

GCM+eHMMs Training Results

Chengze Shen

September 3, 2021

1 Notations

1. **PASTA+UPP** is UPP with PASTA backbone (by default).
2. **MAGUS+UPP** is UPP with MAGUS backbone (in the previous paper, we showed that UPP with MAGUS backbone is better).
3. **PASTA+GCM+eHMMs(unweighted)** is unweighted GCM+eHMMs with PASTA backbone.
4. **PASTA+GCM+eHMMs(weighted)** is weighted GCM+eHMMs with PASTA backbone.
5. **MAGUS+GCM+eHMMs(unweighted)** is unweighted GCM+eHMMs with MAGUS backbone.
6. **MAGUS+GCM+eHMMs(weighted)** is weighted GCM+eHMMs with MAGUS backbone.

The decomposition subset size A ranges from $[1, 2, 5, 10, 50]$, while the number of HMMs used k ranges from $[1, 2, 4, 10]$. For PASTA+UPP and MAGUS+UPP, since they only use the best HMM for a query, they are categorized under $k = 1$.

2 1000M1, 1000M2, 16S.M, 23S.M (high fragmentation)

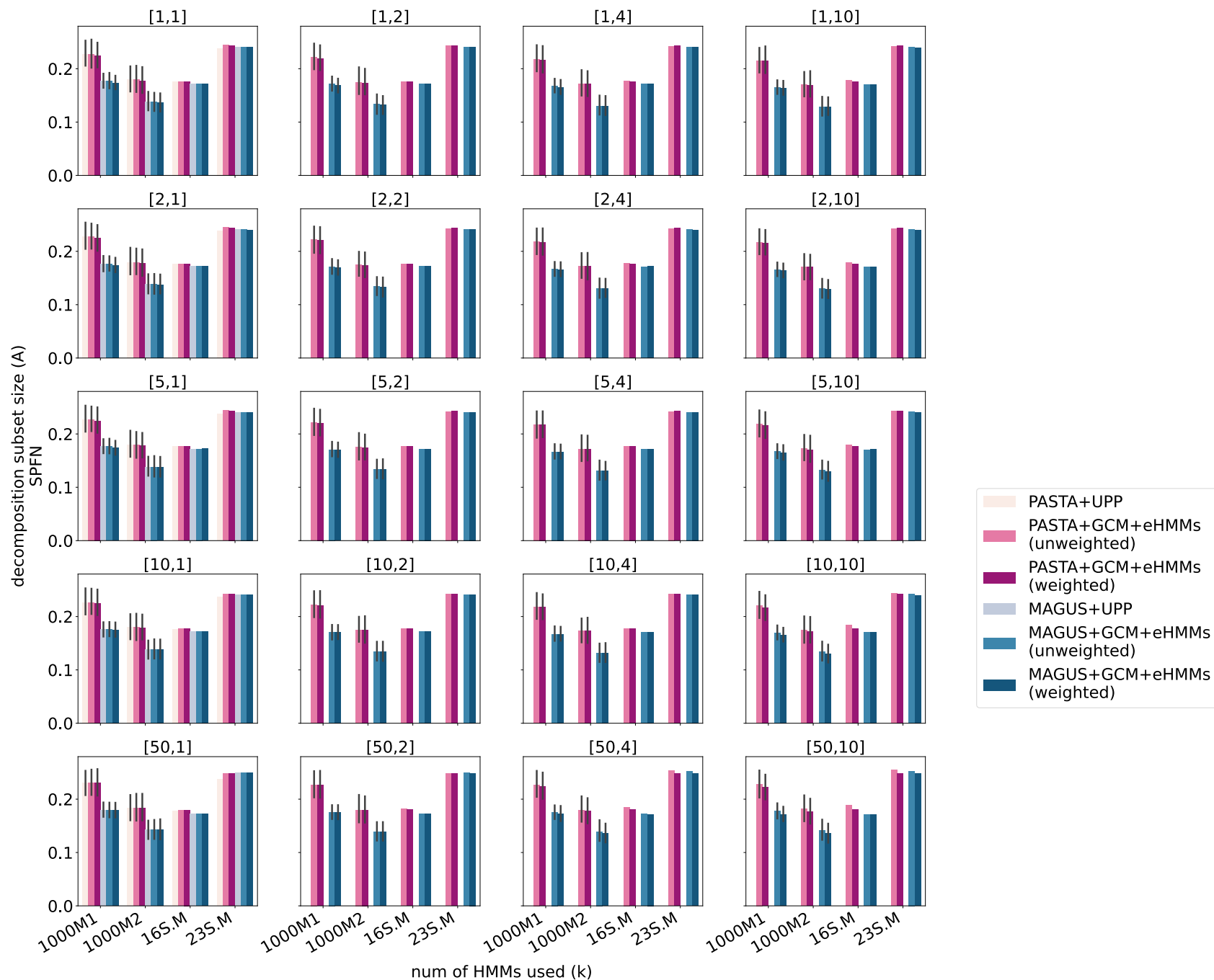


Figure 1: SPFN for the **high fragmentary** 1000M1, 1000M2, 16S.M and 23S.M datasets.

