Supplementary Material for Deep Learning on Chaos Game Representation for Proteins

Hannah F. Löchel, Dominic Eger, Theodor Sperlea and Dominik Heider

March 2019

Abstract

This section contains the AUCs for the test- training splits for the best hidden layer and FCGR configuration (Figure 1 and Figure 2), the mean errors (Figure 3 and Figure 4) and $< \phi, \delta >$ diagrams [Armano and Giuliani, 2018] for the encodings used in Heider et al. [2011] for all drugs (Figure 5), as well as for the CGR encoding for all configurations used in the current study (left column of Figure 7 - Figure 33). The difference between the average FCGR of positive and negative sequences in the datasets are shown in the middle column of Figure 7 to Figure 33, and the significance of the differences in log(p) values as calculated using a Bonferroni-corrected t-test in the right column. Blue pixels are significantly different between positively and negatively labeled sequences. Gray pixels contain no values or zeros.

References

Giuliano Armano and Alessandro Giuliani. A two-tiered 2d visual tool for assessing classifier performance. *Information Sciences*, jun 2018. doi: 10.1016/j.ins.2018.06.052. URL `https://doi.org/10.1016%2Fj.ins.2018.06.052`.

Dominik Heider, Jens Verheyen, and Daniel Hoffmann. Machine learning on normalized protein sequences. *BMC research notes*, 4(1):94, 2011.

Dominik Heider and Daniel Hoffmann. Interpol: An r package for preprocessing of protein sequences. *BioData mining*, 4(1):16, 2011.

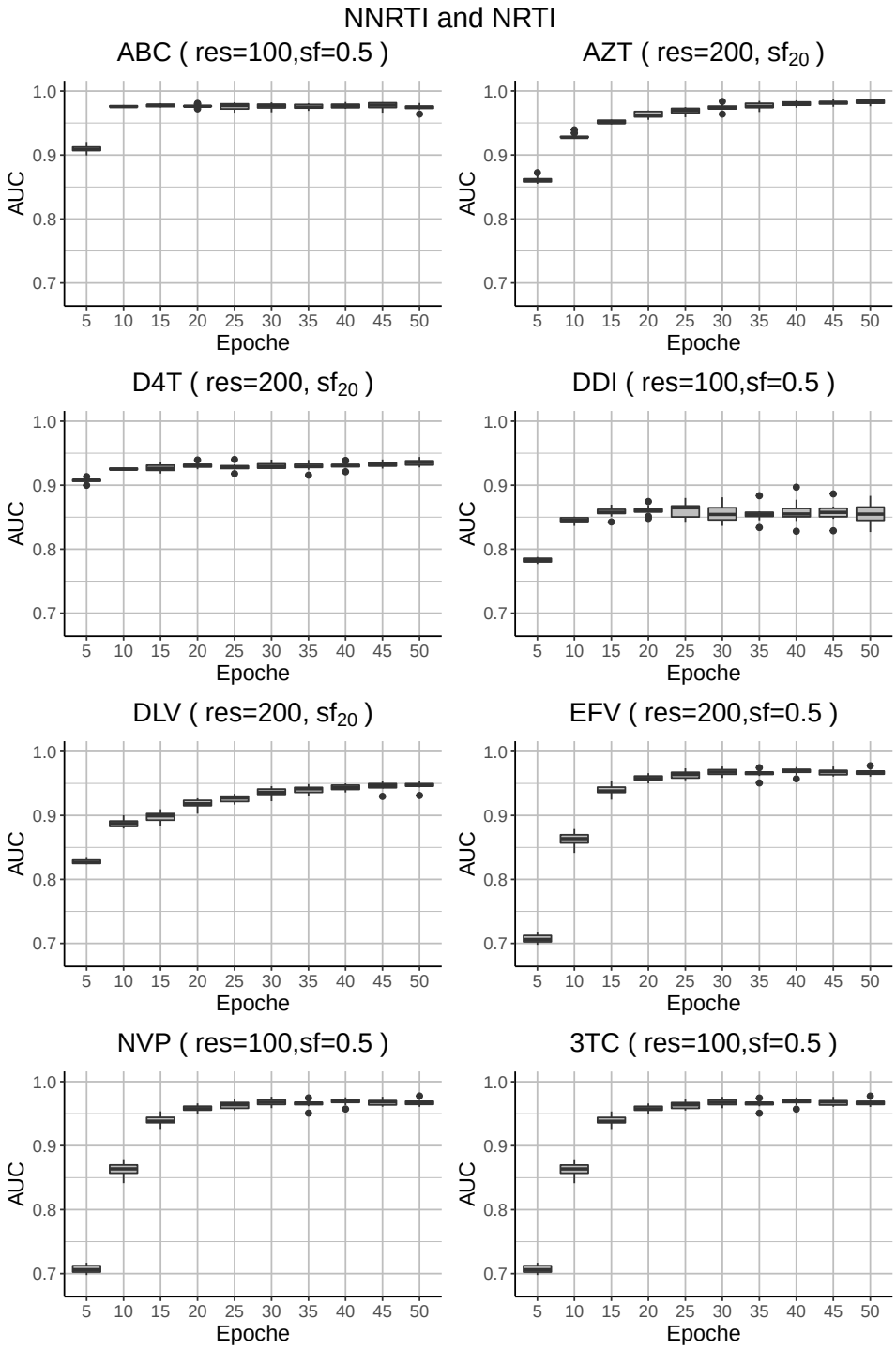

Figure 1: AUCs of NNRTIs and NRTIs for test- trainingsplits within best hiddenlayer and FCGR configurations

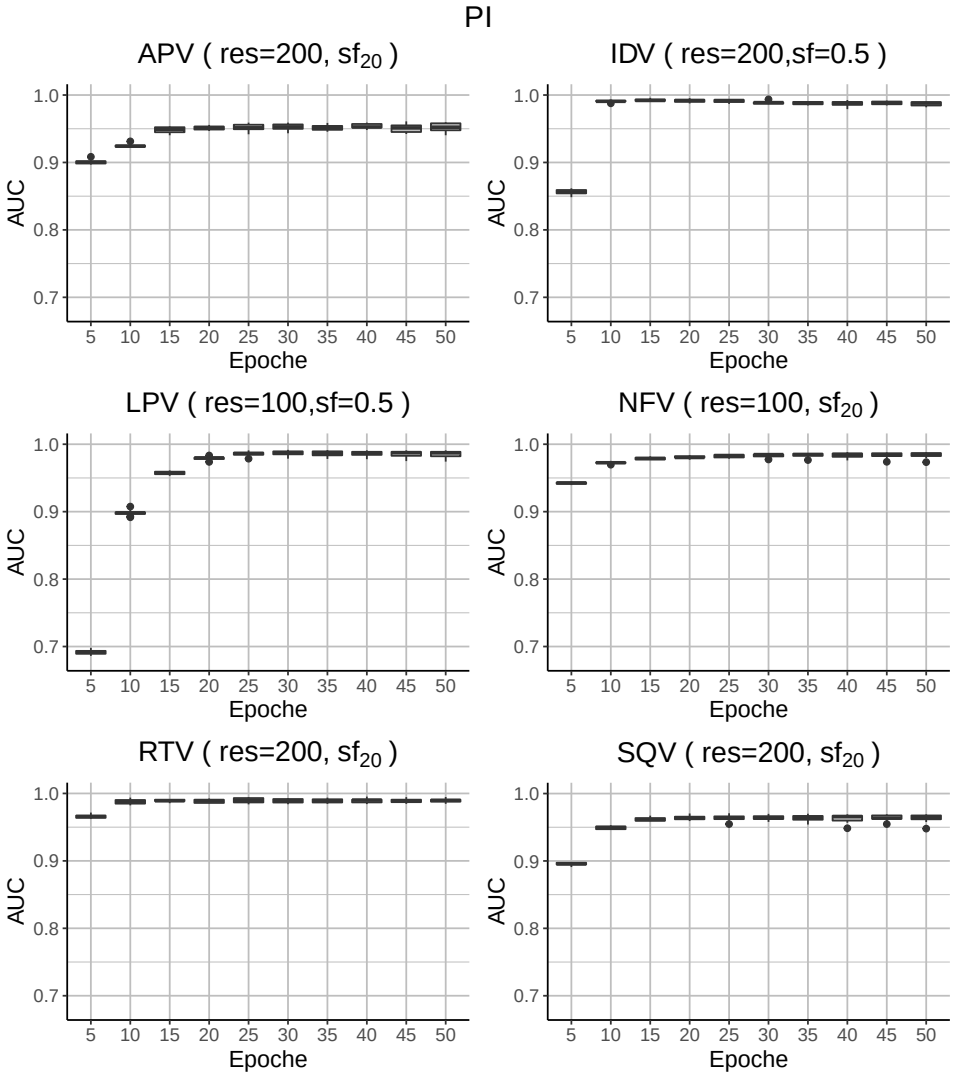

Figure 2: AUCs of PIs for test- trainingsplits within best hiddenlayer and FCGR configurations

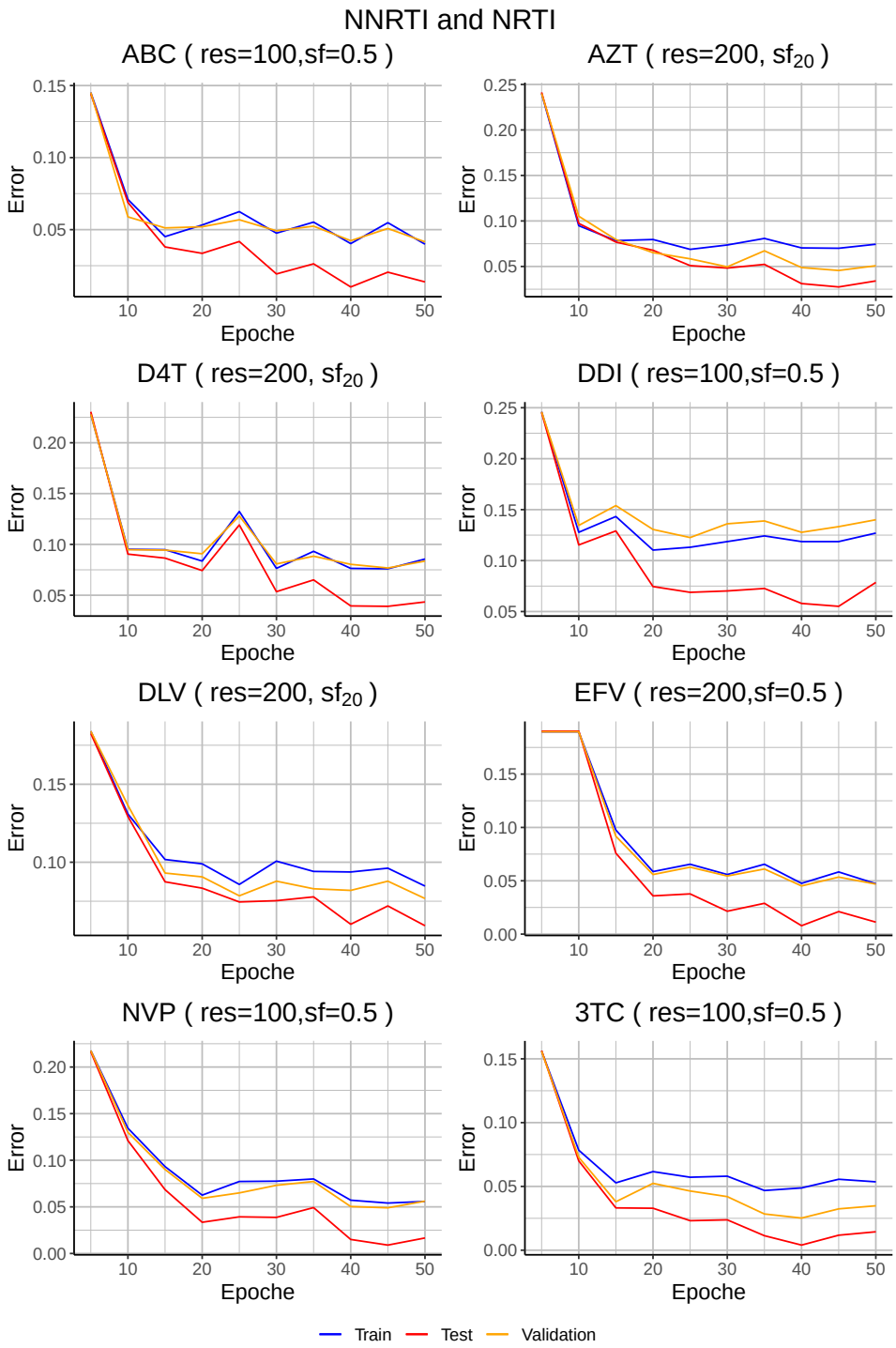

Figure 3: Mean error of NNRTIs and NRTIs for test- trainingsplits within best hiddenlayer and FCGR configurations

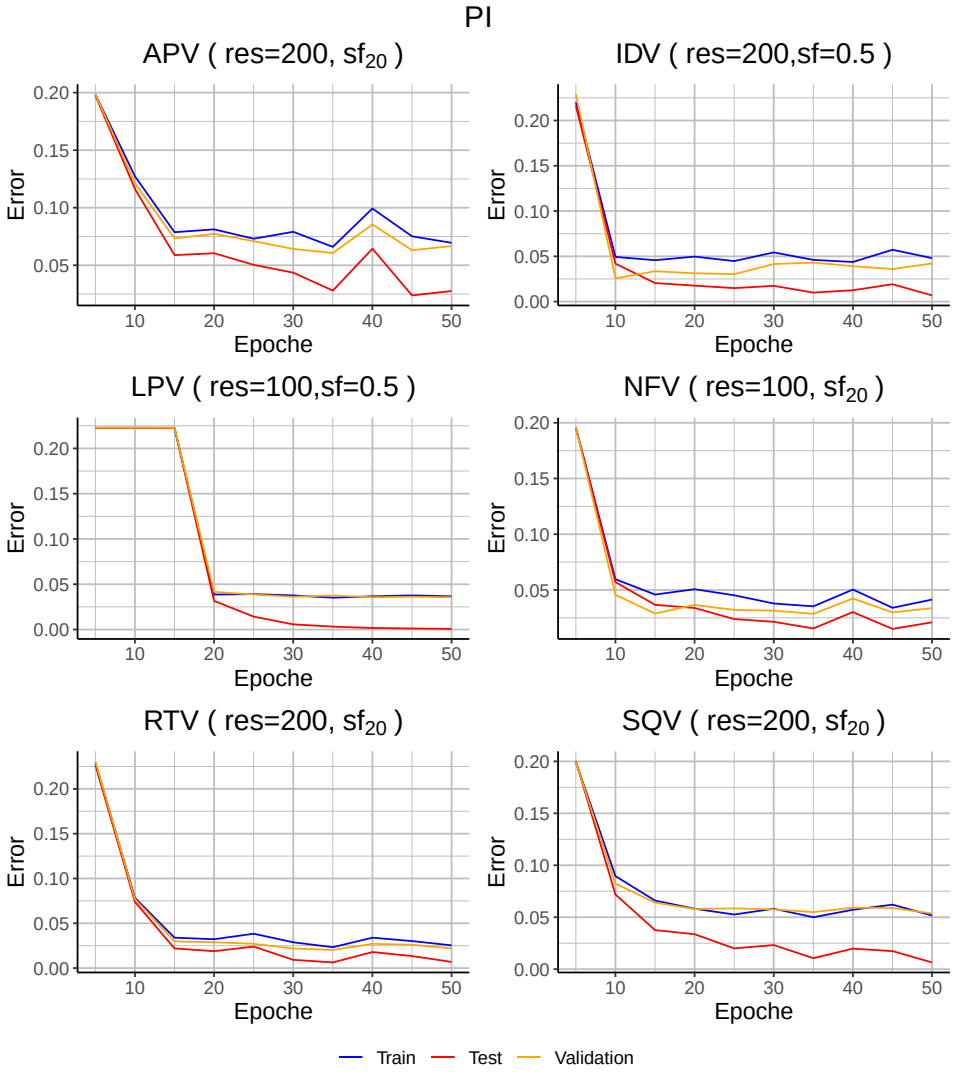

Figure 4: Mean error of PIs for test- trainingsplits within best hiddenlayer and FCGR configurations

