A Conceptual Introduction to Principal Component Analysis (PCA)

Steven Greybush
University of Maryland
AOSC 630

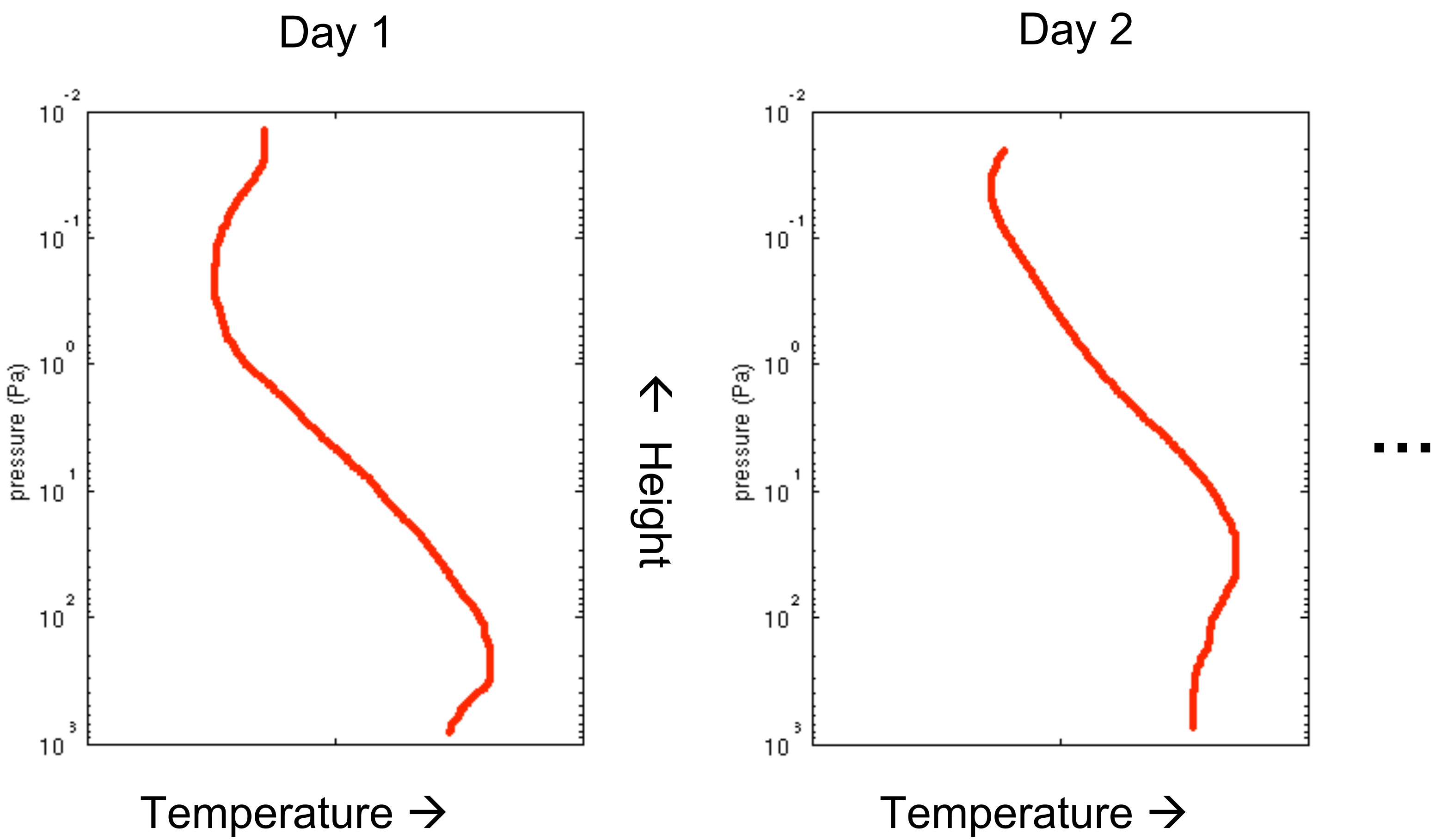
Note: These slides complement Lecture 14 of AOSC 630, taught by Prof. Eugenia Kalnay.

