Two examples of Machine Learning in Geoscience:

Self-Organizing Maps (SOMs) & Feed-Forward Networks (FFNs)

Lydia Keppler
PhD candidate (she/her)

Observations, Analysis and Synthesis (OAS)
Director's Research Group (DRO)

Max-Planck-Institut für Meteorologie

International Max Planck Research School on Earth System Modelling

IMPRS-ESM

Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

Clustering: Self-Organising Maps (SOMs)

- SOMs to cluster data into regions of similar properties (Kohonen 1987, 2001)

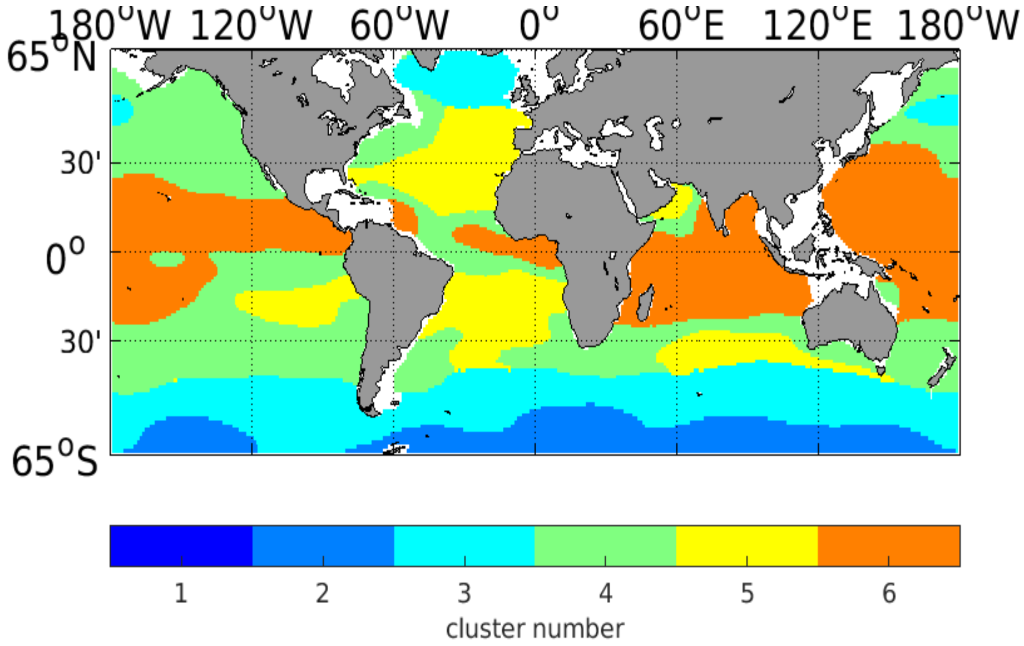


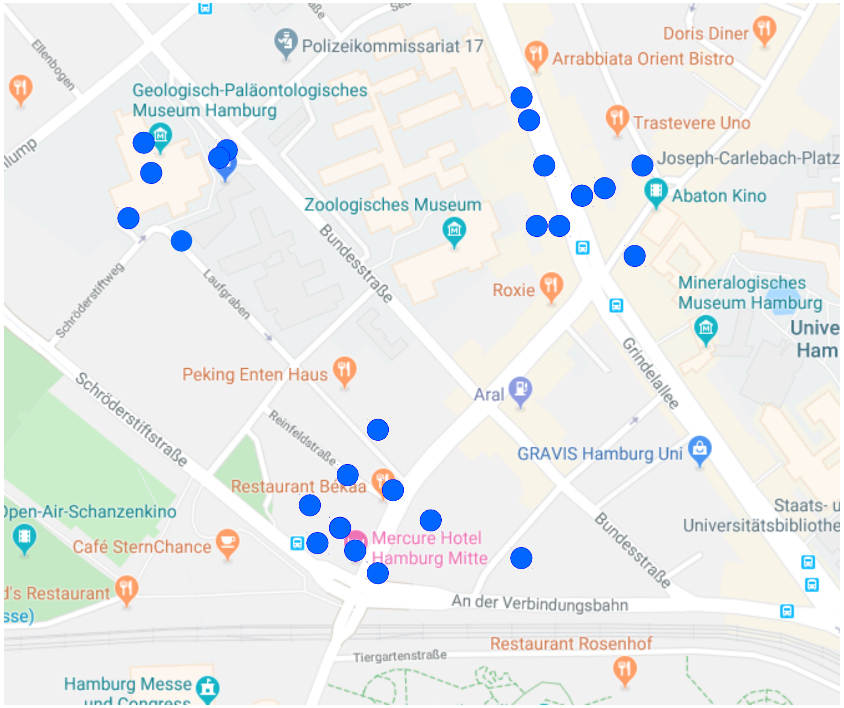
Figure adapted from Keppler et al. (in prep.)