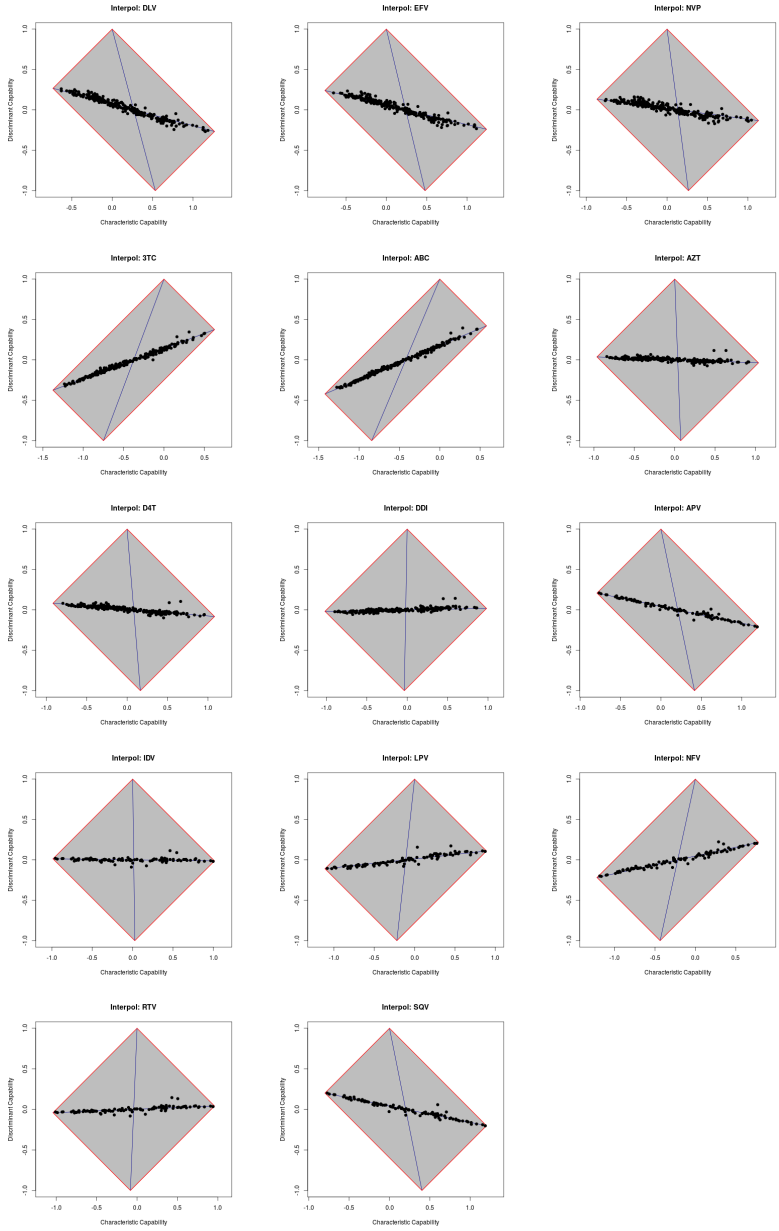

Figure 5: $< \phi, \delta >$ diagrams for feature encodings calculated using the R package Interpol [Heider and Hoffmann, 2011] as described in Heider et al. [2011]

