The First Magnetospheric Space Weather Metrics Challenge

A Report

Michael Hesse
CCMC
NASA/GSFC

GEM
6/19/01

Background:History

- NSWP Iplan calls for evaluation of models
- Metrics study (R. Wolf and T. Fuller-Rowell, chairs) defines parameters
- NSF/R. Behnke calls for evaluation of magnetospheric models
- NSF tasks GGCM Steering Committee with implementation of first metric
- G. Siscoe leads implementation
- CCMC performs evaluation and reports results at SWW
- Additional participants now included in report to GEM

Metrics for This Study

Goal: compare velocity measurements from DMSP F13 with (ExB drift) velocities from models

Precisely:
If v_y is the DMSP cross-track component of the perpendicular (to B, ion) flow velocity, and v_{ym} is the corresponding quantity, derived from model m, then the <u>deviation</u> or <u>metrics</u> $D_{\underline{m}}$

$$D_m = \frac{1}{N}\sum_{i=1}^{N}\left|v_{y,i} - v_{ym,i}\right|$$

<u>Skill Score</u>:
If D_g is the deviation for a given (e.g., commonly used) model g, the the skill score of model m relative to model g is

$$S_{m,g} = 1 - \frac{D_m}{D_g}$$

Note: S<0 possible!

Here: g is the model, which predicts v_y=0 (!)

Event Selection

Based on quality of solar-wind and DMSP F13 data, and consensus

Period 1: 16-17 April 1999, Days of year = 106-107 (prime period)
Period 2: 10-11 December 1998, Days of year = 344-345
Period 3: 05-06 November 1998, Days of year = 309-310

Data from ACE, Geotail, IMP8, Interball Tail, and Wind, collected by M. Heinemann

