Supplementary material for Hydrology and Earth System Sciences (HESS) manuscript submission: Generalization of Deep Learning Models to Ungauged Glacierized Basins: Evidence from Alpine, Patagonian, and North American Catchments

Meelisha Maharjan[1], Colin J. Gleason[1], Casey Brown[1]

[1]University of Massachusetts, Amherst, 01002 Massachusetts, USA

Correspondence to: Meelisha Maharjan (mmaharjan@umass.edu)

Contents of this file

Table S1

List of Large sample datasets and number of respective basins (Kratzert et al., 2023*)*

Dataset	A	M	G	Source
CAMELS-US	462	155	29	(Newman et al., 2015)
CAMELS-AUS	61	30	-	(Fowler et al., 2021)
CAMELS-BR	193	80	-	(Chagas et al., 2020)
CAMELS-CH (extension)	115	94	51	(Höge et al., 2023)
CAMELS-CL	7	7	6	(Alvarez-Garreton et al., 2018)
CAMELS-GB	274	19	-	(Coxon et al., 2020)
HYSETS	1401	420	118	(Arsenault et al., 2020)
LamaH-CE	332	215	79	(Klingler et al., 2021)
	2845	**1020**	**283**	

Table S2a

Meteorological Forcing Data and Model State Variable as Inputs for LSTM & Graph WaveNet (Muñoz-Sabater et al., 2021)

Meteorological Feature	Aggregation	Unit
Precipitation	Daily sum	*mm/day*
Air Temperature	Daily min, Daily max	*℃*
Surface Pressure	Daily mean	*KPa*
Surface net solar radiation	Daily mean	*Wm-2*
Model State Feature	**Aggretation**	**Unit**
Snow water equivalent	Daily mean	*mm/day*

Table S2b
Meteorological Forcing Data for δHBV (Muñoz-Sabater et al., 2021)

Feature	Aggregation	Unit
Precipitation	Daily sum	*mm/day*
Air Temperature	Daily mean	*℃*
Potential evaporation	Daily sum	*mm/day*

Table S3
Static attributes list (Anon, 2019)

Group	Description	Aggregation	Unit
Physiography	Elevation	Spatial mean/min/max	*m above sea level*
	Terrain slope	Spatial mean	*° (x 10)*
	Stream gradient	Mean of reach segments	*dm/km*
Metadata	Area		*km2*
	Gauge latitude		*deg*
	Gauge longitude		*deg*
LandCover	Forest cover extent	Spatial mean	*% cover*
	Cropland extent	Spatial mean	*% cover*
	Wetland extent	Spatial mean	*% cover*
	Permafrost extent	Spatial mean	*% cover*
Soils and Geology	Sand fraction in soil	Spatial mean	*%*
	Silt fraction in soil	Spatial mean	*%*
	Clay fraction in soil	Spatial mean	*%*
	Organic Carbon content in soil	Spatial mean	*tonnes/hectare*
	Soil water content	Annual mean	*%*
	Karst area extent	Spatial mean	*% cover*

Climate attributes	Aridity index	Ratio of mean PET and mean precipitation	-
	Fraction of precipitation falling as snow	-	-
	Mean annual moisture index in range [-1,1]	-	-
	Snow cover percent	Annual max, Annual mean	*% cover*

Table S4

Overview of RGI Classification, Country and Mountain ranges of the Glacial High Mountain Catchments (G)

First Order RGI Region	Second Order RGI Region	Country	Mountain Range
02 – Western Canada and USA	02-2 (S Coast Ranges)	Canada, USA	S Coast Ranges
	02-3 (N Rocky Mts)		N Rocky Mts
	02-4 (Cascade Ra and Sa Nevada)		Cascade Ra and Sa Nevada
	02-5 (S Rocky Mts)		S Rocky Mts
11 – Central Europe	11-1 (Alps)	Austria, Germany, Switzerland	Alps
17 – Southern Andes	17-2 (C Andes)	Chile	C Andes