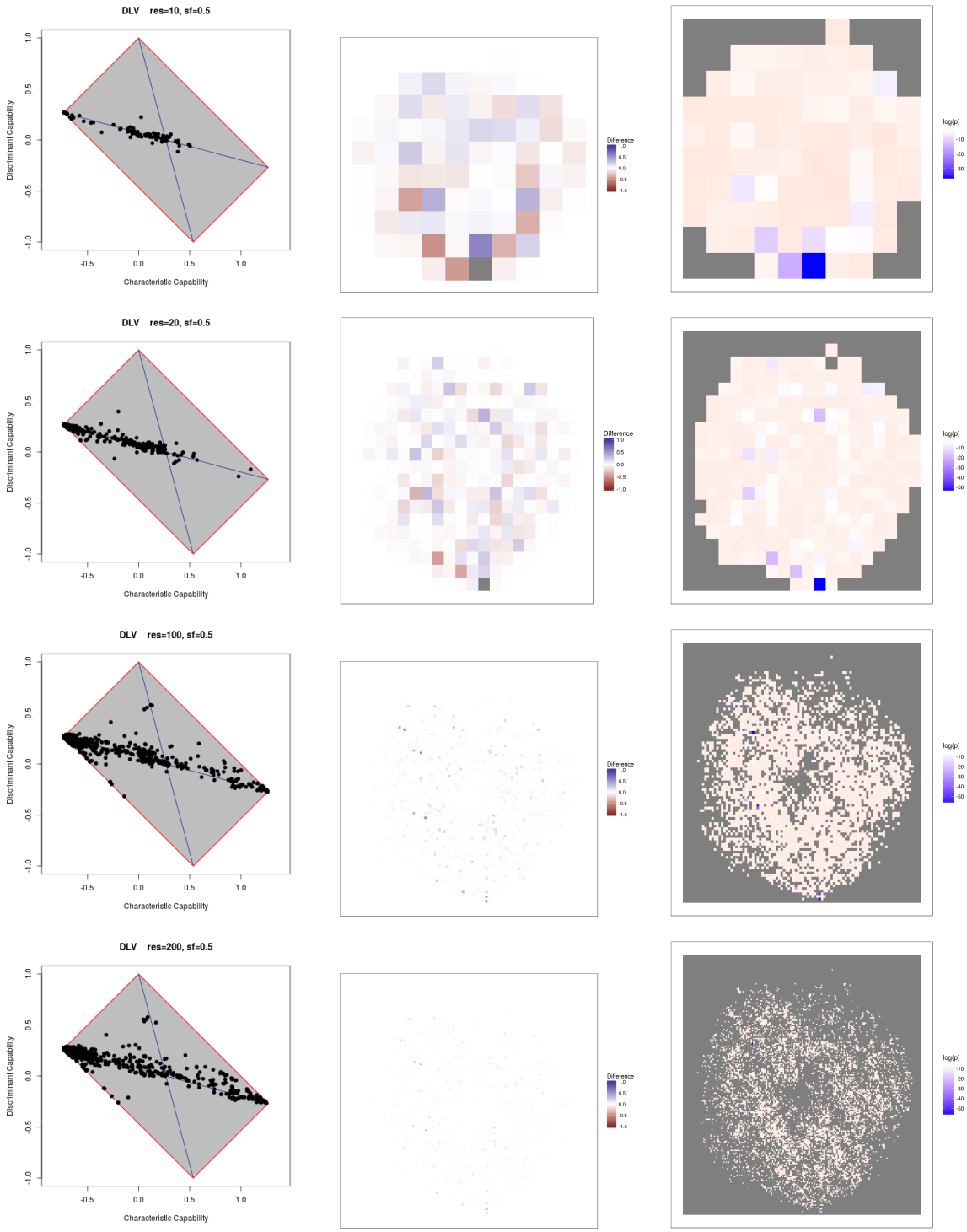

Figure 6: DLV, $sf = 0.5$

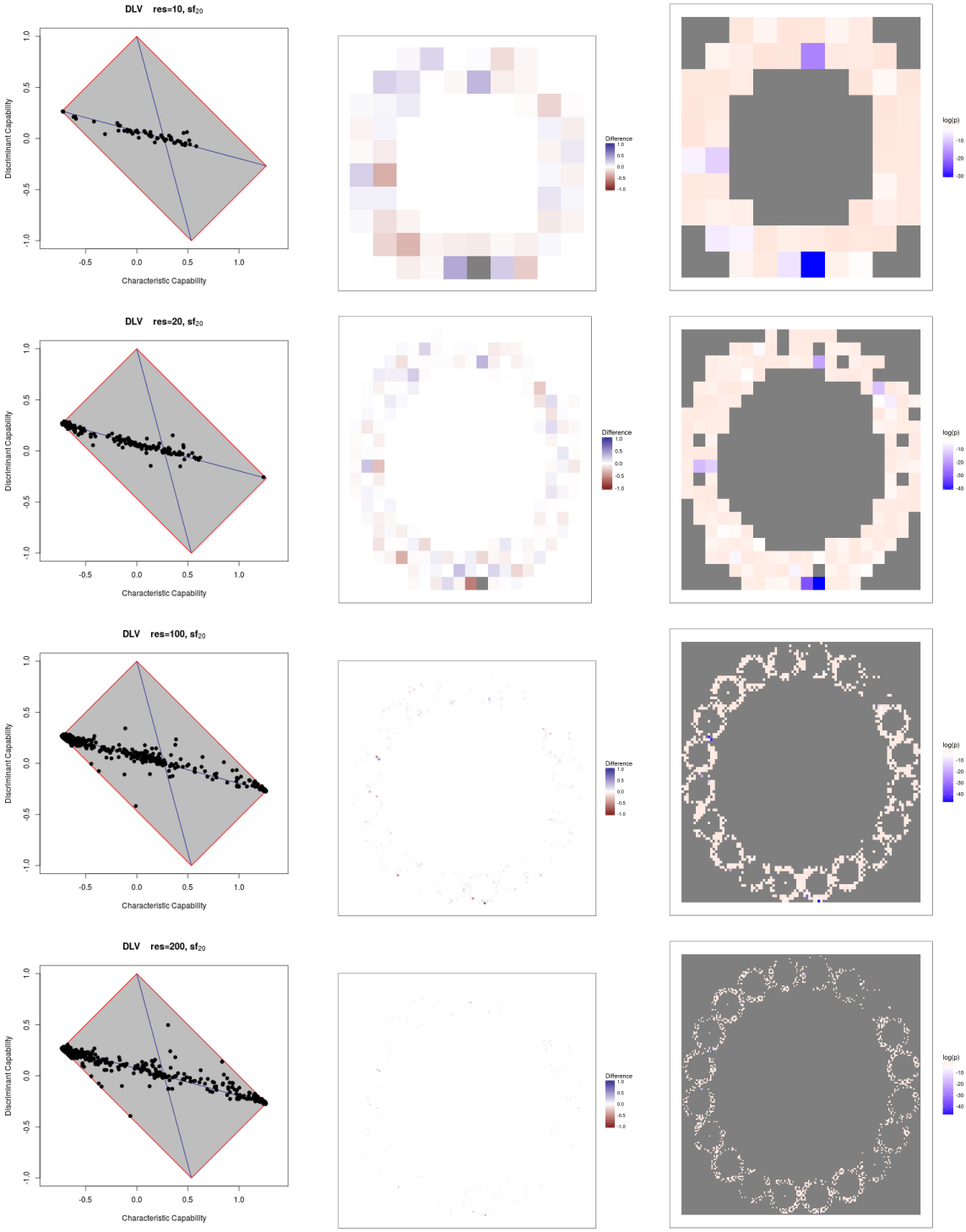

Figure 7: DLV, sf_{20}

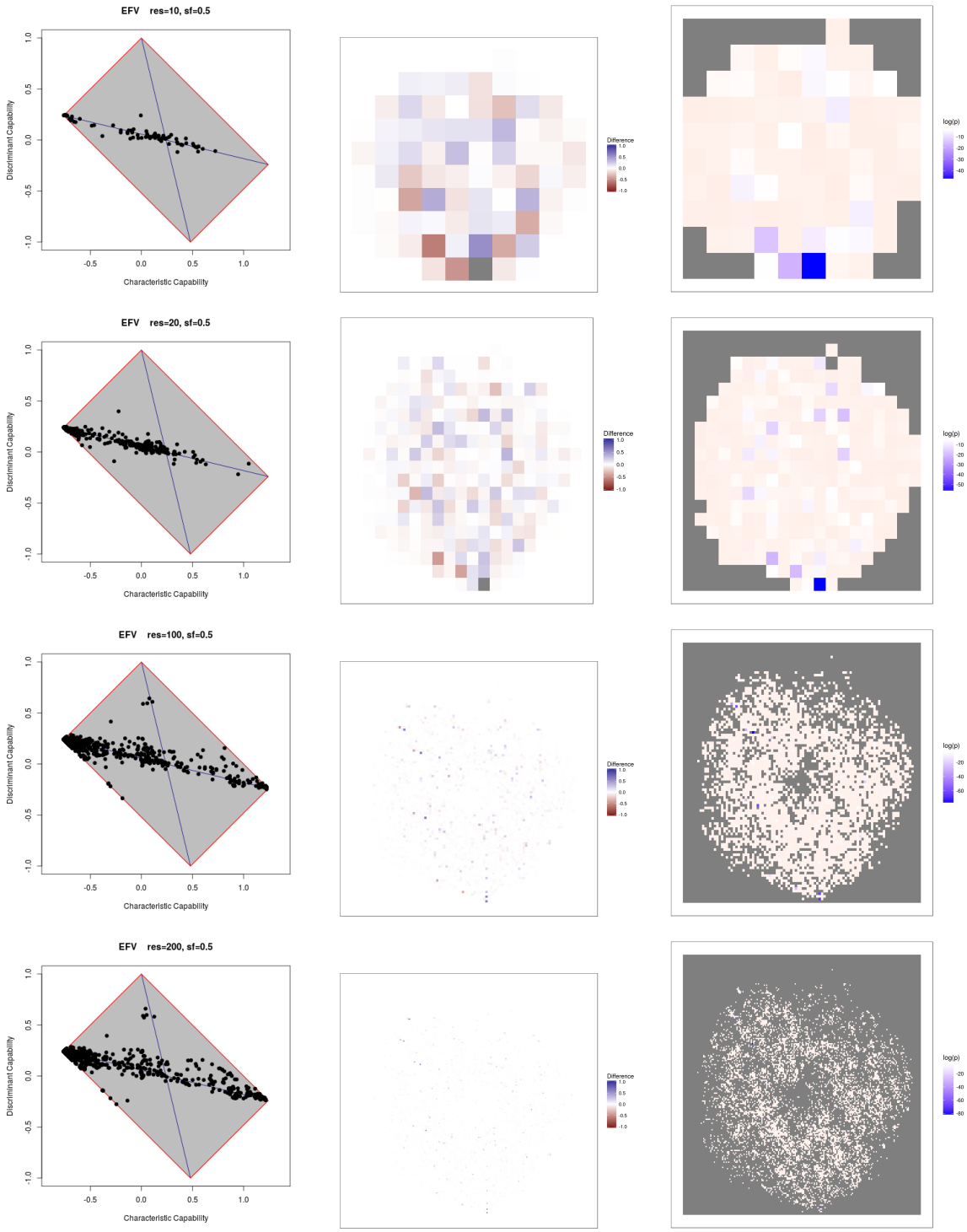

Figure 8: EFV, $sf = 0.5$

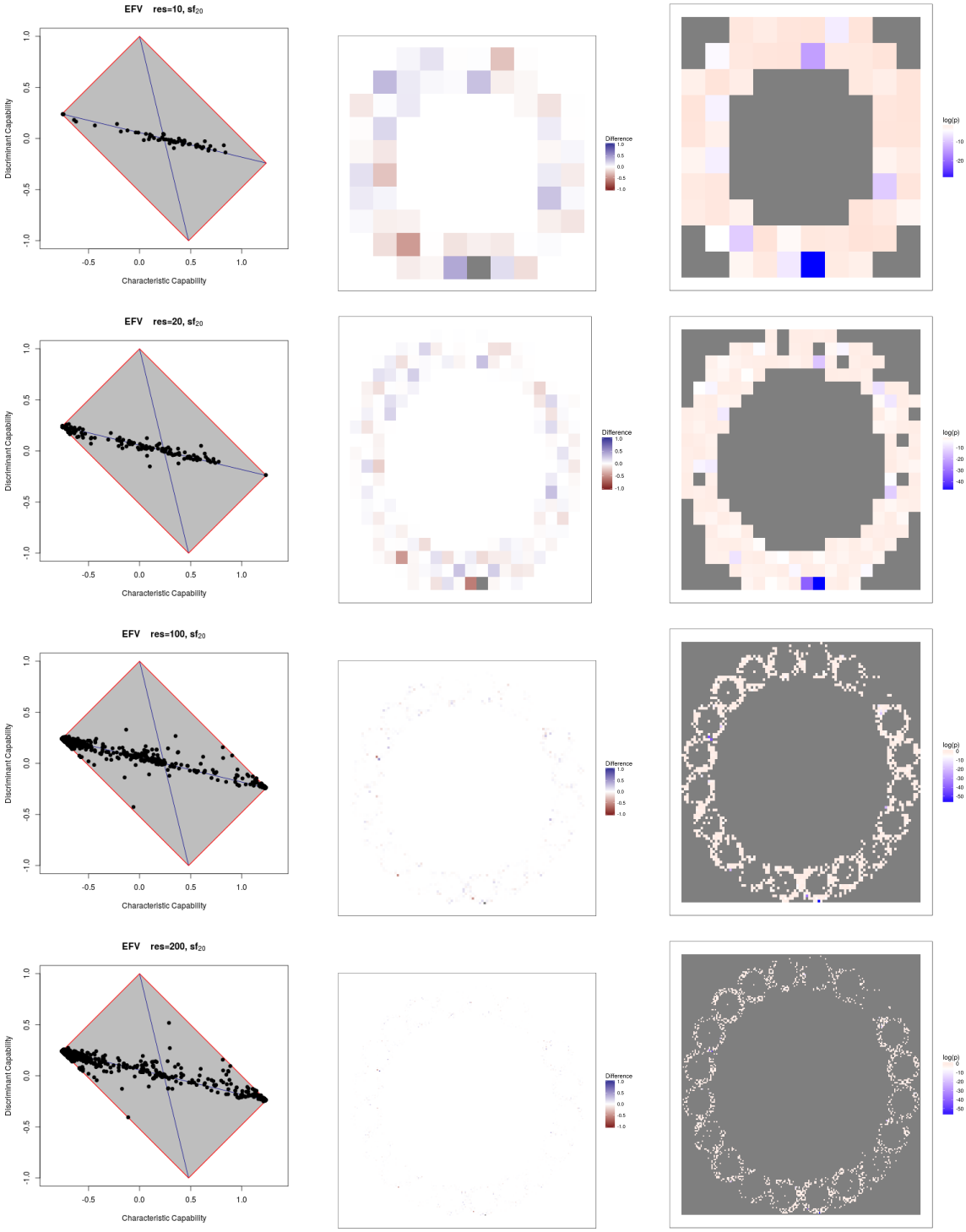

Figure 9: EFV, sf_{20}

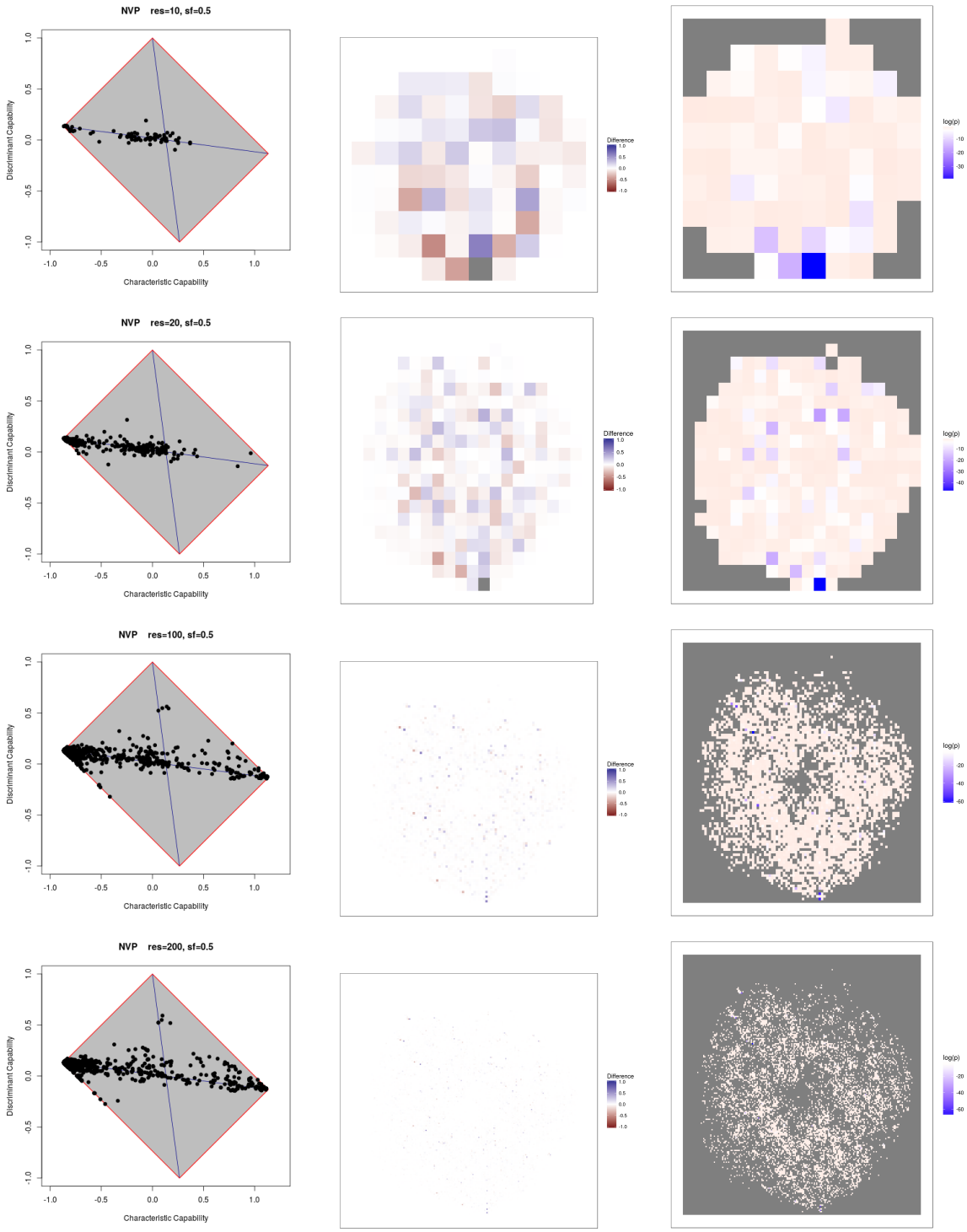