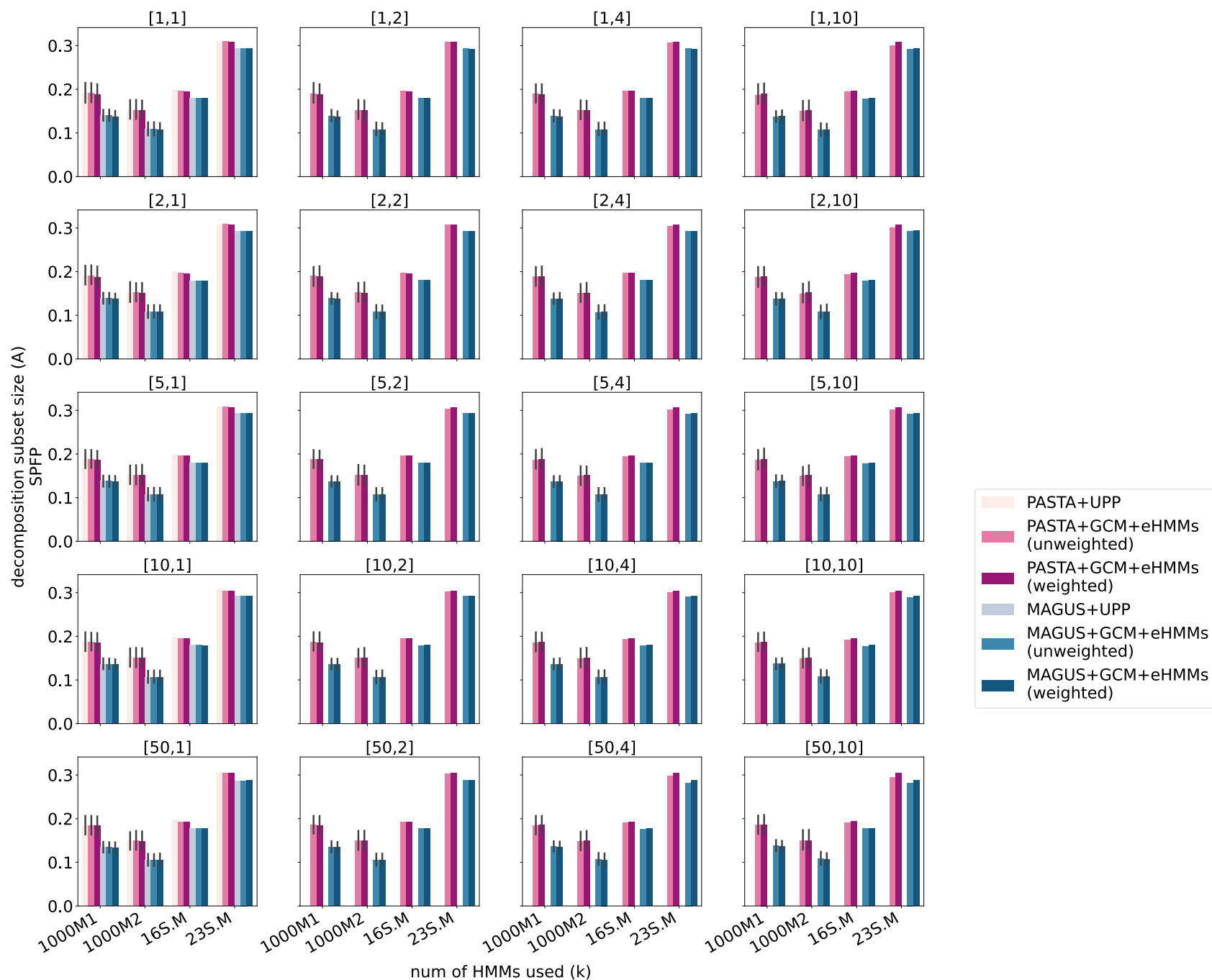


Figure 2: SPFP for the **high fragmentary** 1000M1, 1000M2, 16S.M and 23S.M datasets.

