Scatter plot of SD of models 2 and 4 on 50 test sets