Figure 10: NVP, $sf = 0.5$

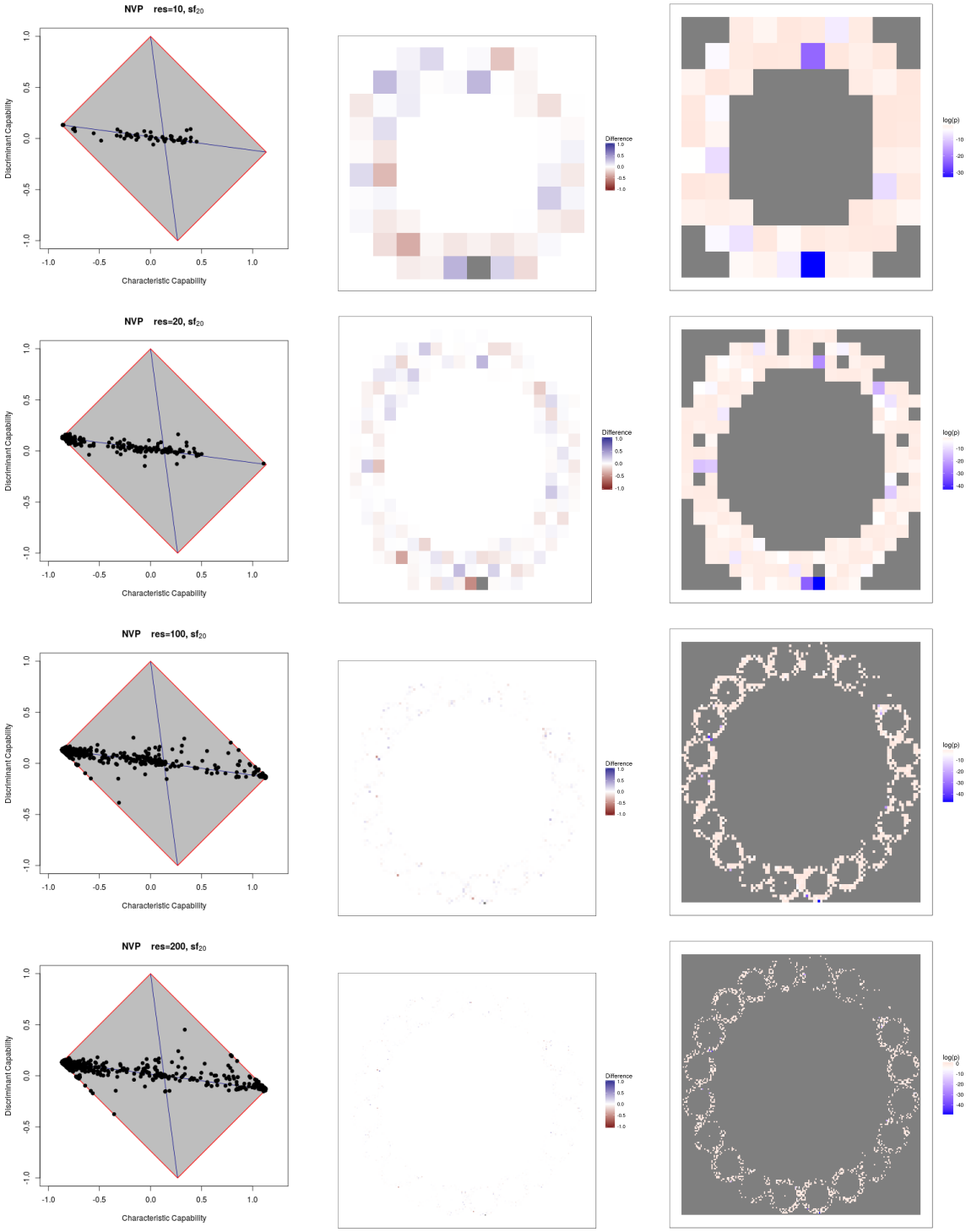

Figure 11: NFV, sf_{20}

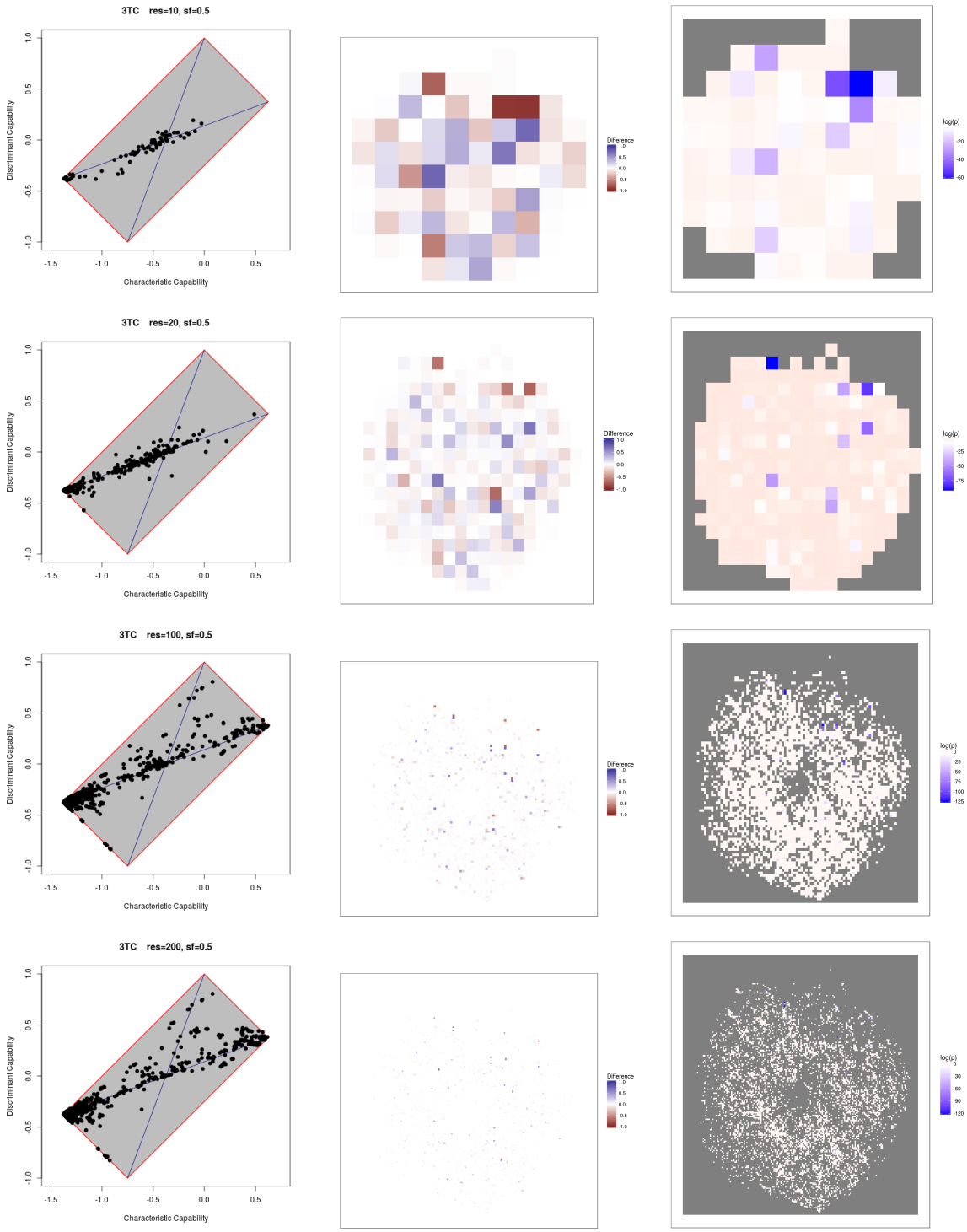

Figure 12: 3TC, $sf = 0.5$

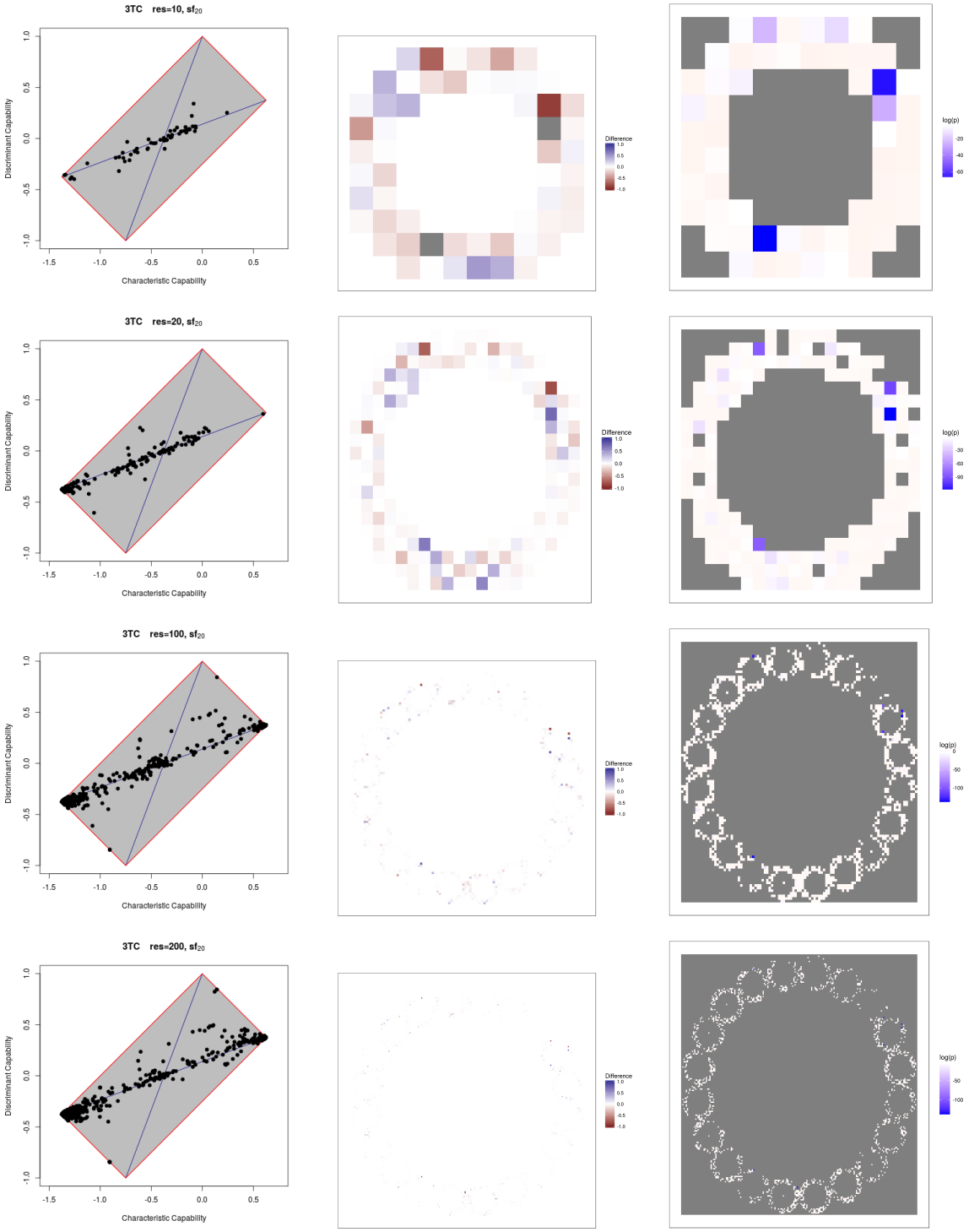

Figure 13: 3TC, sf_{20}

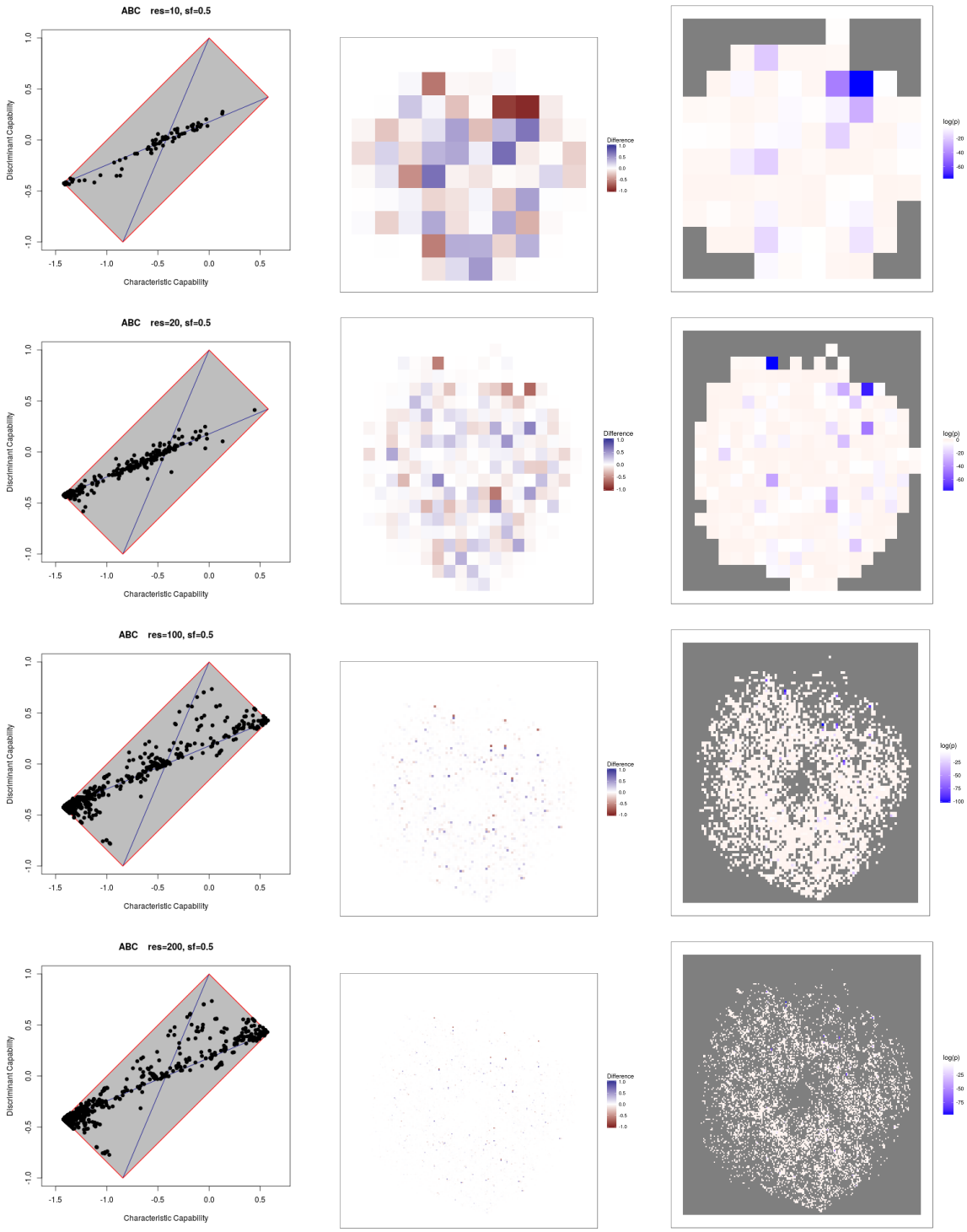

Figure 14: ABC, $sf = 0.5$

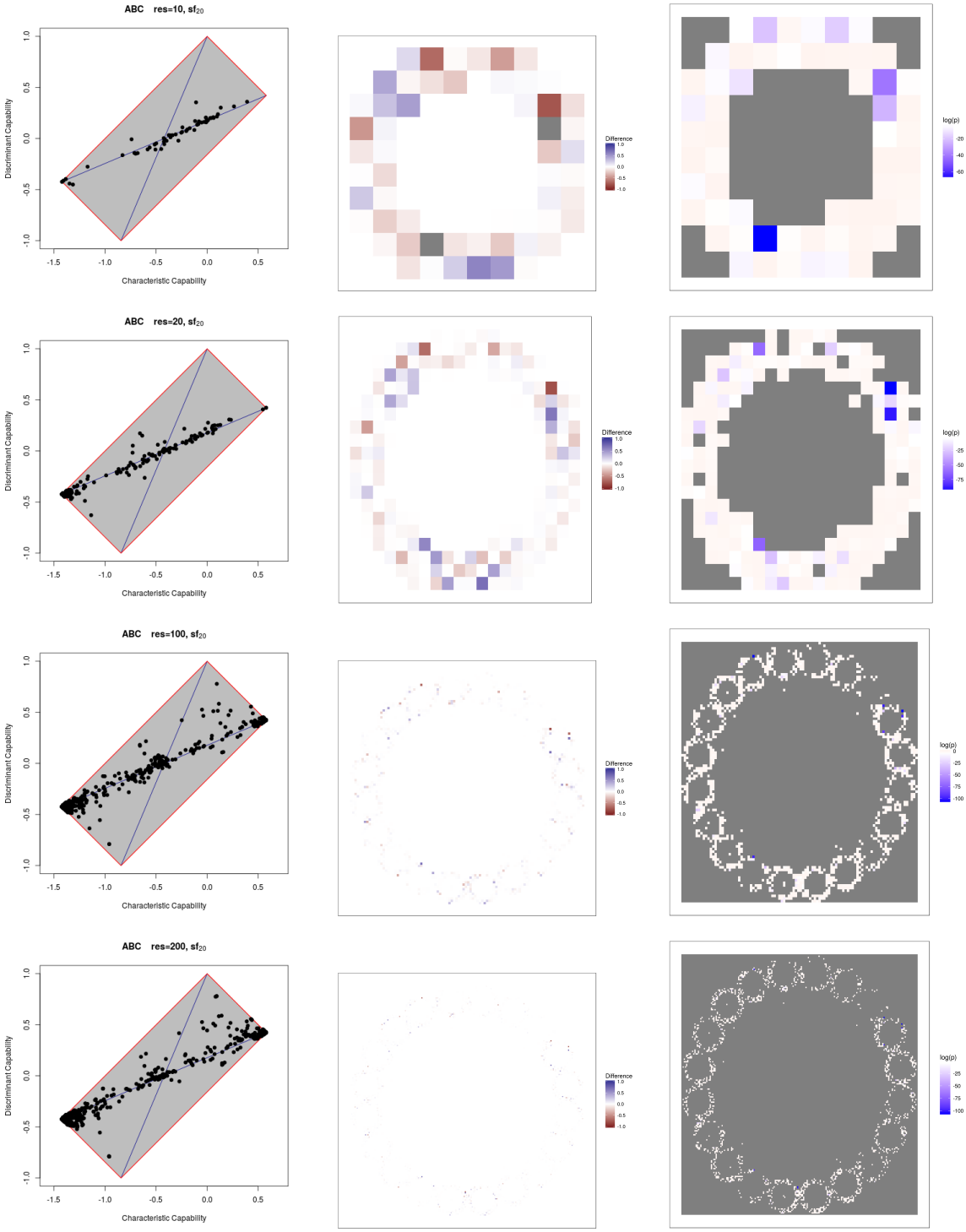

