

Assessing the effect of date and sequence data in phylodynamics

Leo A. Featherstone

2022-06-21

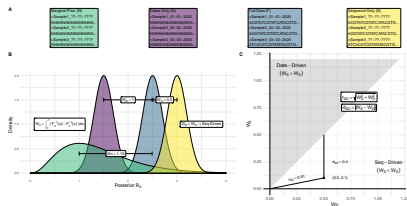
About me

- ▶ PhD candidate (2021-2024) @ University of Melbourne Peter Doherty Institute



- ▶ Working with Dr. Sebastian Duchene, Dr. Timothy Vaughan, and Prof. Ben Phillips
- ▶ Twitter: @LeoPhylostone

Core content

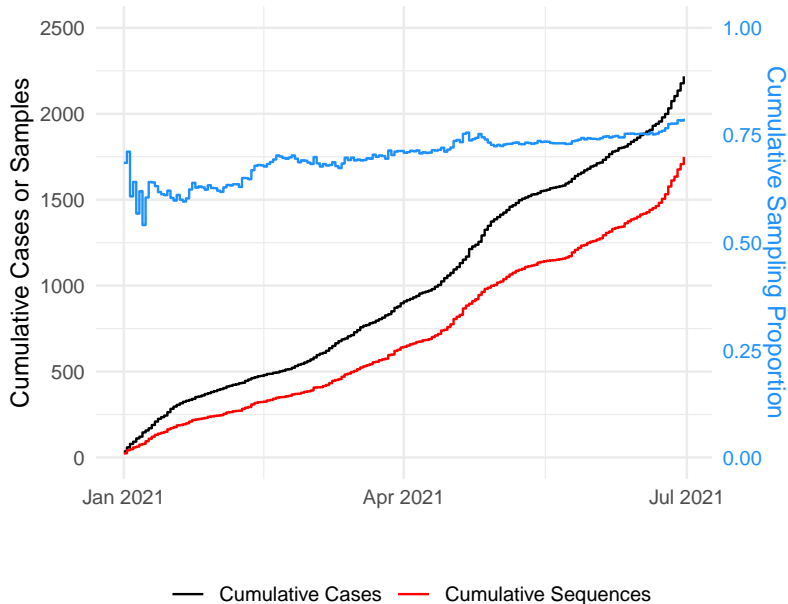


- ▶ Preprint: Assessing the effects of date and sequence data in phylodynamics
- ▶ Under birth-death with serial sampling (λ, δ, p)

Talk breakdown

1. Origins of project
2. Quantify data and sequence effects under birth-death model
3. Validation
4. Context

Project origins (SARS-CoV-2 in Australia)



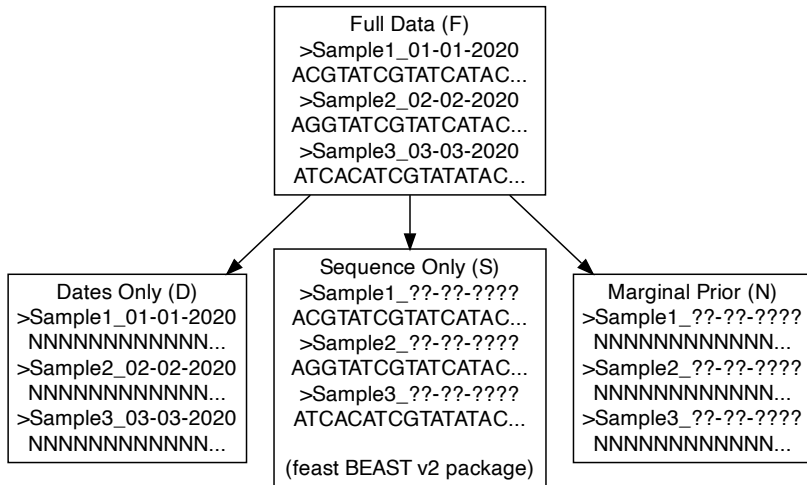
Project origins: literature

- ▶ Volz et al. 2014 → sampling times heavily influential under BD
- ▶ Featherstone et al. 2021 → sampling times help prevent bias