Principal Component (EOF) Analysis

- Have a set of spatial maps (or vectors).
- Would like to describe each map as a sum of pattern vectors (EOFs).
- These pattern vectors are chosen to be orthogonal, and are ordered according to the amount of variance in the dataset that they describe.
- The weights in this sum are known as amplitudes, or principal components.

Principal Component (EOF) Analysis

- Example of spatial maps: vertical temperature profiles.



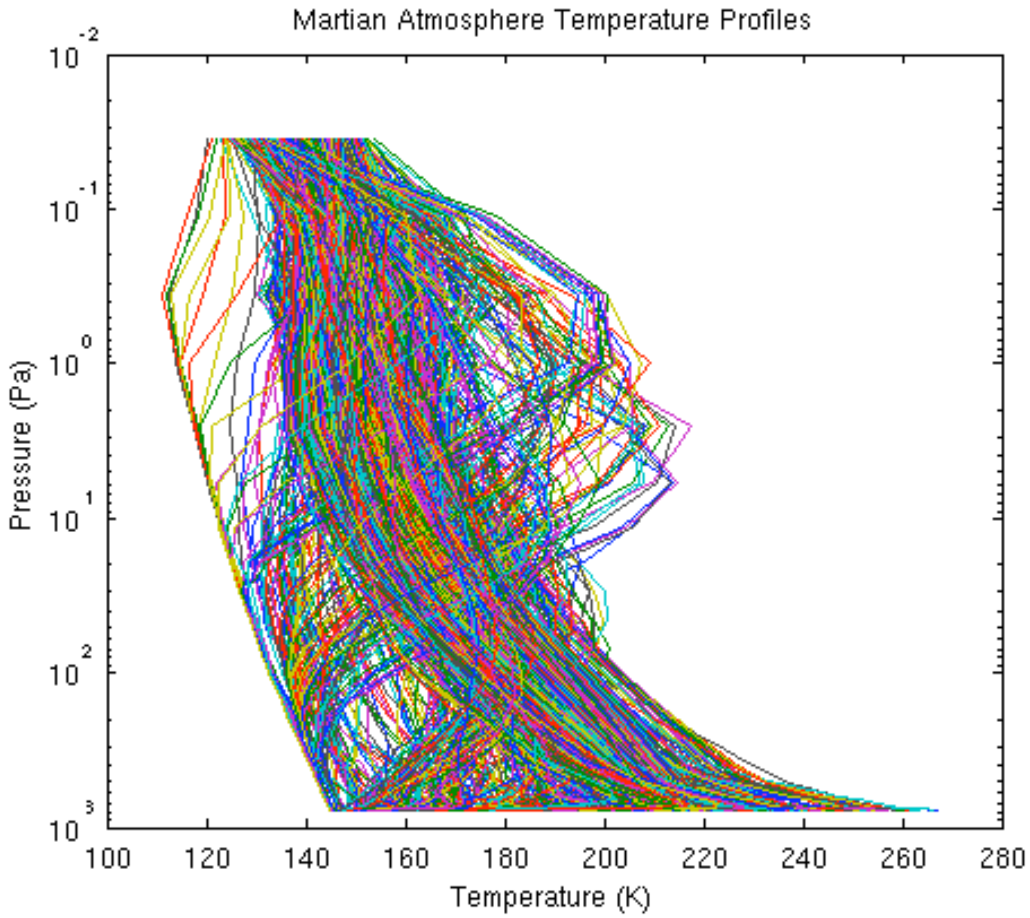
Sample Data: Vertical Temperature Profiles on Mars

- K=28

  (number of vertical grid points)

