The general *p*–compact regions problem

Richard Church
Wenwen Li

Outline

- Problem motivation
- Past work: a synthesis
- Formulating the problem
- Devising a solution technique
- Results
- Conclusions & future directions

Problem motivation

- The MRPI plans to apply RELU-TRAN, a large scale economic/transportation model, to the Los Angeles region.

  Anas, A. and Y. Liu (2007) "A regional economy, land use and transportation model (RELU-TRAN): Formulation, Algorithm design, and Testing," Journal of Regional Science 47; 415-455.

- The feasible size for applying RELU-TRAN is 100 spatial units (planning units)
- The basic planning unit is the Transportation Analysis Zone (TAZ)
- Objective here is to aggregate TAZs into 100 contiguous planning units or model regions

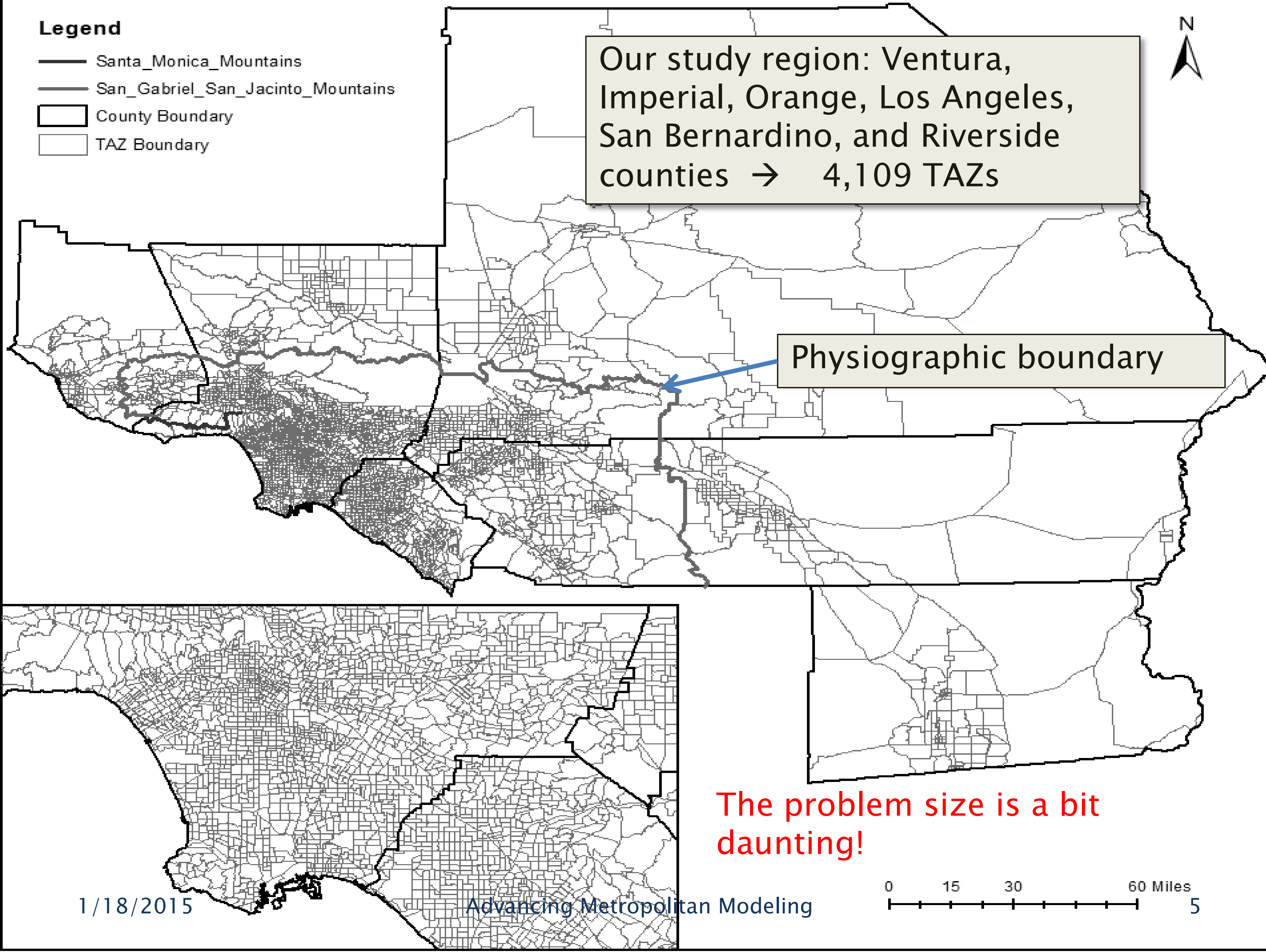
The essence

An aggregation problem: clustering 4,109 TAZs into 100 planning areas/units or regions

