A Appendix

A.1 2D Ising model results

In tables 1 and 2 below, we provide the absolute error between the LBP marginals and true marginals (as computed by junction tree algorithm) for different parameter settings. Table 1 is relevant to Figure 3 where the target distribution has multiple modes. Clearly, the LBP approximation starts to degrade as we increase W, which hurts the AAS and AAG samplers, equalizing them with HMC and CMH (see figure 3).

Table 2 is relevant to Figure 5 where the target distribution is uni-modal. As bias increases, the target distribution becomes more peaked with the LBP approximation being consistently accurate, more so with higher bias values. We are able to leverage LBP prior to drive annular augmentation samplers.

Table 9 is relevant to Figure 4. Here we show RMSE of node marginals estimate for AAS and AAS + RB to highlight the benefit of Rao-Blackwellization. The target distribution in this case is bimodal with the two modes corresponding to $\mathbf{s} = \pm 1$ (all coordinates as $+1$ or -1). When we are at one of the modes, say $\mathbf{s} = 1$, the opposite point $\mathbf{s} = -1$ also lies on the annulus (by over relaxed property of the annular augmentation framework). With Rao-Blackwellization, we are able to take advantage of this fact, thereby leading to a more accurate estimate.

Table 1: LBP error for different values of W. The bias scale c is fixed at 0.1

STRENGTH (W)	LBP APPROX. ERROR
0.1	0.0002
0.2	0.0035
0.3	0.0182
0.4	0.0610
0.5	0.1783
0.6	0.8895
0.7	9.3581
0.8	25.5220

A.2 Heart disease dataset

We take samples of W from 1000 Markov chains, as described in Section 3.2, and fit a normal distribution which is shown in figure 7 for three cases: "true" refers to the MH sampler where exact partition function ratio was used, and the other two refer to the approximate MH samplers for which the partition function ratio was approximated by CMH and AAS respectively. As shown in figure 6, AAS approximated the partition function ratio more accurately than CMH and this effect an be clearly seen in figure 7. The estimated posterior distribution of W using CMH is a bit off as compared to the true posterior. The results for HMC and AAG were almost identical (as approximation error for both was similar (see figure 6), and thus we omit these plots.

Table 2: LBP error for different values of c. Strength W is fixed at 0.2

BIAS SCALE (c)	LBP APPROX. ERROR ($\times 10^{-3}$)
0.2	3.44
0.4	5.60
0.6	5.77
0.8	5.10
1.0	4.88
1.5	1.99
2.0	0.55
3.0	0.10
4.0	0.05

Table 3: RMSE of node marginals estimate for increasing values of W. No bias is applied to any node.

STRENGTH (W)	AAS ($\times 10^{-2}$)	AAS + RB ($\times 10^{-2}$)
0.25	1.31	0.03
0.5	1.18	0.03
0.75	1.19	0.03
1.0	1.13	0.03
2.0	1.29	0.03
3.0	1.08	0.03
4.0	1.01	0.03
5.0	1.23	0.03

A.3 CMH + LBP sampler

We provide details of the CMH + LBP sampler as introduced in section 3, and for which the results are shown in figures 3, 4 and 5. Let $\hat{p}(\mathbf{s})$ be the LBP approximation of the target distribution $p(\mathbf{s})$, defined over a d-dimensional binary vector $\mathbf{s} \in \{-1, +1\}^d$. A simple idea is to use $\hat{p}(\mathbf{s})$ as a proposal distribution for CMH, which can be done as:

1. Let $\mathbf{s}^{(t)}$ be the current point.
2. Sample a coordinate i uniformly from $\{1, ..., d\}$
3. With probability $\hat{p}(-s_i)$, propose flipping s_i i.e. propose $\mathbf{s}^{(t+1)} = (-s_i^{(t)}, \mathbf{s_{-i}}^{(t)})$. With probability $\{1 - \hat{p}(-s_i)\}$ stay at the current point.
4. If a flip is proposed, accept it with MH probability $\alpha = \min\left\{1, \frac{p(-s_i, \mathbf{s}_{-i})\hat{p}(s_i)}{p(s_i, \mathbf{s}_{-i})\hat{p}(-s_i)}\right\}$

