Supplementary 4 – 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)

Test subset of participants (*N* = 50):

Sampling Approach	Median R^2	K-S Statistic
Full multiverse	0.084	
Active learning	0.079	0.051
Random	0.079	0.067
Stratified	0.081	0.073

Lockbox set of participants (*N* = 28):

Sampling Approach	Median R^2	K-S Statistic
Full multiverse	0.138	
Active learning	0.152	0.093*
Random	0.130	0.100
Stratified	0.138	0.095

* = $p < 0.05$; **= $p < 0.01$; *** = $p < 0.001$

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

Test subset of participants (*N* = 50):

Sampling Approach	Median R^2	K-S Statistic
Full multiverse	0.084	
Active learning	0.079	0.083
Random	0.078	0.103
Stratified	0.085	0.080

Lockbox set of participants (*N* = 28):

Sampling Approach	Median R^2	K-S Statistic
Full multiverse	0.138	
Active learning	0.130	0.081
Random	0.123	0.083
Stratified	0.143	0.074

Output 2: Predictive Accuracy of the Pipelines

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

Test subset of participants (*N* = 50):

Best Full Sample	R^2	Best Active Learning	R^2	Best Random Sampling	R^2	Best Stratified Sampling	R^2
b-100, rMas, 700300, Pz	**0.301**	**b-100, rMas, 700300, Pz**	**0.301**	b-100, rCSD, 500200, P3P4CP1CP2	0.237	**b-100, rMas, 700300, Pz**	**0.301**
b-100, rMas, 700300, around	**0.281**	**b-200, rMas, 700300, Pz**	**0.281**	b-200, rMas, 700200, around	0.222	**b-100, rMas, 700600, Pz**	**0.257**
b-200, rMas, 700300, Pz	**0.281**	**b-100, rMas, 700600, Pz**	**0.257**	b-100, rAvg, 500200, P3P4CP1CP2	0.217	**b-200, rMas, 700300, around**	**0.254**
b-100, rMas, 700600, Pz	**0.257**	**b-100, rMas, GAV400, Pz**	**0.248**	b-100, rMas, 700300, FzCzPz	0.212	**b-100, rMas, 700600, around**	**0.249**
b-200, rMas, 700300, around	**0.254**	**b-100, rAvg, 500200, Cz**	**0.243**	b-200, rMas, 450100, Cz	0.176	b-100, rMas, 700300, Cz	0.236
b-100, rMas, 700600, around	**0.249**	b-100, rMas, 700200, around	0.232	b-100, rMas, 500200, CP1CP2	0.169	b-100, rCSD, 500200, CP1CP2PzP3P4	0.222
b-100, rMas, GAV400, Pz	**0.248**	b-200, rMas, 700600, Pz	0.228	b-100, rMas, 600300, around	0.166	b-200, rCSD, 500200, P3P4CP1CP2	0.206
b-100, rMas, GAV400, around	**0.247**	b-100, rMas, 500200, Cz	0.222	b-200, rCSD, 700300, CP1CP2PzP3P4	0.164	b-200, rAvg, 500200, around	0.192
b-100, rCSD, 500200, CP1CP2	**0.246**	b-200, rMas, GAV400, Pz	0.215	b-200, rMas, 700200, FzCzPz	0.162	b-100, rAvg, 500200, Pz	0.184
b-100, rAvg, 500200, Cz	**0.243**	b-100, rMas, 700300, FzCzPz	0.212	b-200, rMas, 600600, Pz	0.159	b-200, rMas, 700200, Pz	0.178

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.

Lockbox set of participants (*N* = 28):

Best Full Sample	R^2	Best Active Learning	R^2	Best Random Sampling	R^2	Best Stratified Sampling	R^2
b-200, rMas, SAV400, CP1CP2PzP3P4	**0.522**	**b-100, rAvg, 700200, CP1CP2PzP3P4**	**0.512**	b-100, rMas, 700200, CP1CP2	0.386	**b-100, rAvg, 700200, CP1CP2**	**0.441**
b-100, rAvg, 700200, CP1CP2PzP3P4	**0.512**	**b-200, rMas, SAV400, CP1CP2**	**0.452**	b-200, rMas, 600300, P3P4CP1CP2	0.367	b-200, rAvg, 700200, P3P4CP1CP2	0.380
b-200, rAvg, 700200, CP1CP2PzP3P4	**0.502**	**b-100, rAvg, SAV400, CP1CP2PzP3P4**	**0.451**	b-200, rAvg, 600300, P3P4CP1CP2	0.338	b-100, rAvg, 600600, CP1CP2PzP3P4	0.376
b-100, rMas, SAV400, CP1CP2	**0.469**	**b-100, rAvg, 600200, CP1CP2PzP3P4**	**0.450**	b-100, rMas, 600200, CP1CP2PzP3P4	0.336	b-200, rAvg, GAV400, CP1CP2PzP3P4	0.369
b-200, rAvg, SAV400, CP1CP2PzP3P4	**0.458**	b-100, rMas, 500200, CP1CP2PzP3P4	0.411	b-200, rMas, 600600, P3P4CP1CP2	0.333	b-200, rAvg, GAV400, CP1CP2	0.362
b-200, rMas, SAV400, CP1CP2	**0.452**	b-200, rAvg, 700300, CP1CP2	0.367	b-200, rAvg, GAV400, P3P4CP1CP2	0.317	b-100, rMas, 600300, CP1CP2PzP3P4	0.344
b-100, rAvg, SAV400, CP1CP2PzP3P4	**0.451**	b-200, rMas, 700600, CP1CP2	0.367	b-200, rAvg, 500200, CP1CP2PzP3P4	0.296	b-100, rMas, 450100, CP1CP2	0.280
b-100, rAvg, 600200, CP1CP2PzP3P4	**0.450**	b-200, rAvg, 600300, CP1CP2	0.353	b-100, rMas, 500200, CP1CP2	0.287	b-100, rMas, 700600, CP1CP2	0.259
b-200, rAvg, 700200, CP1CP2	**0.443**	b-100, rMas, 700300, CP1CP2PzP3P4	0.339	b-100, rMas, 600300, P3PzP4	0.287	b-200, rAvg, 500300, P3PzP4	0.247
b-100, rAvg, 700200, CP1CP2	**0.441**	b-200, rMas, SAV400, P3P4CP1CP2	0.328	b-100, rMas, 450100, CP1CP2	0.280	b-100, rMas, 450100, P3PzP4	0.236

