Adapting meteorological verification techniques for space weather at the UK Met Office

Edward Pope, Michael Sharpe, Sophie Murray, David Jackson, David Stephenson*, **Suzy Bingham**.

ILWS, India, 29th January 2016

* University of Exeter

Outline

Verification. Key aspects in space weather. Terrestrial resources

Met Office Space Weather Operations Centre

Verification of CME arrival time forecasts from WSA-Enlil model

Probabilistic geomagnetic storm forecast skill

Converting research to operational verification using NWP systems

Flare forecast verification plans

Summary

1. Space weather verification: key aspects

- Why verify?
  - To monitor, improve & compare forecast quality
  - Understand strengths/limitations
  - To assess forecaster added value
  - For forecasters, modellers, users & stake-holders to understand skill/value
  - Near real-time verification for operational purposes

- Key aspects:
  - Often issued as categories, as probabilities
  - Interest is in extreme events which occur very rarely
  - Data records are short
  - Observations for comparison can be non-existent
  - Data are strongly modulated by 11-year solar cycle
  - Standardise verification procedures across centres to enable comparisons (working with International Space Environment Service http://www.spaceweather.org/)
  - Can adapt NWP verification methods but be aware of differences between space weather/terrestrial meteorology

Terrestrial weather verification resources

- WMO World Weather Research Programme (WWRP): http://www.wmo.int/pages/prog/arep/wwrp/new/Forecast_Verification.html

JWGFVR WWRP WORLD WEATHER RESEARCH PROGRAMME WCRP World Climate Research Programme

WWRP/WGNE Joint Working Group on Forecast Verification Research

---

New: **Enter the *Challenge to Develop and Demonstrate the Best New User-Oriented Forecast Verification Metric***

The aim of this challenge is to promote user-oriented verification, that is, quantitative assessment of forecast quality in terms that are meaningful to particular forecast users. The scope includes all applications of meteorological and hydrological forecasts. The user-oriented verification metrics will help support the WWRP High Impact Weather Project.
Click here to find out more, or contact verifchallenge@ucar.edu.

Introduction - what is this web site about?

Issues:
Why verify?
Types of forecasts and verification
What makes a forecast good?
Forecast quality vs. value
What is "truth"?
Validity of verification results
Pooling vs. stratifying results

Methods:
Standard verification methods:
- Methods for dichotomous (yes/no) forecasts
- Methods for multi-category forecasts
- Methods for forecasts of continuous variables
- Methods for probabilistic forecasts

Scientific or diagnostic verification methods:
- Methods for spatial forecasts
- Methods for probabilistic forecasts, including ensemble prediction systems
- Methods for rare events
- Other methods

Sample forecast datasets:
- Finley tornado forecasts
- Probability of precipitation forecasts

Freely available verification tools and packages

- Website maintained by WMO verification Working Group, http://www.cawcr.gov.au/projects/verification/, includes:
  - Methods (brief definitions)
  - Verification issues
  - FAQs
  - Links and references
  - Verification discussion group

Info from: B. Brown. ISES Verification Workshop, Apr. 2015

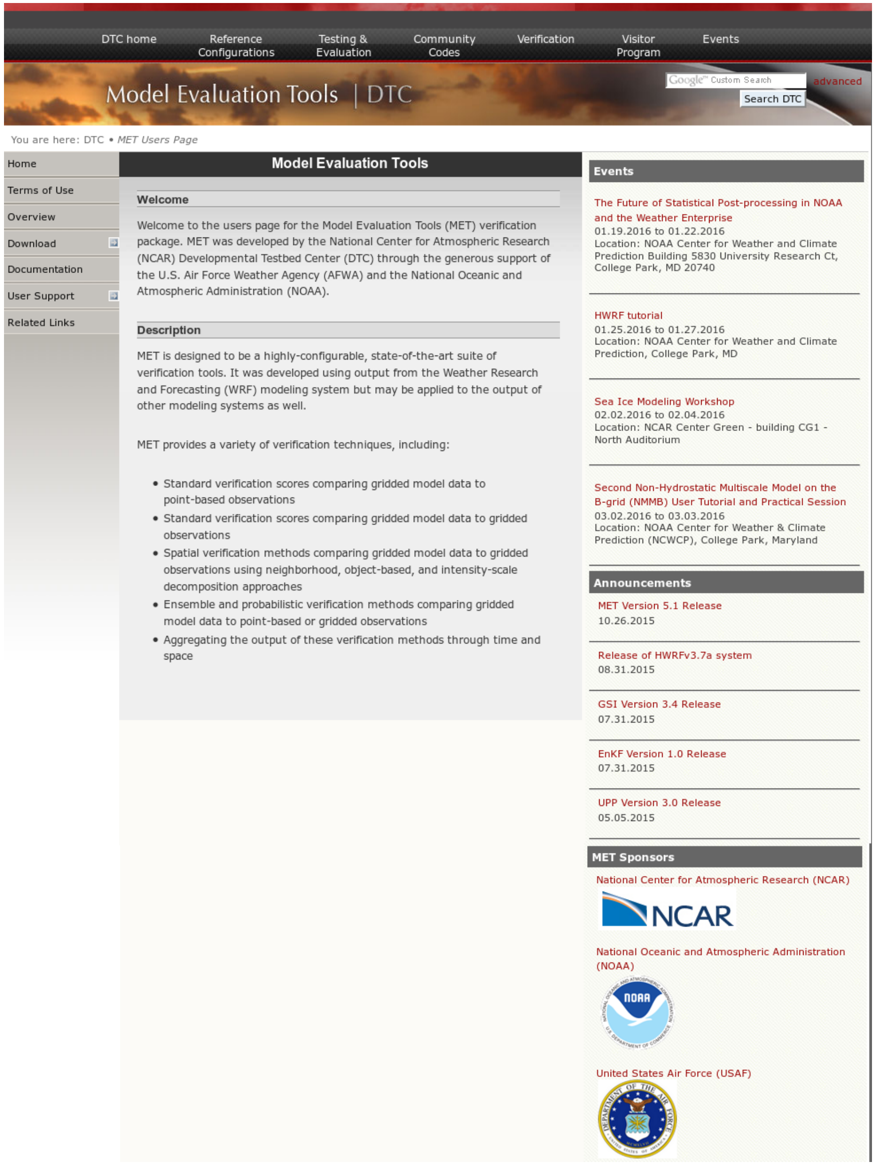
Met Office

Terrestrial weather verification tools: MET & ‘R’ libraries

- Model Evaluation Tools (MET):

http://www.dtcenter.org/met/users/

  - Forecast evaluation tools
  - Implemented & supported by the Developmental Testbed Center (DTC) & Joint Numerical Testbed Program at NCAR/RAL
  - Includes a suite of standard stats & non-traditional stats (e.g. spatial methods)
  - Designed to undertake systematic evaluations
  - Has a database & display system for aggregating & plotting data
  - Provides a standardized evaluation platform for cross-institution comparisons
  - Freely available, highly configurable, “live” tutorials



- NCAR verification stats packages:

http://www.r-project.org/

Info from: B. Brown. ISES Verification Workshop, Apr. 2015

2. Met Office Space Weather Operations Centre (MOSWOC)

- Apr. ‘14: 24x7 operations
- Oct. ‘14: full capability
- Operational collaboration with NOAA SWPC and BGS.
- Products: CME forecasts and guidance on geomagnetic storms, radiation storms and X-ray flares.