Where to build a pizzeria

- Giovanni wants to build three pizzarias near the MPI-M
- Where should he put them?

Where to build a pizzeria

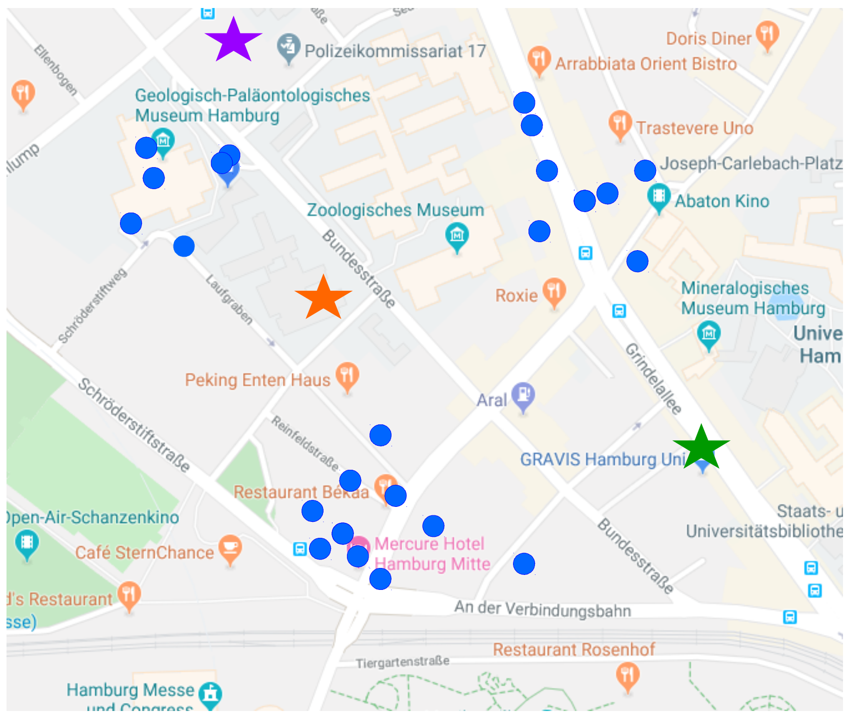
- Giovanni wants to build three pizzarias near the MPI-M
- Where should he put them?