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.

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

Test subset of participants (*N* = 50):

Best Full Sample	R^2	Best Active Learning	R^2	Best Random Sampling	R^2	Best Stratified Sampling	R^2
b-100, rMas, 700300, Pz	**0.301**	**b-100, rMas, 700300, Pz**	**0.301**	**b-200, rMas, 700300, Pz**	**0.281**	**b-100, rMas, GAV400, Pz**	**0.248**
b-100, rMas, 700300, around	**0.281**	**b-200, rMas, 700300, Pz**	**0.281**	**b-100, rMas, GAV400, Pz**	**0.248**	b-100, rCSD, 500200, Cz	0.242
b-200, rMas, 700300, Pz	**0.281**	**b-100, rMas, 700600, Pz**	**0.257**	**b-100, rAvg, 500200, Cz**	**0.243**	b-100, rMas, 700200, around	0.232
b-100, rMas, 700600, Pz	**0.257**	**b-200, rMas, 700300, around**	**0.254**	b-100, rCSD, 500200, P3P4CP1CP2	0.237	b-100, rCSD, 500200, around	0.230
b-200, rMas, 700300, around	**0.254**	**b-100, rMas, GAV400, Pz**	**0.248**	b-100, rCSD, 500200, around	0.230	b-100, rMas, 450100, around	0.229
b-100, rMas, 700600, around	**0.249**	**b-100, rAvg, 500200, Cz**	**0.243**	b-200, rCSD, 500200, around	0.200	b-100, rAvg, 500200, around	0.224
b-100, rMas, GAV400, Pz	**0.248**	b-200, rMas, 500200, P3P4CP1CP2	0.237	b-100, rMas, 450100, Cz	0.194	b-100, rAvg, 500200, CP1CP2	0.213
b-100, rMas, GAV400, around	**0.247**	b-100, rMas, 700200, around	0.232	b-100, rMas, GAV400, Cz	0.184	b-200, rMas, 700300, Cz	0.210
b-100, rCSD, 500200, CP1CP2	**0.246**	b-100, rMas, 450100, around	0.229	b-100, rCSD, 500200, P3PzP4	0.175	b-100, rMas, 600600, around	0.205
b-100, rAvg, 500200, Cz	**0.243**	b-200, rMas, 700200, around	0.222	b-200, rAvg, 700300, CP1CP2PzP3P4	0.141	b-100, rMas, 500200, P3P4CP1CP2	0.204

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.

Lockbox set of participants (*N* = 28):