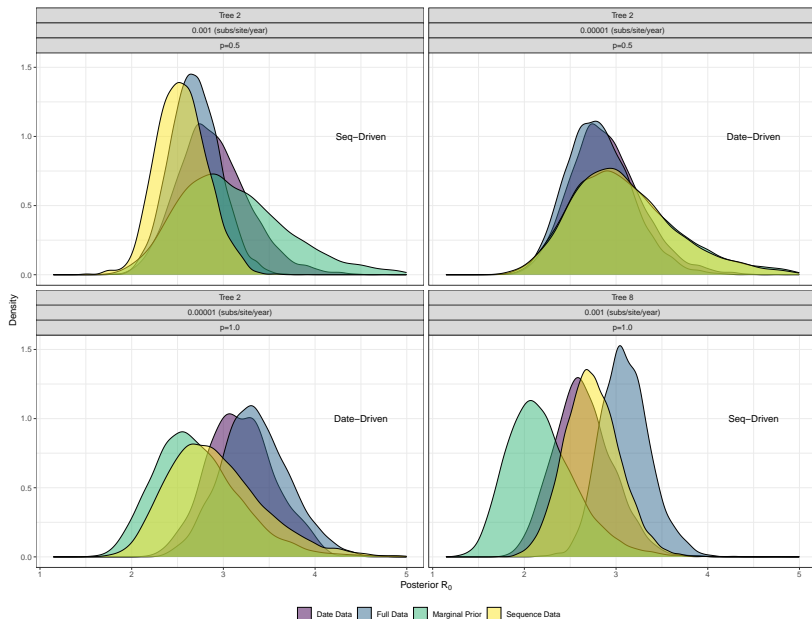
Sapareting sequence and dates

	Dates Included	Dates Excluded
Sequence Included	Combined effect	Sequence Effects
Sequence Excluded	Date effects	Marginal Prior

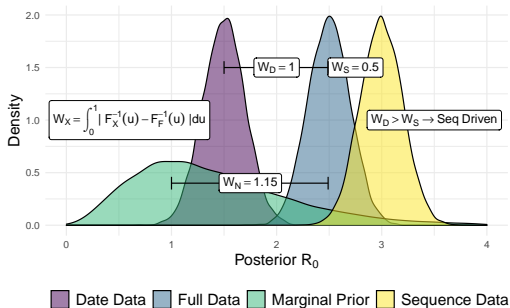
Separating sequence and dates: an example



Output

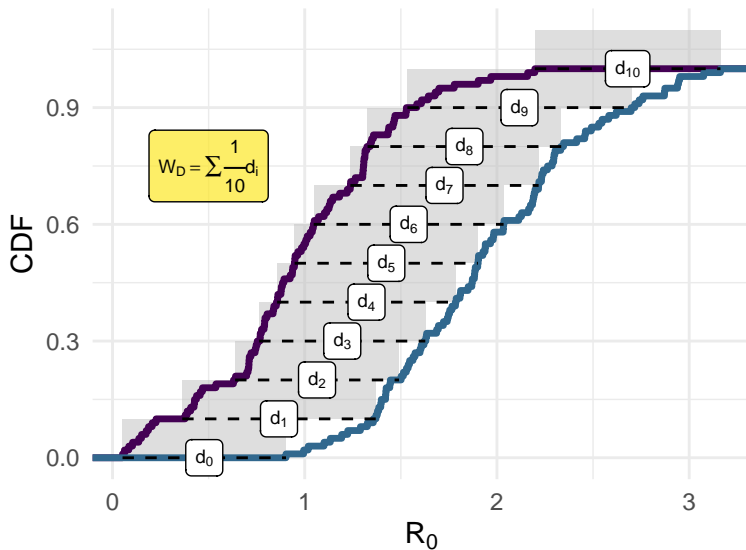


Comparing posterior signal cont.

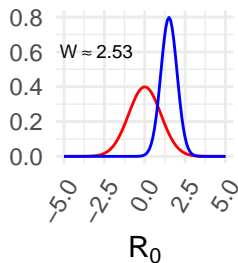
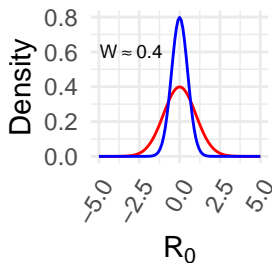
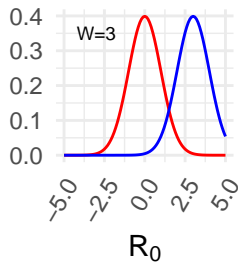
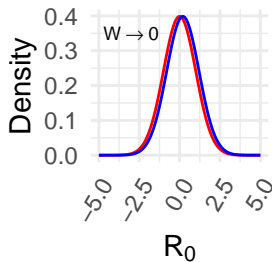


