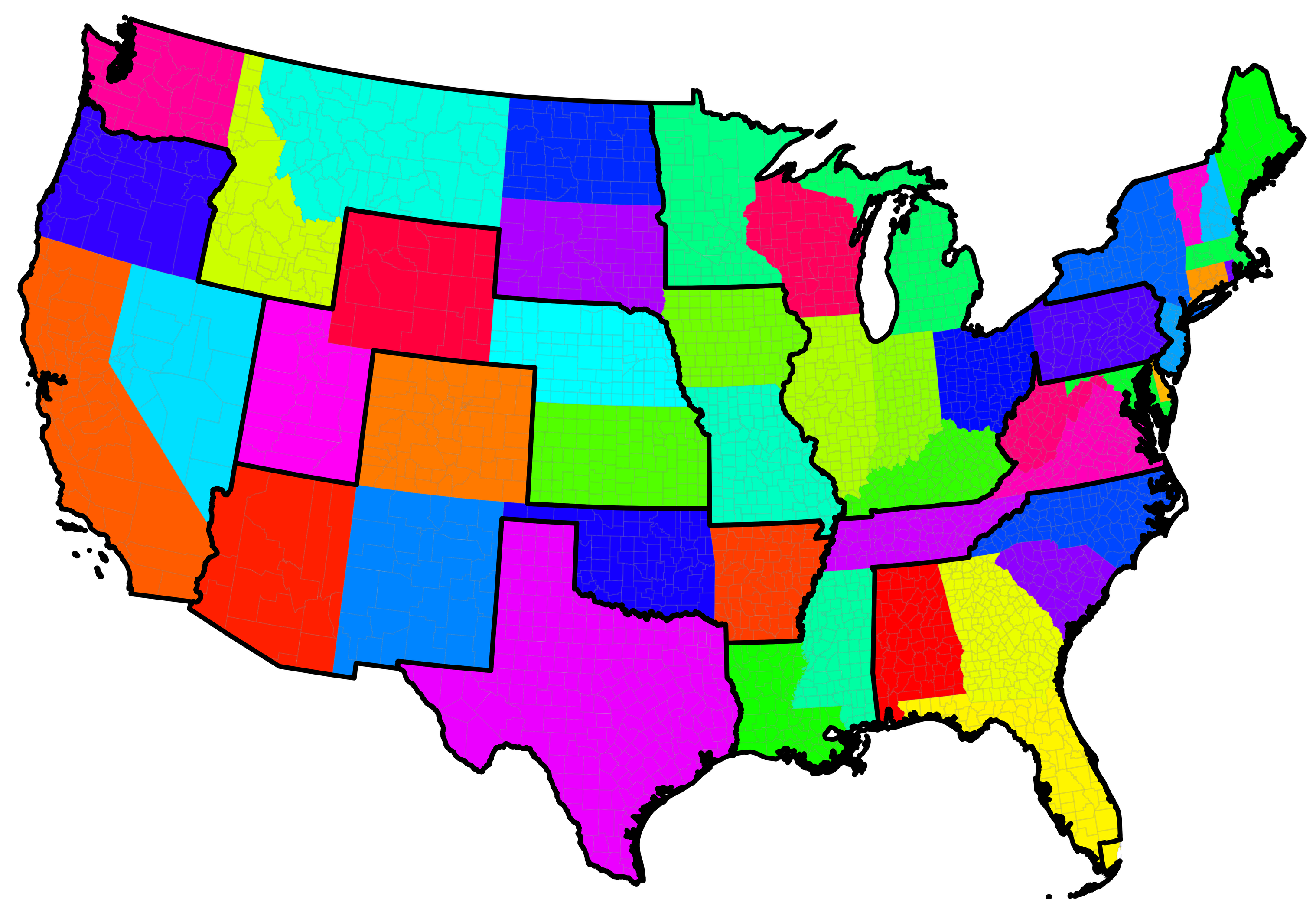
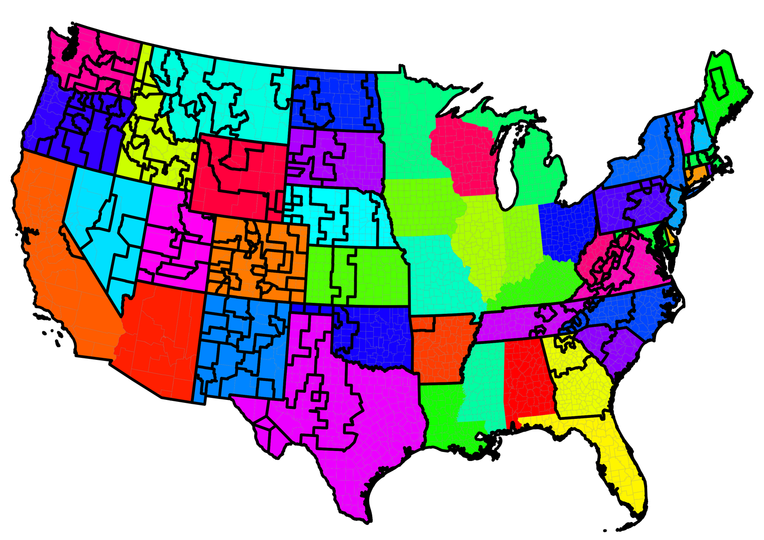
