Journal of Geophysical Research: Oceans

RESEARCH ARTICLE

10.1002/2015JC010989

Special Section:
Forum for Arctic Modeling and Observing Synthesis (FAMOS): Results and Synthesis of Coordinated Experiments

Skill metrics for evaluation and comparison of sea ice models

Dmitry S. Dukhovskoy[1], Jonathan Ubnoske[1], Edward Blanchard-Wrigglesworth[2], Hannah R. Hiester[1], and Andrey Proshutinsky[3]

[1]Center for Ocean-Atmospheric Prediction Studies, Florida State University, Tallahassee, Florida, USA, [2]Department of Atmospheric Sciences, University of Washington, Seattle, Washington, USA, [3]Department of Physical Oceanography, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts, USA

Numerical Models

Dmitry Dukhovskoy

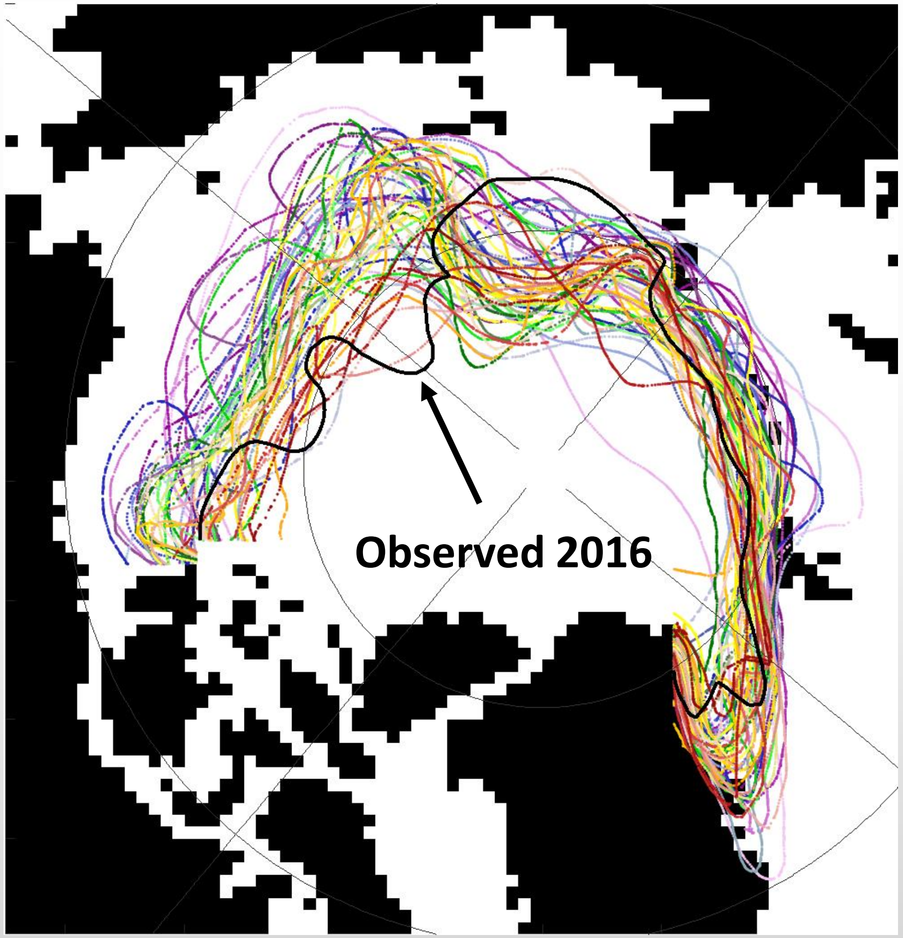
Florida State University
Center for Ocean-Atmospheric Prediction Studies

- **Hana Heister, Eric Chassignet, Steve Morey** implemented skill metrics in real applications
- **Jonathan Ubnoske** provided helpful insights on metric testing and proof of concept.
- The idea of analysis of objective automated metrics for sea ice model comparison was stimulated by ”Minimum Sea Ice Edge Outlooks” at the FAMOS workshops.

This research was funded by **NSF** and a contract M12PC00003 from the U.S. Department of Interior Bureau of Ocean Energy Management (**BOEM**).

Skill Metrics (Sea Ice / Oceanographic Application)

Minimum Sea Ice Edge Contours 2016 (FAMOS "Outlook")



- **Area and Volume Integrated Quantities (sea ice extent, area, volume)**

  Sea ice application: NSIDC Sea Ice Index; Arzel et al., 2006; Hunke and Holland, 2007; Schweiger et al., 2011; Turner et al., 2013; Peterson et al., 2014; SIPN/SIPN2

- **Root Mean Square Error/Deviation (RMSD)**

  Sea ice application: Emery et al., 1991; Cavalieri, 1992; Tietsche et al., 2014

- **Equitable Threat Skill Score (ETSS)**

  Assessment of weather predictions: Mittermaier, 2008

- **Odds Ratio Benefit Skill Score (ORBSS)**

  Assessment of weather predictions: Mittermaier, 2008 ; Gober et al., 2004

Shape Matching/Recognition Algorithms

- **Mean Displacement** (scatter about the centroid of the set of points)

  Shape recognition: Chang et al., 1991; Zhang and Lu, 2004

- **Hausdorff Distance**

  Shape and object recognition: Huttenlocher et al., 1993; Dubuisson and Jain, 1994; Rucklidge, 1997; Daoudi et al., 1999]

Metrics Tested for Model Skill Assessment

2D Application (Contours)

(1) Absolute Deviation (Area-Integrated)

$$D_{AD}(C_i, C_0) = |C_i - C_0|$$

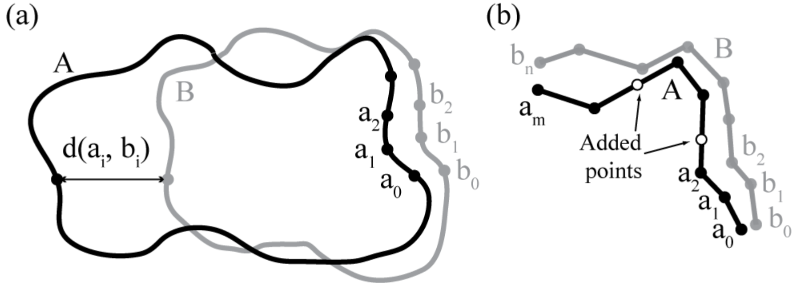
(2) Root Mean Square Deviation

Contour (2D)

$$D_{RMSD}(A,B) = \sqrt{\frac{\sum_{i=1}^{n} [d(a_i, b_i)]^2}{n}}$$

Shape and scalar field (3D)

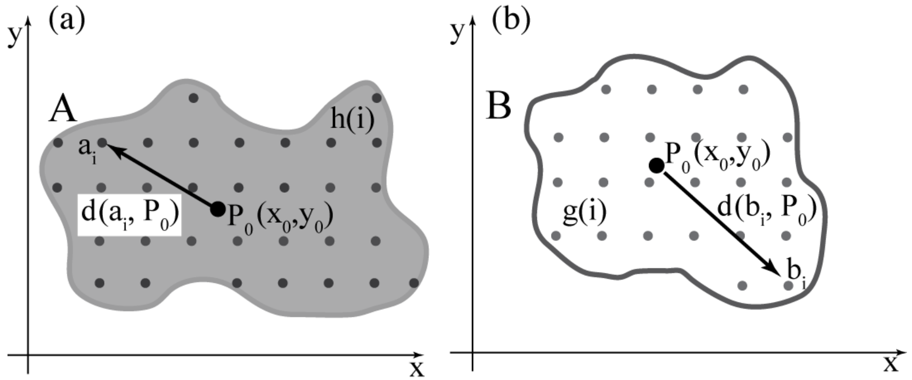
$$D_{RMSD3D} = \frac{\sqrt{\frac{\sum_{j=1}^{m}\sum_{i=1}^{n}[h(i,j)-g(i,j)]^2}{nm}}}{\max(g) - \min(g)}$$



$$(a_i \in A \leftrightarrow b_i \in B)$$

Note the definition of RMSD assumes a correspondence between the points in sets A and B

(3) Mean Displacement



$$\bar{D}_A = \frac{1}{n}\sum_{i=1}^{n} h_i\, d(a_i, P_0),$$

$$\bar{D}_B = \frac{1}{m}\sum_{i=1}^{m} g_i\, d(b_i, P_0),$$

$$D_{MD3D}(A,B) = |\bar{D}_A - \bar{D}_B|$$

(4) Hausdorff Distance

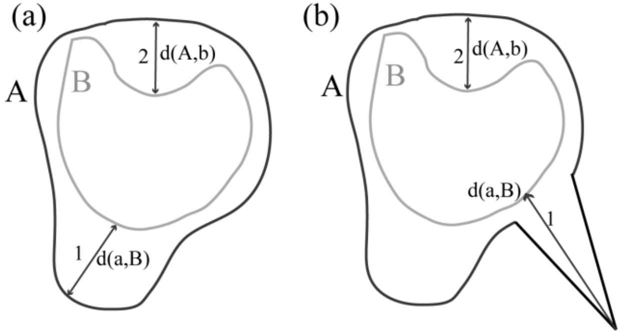
$$D_{HD}(A,B) = \max\left\{\sup_{a\in A} d(a,B),\ \sup_{b\in B} d(A,b)\right\}$$

$$d(a,B) = \inf_{b\in B} d(a,b),\ d(A,b) = \inf_{a\in A} d(a,b)$$

- **Based on a topological shape matching method**
- **Quantifies similarity of shapes/ contours (2D) and spatial patterns (3D)**
- **Can work with any set of points (continuously linked or not, disconnected segments and patches) on any grid (no interpolation required)**