- N>10^5

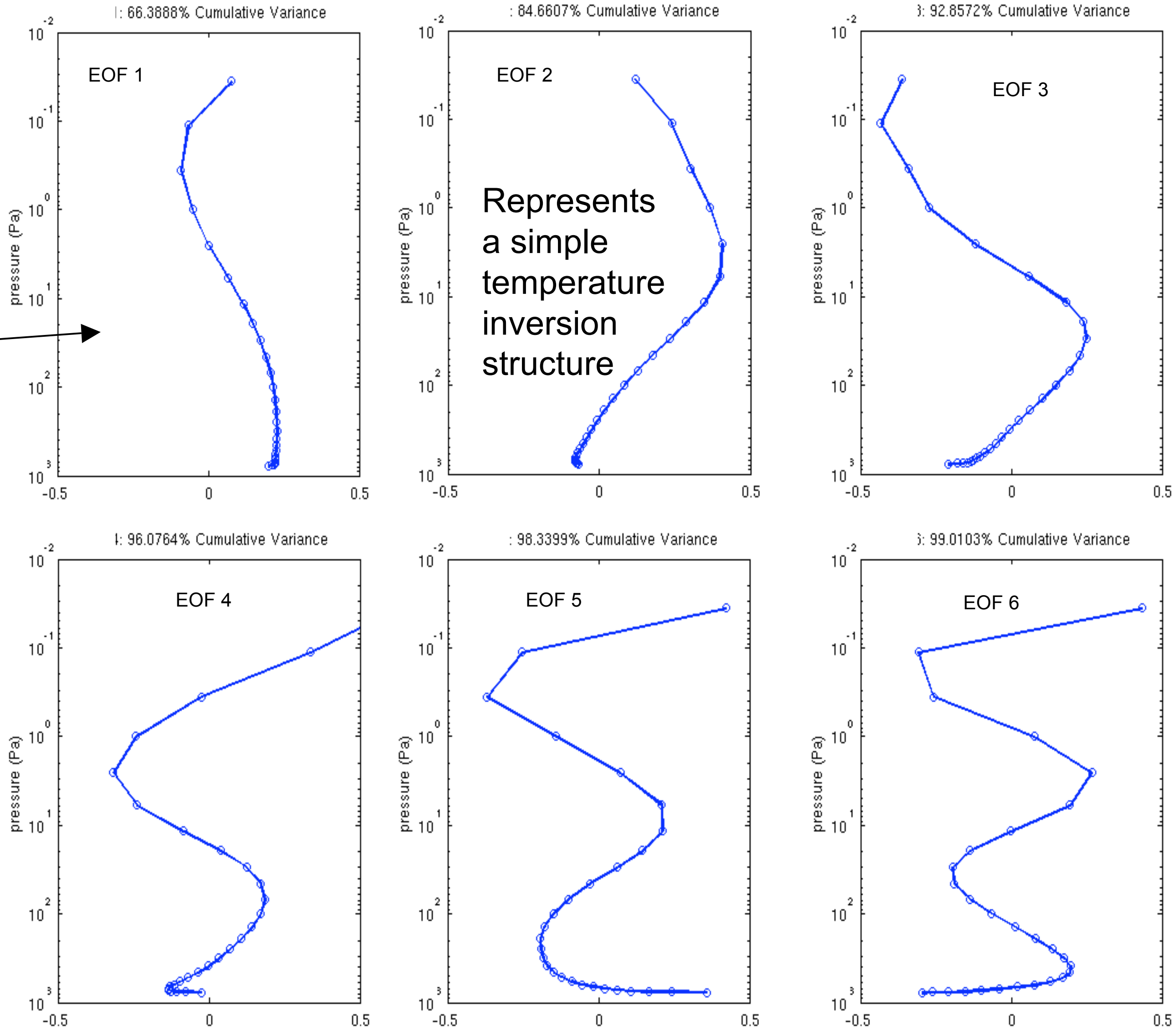
  (number of profiles in space/time)



