Eflows technical team meeting: Tier 1 Update/Discussion

9/6/2018

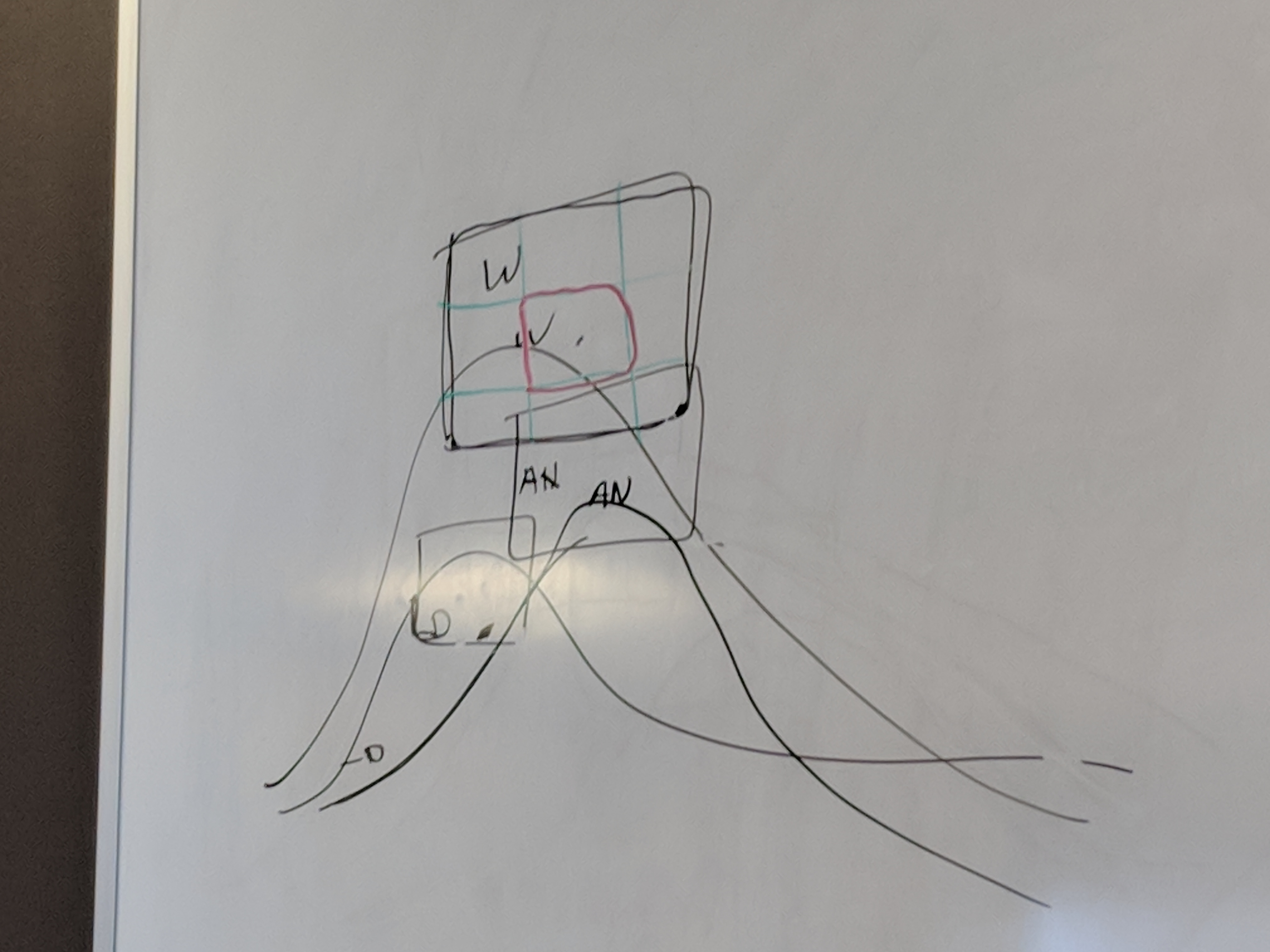
TNC

**Updates**

* October 9 webinar on calculator and detail on the metrics
  + Sam and Belle sent description to DWR to send out
    - Sam will follow up with getting an email out to listerv
  + Avoid scheduling during Bay Delta conference?
  + Record and post this to website
* CWQMC Website
  + Eric will follow up with Dan
* Nature Net postdoc
  + Would start between March and September 2019
  + UCLA/TNC partnership – flow ecology relationships, support tier 2 work, choosing ecological indicators
  + Two-year position
  + Not guaranteed at this point – 1 of 3 projects from UCLA will be chosen
* CWS/SCCWRP joint postdoc
  + Case studies
  + Optimization/modeler would be ideal
* SCCWRP also hiring permanent positions
* Ted will be hiring postdoc for hydrologic modeling at Berkeley
* USFS working to develop flow-sensitive indicators with benthic macroinverts

