UK Boundaries

Jeanne Clelland and Moon Duchin

October 2025

1 Introduction

This short report showcases results of using algorithmic methods to harmonize boundary layers in a geographical hierarchy for the creation of legislative constituencies in the United Kingdom. We focus on England in this report, though the methods here may be directly applied elsewhere in the UK.

1.1 Background and materials

We make use of a shapefile with the following partial geographical hierarchy.

Figure 1: The hierarchical spine of geographical units in England provided by the Boundary Commission and used for this project.

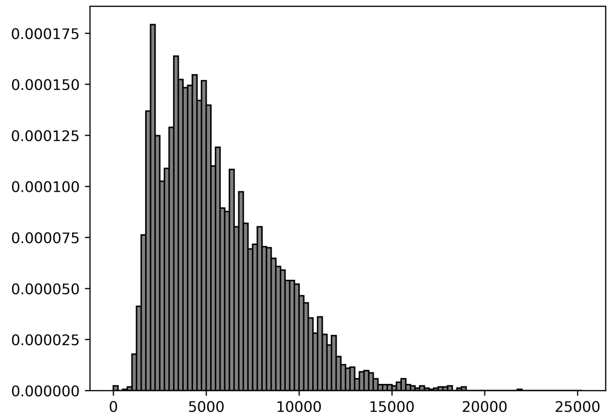
England
|
Regions (9, full coverage)
|
Counties (22, partial coverage)
|
Principal councils (309, full coverage)
|
Wards (7024, full coverage)

Within the wards are 10,464 parishes, but they are often extremely small and are far from full coverage, so may not be useful for this analysis.

The electoral constituencies of England are meant to be balanced with respect to the number of registered voters (called the *electorate* in each unit). Typically the balance when new constituencies are created puts nearly each one within 5% of ideal electorate quota of 73,393 voters, with some exceptions (as for the Isle of Wight). The number of constituencies is controlled by a process described in legislation; there were 533 constituencies in England in 2010, and the number grew to 543 in 2023.

These layers are not *topologically integrated*, meaning that the boundaries do not coincide even when they are close enough to agreement to suggest that they should. We performed some adjustment using the package MAUP (`github.com/mggg/maup`) to reconstruct shapes that do have aligned boundaries. In particular, the analysis below uses **approximations** to the 2010 and 2023 constituencies rebuilt from whole wards. Whole-ward reconstruction certainly alters the population balance, though it should not worsen the hierarchical alignment of the maps. The population of the wards in our shapefile ranges from 66 to 25,508, as shown in Figure 2. Our alternative maps are also built from whole wards, enforced to lie within 10% of ideal electorate size. To finalize these plans to the tighter 5% standard, wards must be split by the Boundary Commission through an electorate estimation process to which we do not have access.

Figure 2: Ward populations.



2 Methods

We conduct runs of the *recombination* Markov chain algorithm for generating districting plans introduced by DeFord et al. [1] and implemented in the GerryChain Python package (`mggg.github.io/GerryChain`).

To facilitate optimization, we introduce metrics of unit splitting. We count the *pieces* into which the constituencies cut each geographical layer. For example, if one county contains parts of constituencies a, b, c while another intersects a and d, then together they contribute five county pieces to the total.

- **C**: the number of county pieces created by the constituencies.
- **P**: the number of principal council pieces created by the constituencies.
- **O**: the number of old-constituency pieces created by the (new) constituencies.
- **D**: the displacement, or the share of current registered voters who do *not* receive the same constituency assignment as they did in 2010, given a geographically optimized matching of old constituencies to new.

We regard O as a computationally efficient indicator for D, the preferred measurement of alignment between old boundaries and new. The O score can be computed in a fraction of a second, whereas each calculation of D takes several seconds.

The 2023 constituency plan has a score of $(283, 792, 994, 0.16)$. The last figure $D = .16$ says that 16% of current registered voters received new constituency assignments in the shift from the 2010 to 2023 boundaries. Equivalently, we say that the 2023 plan has 84% *core retention* with respect to previous boundaries. This figure should be regarded as quite strong, especially considering that the overall number of constituencies has grown (so that 10/543 or about 2% of the voters have necessarily had new assignments because their constituencies did not exist before.)