● people who eat a lot of pizza

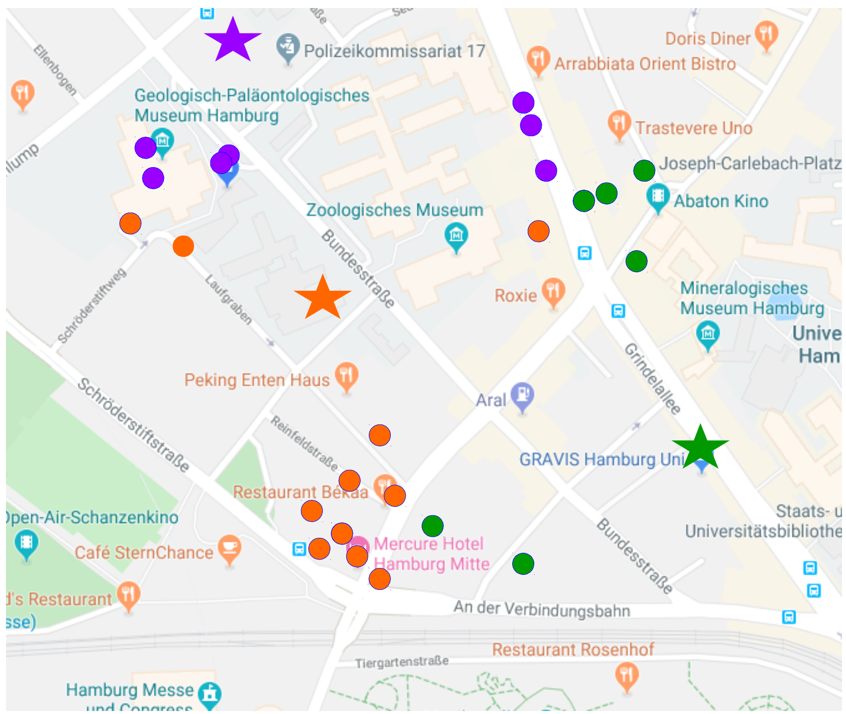
Where to build a pizzeria

- We start with three random locations for the pizzeria



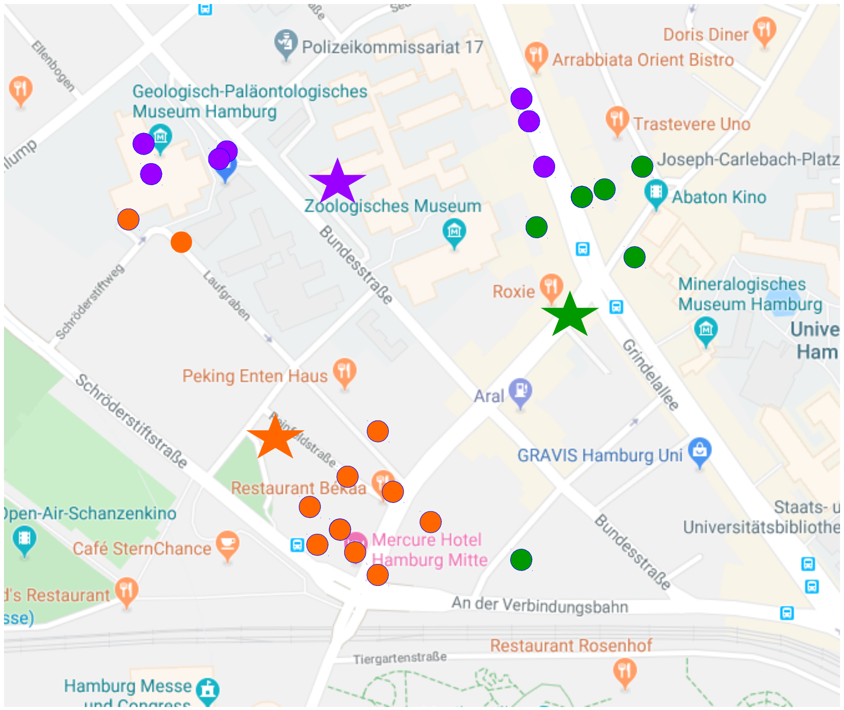
Where to build a pizzeria

- We start with three random locations for the pizzeria
- Everyone goes to the one that is closest



