Predicting Wildfires:

Propositional and Relational Pre-processing Approaches

MARIANA OLIVEIRA

LUÍS TORGO

VÍTOR SANTOS COSTA

DISCOVERY SCIENCE

OCTOBER 2016

BARI, ITALY

* Thanks to Dr. João Torres

Outline

Motivation

Wildfires in Portugal

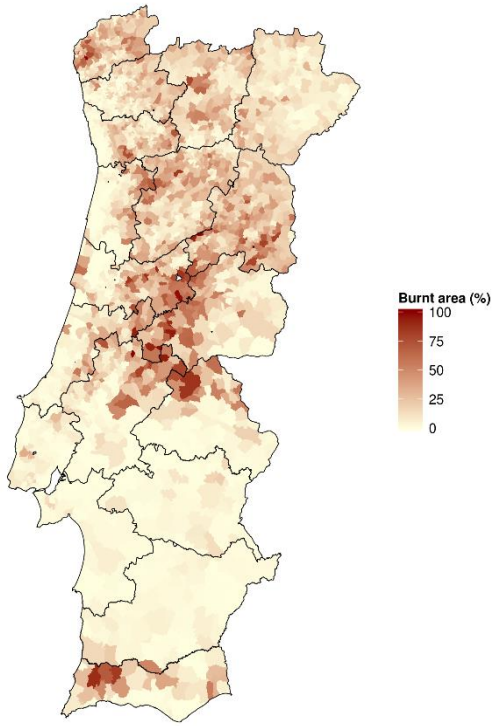
Predicting wildfires

- Propositional and relational pre-processing approaches
- Experimental results

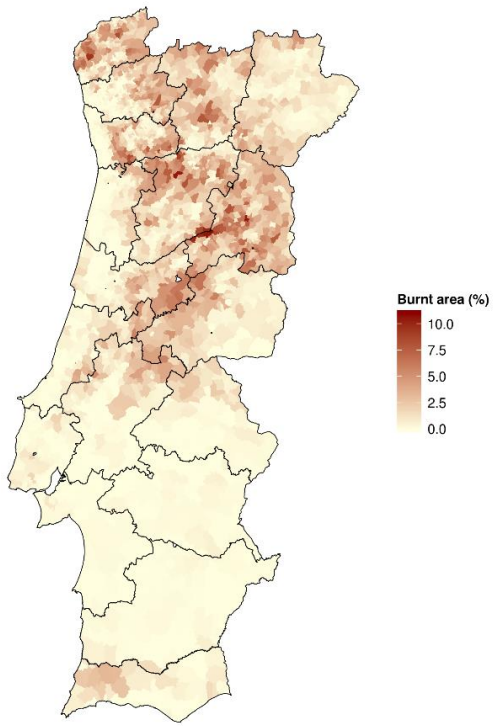
Conclusion

Motivation

Motivation and main goals

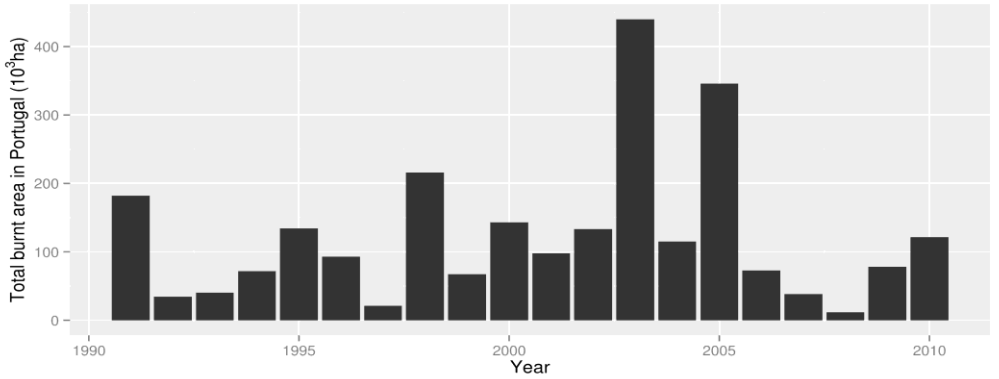


Maximum burnt area (%)



Mean burnt area (%)

Total burnt area in Portugal (10^3ha)



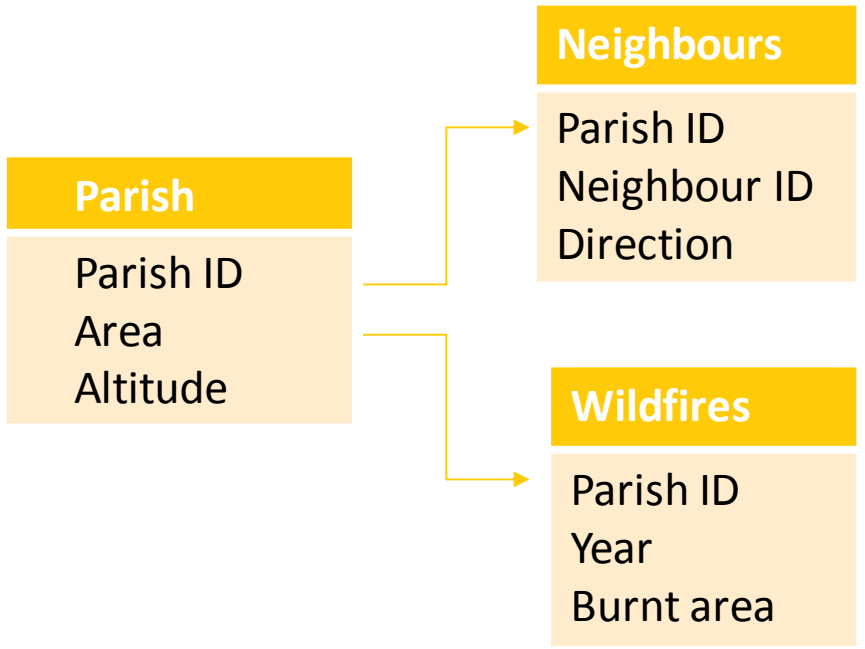
- Apply propositional and relational pre-processing methods to predict yearly burnt area (%) in Portuguese civil parishes.
- Evaluate and compare approaches