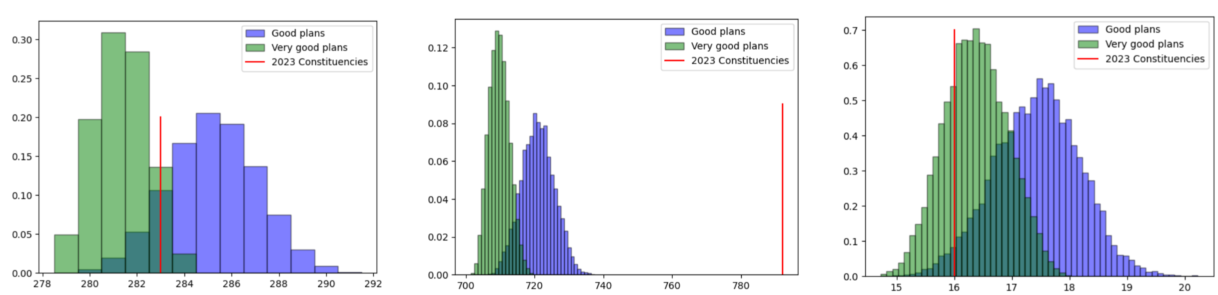
We now sketch the random walk process. Beginning with an initial districting plan for England, we propose a change by taking a "recombination" step, which fuses two of the constituencies and draws on a mathematical structure called a spanning tree to split them in a new way. To find well-aligned examples, we run chains using a combination of two techniques that aim to heuristically improve the C, P, D scores using C, P, O scores. The first is a *region-aware* variant of recombination that applies weights to the selection of spanning trees.[1] In addition, we employ the *short bursts* method of optimization, with an objective function like $z = 20C + 10P + O$ which we view as an aggregate score of boundary misalignment that we will seek to minimize.[2] (Several different combinations were attempted, and all gave similar results.) If the run is set to have a burst length of 50, then 50 recombination steps are taken and the one with the lowest z value is accepted as the next plan in our sequence. Then we iterate, alternating between taking 50 recombination steps and restarting at the one with the lowest z score. In all, we take $500,000$ steps from each of 9 different starting plans in each of the 9 regions. One such run takes on the order of 40 minutes to run on a standard laptop.

By sifting through the plans collected across all of these runs, we choose 9 *good* plans for each region, and from those we highlight 3 *very good* plans. We then combine these to create an assortment of "good" and "very good" plans for England as a whole.

3 Findings

Through a few days of exploratory work, we confirm that it is possible to somewhat improve on the hierarchical integrity of the units available to us, though we also find that the 2023 constituency boundaries already balanced the factors quite successfully.

Figure 3: County pieces, principal council pieces, and displacement scores of our heuristically optimized plans compared to the 2023 constituencies. In all three metrics, lower scores are better.



C: County pieces

P: Principal council pieces

D: Displacement wrt 2010 constituencies

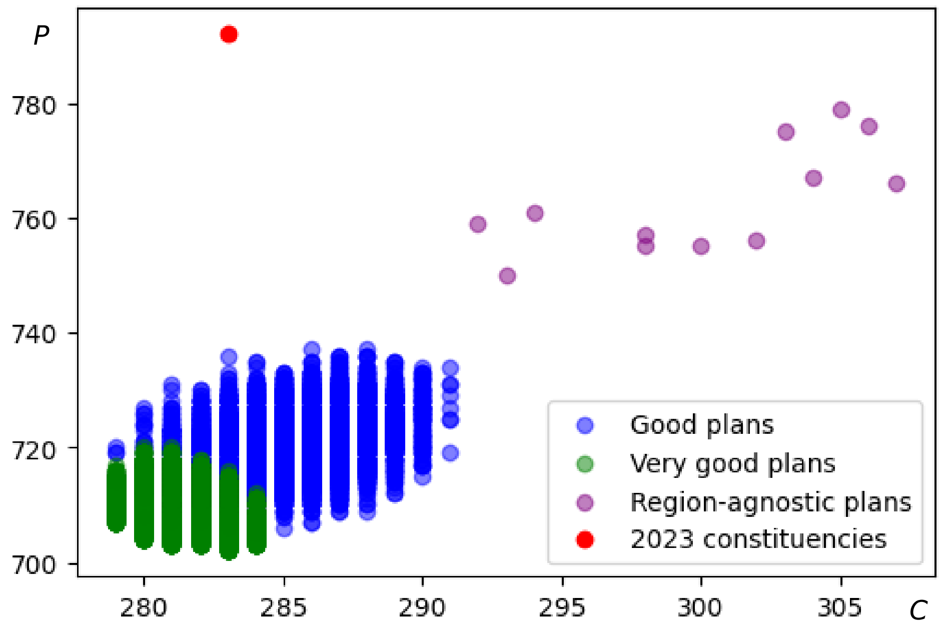
As Figure 3 shows, the optimization runs can modestly improve on the splitting of counties and the displacement with respect to previous constituency boundaries, and can more significantly improve on the intactness of principal councils.