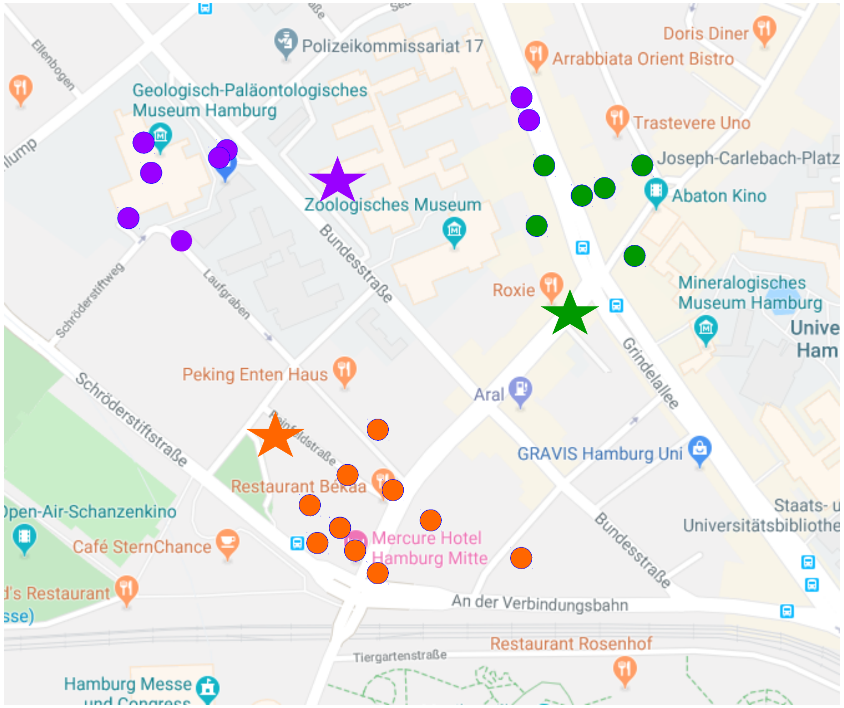
Where to build a pizzeria

- We move the pizzaria to the center of the houses



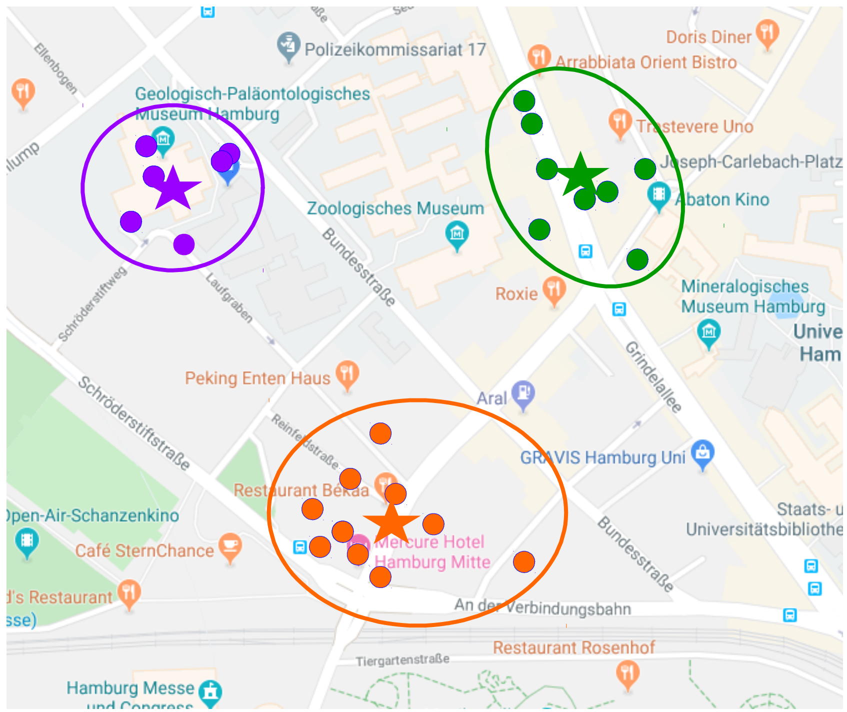
Where to build a pizzeria

- We re-adjust the closest pizzarias for each house



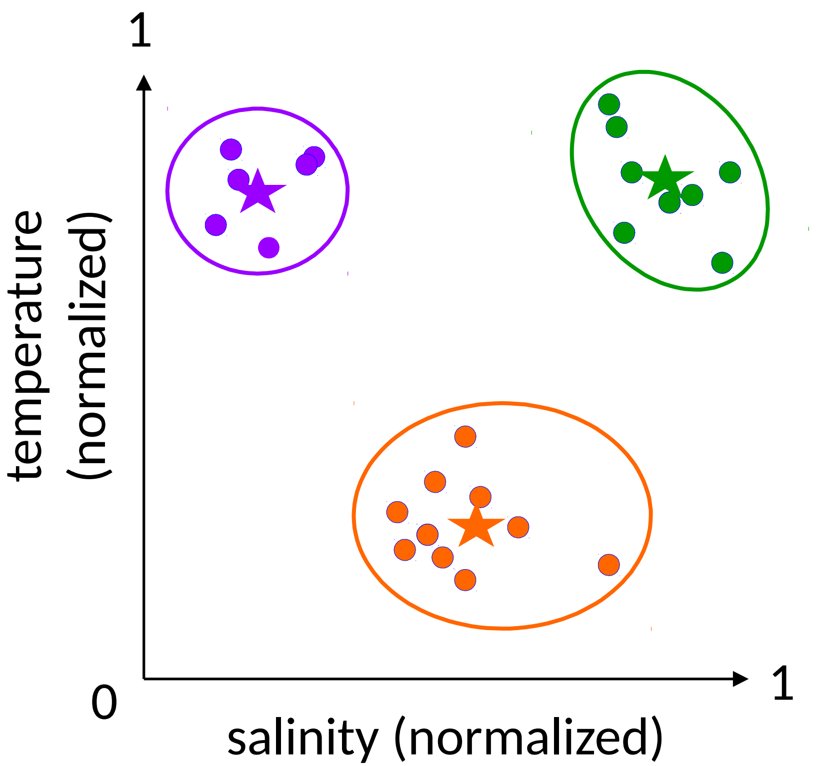
Where to build a pizzeria

- We repeat this process until the distances do not get smaller anymore



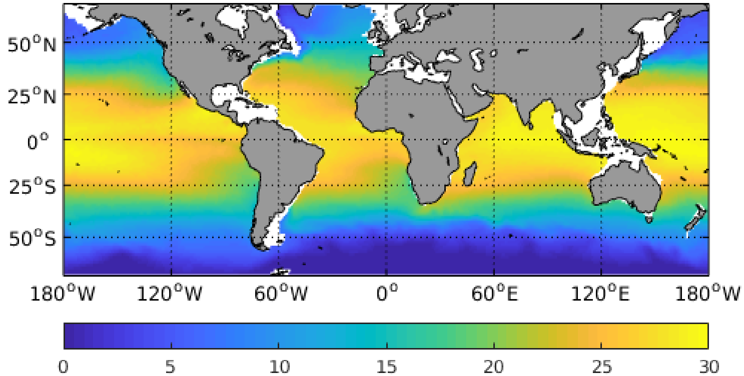