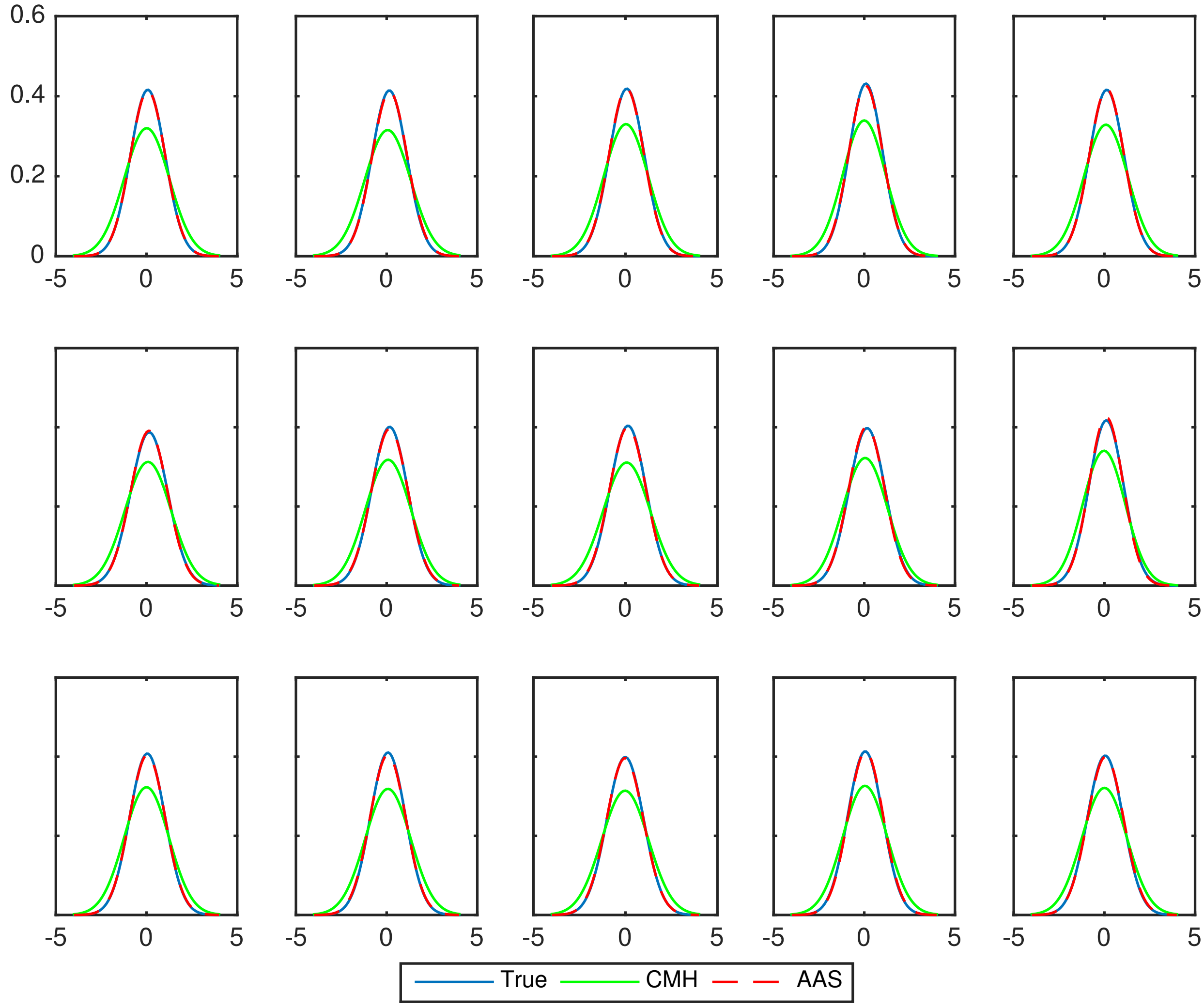

Figure 7: Estimated posterior distribution using approximate MH sampler: relatively large approximation error made by CMH in estimating the partition function ratio affects the quality of the MH sampler. AAS (also AAG and HMC) make a relatively small error and do as well as the ground truth.

It is easy to verify the reversibility of the above algorithm. But this approach has a short-coming. As an illustration, consider a 2D Ising model with a large positive bias applied to all nodes. As explained in section 3.1, the target distribution in this case is uni-modal and highly peaked, with the mode corresponding to $s_i = 1$ for all i. Now suppose that we are at the mode i.e. $s_i^{(t)} = 1$ for all i. Then $\hat{p}(-s_i)$ is quite small for all coordinates i and thus it will take a long time for a flip to be proposed. The sampler gets stuck at the same position for a long time thereby wasting computation. A novel approach to get around this problem is to use the idea of discrete event simulation.

