Notes from Sankar Meeting

Last Modified: 08/04/2021 09:40 AM

My points

- Dropped highly correlated resevoirs within RT groups
- Upstream reservoirs are not highly correlated so the groups were (0.6, 0.5, 0.4, 0.3)
- Downstream reservoirs where much more correlated so groups were (0.9, 0.8, 0.7, 0.6)
- New score format:
  - (F|P) (+| |-)(#.###)
  - Fitted or predicted
  - + or - bias
  - NSE ### Sankar comments
- Check scores for high and low flows for both simple and tree model
- Leave storage roll 7 out of upstream regression
  - It is actually best to include the difference between the previous storage and 7 day rolling mean

```
X["sto_diff"] = X["Storage_pre"] - X["Storage_roll7"]
```

- Use last 9 reservoirs left to fit, predict on other 18
  - Perform the rest of the analysis using this format
    - Including the high and low flow checks
  - Want to know seasonal bias
  - Individual performance
  - Spatial patterns
  - General observations from performance and from coefficients

My work since the meeting

- Wrote abstract for AGU on this work
  - Sent to Sankar for comments on Friday 7/30/2021
- Tested various parameter combinations
  - Diff's of all phyical variables and their weekly means instead of including both as covariates for upstream model
    - Only storage can be diffed
  - Tried leaving out release predictors from upstream and downstream models
    - Release can be left out of downstream with no real impacts in performance
      - Their performance is dictated more by storage and inflow (just inflow for ROR reservoirs)
    - Release cannot be left out of upstream model, NSE for fitted and predicted drops by 0.1 with out it
      - This indicates that the upstream reservoirs follow more of an autoregressive process than the downstream ones
- Seasonal biases (all values in 1000 acre-ft/day)
  - Upstream

Data Set	Jan	Feb	Mar	Apr	May	Jun
Fitted	-0.2034	0.0947	0.5452	0.6577	0.3449	0.0475
Predicted	-0.1136	0.0410	0.2440	0.2951	0.1582	-0.0005

Data Set	Jul	Aug	Sep	Oct	Nov	Dec
Fitted	0.0302	-0.3817	-0.1927	-0.3896	-0.2874	-0.3973
Predicted		-0.0418	-0.2818	-0.1489	-0.2178	-0.1356

  - Downstream

Data Set	Jan	Feb	Mar	Apr	May	Jun
Fitted	-0.8585	-0.4050	1.2094	2.1272	1.1838	0.3014
Predicted	-1.3566	-0.9886	0.1674	4.0158	1.1473	0.7174

Data Set	Jan	Feb	Mar	Apr	May	Jun
Fitted	0.4162	-0.0385	-0.2278	-0.6141	-1.8239	-1.1695
Predicted	0.3046	-0.1344	-0.3275	-0.8217	-1.0478	-1.5437

  - I made actual and relative value plots for these tables for this weeks presentation. They are bar charts.
    * `python fit9_plots.py -p monthly_bias [--relative]`
  - The above biases for the downstream reservoirs do not include monthly intercepts, including those may reduce the bias.

- High-Low flow analysis
  - Can use

```
<array>_quant, <array>_bins = pd.qcut(<array>, 3, labels=False, retbins=True)
```

    to get 3 quantiles (low, medium, high)
  - Quantile Scores (Downstream)

Bin	Train	Test
0	0.012	-0.015
1	0.768	0.567
2	0.961	0.901

  - Quantile Scores (Upstream)

Bin	Train	Test
0	-8.047	-17.916
1	-2.545	-3.149
2	0.732	0.724

  - Quantile RMSE (Downstream)

Bin	Train	Test
0	53.8%	51.6%

Bin	Train	Test
1	29.6%	39.9%
2	12.8%	22.1%

- Quantile RMSE (Upstream)

Bin	Train	Test
0	205.8%	297.2%
1	91.1%	77.1%
2	27.9%	32.4%

- Histogram plots with KDE, labeled with the scores for this weeks meeting
  * Make plots in python, then edit in powerpoint.
  * `python fit9_plots.py -p quants --pmod hist`
  * `python fit9_plots.py -p quants --pmod 1to1` for one to one plots with abline