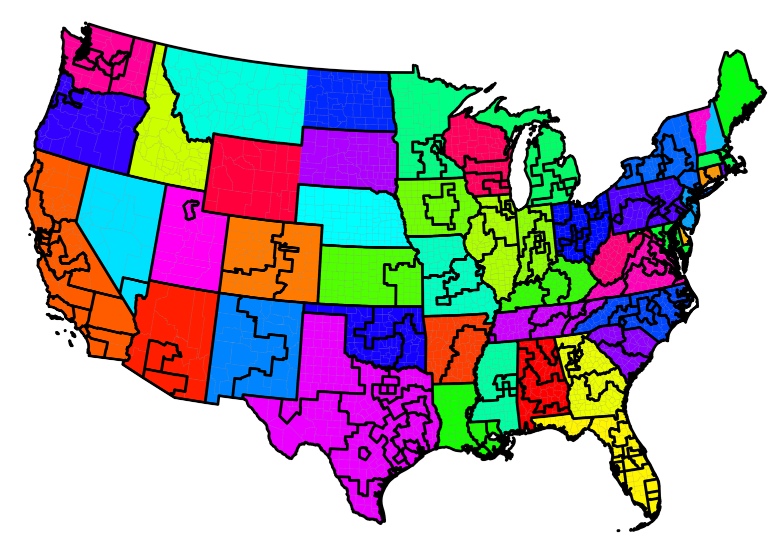