First note that the probability of a flip being proposed in any given iteration is

$$\alpha(\mathbf{s}) = \frac{1}{d}\sum_{j=1}^{d} \hat{p}(-s_j)$$

Therefore the number of iterations until a flip is proposed is a geometric random variable with success probability $\alpha(\mathbf{s})$. The conditional probability that coordinate i is proposed to flip, given that a flip is proposed is:

$$p(s_i^{(t+1)} = -s_i^{(t)} | \mathbf{s}^{(t+1)} \neq \mathbf{s}^{(t)}) \quad = \frac{\hat{p}(-s_i)}{\sum_{i=1}^{d} \hat{p}(-s_i)}.$$

Now the CMH + LBP samplers with discrete event simulation can be summarized as:

1. Let $\mathbf{s}^{(t)}$ be the current point. Compute $\alpha(\mathbf{s}^{(t)}) = \sum_{i=1}^{d} \frac{1}{d} \times \hat{p}(-s_i^{(t)})$
2. Sample a flipping time $k \sim Geometric(\alpha(\mathbf{s}^{(t)}))$. For the next k iterations we stay at the current point i.e. $\mathbf{s}^{(t)} = \mathbf{s}^{(t+1)} = \ldots = \mathbf{s}^{(t+k)}$
3. Propose flipping coordinate i with probability $\frac{\hat{p}(-s_i)}{\sum_{j=1}^{d} \hat{p}(-s_j)}$
4. Accept this MH transition with probability: $\min\left\{1, \frac{p(-s_i, \mathbf{s}_{-i})\hat{p}(s_i)}{p(s_i, \mathbf{s}_{-i})\hat{p}(-s_i)}\right\}$

Intuitively, this saves some computational effort as now we only consider flip events and overall we are able to take more samples.

A.4 Fake data simulations

A.4.1 $d = 10$

We give details of the fake data simulation to get results for Figure 6 (right). Fake data, $[\mathbf{s}_j, \quad j = 1, \ldots, N \in \{0, 1\}^{10}]$, was generated using know parameter values $(\hat{\mathbf{W}}, \hat{\mathbf{b}})$ and then modeled as outputs of a fully connected BM. The connection matrix $\hat{\mathbf{W}}$ was drawn randomly from a unit gaussian i.e. $\hat{W}_{i,j} \sim N(0, 1)$ and a very small bias $\hat{b}_i \sim Unif[-0.1, 0.1]$ was applied to each node i. A data set of 400 binary vectors $\mathcal{D} = \{\mathbf{s}_j, \quad j = 1, \ldots, N\}$ was generated with the likelihood:

$$p(\mathbf{s}_j | \hat{\mathbf{W}}, \hat{\mathbf{b}}) = \frac{1}{Z(\hat{\mathbf{W}}, \hat{\mathbf{b}})} \exp\left\{\sum_{i<j} \hat{W}_{i,j} s_i s_j + \sum_i \hat{b}_i s_i\right\} \tag{6}$$

If no bias is applied, the data essentially exhibits a bi-modal behavior with $\mathbf{s}_j = \{1, \ldots, 1\}$ or $\mathbf{s}_j = \{-1, \ldots, -1\}$; the small bias was applied to break this behavior. As the data is only 10 dimensional, we can sample from the likelihood in (6) exactly. This fake data now is modeled as outputs of a fully connected BM, where a uninformative $N(0, 5)$ prior was placed on the parameters. We ran 1000 Markov chains, each with 100 samples, drawn using the approximate MH sampler. To approximate the partition function ratio in equation (5), we use 500 samples from HMC, CMH, AAS + RB and AAG + RB and plot the average absolute error in approximating the log partition function ratio. We wanted to play around with the number of samples in the inner MCMC loop (to make sure the results are not sensitive to this), and hence 500 samples were used as opposed to 1000 samples used for the real data case.

