Gradient norm per optimizer step
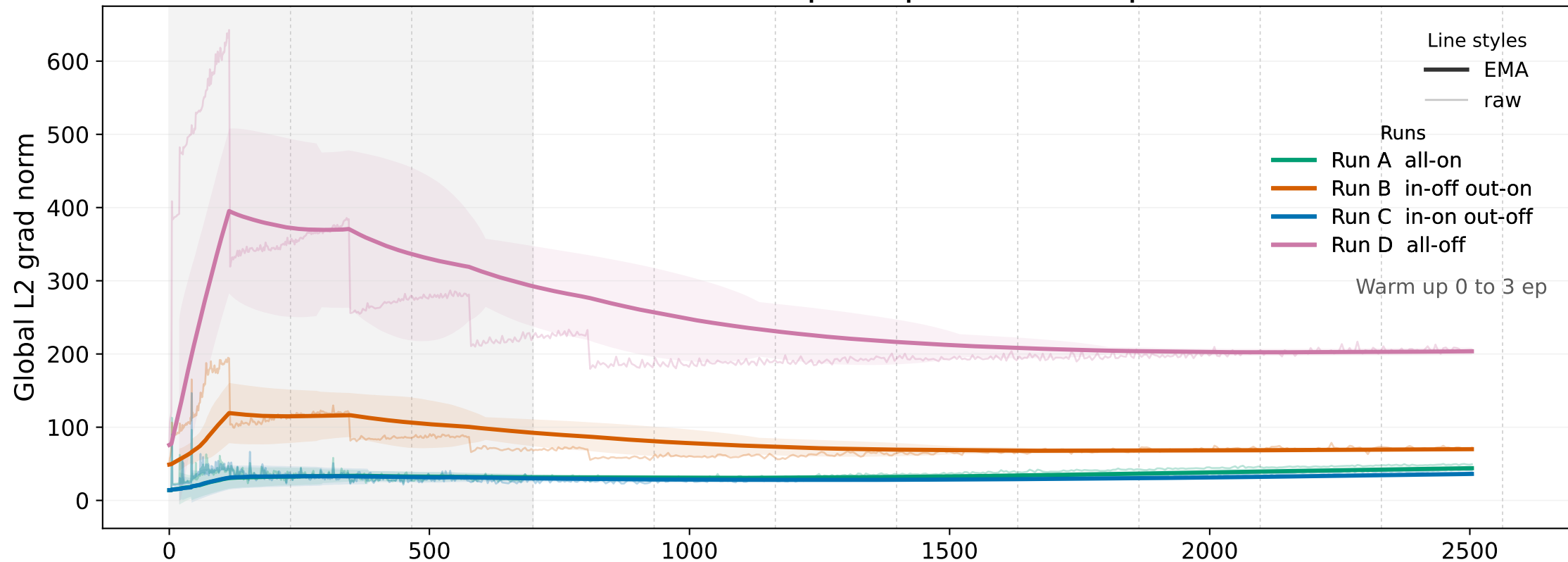
Metric	Run A	Run B	Run C	Run D
Peak warm up	113.1	194.9	146.7	642.6
EMA at warm up end	31.6	92.6	30.2	293.0
Plateau mean ± std last 1000	36.4 ± 8.4	88.3 ± 32.9	32.6 ± 8.3	284.2 ± 124.7
Plateau coefficient of variation	0.232	0.373	0.256	0.439

Lower peak and EMA at warm up end with lower plateau coefficient of variation indicate higher stability.