Understanding a spatio-temporal dataset

PROPOSITIONAL

Parish ID	Altitude	Year	Burnt area

RELATIONAL



Wildfires in Portugal

AN APPLICATION

Wildfires in Portugal

Portugal

18 districts

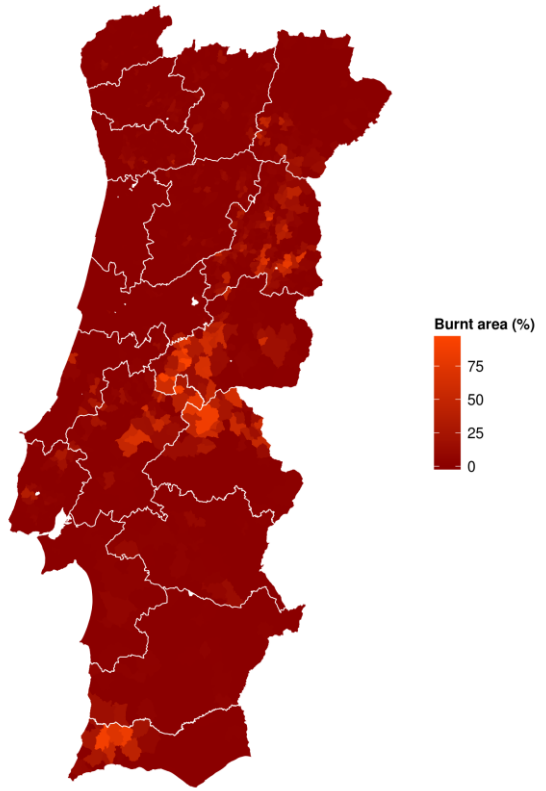
2882 civil parishes

Area: 20 ha – 88 000 ha (median: 1700 ha)