[1] For instance, we "surcharge" edges of the adjacency graph of units that lie between counties relative to edges within counties; then, a minimum-weight spanning tree will be more likely to have large components entirely within counties. This idea was independently proposed by Daryl DeFord and Amy Becker. Documentation of the surcharge method for region-aware chain runs can be found here.

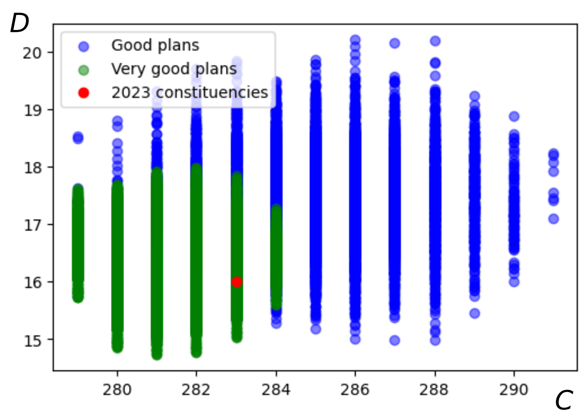
[2] The short bursts idea was initially proposed by Zach Schutzman and is developed in an article by Cannon et al. [2]. Documentation within GerryChain can be found here.

To see another view of the success of these optimization schemes, we now show scatterplots that take the scores two at a time, in order to visualize tradeoffs.

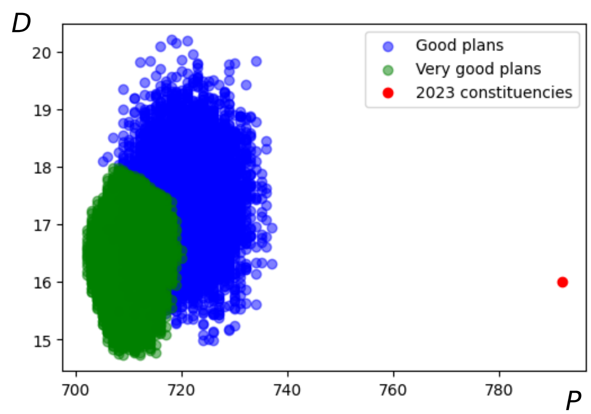
Figure 4: County pieces, principal council pieces, and displacement scores of heuristically optimized plans compared to the 2023 constituencies. In all three metrics, lower scores are better.