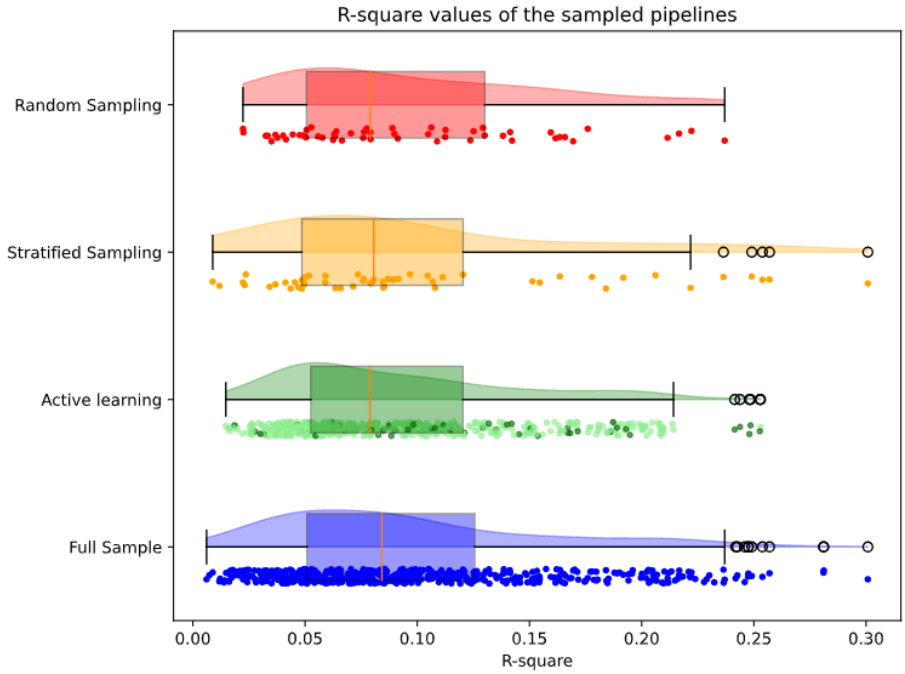
Best Full Sample	R^2	Best Active Learning	R^2	Best Random Sampling	R^2	Best Stratified Sampling	R^2
b-200, rMas, SAV400, CP1CP2PzP3P4	**0.522**	**b-200, rMas, SAV400, CP1CP2**	**0.452**	**b-100, rMas, 500200, CP1CP2PzP3P4**	**0.411**	**b-100, rAvg, 700200, CP1CP2PzP3P4**	**0.512**
b-100, rAvg, 700200, CP1CP2PzP3P4	**0.512**	**b-100, rAvg, SAV400, CP1CP2PzP3P4**	**0.451**	b-200, rMas, 450100, CP1CP2PzP3P4	0.387	**b-100, rAvg, SAV400, CP1CP2PzP3P4**	**0.451**
b-200, rAvg, 700200, CP1CP2PzP3P4	**0.502**	**b-100, rAvg, 600200, CP1CP2PzP3P4**	**0.450**	b-200, rMas, 700600, P3P4CP1CP2	0.387	b-200, rMas, 700300, CP1CP2	0.376
b-100, rMas, SAV400, CP1CP2	**0.469**	**b-100, rAvg, 700200, CP1CP2**	**0.441**	b-200, rAvg, 700200, P3P4CP1CP2	0.380	b-100, rAvg, 600600, CP1CP2PzP3P4	0.376
b-200, rAvg, SAV400, CP1CP2PzP3P4	**0.458**	b-200, rMas, 700300, CP1CP2PzP3P4	0.411	b-200, rMas, 700300, CP1CP2	0.376	b-200, rMas, 700600, CP1CP2	0.367
b-200, rMas, SAV400, CP1CP2	**0.452**	b-200, rMas, 600300, CP1CP2PzP3P4	0.401	b-200, rAvg, 700600, CP1CP2	0.375	b-100, rMas, 700300, CP1CP2PzP3P4	0.339
b-100, rAvg, SAV400, CP1CP2PzP3P4	**0.451**	b-200, rMas, 700300, CP1CP2	0.376	b-200, rAvg, 600300, CP1CP2	0.353	b-200, rAvg, 600300, P3P4CP1CP2	0.338
b-100, rAvg, 600200, CP1CP2PzP3P4	**0.450**	b-200, rAvg, GAV400, CP1CP2	0.362	b-200, rAvg, 700300, CP1CP2PzP3P4	0.343	b-100, rMas, 600200, P3P4CP1CP2	0.314
b-200, rAvg, 700200, CP1CP2	**0.443**	b-200, rAvg, 600300, CP1CP2	0.353	b-200, rAvg, 500300, CP1CP2PzP3P4	0.318	b-100, rMas, SAV400, P3P4CP1CP2	0.309
b-100, rAvg, 700200, CP1CP2	**0.441**	b-200, rAvg, 700300, CP1CP2PzP3P4	0.343	b-100, rMas, 600600, CP1CP2PzP3P4	0.312	b-200, rAvg, 700300, P3P4CP1CP2	0.298

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.

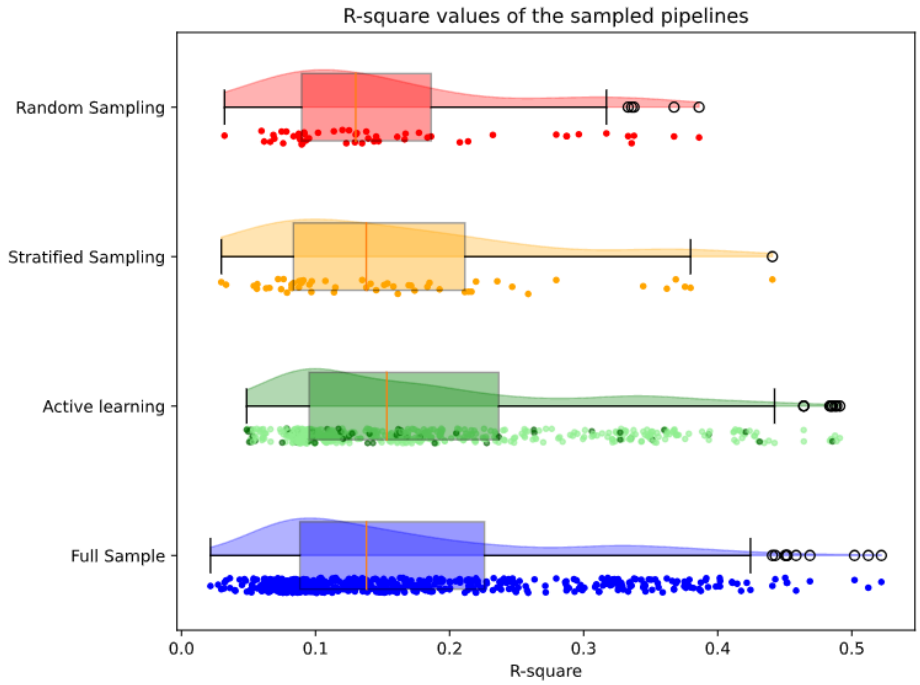
Output 3: Raincloud Plots of Predictive Accuracies

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

Test subset of participants (*N* = 50):