Data

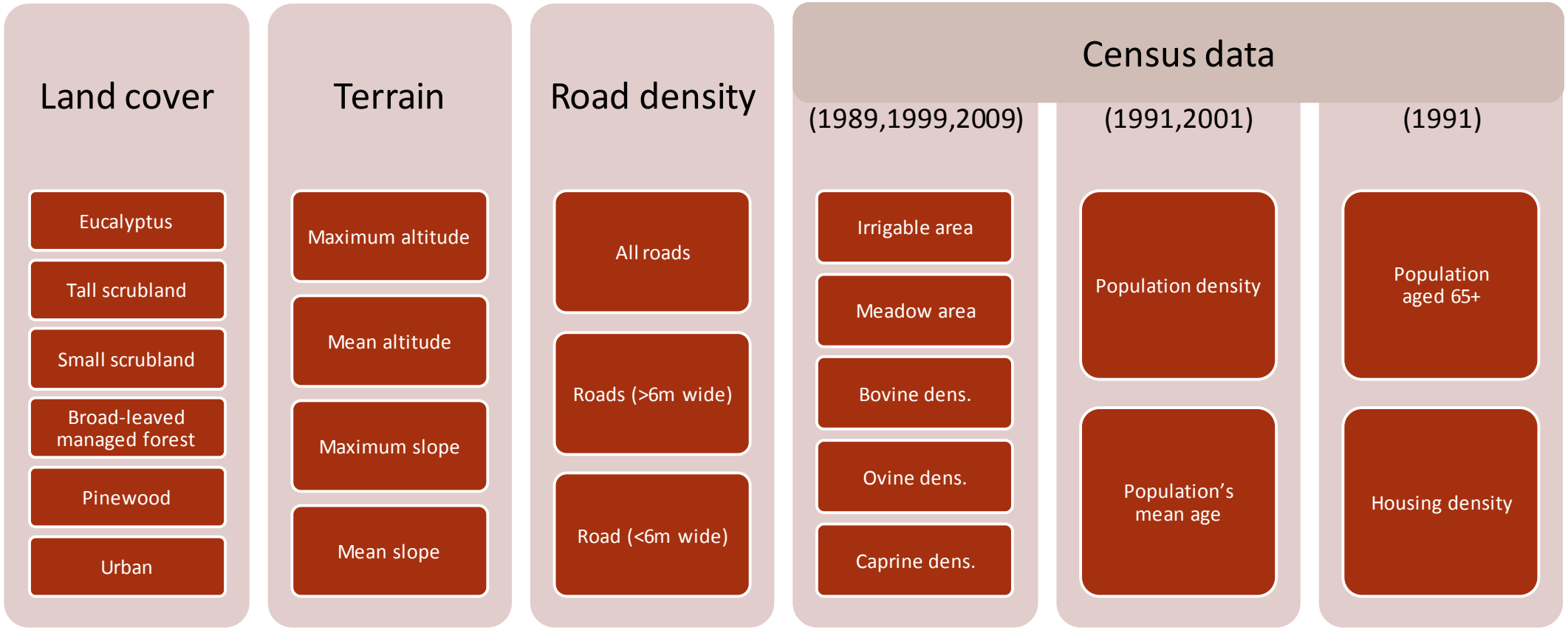
Yearly burn fraction area 1991 to 2010

Background knowledge

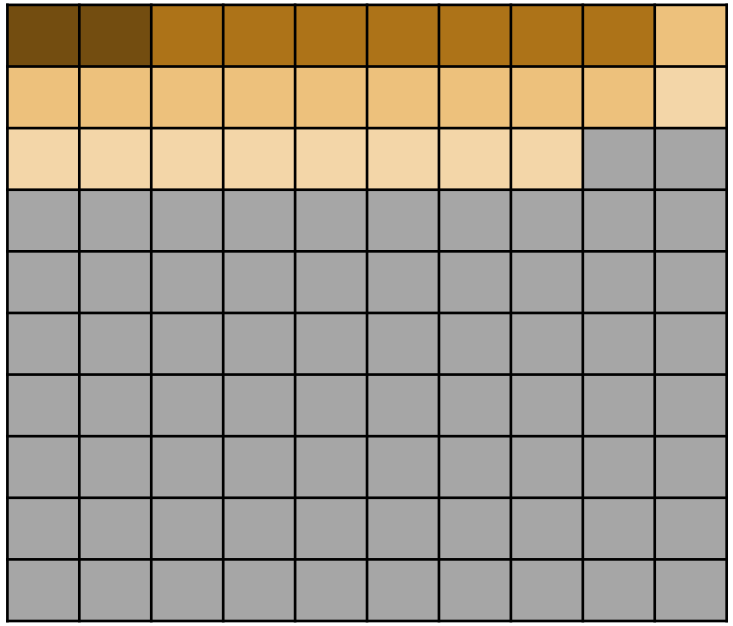


Burn fraction in 2003

Background knowledge



Imbalanced domain



28% burned more than 0%

19% burned 1% or more

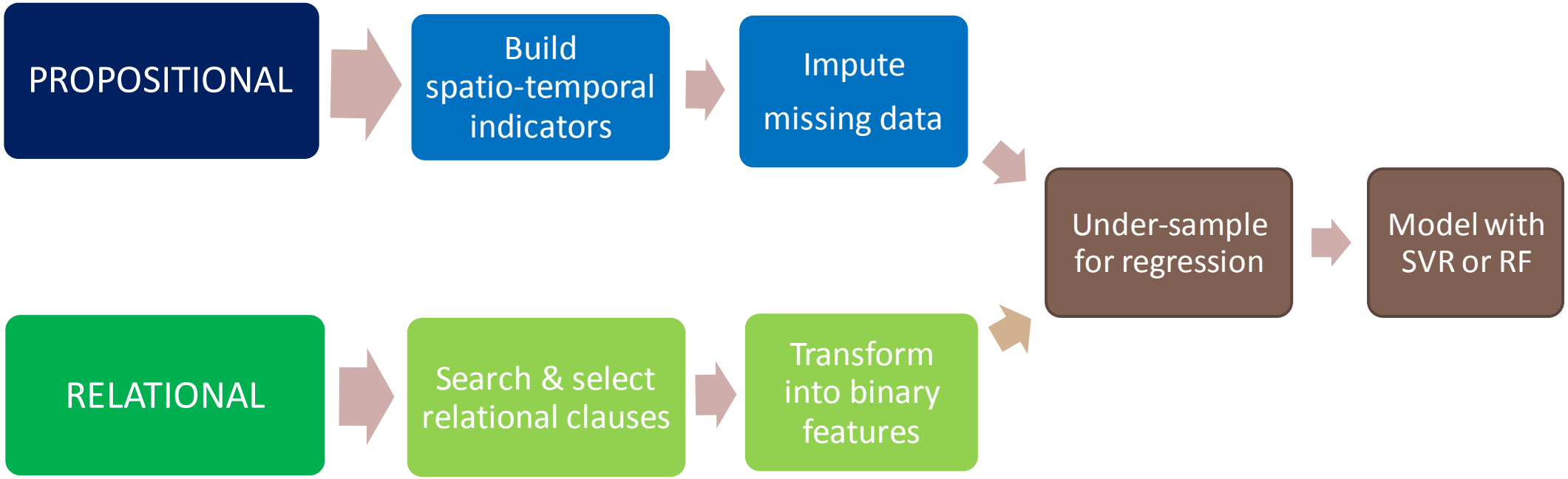
9% burned 5% or more

2% burned 20% or more

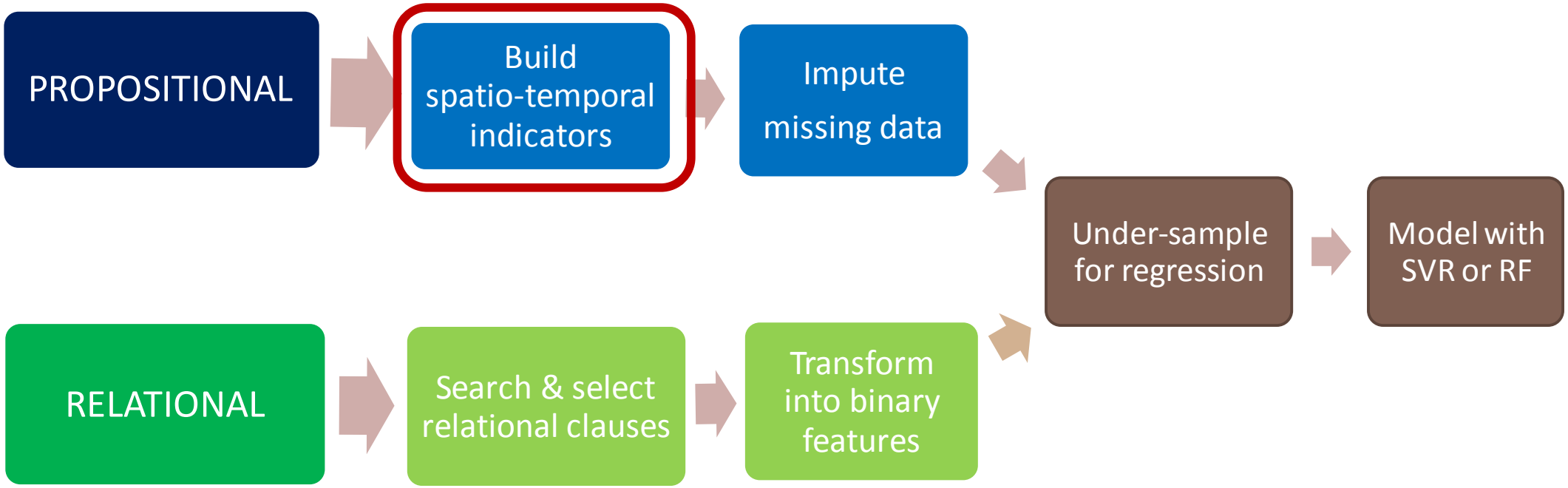
Predicting wildfires