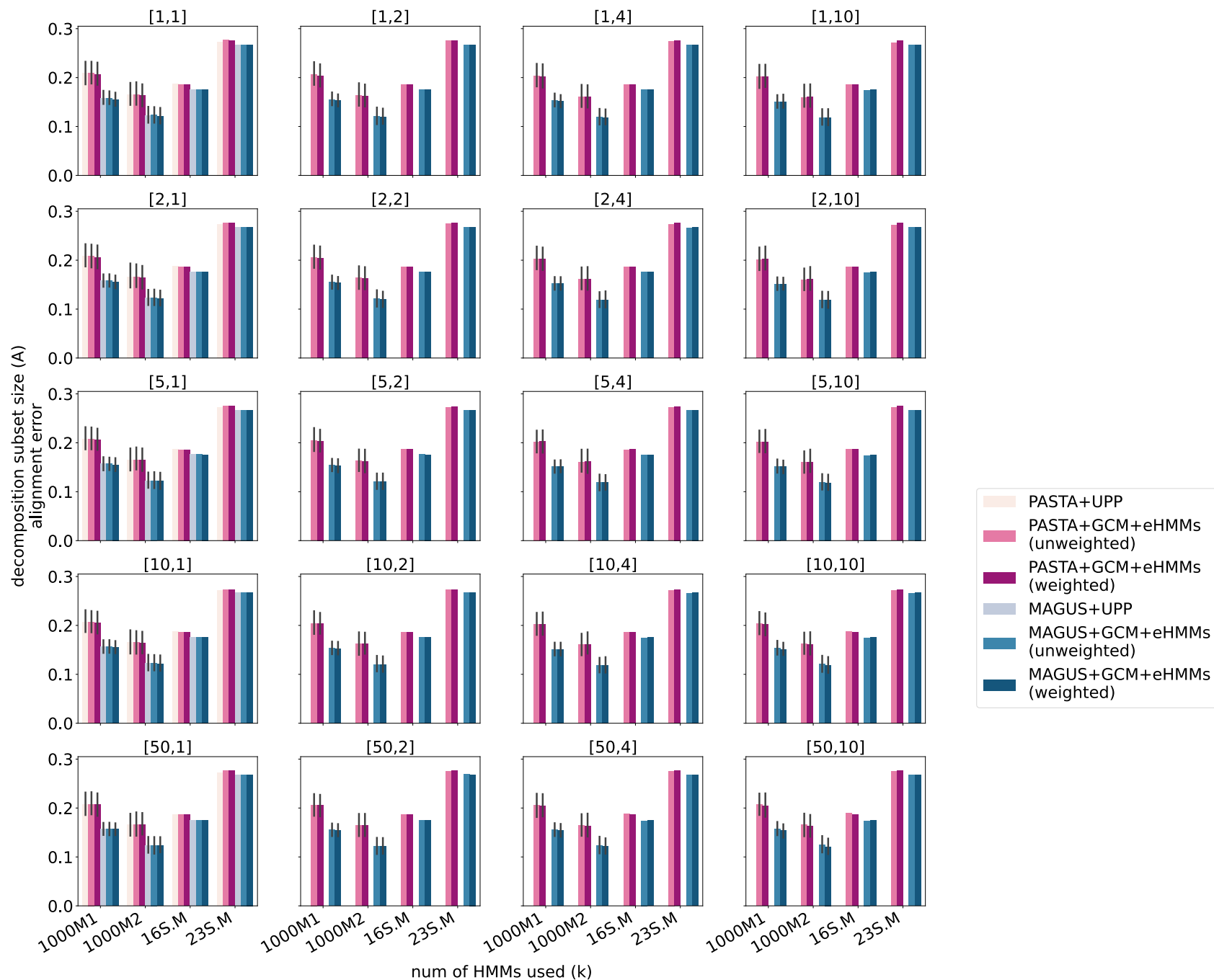


Figure 3: Alignment error for the **high fragmentary** 1000M1, 1000M2, 16S.M and 23S.M dataset.