- Planning regions need to be contiguous
- They should not be too large
- They should be compact
- The traffic within the region should be less than X% of the total commuting traffic generated by the region
- Known subcenters should be part of only one region
- Regions should not cross a physiographic boundary

6

Our study region: Ventura, Imperial, Orange, Los Angeles, San Bernardino, and Riverside counties → 4,109 TAZs



Physiographic boundary

The problem size is a bit daunting!

What's in a name?

This problem belongs to the class of problems called:

- Political Districting & Redistricting

Hess et al. (1965) "Nonpartisan political redistricting by computer," *Operations Research* 13; 998-106.

Garfinkel and Nemhauser (1970) "Optimal Political Districting by implicit enumeration," *Management Science* 16, B495-508.

Plane, D. A. (1982) "Redistricting reformulated: a maximum interaction separation objective," *Socio-Economic Planning Sciences* 16; 241-244.

Morrill, R.L. (1976) "Redistricting Revisited," *Annals of Association of American Geographers* 66, 548-556.

What's in a name?

This problem belongs to the class of problems called:

- Zonation

Openshaw, S. (1977) "A geographical solution to scale and aggregation problems in region-building, partitioning and spatial modeling," *Transactions of the Institute of British Geographers* 2; 459-472.

Openshaw, S. (1977) "Optimal zoning systems for spatial interaction models," *Environment and Planning A* 9;169-184.

Openshaw, S. (1977) "Algorithm 3: a procedure to generate pseudo random aggregations of N zones into M zones where M is less than N," Environment and Planning A 9; 1423-1428.

What's in a name?

This problem is related to:

- Districting and Redistricting
- Zonation
- Turfing and Sales territory alignment
- School Districting
- Multiple land use allocation (MULA)

P-regions

Duque, J.C., Church, R.L., and R. Middleton ((2011) "The p-regions problem," *Geographical Analysis* 43, 104-126.

Max P-regions

Duque, J.C., Anselin, L., and S.J. Rey (2012) "Max P-Regions problem," *Journal of Regional Science* 52; 397-419.

Compact P-regions

Li, W. , Church, R.L., and M. Goodchild (2013) "The compact p-regions problem," in review

p–Regions problem

- Three proposed formulations:

→Order model based upon Cova and Church (2000)

→A tree model inspired by Miller, Tucker and Zemlin

→A flow–based model inspired by Shirabe (2005)

A formulation: notation

I = the set of areas or basic units being aggregating

i, j = indices used to refer to specific areas, where $i, j \in I$

k = index used to refer to regions, where $k = 1, \ldots, p$

o, O = index and set of contiguity order

N_i = set of adjacent areas to unit i

s_{ij} = a measure of dissimilarity between area i and area j

A formulation: notation

Decision variables:

$$t_{ij} = \begin{cases} 1, & \text{if areas } i \text{ and } j \text{ belong to the same region } k \text{ where } i < j \\ 0, & \text{otherwise} \end{cases}$$

$$x_i^{ko} = \begin{cases} 1, & \text{if area } i \text{ is assigned to region } k \text{ in order } o \\ 0, & \text{otherwise} \end{cases}$$

One of three formulations for the p-regions problem

Minimize $$\sum_{i}\sum_{j>i} s_{ij}t_{ij}$$

1) $$\sum_{i} x_i^{k0} = 1 \quad \forall k = 1, 2, \ldots, p$$

2) $$\sum_{k=1}^{p}\sum_{o} x_i^{ko} = 1 \quad \forall i \in I$$

3) $$x_i^{ko} \le \sum_{j \in N_i} x_j^{k(o-1)} \quad \forall i \in I,\ \forall k = 1, 2, \ldots, p, \forall o \in O$$