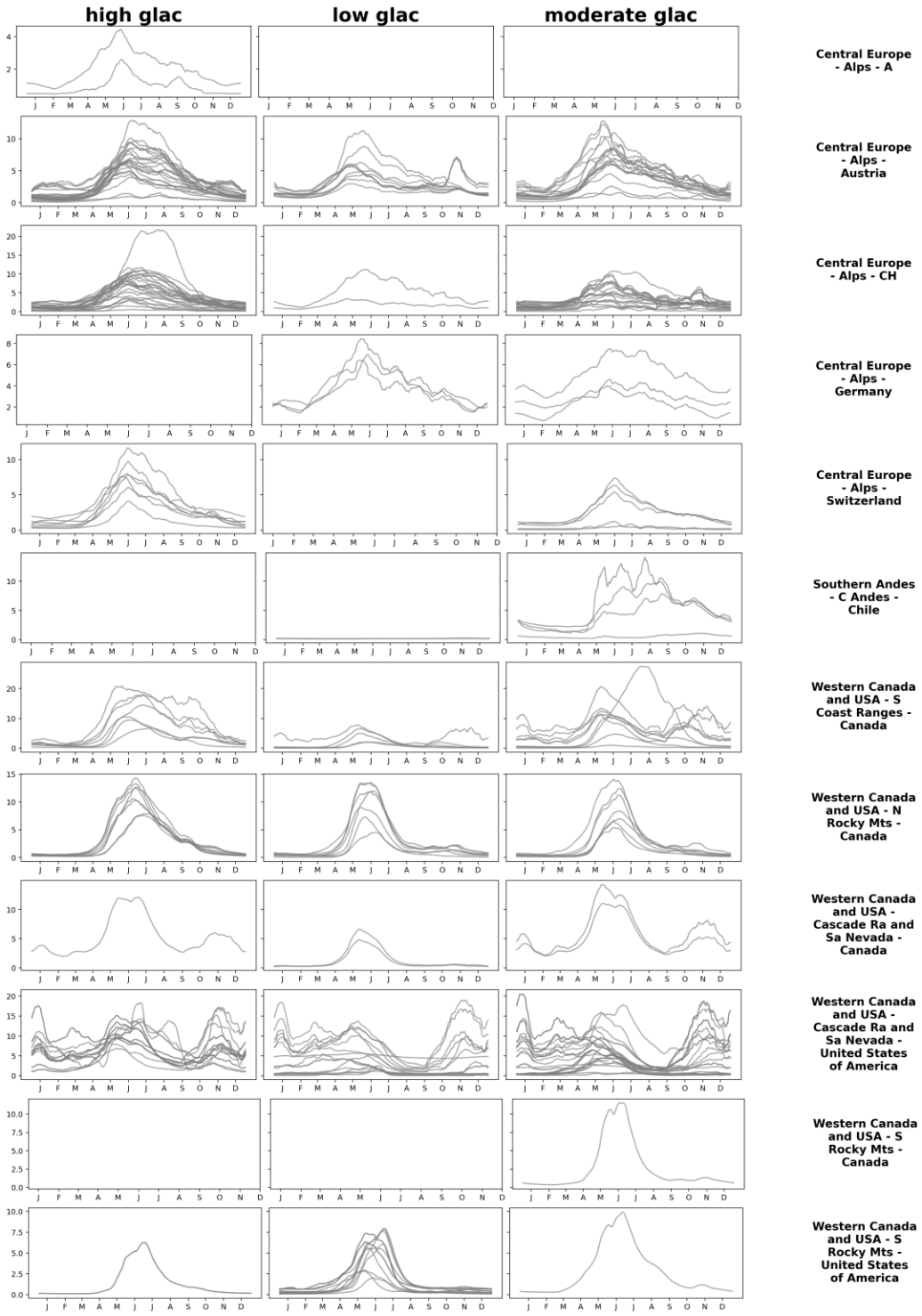

Figure S1. 20-day average of annual mean streamflow (mm/day) for individual G basins, grouped into four glacier-coverage classes—low (<2%), moderate (2–20%), high (20–40%), and very high (>40%)—and categorized by RGI first- and second-order regions (rows: Central Europe – various Alpine subregions; Southern Andes – C Andes – Chile; Western Canada & USA – multiple mountain ranges). Each grey line represents one basin's January–December hydrograph.

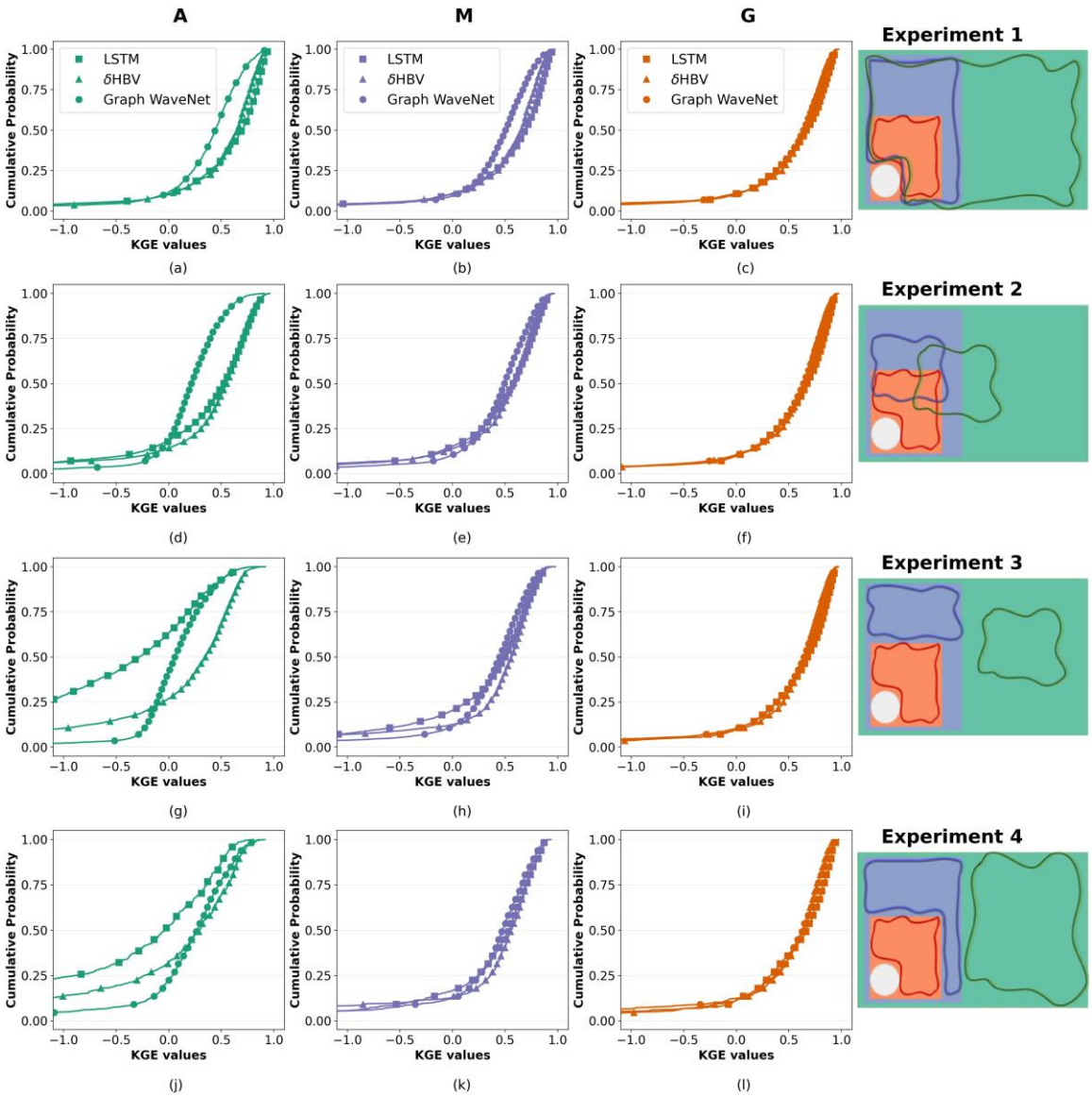

Figure S2. CDF comparison of KGE values of 3 DL model architectures trained with different partitions. Each column represents 3 different data partitions – A, M, and G respectively. Rows 1-4 showing performance for experiments 1-4 respectively. Each line of the CDF plot comprises 5600 sample points from each of the 100 trials with 56 test basins for experiments 1-3, while the lines in experiment 4 consist of 560 sample points from 10 trials with 56 test points each. *Graph WaveNet in experiment 1 for A partition has only ~2300 samples due to excessive resource needs required to train the model.