SOM-clustering with non-pizza variables

- We can e.g., use normalized temperature and salinity to cluster the ocean into water masses

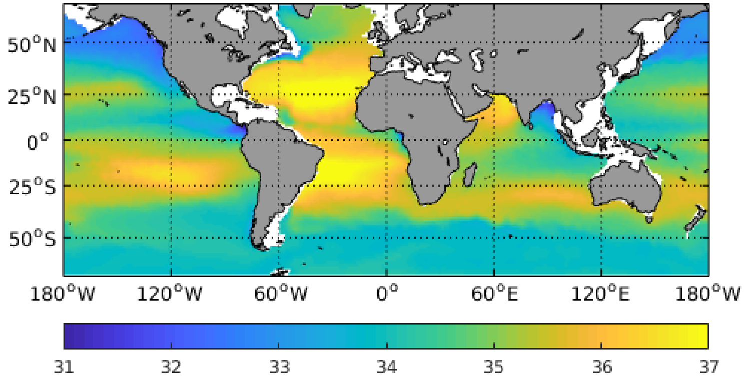


SOM-clustering with non-pizza variables

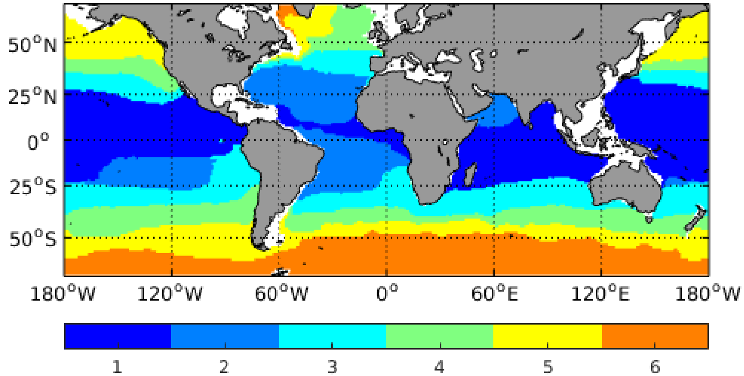
- Example: temperature and salinity at 10 m as input to SOMs



temperature (°C)



salinity



cluster number

Feed-Forward Networks (FFNs)

