Supplementary 5 – Outputs for the 53 pipeline (10%) and 79 pipeline (15%) samples

Output 1: Median Predictive Accuracy and Kolmogorov-Smirnov Statistic

Sample Size = 53 pipelines (10% of full multiverse)

Sampling Approach	Median R^2	K-S Statistic
Full multiverse	0.064	
Active learning	0.066	0.144
Random	0.066	0.088
Stratified	0.077	0.210

Sample Size = 79 pipelines (15% of full multiverse)

Sampling Approach	Median R^2	K-S Statistic
Full multiverse	0.064	
Active learning	0.065	0.093
Random	0.051	0.154
Stratified	0.067	0.068

Output 2: Predictive Accuracy of the Pipelines

Sample Size = 53 pipelines (10% of full multiverse)

Best Full Sample	R^2	Best Active Learning	R^2	Best Random Sampling	R^2	Best Stratified Sampling	R^2
b-200, rMas, 500200, CP1CP2	**0.203**	b-100, rAvg, SAV400, CP1CP2PzP3P4	0.157	**b-200, rMas, 600200, CP1CP2**	**0.173**	**b-200, rAvg, 500200, CP1CP2PzP3P4**	**0.177**
b-100, rMas, 500200, CP1CP2	**0.188**	b-200, rAvg, SAV400, CP1CP2PzP3P4	0.157	b-200, rCSD, 500200, CP1CP2PzP3P4	0.147	**b-200, rMas, 600200, CP1CP2**	**0.173**
b-200, rMas, SAV400, CP1CP2PzP3P4	**0.184**	b-100, rAvg, 500200, CP1CP2PzP3P4	0.155	b-200, rAvg, 500200, P3P4CP1CP2	0.146	**b-100, rMas, 600200, CP1CP2**	**0.169**
b-100, rMas, SAV400, CP1CP2PzP3P4	**0.182**	b-200, rAvg, 500200, CP1CP2PzP3P4	0.155	b-100, rAvg, 500200, CP1CP2PzP3P4	0.141	b-200, rCSD, 500200, CP1CP2PzP3P4	0.147
b-200, rAvg, 500200, CP1CP2PzP3P4	**0.177**	b-100, rCSD, 500200, P3P4CP1CP2	0.151	b-100, rMas, 500300, CP1CP2	0.137	b-200, rAvg, 500200, P3P4CP1CP2	0.146
b-200, rMas, 700200, CP1CP2	**0.173**	b-100, rCSD, 500200, CP1CP2	0.150	b-200, rCSD, 500200, around	0.127	b-100, rAvg, 500200, CP1CP2PzP3P4	0.141
b-200, rMas, 600200, CP1CP2	**0.173**	b-100, rCSD, 500200, Cz	0.149	b-100, rMas, 450100, CP1CP2	0.120	b-100, rMas, 500300, CP1CP2	0.137
b-200, rAvg, SAV400, CP1CP2PzP3P4	**0.171**	b-200, rCSD, 500200, CP1CP2	0.147	b-200, rAvg, GAV400, CP1CP2PzP3P4	0.113	b-200, rCSD, 500200, around	0.127
b-100, rMas, 600200, CP1CP2	**0.169**	b-100, rCSD, 500200, CP1CP2PzP3P4	0.147	b-200, rCSD, 500200, Pz	0.110	b-200, rAvg, 450100, CP1CP2PzP3P4	0.125
b-100, rAvg, SAV400, CP1CP2PzP3P4	**0.168**	b-200, rCSD, 500200, P3P4CP1CP2	0.147	b-100, rAvg, 700300, CP1CP2PzP3P4	0.109	b-200, rAvg, 500200, Cz	0.121

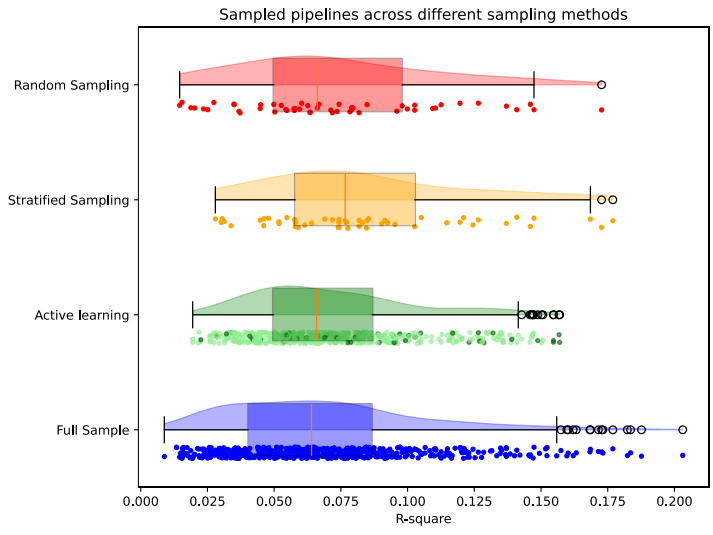
Sample Size = 79 pipelines (15% of full multiverse)