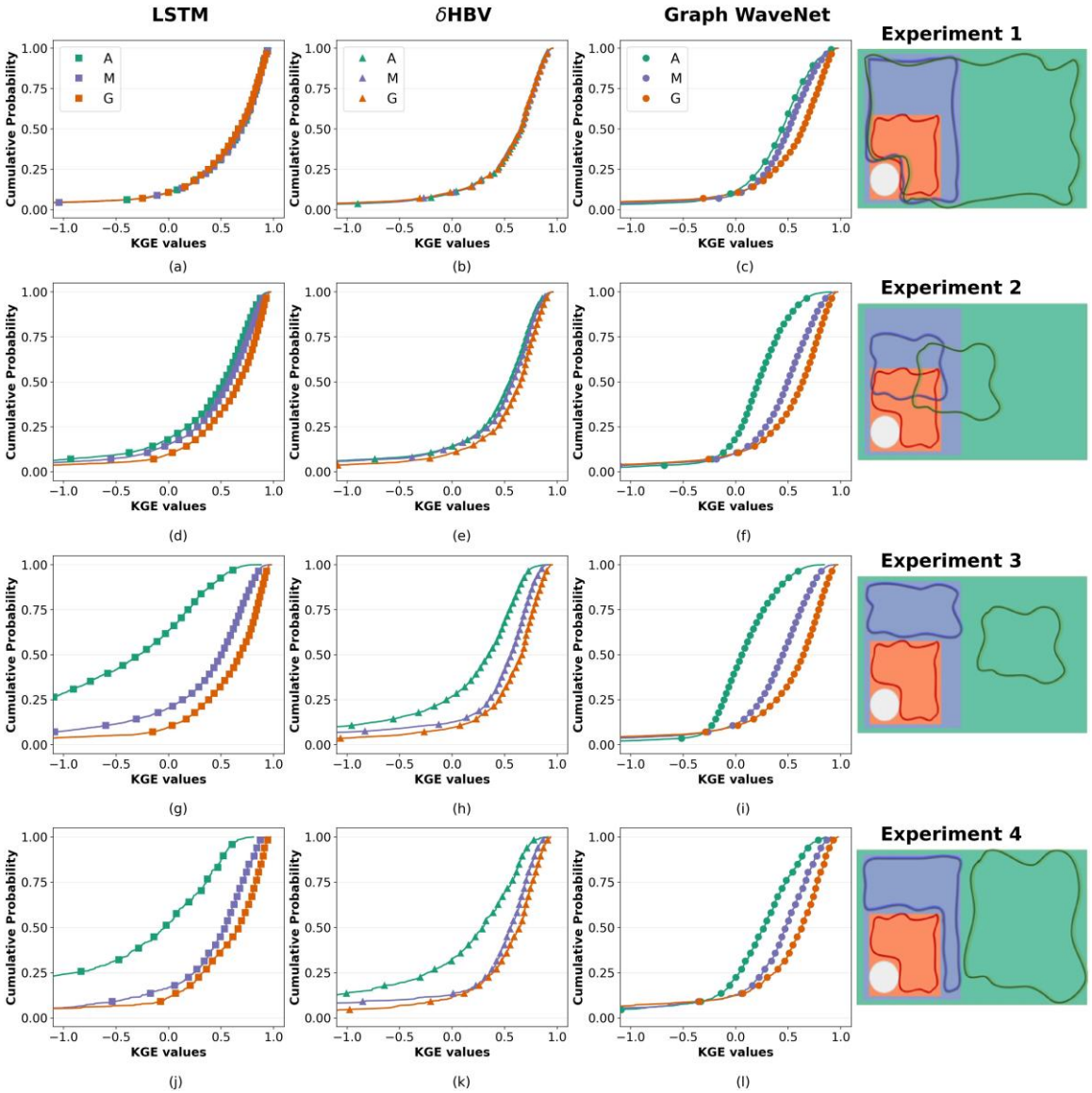

Figure S3. CDF comparison of KGE values of LSTM, δHBV and Graph WaveNet when trained with 3 different data partitions. Rows 1-4 showing performance for experiments 1-4 respectively. Each line of the CDF plot comprises 5600 sample points from each of the 100 trials with 56 test basins for experiments 1-3, while the lines in experiment 4 consist of 560 sample points from 10 trials with 56 test points each. * *Graph WaveNet in experiment 1 for A partition has only ~2300 samples due to excessive resource needs required to train the model.*