Public webpages:

http:// www.metoffice.gov.uk/publicsector/emergencies/space-weather

MOSWOC forecasts

Components of the guidance (issued twice/day)...

Analysis of activity

4 day summary

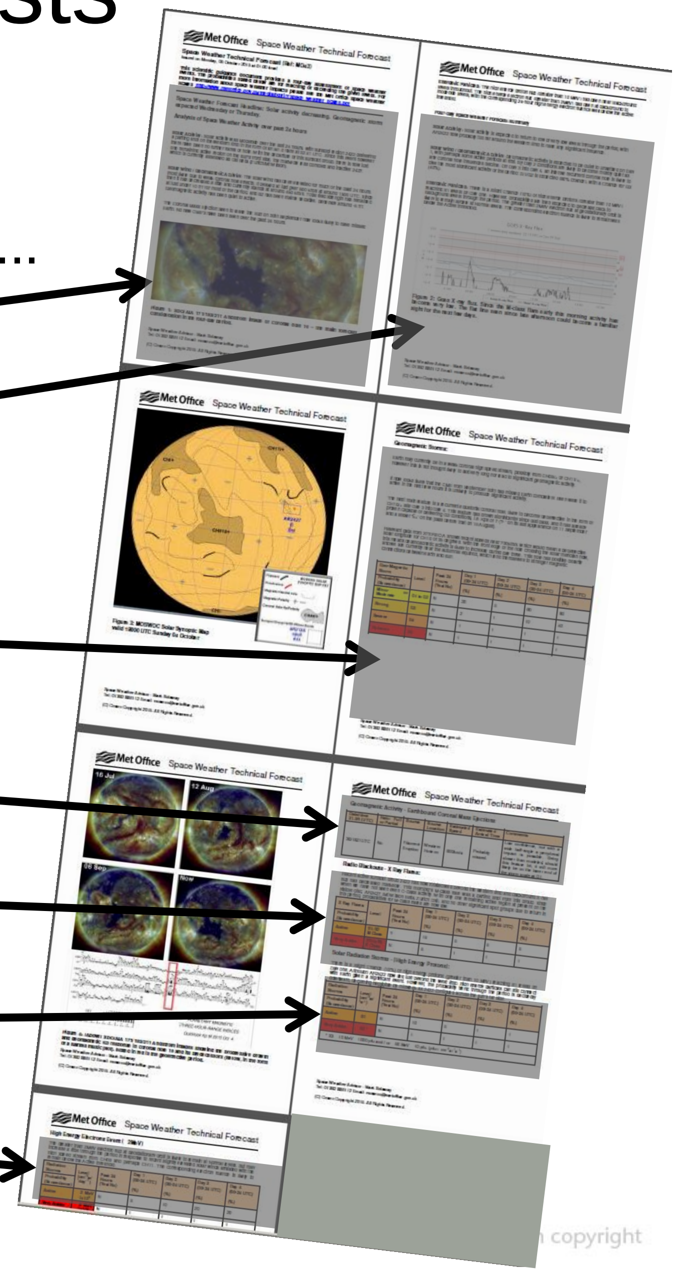
Geo-magnetic storm forecast

Earthbound CME warning

Radio blackout forecast

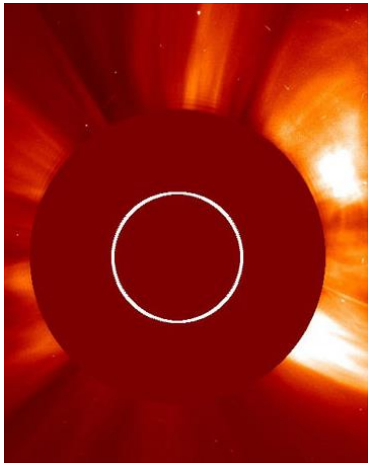
Solar radiation storm forecast

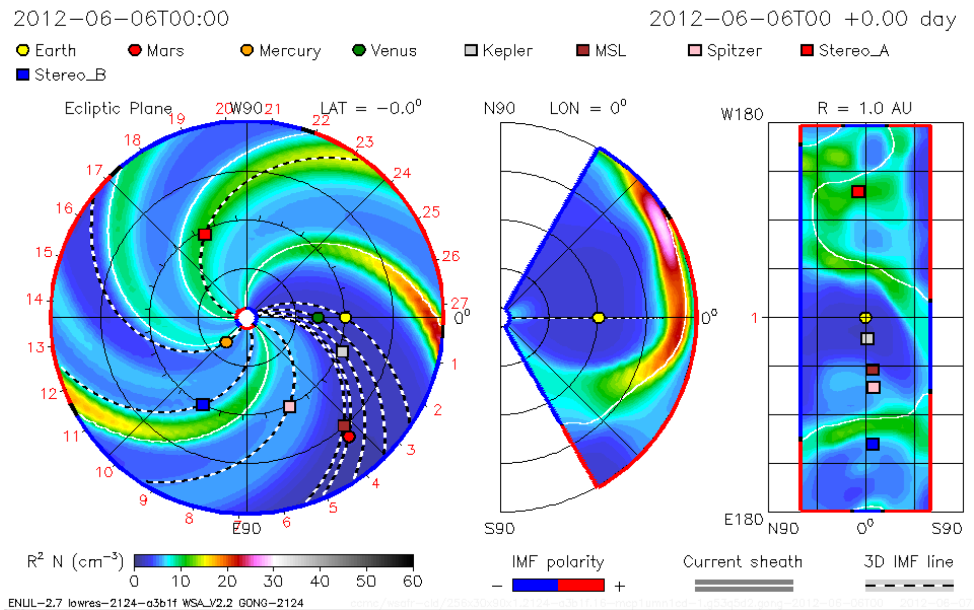
High energy electron event forecast



3. CME forecasts

- CME arrival time forecasts use WSA-ENLIL (3-D MHD) solar wind model:
  - provides 1-4 day warning of geomagnetic storms
- CMEs initialised using coronograph images (SOHO, STEREO) => to estimate basic CME properties (time at 21.5 Rs, source lat/lon, half angle, radial velocity)
- MOSWOC issue forecast arrival times, as well as speed and source region





Date/time 21.5R (UTC)	Halo: Full or Partial	Source	Source Location	Estimated Speed	Estimated Arrival Time	Comments
12/1920Z	Partial	Filament eruption	20S35W	600	16/0200	nil

CME forecast verification

- Compare observed CME arrivals (identified using Advance Composition Explorer (ACE) data) with MOSWOC forecasts:
    - Use verification statistics derived from 2x2 contingency table, e.g. hit rate, false alarm rate, Heidke/Peirce skill scores, etc.