Best Full Sample	R^2	Best Active Learning	R^2	Best Random Sampling	R^2	Best Stratified Sampling	R^2
b-200, rMas, 500200, CP1CP2	**0.203**	b-100, rMas, SAV400, CP1CP2PzP3P4	0.160	b-100, rCSD, 500200, P3P4CP1CP2	0.156	**b-200, rMas, 600200, CP1CP2**	**0.173**
b-100, rMas, 500200, CP1CP2	**0.188**	b-100, rAvg, SAV400, CP1CP2PzP3P4	0.157	b-100, rCSD, 500200, around	0.152	b-100, rCSD, 500200, CP1CP2	0.162
b-200, rMas, SAV400, CP1CP2PzP3P4	**0.184**	b-200, rAvg, SAV400, CP1CP2PzP3P4	0.157	b-200, rAvg, 700200, CP1CP2PzP3P4	0.152	b-200, rCSD, 500200, CP1CP2PzP3P4	0.147
b-100, rMas, SAV400, CP1CP2PzP3P4	**0.182**	b-100, rAvg, 500200, CP1CP2PzP3P4	0.156	b-200, rMas, SAV400, CP1CP2	0.147	b-200, rCSD, 500200, CP1CP2	0.142
b-200, rAvg, 500200, CP1CP2PzP3P4	**0.177**	b-200, rAvg, 500200, CP1CP2PzP3P4	0.156	b-200, rMas, 700200, P3P4CP1CP2	0.137	b-100, rAvg, 500200, CP1CP2	0.138
b-200, rMas, 700200, CP1CP2	**0.173**	b-100, rCSD, 500200, P3P4CP1CP2	0.152	b-200, rMas, 600200, P3P4CP1CP2	0.131	b-100, rMas, 500300, CP1CP2	0.137
b-200, rMas, 600200, CP1CP2	**0.173**	b-100, rCSD, 500200, CP1CP2	0.152	b-200, rAvg, 700200, P3P4CP1CP2	0.116	b-200, rCSD, 500200, Cz	0.128
b-200, rAvg, SAV400, CP1CP2PzP3P4	**0.171**	b-100, rCSD, 500200, Cz	0.150	b-100, rAvg, 600200, CP1CP2PzP3P4	0.115	b-200, rCSD, 500200, around	0.127
b-100, rMas, 600200, CP1CP2	**0.169**	b-200, rMas, SAV400, CP1CP2PzP3P4	0.149	b-100, rAvg, SAV400, P3PzP4	0.103	b-200, rAvg, 500200, around	0.123
b-100, rAvg, SAV400, CP1CP2PzP3P4	**0.168**	b-100, rCSD, 500200, CP1CP2PzP3P4	0.149	b-200, rAvg, 700600, CP1CP2	0.099	b-100, rAvg, 500200, P3P4CP1CP2	0.121

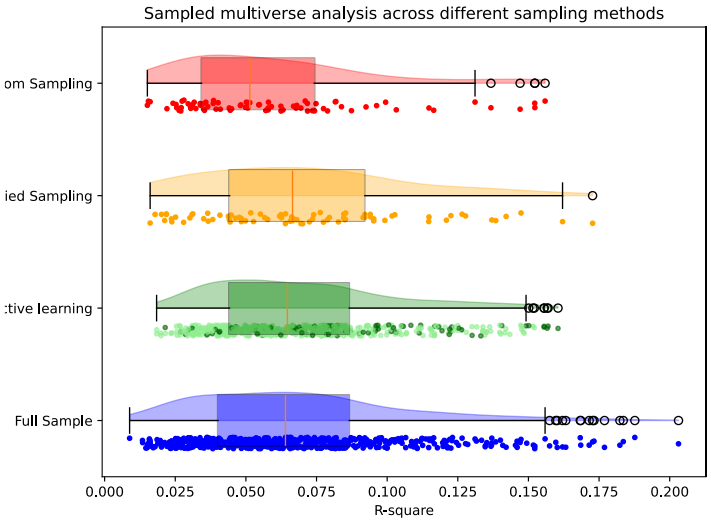
Blue = the Pipeline is included within the best pipelines of the full multiverse. b-100 and b-200 = baseline durations of 100 and 200 ms respectively, rAvg, rMas and rCSD = common average, linked mastoids and common source density reference schemes respectively.