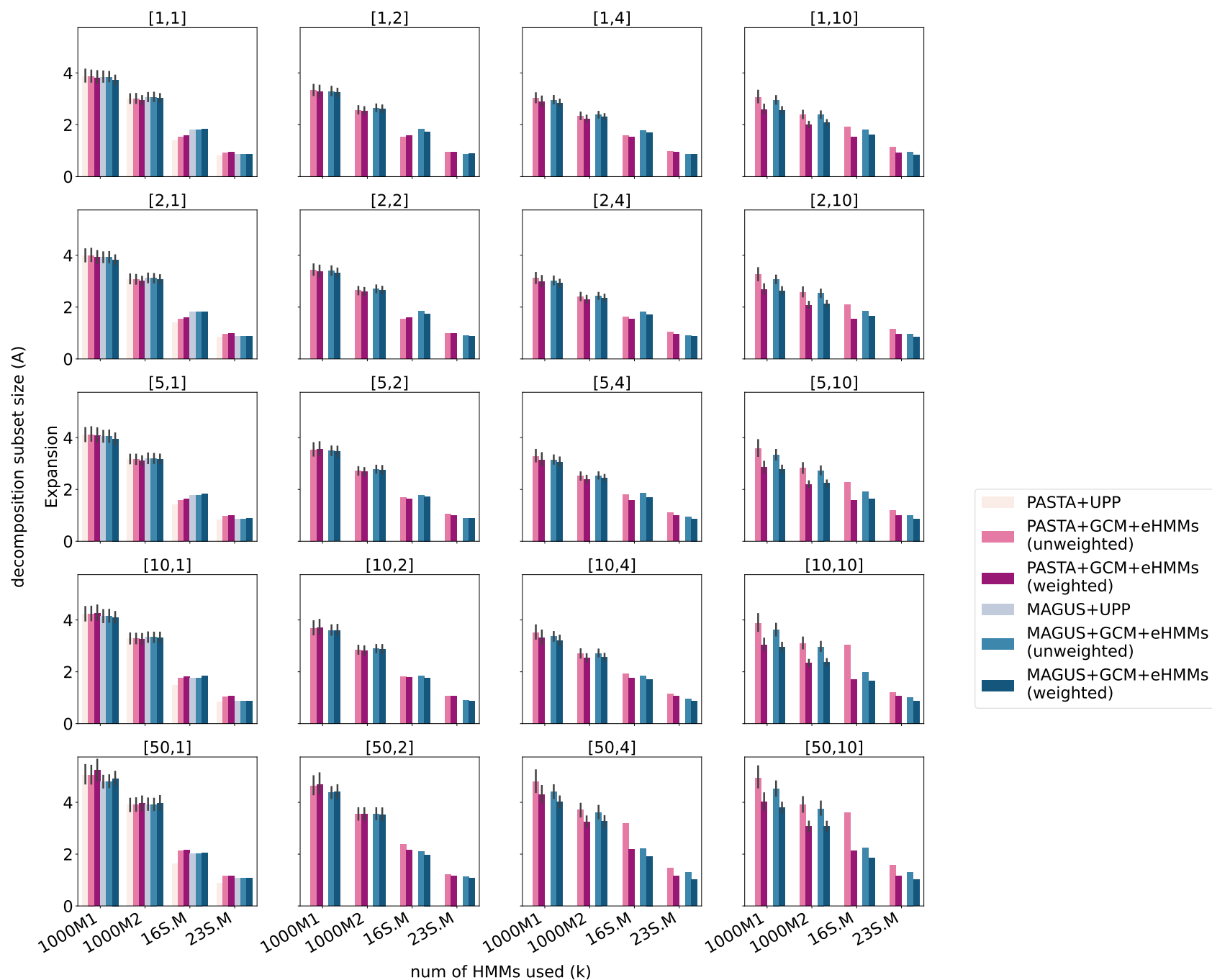


Figure 4: Expansion score for the **high fragmentary** 1000M1, 1000M2, 16S.M and 23S.M dataset.

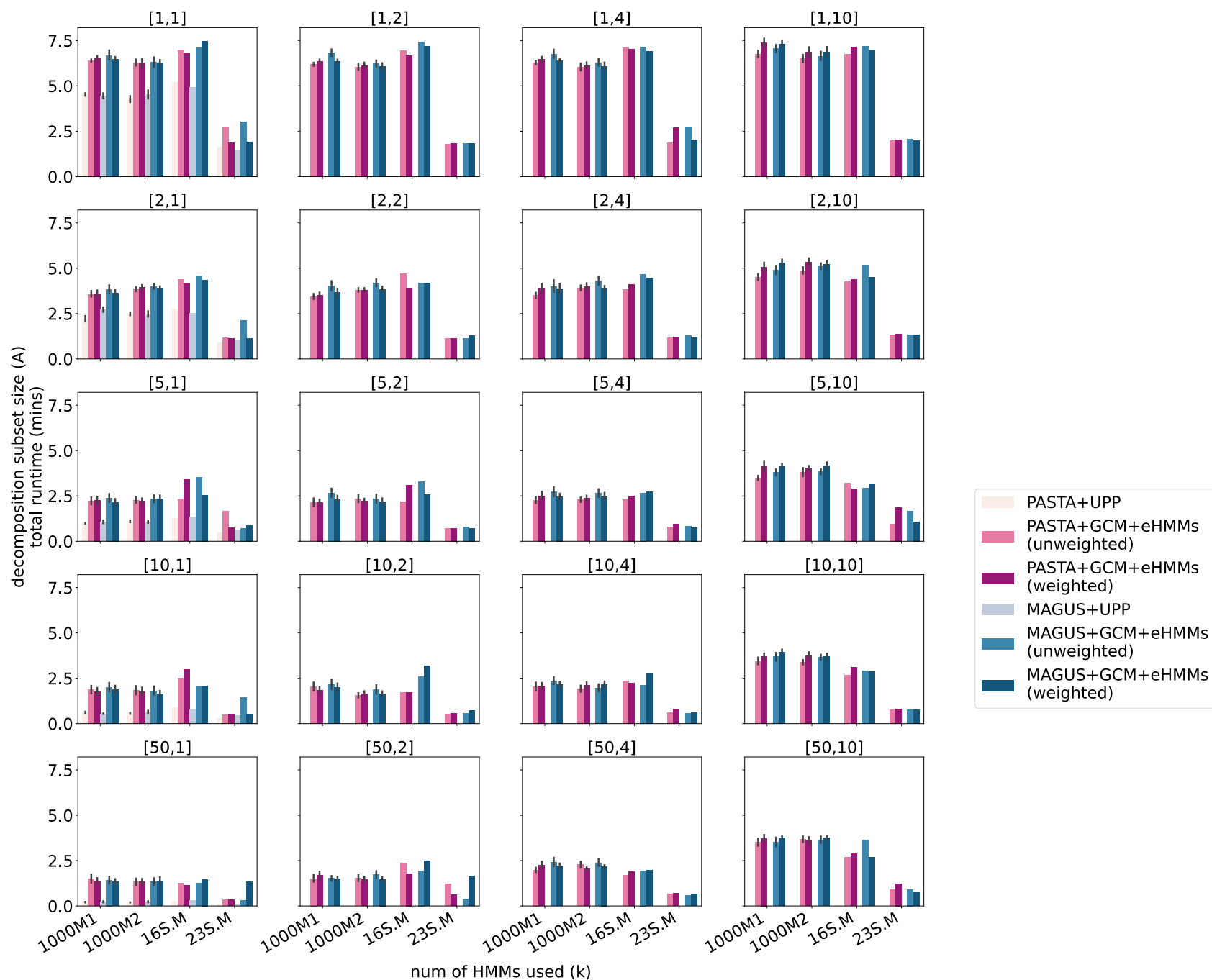


Figure 5: Runtime in minutes for the **high fragmentary** 1000M1, 1000M2, 16S.M and 23S.M datasets