DMSP F13 passes

Period 1:	33
Period 2:	40
Period 3:	37

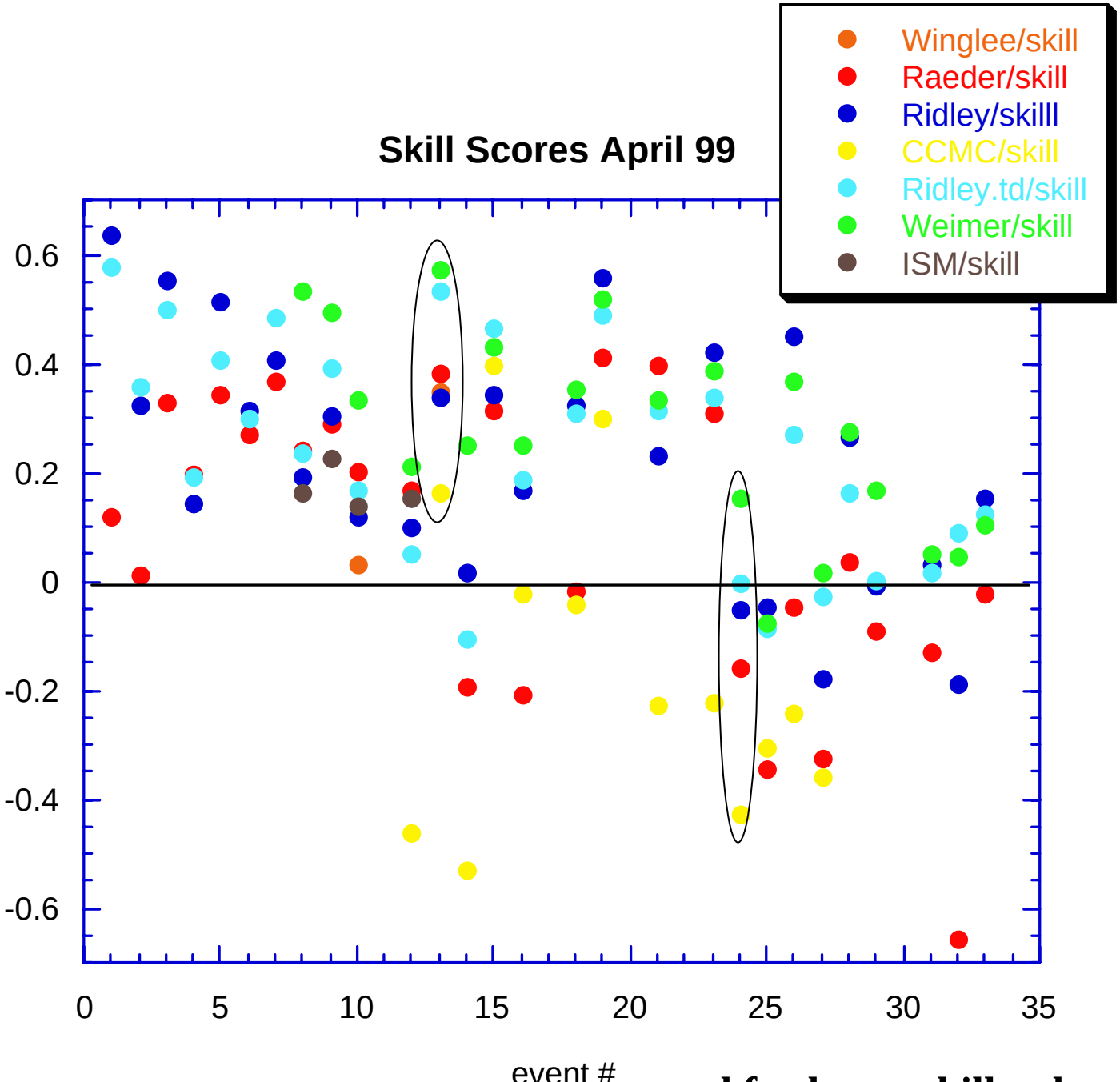
DMSP data provided by M. Hairston

Participants/POCs

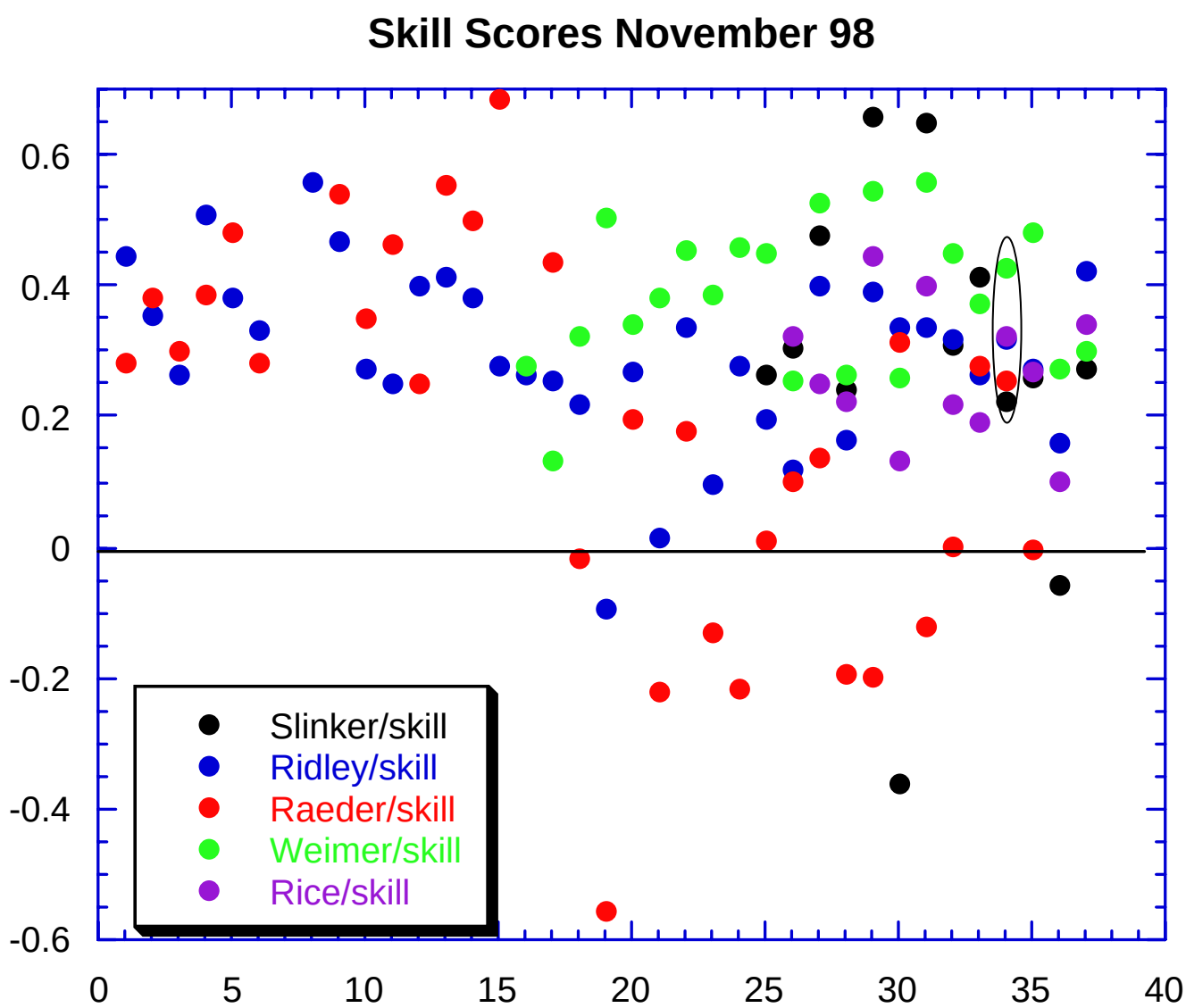
B. Hausman/RiceU:	Toffoletto-Hill Model
J. Raeder/UCLA:	UCLA MHD model
A. Ridley/Umich:	BATSRUS MHD model
S. Slinker/NRL:	LFM MHD model
R. Winglee/Uwash:	Winglee MHD model
D. Weimer/MRC:	Weimer PC model
D. Weimer/MRC-ISM:	ISM MHD model
M. Wiltberger/Dartmouth:	LFM model
CCMC (blind test only):	BATSRUS MHD model

