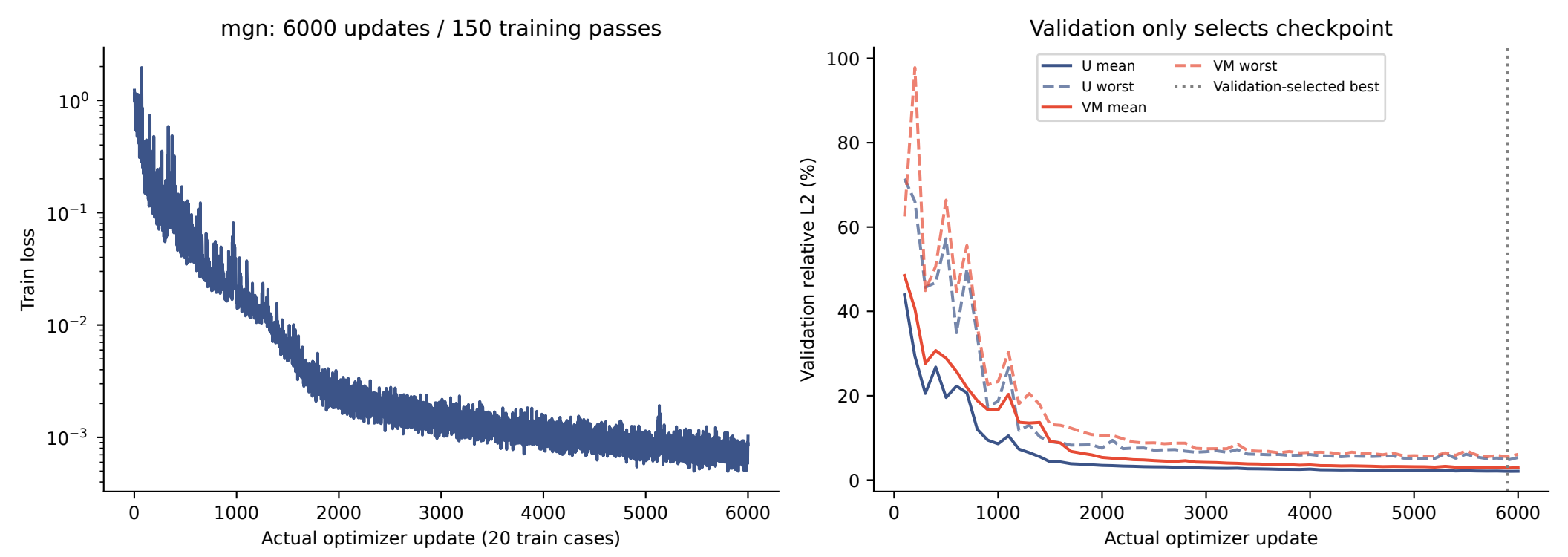

mgn: 6000 updates / 150 training passes
Train loss
10^{0}
10^{-1}
10^{-2}
10^{-3}
0
1000
2000
3000
4000
5000
6000
Actual optimizer update (20 train cases)
Validation only selects checkpoint
Validation relative L2 (%)
100
80
60
40
20
0
U mean
U worst
VM mean
VM worst
Validation-selected best
0
1000
2000
3000
4000
5000
6000
Actual optimizer update