Wasserstein Intuition (in 1D)

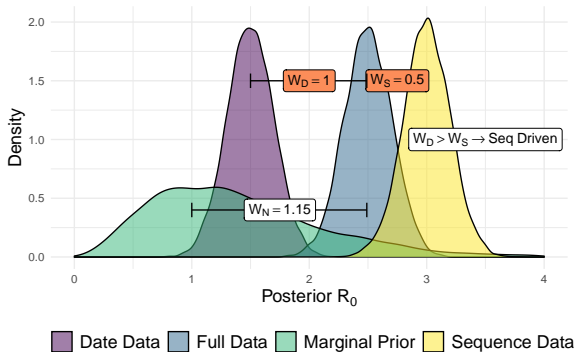
How much **work** to push one distribution into another?



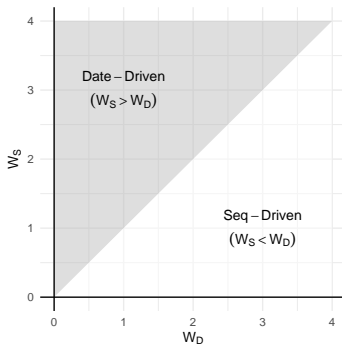
Some examples



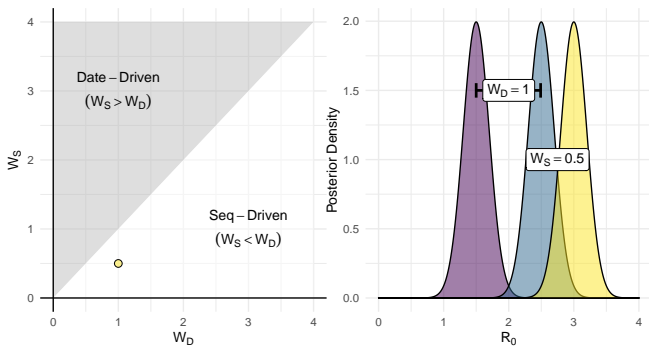
Classification



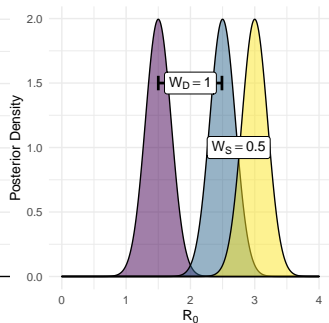
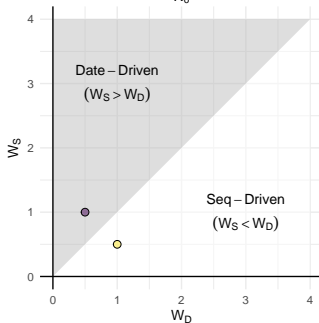
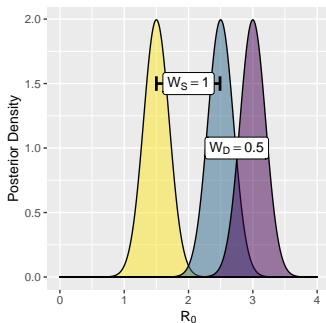
Visualizing Wasserstein Data