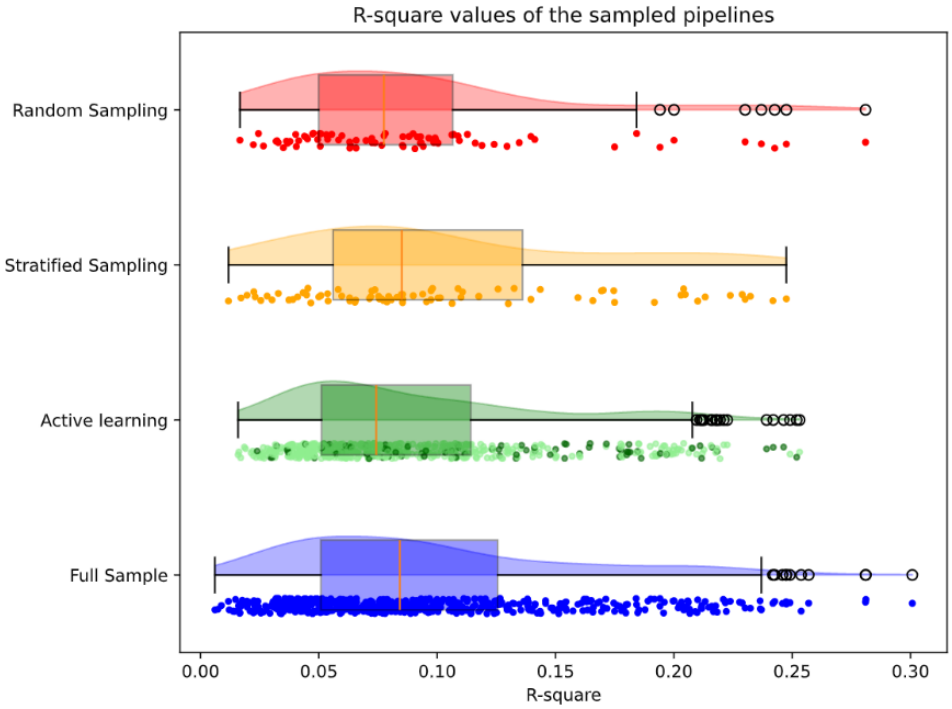


Lockbox subset of participants (*N* = 28):

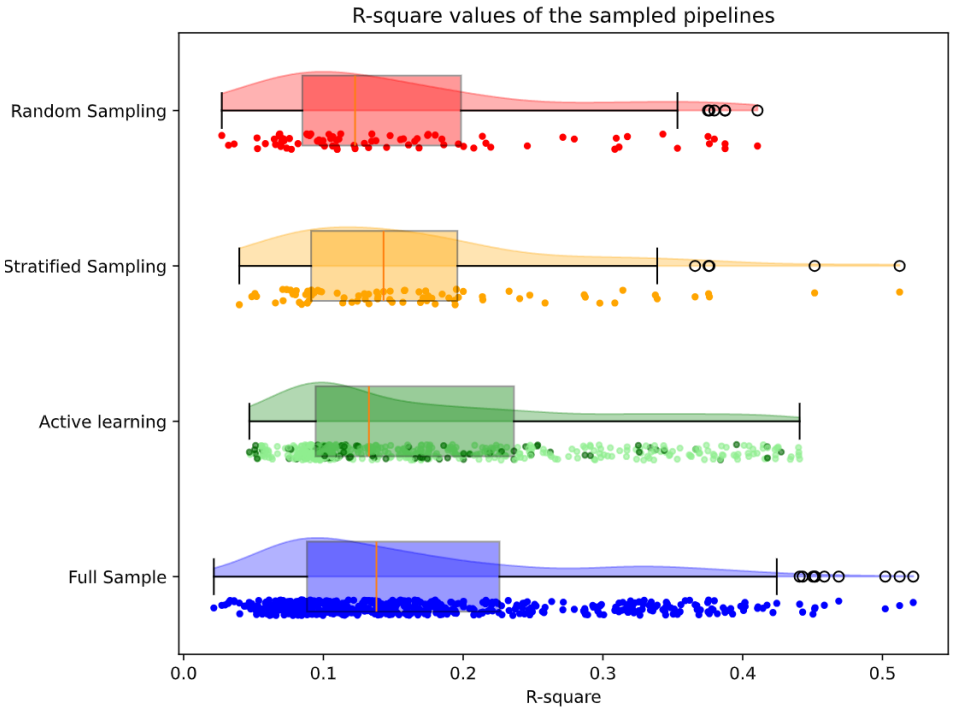


Sample Size = 79 pipelines (15% of full multiverse)

Test subset of participants (*N* = 50):



Lockbox subset of participants (*N* = 28):

