Chat from 5/21/2020

<https://github.com/InteragencyEcologicalProgram/DataScience/tree/master/2020.05.21>

[‎5/‎21/‎2020 1:35 PM]  Flynn, Ted@DWR:

that was amazing!

[‎5/‎21/‎2020 1:35 PM]  Acuna,Shawn C:

ditto

[‎5/‎21/‎2020 1:36 PM]  Travis Hinkelman:

How is this system similar or different than Bay-Delta Live or SacPAS?

[‎5/‎21/‎2020 1:37 PM]  Juliana Spector:

Would it be possible to export spatial visualizations from the data integration system into ArcGIS or other spatial software?

[‎5/‎21/‎2020 1:37 PM]  Rasmussen, Nicholas@DWR:

This is awesome! My impression is that water quality parameters like temperature, turbidity, and salinity would generally be highest priority. Many folks focus on these parameters directly and others use these parameters as proxies for biological parameters

[‎5/‎21/‎2020 1:39 PM]  Rasmussen, Nicholas@DWR:

Also, maybe Chuck could share the link for the website he has been describing

[‎5/‎21/‎2020 1:40 PM]  Acuna,Shawn C:

Currently what data sets are already part of this platform? Will you be integrating the contaminant data? How will you determine your prioritization of what data should be integrated?

[‎5/‎21/‎2020 1:56 PM]  Ted Swift:

I would think patchiness would be a severe challenge: It would cause most of your variation if your sampling sites are fixed, but the SAV and FAV are moving around.

[‎5/‎21/‎2020 1:57 PM]

https://cran.r-project.org/web/packages/pwr/pwr.pdf

[‎5/‎21/‎2020 1:58 PM]  Rasmussen, Nicholas@DWR:

yep, that's what we've used

[‎5/‎21/‎2020 1:58 PM]  Kyle Griffiths:

as I understand, PA is specific to the statistic you're looking at, do you have a model/test in mind already?

[‎5/‎21/‎2020 1:58 PM]  Bashevkin, Sam@DeltaCouncil:

Here's 2 more packages, but I'm not sure if they'd help with multivariate stats: https://besjournals.onlinelibrary.wiley.com/doi/full/10.1111/2041-210X.12748   
  
https://besjournals.onlinelibrary.wiley.com/doi/full/10.1111/2041-210X.12504

\

[‎5/‎21/‎2020 2:03 PM]  291e0f1b-1b09-44be-89d9-79309fe0910b@anonymous.invalid:

Ive been using Principal Tensor Analysis for space-time-community assemblage data

[‎5/‎21/‎2020 2:03 PM]  Travis Hinkelman:

It’s nice to hear from people who have too much data. Logistically speaking, which is more challenging, sampling frequently or spatial subsampling?

[‎5/‎21/‎2020 2:05 PM]

Sampling frequently is usually harder than spatial subsampling.

[‎5/‎21/‎2020 2:07 PM]

And the problem isn't "having too much data" the problem is "doing too much work". Those are some long days in the field!

[‎5/‎21/‎2020 2:08 PM]  Travis Hinkelman:

Well, that’s what I meant. People usually err on the side of not collecting enough data.

[‎5/‎21/‎2020 2:08 PM]  Flynn, Ted@DWR:

yeah it's not like we're Supreme Court justices or anything :D

[‎5/‎21/‎2020 2:08 PM]  Brendan Wakefield:

I have also used the power package for ANOVA's in R, but if you have blocking variables (randomized complete block design/stratified design) you will need to keep that in mind for your number of samples (observational unit is the sample, but experimental unit is the plot) ; there is also the G\*Power software as an alternative

[‎5/‎21/‎2020 2:08 PM]  Ellis, Daniel@Wildlife:

To Travis- we are sampling this much because we knew we would want to pair it back later, and we wanted the stats to back up our choices

[‎5/‎21/‎2020 2:10 PM]  Ellis, Daniel@Wildlife:

Thank you all for your comments and suggestions.

[‎5/‎21/‎2020 2:11 PM]  291e0f1b-1b09-44be-89d9-79309fe0910b@anonymous.invalid:

Hey Dan, check out Cichocki et al. 2015 IEEE for PTA

[‎5/‎21/‎2020 2:12 PM]  291e0f1b-1b09-44be-89d9-79309fe0910b@anonymous.invalid:

tite: Tensor Decompositions for Signal Processing Applications

[‎5/‎21/‎2020 2:18 PM]  Khanna, Shruti@Wildlife:

will we have access to the ppts of today's presentations?

[‎5/‎21/‎2020 2:19 PM]

Yes! https://github.com/InteragencyEcologicalProgram/DataScience/tree/master/2020.05.21

[‎5/‎21/‎2020 2:19 PM]  Khanna, Shruti@Wildlife:

great, thanks

[‎5/‎21/‎2020 2:19 PM]

Vanessa and my presentation are already up there. I'll gather any others that teh presentors are willing to share

[‎5/‎21/‎2020 2:35 PM]  Travis Hinkelman:

I don’t think that Dayflow includes water year type so that must have been added in a separate step.