Level of Participation

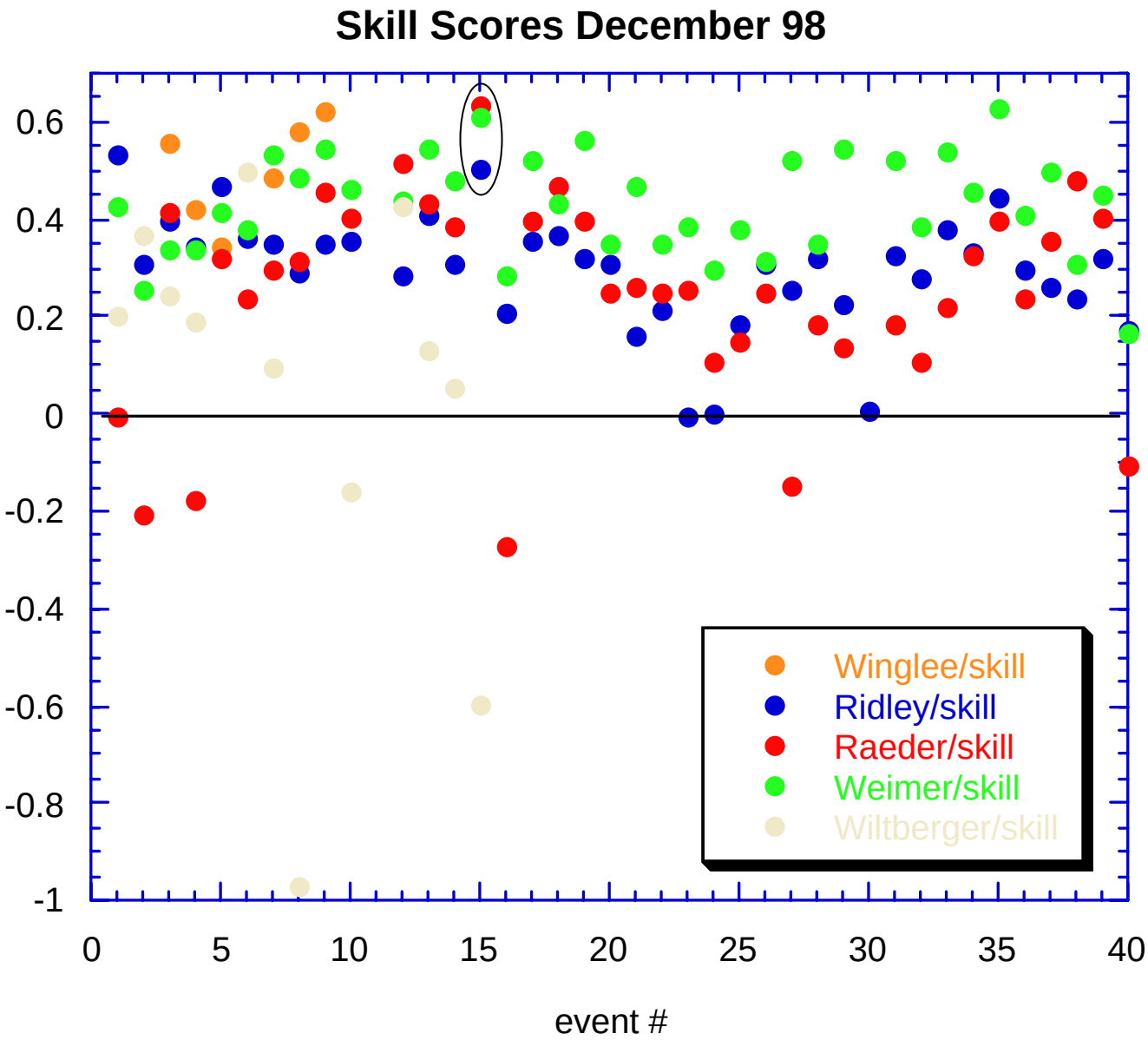
Participant/Period	November passes	December passes	April passes
Hausman	12	-	-
Raeder	34	40	31
Ridley	37	40	33(twice)
Slinker	13 (also F14)	-	-
Winglee	-	6	3
CCMC	-	-	17
Weimer/ PC model	22	40	26
Weimer/ ISM	-	-	5
Wiltberger	-	19	-



