Challenges in Evaluating Lower-Dimensional Features

Challenges in evaluating & utilizing innovations, increments and residuals

- We don't always know how to
- We sometime do but cannot communicate it to other disciplines, stakeholders (e.g., uncertainty)
- "lower dimensional information" (Levy et al., 2010) in climate models is information contained within a bulk simulation even when not directly observed or represented by model variables
- LDI is often misinterpreted, improperly evaluated, and mostly ignored in climate impact analyses.
- Climate models that resolve LDI and the associated physics well, yet imprecisely are penalized by traditional evaluation metrics
- Methods to quantify and measure uncertainty, subjective reasoning, agreement, incremental improvement, etc., along with theory do exist in other fields (e.g., image analysis, human behavior, etc.

Challenges (specific for assimilation)

- The analyzed field, based on the data is not a model realizable state
  - Will produce noise if used to initialize
- Assimilation systems do not have a simple, quantitative measure of representing features contained in bulk simulation (lower dimensional features).
- The observed (remote-sensing) variables do not match the model variables, requiring parameterizations that are not consistent with the data or the physics

Lower-dimensional features

- Features not directly measured/ resolved by observing systems or most models
- Can be related to model resolvable physical or material properties.
- Challenge to assimilate using standard practice of "innovations" (observation increments).
- Challenge to validate (call for fuzzy verification).

Examples (sea-ice)

- Fuzzy metric:

Levy, G., M. Coon, G. Nguyen, and D. Sulsky (2008), Geophys. Res. Lett., doi:10.1029/2008GL035086

- Initialization/assimilation experiment:

Levy, G., M. Coon, G. Nguyen, and D. Sulsky (2010), Geosci. Model Dev., doi:10.5194/gmdd-3-1-2010

http://www.geosci-model-dev-discuss.net/3/517/2010/gmdd-3-517-2010.html

Fuzzy Verification

- A metric to measure model accuracy in representing linear features is used.
- Metric circumvents the requirement that both observations and model variables be commensurate (measured with the same units) by considering features of interest frequencies.
- Substituted in a common skill score

US CLIVAR
Climate Variability & Predictability

Observations vs. Model

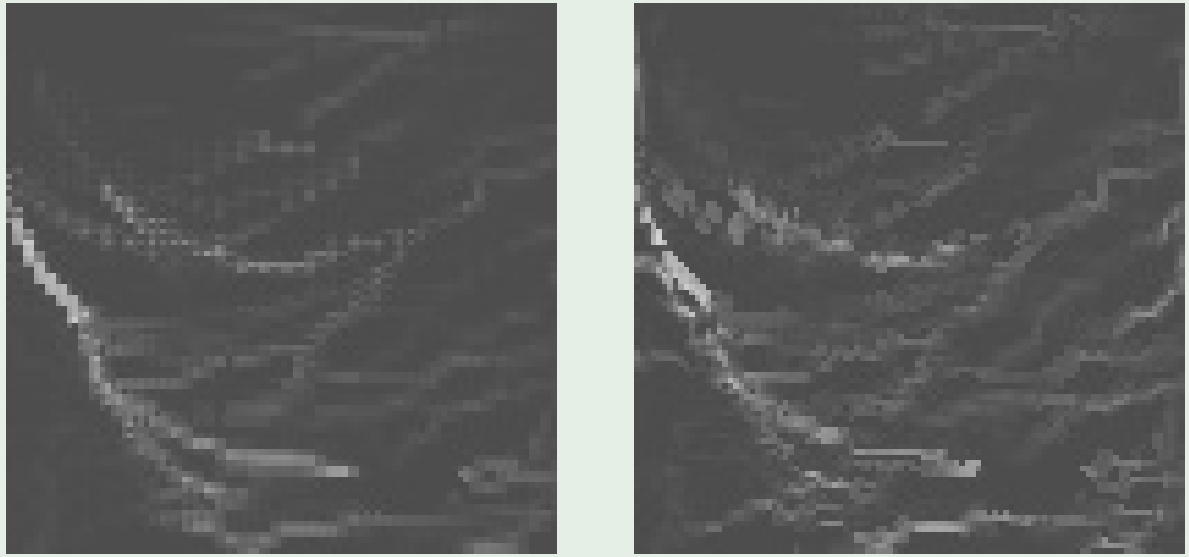


Figure: Illustration of sea-ice (black) and leads (white) of a selected area in the Beaufort Sea. Lower dimensional features (leads) extracted from (left) kinematic interpretation of the RADARSAT Geophysical Processor System (RGPS) data and (right) model simulation (Model 1) on 26 February 2004 (see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