First 6 EOFs of Vertical Temperature Profiles

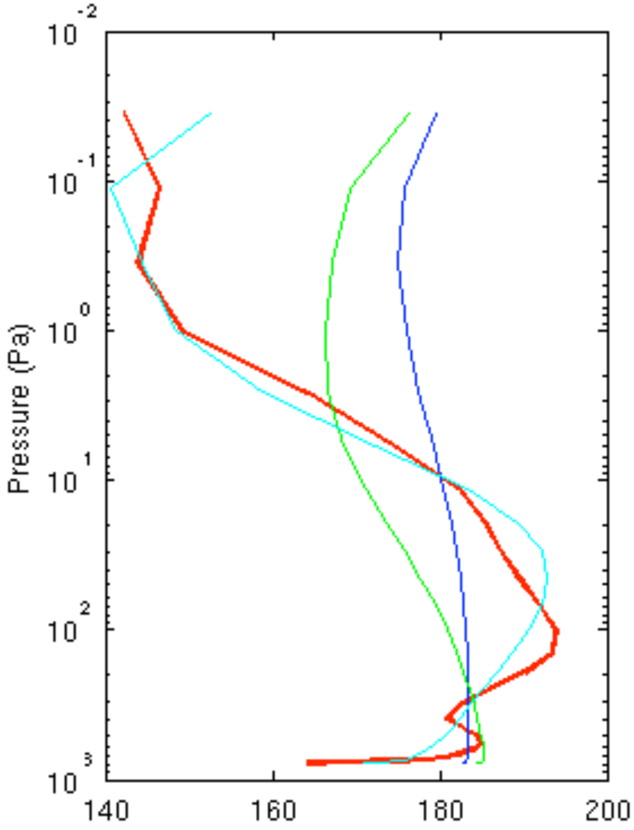
There are 28 EOFs total. We keep only the first 6 because they explain 99% of the variance in the data.

Represents physically the decrease in T with height



Represents a simple temperature inversion structure

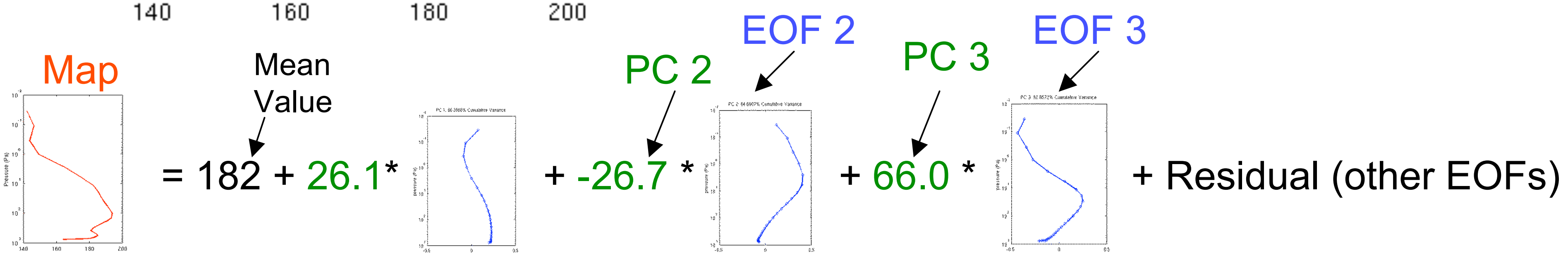
Representing a profile in terms of EOFs



- Red: a temperature profile on a given day.
- Blue: representation using first EOF.
- Green: representation using 2 EOFs.
- Cyan: representation using 3 EOFs.

Notes:

- With 3 EOFs (out of 28), we capture most of the variability (wiggles) in the temperature profile.
- Here, EOF 3 represented most of the variability. Over all profiles, EOF 1 will typically have the highest amplitude.



Map

Mean Value

PC 2

EOF 2

PC 3

EOF 3

= 182 + 26.1* + -26.7 * + 66.0 * + Residual (other EOFs)

Applications of PCA

- To search meteorological data for statistical patterns in space and time (and sometimes provide physical interpretation, such as the North Atlantic Oscillation).

- To represent datasets more efficiently by describing only the most important modes of variability.