	Observed	
Forecast	Hit	False alarm
	Miss	Correct rejection

    - Bootstrap contingency table to get 90% confidence interval for each derived quantity.
- Compare MOSWOC performance against other space weather forecasters (e.g. NASA CCMC: http://kauai.ccmc.gsfc.nasa.gov/CMEscoreboard/).

MOSWOC v CCMC CME arrival time forecast verification

Category	Metric	MOSWOC	CCMC	90% conf. ints.overlap?
Accuracy	Proportion Correct	0.73	0.75	
	Threat Score	0.69	0.69	
Bias	Bias	0.93	1.44	N
Reliability	False Alarm Ratio	0.15	0.31	N
Discrimination	Hit Rate	0.79	1.00	N
	False Alarm Rate	0.46	0.57	N
Skill	Heidke	0.30	0.45	
	Peirce	0.32	0.43	
	Equit. Threat Score	0.18	0.30	

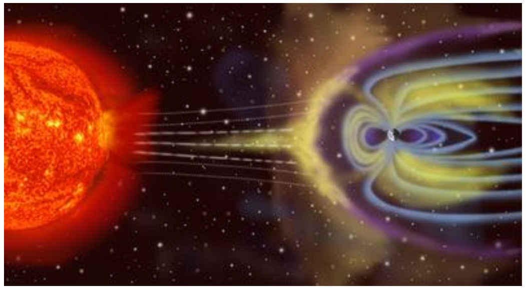
- **Hit rate**: CCMC always predict a hit; false alarm rate and ratio are also higher
- **Bias**: MOSWOC 0.9 - slight under-prediction of events

  CCMC 1.4 - over-prediction of events (consistent with the high hit/false alarm rate)
- **Equitable Threat Score** and **Heidke Skill Scores** are comparable
- Overall, results suggest broadly comparable performance of MOSWOC and CCMC CME forecasts, despite slightly different approaches

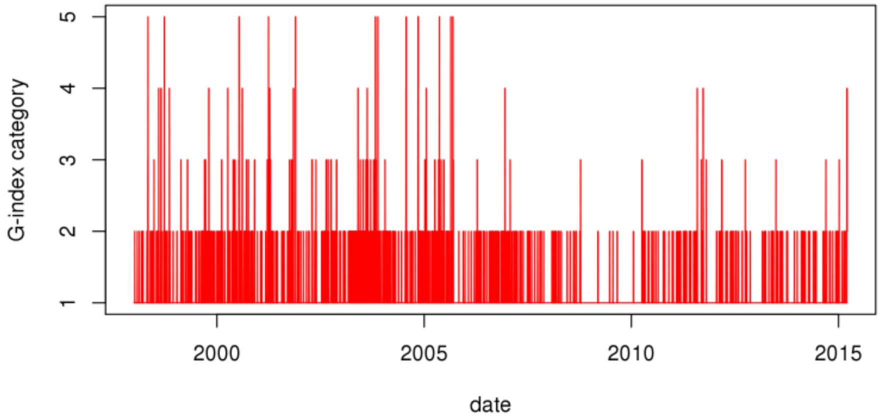
4. Geomagnetic storms

Met Office

- Solar wind can cause disturbances in the Earth's magnetic field via varying compression and/or open field lines.
- Geomagnetic storms can be caused by CMEs or variations in solar wind speed. A southward z-component of CME/solar wind B-field results in stronger storms.



- Planetary K-index (K_p) indicates disturbances in the horizontal geomagnetic field.
- K_p ranges from 0 – 9 (0 = no disturbance; >= 5 indicates the occurrence of a geomagnetic storm) :
  - Storms are characterised using the NOAA G-index, where G = K_p – 5.
- MOSWOC issues probabilistic categorical forecasts for the likelihood of G1-5 disturbances with 24 hour periods, out to 4 days ahead.



Geo-Magnetic Storm Probability (Exceedance)	Level	Past 24 Hours (Yes/No)	Day 1 (00-24 UTC) (%)	Day 2 (00-24 UTC) (%)	Day 3 (00-24 UTC) (%)	Day 4 (00-24 UTC) (%)
Minor or Moderate	G1 to G2	N	5	5	40	30
Strong	G3	N	1	1	5	5
Severe	G4	N	1	1	1	1
Extreme	G5	N	1	1	1	1

Verification of Kp/G-index forecasts

Assess G-index forecasts against observations using:

- Brier scores for each category, i.e. $BS = \frac{1}{N}\sum_{i=1}^{N}(P_i - O_i)^2$
- Ranked Probability Scores to assess the overall performance, i.e.

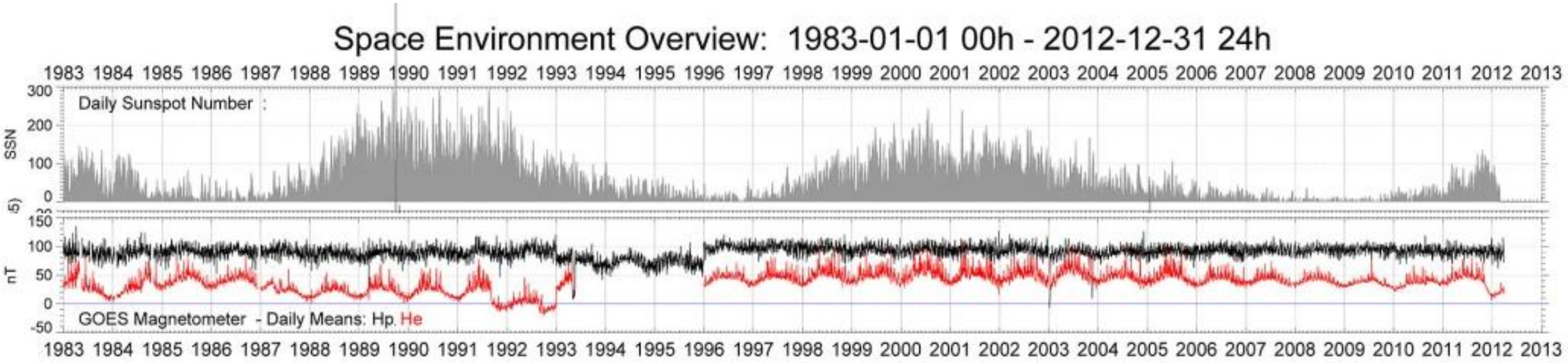
$$RPS = \frac{1}{M-1}\sum_{m=1}^{M}\left[\left(\sum_{k=1}^{m}P_k\right)-\left(\sum_{k=1}^{m}O_k\right)\right]^2$$

Assess G-index forecast **skill** by comparing performance against:

