Evolution Strategies for Optimizing Rectangular Cartograms

Kevin Buchin[1], Bettina Speckmann[1]*, and Sander Verdonschot[2]

[1] Department of Mathematics and Computing Science, TU Eindhoven, Eindhoven, The Netherlands
k.a.buchin@tue.nl, speckman@win.tue.nl
[2] School of Computer Science, Carleton University, Ottawa, Canada
sander@cg.scs.carleton.ca

Abstract. A rectangular cartogram is a type of map where every region is a rectangle. The size of the rectangles is chosen such that their areas represent a geographic variable such as population or GDP. In recent years several algorithms for the automated construction of rectangular cartograms have been proposed, some of which are based on rectangular duals of the dual graph of the input map. In this paper we present a new approach to efficiently search within the exponentially large space of all possible rectangular duals. We employ evolution strategies that find rectangular duals which can be used for rectangular cartograms with correct adjacencies and (close to) zero cartographic error. This is a considerable improvement upon previous methods that have to either relax adjacency requirements or deal with larger errors. We present extensive experimental results for a large variety of data sets.

Keywords: Rectangular cartogram, evolution strategy, regular edge labeling.

1 Introduction

Cartograms [3, 13], also called *value-by-area maps*, are a useful and intuitive tool to visualize statistical data about a set of regions like countries, states, or counties. The size (area) of a region in a cartogram corresponds to a particular geographic variable. A common variable is population: in a population cartogram, the sizes of the regions are proportional to their population. The sizes of the regions in a cartogram are not the true sizes and hence the regions

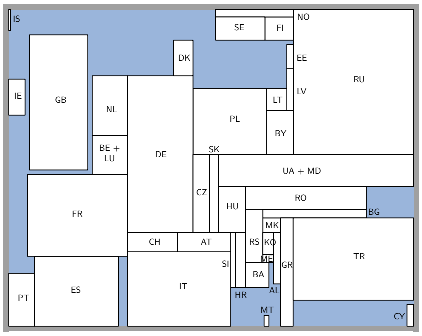


Fig. 1. The population of Europe 2011.

* Supported by the Netherlands Organisation for Scientific Research (NWO) under project no. 639.022.707.

generally cannot keep both their shape and their adjacencies. A good cartogram, however, preserves the recognizability in some way.

Globally speaking, there are four types of cartogram. The standard type—also referred to as contiguous area cartogram—has deformed regions so that the desired sizes can be obtained and the adjacencies kept. The most prominent algorithm for such cartograms was developed by Gastner and Newman [8]. The second type of cartogram is the non-contiguous area cartogram [14]. The regions have the true shape, but are scaled down and generally do not touch anymore. Sometimes the scaled-down regions are shown on top of the original regions. A third type of cartogram was introduced by Dorling [4] and is in its original form based on circles. Dorling cartograms maintain neither correct adjacencies between regions nor correct relative positions. A variant of Dorling cartograms are Demers cartograms which use squares instead of circles. Demers cartograms also do not maintain correct adjacencies and disturb relative positions even more than Dorling cartograms. We concentrate on a fourth type of cartograms, *rectangular cartograms*, as introduced by Raisz in 1934 [15], where each region is represented by a rectangle and adjacencies are maintained as well as possible.

Quality criteria. Whether a rectangular cartogram is good is determined by several factors. One of these is the *cartographic error* [5], which is defined for each region as $|A_c - A_s| / A_s$, where A_c is the area of the region in the cartogram and A_s is the specified area of that region, given by the geographic variable to be shown. Another factor are the *correct adjacencies* of the regions of the cartogram. This requires that the dual graph of the cartogram is the same as the dual graph of the original map. Here the *dual graph* of a map—also referred to as *adjacency graph*—is the graph that has one node per region and connects two regions if they are adjacent, where two regions are considered to be adjacent if they share a 1-dimensional part of their boundaries (see Fig. 3). A third factor is important for the recognizability of a rectangular cartogram: the *relative position* of the rectangles. For example, a rectangle representing the Netherlands should lie west of a rectangle representing Germany. To measure how well a cartogram matches the spatial relations between regions in the input map we use the *bounding box separation distance* (BBSD) [2], which is defined in the next section. Finally, it is important that the *aspect ratio* of the rectangles does not exceed a certain maximum since otherwise the areas become difficult to judge.

Rectangular duals. We follow the general approach set out in previous work [2, 16, 18] and construct rectangular cartograms by first finding a suitable *rectangular dual* of the dual graph of the input map. A rectangular dual is defined as follows. A *rectangular partition* of a rectangle R is a partition of R into a set $\mathcal{R}$ of non-overlapping rectangles such that no four rectangles in $\mathcal{R}$ meet at one common point. A *rectangular dual* of a plane graph G is a rectangular partition $\mathcal{R}$, such that (i) there is a one-to-one correspondence between the rectangles in $\mathcal{R}$ and the nodes in G; (ii) two rectangles in $\mathcal{R}$ share a common boundary if and only if the corresponding nodes in G are connected.