trend for lower skill at later times?



quality has same trend for all models

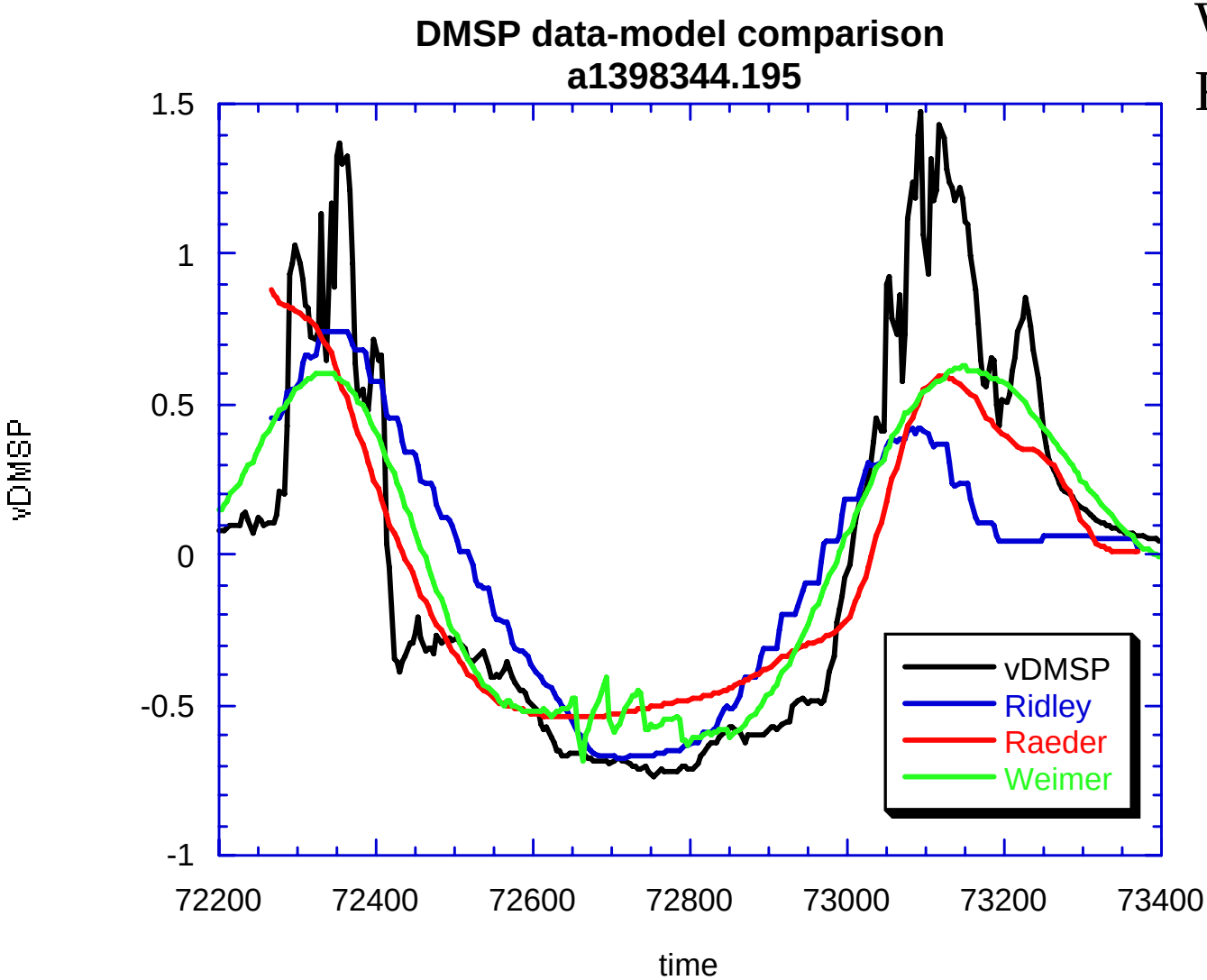

Skill Scores December 98
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
0
5
10
15
20
25
30
35
40
event #
Winglee/skill
Ridley/skill
Raeder/skill
Weimer/skill
Wiltberger/skill

