F1 score evaluation,
MC Unc=Monte Carlo uncertain, MC Crt=Monte Carlo non-uncertain, TI
Unc=Threshold interval uncertain, TI Crt=Threshold interval non-uncertain

