Table 1: Predictive accuracy table, new dataset

Duration	Name	Evaluation metric				
		RMSE [l/s/km2]	CRPS [l/s/km2]	MAE [l/s/km2]	MRE [%]	MAPE [%]
0 hour	floodGAM	**149.7**	**69.7***	**98.8***	**21.3***	**21.3***
	RFFA_2018	158.1	80.3	110.4	29.5	26.5
	XGBoost	-	-	98.8	-	-
1 hour	floodGAM	**150.7**	**69.5***	**98.6***	**21.3***	**21.2***
	RFFA_2018	156.9	79.2	109.3	29.0	26.2
	XGBoost	-	-	95.0	-	-
6 hour	floodGAM	**132.7**	**63.1***	**90.3**	**20.2***	**20.4***
	RFFA_2018	136.4	69.4	97.3	25.4	24.0
	XGBoost	-	-	88.1	-	-
12 hour	floodGAM	**109.6**	**54.6***	**77.5**	**18.9***	**19.5***
	RFFA_2018	113.2	58.5	82.6	21.9	21.2
	XGBoost	-	-	76.9	-	-
18 hour	floodGAM	**100.0**	**49.8**	**70.4**	**18.3**	**18.7**
	RFFA_2018	100.4	51.7	72.9	20.0	19.5
	XGBoost	-	-	69.2	-	-
24 hours	floodGAM	89.8	**45.8**	**65.0**	**17.7**	18.2
	RFFA_2018	**89.4**	46.5	65.3	18.5	18.2
	XGBoost	-	-	67.7	-	-
36 hour	floodGAM	74.5	39.4	55.4	**16.7**	17.3
	RFFA_2018	**72.8**	**38.7**	**54.2**	16.8	**16.3**
	XGBoost	-	-	52.5	-	-
48 hour	floodGAM	65.0	34.9	48.7	16.0	16.4
	RFFA_2018	**63.5**	**34.2**	**48.4**	**15.9**	**15.6**
	XGBoost	-	-	47.2	-	-
72 hour	floodGAM	54.4	29.9	41.6	15.5	15.6
	RFFA_2018	**52.7**	**29.0**	**41.0**	**14.9**	**15.1**
	XGBoost	-	-	39.9	-	-
7 days	floodGAM	42.9	23.4	32.4	15.1	**15.3**
	RFFA_2018	**41.4**	**23.2**	**31.9**	**14.6**	15.4
	XGBoost	-	-	33.3	-	-
14 days	floodGAM	38.4	**19.5**	**26.7**	**14.8**	**14.9***
	RFFA_2018	**37.1**	20.0	26.9	15.3	16.0
	XGBoost	-	-	25.8	-	-
30 days	floodGAM	33.7	**16.0**	**21.9**	**14.8***	**14.6***
	RFFA_2018	**32.7**	17.0	23.1	16.3	17.2
	XGBoost	-	-	20.7	-	-

