errors comparison on the same methods
red: ETCC win, green: AI win, black: equal, logscale