P versus C



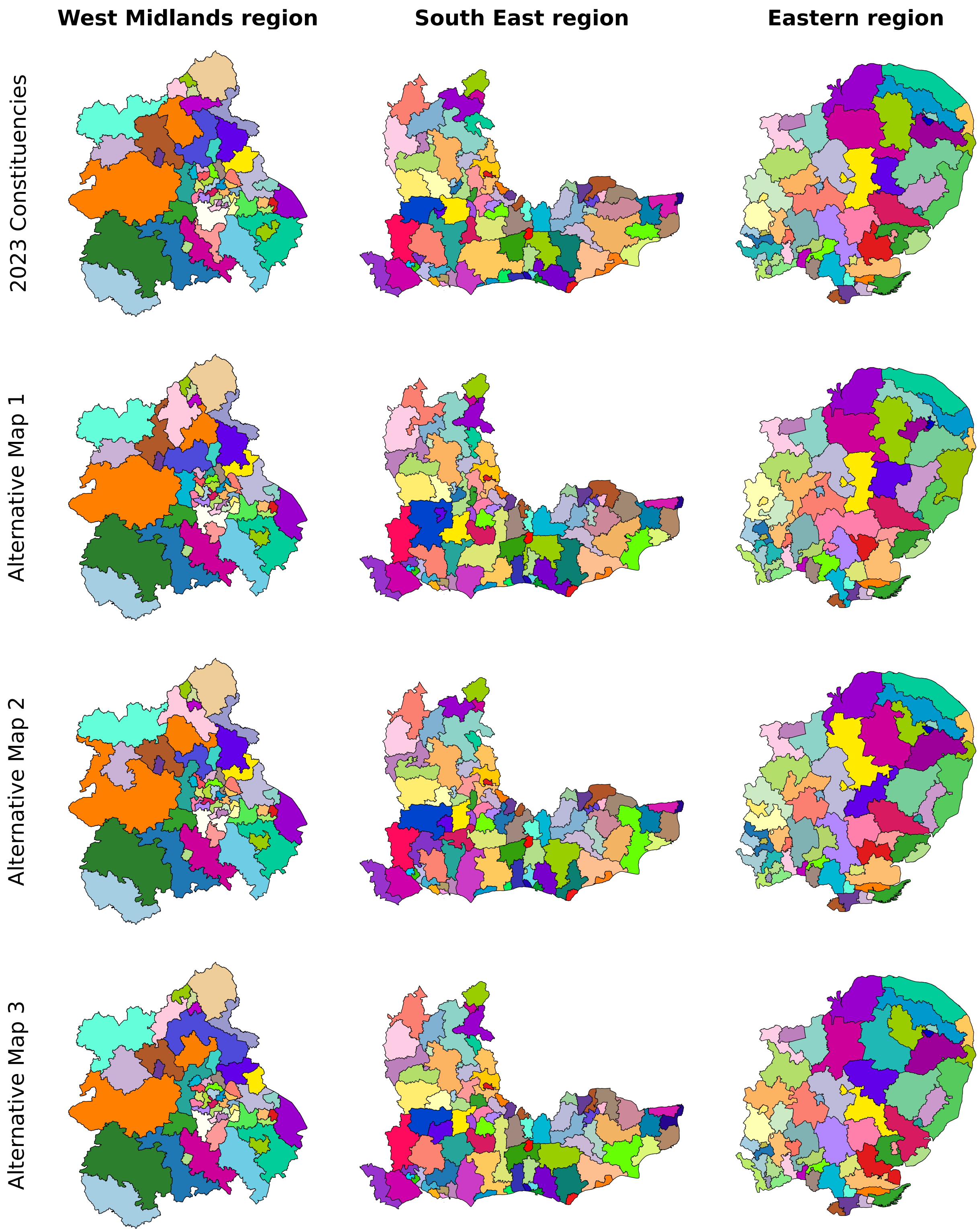
D (percent) **versus C**



D (percent) **versus P**

Note that the *P* vs. *C* plot at top also includes, in purple, a small selection of plans that were optimized *without* first separating the country into its nine regions. Though this is only an initial attempt, it suggests that the use of the regions is helpful for better harmonization of boundaries.

Figure 5: The 2023 constituencies are shown for selected regions in the top row, along with alternative plans that harmonize the hierarchy slightly better.



Finally, we present selected maps in Figure 5 and Table 1 for just three of the nine regions, to give both a visual and a numerical sense of the incremental improvements. Though the randomized optimization runs were started from multiple different starting points, Figure 5 illustrates the similarity of the best plans to the 2023 constituencies. (Colors are aligned by optimal population match.)

Table 1: We compare the county pieces, principal council pieces, old-to-new constituency pieces, and displacement (C, P, O, D) across twelve regional plans: parts of the 2023 constituency map and several alternatives found by the algorithmic searches described above.

region	West Midlands	South East	Eastern
2023 Constituencies	(23, 78, 100, 0.14)	(67, 140, 171, 0.19)	(53, 100, 106, 0.11)
Alternative Map 1	(22, 71, 91, 0.13)	(68, 123, 145, 0.18)	(53, 89, 99, 0.21)
Alternative Map 2	(22, 70, 91, 0.13)	(67, 127, 145, 0.17)	(54, 86, 95, 0.16)
Alternative Map 3	(23, 71, 89, 0.12)	(66, 124, 144, 0.17)	(54, 87, 89, 0.13)

4 Conclusion

Simple runs of the leading districting algorithms employed here show that it is possible to modestly improve on boundary alignment in the construction of electoral constituencies in England. With a bit more time and tuning, we would expect to find more significant improvements. Furthermore, the methods here are adaptable to other areas and to additional layers of geographical hierarchy.

The whole-ward construction has advantages and disadvantages; on one hand, a disadvantage is that the constituencies in these examples would require adjustments below the ward level with a process available to the Boundary Commission. On the other hand, an advantage is that any statistics that are available on wards—total population, demographics, social variables, and so on—can now be included in the comparison of alternative plans.

References

[1] Daryl DeFord, Moon Duchin, and Justin Solomon, *Recombination: A Family of Markov Chains for Redistricting*, Harvard Data Science Review, Vol. 3, No. 1, Winter 2021. https://hdsr.mitpress.mit.edu/pub/1ds8ptxu/release/3.

[2] Sarah Cannon, Ari Goldbloom-Helzner, Varun Gupta, JN Matthews, and Bhushan Suwal, *Voting Rights, Markov Chains, and Optimization by Short Bursts*, Methodology and Computing in Applied Probability, Vol. 25, Article 36 (2023). https://link.springer.com/article/10.1007/s11009-023-09994-1.