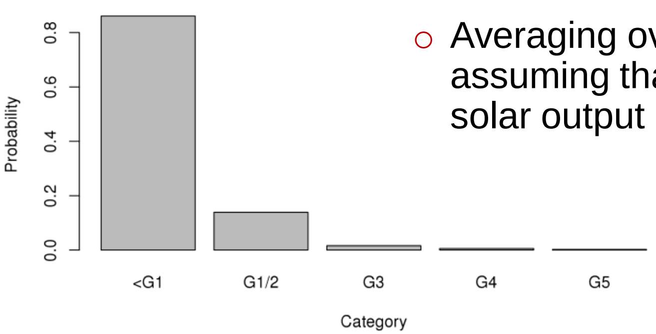
- Climatology, i.e. $BSS = 1 - \frac{BS_{fcast}}{BS_{clim}}$, $RPSS = 1 - \frac{RPS_{fcast}}{RPS_{clim}}$
- Persistence forecast, i.e. $BSS = 1 - \frac{BS_{fcast}}{BS_{pers}}$, $RPSS = 1 - \frac{RPS_{fcast}}{RPS_{pers}}$

Kp index climatology

- In climate science, at least 30 years of data are needed to derive a robust climatology.
- What is the equivalent for solar output which exhibits 11 year cycles? For example, 30 solar cycles = 30 x 11 = 330 years.



- Several options for deriving climatological frequencies, e.g. :
  - Averaging over all available observations (20-30 years = 2-3 solar cycles).
  - Averaging over a recent period of observations (e.g. last 2 years), and assuming that this provides an adequate representation for the climatology of solar output at the present phase of the current solar cycle.



More extreme events (G3-G5) are the most important but are also very rare!

Markov chain persistence model

- When the geomagnetic field is disturbed, the Kp-index time series exhibits an almost instantaneous rise, followed by a decay which occurs over a period of 1-2 days
    - A one-step Markov chain provides an informative description:
        - Use time series of daily maximum Kp/G-index to generate a matrix of transition probabilities (T), i.e. $P_{ji} = \Pr(X_{n+1} = j \mid X_n = i)$
    - Starting from the observed state on a given day, u (e.g. u = (0,1,0,0,0)), the forecast probabilities on the *n*th day are: $u_n = uT^n$
    - Quantify uncertainty in transition matrix (and forecast probabilities) by bootstrapping.
    - For N >=3, $T^n \sim P_{clim}$

Kp verification summary

Results so far indicate the following:

- The performance of the climatological and Markov chain forecasts relative to the standard forecast is significantly affected by the data used to train the models.
    - Both statistical forecasts perform much better when trained on recent data (e.g. the most recent 1-2 years), than a longer time series.
- The Ranked Probability Skill Scores (RPSS) suggests that the Markov chain model can outperform the standard and climatological forecasts on days 1 and 2.
    - For days 3 and 4, the Markov chain and climatological forecast skill is comparable.
- The Brier Scores indicate that Markov chain forecast can perform better than the standard and climatological forecasts in the low Kp/G-index categories, where the vast majority of events occur.
    - In the high Kp/G-index categories the performance of the three forecasts models is almost indistinguishable, primarily due to the rarity of G3,4 and 5 events.

5. Adapting a meteorological verification system

Recently we developed a new verification system to evaluate categorical forecasts in near-real-time.

Originally applied to marine products:

- Shipping forecast
- Inshore waters forecast
- High seas forecast

It is now being used more widely.

This system has been adapted to verify the geomagnetic storm forecast.

Although still in initial stages.


METEOROLOGICAL APPLICATIONS
Meteorol. Appl. **20**: 224–235 (2013)
Published online 22 May 2013 in Wiley Online Library
(wileyonlinelibrary.com) DOI: 10.1002/met.1341

RMetS
Royal Meteorological Society

Verification of marine forecasts using an objective area forecast verification system

Michael A. Sharpe*
Weather Science, Met Office, Exeter, UK

ABSTRACT: Many meteorological service providers issue forecasts that contain short summary statements. These
forecasts attempt to capture all the expected conditions during a forecast period. The area forecast verification system
has been developed to verify these forecasts when they are issued in pre-determined geographical areas. This system
verifies forecasts by comparing them with gridded nowcast model analysis fields. Results are presented for the wind speed
component of the shipping forecast and the inshore waters forecast. Both of these products are issued by the UK Met Office
on behalf of the Maritime and Coastguard Agency. Like most summary forecasts these products contain hand written text.
This text is carefully interpreted into a series of invariant sub-periods. These sub-periods are then compared to the nowcast
model analysis fields. During each forecast sub-period all available hourly nowcast model analyses are accumulated to
create an observation distribution within that sub-period. Verification is performed using a contingency table approach which
is specifically designed to reflect the distribution formed by the observations within each sub-period of the forecast. Various
performance measures are calculated including the Gerrity skill score – which is weighted according to the priorities of
the customer. Shipping forecasts issued during the 12 month period ending on 31 August 2011 had a proportion correct to
within ±1 Beaufort Force of 0.886 and a Gerrity skill score of 0.823. Inshore waters forecasts issued during this period
had a proportion correct to within ±1 Beaufort Force of 0.940 and a Gerrity skill score of 0.848. Analysis indicates that
both forecasts tend to exaggerate the wind speeds and often neglect to include the lower Beaufort Force categories.
© 2013 British Crown copyright, the Met Office. Published by John Wiley & Sons Ltd.

KEY WORDS nowcast; analyses; contingency table; Gerrity skill score; Peirce skill score; Beaufort force

Received 28 November 2011; Revised 23 April 2012; Accepted 21 May 2012

1. Introduction

Forecast products tend to be issued either for a specific site/location or for a predefined geographical area. Forecasts which are issued for specific sites are usually verified against observations taken at those sites and many meteorological forecasting centres also perform area based verification of their numerical models over a defined grid of points at specified times. However, a different approach must be taken for forecast products when they attempt to capture all conditions within specified geographical areas through the use of short summary statements. Such area forecasts attempt to give a range of values for weather variables (e.g. wind speed, wind direction, temperature, visibility) to capture all the observed values expected at every location within a geographical area. Consequently the range of weather values in an area forecast tends to be greater than the equivalent range for a site specific forecast. Often area forecasts are written by hand by a forecaster using various key words and phrases which have specific meanings. Time information is often expressed in this way, i.e. 'at first', 'later', 'increasing'. In the UK the two best known examples of area forecasts are the shipping forecast and the inshore waters forecast. These products have become so well known that they appear in popular literature such as *And Now the Shipping Forecast*, Jefferson (2011). Both of these high profile forecasts are produced by the UK Met Office on behalf of the Maritime and Coastguard Agency (MCA) and are regularly broadcast on the radio and television. Despite the growth in automation these particular forecast products continue to be completely hand written by marine forecasters. It is possible that an automated forecast system may be developed in the future. However, before any such product is adopted it is essential that a verification system can demonstrate that it is at least as accurate as the manual product. Both forecasts are accompanied by warning services, gale warnings accompany the shipping forecast and coastal strong wind warnings accompany the inshore waters forecast. These warning services only predict when the wind speed is expected to exceed a single predefined threshold, they should not be confused with the forecast products.