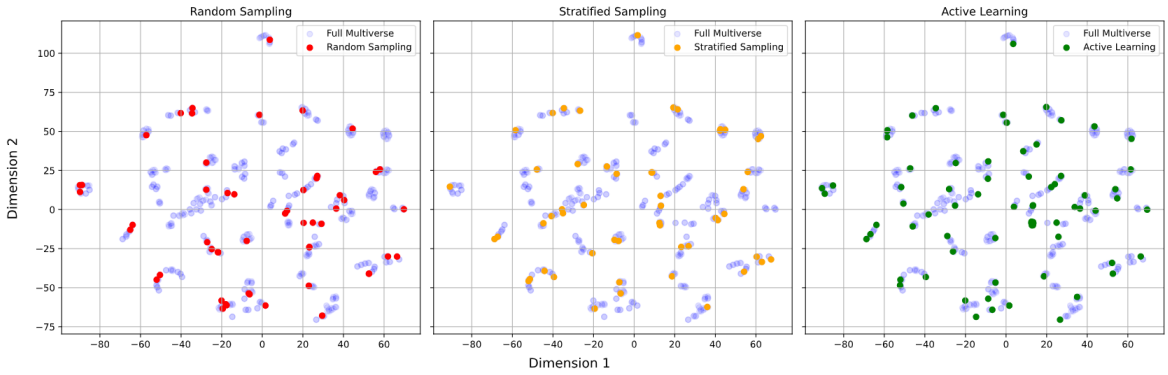


Output 4: Scatter Plots of Spatial Distribution in the Low Dimensional Space

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

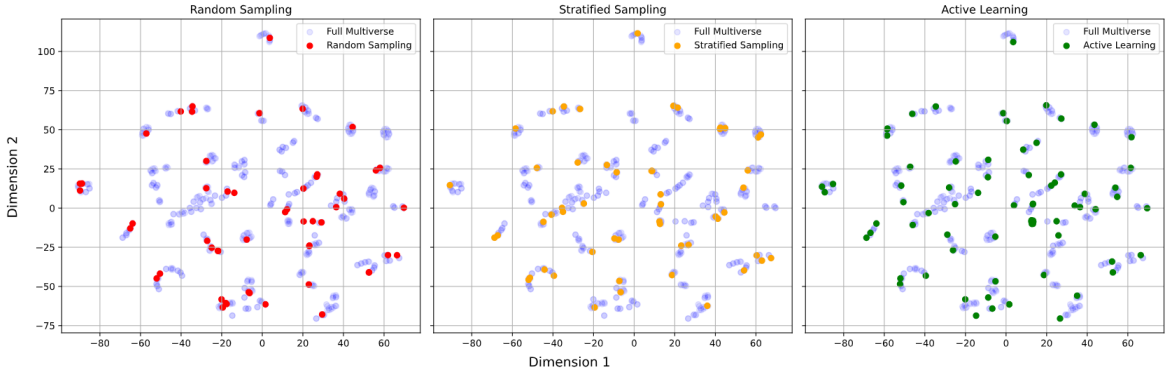
Test subset of participants (*N* = 50):

Sample Size	Full Multiverse vs. Random Sample	Full Multiverse vs. Stratified Sample	Full Multiverse vs. Active Learning Sample
53 pipelines	0.408	0.530	0.439



Lockbox subset of participants (*N* = 28):

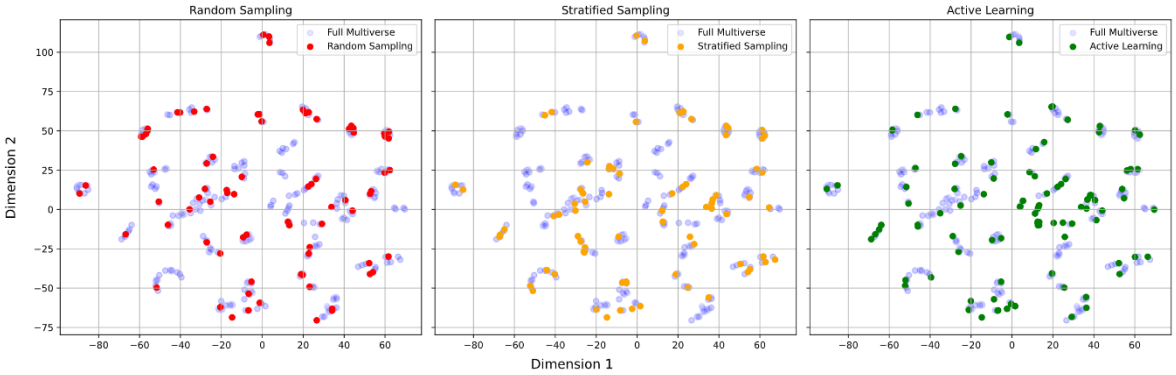
Sample Size	Full Multiverse vs. Random Sample	Full Multiverse vs. Stratified Sample	Full Multiverse vs. Active Learning Sample
53 pipelines	0.408	0.530	0.439



Sample Size = 79 pipelines (10% of full multiverse)

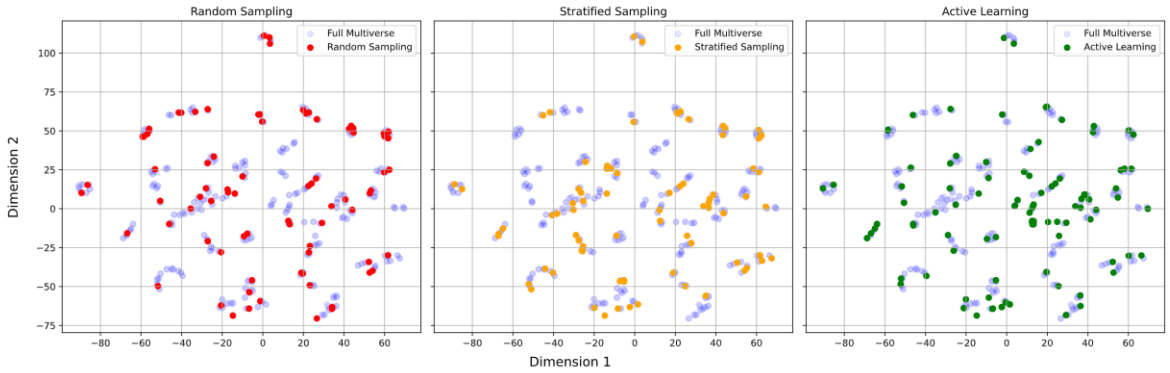
Test subset of participants (*N* = 50):