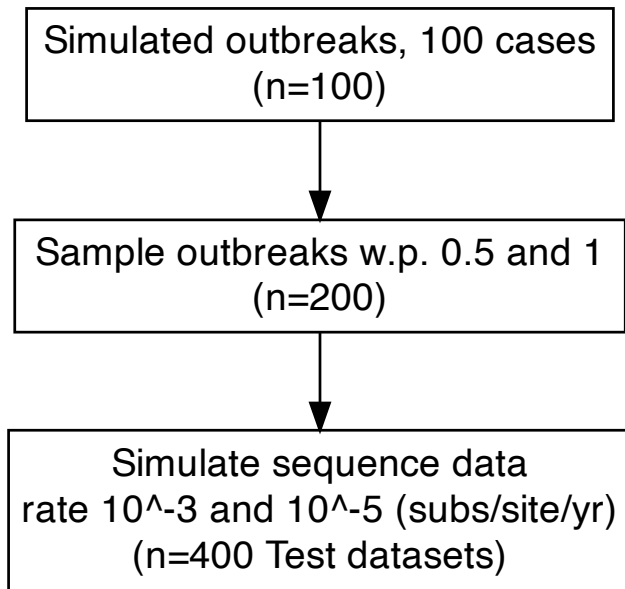
Visualizing Wasserstein Data



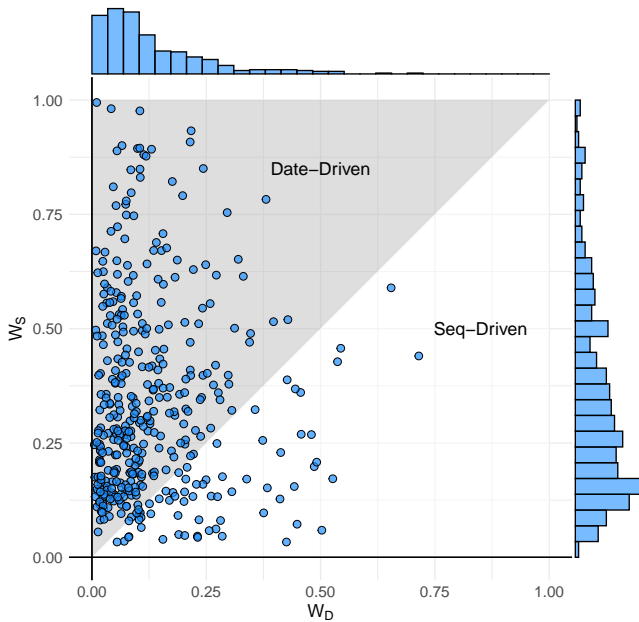
Visualizing Wasserstein Data



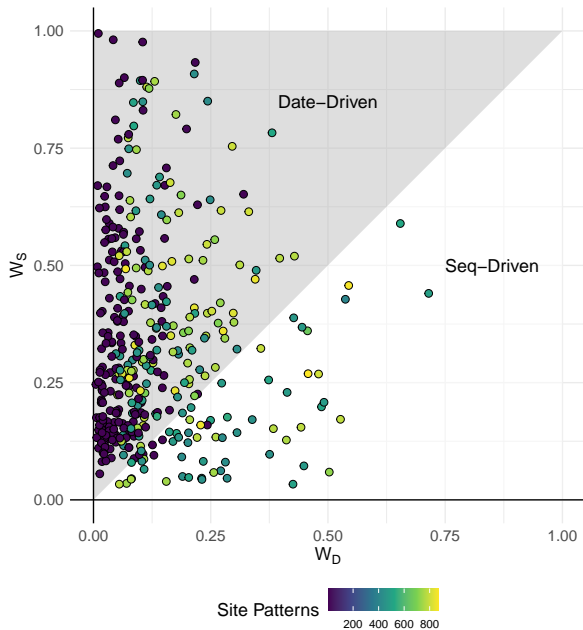
Validation: Simulation study



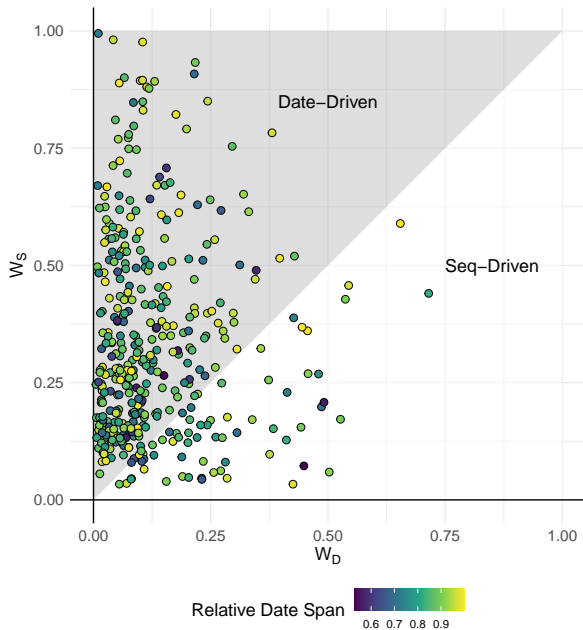
On the plane



Effects of Sequence Patterns



Effects of Date Span



Method summary

1. Separate date and sequence data

Method summary

1. Separate date and sequence data
2. Analyse both under a birth-death