A.4.2 $d = 25$, 50

We extend the above the simulation setting for $d = 25$ and 50. As before, the connection matrix $\hat{\mathbf{W}}$ was drawn randomly from a unit gaussian i.e. $\hat{W}_{i,j} \sim N(0, 1)$ and a very small bias $\hat{b}_i \sim Unif[-0.1, 0.1]$ was applied to each node i. As exact sampling is computationally expensive for $d = 25$, 50 we used Exact-HMC Pakman and Paninski (2013) to sample from the likelihood in equation (6). A large number of samples were generated and appropriately thinned to get 200 fake data points which were then modeled as outputs of a fully connected Boltzmann machine. We ran 200 Markov chains, each with 100 samples, drawn with the approximate MH sampler; with 500 samples used in the inner MCMC loop to approximate the partition function. Following Murray and Ghahramani (2004), in figure we plot the fraction of samples (samples of $W_{i,j}$), within ±20% of the true value for every parameter $\hat{W}_{i,j}$ of this (300 edge) system.

A.5 Numerical Values:

We provide the numerical values used to generate the heat maps in Figures 3, 4, and 5.

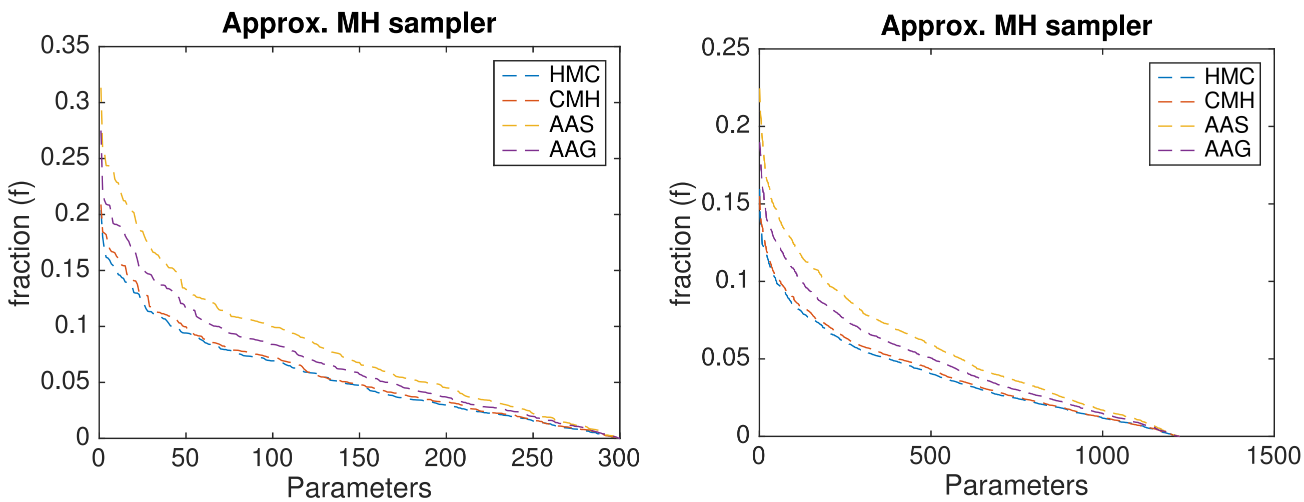

Figure 8: Comparisons for different MCMC samplers used as inner loops in the approximate MH sampler. We plot the fraction of samples, f, within ±20% of the true value for every parameter in the fully connected Boltzmann machine for $d = 25$ (left) and $d = 50$ (right). Higher curves indicate better performance.

Table 4: Numerical values of RMSE (Node Marginals) for Figure 3. Bias scale $c = 0.2$ and we increase the strength (W)

ALGORITHM	W=0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8
HMC	0.0211	0.0220	0.0513	0.1498	0.2907	0.4515	0.4117	0.4749
CMH	0.0134	0.0188	0.0475	0.1583	0.3990	0.4670	0.3992	0.4752
CMH + LBP	0.0132	0.0168	0.0466	0.1576	0.3965	0.4527	0.3992	0.4752
AAS	0.0267	0.0416	0.0761	0.1495	0.1585	0.2238	0.2564	0.0558
AAS + RB	0.0214	0.0385	0.0741	0.1490	0.1587	0.2243	0.2563	0.0547
AAS + RB + LBP	0.0097	0.0221	0.0664	0.1418	0.2358	0.3200	0.3884	0.4741
AAG	0.0225	0.0352	0.0635	0.1566	0.1393	0.2694	0.3017	0.0597
AAG + RB	0.0169	0.0313	0.0609	0.1558	0.1419	0.2692	0.3042	0.0581
AAG + RB + LBP	0.0081	0.0166	0.0600	0.1580	0.1661	0.3559	0.3962	0.4752
AAST	0.0407	0.0359	0.0675	0.1205	0.1605	0.3081	0.3128	0.0598
AAST + RB	0.0404	0.0352	0.0671	0.1201	0.1637	0.3073	0.3128	0.0594
AAST + RB + LBP	0.0285	0.0223	0.0536	0.1766	0.1817	0.3921	0.3971	0.4746

