Fig. 11. Experiments on synthetic data with random reference set, T =12500, k=30, m=500, dcom, kFN.

1. Comparison between models. //5compare\_log，CP PRP 等那5個model的比較
2. Number of candidate sites in each level of MSwave-L. //rand\_site\_v2，Site 減少的狀況

//Description about synthetic data

In order to test the performance on large data, we generated the synthetic data by the same random walk data model used in [16]. For every stream, we generate it by the random walk whose every step size is a normal distributed random number with mean 0 and standard deviation 1. We generate 12500 streams with length 12500 by the model.

//Discussion about comparison among models fig11.(a)

From Fig.11(a), we can see that both MSWave-L and MSWave-S still saved more transmission cost than CP, PRP, and LEEWAVE-M even though the scale of data set increased. Furthermore, we can notice that the difference between these methods became more distinguished than we saw in temperature data as the data were plotted by semi-log graph. That means the larger data we face, the better MSWave-L and MSWave-S work. Moreover, the gap between MSWave-L and MSWave-S also increase­d as |Q| became larger which was the same with our discussion in the previous.

//Discussion about # site reduced. fig11.(b)

Fig.11(b) shows the performance of pruning of MSWave-L. Although the scale of data increased, the performance of pruning was still good as long as the difference between k (Here was 30) and M (Here was 500) was large. Thanks to the performance of pruning, the reduction in the transmission cost was much more significant than temperature data.

[16] <http://www.cs.ucr.edu/~eamonn/SIGKDD_trillion.pdf> or <http://www.cs.ucr.edu/~eamonn/UCRsuite.html>