

# Unsupervised profile classification can help create interpretable oceanographic knowledge

Dan(i) Jones, British Antarctic Survey, NERC, UKRI, UK

Profile classification models (PCMs) attempt to find a set of “profile types” or “classes” in profile datasets (Maze et al., 2017)

As with many classification problems, there is a balance between (1) the ability of the PCM to accurately represent the underlying covariance structure whilst avoiding overfitting and (2) the interpretability of the PCM (Fig. 1)

As an example case, we consider a collection of profiles from south of the Polar Front, including profiles from Argo floats and ship-based CTD casts (Fig. 2)

Fig. 1

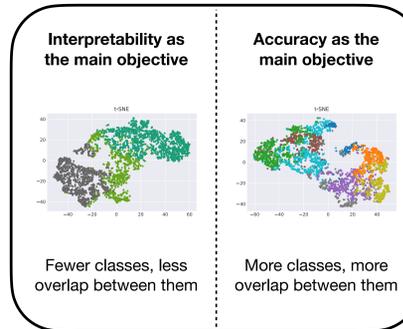


Fig. 2

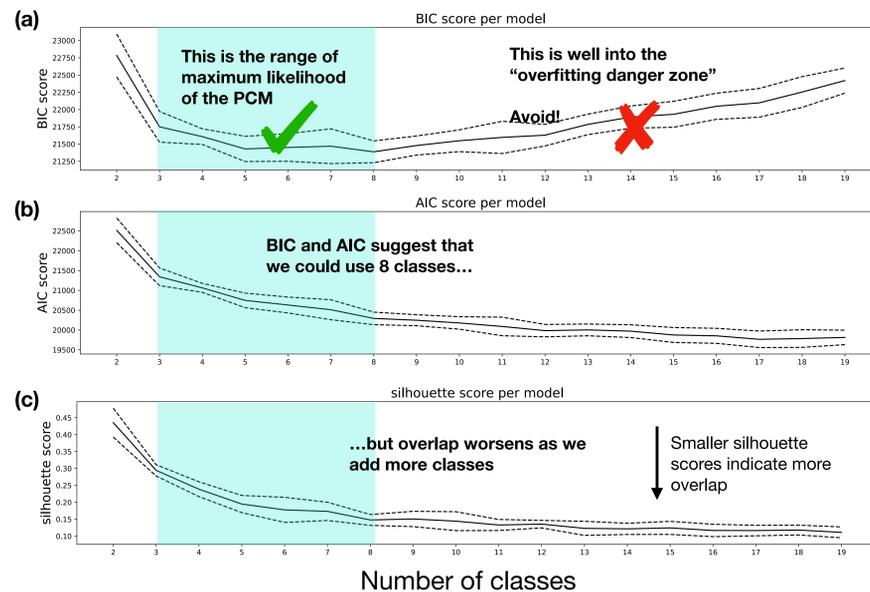
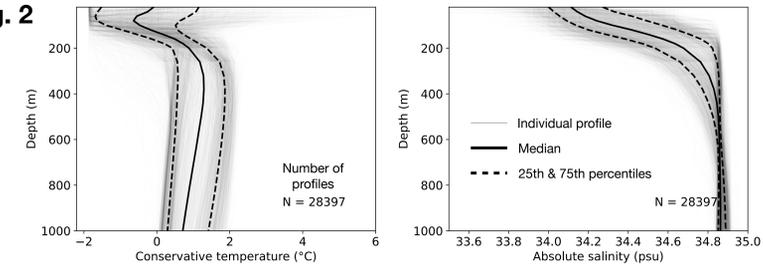


Fig. 3. We use a combination of statistical criteria to guide our choice of the number of classes. This guidance suggests a range between 3-8 profile types; within this range, likelihood increases but overlap worsens