Not every plane graph has a rectangular dual. A plane graph G has a rectangular dual $\mathcal{R}$ with four rectangles on the boundary of $\mathcal{R}$ if G is an *irreducible*

Fig. 2. Two rectangular duals of the dual graph of a map of Europe (from [2]).

triangulation: (*i*) G is triangulated and the exterior face is a quadrangle; (*ii*) G has no separating triangles (a 3-cycle with vertices both inside and outside the cycle) [1, 12]. A plane triangulated graph G has a rectangular dual if and only if we can augment G with four external vertices such that the augmented graph is an irreducible triangulation.

The dual graph F of a typical geographic map can be easily turned into an irreducible triangulation in a preprocessing step. F is in most cases already triangulated. We triangulate any remaining non-triangular faces (for example the face formed by the nodes for Colorado, Utah, New Mexico, and Arizona). It remains to preprocess internal nodes of degree less than four, such as Luxembourg, Moldova, or Lesotho. In these cases, we add the region to one of its neighbors.

A rectangular dual is not necessarily unique. Consider the two rectangular duals of the dual graph G of a map of Europe shown in Fig. 2. To ensure that G is an irreducible triangulation, Luxembourg and Moldova have been removed. Furthermore, "sea regions" have been added to improve the shape of the outline. The left dual will lead to a recognizable cartogram, whereas the right dual (with France east of Germany and Hungary north of Austria) is useless as basis for a cartogram. Most irreducible triangulations have in fact exponentially many different rectangular duals which are described by *regular edge labelings*.

Regular edge labelings. The equivalence classes of the rectangular duals of an irreducible triangulation G correspond one-to-one to the *regular edge labelings* (RELs) of G. An REL of an irreducible triangulation G is a partition of the interior edges of G into two subsets of red and blue directed edges such that: (*i*) around each inner vertex in clockwise order we have four contiguous sets of incoming blue edges, outgoing red edges, outgoing blue edges, and incoming red edges; (*ii*) the left exterior vertex has only blue outgoing edges, the top exterior vertex has only red incoming edges, the right exterior vertex has only blue incoming edges, and the bottom exterior vertex has only red outgoing edges (see Fig. 3, red edges are dashed). Kant and He [11] show how to find a regular edge labeling and construct the corresponding rectangular dual in linear time.

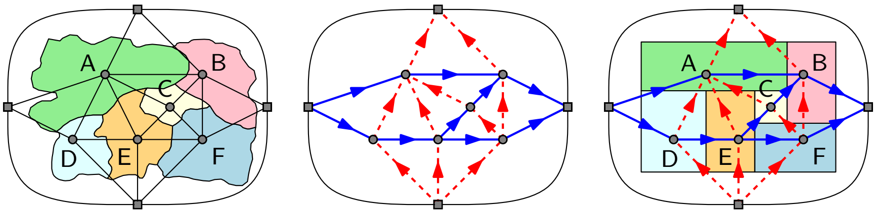
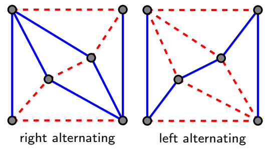


Fig. 3. A subdivision and its augmented dual graph G, a regular edge labeling of G, and a corresponding rectangular dual (from [2]).



An *alternating 4-cycle* is an undirected 4-cycle in which the colors of the edges alternate between red and blue. There are two kinds of alternating 4-cycles, depending on the color of the interior edges incident to the cycle. If these are the same color as the next clockwise cycle edge the cycle is *right alternating*, otherwise it is *left alternating*. Fusy [7] proved that the set of RELs of a fixed irreducible triangulation form a distributive lattice. The flip operation consists of switching the edge colors inside a right alternating 4-cycle, turning it into a left alternating 4-cycle. An REL with no right alternating 4-cycle is called *minimal*; it is at the bottom of the distributive lattice.

Although an irreducible triangulation can have exponentially many RELs and hence exponentially many rectangular duals this does not imply that an error free cartogram for this graph exists. The area specification for every rectangle, as well as other criteria for good cartograms, may make it impossible to realize. The lattice structure of the RELs allows us to traverse the space of all RELs for a given graph and find the best rectangular dual for a given set of input values to be realized. However, already for small graphs it is unfeasible to test all possible rectangular duals: the dual graphs of the countries of Europe or the contiguous states of the US both have over four billion labelings. This calls for search strategies that efficiently explore a significant part of the lattice structure. In this paper we present a new search algorithm based on evolution strategies which clearly outperforms previous approaches.

Related work. The only algorithm for standard cartograms that can be adapted to handle rectangular cartograms is Tobler's pseudo-cartogram algorithm [17] combined with a rectangular dual algorithm. However, Tobler's method is known to produce a large cartographic error and is mostly used as a preprocessing step for cartogram construction [13]. The first method for the automated construction of rectangular cartograms was presented by Van Kreveld and Speckmann [18]. Their cartograms have small cartographic error but require (mildly) disturbed adjacencies to realize most data sets. Their method searches through a comparatively small user-specified subset of the RELs. Every labeling in this subset is considered acceptable with respect to the relative positions of the countries. Speckmann *et al.* [16] improved on their earlier results by using an iterative linear programming method to build a cartogram from an REL. With this methodology world maps could be realized, although small disturbances in the adjacencies were still necessary to obtain acceptable cartographic errors. Speckmann *et al.* [16] used the same user-specified subset of the RELs as Van Kreveld and Speckmann [18]. In a recent paper [2] we presented the first method which uses a heuristic search strategy, namely simulated annealing, on the complete lattice of RELs. We restricted ourselves solely to cartograms with correct adjacencies and nevertheless improved upon the cartographic error of the resulting maps.