Figure 15: ABC, sf_{20}

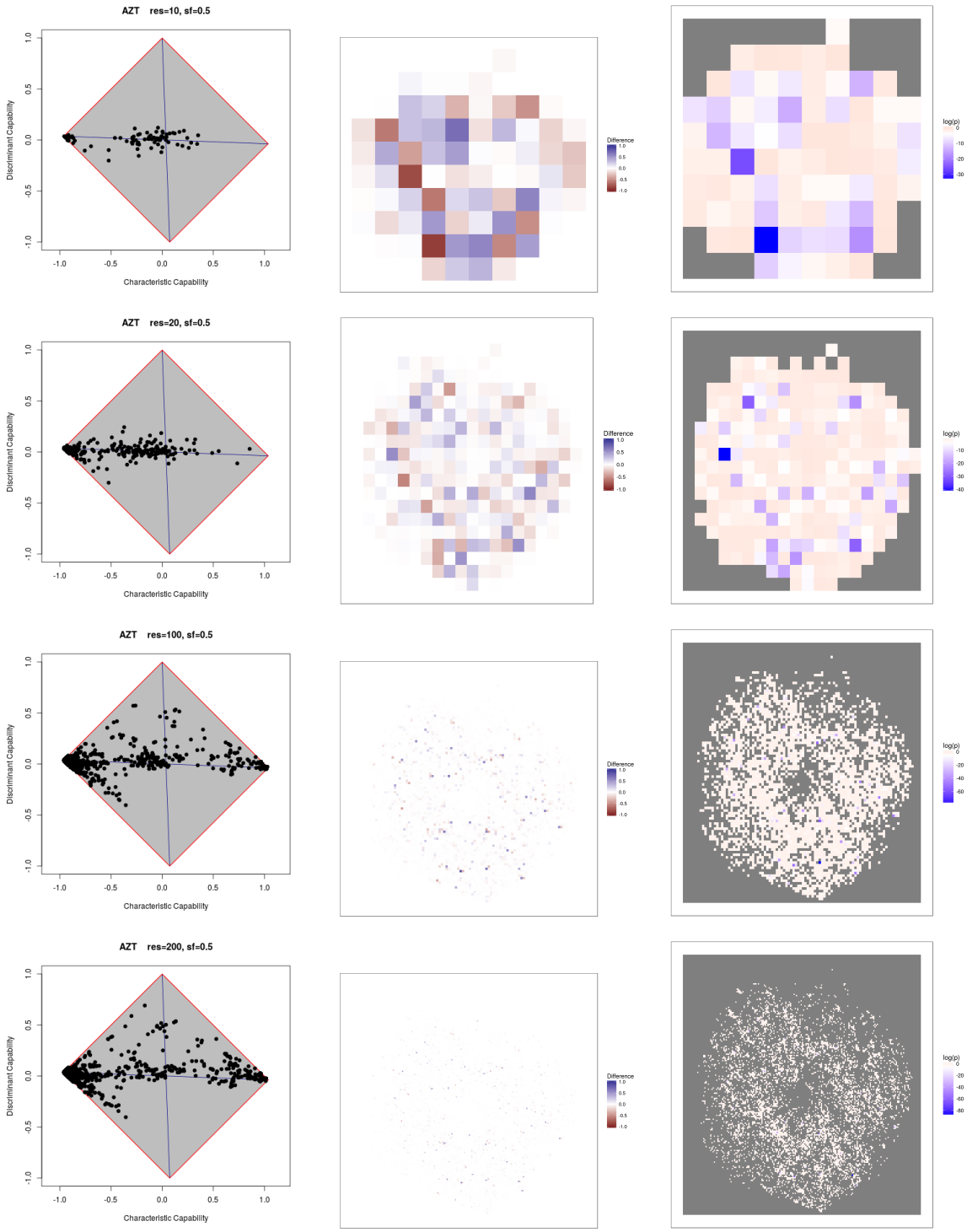

Figure 16: AZT, $sf = 0.5$

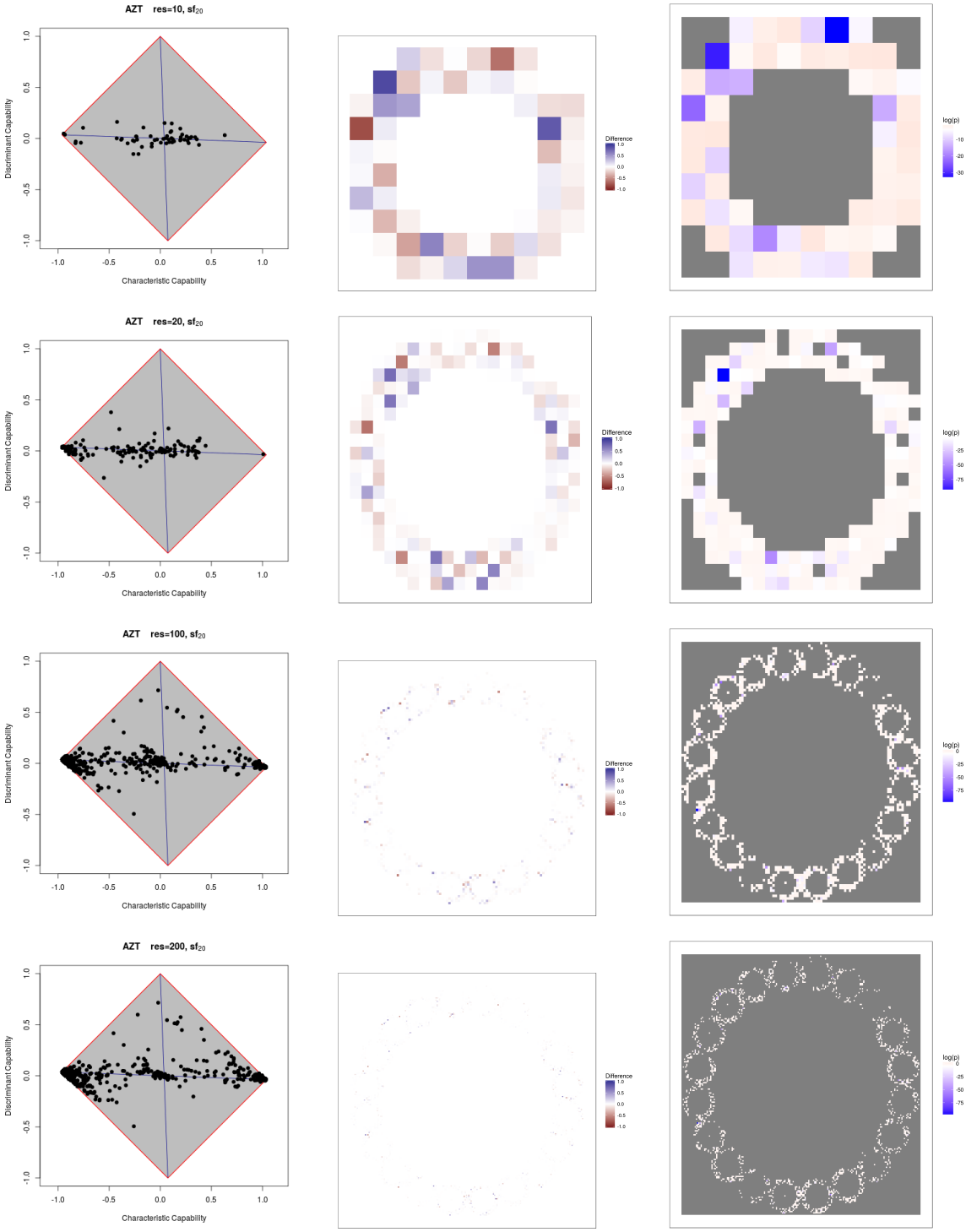

Figure 17: AZT, sf_{20}

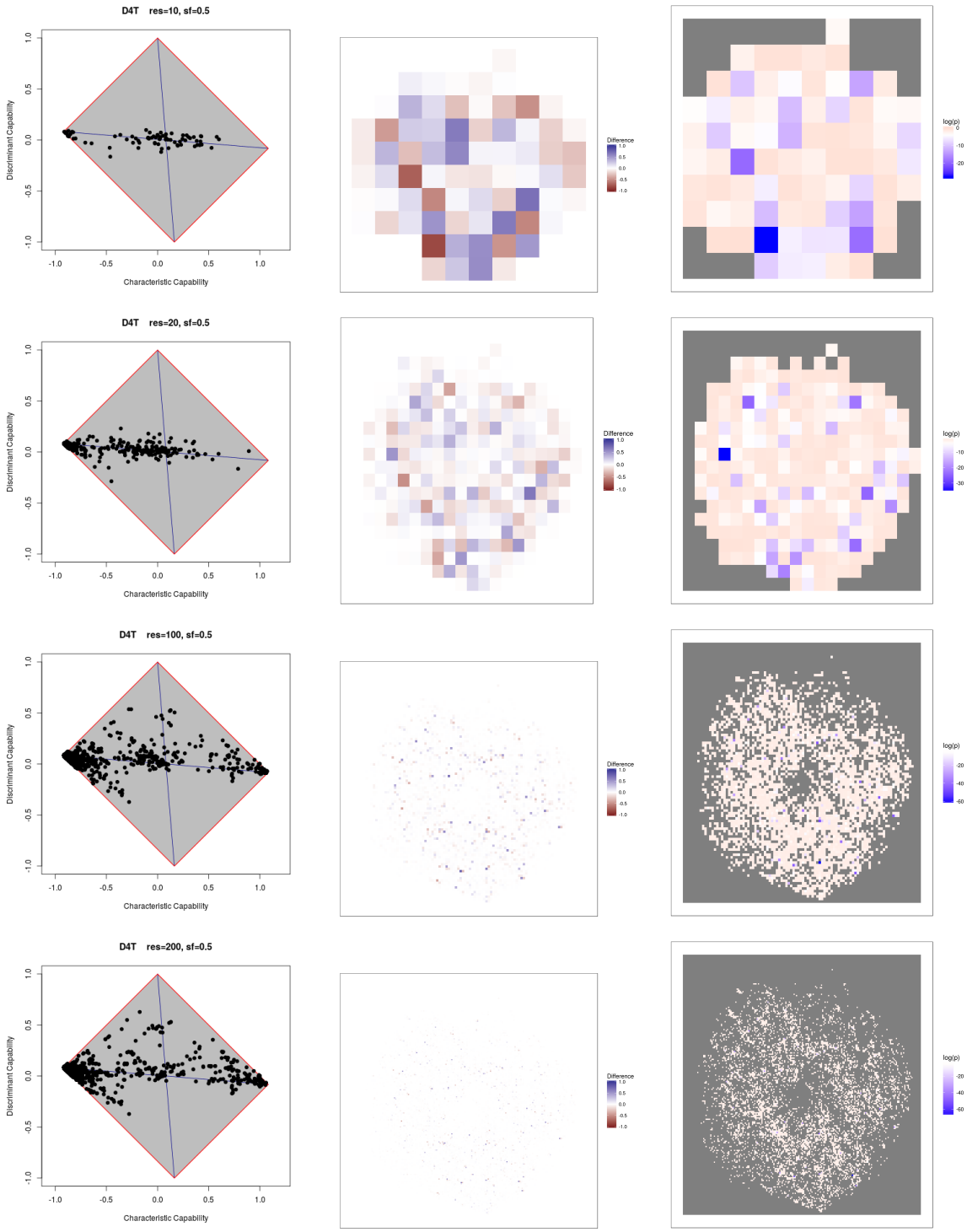

Figure 18: D4T, $sf = 0.5$

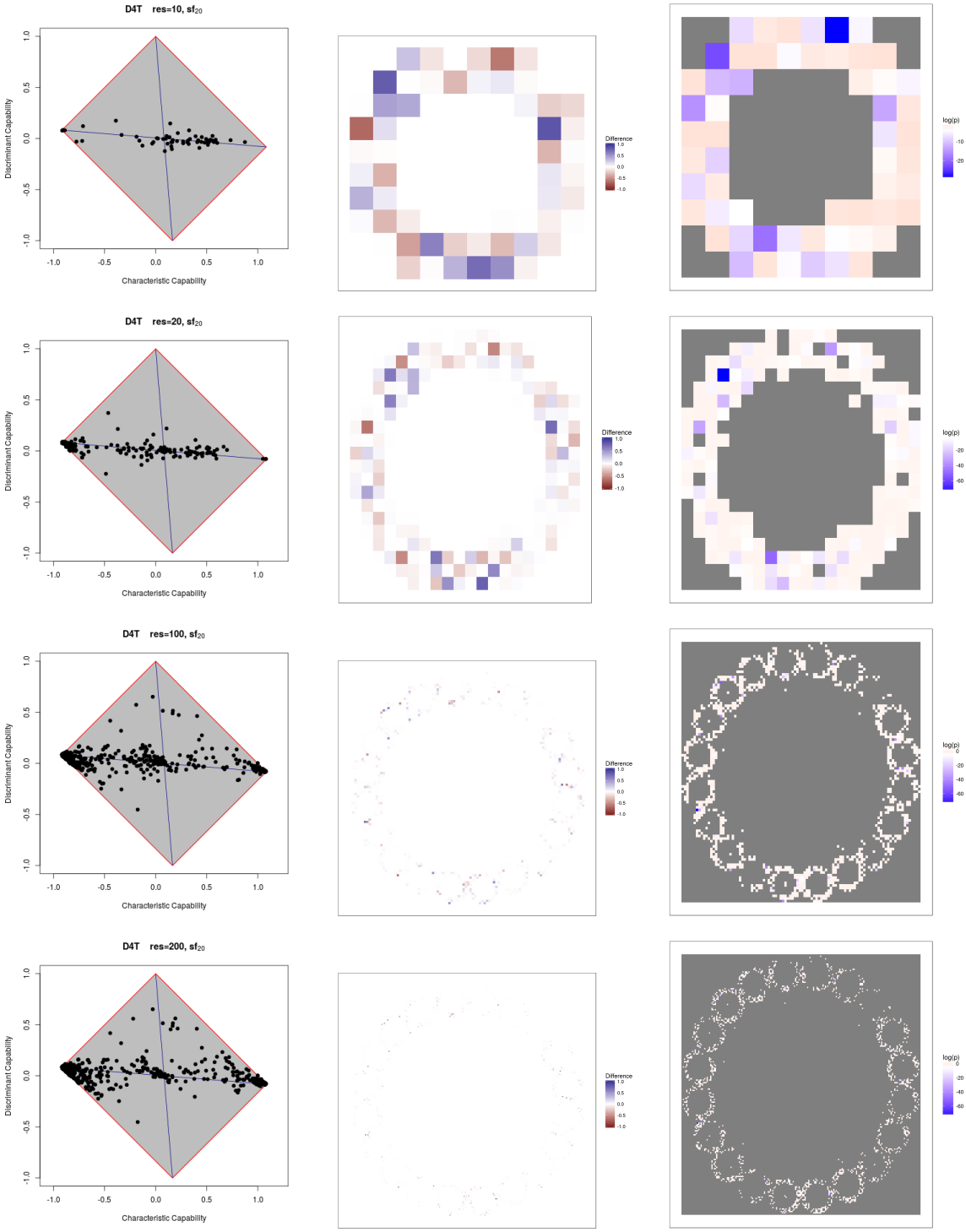

Figure 19: D4T, sf_{20}

