Predictions using fuzzy metrics-based aggregation of climate models

Barnabas Bede
based on joint work with G. Levy, L. Takata-Gomes, B. Liepert, A. Geiss

Digipen Institute of Technology
Redmond, WA

US CLIVAR Summit, 2014

Problem formulation

- Data acquisition tools and climate models have developed considerably in recent years
- A variety of predictions is available to stake holders and the public at large.
- This leads to a typical data analysis problem, i.e., the multitude of data available cannot be efficiently processed, used, or merged in the decision making process.
- Is there a new way to create aggregated projections?
- Can we cluster results of climate models?
- Can we represent the data in a more hierarchical manner?

Discussion

- A multi-model ensemble is a correct approach? Do we calculate the average of comparable quantities?
- Can we improve predictions if we devise a better averaging procedure?
- Can we improve predictions by clustering model outcomes?
- Uncertianty quantification should take into account non-statistical uncertianties, these being uncertainties that are due to the lack of our knowledge (epistemic) rather than measurement errors.
- What areas of science can bring improvements in this direction?
- Can we use expert knowledge and modeling in conjunction? Can we incorporate expert knowledge in the models in a mathematically correct way?

Modeling non-statistical uncertainty

- There are several types of uncertainties beyond measurement errors.
- For example we can consider epistemic uncertainties, which come from our limited understanding of processes and phenomena rather than measurement errors.

Accuracy vs. Uncertainty

- A model that has less uncertainty is not necessarily more accurate. It can be actually more realistic if we have a model that allows uncertainty quantification.

Uncertainty quantification and Communication to steak holders

- Fuzzy sets can naturally handle uncertainty and they are very easy to communicate and be used in the decision making process. Decision support system.

Definition of a fuzzy set

Intuitively, a fuzzy set is a set with uncertain boundaries and it can be mathematically represented as a function

$$A : X \to [0, 1],$$

with the following interpretation: $A(x)$ represents the membership grade of element x in the fuzzy set A, where $A(x) = 1$ means complete membership of x in the fuzzy set A, $A(x) = 0$ means complete non-membership of x in A, while intermediate values show partial membership of x in A.

Fuzzy sets and linguistic variables

A fuzzy set is able to model linguistic concepts such as young, old; low, medium, high; cold, cool, warm, hot, etc. Also it is able to model non-statistical uncertainties.

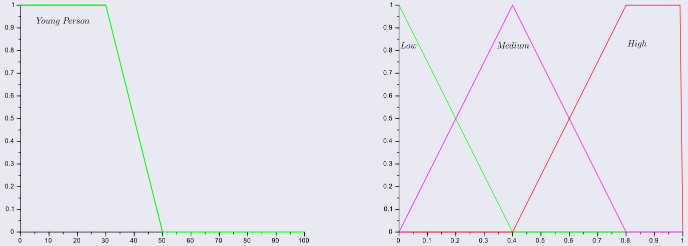


Figure: Examples of linguistic variables "Young Person"(left) and Low, Medium, High model to observation similarity (right), described by fuzzy sets.

Fuzzy rule

A fuzzy rule is able to formulate commonsense knowledge or expert knowledge in mathematical terms. A fuzzy rule can formulate cause effect relationship in the following form:

$$\textbf{If } x \text{ is } A \textbf{ and } y \text{ is } B \textbf{ then } z \text{ is } C,$$

where A, B, C are linguistic variables modeled by fuzzy sets.
Example (similarity in this example means model to observation similarity):

If service is excellent **and** food is delicious **then** tip is generous.

If similarity is high **then** model skill is high.

Fuzzy rule base

One single fuzzy rule is not sufficient for an informed decision. So we need a rule base. A fuzzy rule base can be formulated as

$$\textbf{If } x \text{ is } A_i \textbf{ and } y \text{ is } B_i \textbf{ then } z \text{ is } C_i, i = 1, 2, 3$$

where A_i, B_i, C_i are linguistic variables modeled by fuzzy sets.
Example:

If feature is important **and** similarity is high **then** modell skill is high

If feature is more or less important **and** similarity is average **then** model skill is average

If feature is unimportant **or** similarity is low **then** modell skill is low

where concepts low/average/high can be modeled using fuzzy sets.

Antecedents

Antecedents of fuzzy rules are used for computing the firing levels of various fuzzy rules $\alpha_i = A_i(x) \wedge B_i(y)$.