**Tier 1 Results: Functional Flow Metrics**

* + - Original LOOCV approach: start with all reference sites, for each metric (at annual scale), building a model using geospatial attributes
      * 250 reference sites, generate single forest with all but 1 reference sites, no replacement, boosted trees with full set of reference sites, harvest median prediction across 1000+ trees, use this to measure model performance (compare observation to expected value), median is used to measure sitewise prediction, use percentiles to generate prediction interval
        + Issue with this—large prediction intervals, didn’t conform well to interannual variation or observations themselves, unrealistically large prediction intervals constrained alteration assessment; this approach may have inflated error
    - Ensemble approach: start with all reference sites, generate 1000 different forests for each metric, each uses subset of 250 reference sites (~90%), generate forest from each of those forests, each has 500 trees, then calculate median prediction of all trees in forest, have 1000 median predictions, leaves some reference sites out which gives better prediction of uncertainty. Evaluate model performance – take all forests that excluded a particular reference gage (100)
      * Currently reviewing observation data to throw out any years/data that may be incorrect and impacting results
      * Next step – how to determine which metrics are the most useful where
      * Look at differences regionally which variables drive the performance at each reference site, if it works well in a certain region/class/slope/etc, we move forward with that variable
    - Group discussion: random forest, scaling, predicting based on gaged streams with proportional area – or, just stick with random forest and clarify that there’s uncertainty that might trigger tier 2
* Don’t scale the timing metrics but scale everything else?
* Evaluating the timing metrics
* Evaluate by water year?
* By stream class?
* Cross validate
* Ex: all snowmelt across state gets a particular timing regardless of region, may parse this by water year – evaluate how well this prediction works
* **Take class wide DRH, calculate timing metric, compare this to timing metric calculated at each reference gages within class – see how well this class wide metric falls within range**
* Eventual outcome of this is a prediction of flow metrics for every year and a range of prediction intervals for all FF metrics for every stream segment in the state
  + - Need to go back and gray out areas where there’s uncertainty in the stream classification (SJ classified as GW, etc)
      * Talk about this on Monday, 9/10/18
* Look at performance of metrics (as they vary by water year, stream class, geospatial attributes)
  + Investigate how other factors (above) explain differences in composite performance
  + Decide as a group on a threshold of excluding data or whether to include everything with some qualifiers/foot notes
* Predicting timing: each site in each class calculates each metric for each year, then within each WY type for each stream class take a percentile range of each metrics, use this as the range, then any time you’re in a new segment, use that model to predict range for a particular WY type in a particular stream class. Output of model is percentile value.
* “box discussion”



* Need to evaluate by water year type, create box for each flow component, bounds of box are based on flow metrics, split up by percentiles.
* Sizes of boxes will depend on natural range of variation at reference gages for different water year types
* 10, 25, 50 percentiles
* Repeat this for prediction values, overlay the reference and see how these compare. This will help inform where the percentile thresholds will be.
* This method only accounts for 64% of the variability – something to think about.

**Water Year Types**

* Ask Daniel Swain about water year type
* Precipitation and temperature to determine this? Predicted mean annual discharge?
* Use “unimpaired water year type” when referring to this
* For every segment, have a prediction for mean annual discharge based on water year type
* Potentially compare this to 8 station index?
* Is tier 1 always retrospective?
* Use climate projections for water year type
* **Decision:** quartiles - dry, below moderate, above moderate, wet??
  + 30 or 50-year record for every segment, calculate percentiles, look at how consistent water year types are across space
  + For each reference gage, determine what water year types are classified as by DWR – student project?
  + Want to apply a record requirement to this analysis – 20 years
  + Student will look at MAF for each reference gage site and classify water year type and pull percentiles, and do a comparison between unimpaired water type and DWR water type
  + Within observed data, can look at variations between stream gages over time
  + With predicted data, can use monthly model to calculate MAF
  + Some of these details can be put into tier 2

**Natural flows database tool**

flowline.codefornature.org

* Technical team – review this site and provide feedback to Julie
* Need “landing page” that describes CEFF and directs people to other sites/pages
* Apply for funding for website to house data
* Ask Greg about this (Eric), then talk to Dan, then reach out to foundations