PROPOSITIONAL AND RELATIONAL PRE-PROCESSING APPROACHES

Propositional and relational approaches



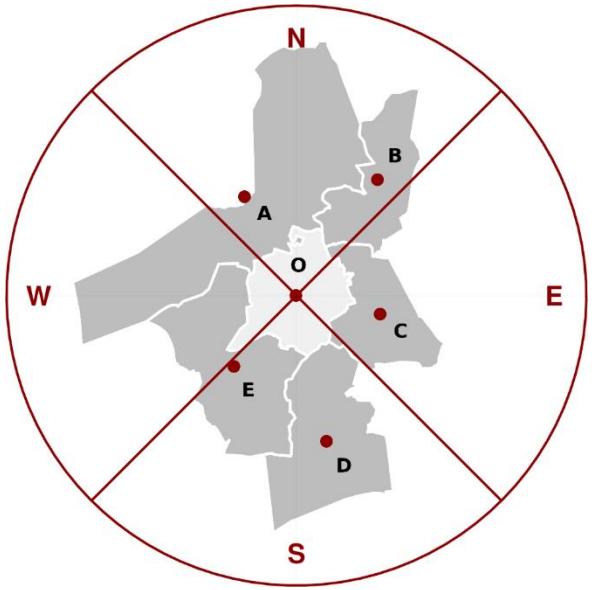
Propositional approach

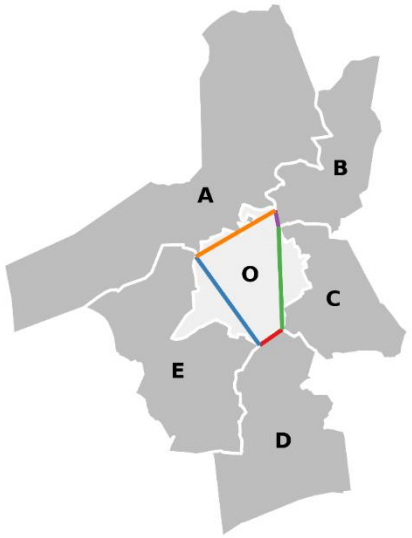


Propositional approach

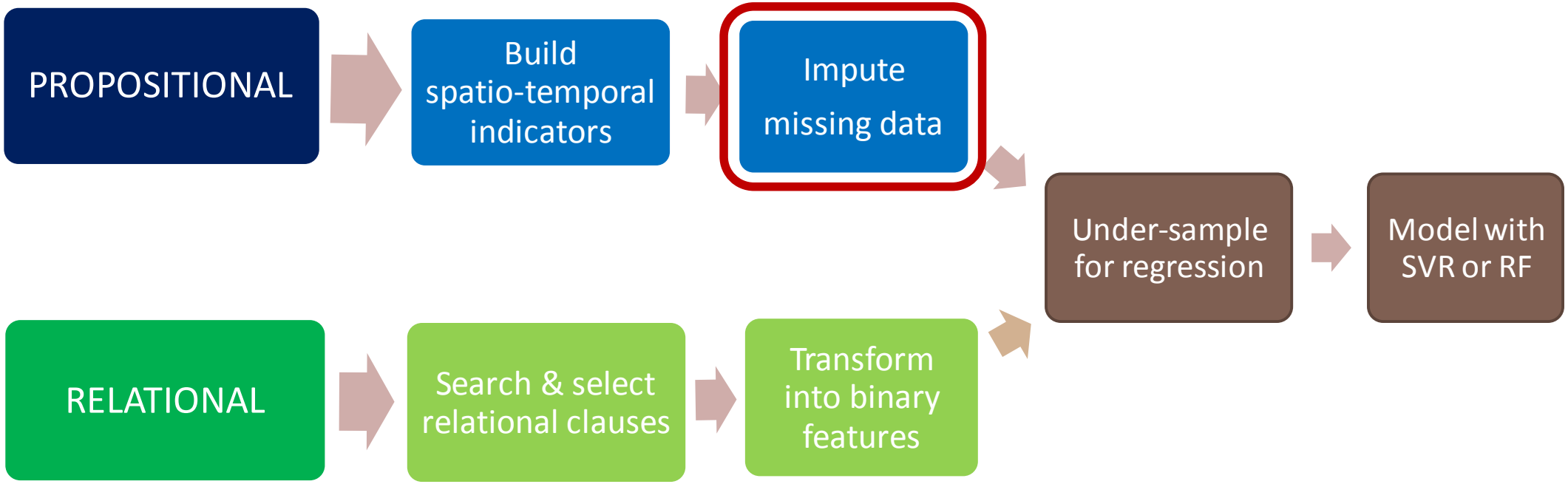
BUILD SPATIO-TEMPORAL INDICATORS

- Calculate exponential average of target for each neighbour
- Use average of neighbours in each direction weighted by simplified border