For L features of interest and N spatial segments of the domain

frequency distance function

$$D_{i,j} = \frac{\left(p_{i,j} - o_{i,j}\right)^2}{\left(p_{i,j} + o_{i,j}\right)^2}$$

$P_{i,j}$ and $o_{i,j}$ are predicted & observed feature frequencies,

RMS Index of Agreement

$$I_A = 1 - \sqrt{\left(\sum_{i=1}^{L} w_i \sum_{j=1}^{N} w_j D_{ij}\right) / \left(\sum w\right)}$$

$D_{i,j}$ is a normalized frequency function, w are weights

Define a general skill score

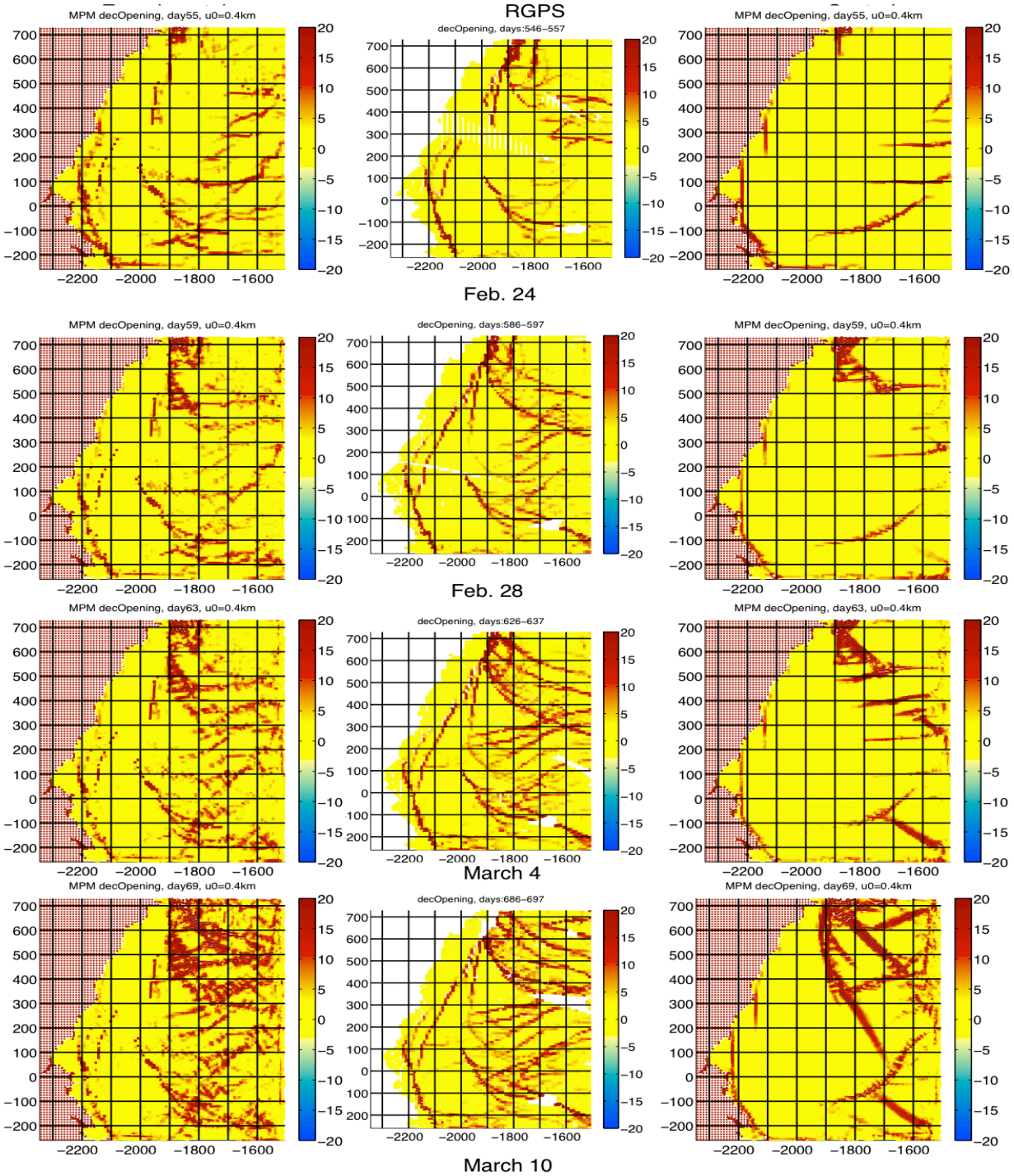
$$SS = \frac{A_f - A_r}{A_p - A_r} = \frac{A_f - A_r}{1 - A_r}$$