Figure 3: Using the standard agglomeration configuration on right, as does the world map below.

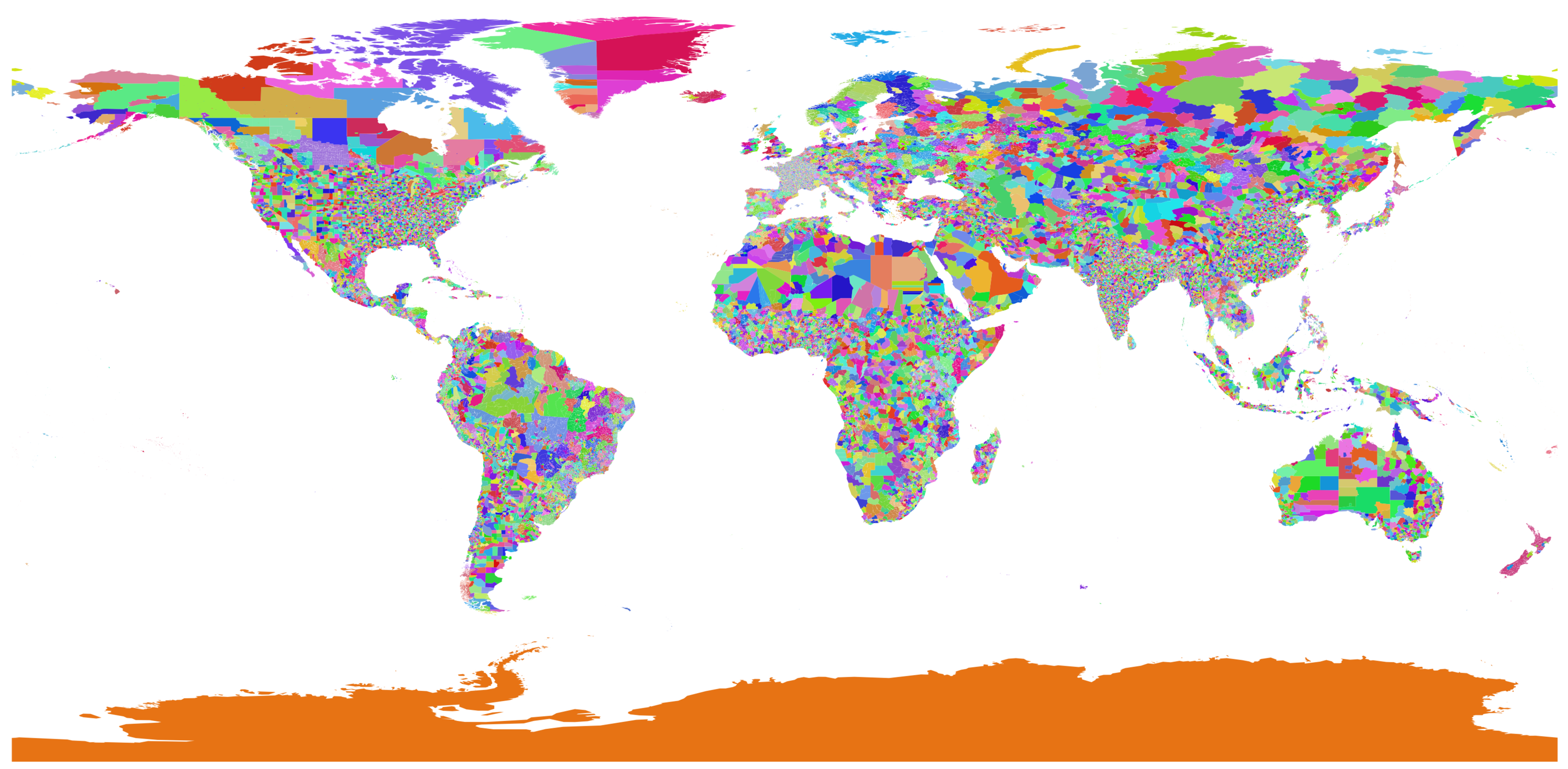
Elevation-only criteria



Population-only criteria



Redrawing the world



Country	Regions	Agglomerated	Type
China	2413	2413 (3760)	each
India	2300	2300 (3233)	each
United States	3150	3150 (*)	each
Russia	2574	500 (708)	
Brazil	10195	500 (714)	
Indonesia	444	444 (687)	each
Bangladesh	463	400 (425)	
Nigeria	775	400 (460)	
Japan	1813	300 (345)	

Country	Regions	Agglomerated	Type
Mexico	1853	300 (343)	
Iran	268	268 (341)	each
Canada	5589	200 (277)	
Ethiopia	567	200 (263)	
Vietnam	10805	200 (247)	
Australia	1396	200 (216)	
Philippines	42009	200 (276)	
Germany	434	200 (224)	
Thailand	5926	180 (188)	

A spatial agglomeration algorithm

The method requires two components specific to the region and values to be balanced: a **Agglomeration Proposer**, which suggests incremental possible agglomerations, and a **Region Evaluator**, which determines the entropy of a region or agglomeration. The algorithm iteratively chooses the agglomeration that increase entropy the least.