4) $$t_{ij} \ge \sum_{o} x_i^{ko} + \sum_{o} x_j^{ko} - 1 \quad \forall k = 1, 2, \ldots, p, \forall i \in I, \text{and}$$

$$\forall j \in I,\ \text{where } i < j$$

Plus integer restrictions on variables

The compact *p*–regions problem

Aggregate *n* basic spatial units into *p*–regions in order to optimize compactness and other metrics (e.g. within region similarities) and while meeting a number of conditions:

Example constraints include:

- regions cannot cross prespecified physiographic boundaries,
- regions must be of a certain size or no larger than a certain size or measure,

….in our case it was a constraint of within zone traffic generated as a percentage of total traffic generated by the zone

Compact *p*–regions problem

Simply put, the CPRP is:

- A multi–objective form of the *p*–regions problem (compactness) and
- Possibly other constraints, e.g. maximum size, can't cross physiographic boundaries, etc.

Compactness Metric

ranges from 0 in case of an infinitely extended shape to 1 in the case of the most compact figure, a circle.

$$C(Z_k) = A_{Z_k}^2 \Big/ 2\pi I_{Z_k}^G$$

the second moment of a region about an axis perpendicular to it and passing through its centroid *G*. (the moment of inertia)

Li, W., Goodchild, M.F., and R.L. Church (2013) "An efficient measure of compactness for two-dimensional shapes and its application in regionalization problems," *International Journal of Geographic Information Systems*

The moment of inertia….

G represents the centroid of unit *i*

for basic unit *i* is:

$$I_i^{G_i} = \rho_i \int d^2 da_i$$

Distance to centroid

Density of some attribute across a unit

So, we'll add to the *p*–regions model a second objective……

$$Min \quad \sum_{k=1}^{p} C(Z_k)$$

Unfortunately, the Compact *p*–Regions problem is a messy nonlinear integer programming problem.

Our solution tack has been to design a heuristic……………

The MERGE heuristic

MEmory–based Randomized Greedy and Edge-reassignment

A heuristic that is designed to be restarted a number of times, where the best solution found among all runs is kept as the best.

3 main parts:

1) *greedy dealing*
2) *GRASP assignment*
3) *Edge reassignment*

1) and 2): Building a solution

3): Improving the solution

Greedy in a nut shell

Each model zone starts as a single seed TAZ or sub-center set)….at each step

- Identify TAZs touching the perimeter of a given model zone that have yet to be included in a MZ:
  - For each such TAZ calculate the change in the objective associated with adding this TAZ to that model zone
- Find the TAZ which yields the best change in objective when added to a specific MZ
- Add that TAZ to the specific TAZ
- repeat process until all TAZs are assigned

Greedy Dealing Heuristic

- The behavior of the shape index can be so overpowering that some zones will continue to grow at the expense of others.
- This led to the development of the Dealing heuristic, so that each MZ has at least a moderate size before, moving to GRASP assignment

GRASP improvement

- Moving to real GRASP,
- In the greedy step, we develop a candidate list of the top *X* assignments.
- From the top *X* assignments, we pick at random, one of the top *X*.

- Reason: Greedy will always determine the same set of MZs given a set of seeds. Randomized greedy will not.