(5) Modified Hausdorff Distance

$$D_{MHD}(A,B) \equiv D(A,B) = \max\left\{\frac{1}{|A|}\sum_{a\in A} d(a,B),\ \frac{1}{|B|}\sum_{b\in B} d(A,b)\right\}$$



Metrics Based on Integrated Characteristics (Sea Ice Extent / Area/ Volume)

Absolute Deviation (absolute difference of the areas inside the contours)

Used in several shape comparison techniques

$$D_{AD}(C_i, C_0) = |C_i - C_0|$$

Sea Ice Extent: Area inside the sea ice edge contour

$$C = \int_{\Omega} dA$$

Sea Ice Area: Area weighted by sea ice concentration inside the sea ice edge contour

$$C_i = \int_{\Omega} g_i(x, y)\, dA,$$

gi – ice concentration inside the contour Ci

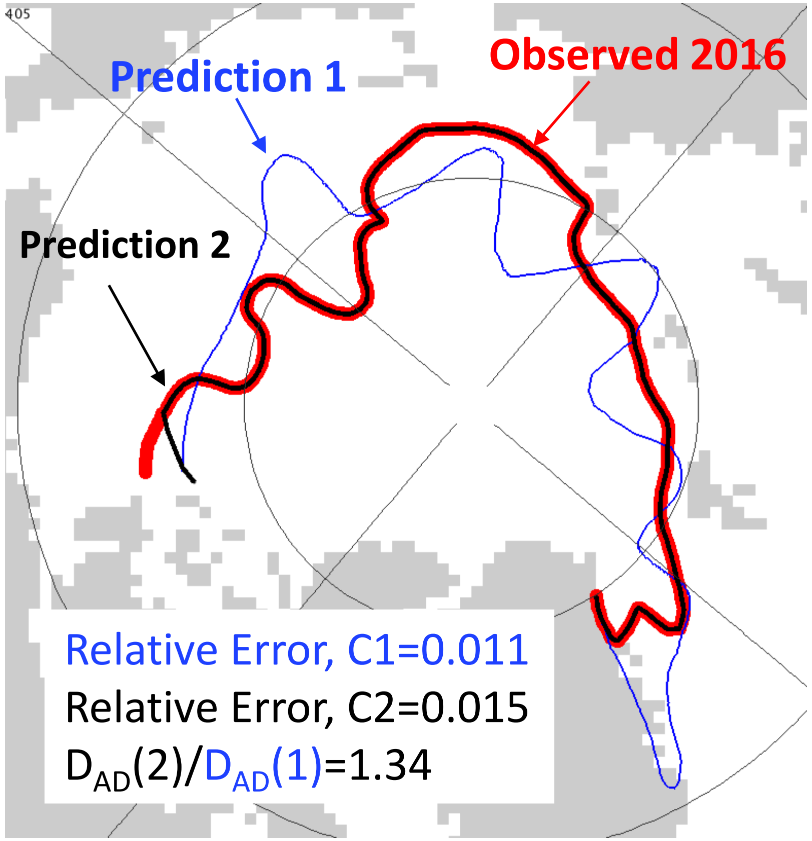
$$C_0 = \int_{\Omega} g_0(x, y)\, dA$$

g0 – ice concentration inside the contour C0

- Used in several shape comparison techniques
- Most of the sea ice assessment and comparison is based on the ice area- and volume-integrated scores
- **Necessary but Not Sufficient Condition for Object Similarity**
- **Do not provide robust skill assessment**

Prediction 1 is higher ranked (better skills) by the AD method because of a closer match to the observed sea ice extent

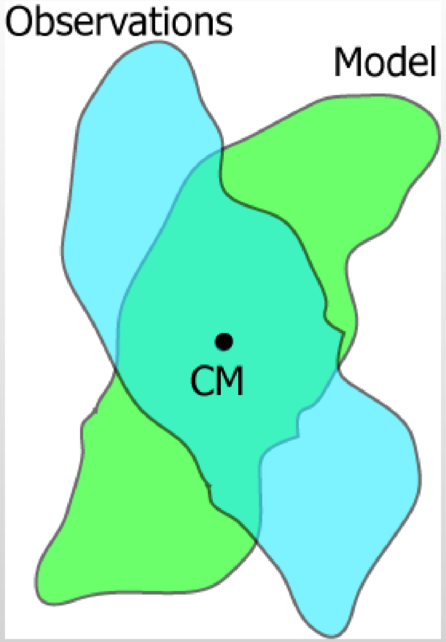
Minimum Sea Ice Edge Contours 2016



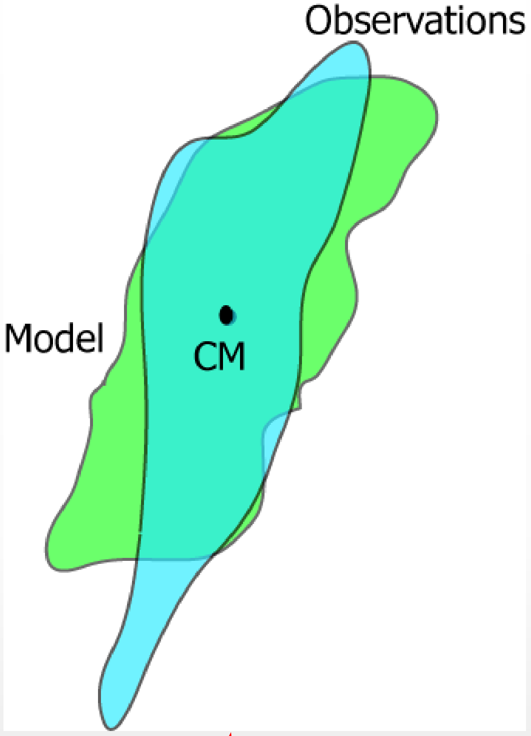
Requirements for an Objective Validation Metric

Attributes that are considered important for a skill metric in the geophysical applications:

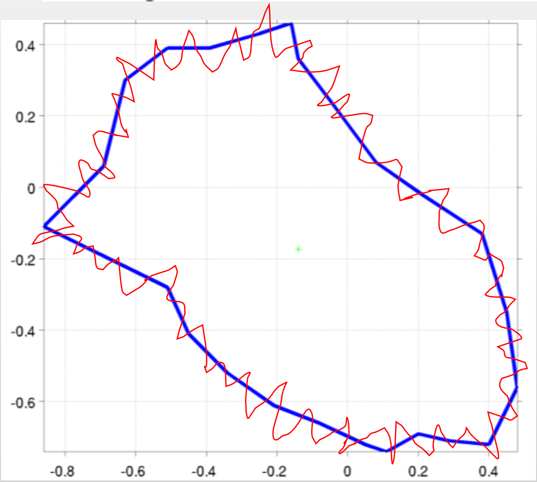
- Scale
- Translation (shift in location)
- Rotation
- Robustness to noise



- **Similar dispersion and center location**
- **Different shape**

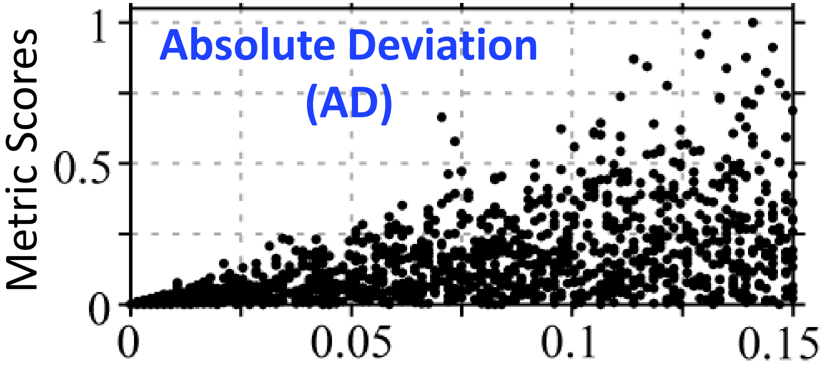


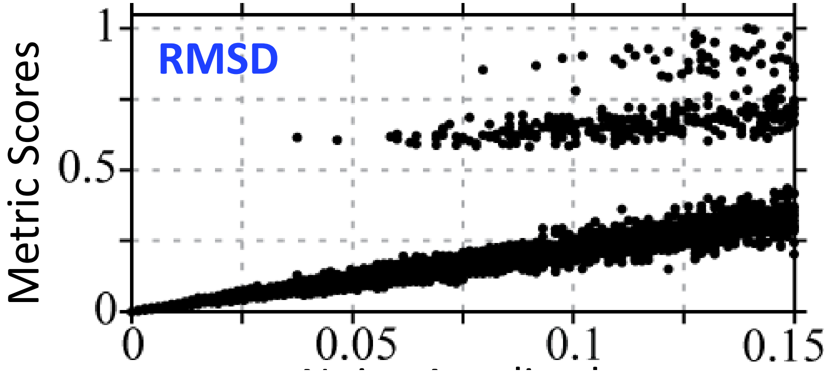
- **Similar dispersion, shape, and center location**
- **Different shape orientation (rotation)**