Output 3: Raincloud Plots of Predictive Accuracies

Sample Size = 53 pipelines (10% of full multiverse)



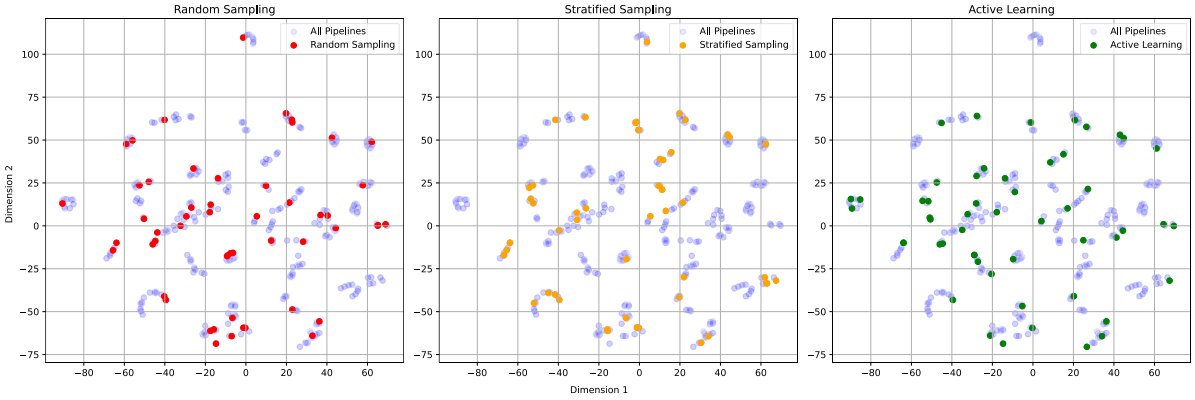
Sample Size = 79 pipelines (15% of full multiverse)



Output 4: Scatter Plots of Spatial Distribution in the Low Dimensional Space

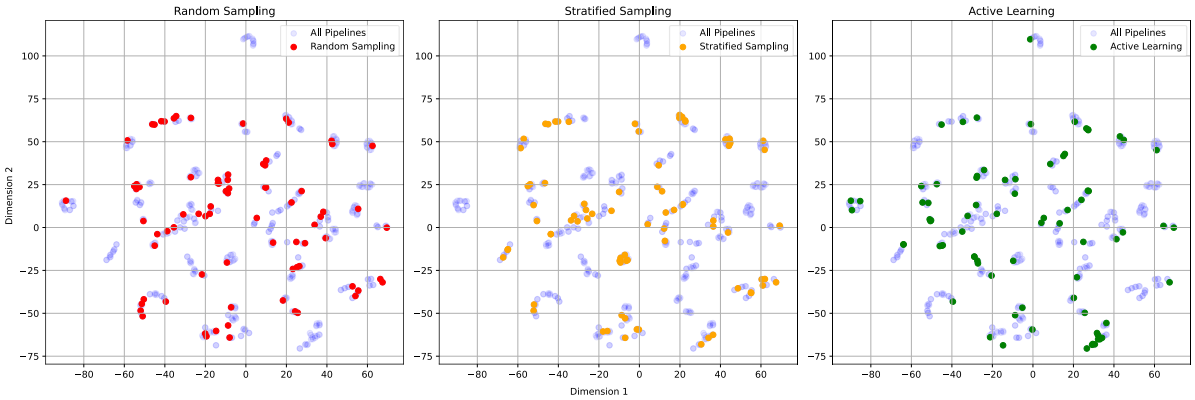
Sample Size = 53 pipelines (10% of full multiverse)

Sample Size	Full Multiverse vs. Random Sample	Full Multiverse vs. Stratified Sample	Full Multiverse vs. Active Learning Sample
53 pipelines	70.90	72.53	70.78