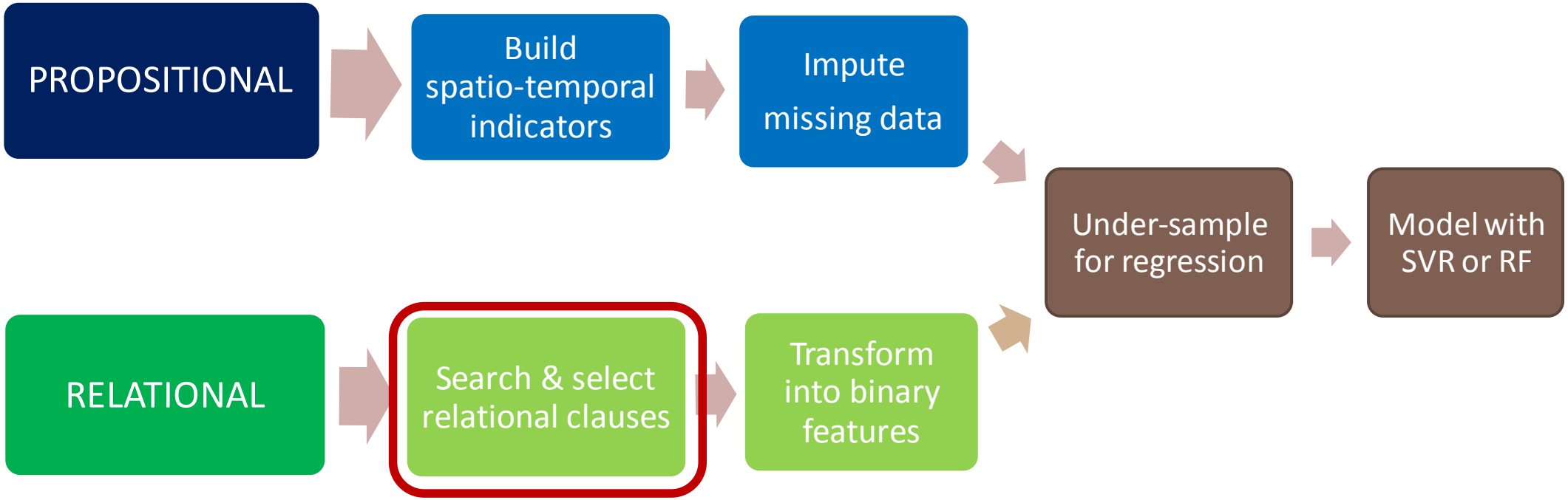




Propositional approach



Relational approach

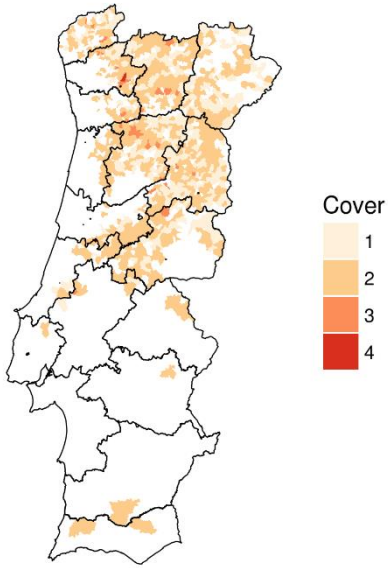


Relational approach

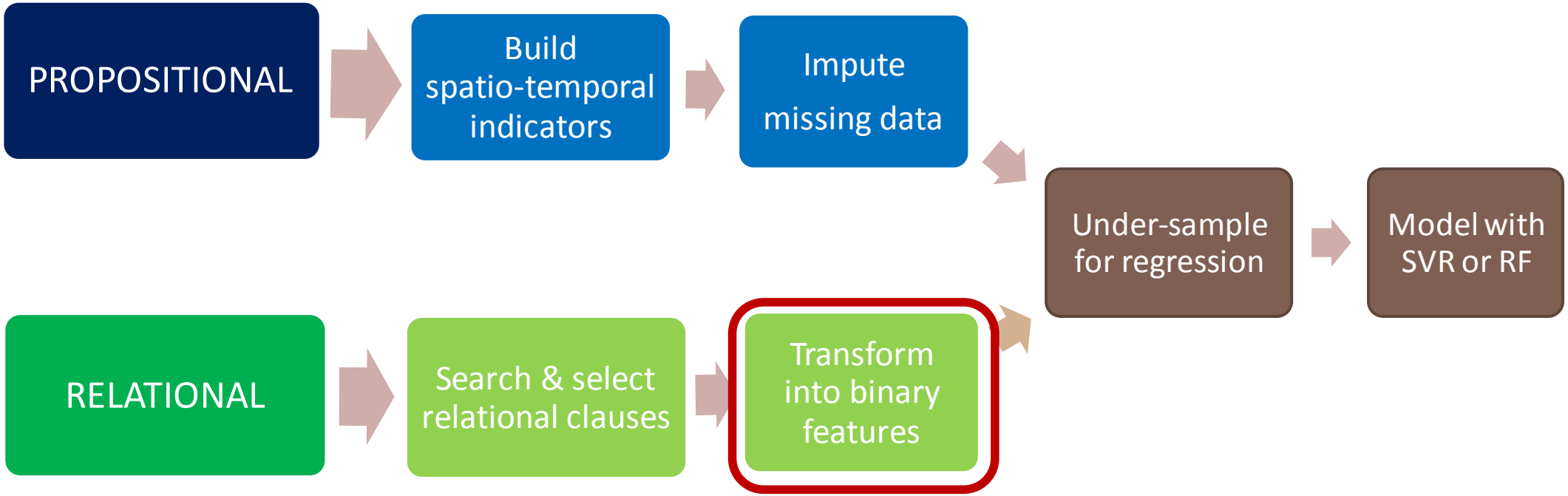
SEARCH AND SELECT RELATIONAL CLAUSES

- Use random example as seed
- Saturate and reduce using $\boldsymbol{F_\beta - measure}$
- Save and select **best so far**
- Repeat 60 times for each $\beta \in \{0.75, 0.9, 1.0, 1.1, 1.25\}$

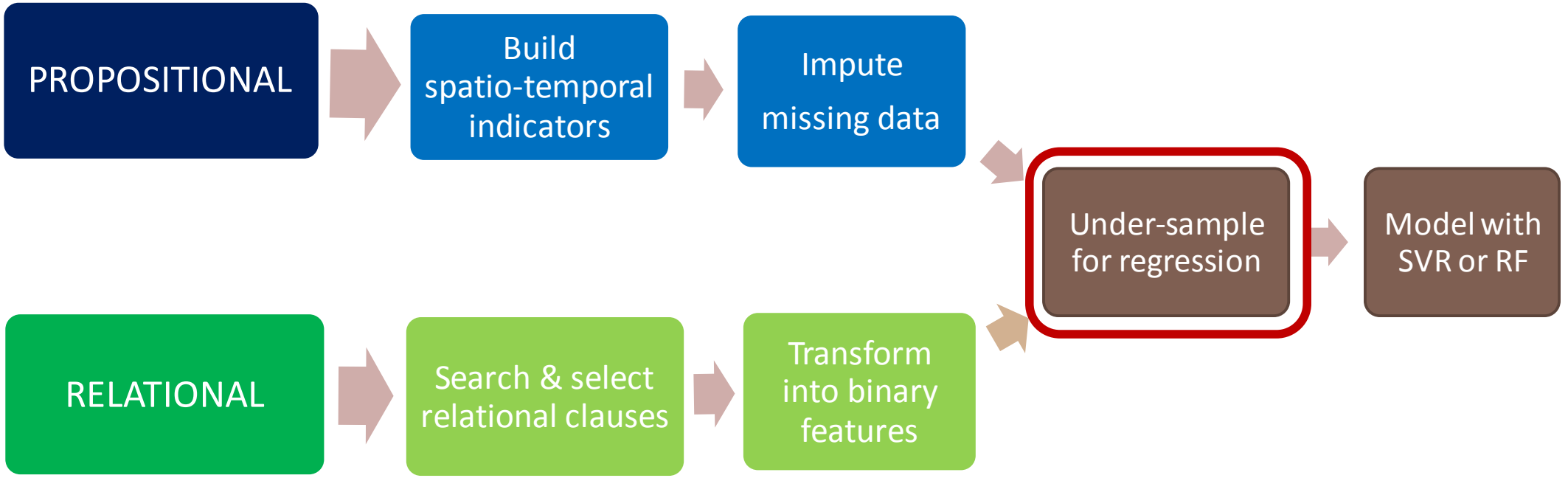
burnt(ParishA, Year) :-
maxAltitudeGE(ParishA, 507) ,
neighbourDirection(ParishA, ParishB, south) ,
yearsSinceFireLE(ParishB, 5) .