3 1000M1, 1000M2, 16S.M, 23S.M (low fragmentation)

∞

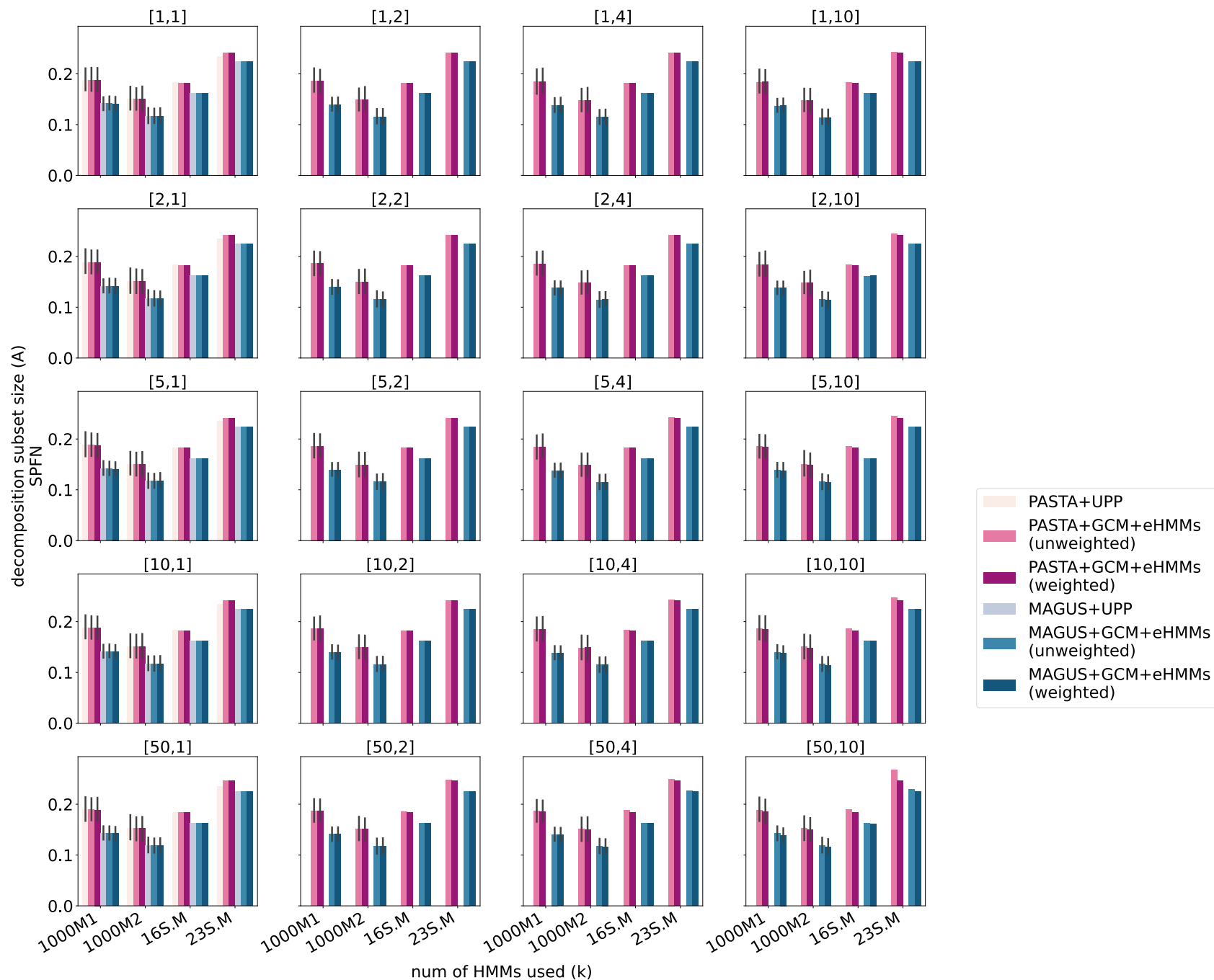


Figure 6: SPFN for the **low fragmentary** 1000M1, 1000M2, 16S.M and 23S.M datasets.