A Few Select Passes

- High skill
- Low skill
- Medium skill 1
- Medium skill 2

A Particularly Good Case

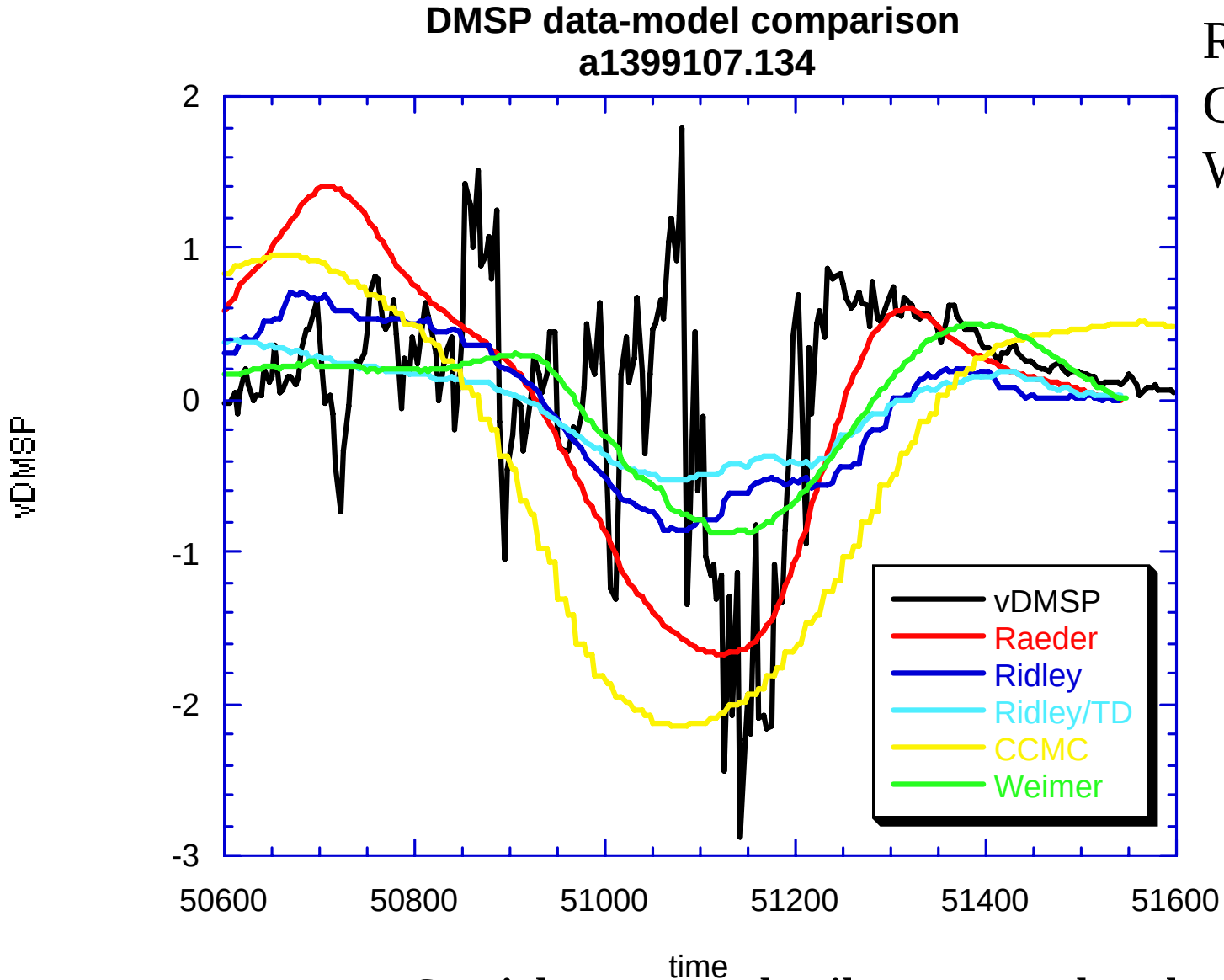
Raeder: 0.63
Weimer: 0.61
Ridley: 0.50



models appear to match smoothed v_y (100s) best

A Particularly "Bad" Case

Raeder: -0.19
Ridley: -0.05
Ridley/td: 0.0
CCMC: -0.43
Weimer: 0.16

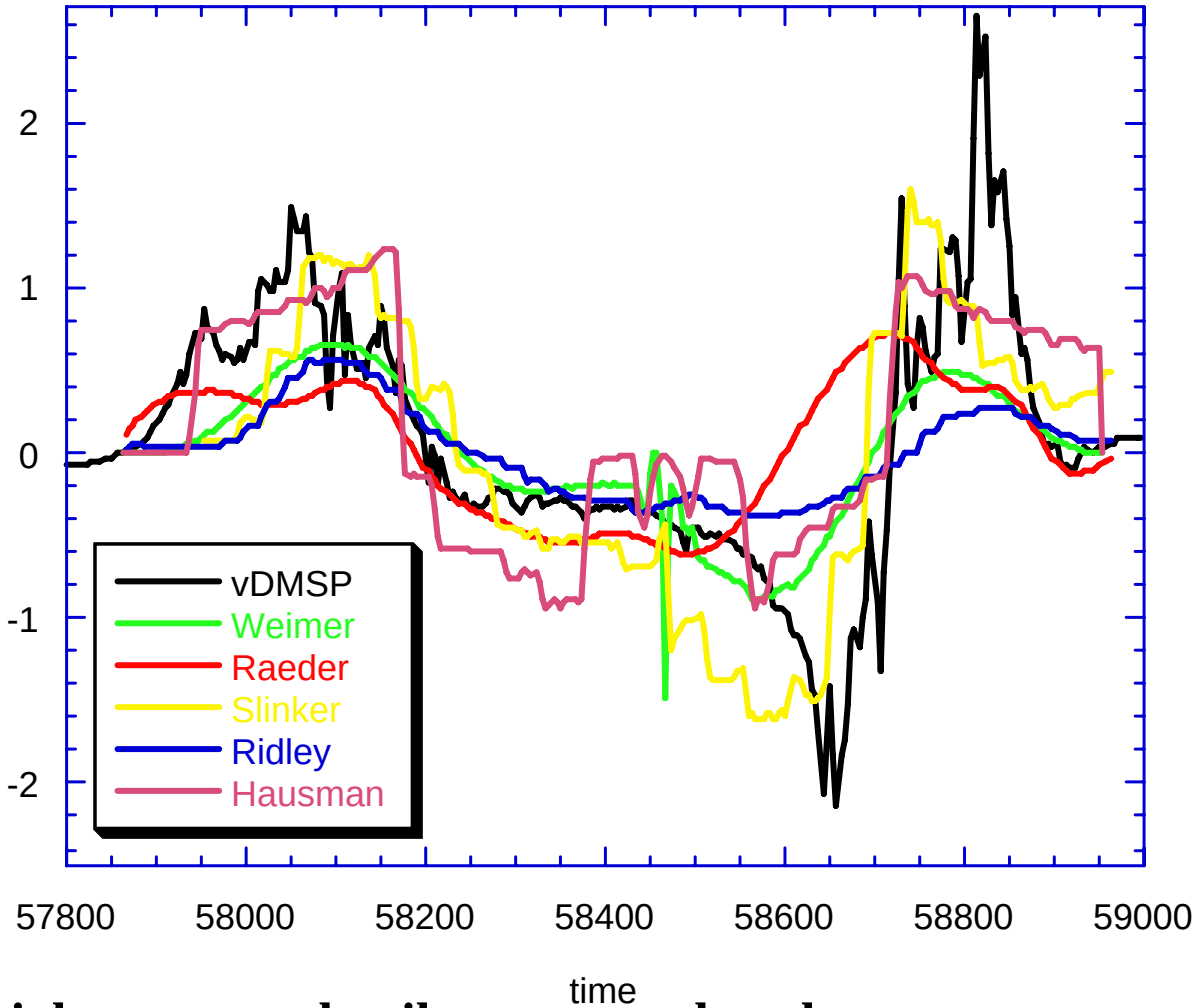