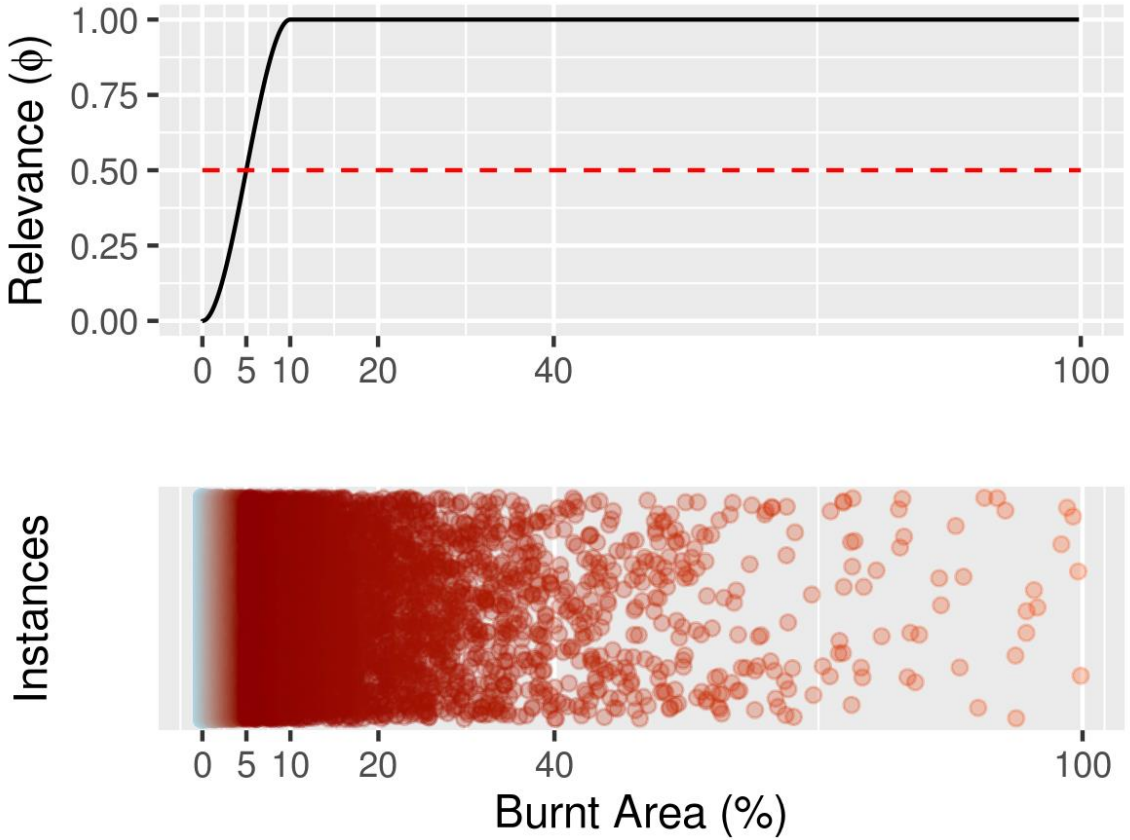


Relational approach



Propositional and relational approaches

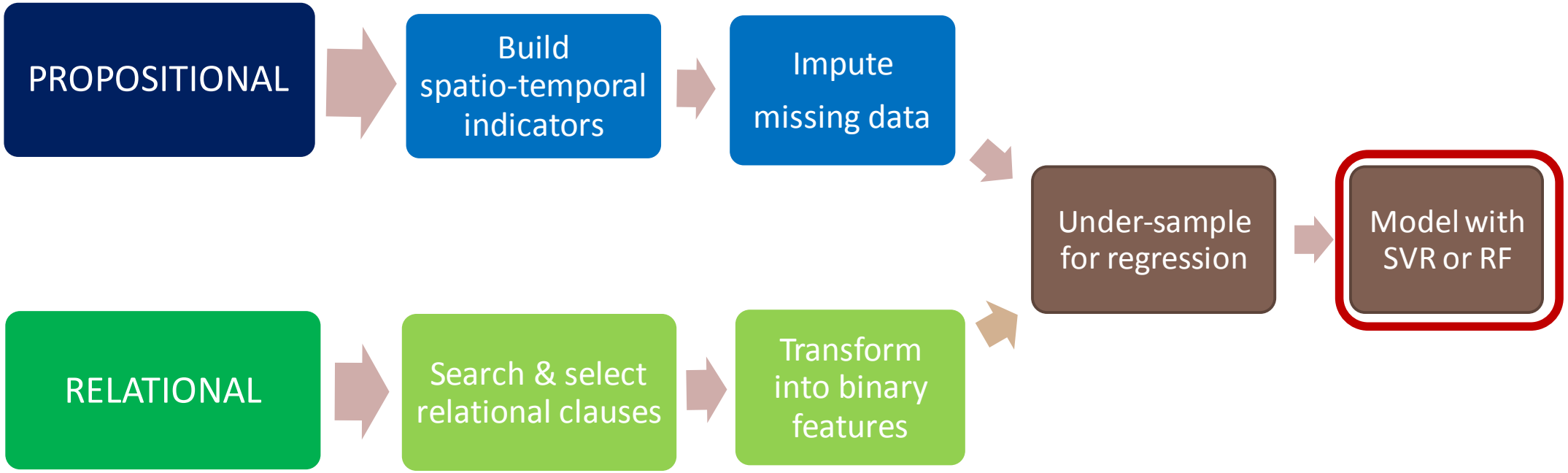




Under-sampling for regression

Proposed by Torgo *et al.* (2013) and implemented in R package UBL

Propositional and relational approaches



Predicting wildfires

EXPERIMENTAL RESULTS

