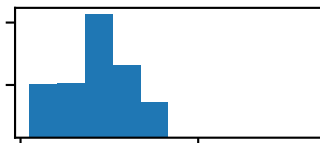


number of model predictions

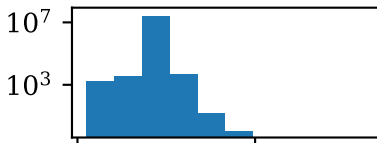
weakening



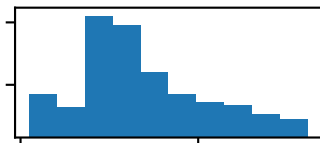
weakening
baseline



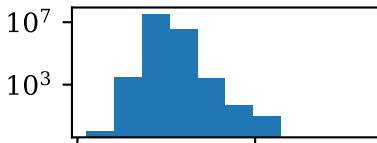
weakening
gate+_post+



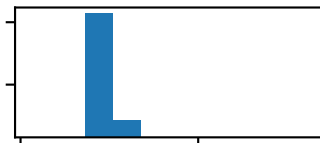
weakening
gate+_post-



weakening
gate-_post+



weakening
gate-_post-



$\text{scale(clean)} / \text{scale(ablated)}$