Karlsruhe Institute of Technology

Steinbuch Centre for Computing and
Institute of Meteorology and Climate Research

D1: A MODULAR SOFTWARE FRAMEWORK FOR COMPRESSION OF STRUCTURED CLIMATE DATA

Ugur Cayoglu, Jennifer Schröter, Jörg Meyer, Achim Streit, and Peter Braesicke

- Through the introduction of next-generation models the climate sciences have experienced a breakthrough in high-resolution simulations, which calculate global simulations with a resolution of five kilometers (e.g. ICON-ART)
- The new models produce an unprecedented volume of data in climate research, so that future studies are limited by the storage capacity rather than numerical calculations
- We propose a modular software framework for the development of a customized prediction-based compression algorithm for structured spatio-temporal data
- The framework helps with the development of a prediction-based compression method by providing a strictly defined interface, concurrent compression for fast testing, implementation of already established prediction models, the possibility to generate ensemble predictors and fast iteration via multi-dimensional subsetting of datasets

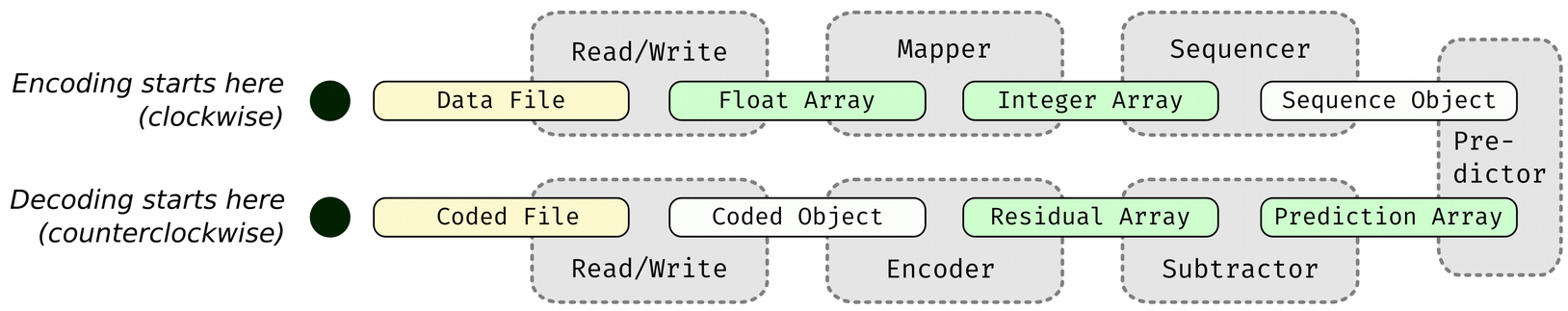
How does prediction-based compression work?

How is compression achieved?

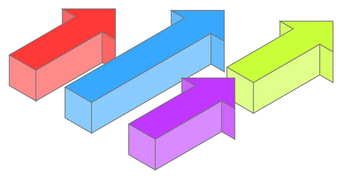
Read the data → Map data to integers → Traverse data using a predefined route → Predict each value based on past values → Calculate residual → Encode residuals and write to disk

```
10111010101111100110101  // prediction          3059509
10111010101111100100110  // truth               3059494
00000000000000000110011  // prediction^truth         19 (XOR)
**********************   // LZC = 16 + 1
                         // (only the last five bits are written on disk)
```

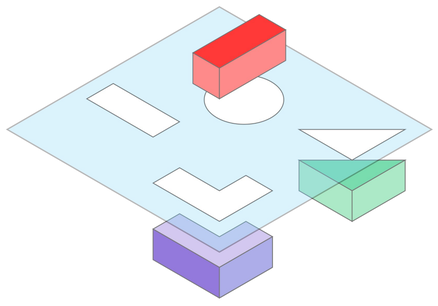
How are these steps implemented in the framework?



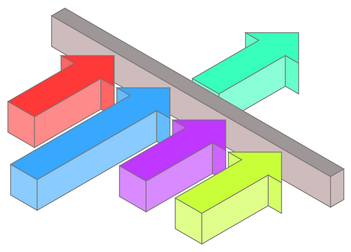
The framework consists of two components: Modifiers and Objects. A modifier defines the transformation process from one object to the other. An object defines the interface to the data.



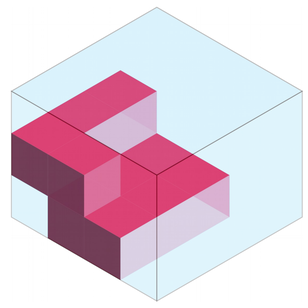
Concurrent compression of a single dataset using several predictors or using a single predictor with several datasets



Interaction of modifiers and objects follow strictly defined interfaces to guarantee interoperability



Several predictors can be merged to a single entity to improve prediction accuracy



Support for random subsetting of multi-dim. data for fast iteration

cayoglu@kit.edu
http://github.com/ucyo/cframework