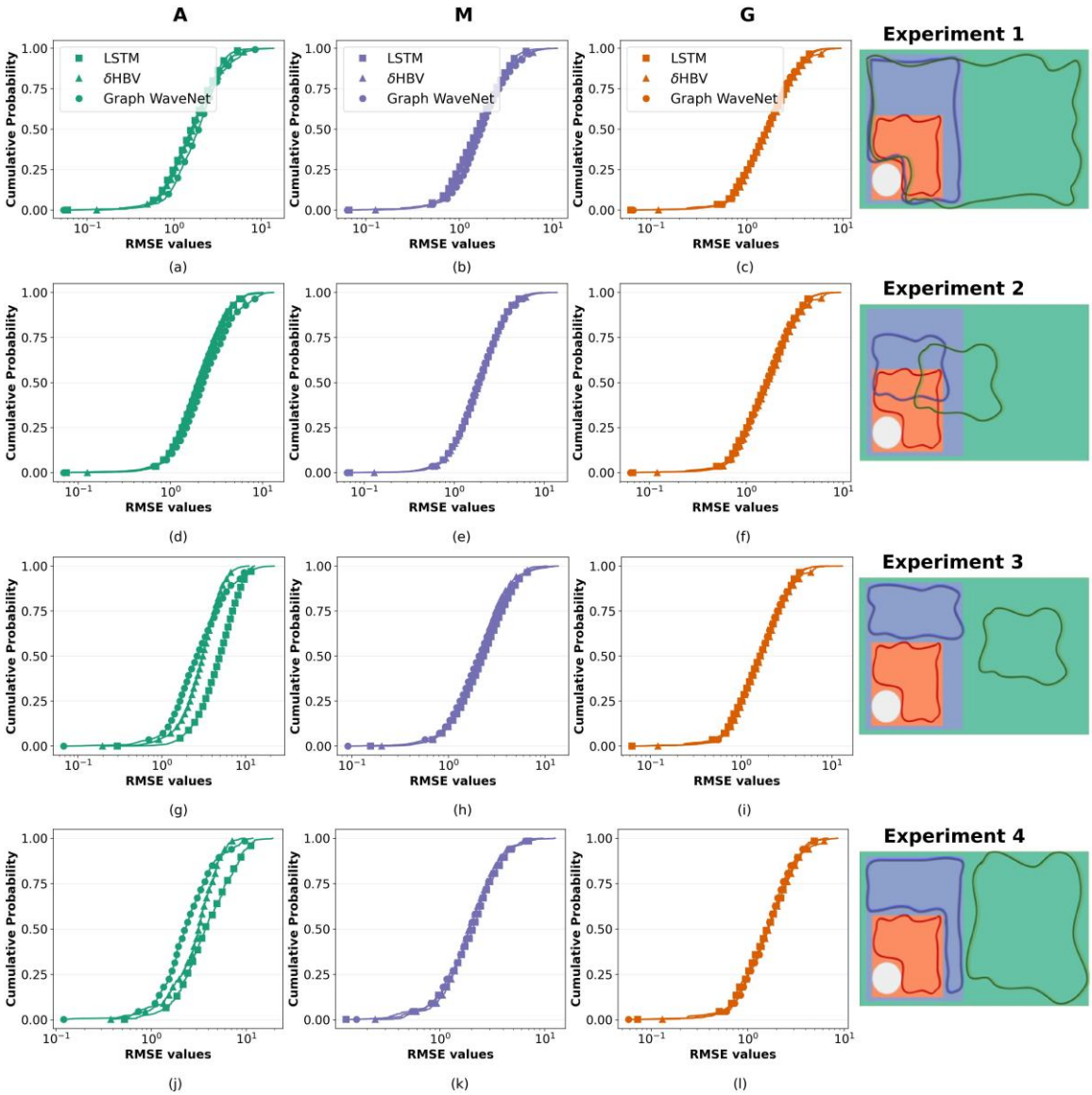

Figure S4. CDF comparison of RMSE values of 3 DL model architectures trained with different partitions. Each column represents 3 different data partitions – A, M, and G respectively. Rows 1-4 showing performance for experiments 1-4 respectively. Each line of the CDF plot comprises 5600 sample points from each of the 100 trials with 56 test basins for experiments 1-3, while the lines in experiment 4 consist of 560 sample points from 10 trials with 56 test points each. *Graph WaveNet in experiment 1 for A partition has only ~2300 samples due to excessive resource needs required to train the model.

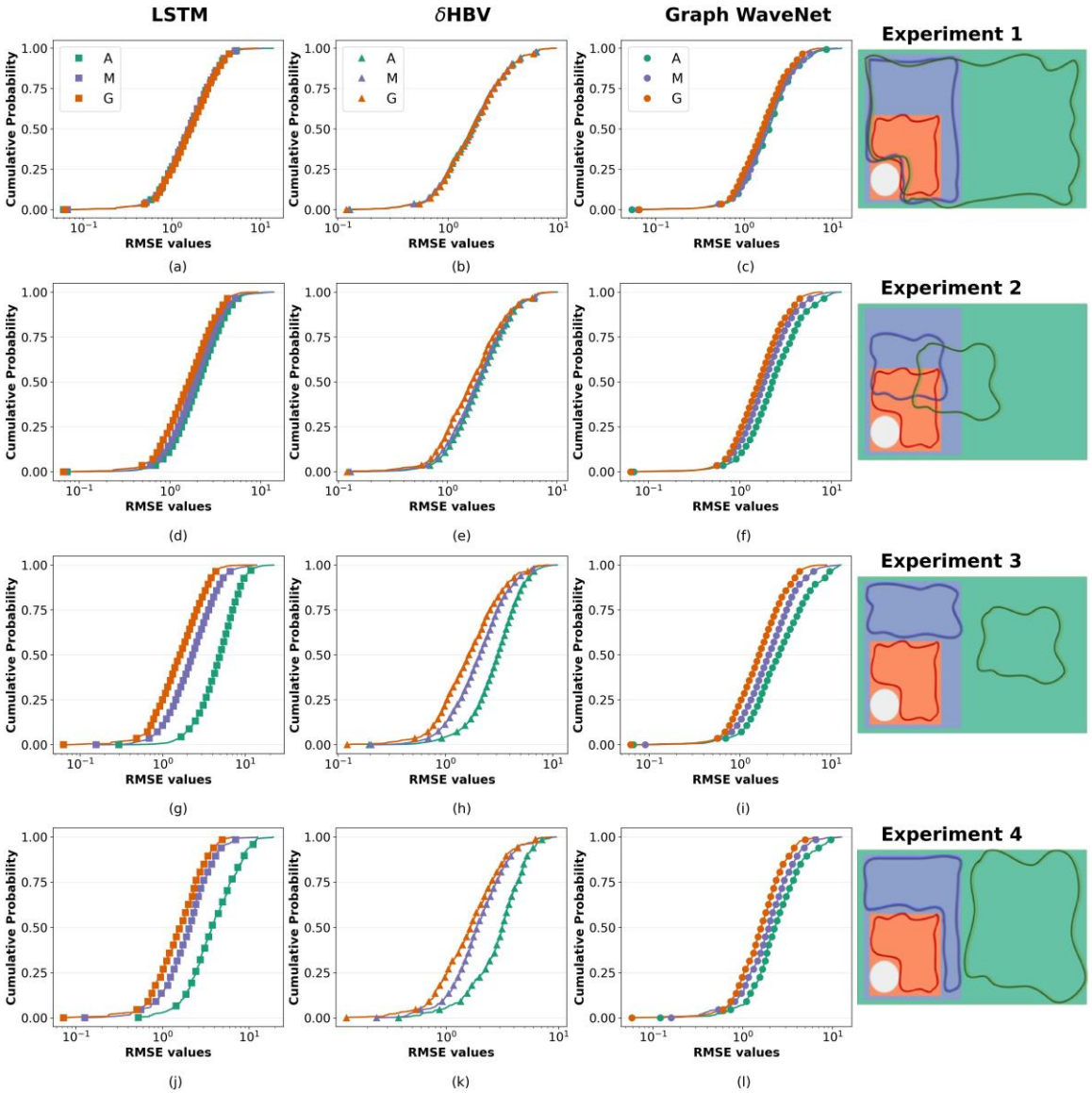

Figure S5. CDF comparison of RMSE values of LSTM, δHBV and Graph WaveNet when trained with 3 different data partitions. Rows 1-4 showing performance for experiments 1-4 respectively. Each line of the CDF plot comprises 5600 sample points from each of the 100 trials with 56 test basins for experiments 1-3, while the lines in experiment 4 consist of 560 sample points from 10 trials with 56 test points each. * Graph WaveNet in experiment 1 for A partition has only ~2300 samples due to excessive resource needs required to train the model.