Spatial structure details not reproduced

Better match with smoothed v_y

Medium Skill Case 1

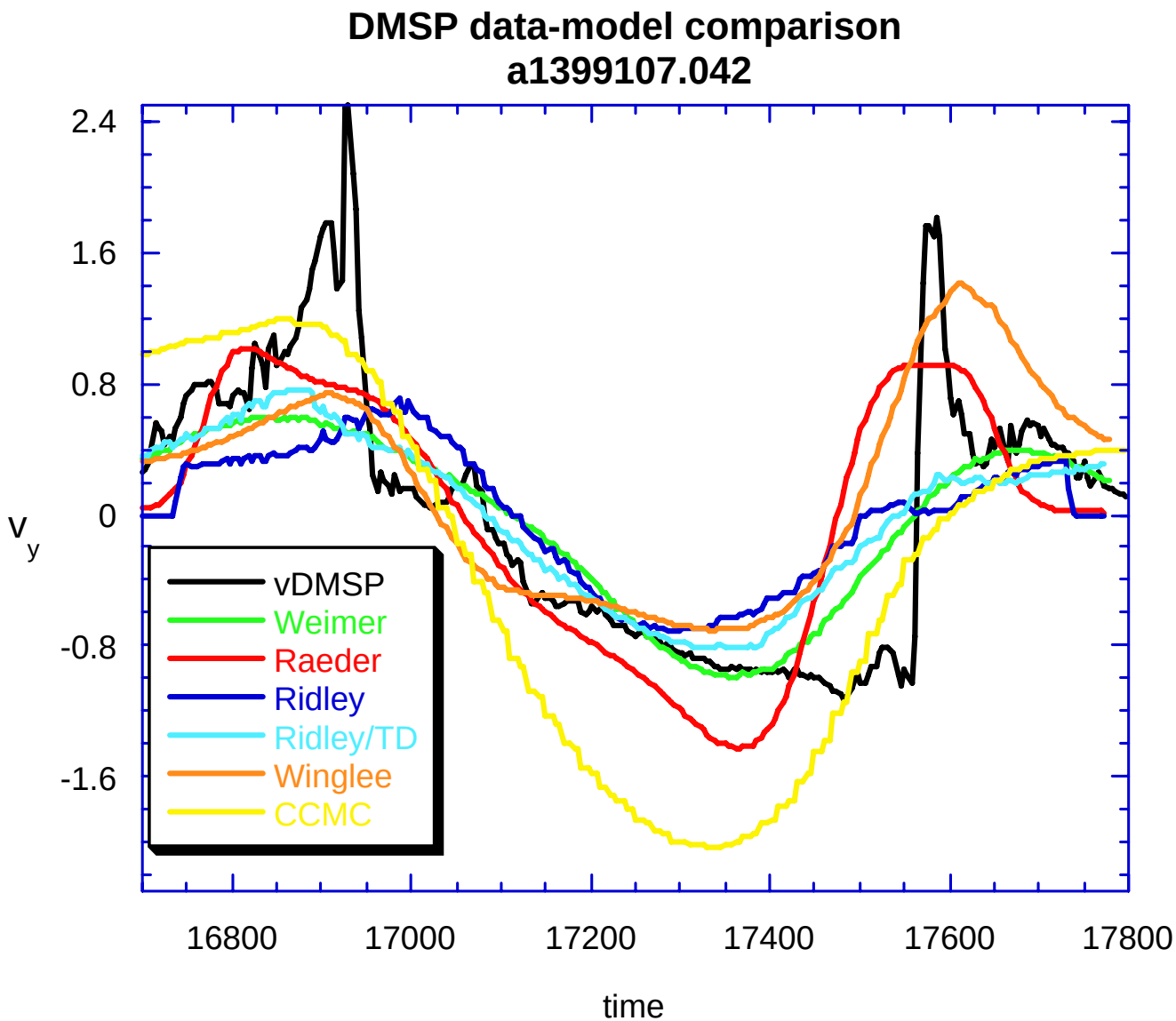
DMSP data-model comparison
a1398310.154



Raeder: 0.27
Ridley: 0.32
Rice: 0.32
Slinker: 0.23
Weimer: 0.43

Spatial structure details not reproduced
More detail (Rice, Slinker) does not necessarily lead to higher skill, although it is desirable