A different approach was taken by Inoue *et al.* [10] who compute rectangular and rectilinear cartograms by triangulating the regions and transforming the triangles to meet the desired area requirements. Their rectilinear cartograms have high region complexity and their rectangular cartograms exhibit large cartographic errors. Finally, Heilmann *et al.* [9] gave an algorithm that always produces regions with the correct areas; but the adjacencies can be disturbed badly.

Results and organization. In this paper we show how to employ evolution strategies to search effectively in the exponentially large lattice of RELs. We find rectangular duals that allow us to realize rectangular cartograms with correct adjacencies and (close to) zero cartographic error. This is a considerable improvement over previous methods. In Section 2 we describe our evolution strategies and in Section 3 we present and discuss an extensive set of experiments.

2 Evolution strategies

The dual graph of a map can have an exponential number of valid RELs, hence we turn to meta-heuristics to find good solutions in this huge search space. In this section, we present a new approach based on evolution strategies that performs significantly better than our previous method based on simulated annealing [2].

Evolution strategies are an optimization technique that is heavily inspired by natural selection. They use a population of candidate solutions, from which the next generation is constructed by selecting promising individuals and mutating these. If the population is initialized with random solutions, this leads to a broad initial search that quickly focuses on promising regions of the search space. The individuals for our problem consist of valid RELs of the augmented dual graph of our input map. The validity requirement is important, as it reduces the search space by an exponential factor. The population is initialized with semi-random individuals, by starting at the minimum labeling and flipping $d\left(\frac{1}{2}+\frac{r}{8}\right)$ random left alternating 4-cycles, where d is the diameter of the lattice and r is a standard normal distributed random number. Since the lattice of RELs is distributive, every upward path between the same two RELs has the same length and therefore the diameter is simply the number of left alternating 4-cycles we need to flip until we reach the maximum labeling from the minimum labeling. We compute the minimum labeling using a linear-time algorithm by Fusy [6].

After this initialization, every generation follows the same three steps:

1. Compute the fitness scores of all individuals. If the quality measure gives a higher score to better individuals, use this score directly, otherwise (as is

the case with cartographic error and bounding box separation distance) use $1/m$, where m is the score given by the measure.
2. Copy the best 4% of the current population to the next generation. This ensures that the best solutions stay in the population unmodified.
3. Fill the remainder of the next generation by repeating the following process:
   - Use rank selection to select an individual from the current population. The individuals are sorted by fitness in decreasing order. Each individual is assigned a score of 0.9^i, where i is the individual's rank, so the best individual gets a score of 0.9 and so on. Then each individual is selected with probability equal to the proportion of their score to the total score. Since the selection depends only on the rank of the individuals and not on the fitness values themselves, it is a good choice for optimization using user-specified fitness measures.
   - With probability 0.05, generate a standard normal distributed random number r. If r is positive, move $\frac{dr}{6}$ steps up the lattice, by flipping random left alternating 4-cycles. If r is negative, move $\frac{dr}{6}$ steps down the lattice, by flipping random right alternating 4-cycles. This is a drastic mutation that is used to keep the population from stagnating too much.
   - With probability 0.9, flip a random alternating 4-cycle. This is a small mutation, used for local exploration of the neighbourhood of the selected individual.

Finally, the best REL found during the process is returned. The parameter values presented can be slightly changed to increase performance on various maps and quality measures, but the presented values were found to work well for our instances.

Quality measures. We now explain how we capture the quality criteria for rectangular cartograms in our evolution strategy. To create a cartogram from an REL we follow the iterative linear programming method presented in [16] with correct adjacencies. Since we consider only valid RELs of the dual graphs of our input maps, this implies that all cartograms we generate have correct adjacencies. That is, all regions that share borders on the geographic input map will do so in the cartogram and regions that do not share borders will not be adjacent in the cartogram. Furthermore, we bound the aspect ratio of all rectangles by 12.

To make a rectangular cartogram as recognizable as possible, it is important that the directions of adjacency between the rectangles of the cartogram follow the spatial relation of the regions of the geographical map as closely as possible. Since these directions of adjacency are specified by the REL, we can assess the recognizability of a rectangular cartogram by looking at its REL. We use the *bounding box separation distance* (BBSD) [2] to quantify how well the directions of adjacency match the geographical directions. The BBSD measures the distance the bounding boxes of the regions would need to be moved to separate them in the direction indicated by the edge label (see Fig. 4).

Finally, to compute the fitness score of an individual we used the weighted sum of 0.7 times the average of squared cartographic errors and 0.3 times the average of squared bounding box separation distances of its regions.

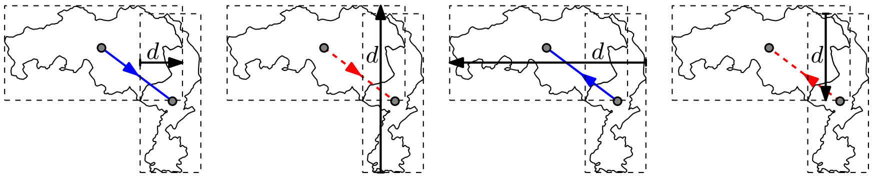


Fig. 4. The BBSD measures the distance d which the bounding boxes of the regions need to be moved to separate them in the direction indicated by the edge label (arrow).