- FFNs compute and apply statistical relationships between multiple predictor and target variables **to approximate a function**

  → like a MLR, but the relationships don't have to be linear

- Here: from sparse data with gaps to mapped data

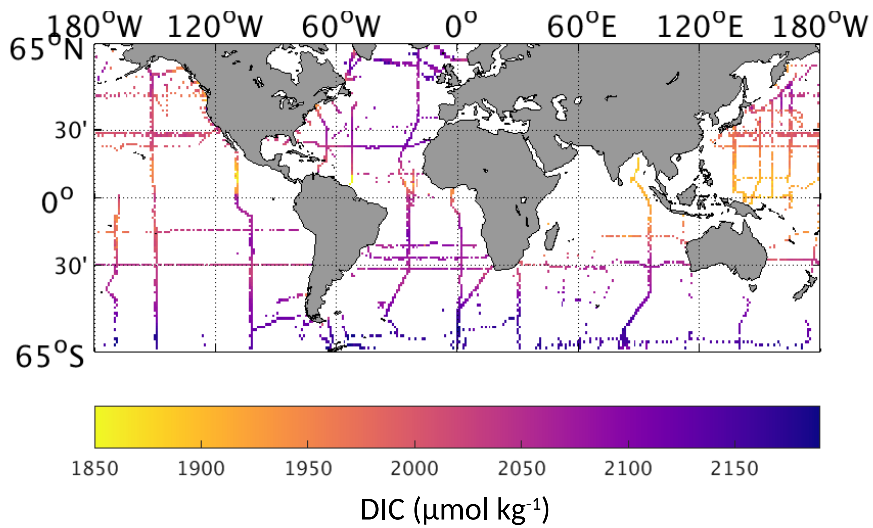
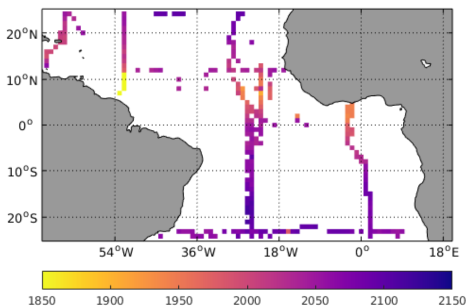


Figure adapted from Keppler et al. (in prep.)

Feed-Forward Networks (FFNs)

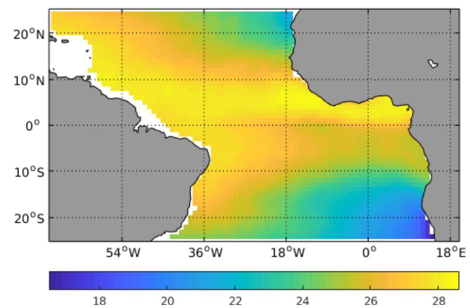
- We have sparse ship data which we want to have mapped (target data)
- We need predictor data (mapped global data; e.g., temperature / salinity)
- The network establishes the statistical relationship between the predictor and the target data and then applies this relationship to map the target data

target data

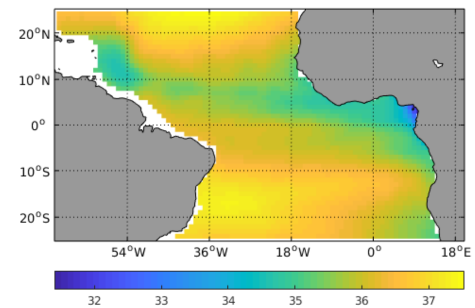


DIC (µmol kg⁻¹)

predictor data



temperature (°C)



salinity

Establishing the relationship (training the FFN)