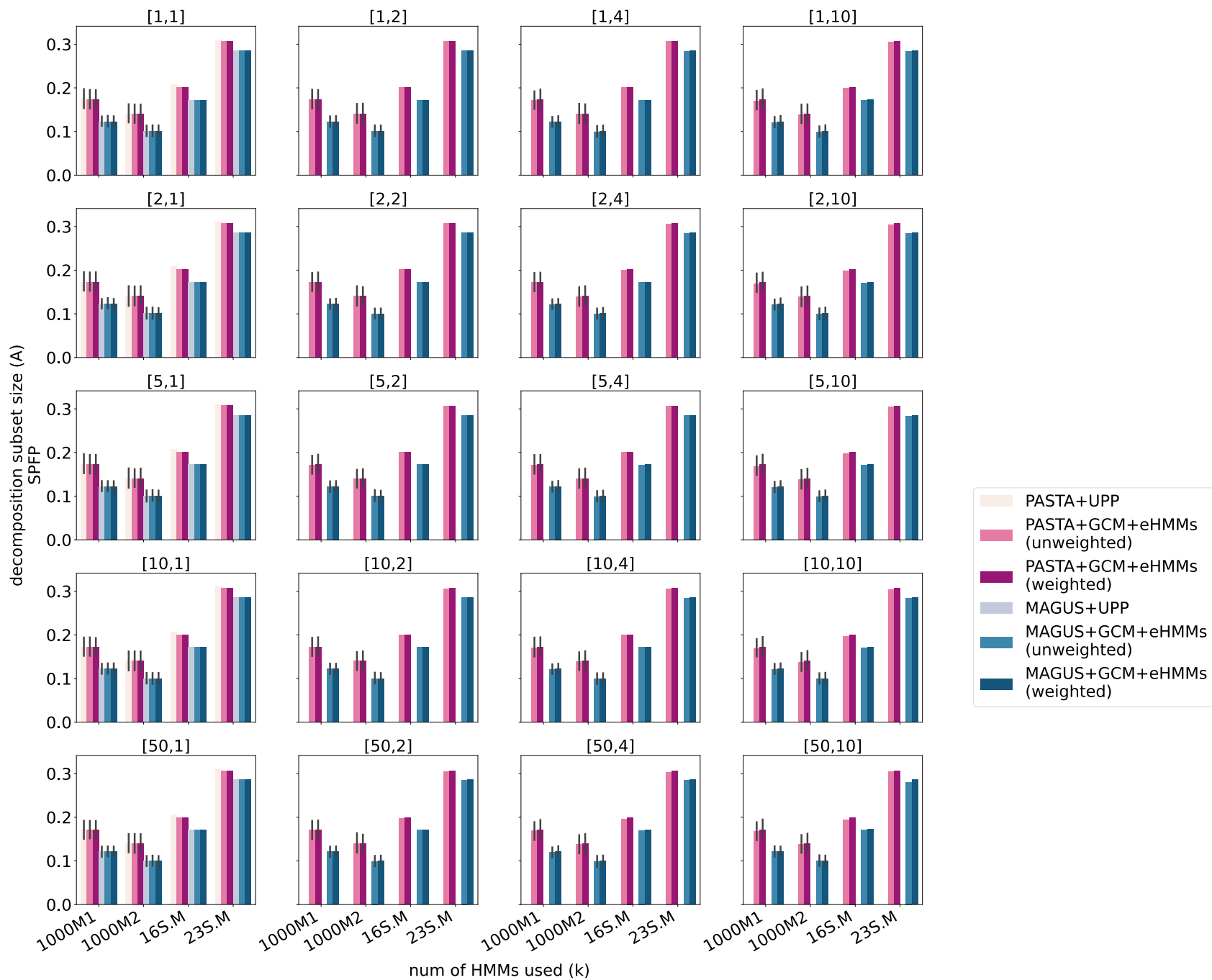
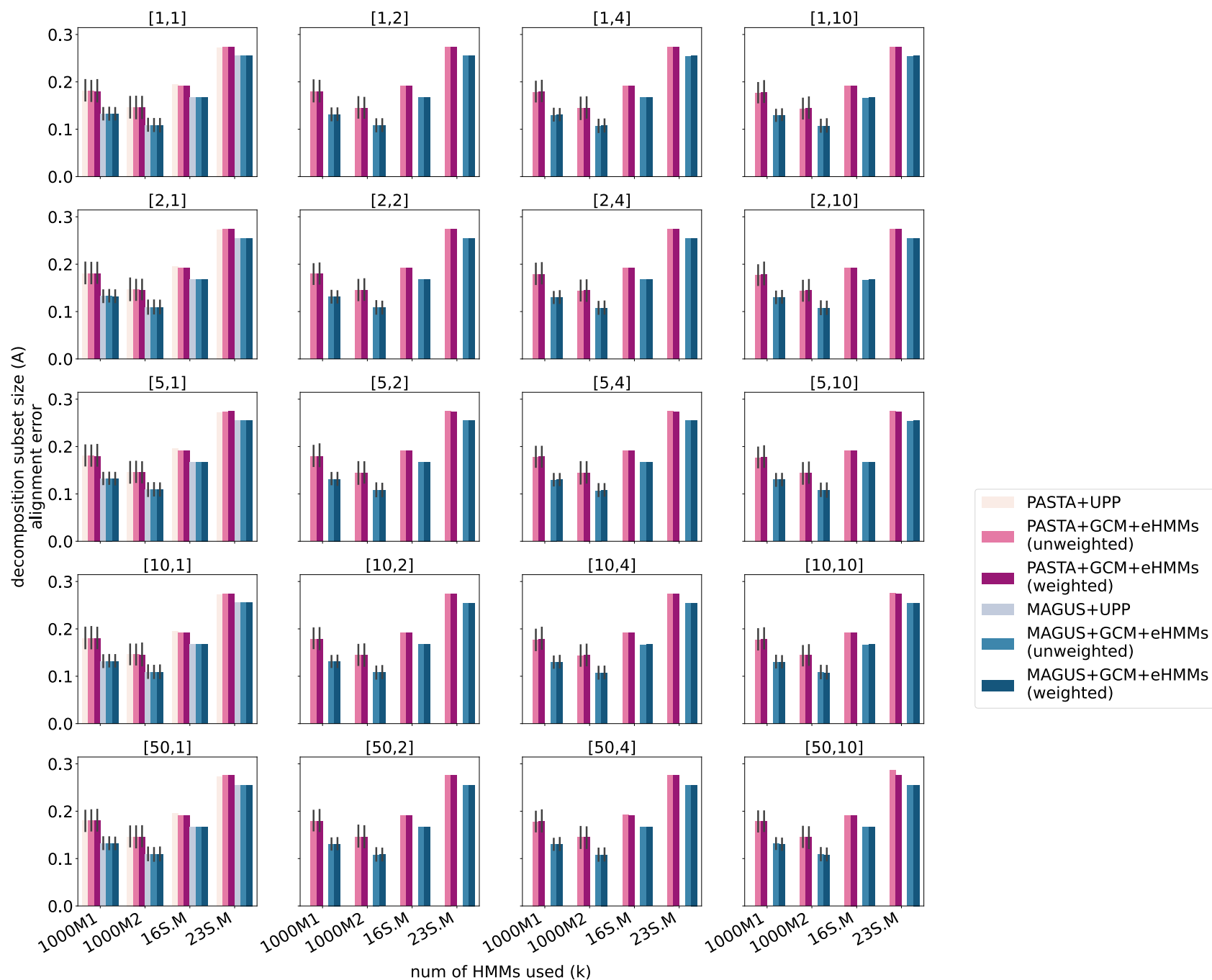