3 Experimental Results

We evaluated our method on a large variety of data sets. For each data set, we measured the cartographic error, bounding box separation distance and running time. We generated cartograms based on three different geographical maps: the contiguous states of the US, the countries of Europe and the countries of the world with a population over 1 million. For the US we used data from the US Census Bureau State and County quickfacts[3]. Since cartograms can not easily represent negative or zero values, we used all 45 data sets from the 2010 census where each state was assigned a positive value. Additionally we used the results of the US presidential election of 2008[4]. For Europe we used data from the ranked CIA World Fact Book data sets[5]. We used all 19 ranked WFB data sets that have data for all countries of Europe included in our cartograms[6]. Our final cartogram uses the world population data from Worldmapper[7]. We conclude with a direct comparison with our previous method [2].

We generated 20 cartograms for each data set. For each run we recorded the average cartographic error, the maximum cartographic error, and the bounding box separation distance. We summarized these results by taking the average, minimum and maximum over all runs per data set in Table 1. For the US census data we included only the population and geography data sets in the summary, the other data sets show similar trends. The columns 'min' give the average cartographic error, maximum cartographic error and the bounding box separation distance of the best cartograms generated for the data set. Since we need only one cartogram per data set, we focus on the values in the 'min' columns.

The rectangular cartograms in the figures have regions that are colored based on their error. Shades of red show that a region is too small and shades of blue

---

[3] `http://quickfacts.census.gov/qfd/index.html`, accessed 2011/11/22.

[4] `http://elections.nytimes.com/2008/results/president/votes.html`, accessed 2012/02/06.

[5] `https://www.cia.gov/library/publications/the-world-factbook/index.html`, accessed 2011/12/10.

[6] For the area cartogram we use the area of Russia within Europe, `http;//en.wikipedia.org/wiki/European_Russia`, accessed 2012/02/06.

[7] `http://www.worldmapper.org/data/nomap/2_worldmapper_data.xls`, accessed 2012/02/01.

Table 1. Average cartographic error (ACE), maximum cartographic error (MCE) and average squared bounding box separation distance (BBSD) for 2010 US census data (people + geography) and World Factbook data of Europe. Average (avg), minimum (min) and maximum (max) taken over 20 runs of our algorithm.

data set description	**ACE**			**MCE**			**BBSD**		
	avg	**min**	**max**	**avg**	**min**	**max**	**avg**	**min**	**max**
US census data 2010									
Resident total population	0.04	0.01	0.08	0.29	0.02	0.76	0.05	0.03	0.10
Resident population (RP) 2000	0.05	0.01	0.14	0.34	0.06	0.74	0.05	0.03	0.10
RP $<$ 5 years, percentage (%)	0.04	0.00	0.09	0.16	0.02	0.34	0.04	0.02	0.06
RP $<$ 18 years, %	0.03	0.02	0.07	0.18	0.06	0.24	0.05	0.02	0.13
RP $\geq$ 65 years, %	0.04	0.01	0.08	0.18	0.04	0.43	0.05	0.02	0.10
RP: total females, %	0.03	0.00	0.05	0.16	0.00	0.32	0.04	0.02	0.07
RP: White alone, %	0.04	0.00	0.08	0.17	0.02	0.40	0.05	0.03	0.07
RP: Black alone, %	0.06	0.01	0.13	0.36	0.05	0.70	0.05	0.03	0.09
RP: Amer. Indian + Alaska Na., %	0.07	0.02	0.14	0.44	0.21	0.89	0.04	0.03	0.07
RP: Asian alone, %	0.05	0.00	0.11	0.32	0.02	0.73	0.05	0.02	0.09
RP: Two or more races, %	0.03	0.00	0.06	0.15	0.00	0.31	0.06	0.03	0.09
RP: Hispanic or Latino Origin, %	0.05	0.00	0.10	0.27	0.02	0.82	0.05	0.03	0.09
RP: Not Hisp., White alone, %	0.04	0.00	0.08	0.19	0.00	0.49	0.05	0.03	0.13
Same househ. 1 yr ago, % '05–'09	0.04	0.00	0.06	0.16	0.00	0.28	0.04	0.02	0.08
Pl. of birth, foreign born,% '05–'09	0.06	0.01	0.10	0.31	0.04	0.77	0.05	0.03	0.09
Pop. $\geq$ 5 yrs, % lang. other '05–'09	0.04	0.00	0.08	0.22	0.00	0.55	0.05	0.03	0.08
$\geq$ 25 yrs % high sch. grad. '05–'09	0.03	0.00	0.06	0.15	0.00	0.25	0.04	0.02	0.10
$\geq$ 25 yrs % bachelor's deg. '05–'09	0.04	0.00	0.08	0.21	0.00	0.48	0.05	0.02	0.12
Veterans - total '05-'09	0.03	0.00	0.08	0.14	0.01	0.45	0.04	0.03	0.09
Land area in square miles	0.00	0.00	0.01	0.01	0.00	0.04	0.02	0.02	0.04
Population per square mile	0.08	0.02	0.14	0.65	0.11	1.00	0.05	0.03	0.12
World Factbook: Europe (Dec. 2011)									
GDP (purchasing power parity)	0.00	0.00	0.01	0.01	0.00	0.07	0.08	0.08	0.09
GDP real growth rate	0.11	0.09	0.14	0.60	0.50	1.00	0.09	0.08	0.10
GDP - per capita (PPP)	0.07	0.03	0.10	0.34	0.08	0.67	0.09	0.07	0.11
Electricity - production	0.00	0.00	0.01	0.03	0.00	0.11	0.08	0.07	0.10
Electricity - consumption	0.00	0.00	0.01	0.02	0.00	0.15	0.08	0.07	0.10
Airports	0.00	0.00	0.01	0.02	0.00	0.08	0.08	0.07	0.10
Exports	0.01	0.00	0.04	0.03	0.00	0.25	0.08	0.08	0.09
Roadways	0.01	0.00	0.02	0.03	0.00	0.11	0.09	0.08	0.09
Imports	0.00	0.00	0.02	0.01	0.00	0.08	0.09	0.08	0.10
Inflation rate (consumer prices)	0.05	0.01	0.11	0.30	0.09	0.54	0.09	0.08	0.11
Labor force	0.00	0.00	0.01	0.01	0.00	0.04	0.08	0.08	0.10
Population	0.00	0.00	0.01	0.01	0.00	0.11	0.09	0.08	0.10
Unemployment rate	0.10	0.03	0.18	0.54	0.16	0.94	0.09	0.07	0.10
Area	0.00	0.00	0.01	0.02	0.00	0.05	0.08	0.07	0.09
Telephones - main lines in use	0.00	0.00	0.00	0.01	0.00	0.03	0.08	0.08	0.10
Telephones - mobile cellular	0.00	0.00	0.01	0.01	0.00	0.08	0.08	0.07	0.11
Distr. of family income - Gini Ind.	0.03	0.00	0.05	0.22	0.00	0.65	0.09	0.07	0.09
Current account balance	0.05	0.03	0.11	0.49	0.20	0.81	0.09	0.08	0.10
Commercial bank prime lend. rate	0.05	0.02	0.07	0.29	0.10	0.43	0.09	0.08	0.10