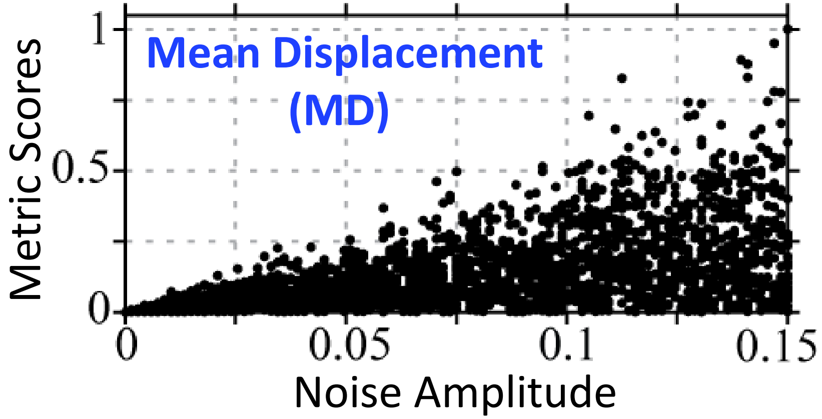


- **Two similar sea ice/ oceanographic contours may disagree on small details**
- **However, as noise increases, the similarity between the objects decays**

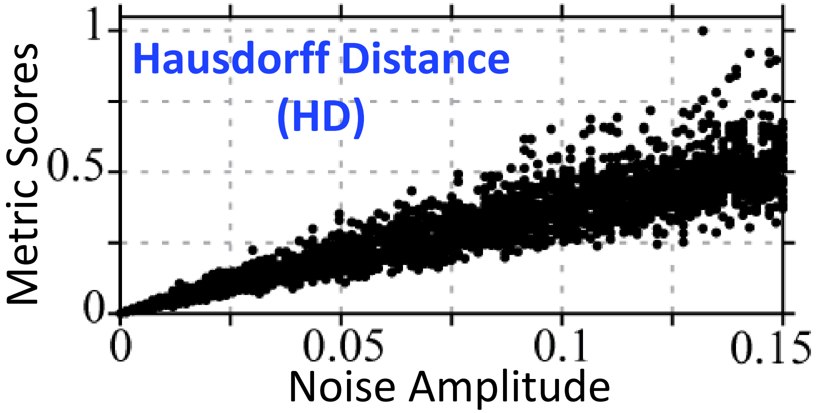
Random Noise Test

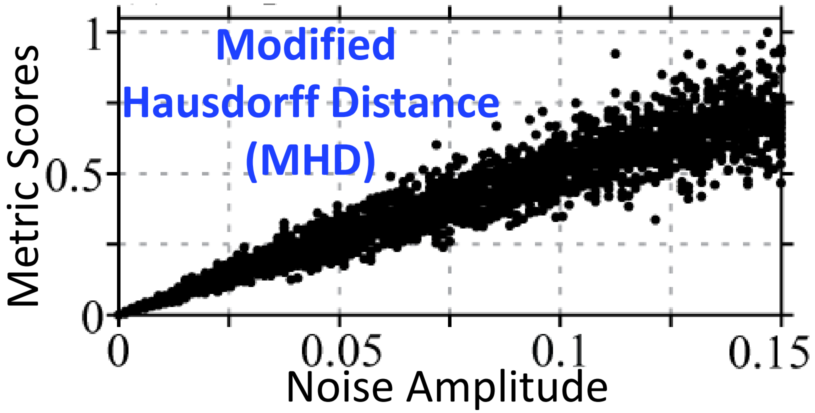






The control shape is perturbed by Gaussian noise with differing noise amplitudes ($\boldsymbol{\sigma}^2$) linearly increasing from 0 to 0.15.





Results:

- **AD, MD failed the test**

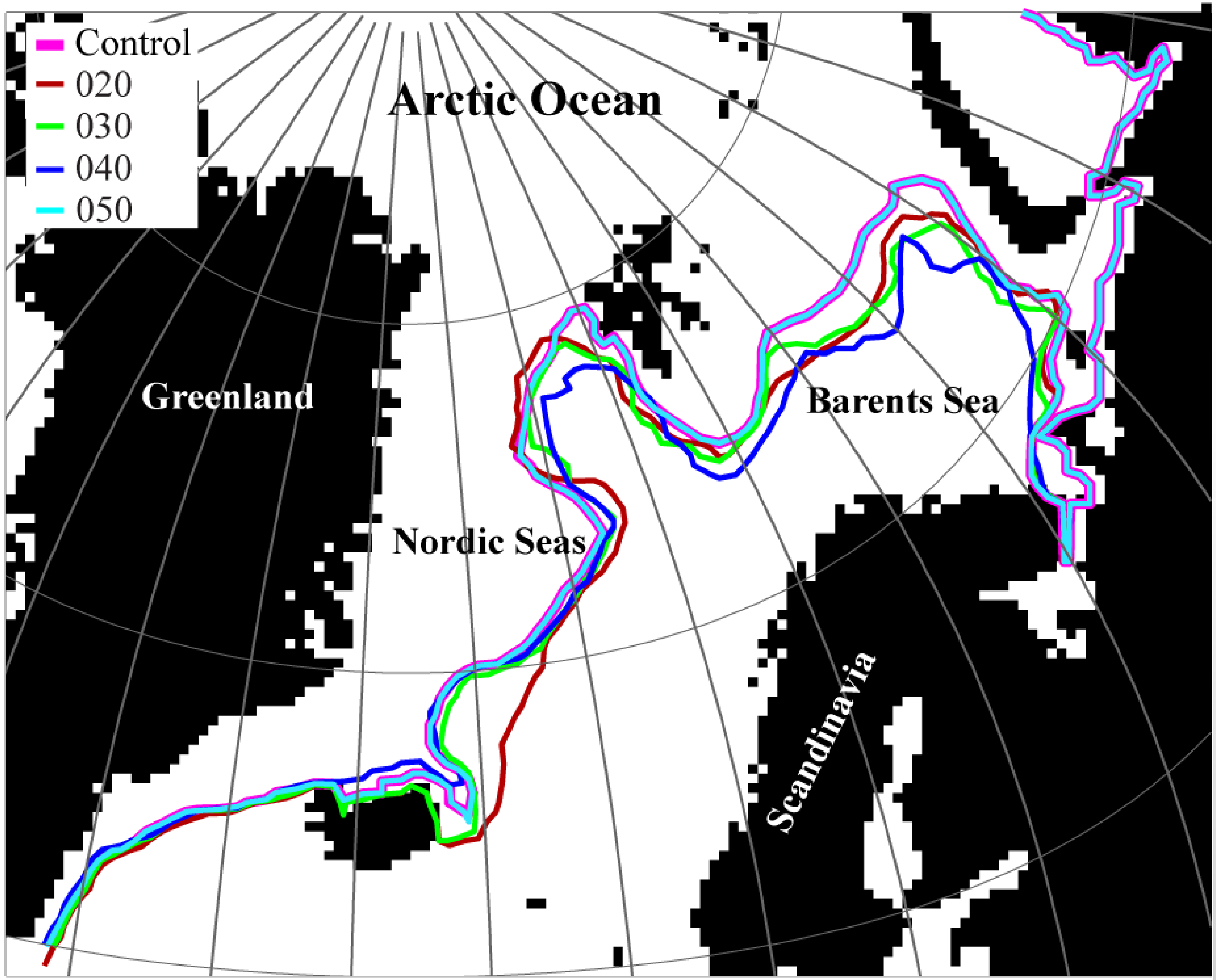
  - The uncertainty range increases rapidly with noise amplitude (no robustness to noise).

  - For large noise amplitude scores can be 0 (perfect match).