Sample Size	Full Multiverse vs. Random Sample	Full Multiverse vs. Stratified Sample	Full Multiverse vs. Active Learning Sample
79 pipelines	0.544	0.449	0.430



Lockbox subset of participants (*N* = 28):

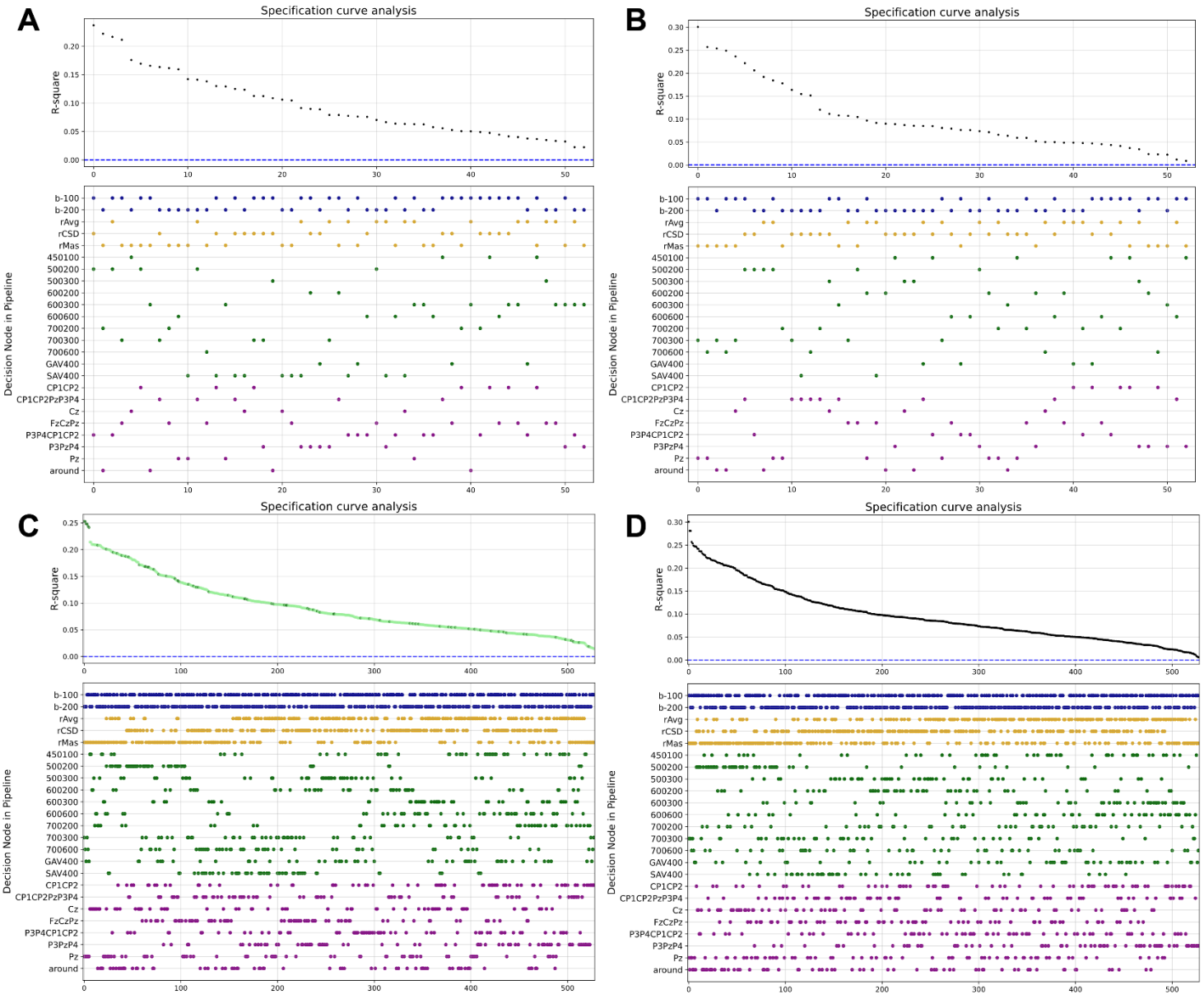
Sample Size	Full Multiverse vs. Random Sample	Full Multiverse vs. Stratified Sample	Full Multiverse vs. Active Learning Sample
79 pipelines	0.544	0.449	0.430