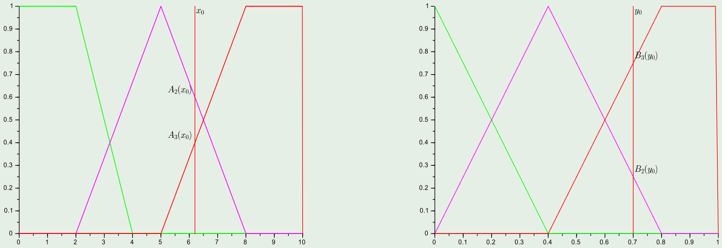


Figure: Examples of linguistic variables "Importance"(left) and Low, Medium, High model to observation similarity (right), described by fuzzy sets.

Consequences

The output of the fuzzy rule base provides a fuzzy inference

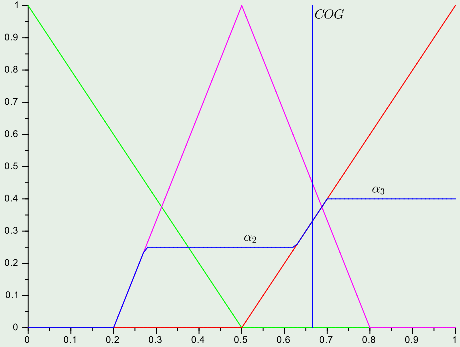
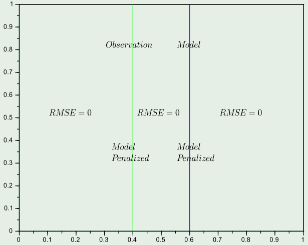


Figure: The model skill as a fuzzy output and its defuzzification

Fuzzy metrics

In order to produce the required similarity measures we need to define fuzzy similarities between climate models. For example we can consider $Sim(A, B) = \frac{V(A \wedge B)}{V(A \vee B)}$.



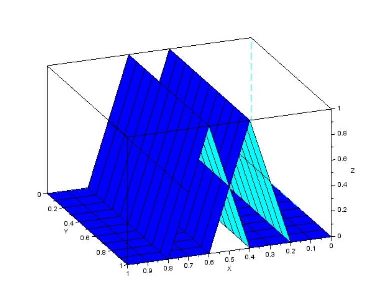


Figure: RMSE penalises models that are not accurate in the location twice. A proposed alternative approach, fuzzification and fuzzy similarities

Observation to Model Comparison

We will compare using RMSE and fuzzy metrics the performance of different models of ice leads

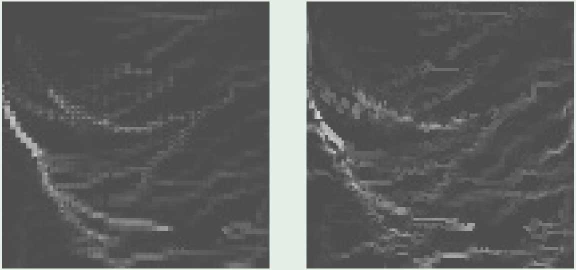
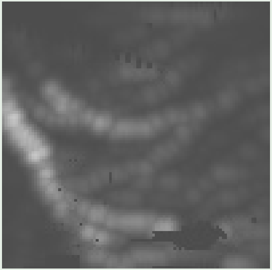


Figure: Illustration of sea-ice (black) and leads (white) of a selected area in the Beaufort Sea. Lower dimensional features (leads) extracted from (left) kinematic interpretation of the RADARSAT Geophysical Processor System (RGPS) data and (right) model simulation (Model 1) on 26 February 2004 (see Levy et al., 2008, 2010).

Observation to Model Comparison

Fuzzifications of the observation and simulation, and their fuzzy similarity $Sim(A, B) = 0.6700731$.



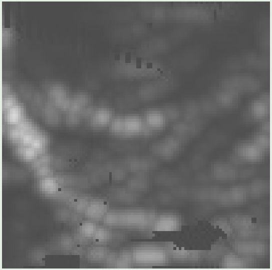


Figure: Fuzzifications of sea-ice observation (left), and fuzzyfication of the sea-ice prediction given by Model 1 (right).

Final Aggregation

The final aggregation of climate models will be an ensemble like prediction that would take into account different model skills as weights for aggregation.

Clustering

How can we use the comparison results in clustering of models?

Learning

Data mining community is nowadays exploiting machine learning techniques. What methodologies from that community could be used for climatology

Predicting the uncertainty

Could we follow the same procedure to predict the uncertainty?

Further mathematical ideas

How can we combine new mathematical ideas at the boundaries of traditional disciplines in order to improve climate models?

Learning and adaptivity

Can we use a more data driven approach? Should (could) we use learning algorithms, Neuro-fuzzy approach?

Other questions

....?