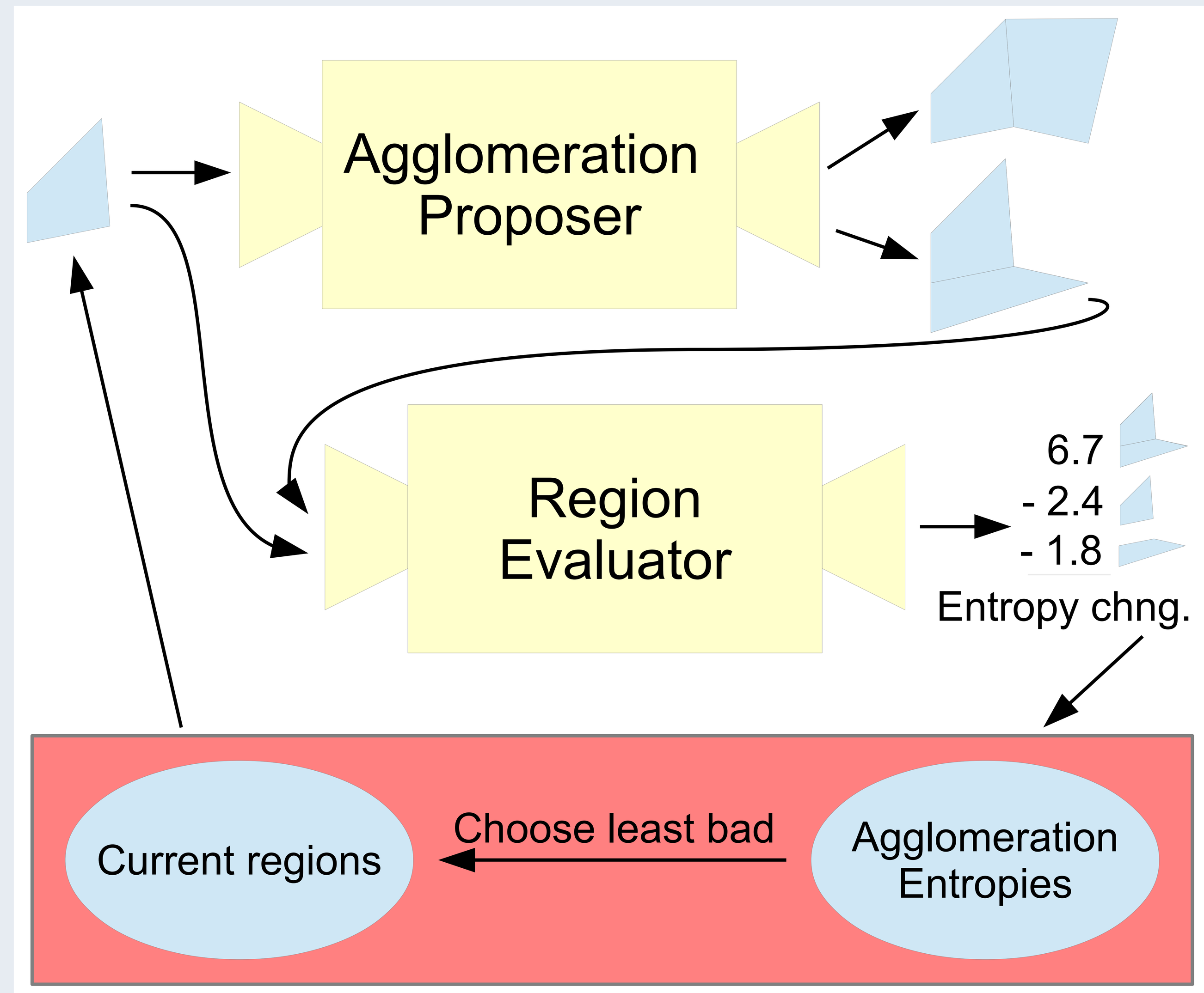


Figure 4: Diagram of the agglomeration algorithm.

Standard Region Properties

(Parenthetical variables are used in the Region Evaluator below)

- Containing regions ($\# = M$)
- Neighbors ($\# = N$)
- Containing administrative regions
- Population (P) and area (A)
- Socioeconomic and Climatic traits, e.g., income, urban fraction, temperatures, biomes ($\{T\}$)
- Containing region centroids (Lat, Lon)

Standard Agglomeration Proposer

Given region R within a containing administrative region S, proposes:

1. The combination of R with each of its neighbors within S.
2. The next higher administrative region, S.
3. If neither of the above is available (e.g., an island state), the combination of R and the closest also at the first administrative level.

Standard Region Evaluator

Let R_i contain subregions indexed by j. Combine scaled scores for the following:

Area $-(\sum_j A_j)^2$

Population $-(\sum_j P_j)^2$

Dispersion $-Var[Lat] - Var[Lon \cos E[Lat]]$

Other traits $-\sum_T Var[T_i] / E[Var[T]]$

Circumference $-M^N / 6\sqrt{M}$

References

- Global Administrative Areas (2012). GADM database of Global Administrative Areas, version 2.8. [online] URL: www.gadm.org.
- Hijmans, R. J., S. E. Cameron, and J. L. Parra. 'WorldClim version 1.4. Museum of vertebrate zoology.' University of California, Berkeley. Available at: http://www.worldclim.org/.
- Bright, E. A., P. R. Coleman, A. N. Rose, and M. L. Urban (2012). 'LandScan 2011.' Digital dataset, Oakridge National Laboratory, Oakridge, TN, USA, web.ornl.gov/sci/landscan/index.shtml.

Thanks to Solomon Hsiang and Bob Kopp and the whole GCP team!