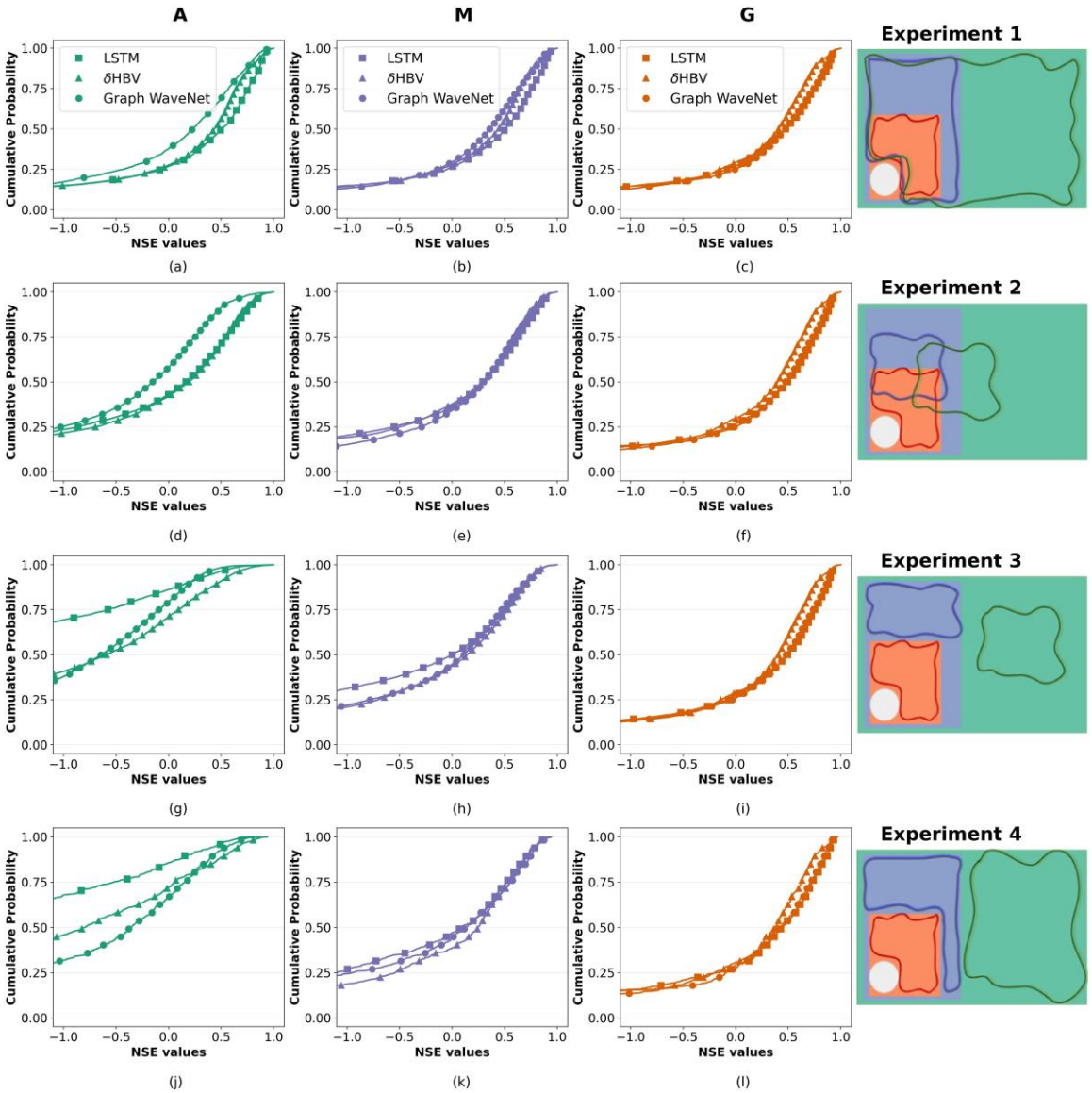

Figure S6. CDF comparison of NSE values of 3 DL model architectures trained with different partitions over the summer months of the test period. Each column represents 3 different data partitions – A, M, and G respectively. Rows 1-4 showing performance for experiments 1-4 respectively. Each line of the CDF plot comprises 5600 sample points from each of the 100 trials with 56 test basins for experiments 1-3, while the lines in experiment 4 consist of 560 sample points from 10 trials with 56 test points each. *Graph WaveNet in experiment 1 for A partition has only ~2300 samples due to excessive resource needs required to train the model.

Table S5. Median KGE Values

	All Global (A)			Mixed High Mountain (M)			High Mountain with Glacier (G)		
	LSTM	δHBV	Graph WaveNet	LSTM	δHBV	Graph WaveNet	LSTM	δHBV	Graph WaveNet
Exp 1	**0.70**	0.65	0.44	**0.69**	0.65	0.50	**0.66**	**0.66**	0.64
Exp 2	0.52	**0.55**	0.21	**0.58**	**0.58**	0.49	**0.67**	0.65	0.64
Exp 3	0.28	**0.35**	0.06	0.50	**0.56**	0.45	**0.68**	0.66	0.64
Exp 4	0.04	**0.29**	0.27	0.54	**0.56**	0.47	**0.67**	0.64	0.63

Table S6. Median RMSE Values

	All Global (A)			Mixed High Mountain (M)			High Mountain with Glacier (G)		
	LSTM	δHBV	Graph WaveNet	LSTM	δHBV	Graph WaveNet	LSTM	δHBV	Graph WaveNet
Exp 1	**1.57**	1.64	1.91	**1.57**	1.66	1.74	1.64	1.66	**1.61**
Exp 2	2.07	**1.97**	2.26	1.90	1.88	**1.82**	1.63	1.66	**1.61**
Exp 3	4.95	3.10	**2.64**	2.31	2.09	**2.07**	**1.62**	1.66	**1.62**
Exp 4	3.72	3.15	**2.26**	2.06	**1.87**	1.94	**1.59**	1.65	1.60