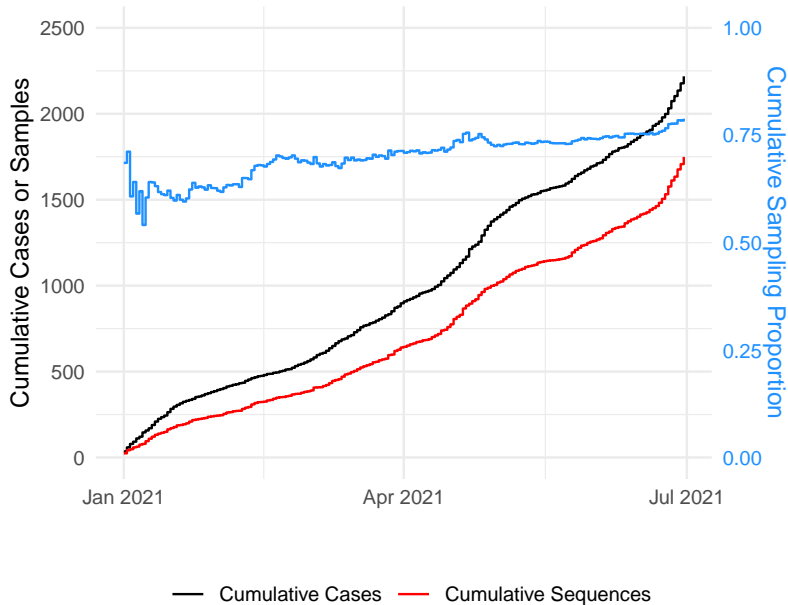
Method summary

1. Separate date and sequence data
2. Analyse both under a birth-death
3. Compare posteriors with Wasserstein metric to find drivers

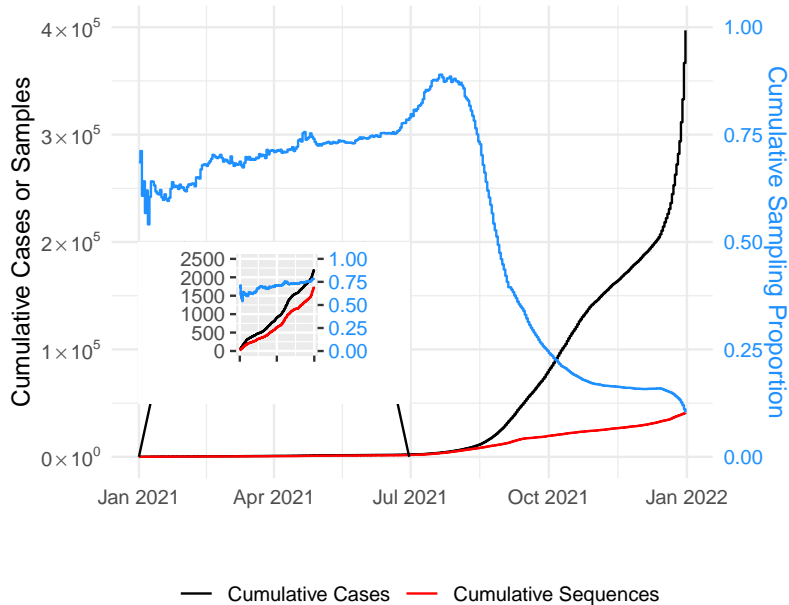
Method summary

1. Separate date and sequence data
2. Analyse both under a birth-death
3. Compare posteriors with Wasserstein metric to find drivers
4. Driver-classification agrees with prior literature on BD