Experimental design

- A control run, and an experimental run of sea-ice in Beaufort Sea over 16 days at 10 km resolution
- Initial and boundary values of all model variables and parameters are the same
- Experimental simulation goes through a one-step assimilation during day 54, where information on the pattern of leads from RGPS at 10-km resolution is used to predispose the model

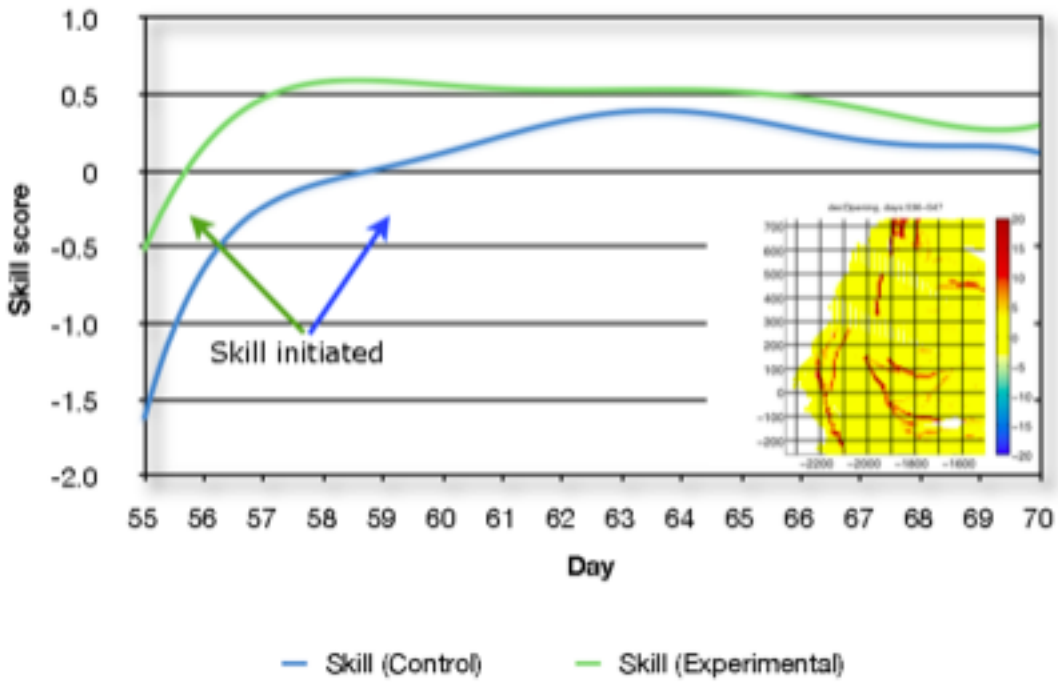
Sixteen day run - B. Sea (Feb.-March, 2004)

- Observations from SAR (RADARSAT 1)
- Model run w/o assimilation
- Model run with assimilation
- Observations from SAR



US CLIVAR
Climate Variability & Predictability

Skill Score



- New assimilation/initialization results in positive prediction skill within one day as opposed to 5 days w/o assimilation
- Skill score significantly improves

Summary/Conclusions

- Much lower dimensional information is contained in bulk simulation that is not specifically resolved by models or observations
- Proper evaluation of such information is challenging, but possible.
- Innovations, increments and residuals need to be differently evaluated in such cases to fully utilize information.
- Methods to quantify and measure uncertainty, subjective reasoning, agreement, incremental improvement, etc., along with theory do exist in other fields and can be adapted

US CLIVAR
Climate Variability & Predictability