Figure 7: SPFP for the **low fragmentary** 1000M1, 1000M2, 16S.M and 23S.M datasets.



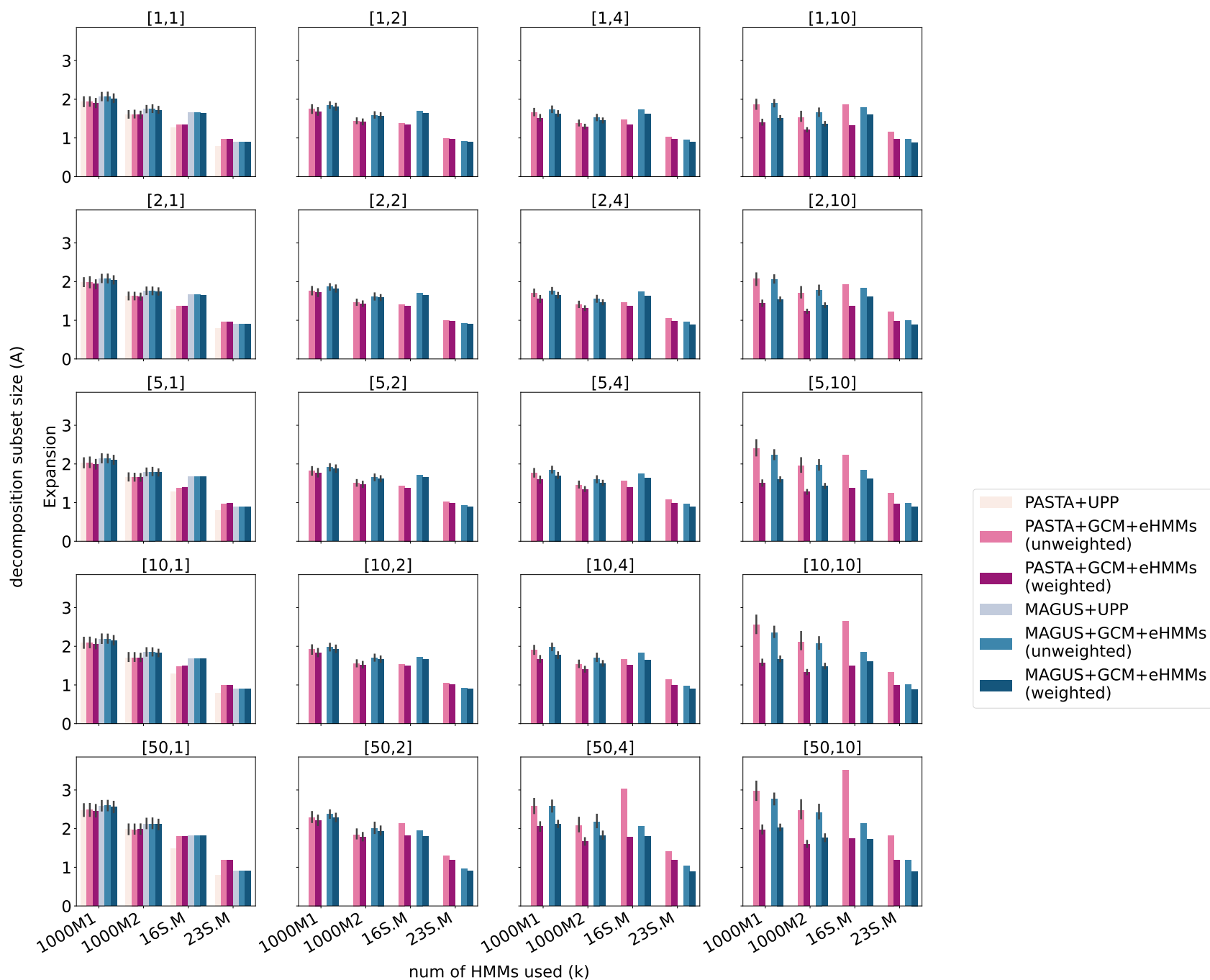


Figure 9: Expansion score for the **low fragmentary** 