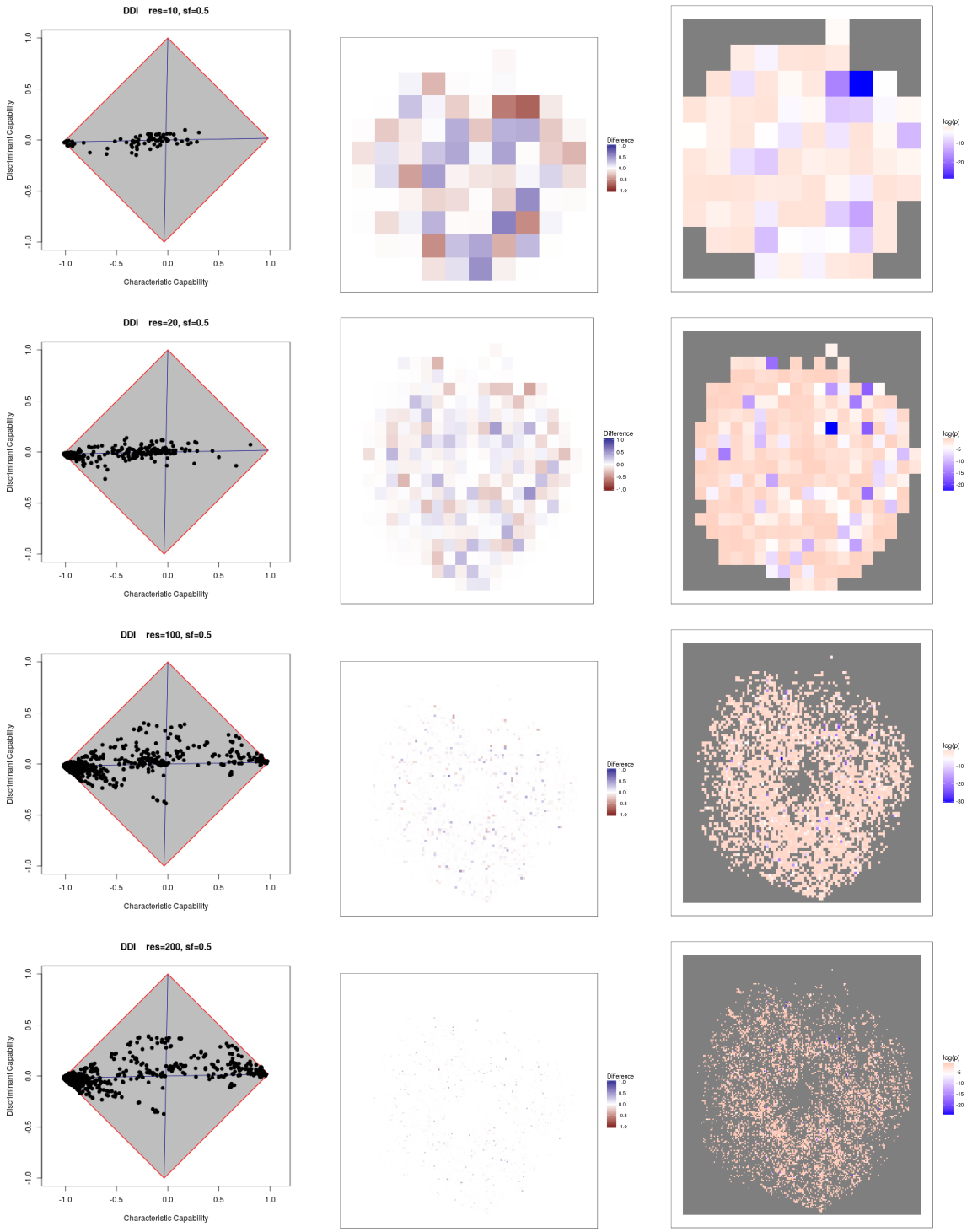

Figure 20: DDI, $sf = 0.5$

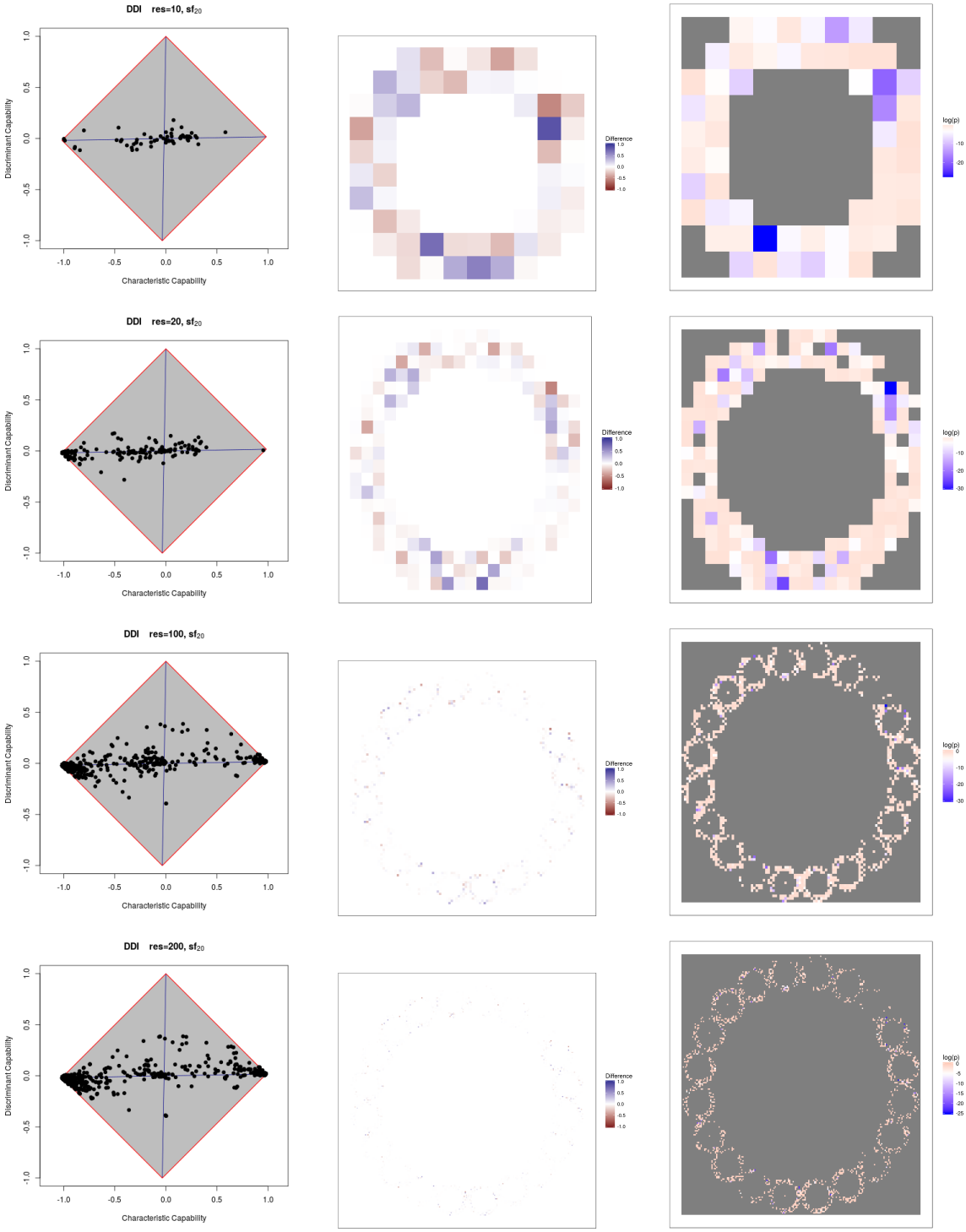

Figure 21: DDI, sf_{20}

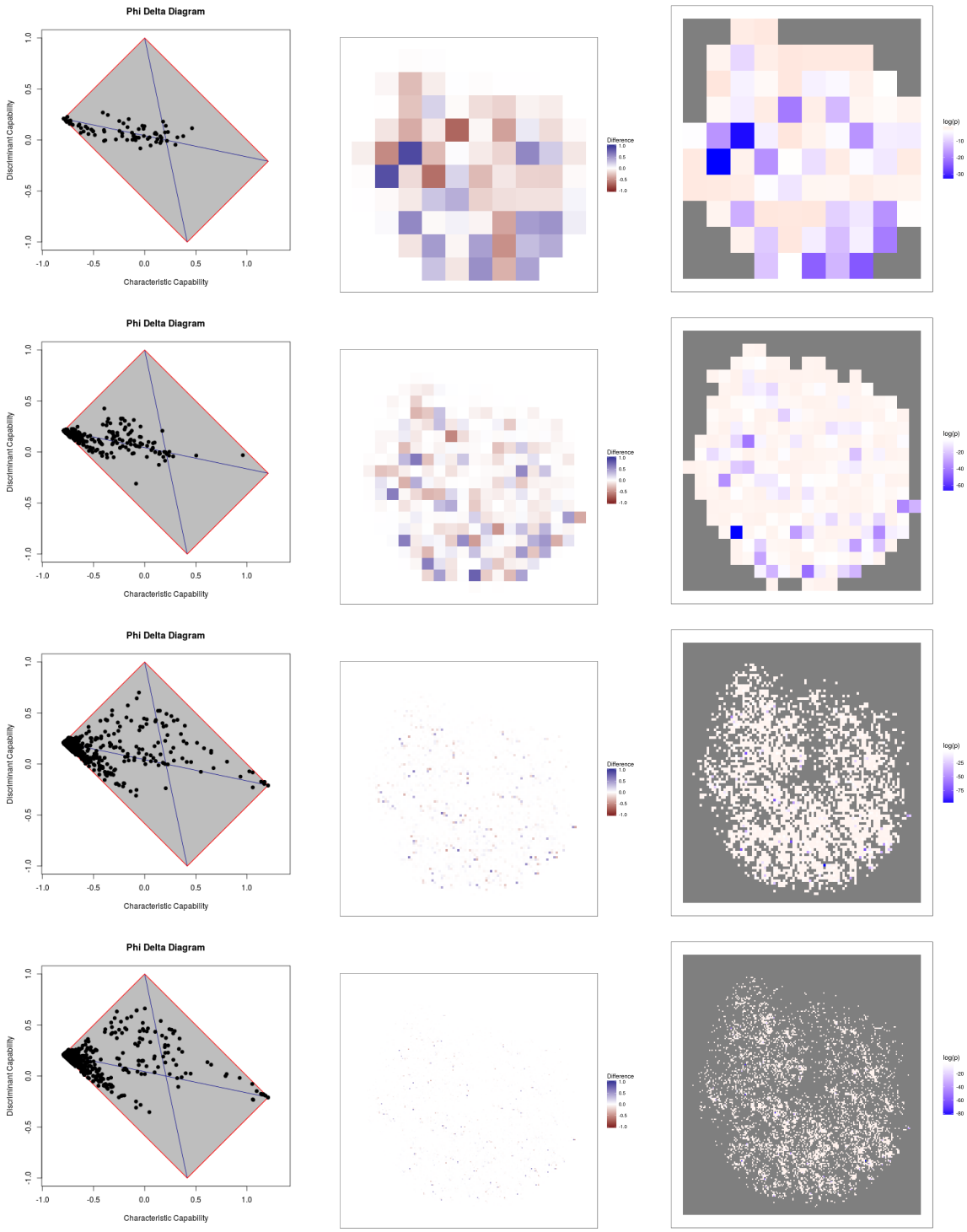

Figure 22: APV, $sf = 0.5$

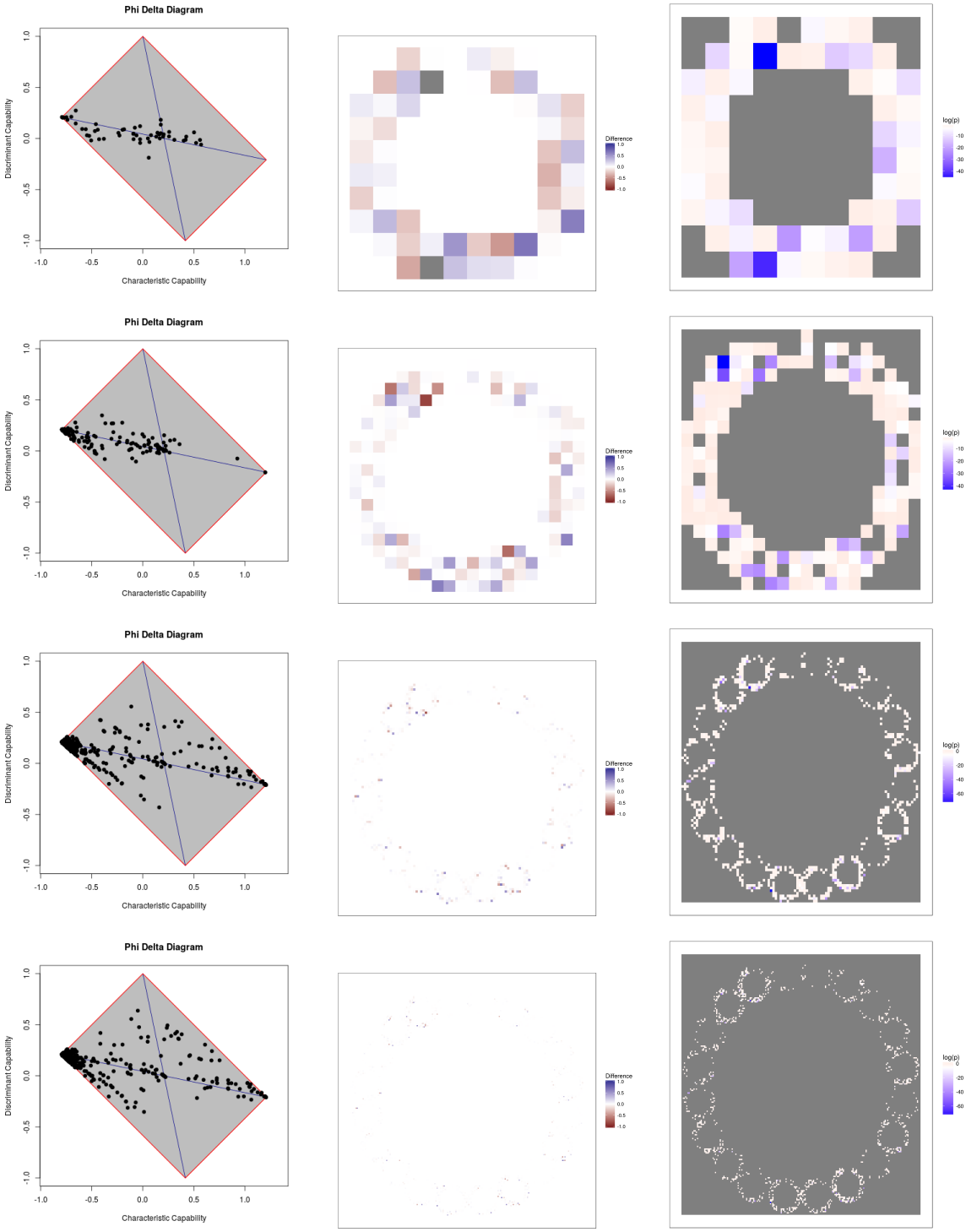

Figure 23: APV, sf_{20}

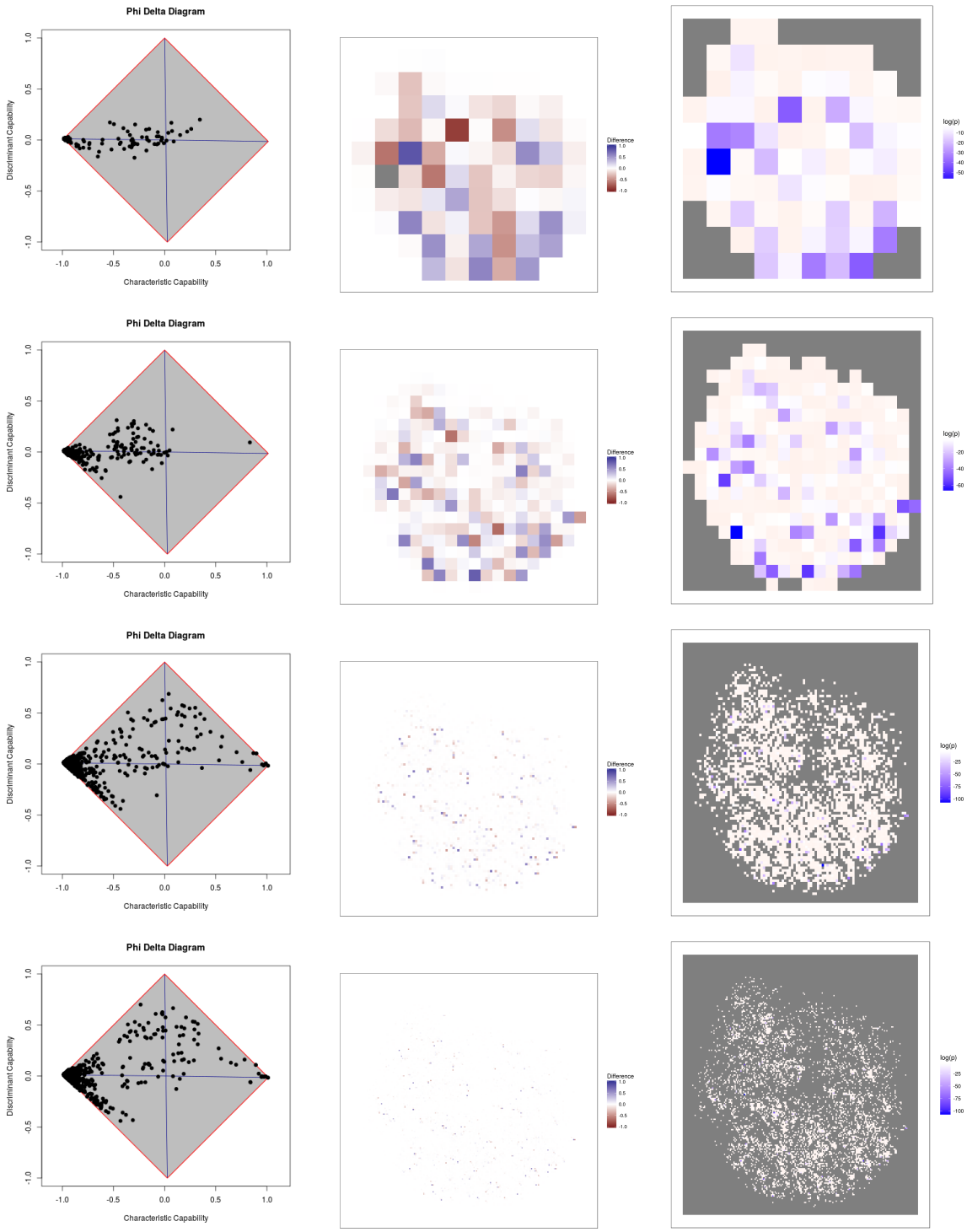