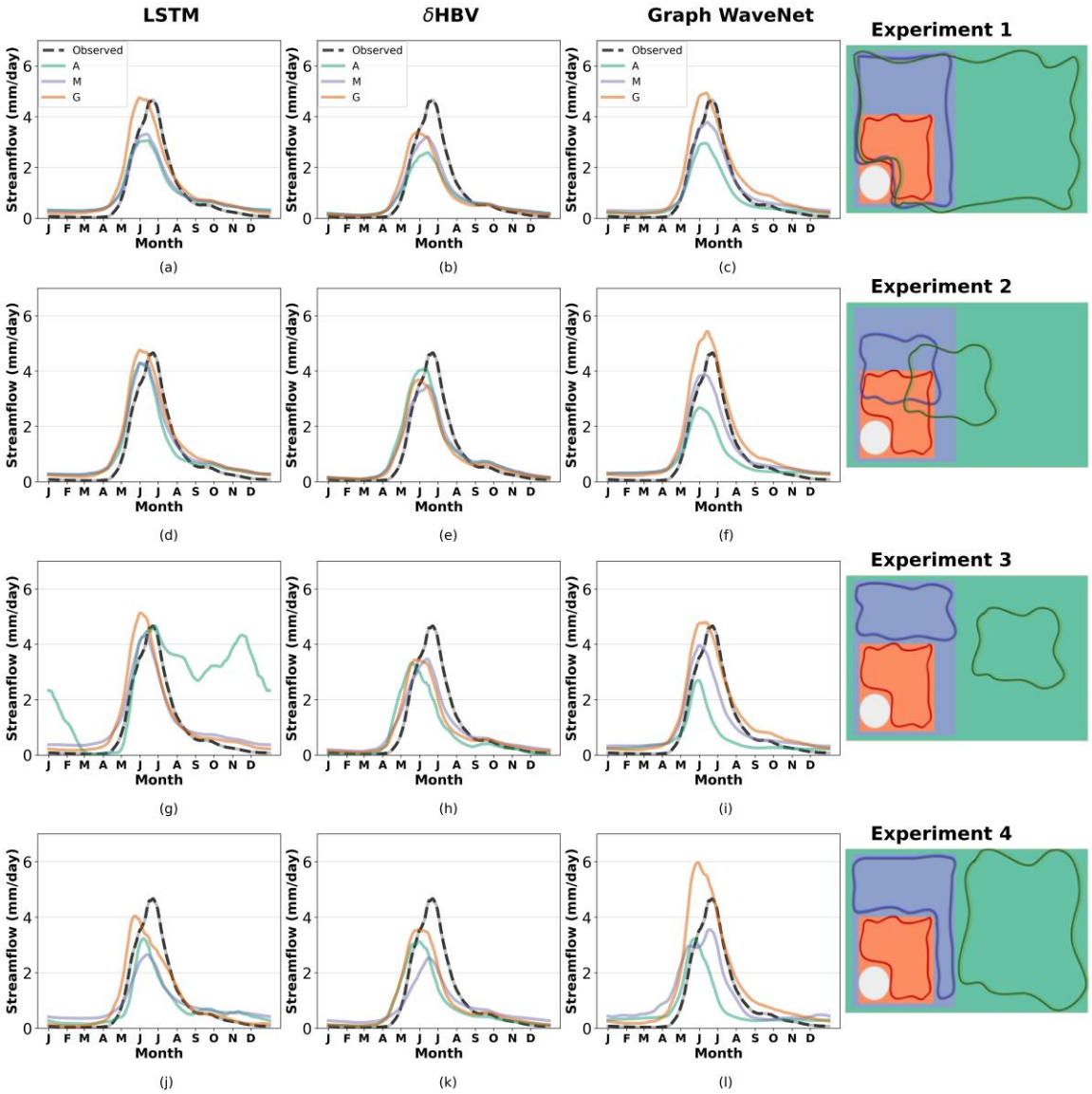

Figure S7. Observed vs. Simulated hydrograph for a low category of glacier (<2%) basin [hysets_08NF001].

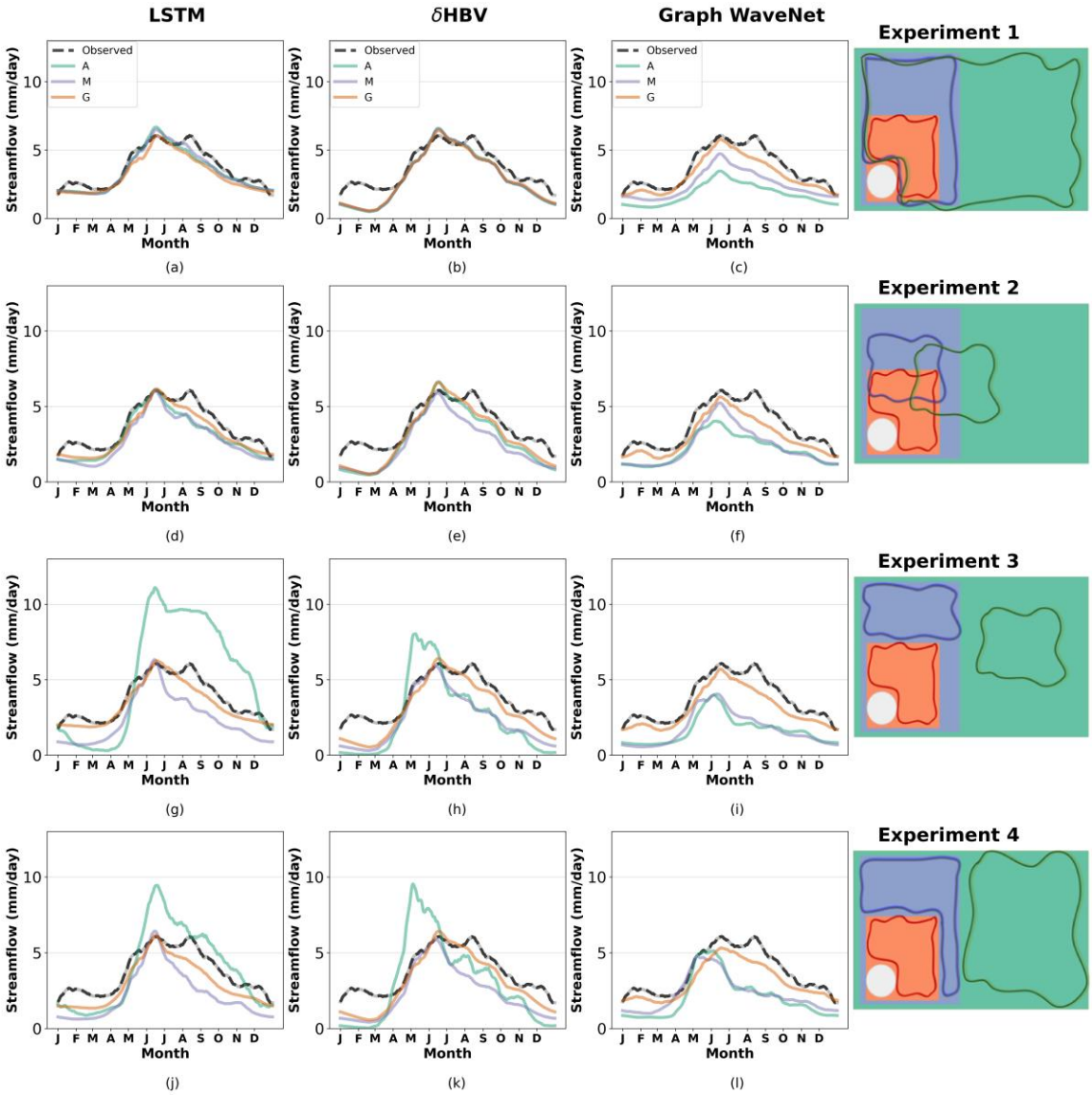

Figure S8. Observed vs. Simulated hydrograph for a moderate category of glacier (2-20%) basin [lamah_201780]