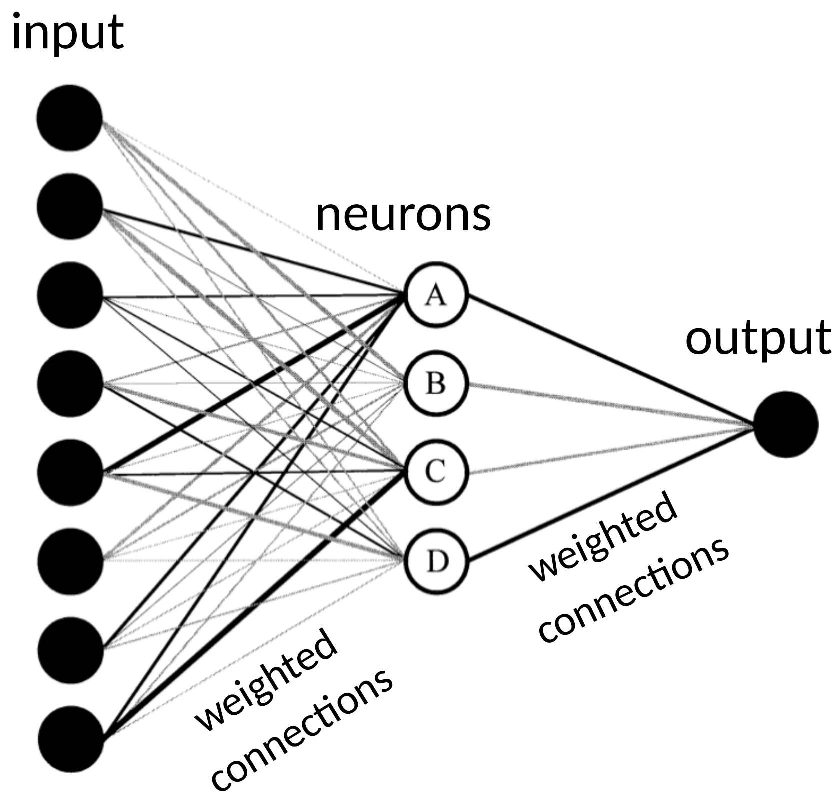
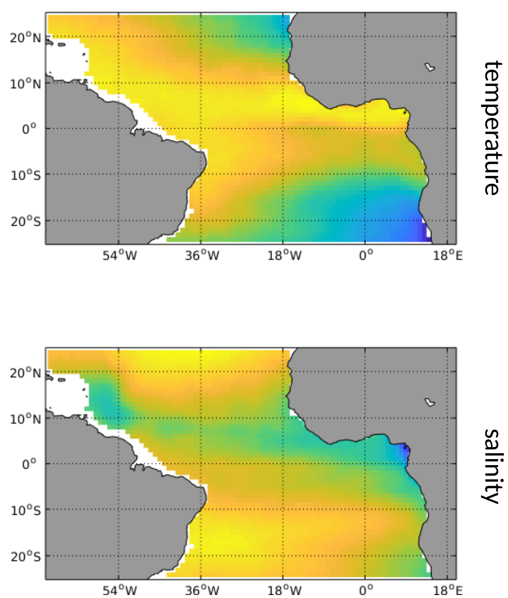


Figure adapted from Olden& Jackson, 2002

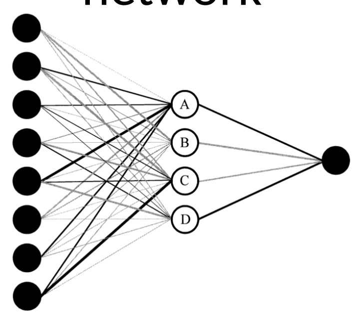
Applying the relationship

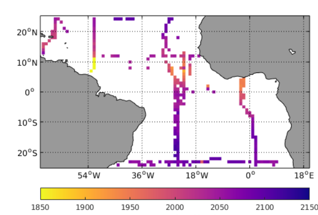
- Now the established relationship between the target data and the predictor data is applied to map the target data globally (disclaimer: the resulting field is not realistic, because of the simplified set-up, e.g. only SST and SSS as predictors; testing of the result is important)