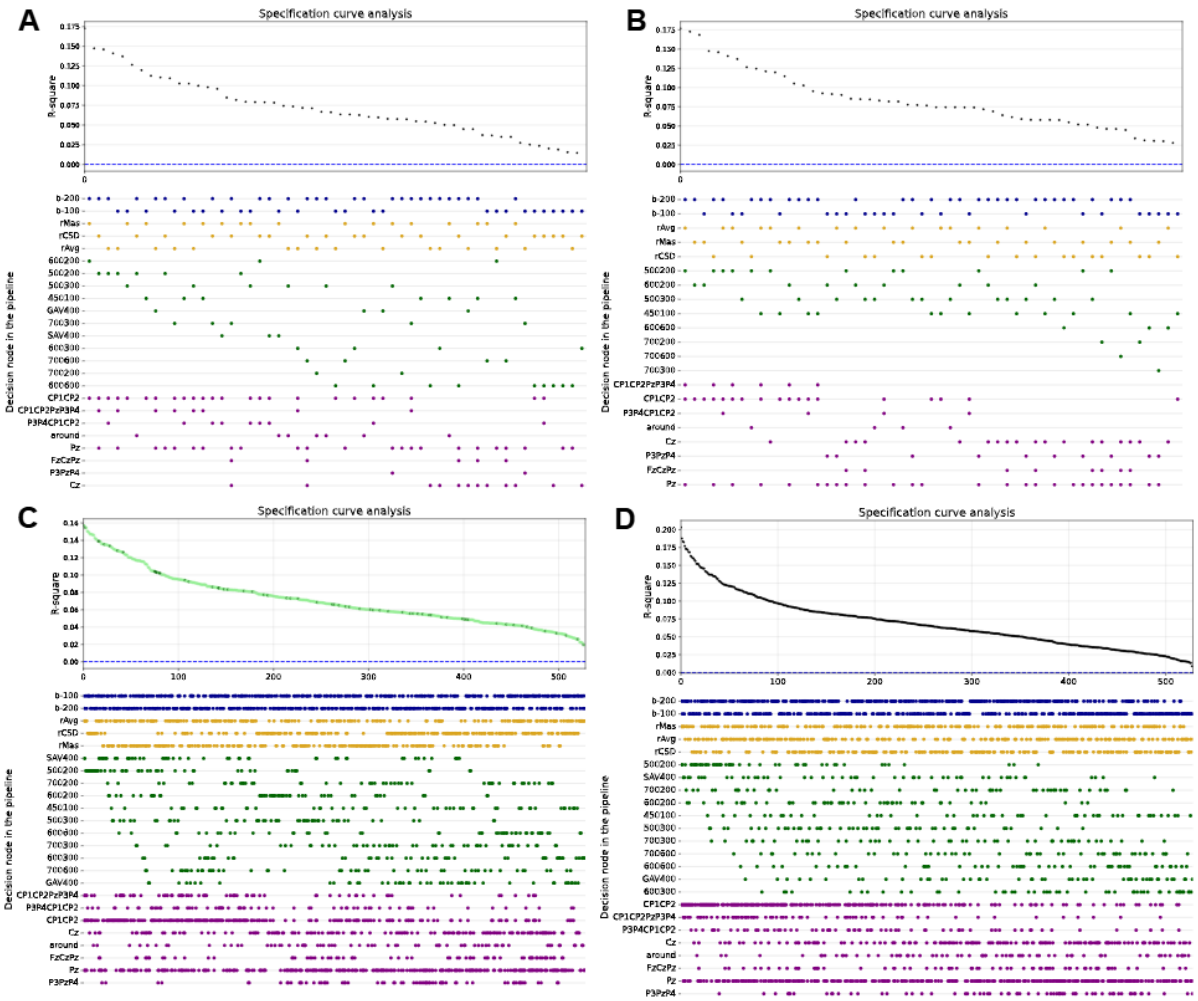
Sample Size = 79 pipelines (15% of full multiverse)

Sample Size	Full Multiverse vs. Random Sample	Full Multiverse vs. Stratified Sample	Full Multiverse vs. Active Learning Sample
79 pipelines	69.09	70.92	71.63



