

Evaluation results: Advanced statistics and machine learning for data analysis (January 22-24, 2018)

Course details

Instructors: Fabian Gans, Paul Bodesheim, Guido Kraemer, Thomas Wutzler

The course aims at giving an overview on concepts of (some advanced) applied statistics and machine learning methods for data analysis. We will cover topics such multivariate explorations, dimensionality reduction, data visualization, multivariate predictions, and time series analysis. The doctoral candidate should obtain a broad overview on the currently used techniques, they must be able to “read” results produced by most important methods, and interpret the statistics correctly as well as with caution. We will provide the participants with perspectives offered by state-of-the-art methods and give orientations where to start their own analyses. Exercises will emphasize some techniques that we think are most suitable in the context of Earth system sciences (and depending on the demand: ecology).

Detailed information is provided on the webpage:

<http://www.imprs-gbgc.de/index.php/Courses/ASDA2017>

6 out of 10 participants filled in the survey by February 1, 2018.

Survey results

Please assess the course in general.

The course stimulated my interest in this topic.

0% Strongly disagree
0% Disagree
0% Neither agree nor disagree
17% Agree
83% Strongly agree

I think that the level of difficulty of this course is appropriate.

0% Strongly disagree
0% Disagree
0% Neither agree nor disagree
83% Agree
17% Strongly agree

I think the lectures were organized in a logical way.

0% Strongly disagree
0% Disagree
0% Neither agree nor disagree
100% Agree
0% Strongly agree

Overall, I am satisfied with this course.

0% Strongly disagree
0% Disagree
0% Neither agree nor disagree
83% Agree
17% Strongly agree

Which parts of the workshop were especially good (and why)?

- The overall level of teaching and exercises was really good. Complex topics were broken down nicely, I feel I could understand most of the basics behind some advanced statistical techniques covered. There was a good student-to-teacher ratio and a lot of flexibility in the course structure to adapt to individual questions and topic details.
- I liked the part by Paul, he really broke the topic down to an easy to understand way and took plenty of time to explain everything and understand questions. Further I liked the student presentations (despite differing quality) and the exercises as they made it more interactive
- In particular Paul's presentations were very good. He explained very complex methods in a way that I could follow, although I was not familiar with them. I think a lot of preparation is required for this.
- Paul's machine learning lecture was great, both because of the subject and because of his teaching skills. Guido's lecture was also very good.
- Paul had very good slides. He addressed a currently relevant topic.
- Machine learning and dimension decomposition They're clearly conveyed

Which parts of the workshop were not so good / not so fitting / not well enough presented?

- For me a short recap on matrix calculation/operations and complex numbers could have been useful to be able to follow all the way through (especially on Day 1); or a suggestion to look into this prior to the course. It was hard to know which level of detail the participant presentation should be made at, and the level of detail was very heterogeneous.
- I missed a short general introduction about the topics at the beginning, linking the topics of each of the 3 teachers. Further, I found it difficult to follow Guido's explanations of PCA, it seemed a little unstructured
- All parts were very interesting and quite good presented.
- Fabian's lecture was not clear enough for me, but he was very helpful during the exercises.
- a bit more background and structure for dim. Red. part would be helpful.
- Maybe the time series analysis could be a little more explained as it's quite important

Do you have other suggestions for a future course?

- The course content could easily have fitted a 4-5 day course, especially if there should be enough time for all the exercises.
- we needed half a day more, so increasing the length of the course to 4-5 days would be appropriate to go through the topics more slowly and have time for (understanding) questions
- more time