This paper provides a description of the Area Forecast Verification System (AFVS) which has been specifically designed to verify area-based summary forecasts. The AFVS is currently being used to verify the wind speed component of the shipping forecast and the inshore waters forecast. Wind speed was identified by the MCA as the most important part of the forecast. However, it is hoped that all components of these forecasts will eventually be verified. The MCA uses information provided by the AFVS to measure the performance of the service they receive from the UK Met Office. This verification information is used to validate the standards in the memorandum of understanding which exists between these two organizations.

Ideally, real observations should be used to verify forecasts. However, real observations can only be used if they are

*Correspondence to: M. A. Sharpe, Weather Science, Met Office, Exeter EX1 3PB, UK. E-mail: michael.sharpe@metoffice.gov.uk

Verification of Kp

Probabilities are cumulative
Probability ≥G0 is always 100%
Min probability = 1%

The probability density function gives the probability each category will occur

Insignificant	G0	Y	100 85	100 70	100 70	100 70
Geo-Magnetic Storm / Probability (Exceedance)	**Level**	**Past 24 Hours (Yes/No)**	**Day 1 (00-24 UTC) (%)**	**Day 2 (00-24 UTC) (%)**	**Day 3 (00-24 UTC) (%)**	**Day 4 (00-24 UTC) (%)**
Minor or Moderate	**G1 to G2**	N	15 15	30 30	30 30	30 30
Strong	**G3**	N	1 0	1 0	1 0	1 0
Severe	**G4**	N	1 0	1 0	1 0	1 0
Extreme	**G5**	N	1 0	1 0	1 0	1 0

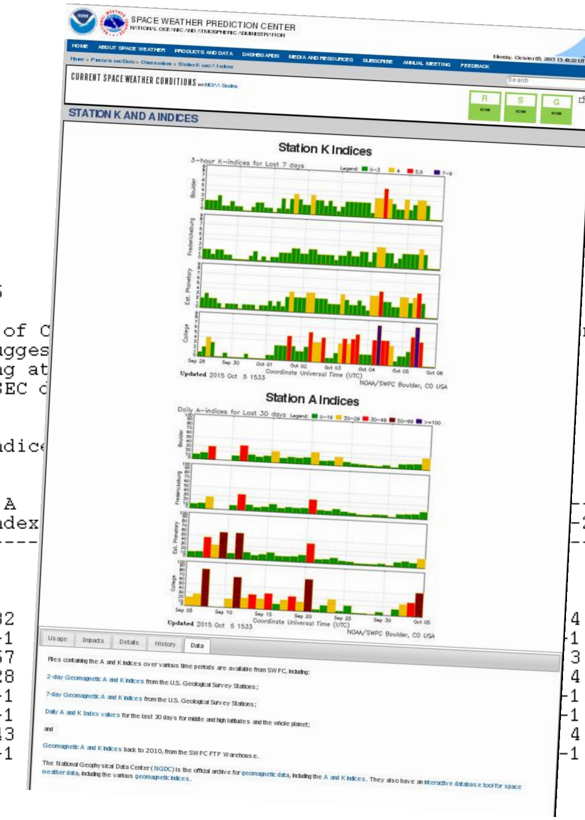
Verification of Kp

To verify GM Storm forecast observations are needed in near real-time.

SWPC’s 7day_AK.txt contains:

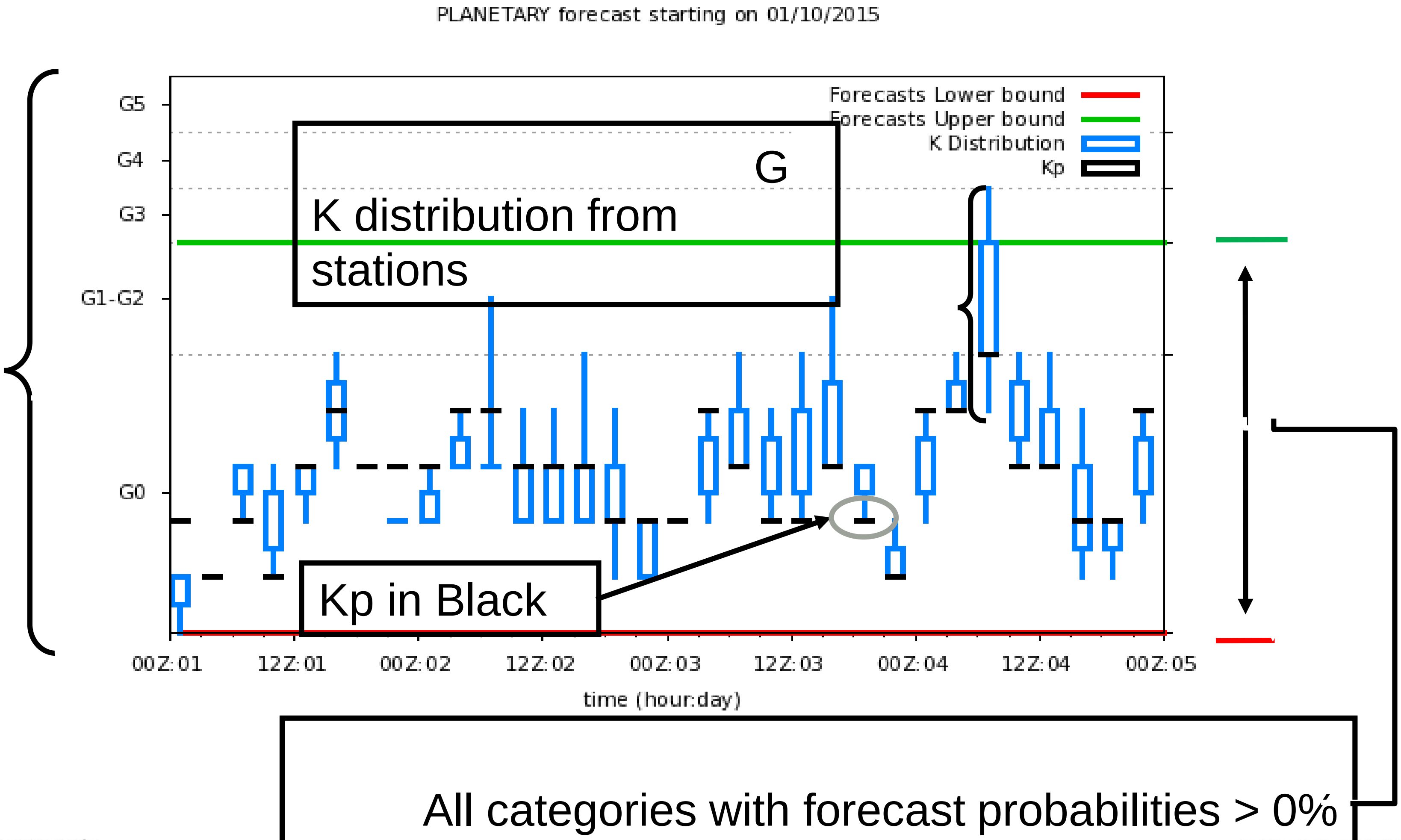
- Data from the past 7 days
- 3-hourly values of...
  - Kp
  - 7 station K values

```
:Product: Geomagnetic Data
:Issued: 0933 UTC 04 Sep 2015
#
# Prepared by the U.S. Dept. of C
# Please send comments and sugges
# Updated every hour beginning at
# Values shown as reported, SEC c
# Missing Data: -1
#
#       Geomagnetic A and K indice
#
#                 Geomagnetic
#                   Dipole      A
# Station          Lat Long   Index
#-----------------------------------

2015 Aug 28

Boulder            N49 W 42    32
Chambon-la-foret   N-- E---    -1
College            N65 W102    57
Fredericksburg     N38 W 78    28
Kergulen Island    S57 E130    -1
Learmonth          S22 E114    -1
Planetary(estimated Ap)        43
Wingst             N54 E 95    -1


2015 Aug 29

Boulder            N49 W 42    15     4     2     2     4     2     4     2     2
Chambon-la-foret   N-- E---    17    -1    -1     2     3     2     3     3     2
College            N65 W102    20     3     3     3     6     3     2     2     2
Fredericksburg     N38 W 78    13     4     2     2     4     3     2     2     2
Kergulen Island    S57 E130    -1    -1    -1    -1    -1    -1    -1    -1    -1
Learmonth          S22 E114    -1    -1    -1    -1    -1    -1    -1    -1    -1
Planetary(estimated Ap)        16     5     3     2     4     2     3     2     2
Wingst             N54 E 95    -1    -1    -1    -1    -1    -1    -1    -1    -1


2015 Aug 30

Boulder            N49 W 42     6     2     1     2     1     2     2     1     2
Chambon-la-foret   N-- E---     8     1     1     1     1     1     1     2     1
College            N65 W102     4     2     2     1     1     1     1     1     1
Fredericksburg     N38 W 78     5     1     1     2     2     2     2     1     1
Kergulen Island    S57 E130    -1    -1    -1    -1    -1    -1    -1    -1    -1
Learmonth          S22 E114    -1    -1    -1    -1    -1    -1    -1    -1    -1
Planetary(estimated Ap)         5     2     1     2     1     1     1     1     1
Wingst             N54 E 95    -1    -1    -1    -1    -1    -1    -1    -1    -1


2015 Aug 31

Boulder            N49 W 42     7     2     1     1     1     2     3     2     2
Chambon-la-foret   N-- E---    11     1     1     1     2     2     1     2     2
College            N65 W102     2     2     1     0     0     0     1     1     1
Fredericksburg     N38 W 78     5     1     1     1     2     2     1     2     2
Kergulen Island    S57 E130    -1    -1    -1    -1    -1    -1    -1    -1    -1
Learmonth          S22 E114    -1    -1    -1    -1    -1    -1    -1    -1    -1
```