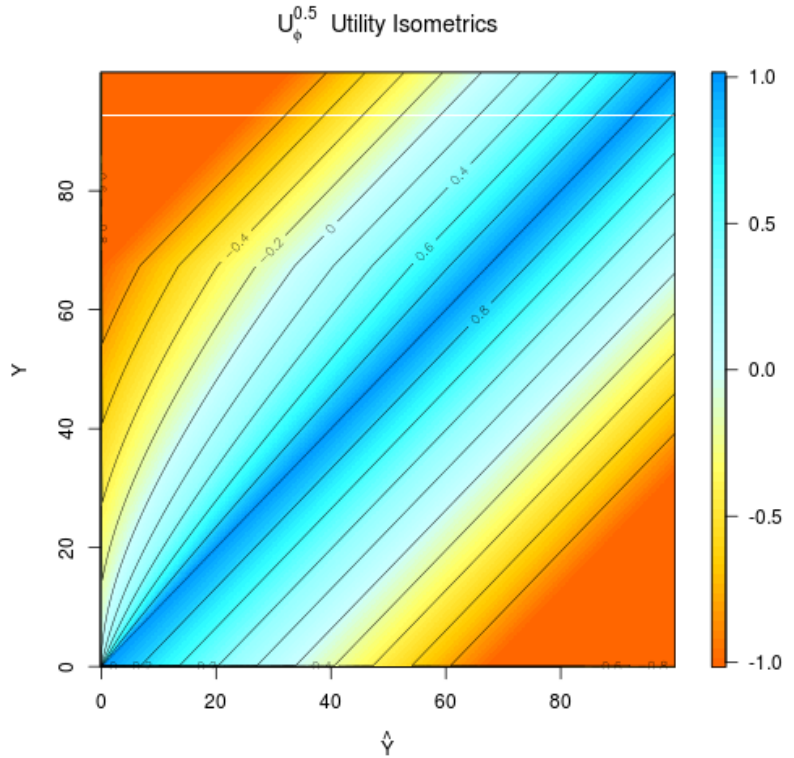
Experimental setup

Branco et. al

$$precision_R = \frac{\sum_{\phi(\widehat{y_i})>t_R}(1+u_i)}{\sum_{\phi(\widehat{y_i})>t_R}(1+\phi(\widehat{y_i}))}$$

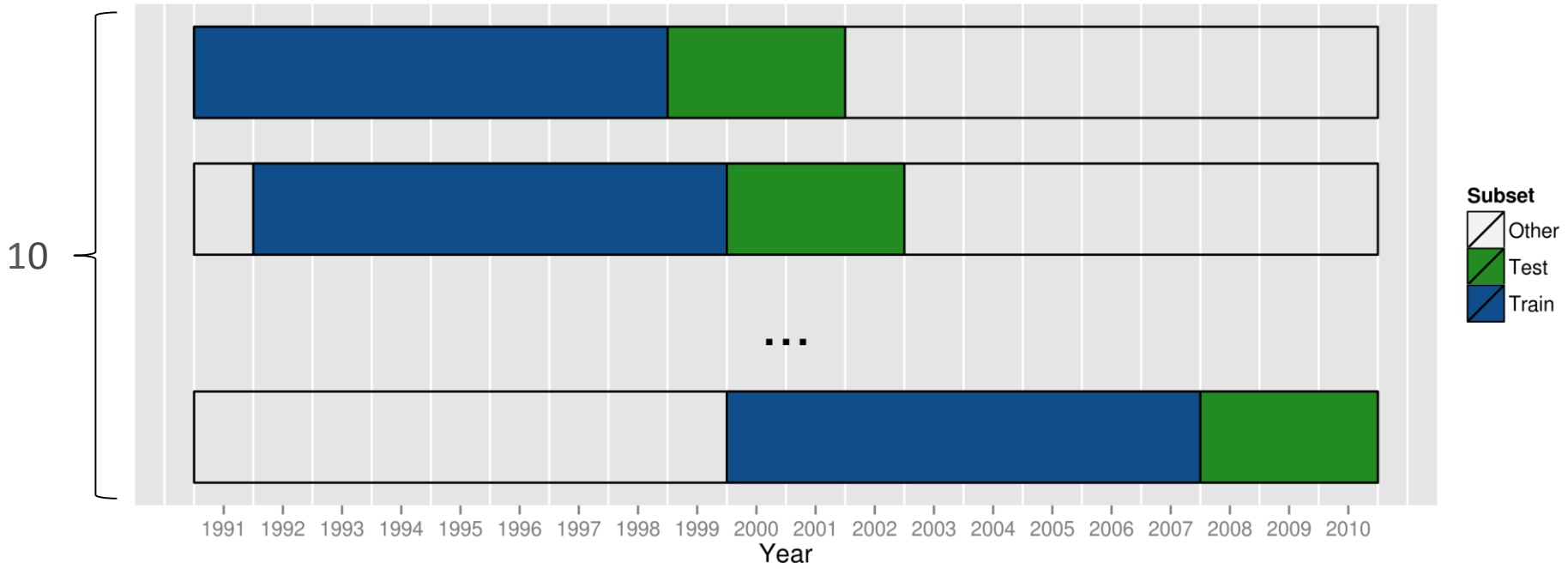
$$recall_R = \frac{\sum_{\phi(y_i)>t_R}(1+u_i)}{\sum_{\phi(y_i)>t_R}(1+\phi(y_i))}$$