Output 5: Specification Curves

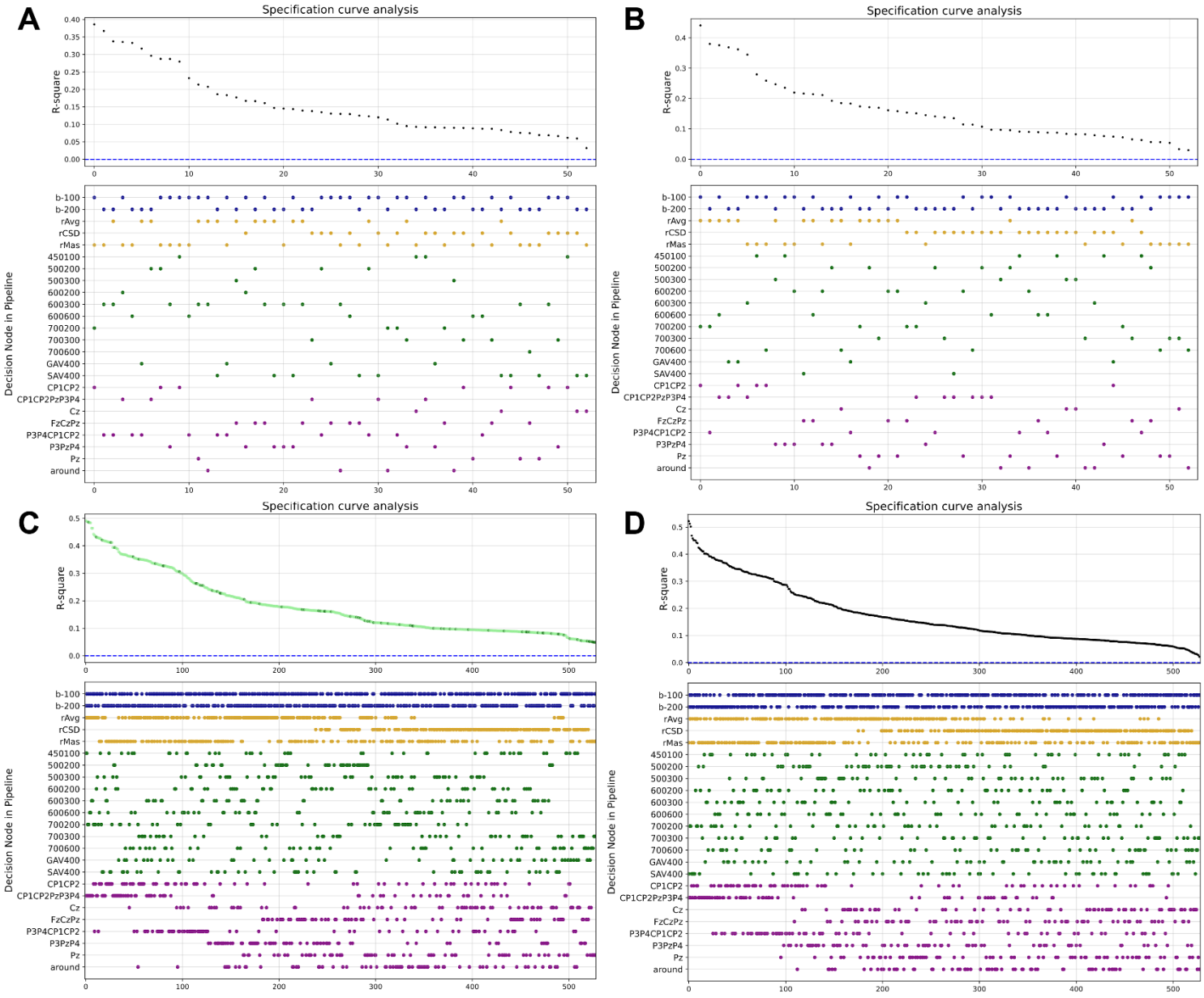
Sample Size = 53 pipelines (10% of full multiverse)

Test subset of participants (*N* = 50):



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.

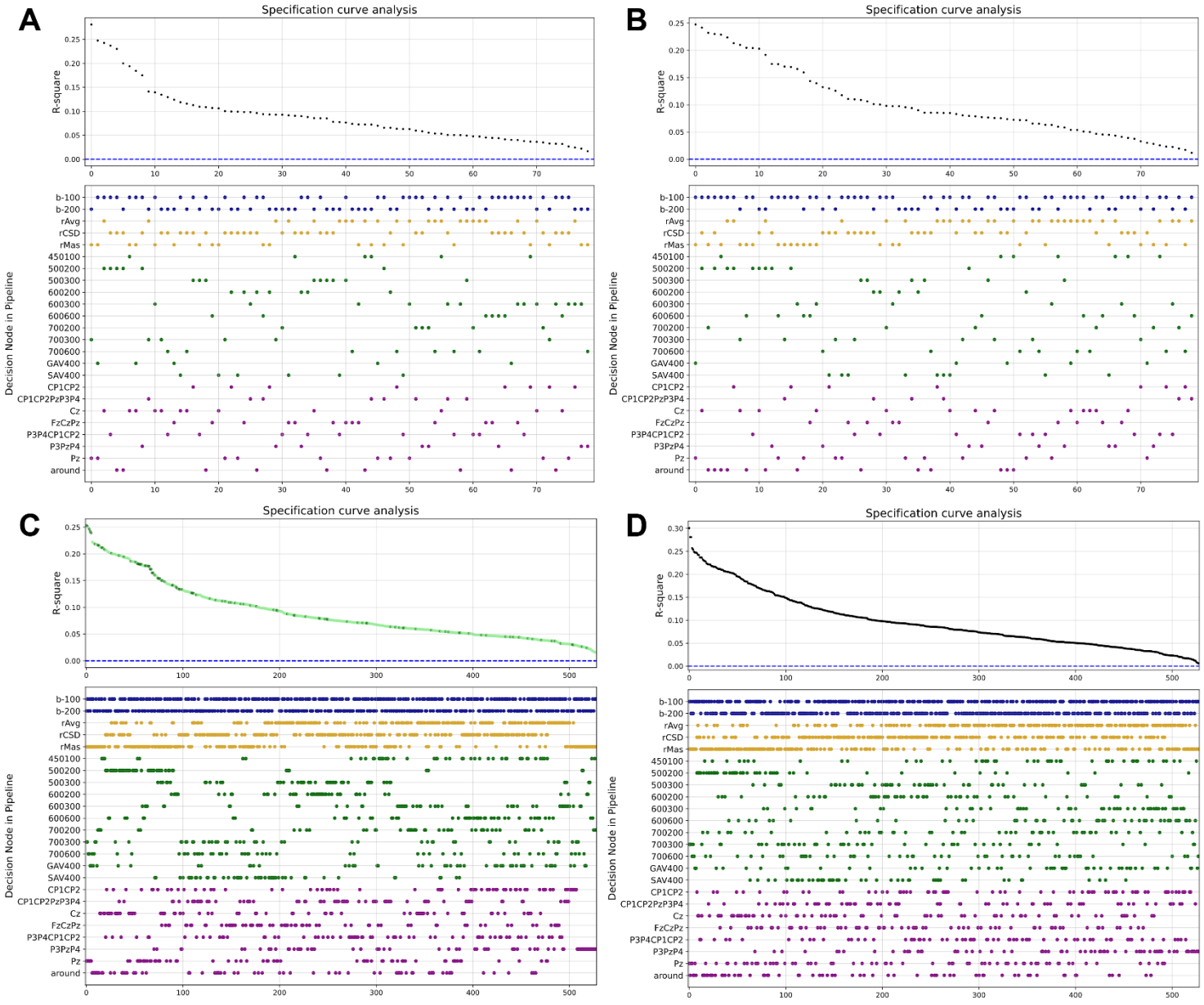
Lockbox subset of participants (*N* = 28):