- **RMSD partially responsive to noise and demonstrates good performance for noise with small amplitude**

- **HD and MHD demonstrate a very robust response to added noise**

Comparing Sea Ice Edge from 0.08 HYCOM-CICE Experiments



Model Experiments with 0.08 HYCOM-CICE:

Control run – forced with NCEP CFSR

020 – NCAR/DOE NCEPR2

030 – Cross-Calibrated Multi-Platform Ocean Surface Wind Components (CCMPv1.0)

040 – Arctic System Reanalysis (ASR)

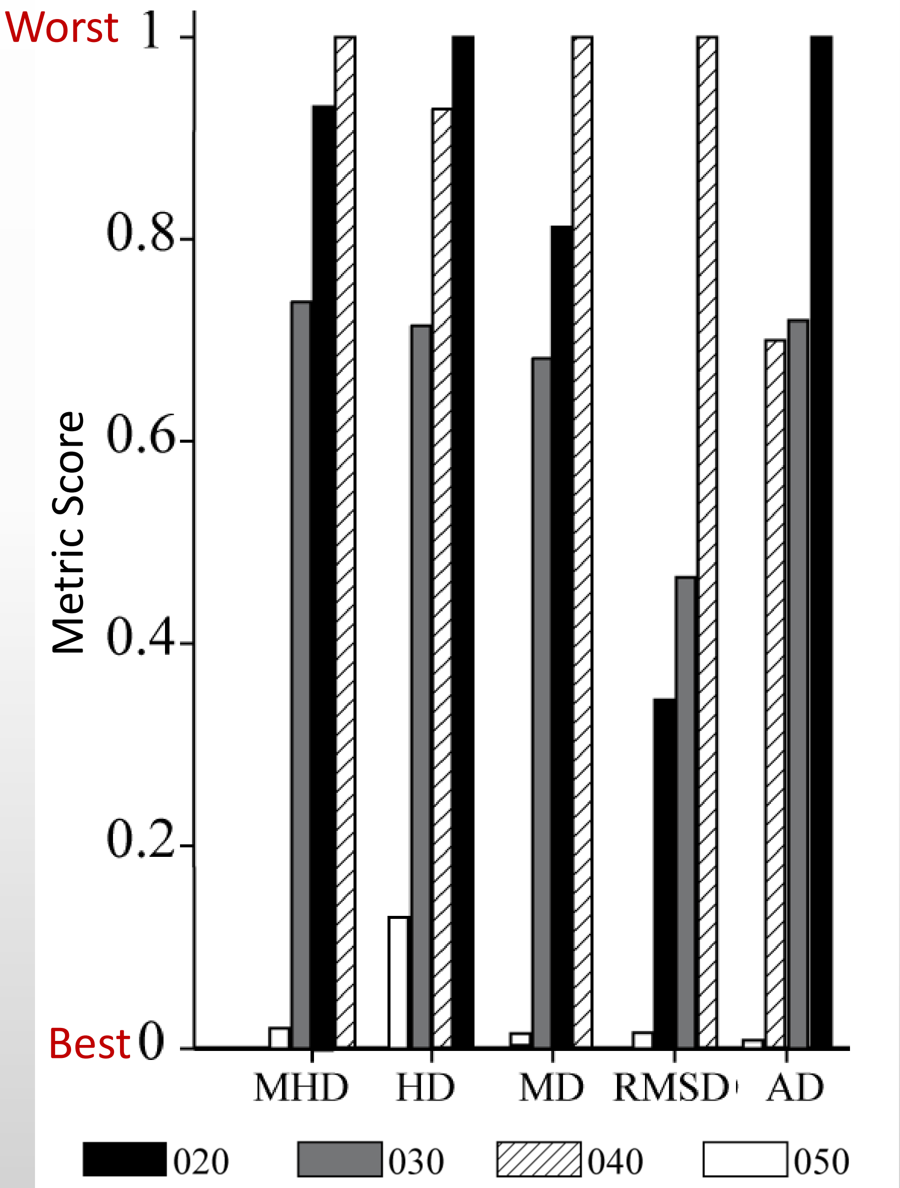
050 – NCEP CFSR (same as Control run) + Greenland runoff (Bamber et al., 2012)

Expected ranking with respect to the Control run:

(1) 050 – the closest resemblance (identical winds)
(2) 030 – based on the visual analysis
(3 & 4) 020 and 040 – based on the visual analysis

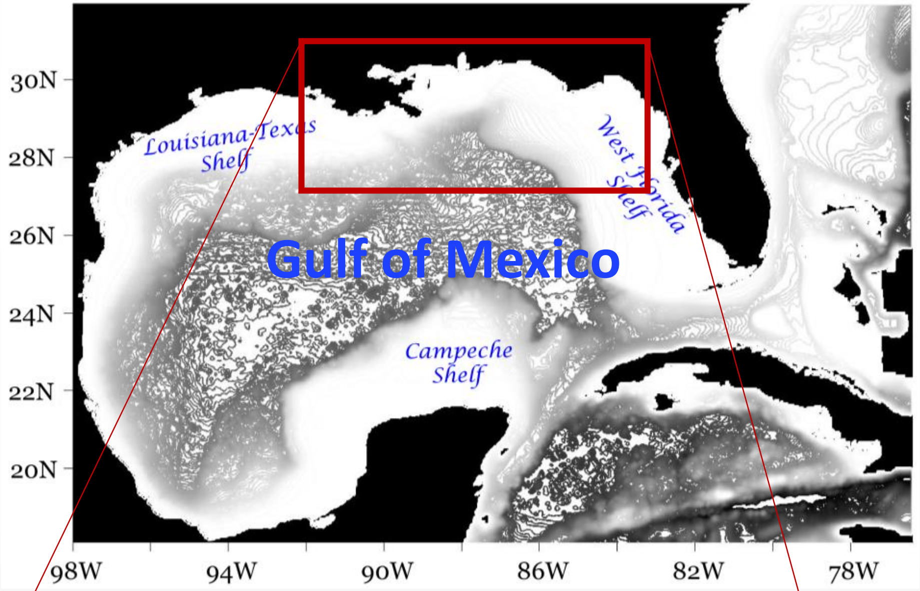
Details of the experiments are in: *Dukhovskoy, D.S. , M.A. Bourassa, G.N. Petersen, and J. Steffen. "Comparison of the ocean surface vector winds from atmospheric reanalysis and scatterometer-based wind products …", JGR, 122(3), 1943-1973, 2017*

Skill Metric Scores on the Ice Edge Contour from the HYCOM-CICE Experiments

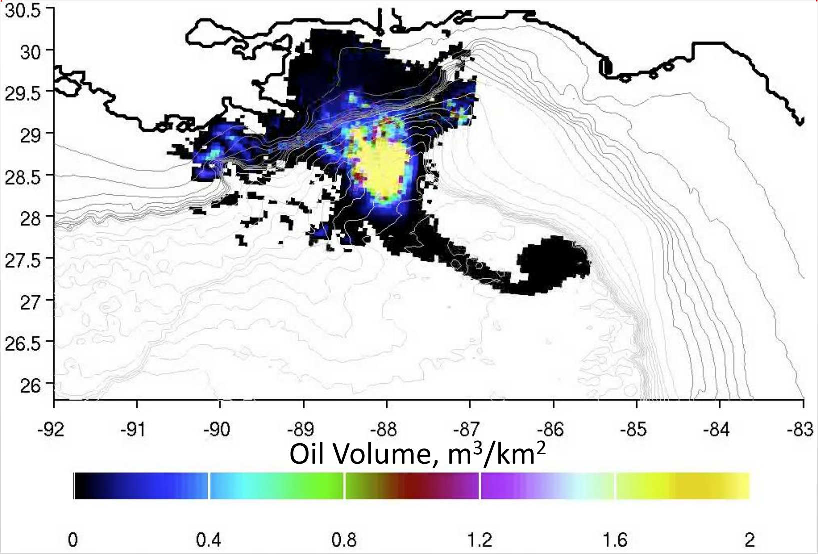


- All metrics correctly rank experiment 050 as the closest simulation to the control run
- Modified Hausdorff Distance (MHD), Hausdorff Distance (HD), and Mean Displacement (MD) identify 030 as the second closest to the control run (in agreement with visual analysis)
- RMSD ranks 020 as the second closest and Absolute Deviation (AD) ranks 040 as the second closest to the Control run (in contradiction to the visual analysis)

MHD Application for Estimating Parameters in a Surface Oil Drift Model



Surface Oil from SAR Images, May 9 2010



The oil particle trajectories are computed as a superposition of advective processes and turbulent diffusion

$$\frac{d\mathbf{x}}{dt} = \mathbf{u}_a(\mathbf{x},t) + \mathbf{u}_d(\mathbf{x},t)$$

- Advective velocity:

$$\mathbf{u}_a = \mathbf{u}_c + C_w \|\mathbf{u}_{10}\| \Theta + C_s \mathbf{u}_s$$

Wind drag coefficient

Unit vector directed at some angle from the wind

Stokes Drift

Ocean surface current

Wind speed

Wave Coefficient

- Surface currents: 1/25° Gulf of Mexico HYbrid Coordinate Ocean Model (HYCOM) Analysis/Reanalysis *(hycom.org/dataserver/goml0pt04):*

Random variable ~N(0,1)