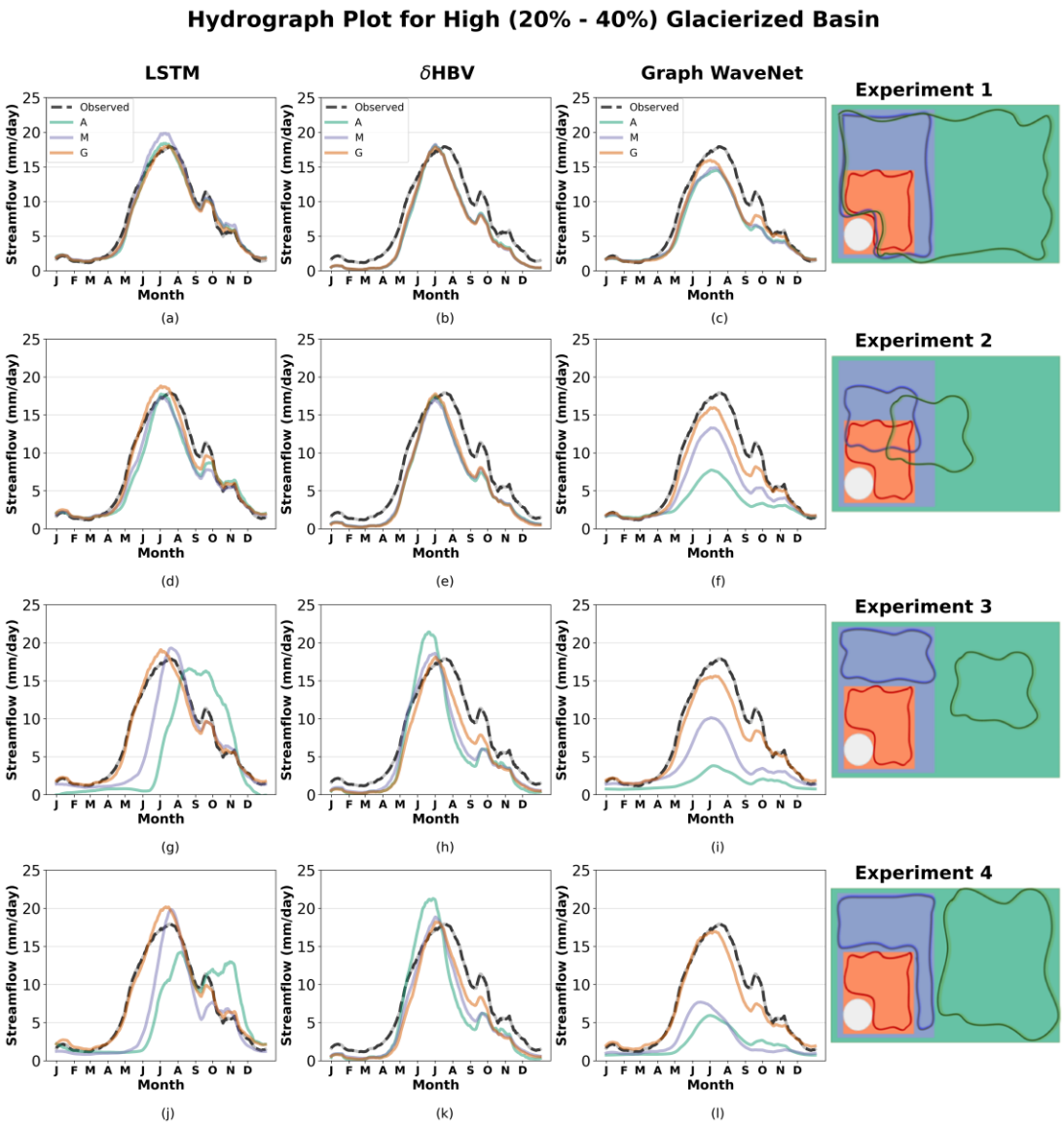

Figure S9. Observed vs. Simulated hydrograph for a high category of glacier (20-40%) basin [hysets_08GA071]