$$f_1 = 2 \cdot \frac{precision_R \cdot recall_R}{precision_R + recall_R}$$



$$U_\phi^p(\hat{y}, y) = B_\phi(\hat{y}, y) - C_\phi^p(\hat{y}, y)$$

Experimental setup



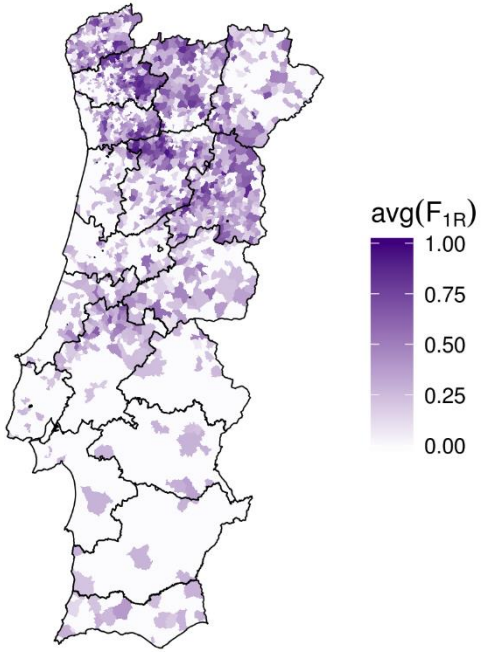
Results

Method	Re-sample	Model	Precision_R med±IQR	Precision_R p-val.	Recall_R med±IQR	Recall_R p-val.	F1-measure_R med±IQR	F1-measure_R p-val.
Propositional	None	RF	0.70 ± 0.13	(**0.002**)	0.22 ± 0.13	(**0.002**)	0.33 ± 0.13	(**0.002**)
		SVR	0.68 ± 0.10	(**0.002**)	0.49 ± 0.10	(**0.002**)	0.56 ± 0.10	(**0.002**)
	Under	RF	0.81 ± 0.13	(**0.002**)	0.67 ± 0.13	(**0.002**)	0.72 ± 0.13	(**0.002**)
		SVR	*0.84 ± 0.07*	(**0.002**)	*0.76 ± 0.07*	(0.01)	*0.80 ± 0.07*	(**0.002**)
Relational	None	RF	0.71 ± 0.12	(**0.002**)	0.18 ± 0.12	(**0.002**)	0.29 ± 0.12	(**0.002**)
		SVR	0.68 ± 0.09	(**0.002**)	0.50 ± 0.09	(**0.002**)	0.57 ± 0.09	(**0.002**)
	Under	RF	0.80 ± 0.09	(**0.002**)	0.58 ± 0.09	(**0.002**)	0.66 ± 0.09	(**0.002**)
		SVR	*0.85 ± 0.06*	(0.02)	*0.76 ± 0.06*	(0.04)	*0.80 ± 0.06*	(**0.002**)
Propositional + Relational	None	RF	0.72 ± 0.11	(**0.002**)	0.22 ± 0.11	(**0.002**)	0.33 ± 0.11	(**0.002**)
		SVR	0.70 ± 0.10	(**0.002**)	0.52 ± 0.10	(**0.002**)	0.59 ± 0.10	(**0.002**)
	Under	RF	0.80 ± 0.12	(**0.002**)	0.65 ± 0.12	(**0.002**)	0.70 ± 0.12	(**0.002**)
		SVR	**0.85 ± 0.06**	–	**0.77 ± 0.06**	–	**0.81 ± 0.06**	–

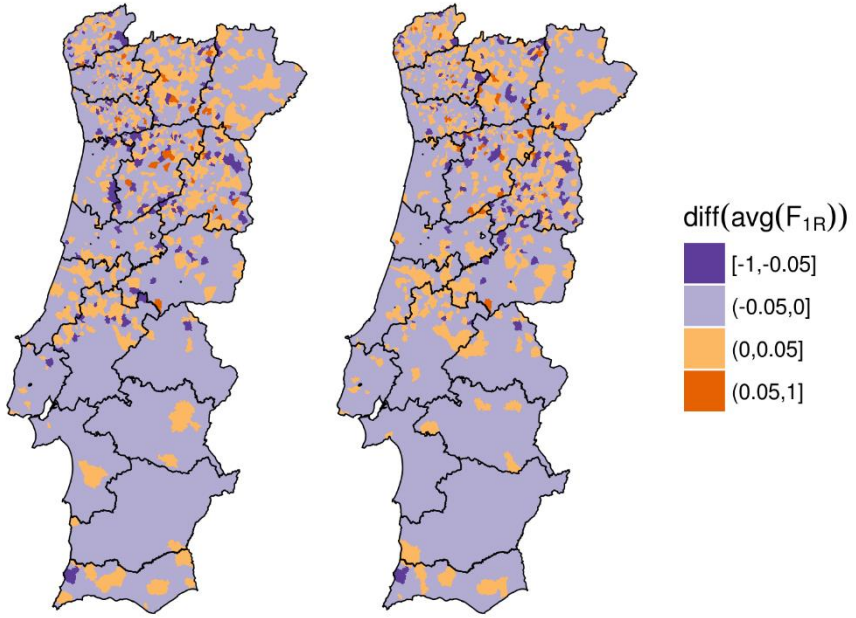
Difference in results

BEST F1-MEASURE:

PROPOSITIONAL + RELATIONAL



PROPOSITIONAL RELATIONAL



Computation time

Method	Re-sample	Model	**Pre-proc.**	**Training**	**Testing**	**Total time**
Propositional	None	RF	2.2e-3	5.8e-1 ± 6e-2	8.0e-4 ± 4e-4	5.8e-1
		SVR	1.7e-3	8.5e-3 ± 5e-4	6.7e-4 ± 7e-5	1.1e-2
	Under	RF	6.8e-3	2.6e-2 ± 6e-3	3.3e-4 ± 6e-5	3.3e-2
		SVR	3.1e-3	1.8e-4 ± 6e-5	2.1e-4 ± 4e-5	**3.5e-3**
Relational	None	RF	1.7	2.1e-1 ± 7e-2	3.6e-4 ± 7e-5	1.9
		SVR	1.7	2.0e-2 ± 1e-2	2.7e-3 ± 6e-4	1.7
	Under	RF	1.7	2.2e-2 ± 6e-3	5.0e-4 ± 4e-4	1.7
		SVR	1.7	6.0e-4 ± 1e-4	7.0e-4 ± 2e-4	*1.7*
Propositional + Relational	None	RF	1.7	1.5e-1 ± 2e-2	2.8e-4 ± 5e-5	1.9
		SVR	1.7	7.0e-2 ± 1e-2	6.0e-3 ± 2e-3	1.8
	Under	RF	1.7	1.9e-2 ± 7e-3	3.2e-4 ± 8e-5	1.7
		SVR	1.7	1.0e-3 ± 3e-4	1.0e-3 ± 1e-3	*1.7*

Conclusion

Summary

- Propositional and relational approach achieve comparable results
  - Propositional approach is faster
  - Relational approach works well though optimized for classification
- Combination of both approaches works best
- Under-sampling for regression greatly improved results

Future research directions

- Explore other propositional approaches:
  - Use clustering to select neighbourhoods as proposed by Appice *et al.* (2013).
- Explore other relational approaches:
  - Use graphical models such as Markov Logic Networks.
- Compare results in different domains to generalise our findings.

Thank you!

QUESTIONS?