- Turbulent diffusive velocity parameterized as a random fluctuation

$$\mathbf{u}_d = \|\mathbf{u}_d\| \xi \cdot \exp(i2\pi\xi)$$

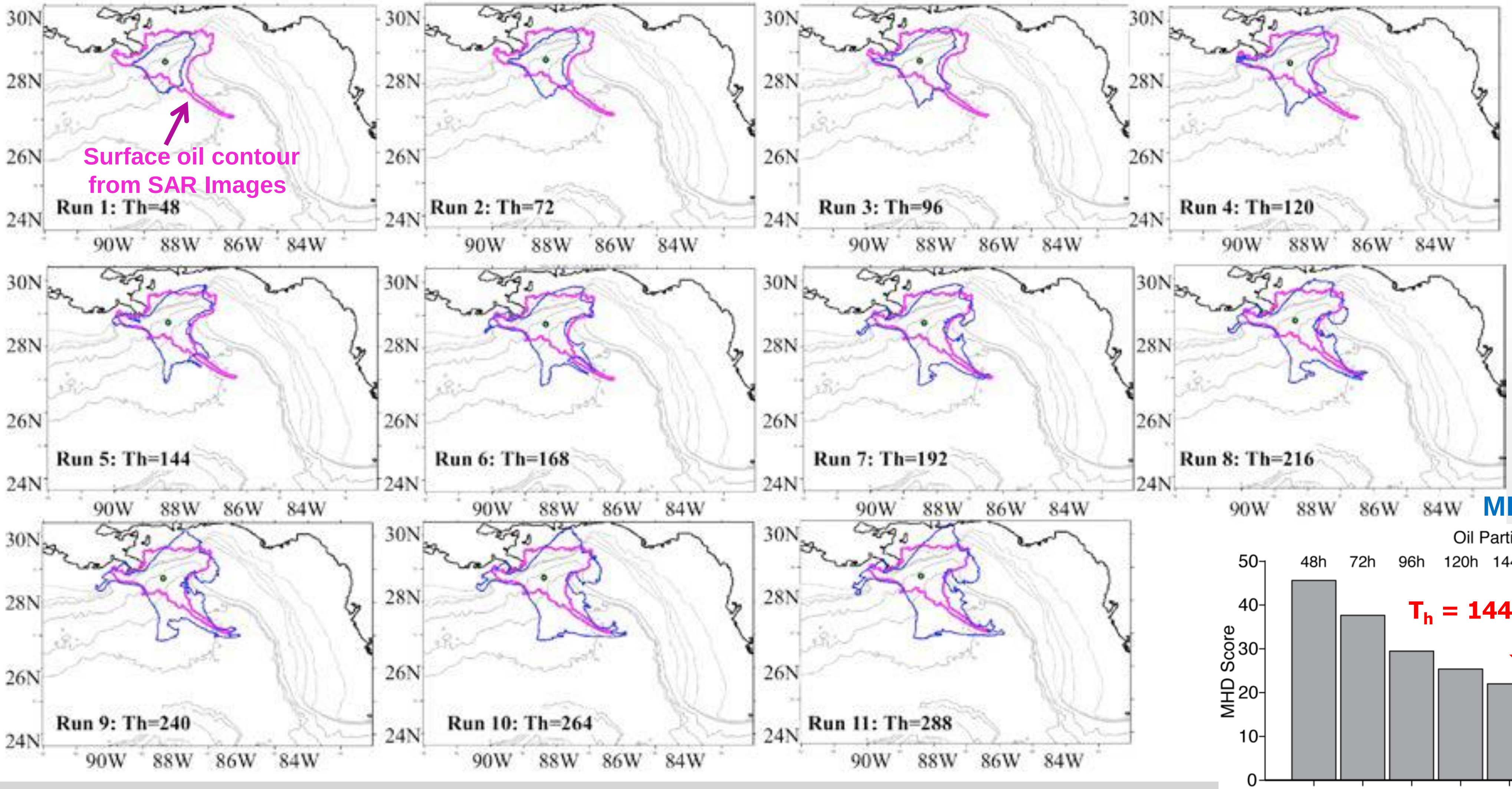
$$\|\mathbf{u}_d\| = \sqrt{\frac{6\nu}{\Delta t}}$$

Free parameters adjusted by minimizing MHD:

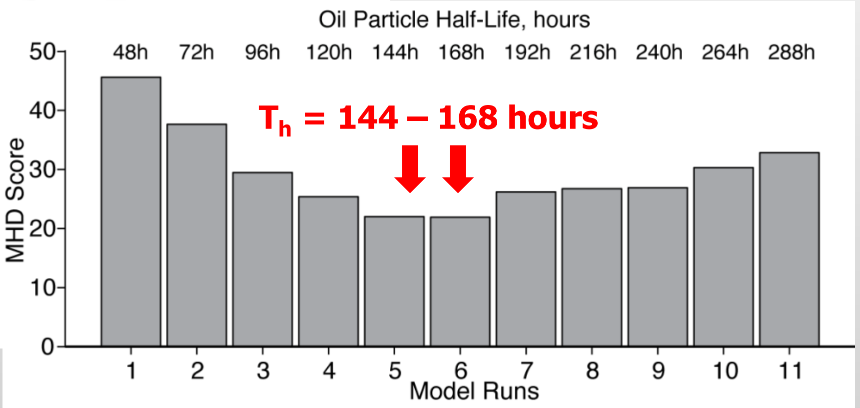
- Half-life: Oil particles are removed randomly based on a prescribed half-life
- Wind drag coefficient: Ranges within 0.02 to 0.045

Estimation of Half-Life (T_h) from SAR Observations and Oil Drift Model

Oil Spill Contours from the Experiments with Different Half-Life vs SAR



MHD Scores



Summary

- The Absolute Deviation metric is the most basic technique for sea ice skill assessment and has demonstrated the weakest performance for the considered application.
  - The major drawback of this technique is the likelihood of equally ranking two geometrically different fields or fields with dissimilar spatial distribution of sea ice characteristics but equal area
- RMSD has shown unsatisfactory performance for the sensitivity tests and sea ice applications.
  - The requirement of point-to-point correspondence between the model and the observations is an inherent weakness of this method.
- The Mean Displacement method is better suited for spatial analysis and it has demonstrated reasonable skills in shape and contour comparison.
  - This metric can provide unrealistic scores when the sea ice fields differ in small details, and this is related to the faulty response to noise.
  - MD can give identical scores to very different contours or patterns when the data points have similar dispersion around the reference point
- Both Hausdorff and Modified Hausdorff Distance methods have demonstrated the best sensitivity and robustness to tested differences and good assessment skills in realistic applications.

Testing MHD: Facial Recognition Test

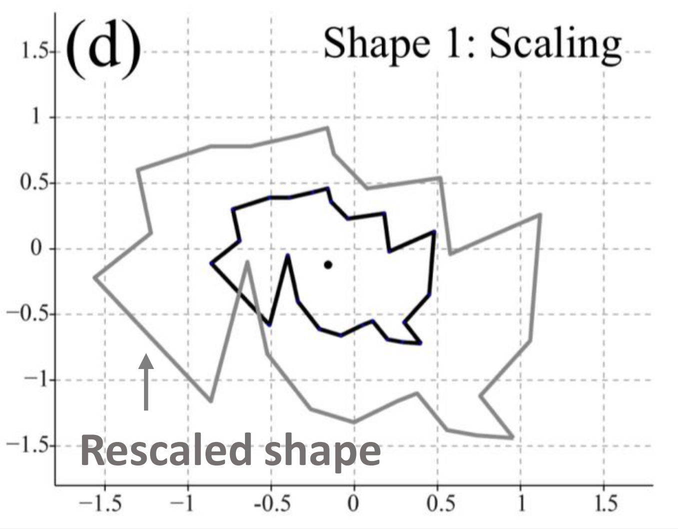
8 "Control Shapes"

MHD Ranking of the Facial Images

Courtesy of J. Ubnoske

Scaling Test

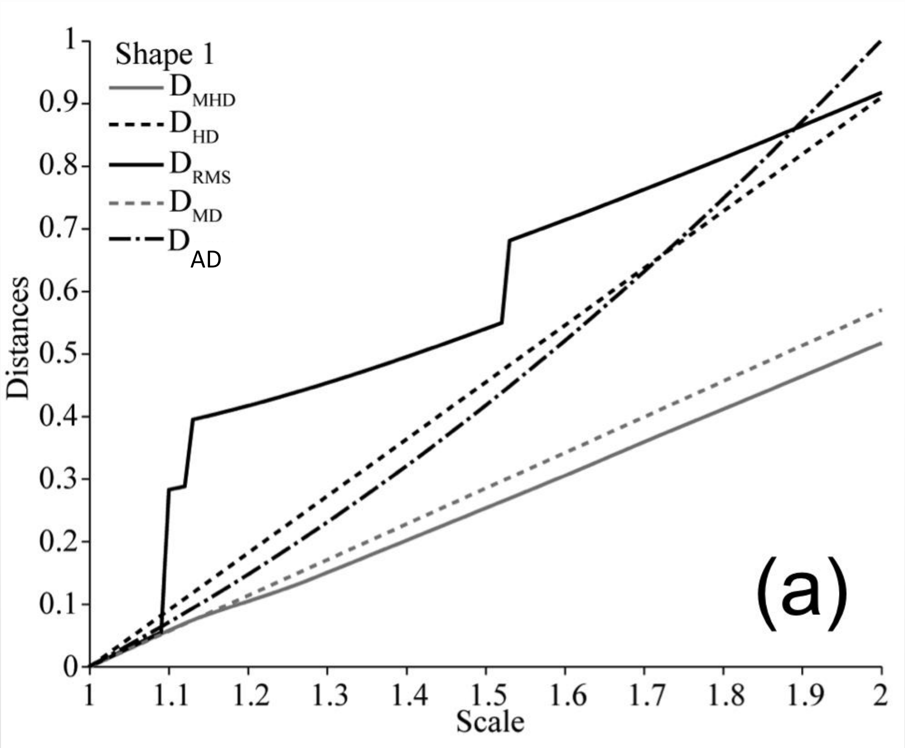
The randomly generated shapes (>100 shapes) undergo a linear increase in size with the maximal scale twice the original size