Table 5: Numerical values of RMSE (Pairwise Marginals) for Figure 4. Bias scale $c = 0.2$ and we increase the strength (W)

ALGORITHM	W=0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8
HMC	0.0211	0.0220	0.0513	0.1498	0.2907	0.4515	0.4117	0.4749
CMH	0.0134	0.0188	0.0475	0.1583	0.3990	0.4670	0.3992	0.4752
CMH + LBP	0.0132	0.0168	0.0466	0.1576	0.3965	0.4527	0.3992	0.4752
AAS	0.0267	0.0416	0.0761	0.1495	0.1585	0.2238	0.2564	0.0558
AAS + RB	0.0214	0.0385	0.0741	0.1490	0.1587	0.2243	0.2563	0.0547
AAS + RB + LBP	0.0097	0.0221	0.0664	0.1418	0.2358	0.3200	0.3884	0.4741
AAG	0.0225	0.0352	0.0635	0.1566	0.1393	0.2694	0.3017	0.0597
AAG + RB	0.0169	0.0313	0.0609	0.1558	0.1419	0.2692	0.3042	0.0581
AAG + RB + LBP	0.0081	0.0166	0.0600	0.1580	0.1661	0.3559	0.3962	0.4752
AAST	0.0407	0.0359	0.0675	0.1205	0.1605	0.3081	0.3128	0.0598
AAST + RB	0.0404	0.0352	0.0671	0.1201	0.1637	0.3073	0.3128	0.0594
AAST + RB + LBP	0.0285	0.0223	0.0536	0.1766	0.1817	0.3921	0.3971	0.4746

Table 6: Numerical values of RMSE (Node Marginals) for Figure 4. No bias is applied to any node and we increase the strength (W).

ALGORITHM	W=0.25	0.5	0.75	1	2	3	4	5
HMC	0.0744	0.1183	0.2173	0.4947	0.8978	1.1969	1.0837	1.1509
CMH	0.0590	0.1148	0.2131	0.5072	1.0911	1.2287	1.0594	1.1517
CMH + LBP	0.0576	0.1123	0.2124	0.5030	1.0848	1.2015	1.0595	1.1519
AAS	0.0913	0.1595	0.2757	0.4734	0.5462	0.6680	0.7240	0.3864
AAS + RB	0.0776	0.1520	0.2711	0.4721	0.5464	0.6687	0.7237	0.3853
AAS + RB + LBP	0.0614	0.1273	0.2590	0.4652	0.6926	0.8719	1.0223	1.1478
AAG	0.0818	0.1434	0.2465	0.5005	0.5497	0.7834	0.8410	0.3986
AAG + RB	0.0676	0.1350	0.2416	0.4988	0.5544	0.7830	0.8460	0.3978
AAG + RB + LBP	0.0585	0.1212	0.2435	0.4970	0.5961	0.9621	1.0493	1.1518
AAST	0.1240	0.1453	0.2545	0.4267	0.5883	0.8768	0.8699	0.3966
AAST + RB	0.1231	0.1436	0.2532	0.4247	0.5915	0.8749	0.8698	0.3964
AAST + RB + LBP	0.0598	0.1217	0.2286	0.5576	0.6239	1.0547	1.0522	1.1495

Table 7: Numerical values of RMSE (Pairwise Marginals) for Figure 4. No bias is applied to any node and we increase the strength (W).