Figure 24: IDV, $sf = 0.5$

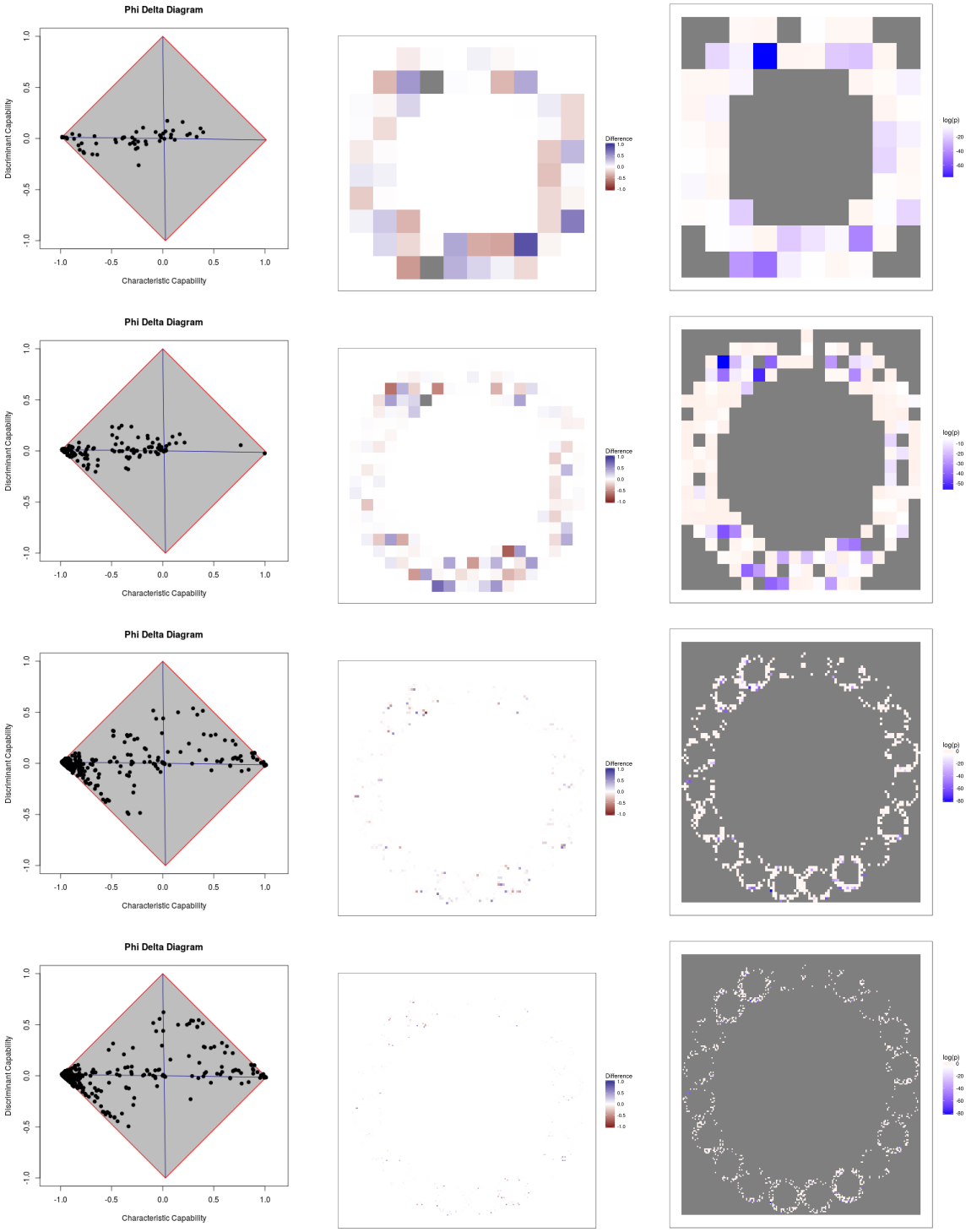

Figure 25: IDV, sf_{20}

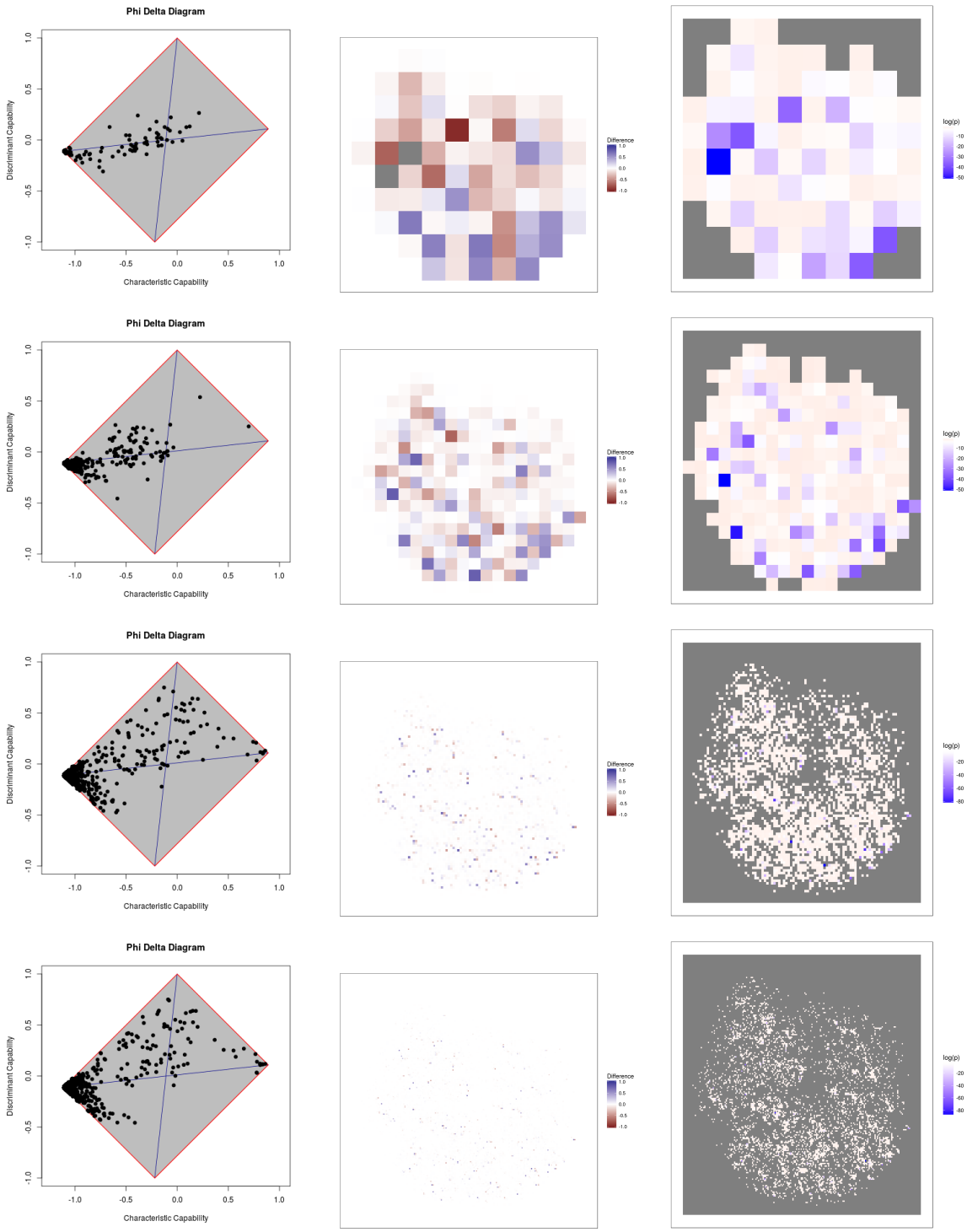

Figure 26: LPV, $sf = 0.5$

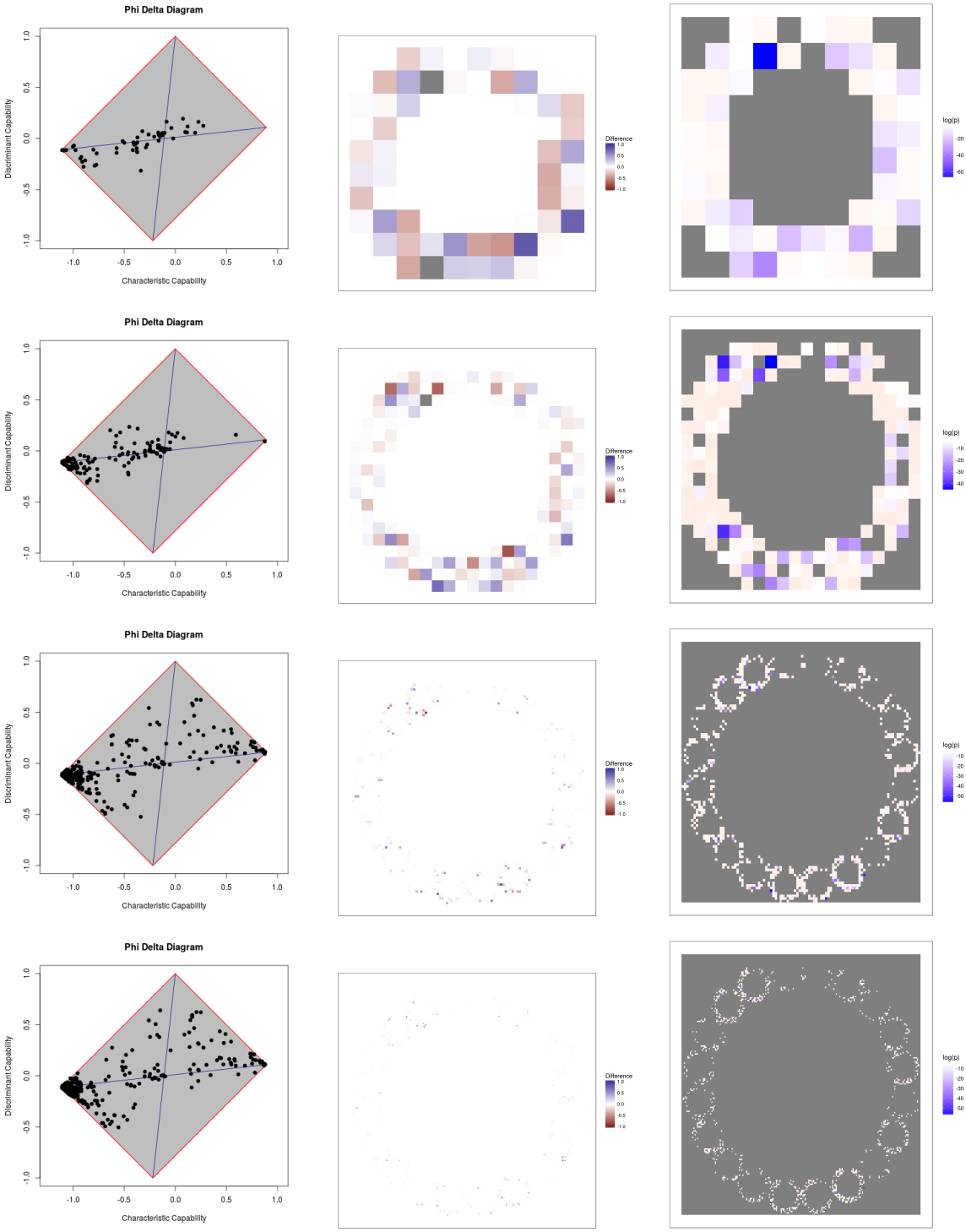

Figure 27: LPV, sf_{20}

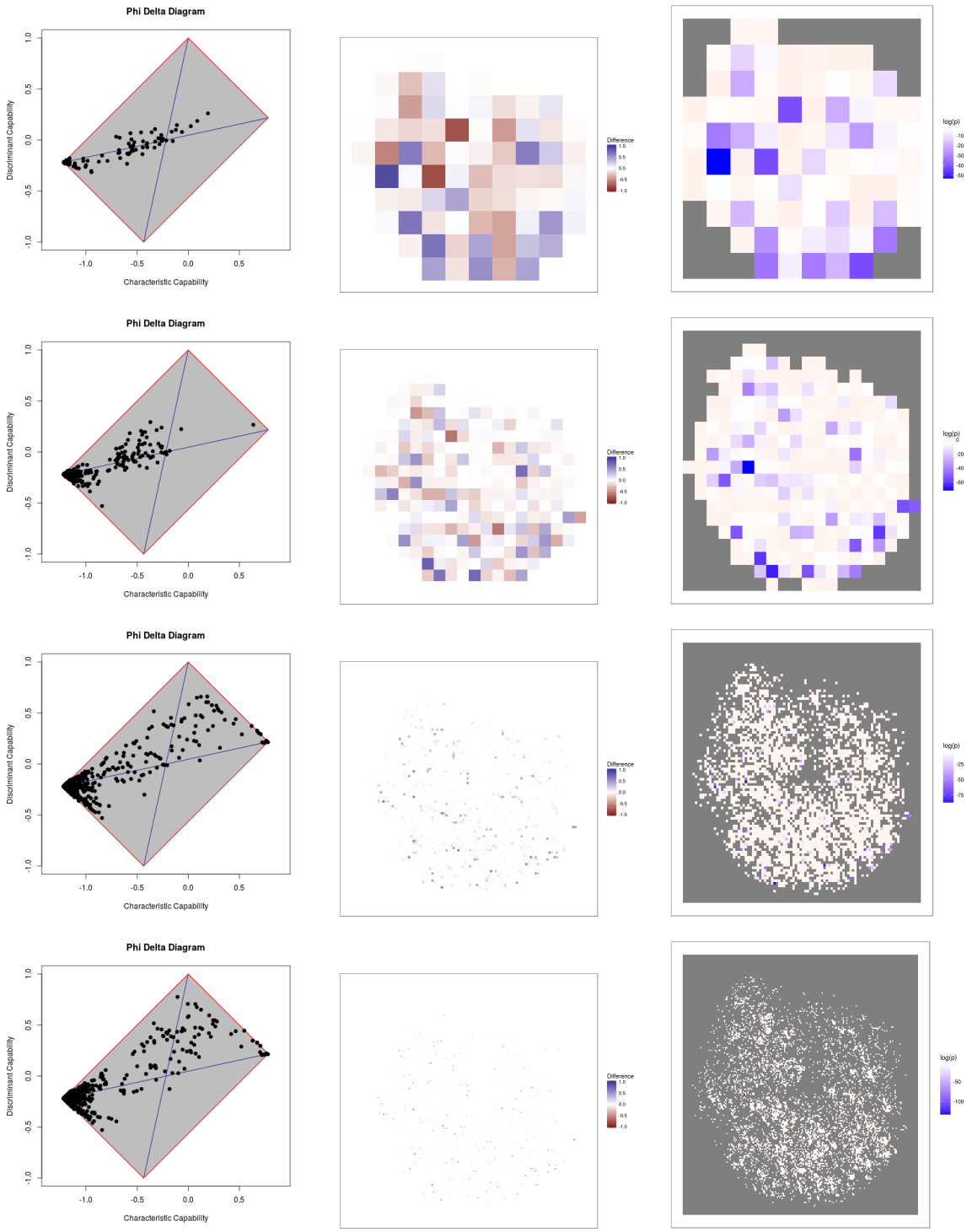

Figure 28: NFV, $sf = 0.5$