A continuous deformation that performs rescaling is

$$f_t(\mathbf{x}) = (1 + kt)\mathbf{x}$$

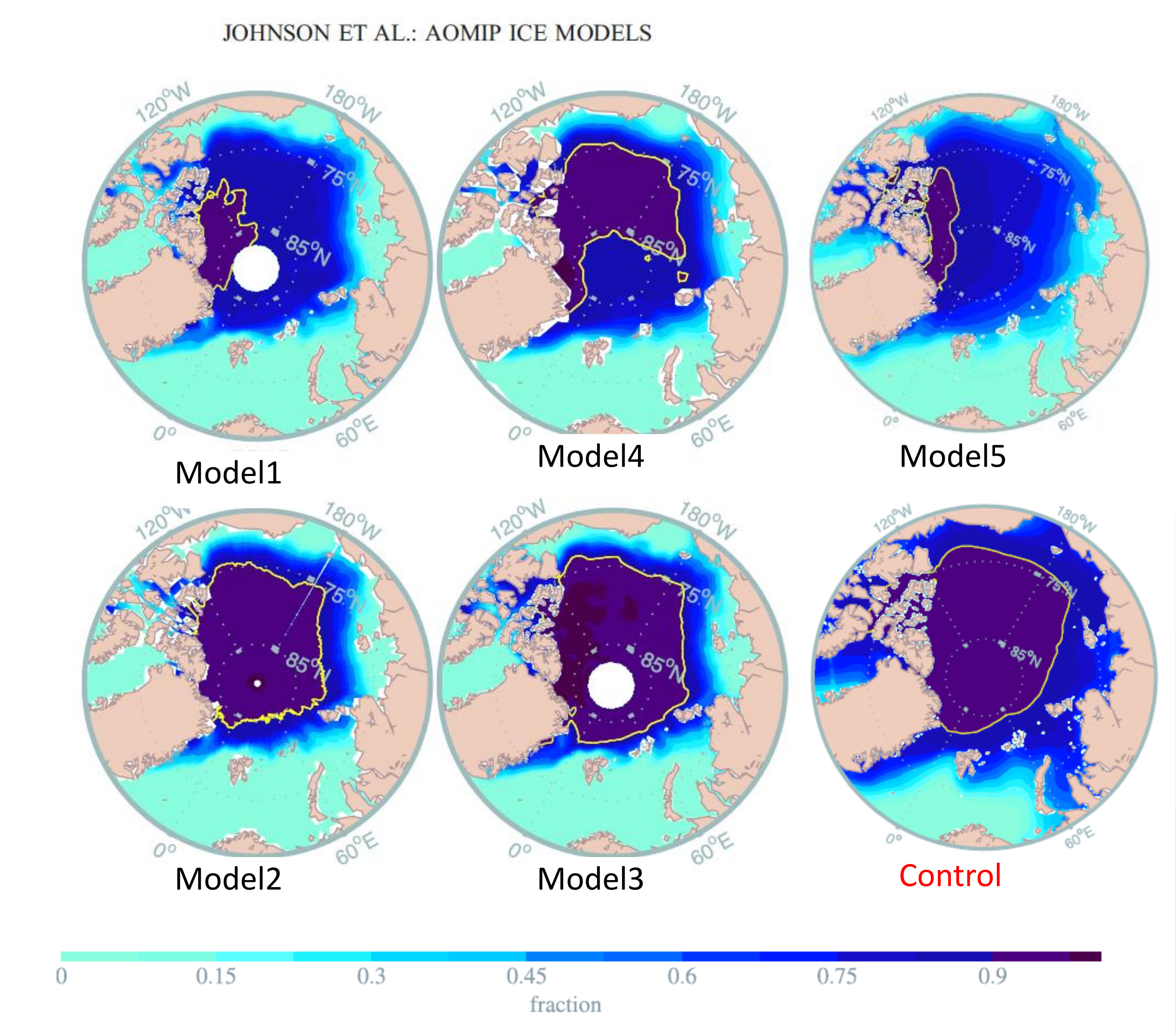
Metric Scores versus Scale



Results:

- **Each metric exhibits the desired behavior of being strictly increasing.**
- **Although RMSD is generally responsive to the scaling, it is not responsive in a robust way.**

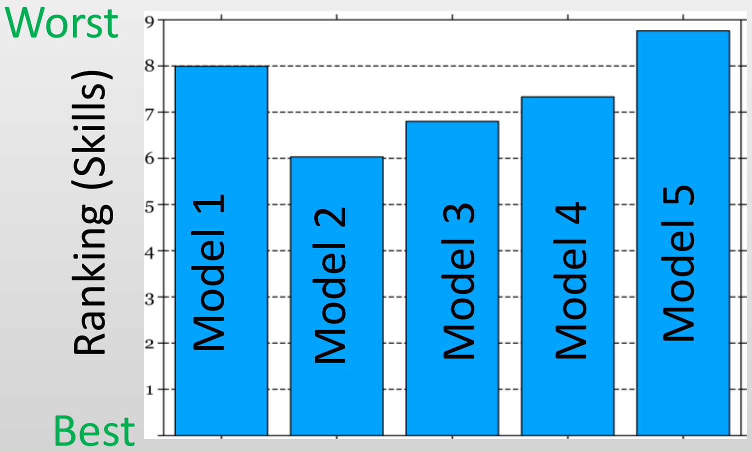
Model Skill Assessment: Quantification of Similarity/Dissimilarity



Skill assessment can be viewed as automated objective way to identify and quantify dissimilarity between the object

Quantitative assessment:

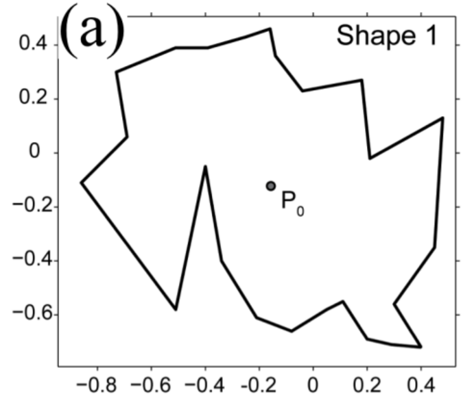
- Model-model or model-data comparison
- Tracking divergence of numerical solution in time
- Assessment of model sensitivity
- Model adjustment



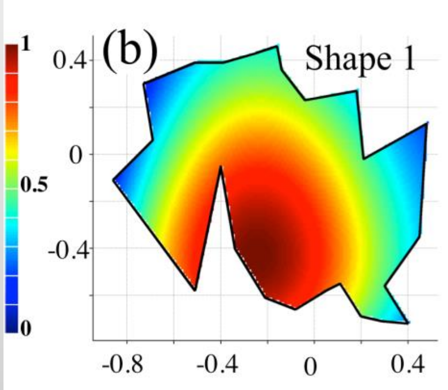
Sensitivity and Robustness Testing of the Skill Metrics