ALGORITHM	W=0.25	0.5	0.75	1	2	3	4	5
HMC	0.1524	1.0189	1.2243	1.0591	0.9981	1.0291	1.0194	1.0089
CMH	0.1515	1.0731	1.2237	1.0542	0.9979	1.0045	1.0506	1.0296
CMH + LBP	0.1514	1.1662	1.2215	1.0465	0.9997	1.0505	1.0402	1.0300
AAS	0.1735	0.4281	0.3581	0.2211	0.3481	0.2603	0.2992	0.2797
AAS + RB	0.1695	0.4272	0.3574	0.2193	0.3466	0.2590	0.2984	0.2781
AAS + RB + LBP	0.1739	0.4158	0.3191	0.2608	0.2632	0.2902	0.3504	0.3653
AAG	0.1682	0.4943	0.3828	0.2943	0.2102	0.2206	0.2687	0.2803
AAG + RB	0.1647	0.4938	0.3823	0.2934	0.2061	0.2196	0.2660	0.2792
AAG + RB + LBP	0.1659	0.4812	0.3778	0.1813	0.2565	0.2687	0.2659	0.3266
AAST	0.1674	0.4869	0.3788	0.1151	0.2658	0.1427	0.2347	0.2527
AAST + RB	0.1701	0.5010	0.3947	0.1566	0.1647	0.2753	0.2224	0.2688
AAST + RB + LBP	0.1666	0.4624	0.3915	0.1518	0.2993	0.2452	0.2323	0.2302

Table 8: Numerical values of RMSE (Node Marginals) for Figure 5. We fix $W = 0.2$ and the bias scale c is increased.

ALGORITHM	c=0.2	0.4	0.6	0.8	1	1.5	2	3	4
HMC	0.0262	0.0360	0.0381	0.0444	0.0444	0.0432	0.0442	0.0502	0.0351
CMH	0.0233	0.0347	0.0351	0.0430	0.0437	0.0413	0.0421	0.0489	0.0327
CMH + LBP	0.0217	0.0332	0.0333	0.0421	0.0423	0.0403	0.0418	0.0482	0.0318
AAS	0.0462	0.0681	0.0754	0.0766	0.0801	0.0789	0.0806	0.0819	0.0637
AAS + RB	0.0431	0.0658	0.0731	0.0749	0.0785	0.0777	0.0797	0.0811	0.0630
AAS + RB + LBP	0.0244	0.0419	0.0416	0.0525	0.0476	0.0423	0.0418	0.0479	0.0319
AAG	0.0406	0.0492	0.0666	0.0694	0.0677	0.0626	0.0688	0.0774	0.0645
AAG + RB	0.0386	0.0471	0.0641	0.0674	0.0660	0.0616	0.0678	0.0767	0.0636
AAG + RB + LBP	0.0229	0.0365	0.0381	0.0461	0.0463	0.0410	0.0423	0.0483	0.0316
AAST	0.0312	0.0573	0.0589	0.0619	0.0702	0.0732	0.0572	0.0659	0.0555
AAST + RB	0.0445	0.0597	0.0675	0.0649	0.0786	0.0782	0.0499	0.0586	0.0430
AAST + RB + LBP	0.0230	0.0462	0.0463	0.0459	0.0442	0.0513	0.0324	0.0505	0.0337

Table 9: Numerical values of RMSE (Pairwise Marginals) for Figure 5. We fix $W = 0.2$ and the bias scale c is increased.

ALGORITHM	c=0.2	0.4	0.6	0.8	1	1.5	2	3	4
HMC	0.1229	0.1327	0.1313	0.1430	0.1429	0.1297	0.1313	0.1477	0.1037
CMH	0.1187	0.1300	0.1226	0.1384	0.1399	0.1238	0.1237	0.1427	0.0962
CMH + LBP	0.1165	0.1269	0.1188	0.1363	0.1365	0.1206	0.1224	0.1403	0.0934
AAS	0.1635	0.2085	0.2226	0.2269	0.2346	0.2326	0.2380	0.2406	0.1882
AAS + RB	0.1554	0.2015	0.2159	0.2213	0.2294	0.2285	0.2349	0.2383	0.1862
AAS + RB + LBP	0.1278	0.1503	0.1392	0.1648	0.1505	0.1272	0.1227	0.1397	0.0934
AAG	0.1498	0.1781	0.1902	0.2001	0.2155	0.2045	0.2116	0.2186	0.1724
AAG + RB	0.1425	0.1711	0.1833	0.1941	0.2099	0.2003	0.2081	0.2161	0.1703
AAG + RB + LBP	0.1251	0.1375	0.1327	0.1498	0.1482	0.1234	0.1239	0.1406	0.0930
AAST	0.1479	0.1814	0.1892	0.2080	0.1980	0.2102	0.1772	0.1940	0.1527
AAST + RB	0.1563	0.1852	0.1936	0.1918	0.2283	0.2297	0.1474	0.1774	0.1250
AAST + RB + LBP	0.1225	0.1379	0.1411	0.1466	0.1354	0.1409	0.0977	0.1227	0.0937