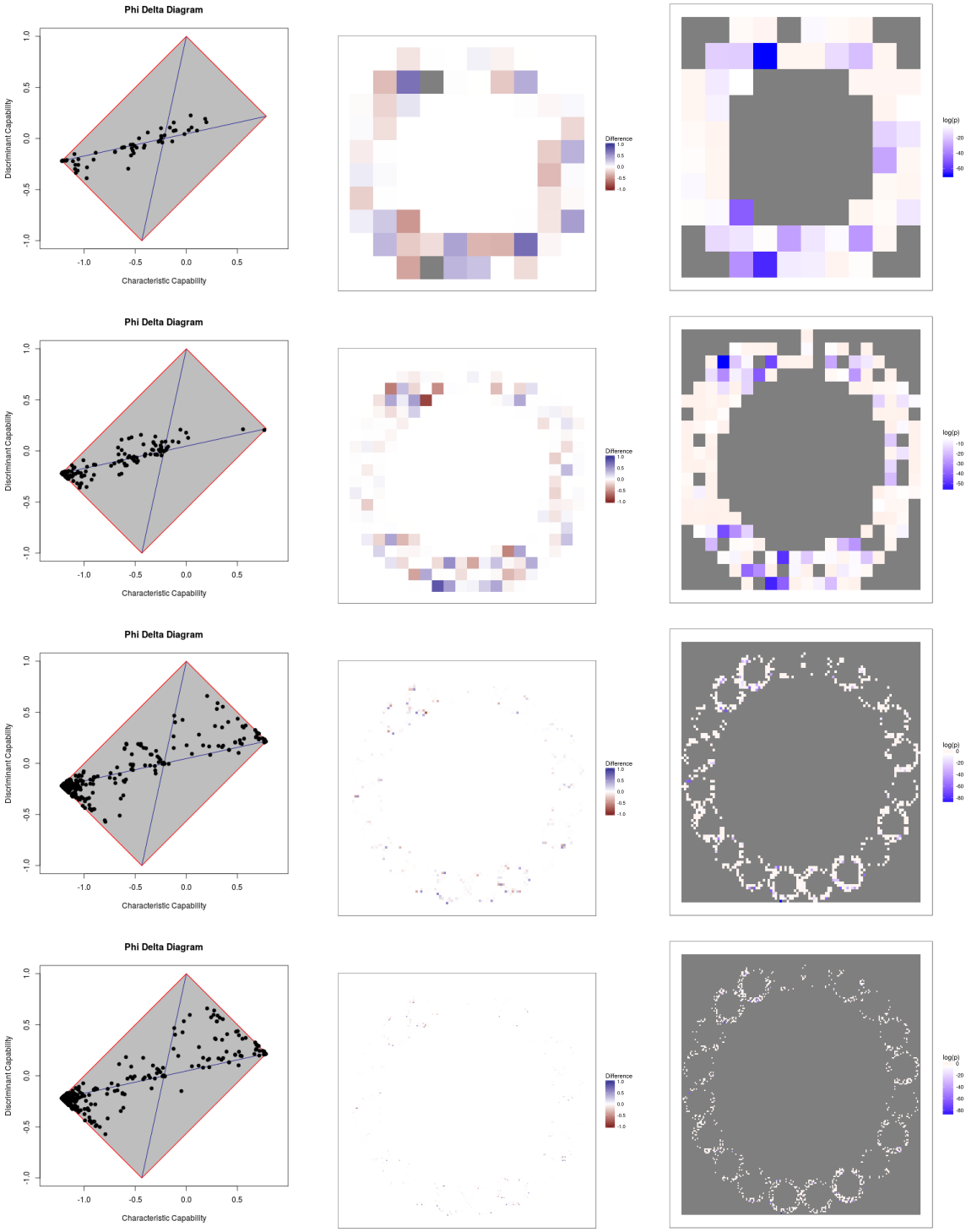

Figure 29: NFV, sf_{20}

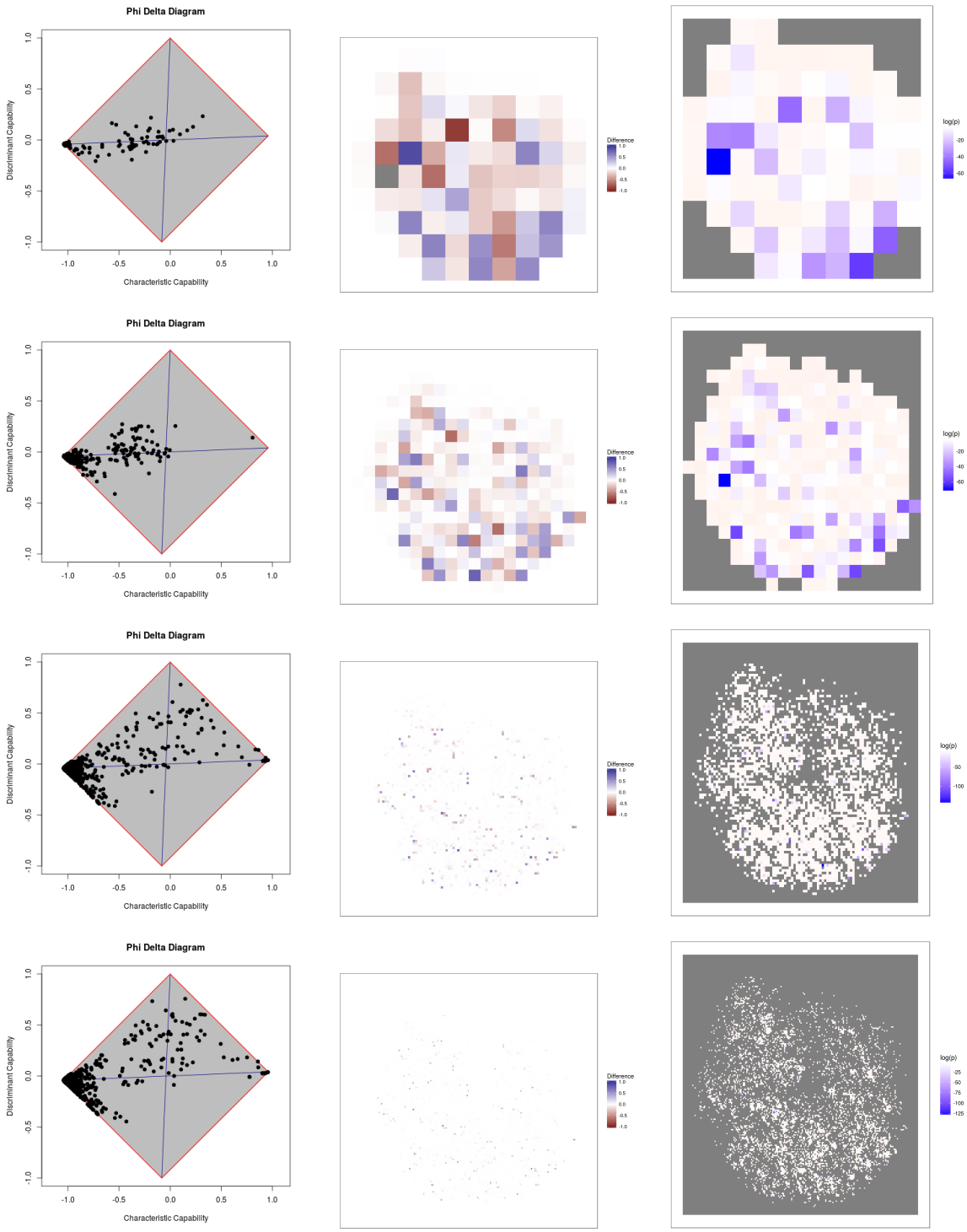

Figure 30: RTV, $sf = 0.5$

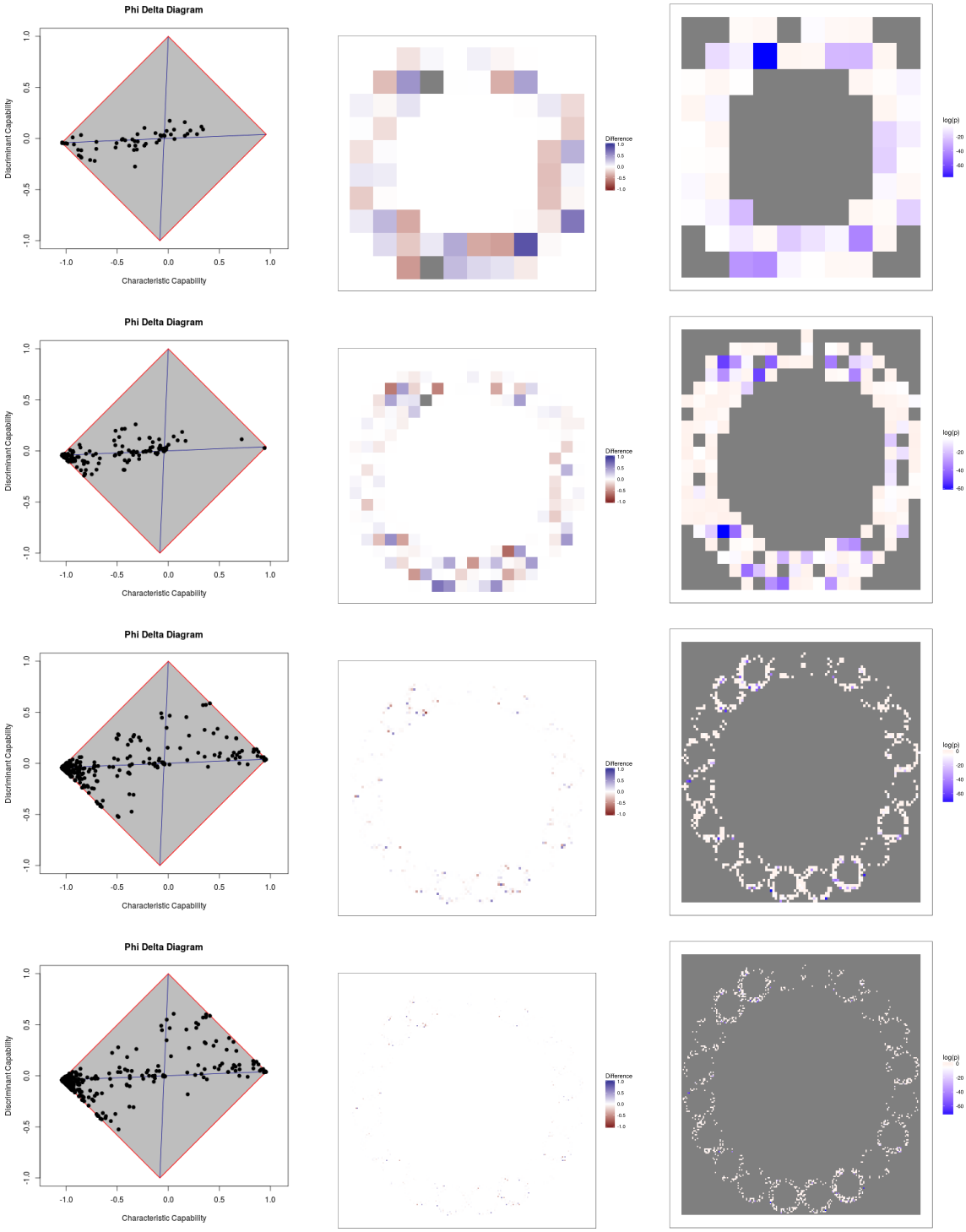

Figure 31: RTV, sf_{20}

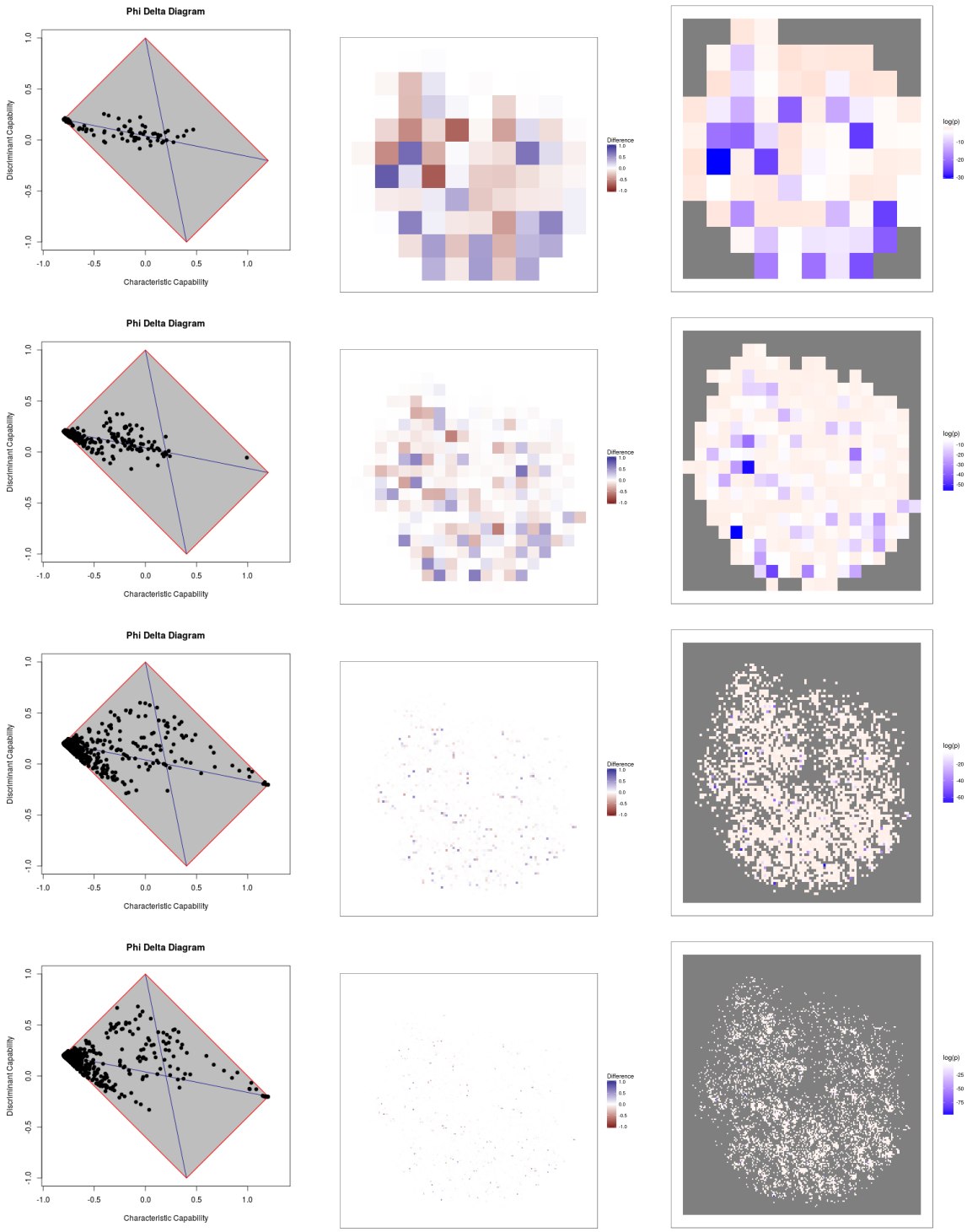

Figure 32: SQV, $sf = 0.5$

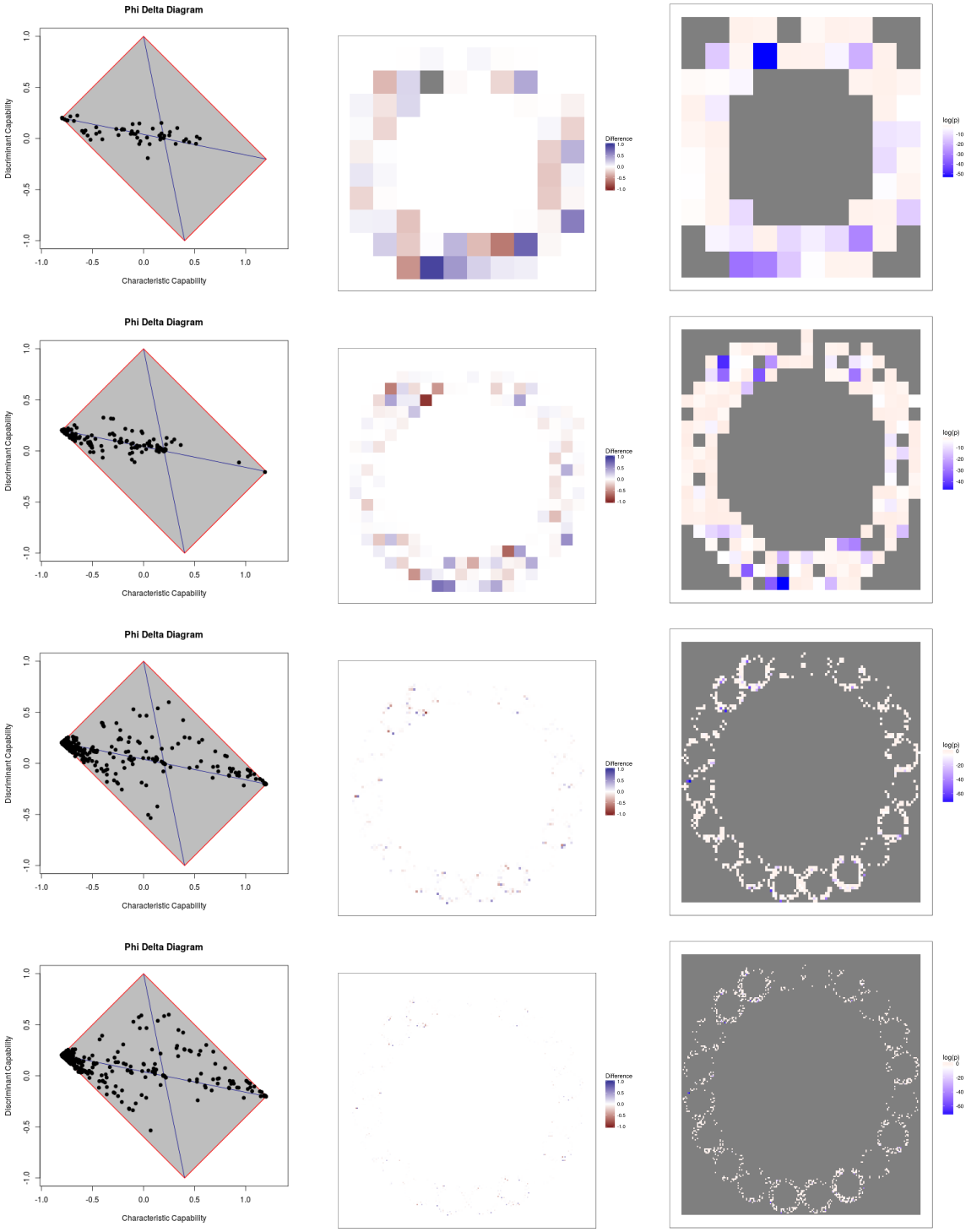

Figure 33: SQV, sf_{20}