Notes from Sankar Meeting

Last Modified: 08/11/2021 08:34 AM

This weeks TODO:

- For the 6 bad plots in the `thirds_1to1` plot:
  - Lower third, train and test, upstream and downstream. Middle third, train and test, upstream
  - Check if the poor performance is due to bias or variance
  - Can look at correlation between modeled and actual values and bias as an alternative to the above
  - This should be done at the reservoir level, not the aggregate level
    * (i.e.) with in each third, look at bias/variance/correlation for all the samples for each reservoir in that third
    * Train Bins : `(0, 2.58], (2.58, 16.78], (16.78, 844.6]`
    * Test Bins : `(0, 1.91], (1.91, 18.8], (18.8, 870.6]`

Reservoir	Bin	Model	Group	NSE	Bias	Corr.	#
Pickwick	0	downstream	test	-1274684.617	24.653	-0.555	8
Cherokee	0	upstream	train	-5936.453	1.681	0.042	1794
WattsBar	0	downstream	test	-1818.707	11.746	-0.238	3
Wheeler	0	downstream	train	-473.781	31.670	-0.587	5
Douglas	0	upstream	test	-181.109	3.353	0.154	19
Guntersville	0	downstream	train	-156.428	23.465	-0.097	27
Norris	0	upstream	test	-140.395	1.827	0.003	661
Fontana	0	upstream	train	-61.862	1.510	0.099	658
Nikajack	0	downstream	test	-54.291	5.801	0.354	29
Wilson	0	downstream	train	-49.917	9.512	-0.862	7
Chikamauga	0	downstream	train	-44.006	12.572	0.202	13
Cherokee	1	upstream	train	-29.409	2.458	0.273	1347
Hiwassee	0	upstream	train	-24.277	1.047	0.043	726
SHolston	0	upstream	train	-14.498	0.445	0.232	3006
TimsFord	0	upstream	test	-13.513	0.563	0.096	5479
Douglas	1	upstream	test	-10.221	3.032	0.302	2546
Fontana	1	upstream	train	-9.483	1.347	0.441	1441
FtLoudoun	0	downstream	train	-9.285	5.246	0.268	227
Watauga	0	upstream	train	-6.913	0.345	0.253	3952
Hiwassee	1	upstream	train	-3.153	0.694	0.417	2506
Norris	1	upstream	test	-1.497	1.141	0.570	3580
Chatuge	0	upstream	train	-1.263	0.165	0.474	5490
MeltonH	0	downstream	train	-1.023	0.780	0.477	3476
Nottely	0	upstream	train	-0.832	0.146	0.530	5870
BlueRidge	0	upstream	test	-0.177	0.148	0.600	6117
TimsFord	1	upstream	test	-0.154	-0.607	0.551	3051
SHolston	1	upstream	train	-0.149	-0.059	0.498	4569

Reservoir	Bin	Model	Group	NSE	Bias	Corr.	#
Boone	0	downstream	train	-0.144	0.272	0.586	6636
Watauga	1	upstream	train	-0.055	-0.169	0.561	4562
Nottely	1	upstream	train	0.042	-0.270	0.675	3320
Chatuge	1	upstream	train	0.208	-0.253	0.737	3700
BlueRidge	1	upstream	test	0.259	-0.304	0.757	3084
Ocoee1	0	downstream	train	0.681	0.036	0.836	8765
FtPatrick	0	downstream	train	0.744	0.143	0.892	6567
Ocoee3	0	downstream	test	0.821	0.010	0.911	7117
Apalachia	0	downstream	train	0.869	0.040	0.935	8041
Wilbur	0	downstream	test	0.983	0.007	0.991	8179

  - Bias and correlation are highly negatively correlated (-7.14)
  - NSE is negatively correlated with bias and positivily with corr though they are not perfect, (-0.46 and 0.35, respectively).
  - In general, it seems that we capture correlation better than bias.