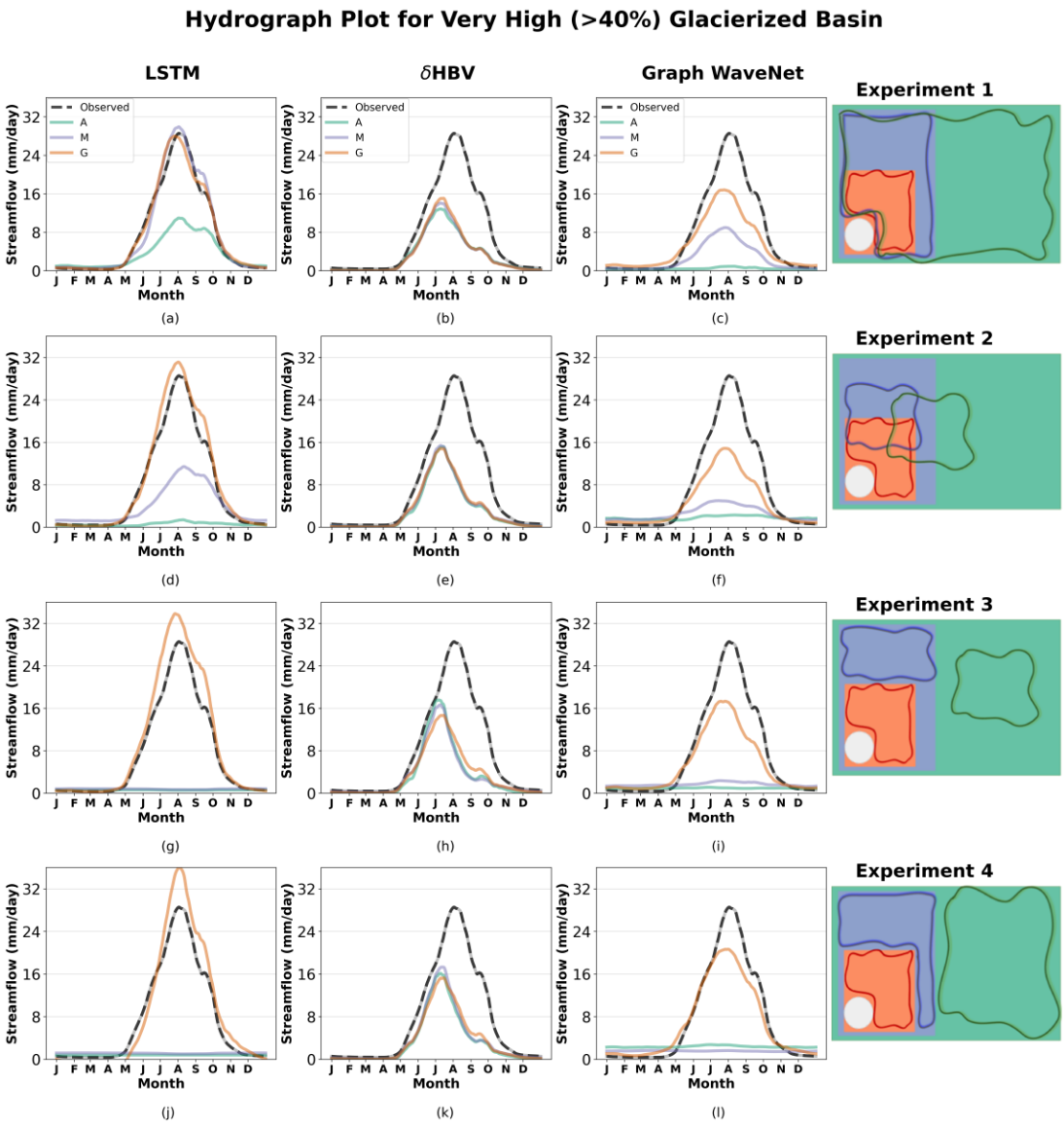

Figure S10. Observed vs. Simulated hydrograph for a very high category of glacier (>40%) basin [hysets_08ME023]

References

Alvarez-Garreton, C., Mendoza, P. A., Boisier, J. P., Addor, N., Galleguillos, M., Zambrano-Bigiarini, M., Lara, A., Puelma, C., Cortes, G., Garreaud, R., McPhee, J., and Ayala, A.: The CAMELS-CL dataset: catchment attributes and meteorology for large sample studies – Chile dataset, Hydrology and Earth System Sciences, 22, 5817–5846, https://doi.org/10.5194/hess-22-5817-2018, 2018.

Anon: HydroATLAS version 1.0, https://doi.org/10.6084/m9.figshare.9890531.v1, 2019.

Arsenault, R., Brissette, F., Martel, J.-L., Troin, M., Lévesque, G., Davidson-Chaput, J., Gonzalez, M. C., Ameli, A., and Poulin, A.: A comprehensive, multisource database for hydrometeorological modeling of 14,425 North American watersheds, Sci Data, 7, 243, https://doi.org/10.1038/s41597-020-00583-2, 2020.

Chagas, V. B. P., Chaffe, P. L. B., Addor, N., Fan, F. M., Fleischmann, A. S., Paiva, R. C. D., and Siqueira, V. A.: CAMELS-BR: hydrometeorological time series and landscape attributes for 897 catchments in Brazil, Earth System Science Data, 12, 2075–2096, https://doi.org/10.5194/essd-12-2075-2020, 2020.

Coxon, G., Addor, N., Bloomfield, J. P., Freer, J., Fry, M., Hannaford, J., Howden, N. J. K., Lane, R., Lewis, M., Robinson, E. L., Wagener, T., and Woods, R.: CAMELS-GB: hydrometeorological time series and landscape attributes for 671 catchments in Great Britain, Earth System Science Data, 12, 2459–2483, https://doi.org/10.5194/essd-12-2459-2020, 2020.

Fowler, K. J. A., Acharya, S. C., Addor, N., Chou, C., and Peel, M. C.: CAMELS-AUS: hydrometeorological time series and landscape attributes for 222 catchments in Australia, Earth System Science Data, 13, 3847–3867, https://doi.org/10.5194/essd-13-3847-2021, 2021.

Höge, M., Kauzlaric, M., Siber, R., Schönenberger, U., Horton, P., Schwanbeck, J., Floriancic, M. G., Viviroli, D., Wilhelm, S., Sikorska-Senoner, A. E., Addor, N., Brunner, M., Pool, S., Zappa, M., and Fenicia, F.: CAMELS-CH: hydro-meteorological time series and landscape attributes for 331 catchments in hydrologic Switzerland, Earth System Science Data Discussions, 1–46, https://doi.org/10.5194/essd-2023-127, 2023.

Klingler, C., Schulz, K., and Herrnegger, M.: LamaH-CE: LArge-SaMple DAta for Hydrology and Environmental Sciences for Central Europe, Earth System Science Data, 13, 4529–4565, https://doi.org/10.5194/essd-13-4529-2021, 2021.

Kratzert, F., Nearing, G., Addor, N., Erickson, T., Gauch, M., Gilon, O., Gudmundsson, L., Hassidim, A., Klotz, D., Nevo, S., Shalev, G., and Matias, Y.: Caravan - A global community dataset for large-sample hydrology, Sci Data, 10, 61, https://doi.org/10.1038/s41597-023-01975-w, 2023.

Muñoz-Sabater, J., Dutra, E., Agustí-Panareda, A., Albergel, C., Arduini, G., Balsamo, G., Boussetta, S., Choulga, M., Harrigan, S., Hersbach, H., Martens, B., Miralles, D. G.,

Piles, M., Rodríguez-Fernández, N. J., Zsoter, E., Buontempo, C., and Thépaut, J.-N.: ERA5-Land: a state-of-the-art global reanalysis dataset for land applications, Earth System Science Data, 13, 4349–4383, https://doi.org/10.5194/essd-13-4349-2021, 2021.

Newman, A. J., Clark, M. P., Sampson, K., Wood, A., Hay, L. E., Bock, A., Viger, R. J., Blodgett, D., Brekke, L., Arnold, J. R., Hopson, T., and Duan, Q.: Development of a large-sample watershed-scale hydrometeorological data set for the contiguous USA: data set characteristics and assessment of regional variability in hydrologic model performance, Hydrology and Earth System Sciences, 19, 209–223, https://doi.org/10.5194/hess-19-209-2015, 2015.