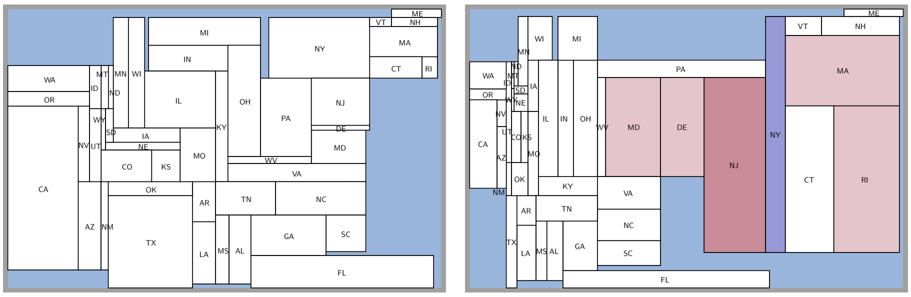


Fig. 5. US population (left) and US population per square mile (right).

show that a region is too large. If the error is below 0.05, the region is white. Note that only Fig. 5 (right) has non-white regions.

All code was written in Java and executed single-threaded, using the Open-JDK Runtime Environment IcedTea6 1.9.9, corresponding to java version 1.6.0_20. For solving linear programs we used IBM ILOG CPLEX 12.0. The measurements of the running time were performed on a 64-bit quad core 1.86GHz Intel Xeon E5320 server with 8 GB RAM, running Ubuntu 10.04.3. On average it took 476 seconds to generate a cartogram for the US, 354 seconds for Europe and 207 minutes for the world. Since the running times showed little variation between data sets, we do not discuss them further.

For all data sets from the US census in the table our algorithm generated at least one map with average cartographic error (ACE) of 2% or less. The average ACE over all runs of the algorithm is between 0% and 8%, where *land area in square miles* has the lowest average and *population per square mile* the highest. For all except two data sets (percentage of American Indian and Alaska Native population and population per square mile) our algorithm generated maps with a maximum cartographic error (MCE) of at most 6% (and an average over all runs of at most 36%). The bounding box separation distance (BBSD) does not vary much and the average over all runs was between 0.02 and 0.06 depending on the data set. For the data sets not included in the table the results are similar, except that there is one data set with a minimum ACE of 4% (wholesale trade) and 4 data sets with a minimum MCE above 7% (Hispanic-owned firms, manufacturing, wholesale trade, and accommodation and food services).

Our rectangular cartogram of the US population in Fig. 5 (left) has an ACE of 0.5%, a MCE of 2.2%, and a BBSD of 0.365. Our results considerably improve on previous work: Van Kreveld and Speckmann [18] obtained a cartogram with an ACE of 8.6% and a MCE of 87.3%, Buchin et al. [2] one with an ACE of 10.2% and a MCE of 59.7%. Inoue et al. [10] don't report on these errors specifically but obtain a rectangular cartogram in which 22 states have a cartographic error between 5% and 20%, and 7 states have a cartographic error larger than 20%.