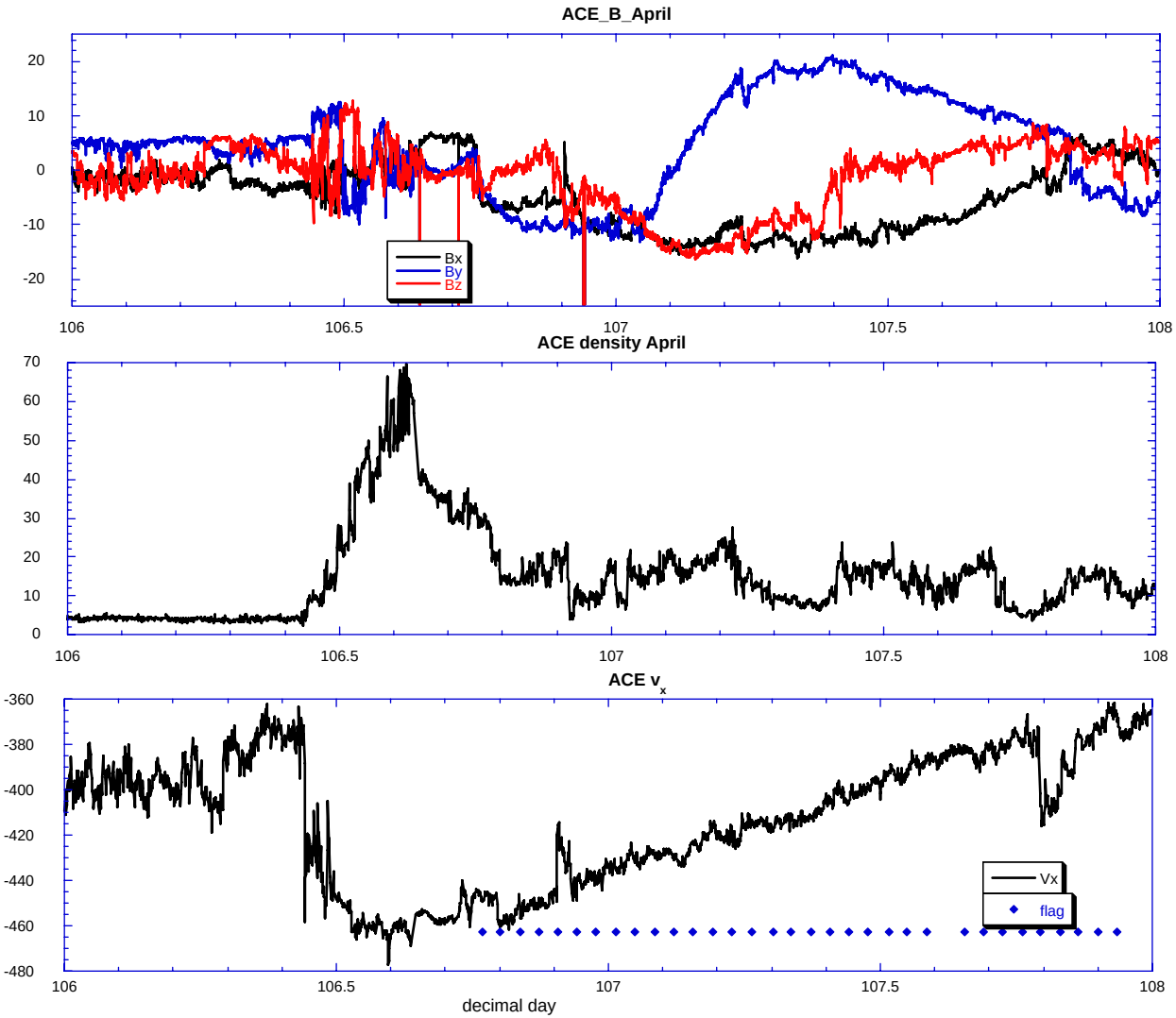
Medium Skill Case 2



Raeder: 0.38
Ridley: 0.34
Ridley/td: 0.53
Winglee: 0.35
Weimer: 0.57
CCMC: 0.17

Sharp boundaries not reproduced
Magnitudes are not too bad

April Period: ACE data



Lower model performance related to

- **Large B_z?**
- **Large B_y?** ←
- **Density fluctuations?**

Lessons learned

This was an interesting, very successful study

It went very well

For future metrics studies, we should

- precisely define the model data format early
- precisely define the relevant time intervals early
- define procedure of handling bad data points
- define procedure for comparing to runs w. bad data points
- precisely define who is responsible for what
- perform blind studies (my suggestion)

Summary/1

- There was no pass modeled by all participants
- All models are performing well, and better than v=0 prediction
- The same model can produce very different quality results for different passes
- There is no general trend for one model doing better under one condition and worse then another under a different condition
- It is very difficult to find passes where all models did well or all did badly
- All models show the same quality trends

Summary/2

- All models are roughly equivalent (bad statistics in some cases)

  The Michigan model provides roughly the same skill whether time dependent or not

  "Relatively simple" models (Weimer, Toffoletto-Hill) and the MHD codes provide similar quality results

- Models appear better at predicting flow magnitude than details such a flow reversal boundaries

- This was a great step in the right direction

- The results could be used for (friendly) in-depth studies and comparisons to improve overall skill scores (even further)