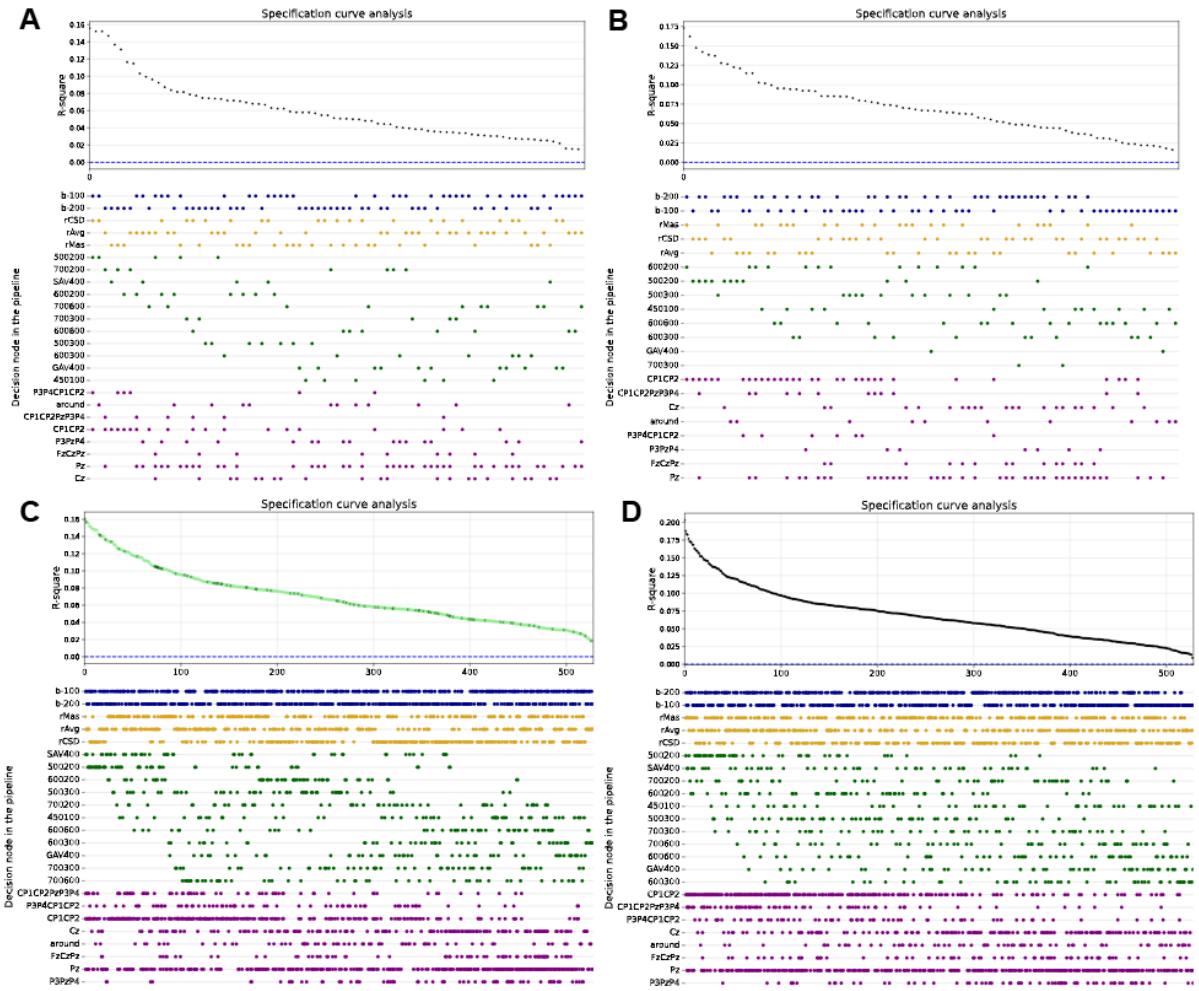
Output 5: Specification Curves

<u>Sample Size = 53 pipelines (10% of full multiverse)</u>



Specification curves displaying the variability in the R^2 across each sample, in vertical alignment with the respective pipeline options. Panel A = random sample, Panel B = stratified sample, Panel C = active learning sample, Panel D = full multiverse. Each colour in the lower specification panel corresponds to one decision node. Blue = baseline duration, yellow = reference scheme, green = time window, purple = electrode cluster. Each row within a colour represents a different option at that decision node. In the top panel of the active learning plot, dark green points denote pipelines that were directly sampled and light green points denote pipelines that were estimated.

Sample Size = 79 pipelines (15% of full multiverse)



Specification curves displaying the variability in the R^2 across each sample, in vertical alignment with the respective pipeline options. Panel A = random sample, Panel B = stratified sample, Panel C = active learning sample, Panel D = full multiverse. Each colour in the lower specification panel corresponds to one decision node. Blue = baseline duration, yellow = reference scheme, green = time window, purple = electrode cluster. Each row within a colour represents a different option at that decision node. In the top panel of the active learning plot, dark green points denote pipelines that were directly sampled and light green points denote pipelines that were estimated.