The data set on population per square mile is one of the few data sets where the MCE obtained is still high (above 7%). Our cartogram in Figure 5 (right) has an ACE of 2%, a MCE of 11.3%, and a BBSD of 0.376. In the cartogram

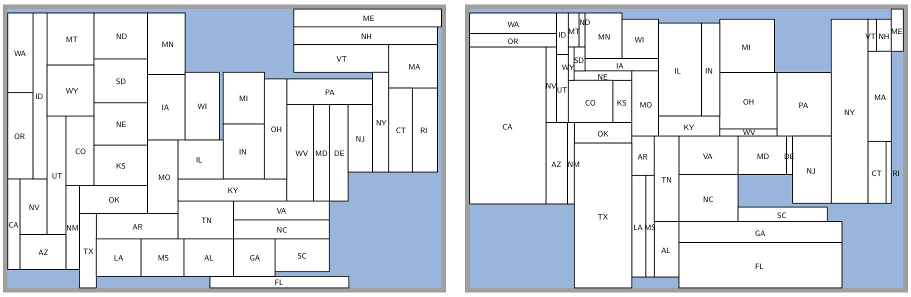
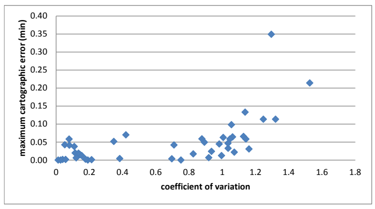


Fig. 6. Percentage of non-Hispanic, white population (left) and number of businesses without payed employees (right). In the left cartogram the correlation to land area is negative, while in the right cartogram the coefficient of variation is high.

we see several causes for the comparably high MCE. In terms of the global layout, the northwest requires so much space (relative to its actual size) that little room is left for the remaining states. The northwest still has not enough space, while the remaining states are depicted with fairly narrow rectangles. More locally, the largest problems seem to be around Pennsylvania, which has to accommodate 4 neighbors with very high population density (and 2 neighbors with lower population density).

In the following we analyze the causes for high MCE further. In terms of the global layout, population density bears several challenges: it is negatively correlated to land area and has a large variation.

Typically cartograms for land area can be generated easily because regions use nearly the same area as on a regular map. It seems natural that data which is negatively correlated to land area is difficult to depict in a cartogram. In our results, however, there does not seem to be a such a relation. Fig. 6 (left) shows a typical cartogram for which land area and the variable depicted have a high negative correlation. The variable is the percentage of non-Hispanic, white population. The cartogram has 0% ACE and MCE, and a BBSD of 0.381.



Generally, high variation in a variable does not necessarily make a variable difficult to depict in a rectangular cartogram. Land area has high variation but can typically be depicted well. Our experiments, however, do indicate a relation between variation and high cartographic error. The scatterplot on the right shows the coefficient of variation (standard deviation divided by mean) for the data sets from the US census plotted against the best MCE error achieved. While the MCE does not seem to change for coefficients up to about 1, beyond that the maximum cartographic error increases considerably. The population density has a coefficient of 1.3. Another data set with a high coefficient is nonemployer businesses (typically self-employed individuals). The coefficient of variation for this data set is 1.2. For this data set we did obtain a cartogram shown in Fig. 6 (right) with low cartographic error. Here the ACE is 0.7%, the MCE 3.1% and the BBSD 0.371.

Our final cartogram of the US is a rectangular cartogram showing the results of the US presidential election of 2008. The area of each state corresponds to the number of electoral votes. States won by the Republicans are depicted in red, while states won by the Democrats are depicted in blue. Note that Nebraska does not have a winner-takes-all system, and therefore is two-colored.

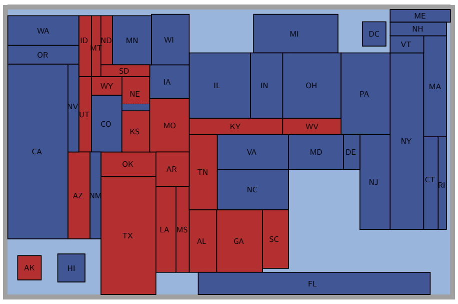


Fig. 7. The US electoral college 2008.

We next turn to the data sets for Europe. To ensure that the dual graph of the map is an irreducible triangulation we joined Luxembourg and Belgium, and Moldova and Ukraine. For most data sets we obtained cartograms without cartographic error, see, for example, the population cartogram on page 1. For 6 data sets, however, the MCE was relatively high, namely between 8% and 50%. This is caused by either unproportionately high or unproportionately low values for the countries in the southeast.

For the maps with very low cartographic error, there is still variation in terms of the layout. Fig. 8 shows two cartograms for European exports. The cartogram on the left-hand side has no cartographic error and a BBSD of 0.088. The cartogram on the right-hand side has ACE 0.2%, MCE 1.6% and a BBSD of 0.078. It seems that it easier to recognize Europe in the cartogram on the right. Hence this cartogram might be preferable despite a small cartographic error.