Files are extracted & processed every 3 hours

Distribution of K observations and Kp from 1-4 Oct 2015.



Skill score to measure Kp forecasts

Need a score to measure performance...

GM storm forecast is categorical & probabilistic

.... Ranked Probability Score is the obvious choice

$$RPS = \sum_{i=0}^{5} \left(P(G_i) - O(G_i) \right)^2$$

Where

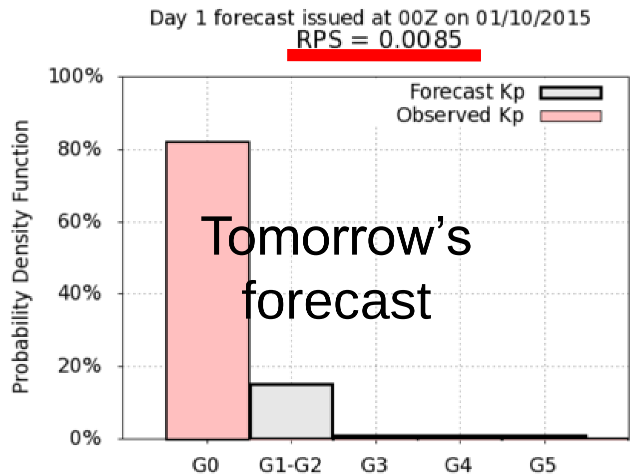
P(Gi) = probability that the observed category is ≤ Gi

$$O(G_i) = \begin{cases} 0 & \text{if observed category} < G_i \\ 1 & \text{if observed category} \geq G_i \end{cases}$$

RPS range is [0,1]
0 is a perfect score

RPS calculated for forecast on 1 Oct. ‘15

Today’s forecast

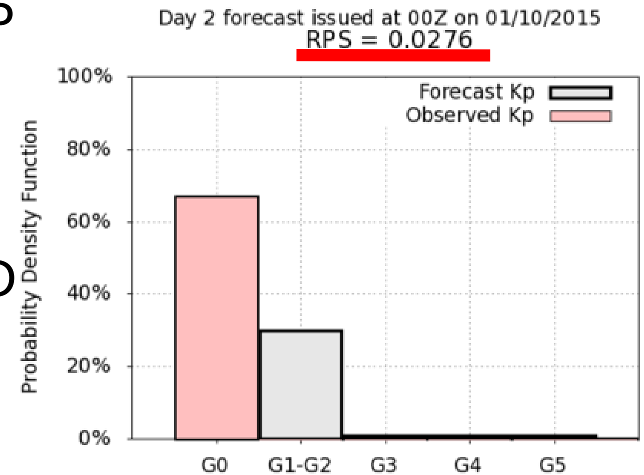


Tomorrow’s forecast

P

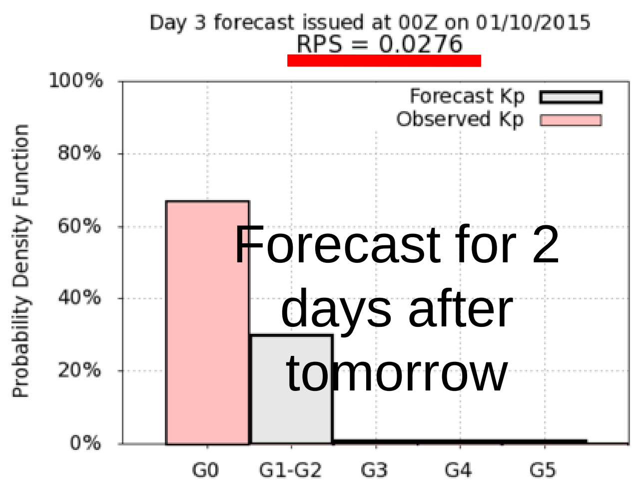
M

D



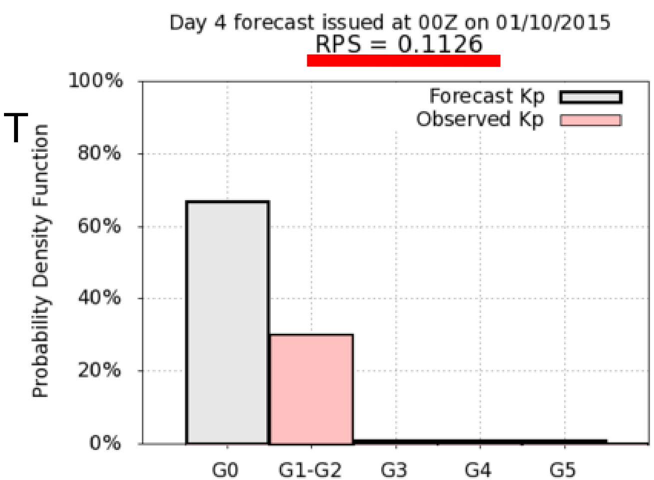
T

Day after tomorrow’s forecast



Forecast for 2 days after tomorrow

T



D

Compare forecast to a benchmark

To determine what is a 'good' forecast:

- Compare the performance to a reference forecast, e.g.:
  - random chance
  - persistence
  - climatology

SWPC ftp site has data from January 2010.....so a 5-year GM storm climatology (2010-4) was created.

- Then calculate a Skill Score, e.g. RPSS
-

$$RPSS = 1 - \frac{\overline{RPS}}{\overline{RPS}_{ref}}$$

RPSS range is (-∞,1)

1= perfect score

0= no additional skill compared to the reference

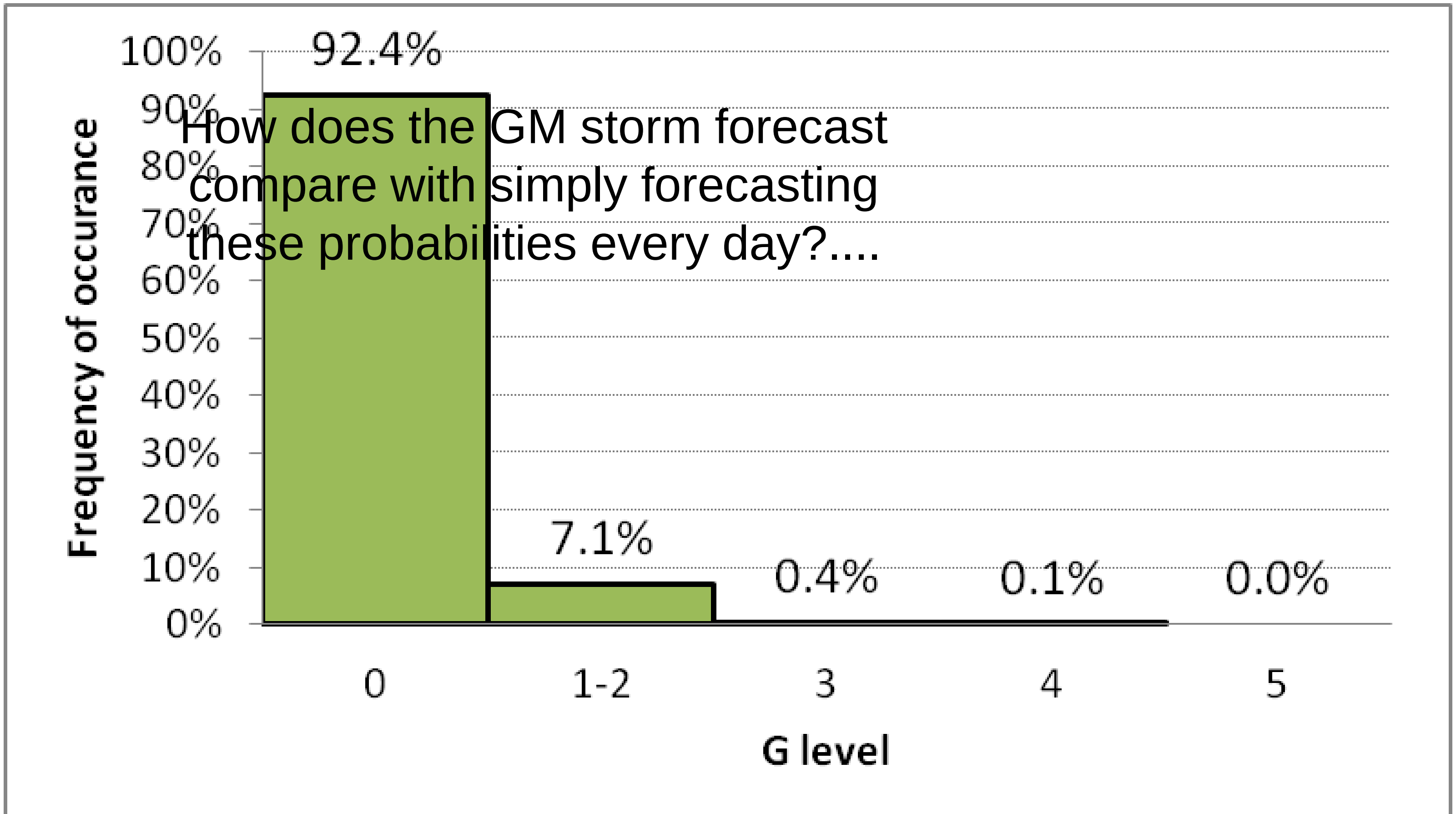
Directory /pub/lists/geon

Up to higher level directory

Last Modified	Size	Name
Sun Feb 28 00:00:00 2010	55012	201001AK.txt
Wed Mar 31 00:00:00 2010	49753	201002AK.txt
Fri Apr 30 00:00:00 2010	55012	201003AK.txt
Sun May 30 00:00:00 2010	53259	201004AK.txt
Wed Jun 30 00:00:00 2010	55012	201005AK.txt
Fri Jul 30 00:00:00 2010	53259	201006AK.txt
Tue Aug 31 00:00:00 2010	55012	201007AK.txt
Thu Sep 30 00:00:00 2010	55012	201008AK.txt
Sat Oct 30 00:00:00 2010	53259	201009AK.txt
Tue Nov 30 00:00:00 2010	55012	201010AK.txt
Thu Dec 30 00:00:00 2010	53259	201011AK.txt
Mon Jan 31 00:00:00 2011	55012	201012AK.txt
Mon Feb 28 00:00:00 2011	55012	201101AK.txt
Mon Mar 28 00:00:00 2011	49753	201102AK.txt
Sat Apr 30 00:00:00 2011	55012	201103AK.txt
Mon May 30 00:00:00 2011	53259	201104AK.txt
Thu Jun 30 00:00:00 2011	55012	201105AK.txt
Sat Jul 30 00:00:00 2011	53259	201106AK.txt
Wed Aug 31 00:00:00 2011	55012	201107AK.txt
Fri Sep 30 00:00:00 2011	55012	201108AK.txt
Sun Oct 30 00:00:00 2011	53259	201109AK.txt
Wed Nov 30 00:00:00 2011	55012	201110AK.txt
Fri Dec 30 00:00:00 2011	53259	201111AK.txt
Tue Jan 31 00:00:00 2012	55012	201112AK.txt
Wed Feb 29 00:00:00 2012	55012	201201AK.txt
Thu Mar 29 00:00:00 2012	51506	201202AK.txt
Mon Apr 30 00:00:00 2012	55012	201203AK.txt
Wed May 30 00:00:00 2012	53259	201204AK.txt
Sat Jun 30 00:00:00 2012	55012	201205AK.txt
Mon Jul 30 00:00:00 2012	53259	201206AK.txt
Fri Aug 31 00:00:00 2012	55012	201207AK.txt
Sun Sep 30 00:00:00 2012	55012	201208AK.txt
Tue Oct 30 00:00:00 2012	53259	201209AK.txt
Fri Nov 30 00:00:00 2012	55012	201210AK.txt
Sun Dec 30 00:00:00 2012	53259	201211AK.txt