Edge reassignment

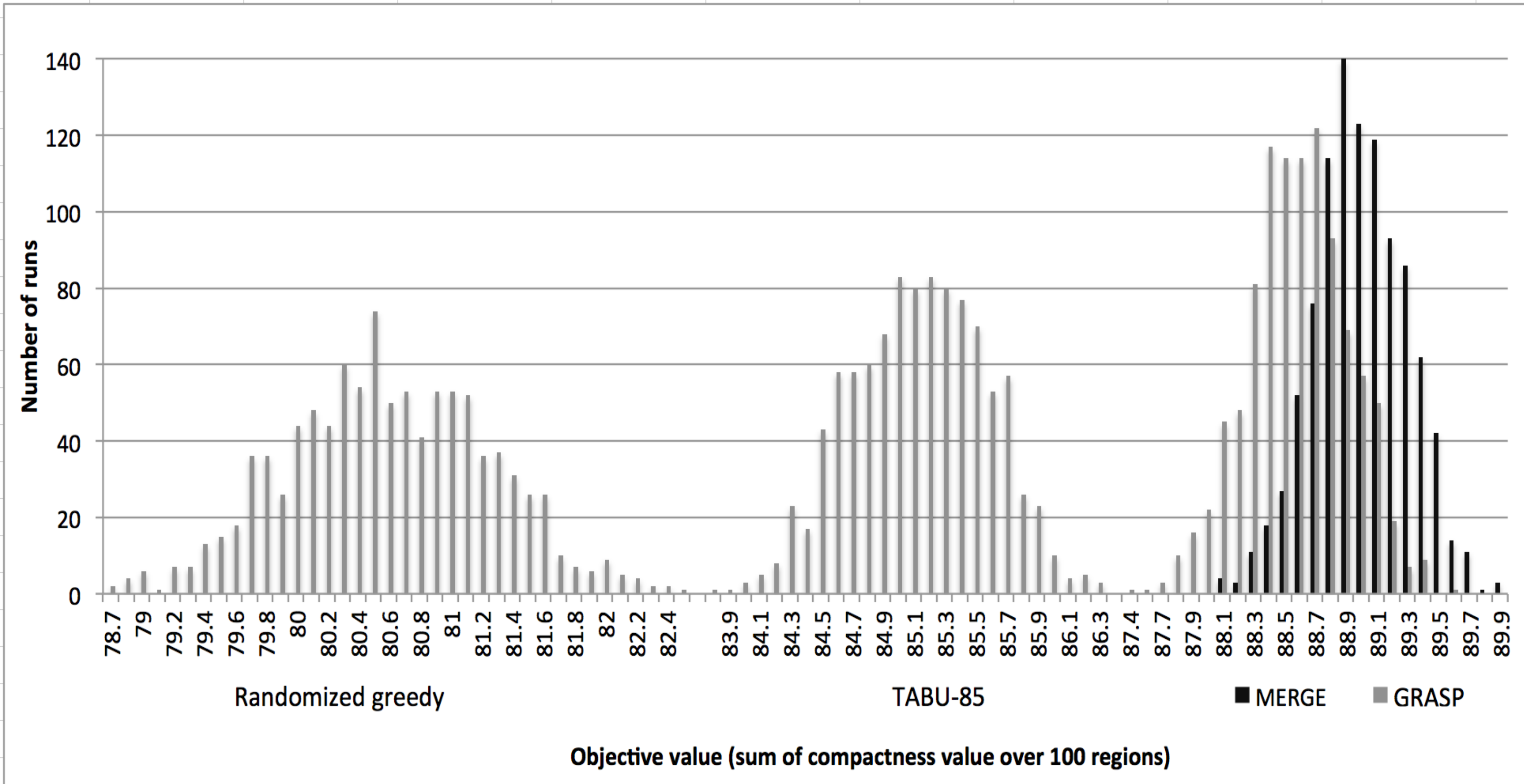
All TAZs have been assigned to a Model Zone (MZ)

- Choose a TAZ at random. If this TAZ is on the boundary of its MZ, identify bordering MZs that touch this TAZ.
  - For each such neighboring MZ, calculate the change in the overall objective, when moving this TAZ to this neighboring MZ
  - If any changes produce an improvement, move the TAZ to the MZ which results in the best improvement.
- Have all TAZs been chosen, it not, return to pick another TAZ for consideration.
- Have any changes been made in MZ delineation? If so repeat above process.

The test

- GREEDY: GD followed by Randomized GREEDY
- GRASP: GD followed by Randomized GREEDY followed by Edge reassignment
- MERGE: GD followed by Randomized Greedy followed by SA based edge reassignment
- TABU-85: GD followed by Randomized Greedy followed by TABU based edge reassignment (TABU list length is 85, same as Duque, Anselin, and Rey, 2012)