Our final cartogram shows the world population in 2002. Fig. 9 compares the rectangular cartogram generated by our method to a non-rectangular cartogram from the Worldmapper collection. Our cartogram has ACE 1.17% and MCE 18.5%. Note that we also use a lower percentage of sea area. Overall, recogniz-

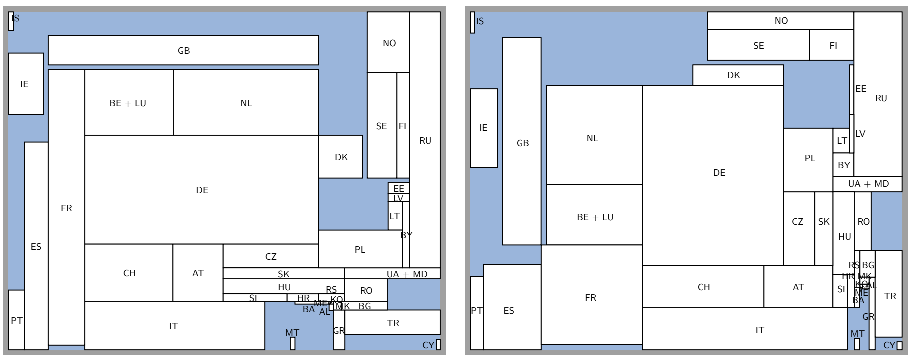


Fig. 8. EU exports: no cartographic error (left) and low cartographic error (right).

Table 2. Evolution strategy vs. simulated annealing approaches. The values are average (avg), minimum (min) and maximum (max) of the average squared bounding box separation distances of the world over 100 runs.

Algorithm	avg	min	max
Simulated annealing	0.101	0.064	0.117
Bootstrapped simulated annealing	0.041	0.019	0.096
Evolution strategy	0.017	0.013	0.025

ability is high for this cartogram, with the most noticeable distortion being the abnormal orientation of Russia. This is a change we noticed in all low-error world population maps. It is unlikely that these orientations would have been considered for a hand-picked set of directions, which demonstrates the clear advantage of searching the entire lattice.

We now compare our previous simulated annealing approach [2] to our new evolution strategy. Both use a probabilistic walk over the lattice of regular edge labelings (RELs) to find good solutions, using the fact that neighbouring labelings are likely to be similar in quality. The largest difference is that the evolution strategy starts many random walks simultaneously and concentrates on the promising ones, while simulated annealing performs a single guided walk.

Results of a comparison are given in Table 2. The goal was to find a REL of the world with a low average squared bounding box separation distance. Simulated annealing was run for 10000 steps, while the evolution strategy was given a population size of 50 with 200 generations, resulting in the same number of fitness evaluations. The original simulated annealing was started at the minimum labeling each time. We also include a bootstrapped version of the simulated annealing approach in the comparison that starts at a random labeling. This random labeling was chosen in the same way as labelings in the initial population of the evolution strategy. The evolution strategy significantly outperforms both simulated annealing versions. Not only is the best REL it finds better than the best RELs found by the simulated annealing versions, its average quality is even better than the best quality found by the others. This is caused mainly by improved reliability, which can be seen from the far smaller range of qualities. The evolution strategy has only a 0.012 difference between the best and worst REL, compared to 0.053 and 0.077 for the simulated annealing versions.

4 Conclusion

We presented a new method based on evolution strategies for generating rectangular cartograms with correct adjacencies. The resulting cartograms—for a large range of data sets for Europe, US, and the world—have (close to) zero cartographic error and high visual quality. This is a considerable improvement over previous methods. Nevertheless, several challenges remain. Data sets with extremely high variability still prove difficult to realize as cartograms with correct adjacencies, low error, and reasonable aspect ratio. Generally speaking, we