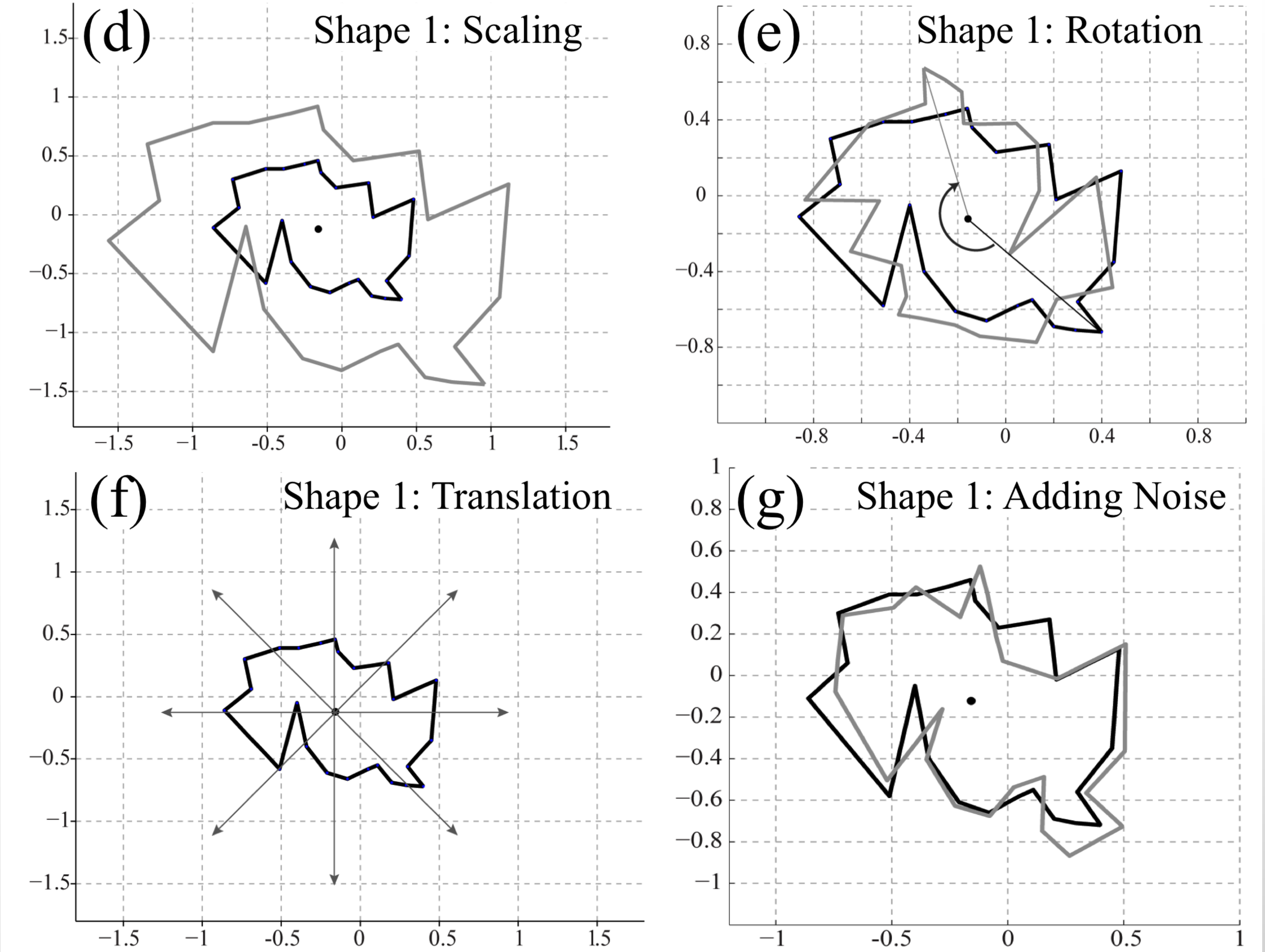
Testes were performed for 100 randomly generated shapes

Randomly Generated Shape



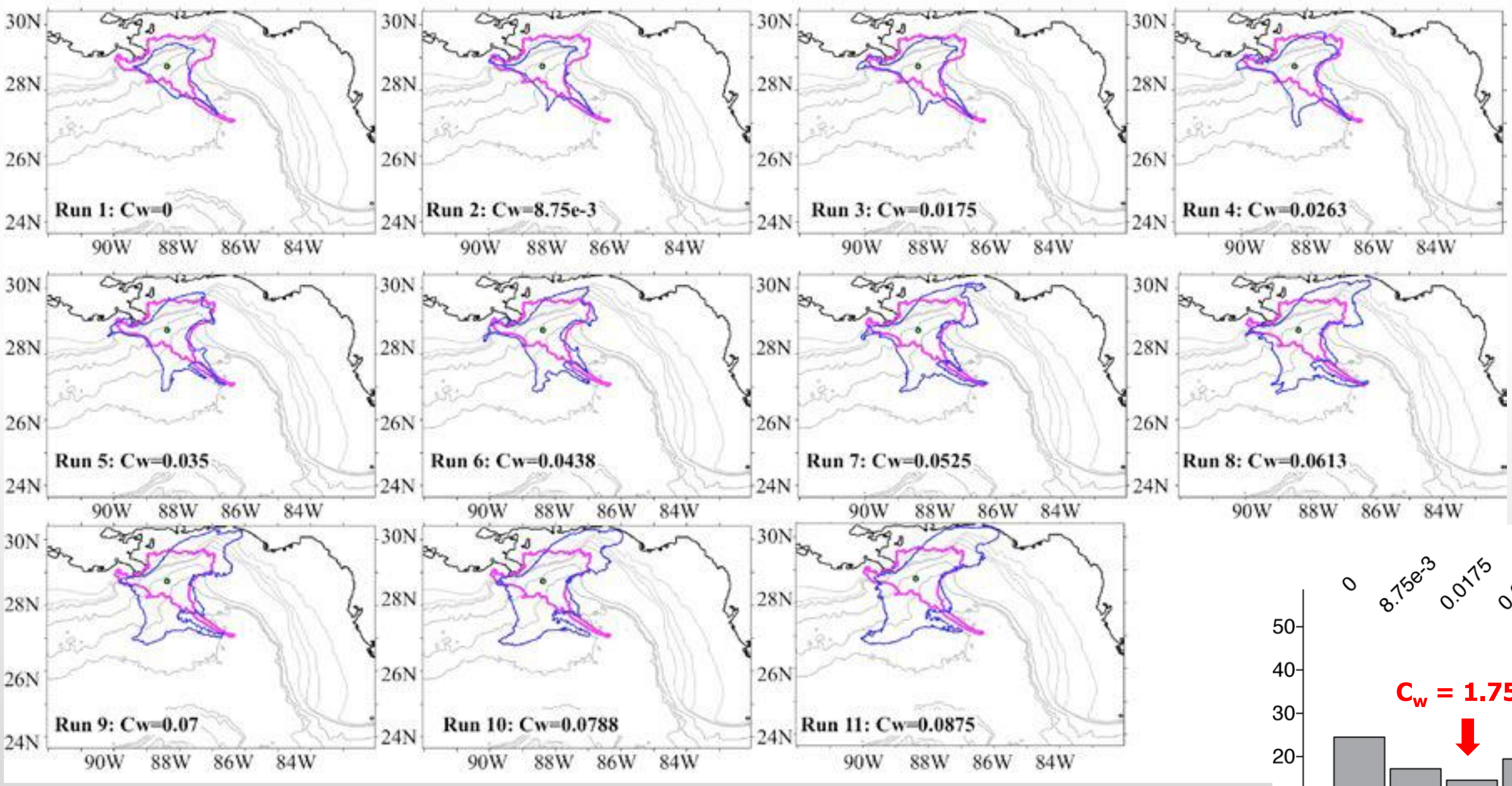
Randomly Generated Shape with a concentration field inside