Comparison of Approaches

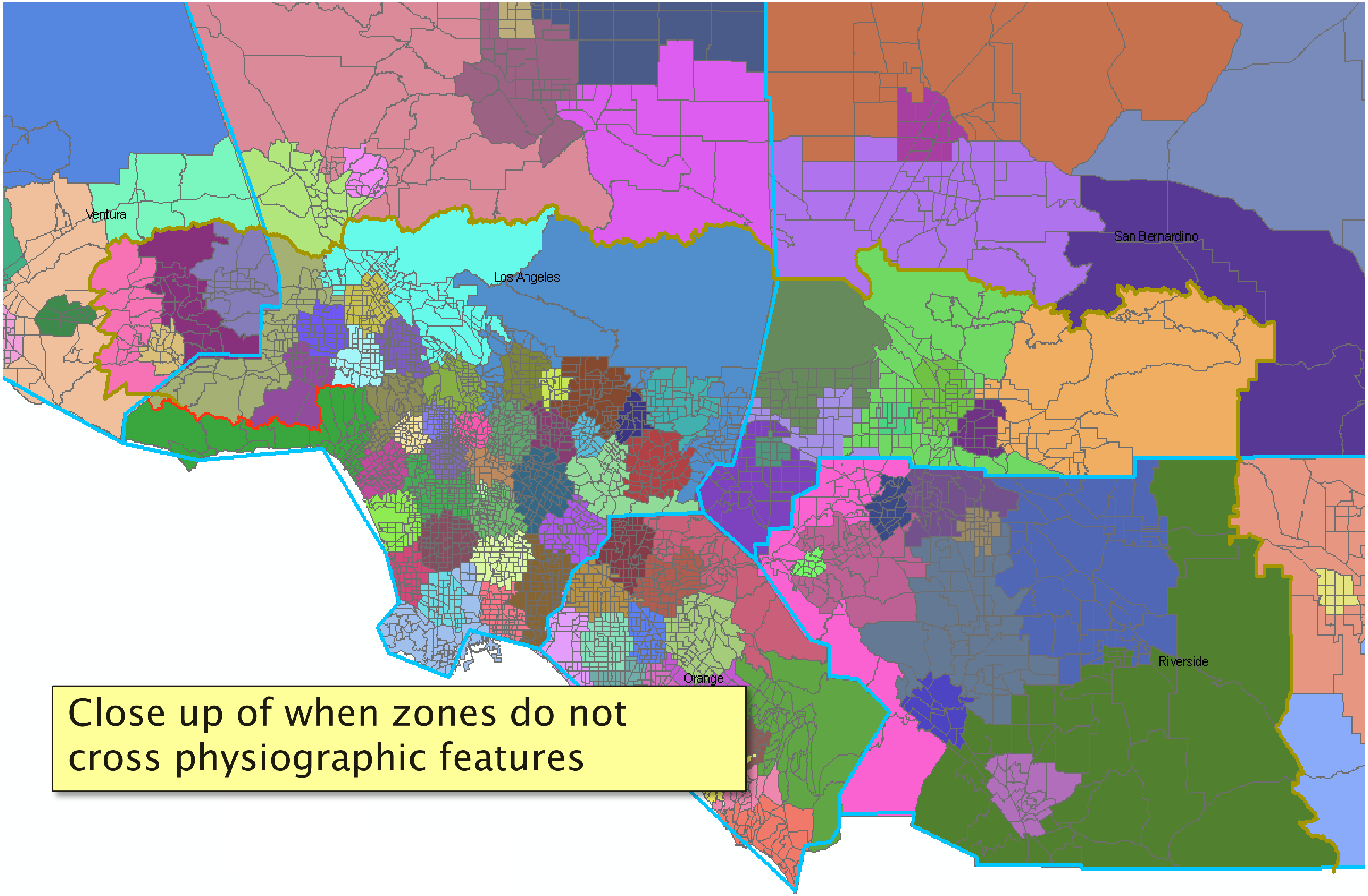


MERGE outperforms GRASP, TABU & GREEDY

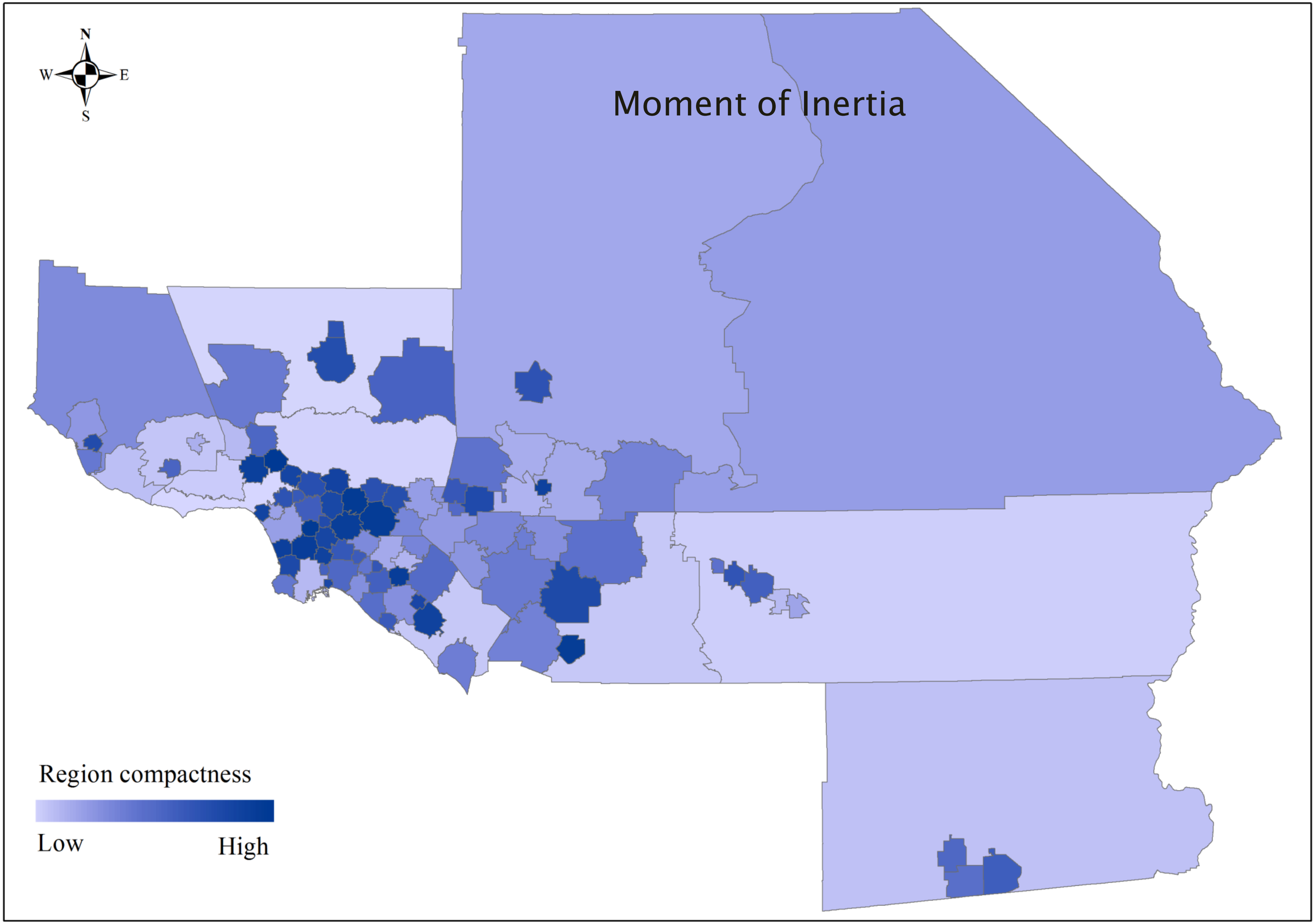
Comparing to Duque, Church, & Middleton

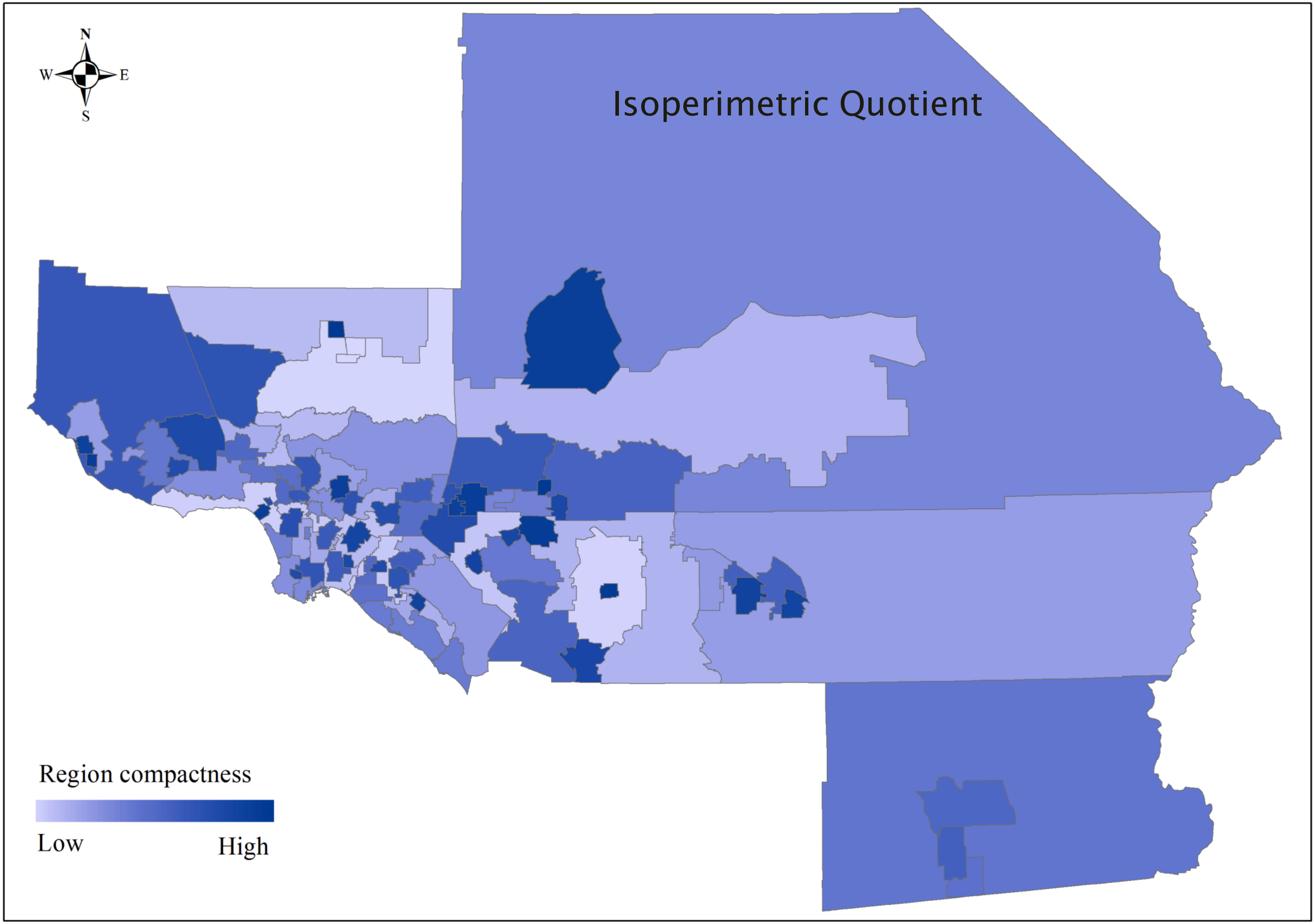
Problem	N	P	Best Known		Run Time (Sec)		
			Duque	Ours	Duque	Ours	Speed-up
1	16	3	27.42*	27.42*	0.84	0.30/run	3
2	16	4	17.34*	17.34*	3.28	0.22/run	15
3	16	5	10.99*	10.99*	0.61	0.23/run	3
4	25	3	59.72*	59.716*	296.38	0.26/run	1140
5	25	4	37.52*	37.512*	4594.53	0.27/run	17017
6	25	6	19.01*	19.009*	1439.73	0.18/run	7999
7	49	3	416.11	378.51	-	0.5/run	21600
8	49	5	226.32	169.83	-	0.36/run	30000
9	49	7	101.36	101.36	-	0.28/run	38571
10	49	10	55.32	55.32	-	0.24/run	45000

*Optimal (by CPLEX) - Run stopped after 3 hours



Close up of when zones do not cross physiographic features


N
W
E
S
Moment of Inertia
Region compactness
Low
High



N
W
E
S
Isoperimetric Quotient
Region compactness
Low
High

Summary & Conclusions

- An expanded form of the p-regions problem has been proposed
- A nonlinear-integer programming model has been developed
- A new heuristic has been developed which appears to outperform other approaches for the compact case.
- The best solution generated by this process formed the basis for which the final set of RELU-TRANS model regions were defined

Acknowledgements

- Thanks are due to Alex Anas, Richard Arnott and many others in the MRPI project team