[‎5/‎21/‎2020 2:36 PM]  Ted Swift:

Yes, I think WY classification must have come from, e.g., the CDEC listing.

[‎5/‎21/‎2020 2:37 PM]

THe Dayflow data is a bit of a mess. Some of the historic data includes it, some doesn't. Vanessa's dataset has been cleaned up considerably.

[‎5/‎21/‎2020 2:38 PM]  Ted Swift:

Is there a statistical test to determine whether your choosing too large a “k”? Something like an Akaiki Information Criterion, or the d.f. you actually have?

[‎5/‎21/‎2020 2:42 PM]  Bashevkin, Sam@DeltaCouncil:

There's a gam.check function to help with that

it includes a statistical test

[‎5/‎21/‎2020 2:45 PM]  Brendan Wakefield:

Hey Ted, the AIC (as well as ANOVAs and the BIC) will allow you to compare fit qualities of your model objects, and this is certainly dependent on different levels of k

[‎5/‎21/‎2020 2:48 PM]  Travis Hinkelman:

Don’t worry about running out of time. We can bump the best practices discussion.

[‎5/‎21/‎2020 2:48 PM]  Bashevkin, Sam@DeltaCouncil:

I think the issue with setting K too high is more related to the computational cost of that choice, than the risk of creating a "too wiggly" model, since mgcv will adaptively choose how many knots it really should have, as Venessa said

[‎5/‎21/‎2020 2:51 PM]  291e0f1b-1b09-44be-89d9-79309fe0910b@anonymous.invalid:

It may be worth noting that you can set “fx=FALSE, k=-1” and mccv will use cross-validation to optimize smoothing. This prevents having to mess with k

[‎5/‎21/‎2020 2:51 PM]

It also depends on the reason behind your analysis. IF you want to describe every "wiggle", than you can have a high K value, no problem. However, if you want to describe the big wiggles and ignore the small wiggles, you will want to set k lower.

[‎5/‎21/‎2020 2:55 PM]  Khanna, Shruti@Wildlife:

Excellent ppt Vanessa. Nice intro to basics behind the method

[‎5/‎21/‎2020 2:55 PM]  291e0f1b-1b09-44be-89d9-79309fe0910b@anonymous.invalid:

gamm4 is an lme4-based mixed effects gam package that uses syntax similar to lmer

Mccv is great for autocorrelation and weighting; gamm4 is great for >2-level models. FYI.

[‎5/‎21/‎2020 2:56 PM]  Khanna, Shruti@Wildlife:

Jereme, you need to follow up with how mixed-effects models fit in the big GAM picture

[‎5/‎21/‎2020 2:57 PM]

I love MuMIn. I use it for linear models all the time. the 'dredge' function is magical.

[‎5/‎21/‎2020 2:57 PM]  291e0f1b-1b09-44be-89d9-79309fe0910b@anonymous.invalid:

sorry MGCV

[‎5/‎21/‎2020 3:00 PM]  Ted Swift:

Sounds like your basis functions can have actual analytical meaning, where loess is just smoothing.

[‎5/‎21/‎2020 3:01 PM]  Bashevkin, Sam@DeltaCouncil:

I just came across this today on LOESS vs gam: https://fromthebottomoftheheap.net/2016/04/10/loess-revisited/

[‎5/‎21/‎2020 3:02 PM]  Vanessa Tobias:

Yes! I'm happy to answer more questions later!

[‎5/‎21/‎2020 3:03 PM]  Deanna Sereno:

Will this chat, will all the wonderful links, be posted on github after the meeting?

[‎5/‎21/‎2020 3:04 PM]

I think I can post the chat! I'll try at least.

[‎5/‎21/‎2020 3:36 PM]  Ted Swift:

If you’re notebook needs to represent the project, why not put the code in it?

[‎5/‎21/‎2020 3:36 PM]

Very long lines of code

[‎5/‎21/‎2020 3:37 PM]  Juliana Spector:

Messy formatting

[‎5/‎21/‎2020 3:40 PM]

No comments!

[‎5/‎21/‎2020 3:41 PM]  Vanessa Tobias:

I second the "no comments".

[‎5/‎21/‎2020 3:43 PM]  Vanessa Tobias:

Comments are best for why you're doing something.

[‎5/‎21/‎2020 3:45 PM]

Or I might need it later...

[‎5/‎21/‎2020 3:50 PM]  Ted Swift:

https://medium.com/better-programming/string-case-styles-camel-pascal-snake-and-kebab-case-981407998841

[‎5/‎21/‎2020 3:51 PM]  Ted Swift:

From the article:

[‎5/‎21/‎2020 3:51 PM]  Ted Swift:

TLDR;

    camelCase

    PascalCase

    snake\_case

    kebab-case

[‎5/‎21/‎2020 3:52 PM]

Being internally consistant is the most important

[‎5/‎21/‎2020 3:55 PM]  Vanessa Tobias:

Friently tip for nor new coders: if it works, it's good code. Don't worry if it's ugly or stinky. You can adopt these style things as you get more comfortable. :)

Just in case anyone was getting nervous