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.

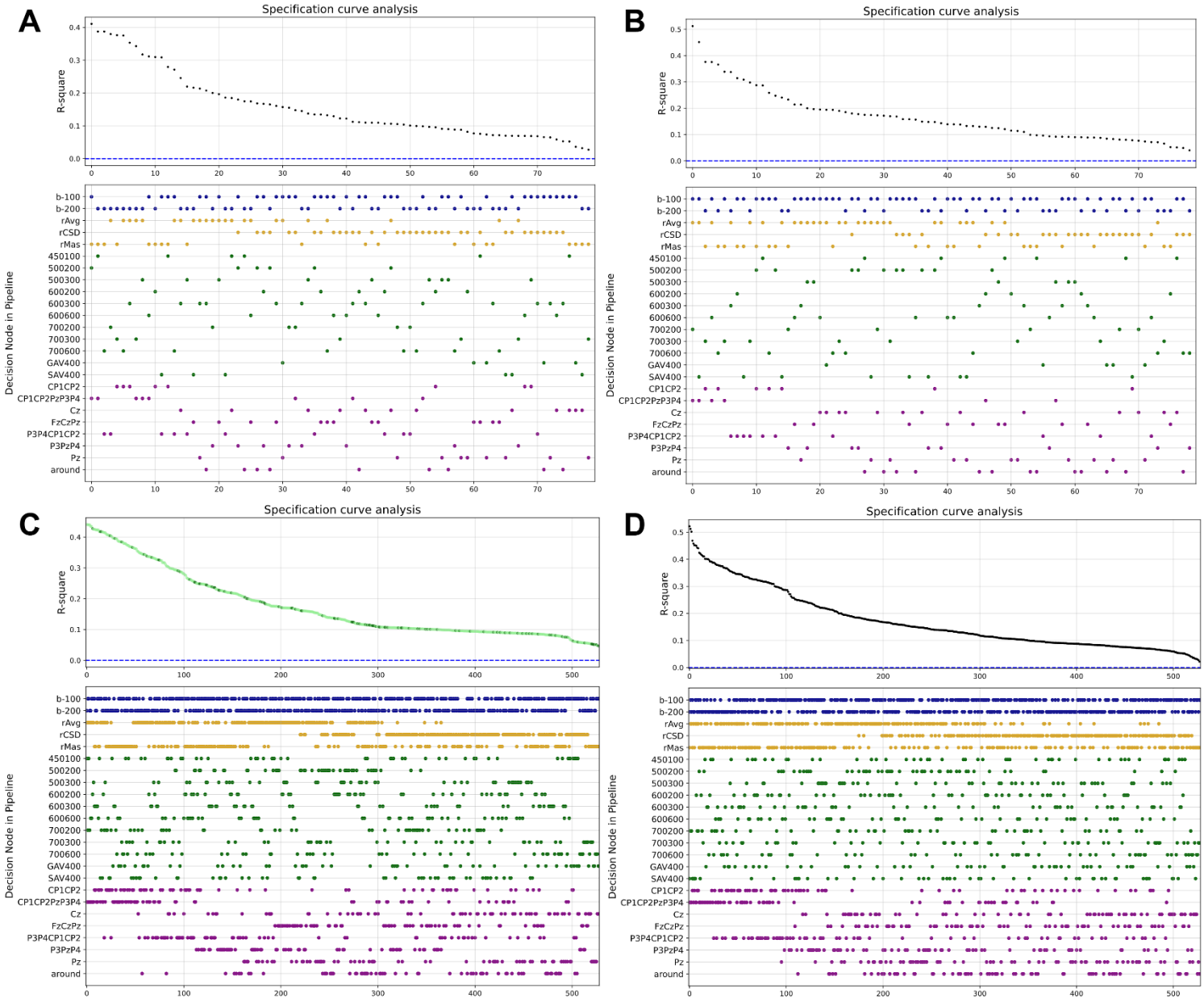
Sample Size = 79 pipelines (15% of full multiverse)

Test subset of participants (*N* = 50):



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.

Lockbox subset of participants (*N* = 28):



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.