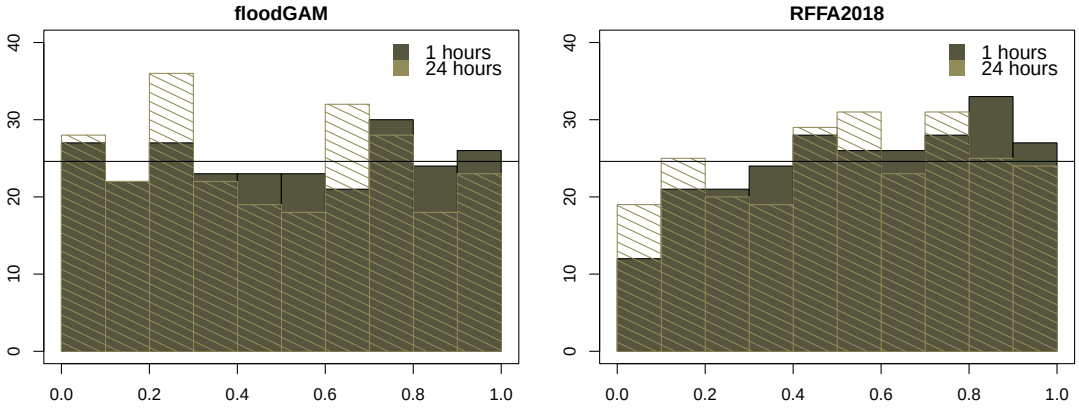

Figure 1: PIT histograms, with new dataset

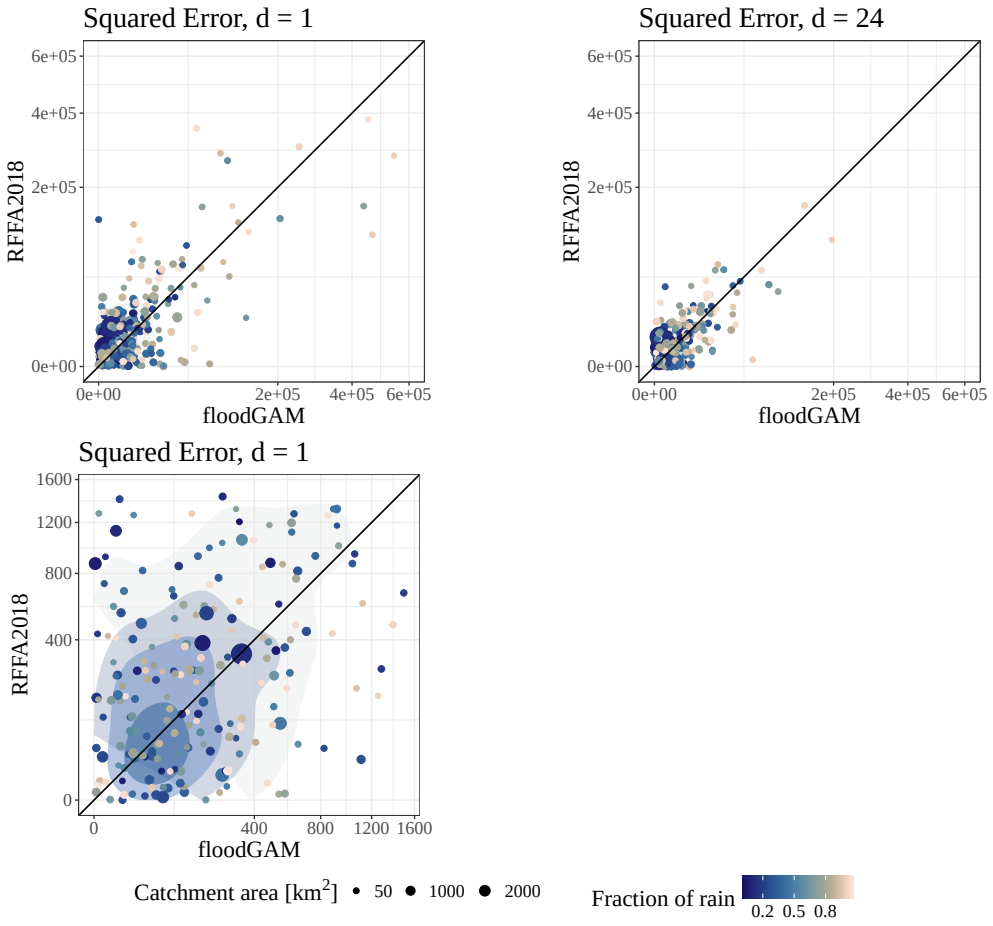

Figure 2: squared error dotplots, new dataset

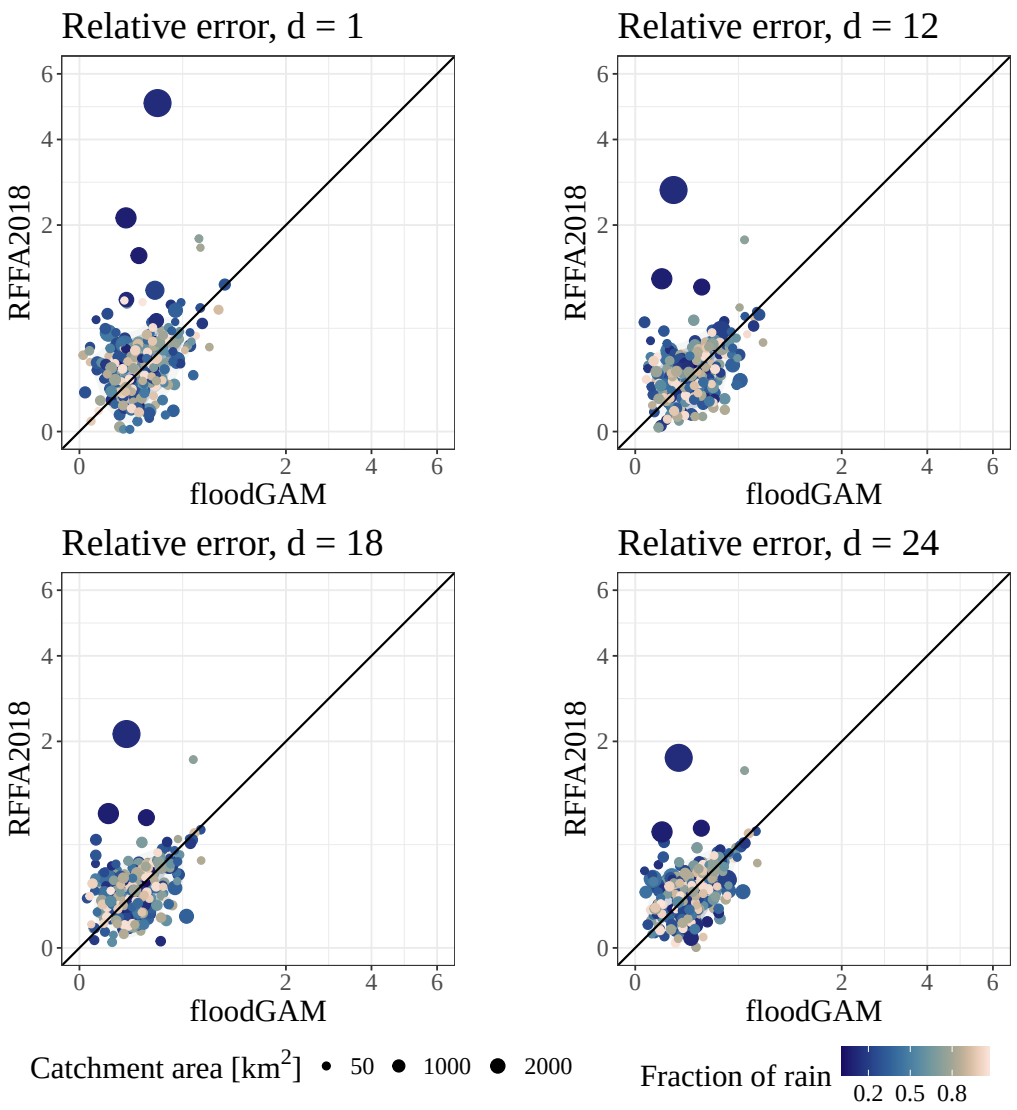

Figure 3: We still see underestimation from RFFA2018 at large, snowmelt driven catchments