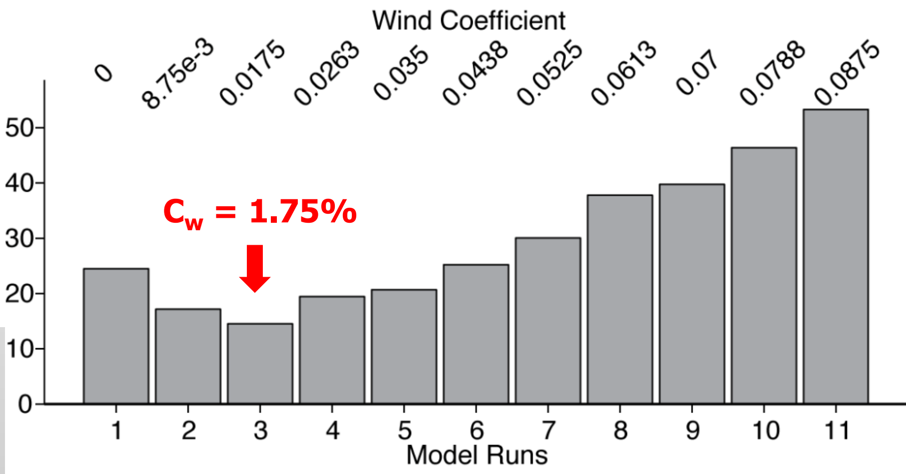


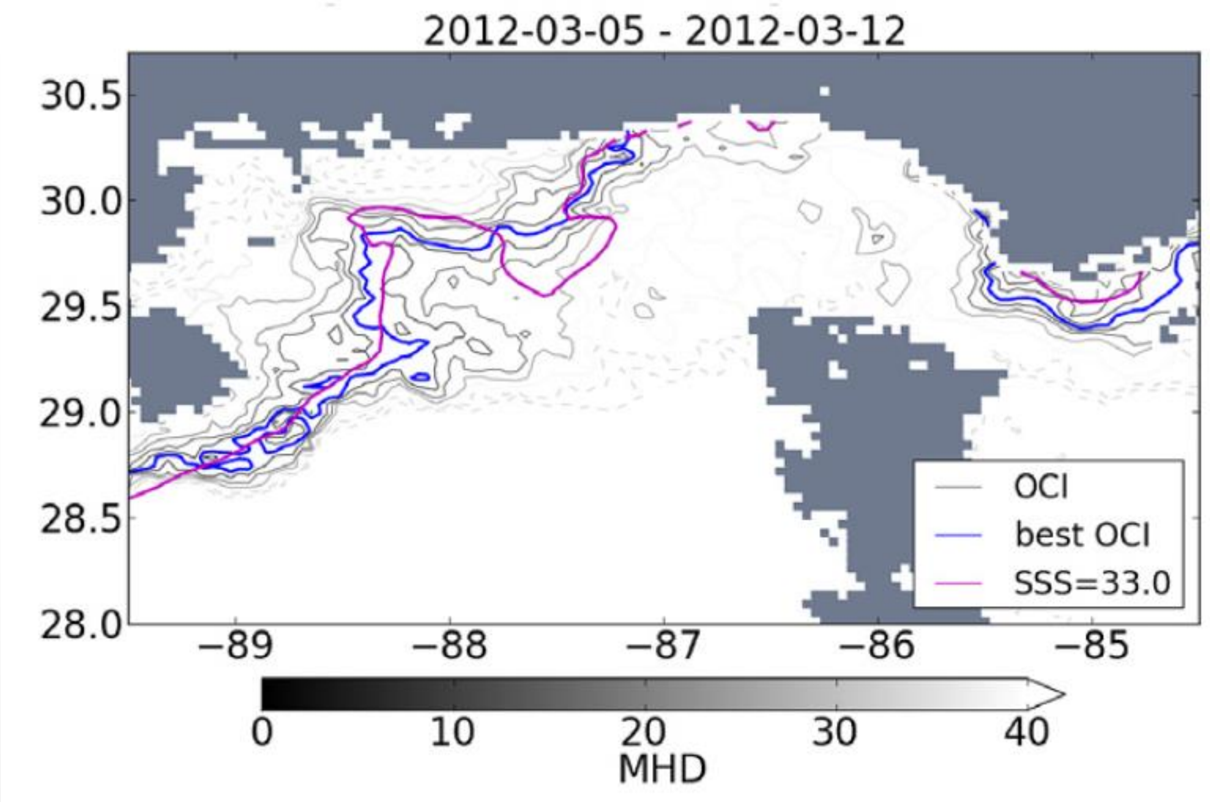
Estimation of Wind Drag Coefficient (C_w) from SAR Observations and Oil Drift Model

Oil Spill Contours from the Experiments with Different Wind Drag vs SAR



MHD Scores

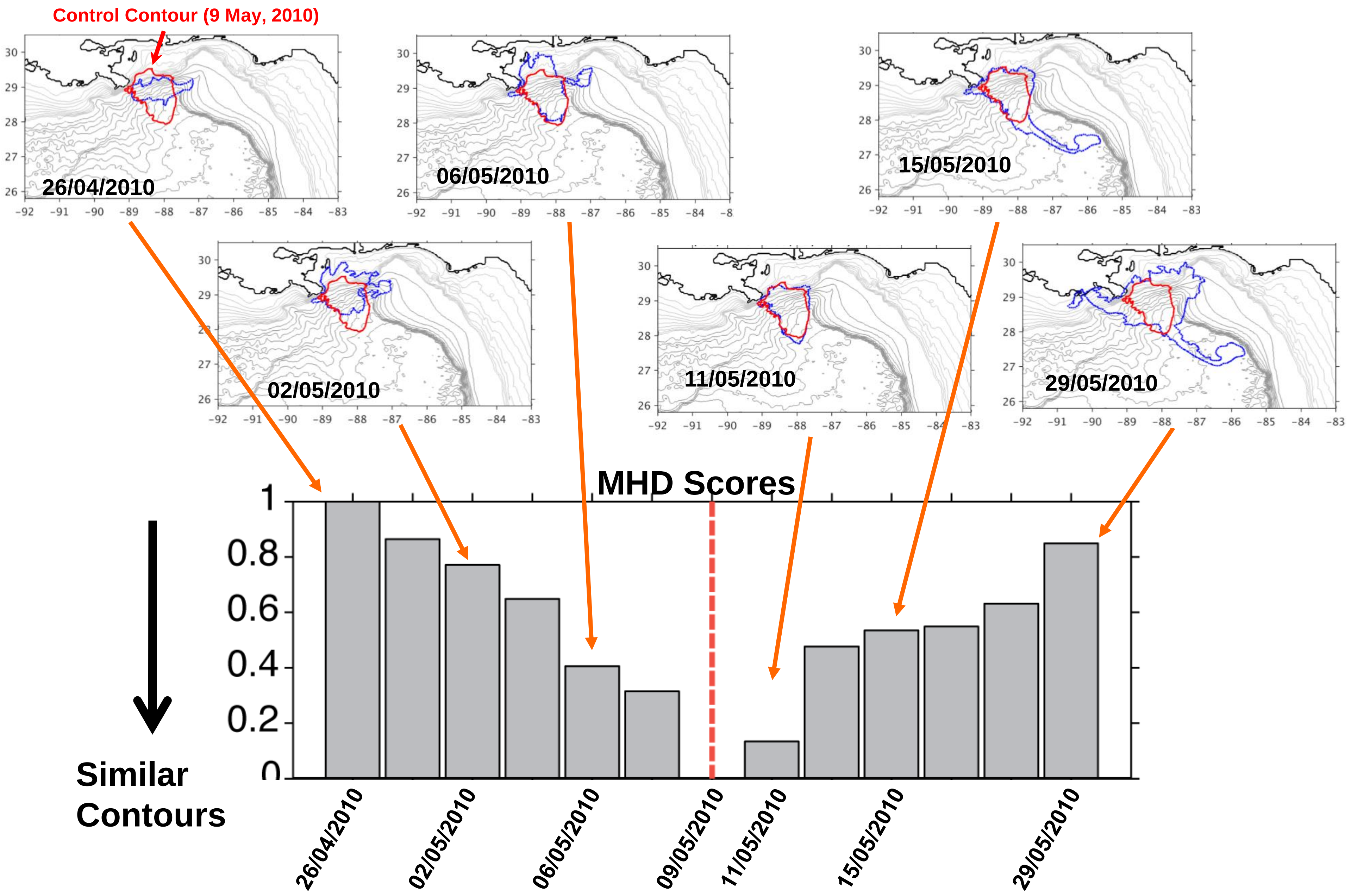




3.4.3. *Application of the Modified Hausdorff Distance*

The similarity between contours of OCI from the satellite data and contours of SSS from each of the models is quantified by calculating the MHD. An example illustrating application of the MHD to compare SSS and OCI contours is shown in Fig. 6. In this example, the SSS = 33 contour from a particular 8-day average salinity field from the DSC-ROMS simulation processed as described in Section 3.4.2 is shown by the magenta line. Contours for several satellite OCI values are overlain in gray, with the shading of those contours indicating the MHD computed between them and the model SSS = 33 contour. (OCI contours that correspond to an MHD > 40 are dashed.) The OCI contour (OCI = 1.65) that most closely matches the model SSS = 33 contour determined by the lowest (best) MHD score (12.20) is shown by the blue line.

Oil Volume Contours Compared to the Contour on 9 May, 2010



Journal of Geophysical Research: Oceans

RESEARCH ARTICLE
10.1002/2015JC010989

Special Section:
Forum for Arctic Modeling and Observing Synthesis (FAMOS): Results and Synthesis of Coordinated Experiments

Skill metrics for evaluation and comparison of sea ice models

Dmitry S. Dukhovskoy[1], Jonathan Ubnoske[1], Edward Blanchard-Wrigglesworth[2], Hannah R. Hiester[1], and Andrey Proshutinsky[3]

[1]Center for Ocean-Atmospheric Prediction Studies, Florida State University, Tallahassee, Florida, USA, [2]Department of Atmospheric Sciences, University of Washington, Seattle, Washington, USA, [3]Department of Physical Oceanography, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts, USA

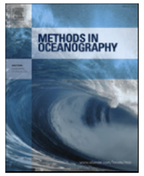
Methods in Oceanography 17 (2016) 232–250

Contents lists available at ScienceDirect

Methods in Oceanography

journal homepage: www.elsevier.com/locate/mio

ELSEVIER



Full length article

A topological approach for quantitative comparisons of ocean model fields to satellite ocean color data

CrossMark

Hannah R. Hiester[a,*], Steven L. Morey[a], Dmitry S. Dukhovskoy[a], Eric P. Chassignet[a], Vassiliki H. Kourafalou[b], Chuanmin Hu[c]

MacDonald I.R., D. Dukhovskoy, M. Bourassa, S. Morey, O. Garcia-Pineda, Daneshgar Asl, S. C. Hu, M. Reed, and J. Skancke. 2017. Remote Sensing Assessment of Surface Oil Transport and Fate During Spills in the Gulf of Mexico. US Dept. of the Interior, Bureau of Ocean Energy Management, Gulf of Mexico Region, New Orleans, LA. OCS Study BOEM 2017-030. 129 p.

MHD Values for Sea Ice Contours from a HYCOM-CICE Simulation

TITLE