- Spatially-varying parameters
  - Rather than grouping reservoirs when fitting models, we can use each reservoir as its own group
  - Given the current setup this will basically be fitting a model for each reservoir.
  - We can either:
    * vary intercepts for each reservoir
      · This should help with bias
    * vary slopes for each reservoir and each covariate
      · This should help with variance
    * Vary both
      · Improving bias and variance.
  - Results (because bias and correlation are bad when the score is bad, I vary slopes and intercepts)
    * Thirds score correlated with RT and MStL

Bin	Pearson's *r* (RT)	Pearson's *r* (MStL)
0	-0.423	-0.063
1	-0.764	0.021
2	-0.750	0.077

    * Regress parameters against categorical variables (provide insight into important reservoir characteristics and release parameterizations)

	Intercept		Inflow		Pre. Storage		Pre. Release		7 Day Storage		7 Day Inflow		7 Day Release		$S_{t-1} \times I_t$	
Parameter	Coef.	P-value	Coef.	P-value	Coef.	P-value	Coef.	P-value	Coef.	P-value	Coef.	P-value	Coef.	P-value	Coef.	P-value
Intercept	-0.031	*0.069*	*0.330*	*0.420*	*2.198*	0.036	0.930*	0.000	-2.015	*0.052*	*-0.036*	*0.743*	*0.215*	0.026	-0.556	*0.086*
C(RunOfRiver)[T.1]	*0.034*	0.024	0.416	0.232	-1.052	0.216	-0.544	*0.000*	*0.973*	*0.251*	*0.022*	*0.811*	*-0.112*	*0.155*	*0.277*	*0.299*
C(NaturalOnly)[T.1]	*-0.000*	*0.983*	*-0.055*	*0.875*	*1.326*	*0.130*	*0.129*	*0.314*	*-1.371*	*0.118*	*-0.035*	*0.714*	*0.089*	*0.260*	*-0.124*	*0.643*
C(PrimaryType)[T.1]	*0.003*	*0.714*	*-0.021*	*0.913*	*0.235*	*0.617*	*-0.053*	*0.448*	*-0.233*	*0.620*	*-0.028*	*0.601*	*0.069*	*0.119*	*0.017*	*0.911*
rts	*-0.000*	*0.152*	*-0.003*	0.084	0.004	0.354	0.001	0.253	-0.004	0.365	0.000	0.932	0.001	0.106	0.001	0.333
mstl	0.054*	0.023	0.194	0.725	-2.612	*0.063*	*-0.859*	0.000	2.453*	0.078	0.036	0.811	-0.208	0.103	0.962*	0.032

- Thirds Model (spatial model is still bad for small release values)
  - Fit a model for the lower, middle, and upper thirds (based on all release values, not reservoir specific)

Parameter	(-0.001, 2.311]	(2.311, 17.595]	(17.595, 870.587]
Intercept	-0.172	0.129	0.043
Net Inflow	-0.030	0.189	0.340
Previous Storage	0.099	0.186	0.177
Previous Release	0.451	0.520	0.556
Ave. Storage (7 day)	-0.074	-0.160	-0.137
Ave. Inflow (7 day)	0.026	-0.038	-0.079
Ave. Release (7 day)	0.180	0.171	0.111
Storage and Inflow	0.074	0.130	0.078
Jan	-0.055	0.037	0.047
Feb	-0.098	-0.023	0.034
Mar	-0.145	-0.084	-0.008
Apr	-0.165	-0.161	-0.078
May	-0.127	-0.140	-0.055
Jun	-0.029	-0.114	-0.043
Jul	0.008	-0.082	-0.042
Aug	0.099	-0.018	-0.035
Sep	0.025	-0.010	-0.017
Oct	0.032	0.034	0.005
Nov	-0.024	0.052	0.027
Dec	-0.026	0.063	0.059

  - This model is worse than the spatial model for all thirds
- Seasonal bias for each reservoir.
  - Take the monthly bias plots and break them out in to each reservoir
  - Maybe reservoirs on the y axis, months on the x axis, and color bias
  - (i.e.) a heat map
- I also need to write down the model formulation mathematically

$$R_{r,t} = \Gamma_g \times [R_{r,t-1}, S_{r,t-1}, I_{r,t}, S_{r,t-1} \times I_{r,t}, \bar{R}_{7,r,t}, \bar{S}_{7,r,t}, \bar{I}_{7,r,t}] + M_{g,t} + \epsilon_{r,t}$$

$$t \in [10/23/1990 - 12/31/2015]\ (9,201 \text{ days})$$

$$r \in 27 \text{ reservoirs in Tennessee River Basin}$$

$$g \in \{Groups\} \text{ These groups can be characteristic, tree, reservoir, or quantile based}$$

- Reach out to committee members to set up a date for my preliminary exam during the first week of October
  - October 4th - 8th