Results in context



Results in context



Acknowledgements

- ▶ cEvo lab & Swissnex

Acknowledgements

- ▶ cEvo lab & Swissnex
- ▶ Lionell Gell Foundation

Acknowledgements

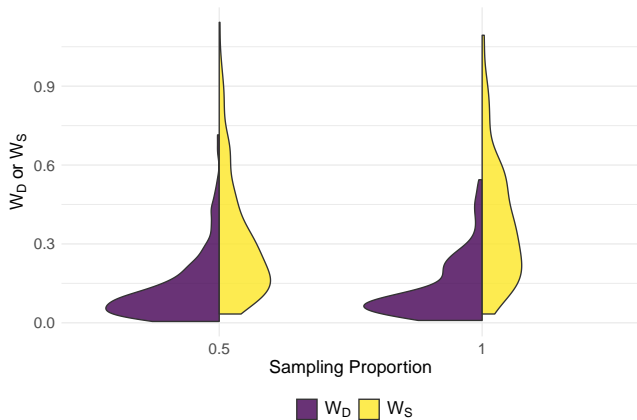
- ▶ cEvo lab & Swissnex
- ▶ Lionell Gell Foundation
- ▶ Wonderful supervisors!

All questions welcome

References

- 1) Volz Erik M. and Frost Simon D. W. 2014 Sampling through time and phylodynamic inference with coalescent and birth–death models J. R. Soc. Interface. 11: 20140945–20140945 <http://doi.org/10.1098/rsif.2014.0945> Section A
- 2) Featherstone, LA, Di Giallonardo, F, Holmes, EC, Vaughan, TG, Duchêne, S. Infectious disease phylodynamics with occurrence data. Methods Ecol Evol. 2021; 12: 1498– 1507. <https://doi.org/10.1111/2041-210X.13620>

Sampling effects

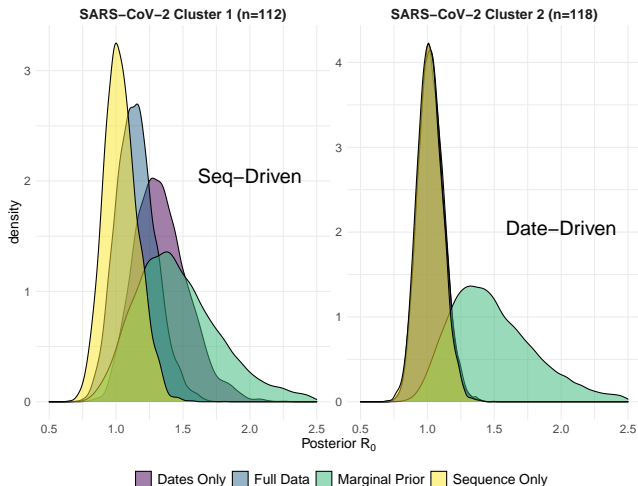


- ▶ Two SARS-CoV-2 clusters from 2020 in Victoria, Australia

Empirical Data

- ▶ Two SARS-CoV-2 clusters from 2020 in Victoria, Australia
- ▶ Fixed parameters except origin and R_0

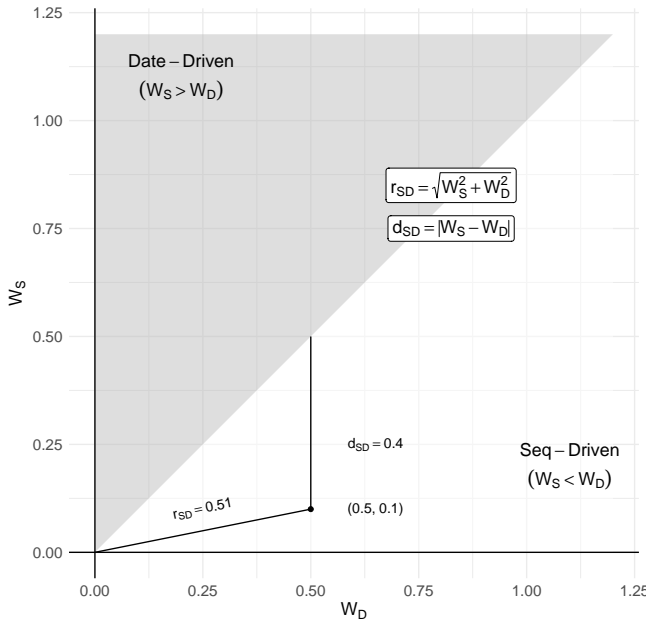
Empirical Data: Posteriors



Empirical Data: Values

	Cluster 1	Cluster 2
n	112	188
Classification	Seq-Driven	Date-Driven
Site Patterns	183	126
r_{SD}	0.223	0.009
d_{SD}	0.078	0.008
W_D	0.192	0.001
W_S	0.114	0.009
W_N	0.325	0.481

Strength of Classification



Error