1000M1, 1000M2, 16S.M and 23S.M dataset.

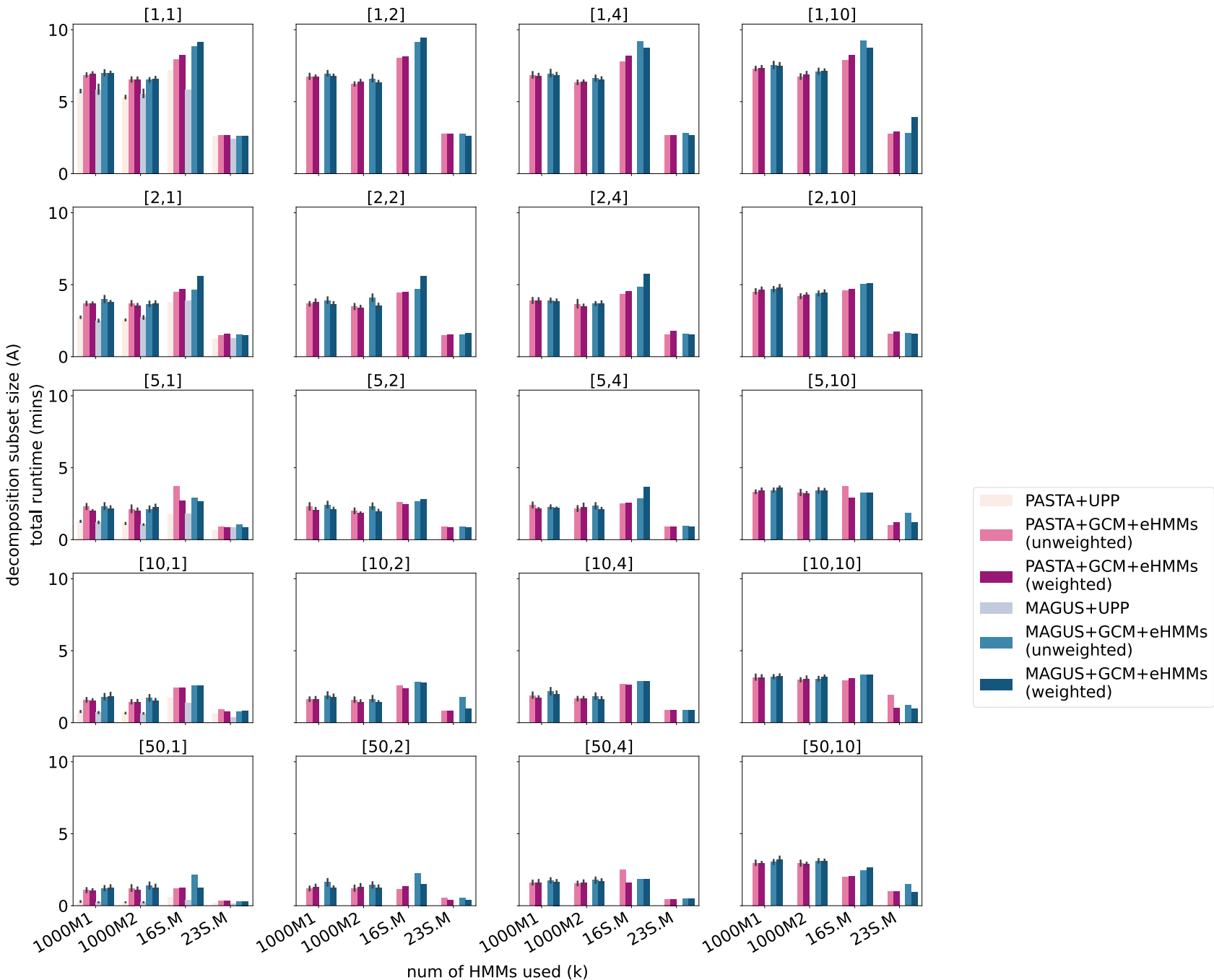


Figure 10: Runtime in minutes for the **low fragmentary** 1000M1, 1000M2, 16S.M and 23S.M datasets