predictor data

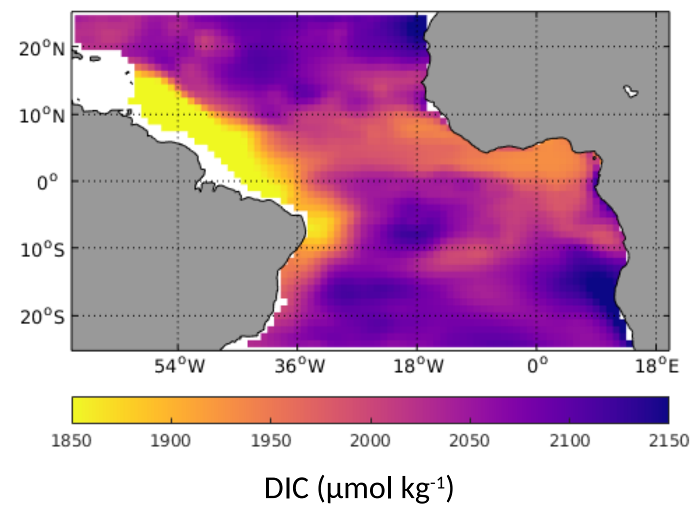


trained network



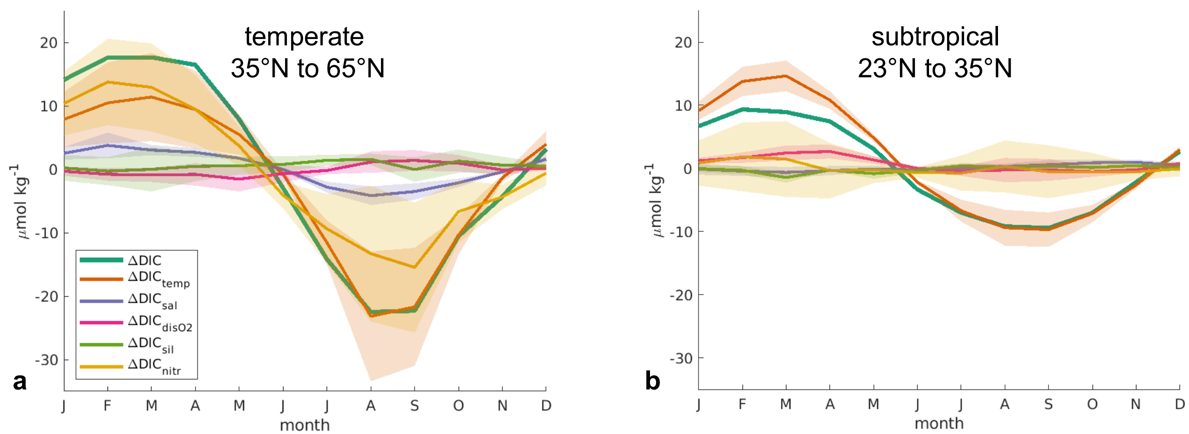


output



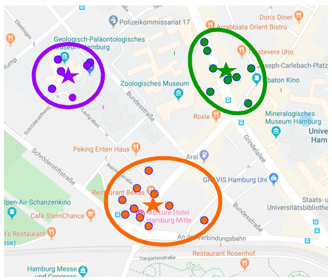
Understanding the resone

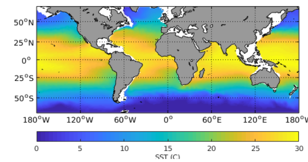
- We can see how each of the predictors contributes to the FFN (Similar to profile method by Gevrey et al., (2003))
- We train the network as usual, and in the simulation-step, we hold one predictor constant in time, and vary the others (iteratively for all predictors)
- We get the change in DIC due to each predictor

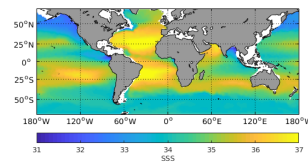


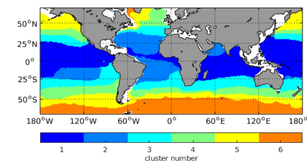
Summary

- **SOMs** can **cluster** data into (e.g., into regions of similar properties)









- **FFNs** can compute and apply statistical relationsships between multiple predictor and target variables **to approximate a function**