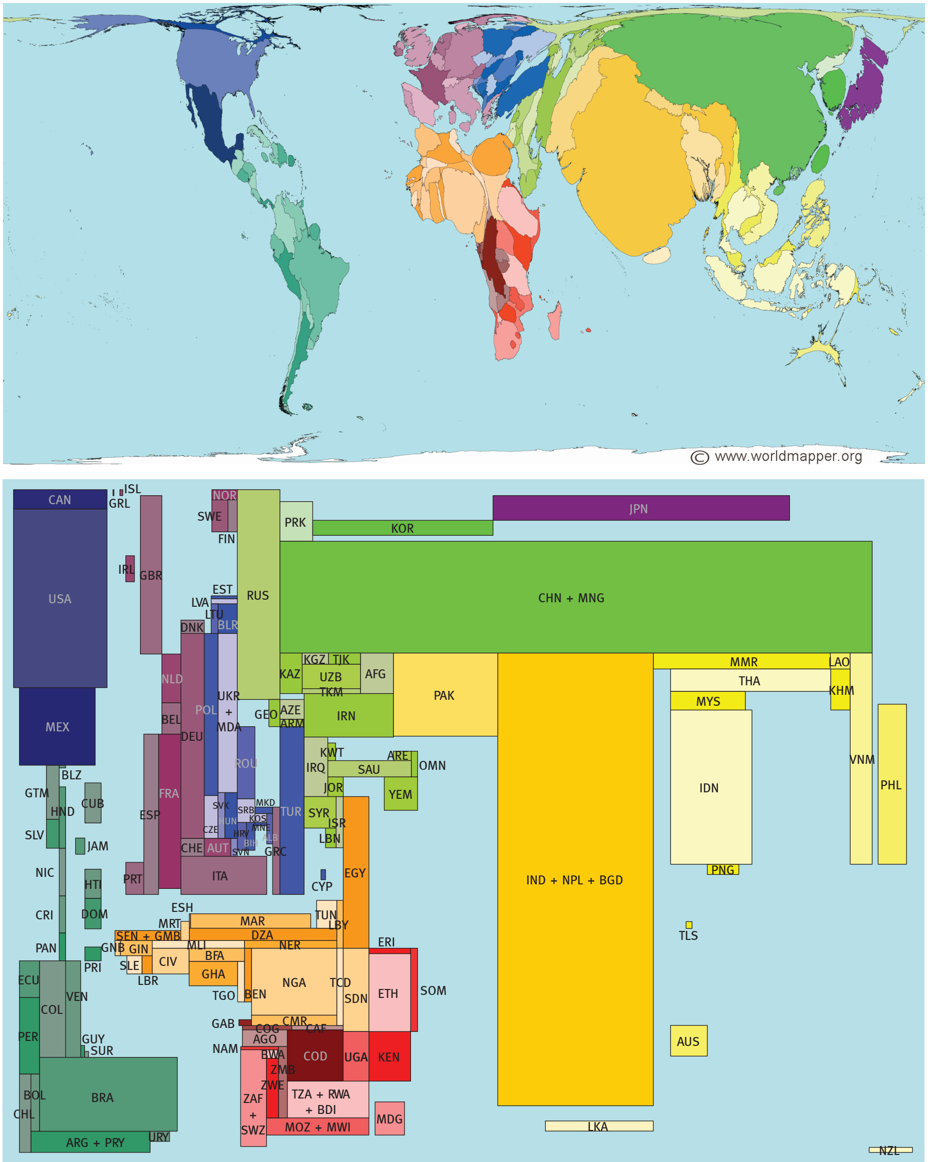


Fig. 9. World population 2002. Top: © Copyright 2006 SASI Group (University of Sheffield) and Mark Newman (University of Michigan).

would like to be able to search in the lattice of RELs for cartograms with the best visual properties. These require different trade-offs between adjacencies, relative positions, aspect ratio and error for every data set and it is a challenge to automatically adapt the fitness function to the requirements of each input.

References

1. J. Bhasker and S. Sahni. A linear algorithm to check for the existence of a rectangular dual of a planar triangulated graph. *Networks*, 7:307–317, 1987.
2. K. Buchin, B. Speckmann, and S. Verdonschot. Optimizing regular edge labelings. In *Proc. 18th International Symposium on Graph Drawing (GD 10)*, LNCS 6502, pages 117–128, 2011.
3. B. Dent. *Cartography - thematic map design.* McGraw-Hill, 5th edition, 1999.
4. D. Dorling. *Area Cartograms: their Use and Creation.* Number 59 in Concepts and Techniques in Modern Geography. University of East Anglia, Environmental Publications, Norwich, 1996.
5. J. A. Dougenik, N. R. Chrisman, and D. R. Niemeyer. An algorithm to construct continous area cartograms. *Professional Geographer*, 3:75–81, 1985.
6. É. Fusy. *Combinatoire des cartes planaires et applications algorithmiques.* PhD thesis, École Polytechnique, 2007.
7. É. Fusy. Transversal structures on triangulations: A combinatorial study and straight-line drawings. *Disc. Math.*, 309(7):1870–1894, 2009.
8. M. Gastner and M. Newman. Diffusion-based method for producing density-equalizing maps. *Proceedings of the National Academy of Sciences of the United States of America (PNAS)*, 101(20):7499–7504, 2004.
9. R. Heilmann, D. A. Keim, C. Panse, and M. Sips. Recmap: Rectangular map approximations. In *Proceedings of the IEEE Symposium on Information Visualization (INFOVIS)*, pages 33–40, 2004.
10. R. Inoue, K. Kitaura, and E. Shimizu. New solution for construction of rectilinear area cartogram. In *Proceedings of 24th International Cartography Conference*, 2009. CD-ROM.
11. G. Kant and X. He. Regular edge labeling of 4-connected plane graphs and its applications in graph drawing problems. *Theoretical Computer Science*, 172(1–2):175–193, 1997.
12. K. Koźmiński and E. Kinnen. Rectangular dual of planar graphs. *Networks*, 5:145–157, 1985.
13. NCGIA / USGS. Cartogram Central, 2002. `http://www.ncgia.ucsb.edu/projects/Cartogram_Central/index.html`.
14. J. Olson. Noncontiguous area cartograms. *Professional Geographer*, 28:371–380, 1976.
15. E. Raisz. The rectangular statistical cartogram. *Geographical Review*, 24:292–296, 1934.
16. B. Speckmann, M. van Kreveld, and S. Florisson. A linear programming approach to rectangular cartograms. In *Progress in Spatial Data Handling: Proc. 12th International Symposium on Spatial Data Handling*, pages 529–546. Springer, 2006.
17. W. Tobler. Pseudo-cartograms. *The American Cartographer*, 13:43–50, 1986.
18. M. van Kreveld and B. Speckmann. On rectangular cartograms. *Computational Geometry: Theory and Applications*, 37(3):175–187, 2007.