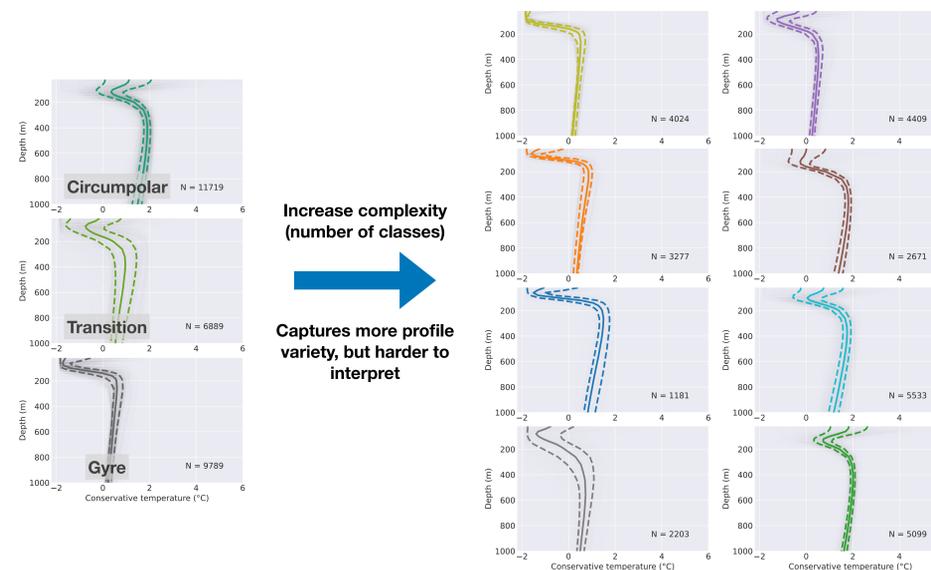


Fig. 4. As we increase the number of classes, and thereby the complexity of the PCM, we represent more variability at the expense of interpretability

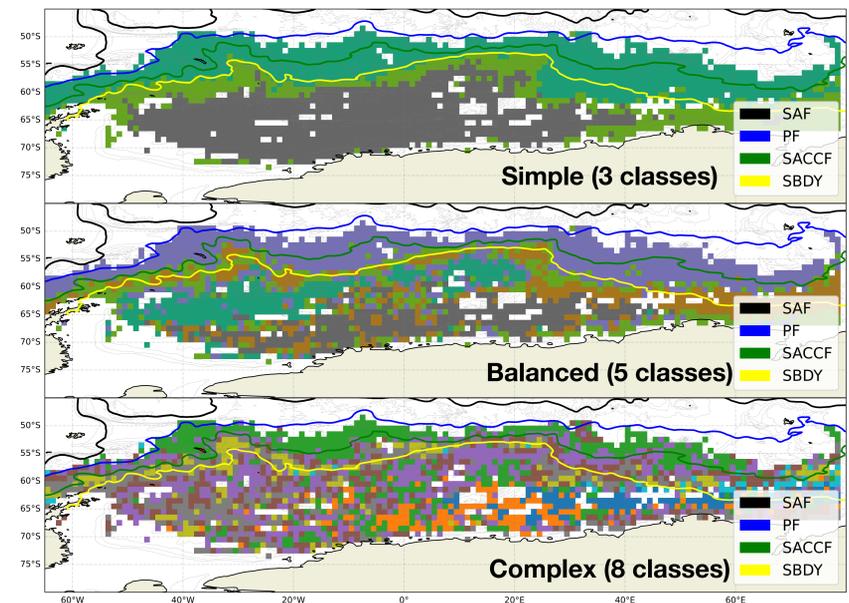


Fig. 5. The spatial distribution of the classes shows worsening overlap as we increase the number of classes. Although the eight-class model is “allowed”, it is not trivial to interpret

Summary:

Profile classification models can be useful discovery tools, highlighting both expected and under-appreciated structures in profile datasets

Statistical guidance can help us design a suitable classification model, but the ability of a model to accurately represent variability may come at the expense of interpretability. Expert judgment is needed to balance accuracy and interpretability; the ideal balance depends on the objective of the application

For an overview of unsupervised classification applications in oceanography, as well as machine learning in general, see Sonnewald et al. (2021)

[dannes@bas.ac.uk](mailto:dannes@bas.ac.uk)  
[@DanJonesOcean](https://twitter.com/DanJonesOcean)  
#AGU2022

DJ is funded by a UKRI Future Leaders Fellowship (MR/T020822/1)

References

Maze et al. (2017), <https://doi.org/10.1016/j.pocan.2016.12.008>  
Sonnewald et al. (2021), <https://doi.org/10.1088/1748-9326/ac0eb0>