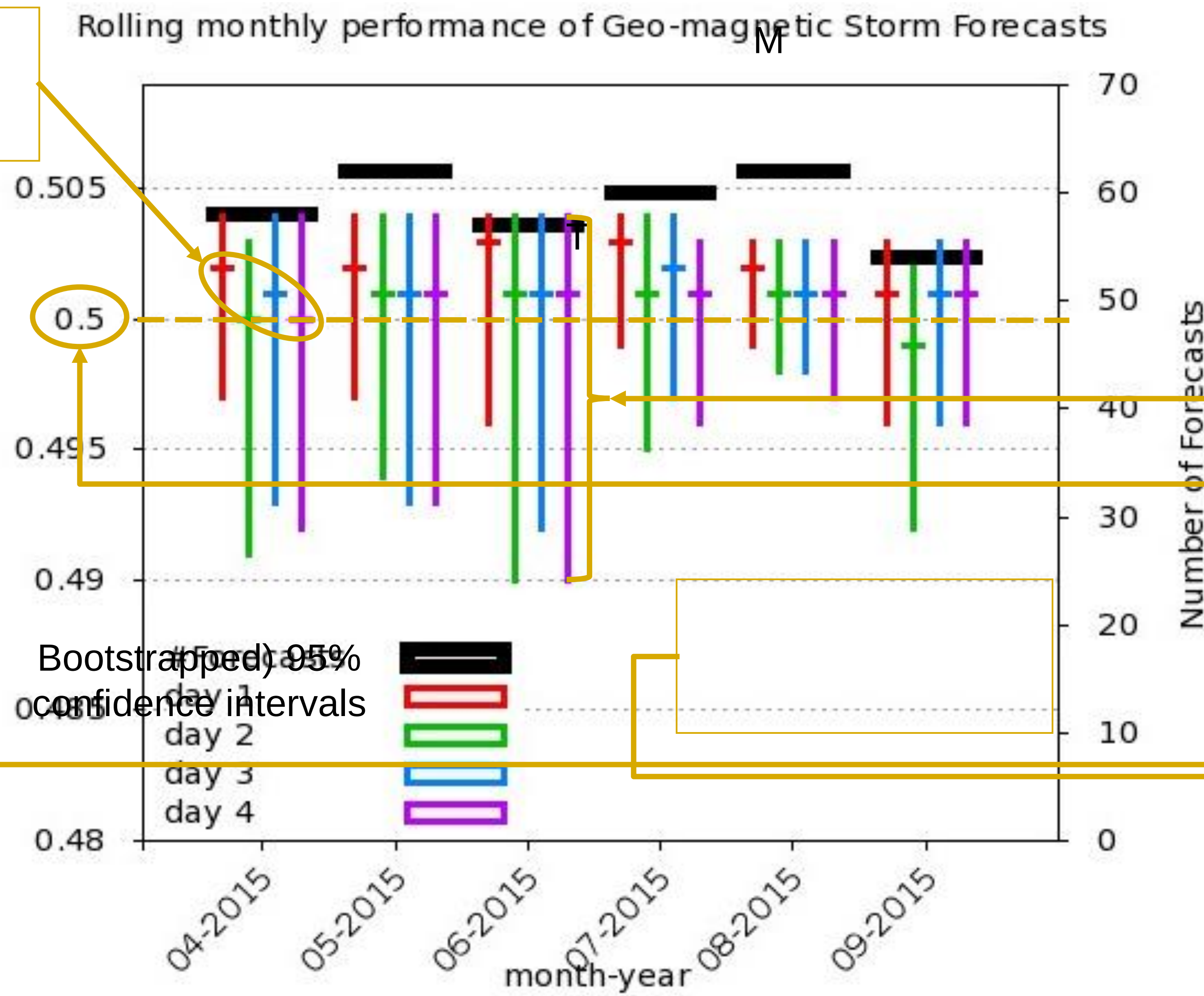
G-level climatology benchmark

5-year G-level climatology



How does the GM storm forecast compare with simply forecasting these probabilities every day?....

Kp forecast v climatology



Bootstrapped 95% confidence intervals

Score of 0.5:

Skill of forecast

= Skill of reference

Summary: adapting a meteorological system for Kp

Conclusions so far...

- Median RPSS on day 1 very slightly > RPSS on days 2-4
  - but no evidence (at 95% level) to suggest any difference
- Almost all median values > 0.5
  - but no evidence (at 95% level) to suggest forecast better than climatology

Analysis for the future...

- How do MO forecasts compare with SWPC/other forecasts?
- How do the Markov chain 1st guess GM Storm forecasts compare?

In the mean-time...

- Near real-time verification of Kp forecasts are available to forecasters.

6. Verification of flare forecasts

- Will develop in-house flare verification in similar manner to K_p (e.g., ranked probability scores).

Numerous collaborative projects also ongoing:

- International Space Environment Services
  - Internationally consistent verification.
  - ROC curves and reliability diagrams.
- NASA CCMC Flare scoreboard:
  - Visualisation of real-time forecasts with verification.
- FLARECAST project:
  - Automated ensemble forecasting system will be compared with our current forecasting methods.
  - Met Office involvement with verification and dissemination.

X Ray Flares / Probability (Exceedance)	Level	Past 24 Hours (Yes/No)	Day 1 (00-24 UTC) (%)	Day 2 (00-24 UTC) (%)	Day 3 (00-24 UTC) (%)	Day 4 (00-24 UTC) (%)
Active	R1-R2 M Class	N	5	10	10	15
Very Active	R3 to R5 X Class	N	1	1	1	5

7. Summary

- MOSWOC produce twice daily forecasts containing CME arrival time predictions and probabilistic 4-day forecasts for geomagnetic storms, flares and electron/proton events.

- Initial verification has focused on:
    - CME arrival time prediction
    - Kp probabilistic forecasts
    - Adapting a near real-time verification system for space weather purposes

- Verification of CME arrival time forecasts show good agreement with CCMC.

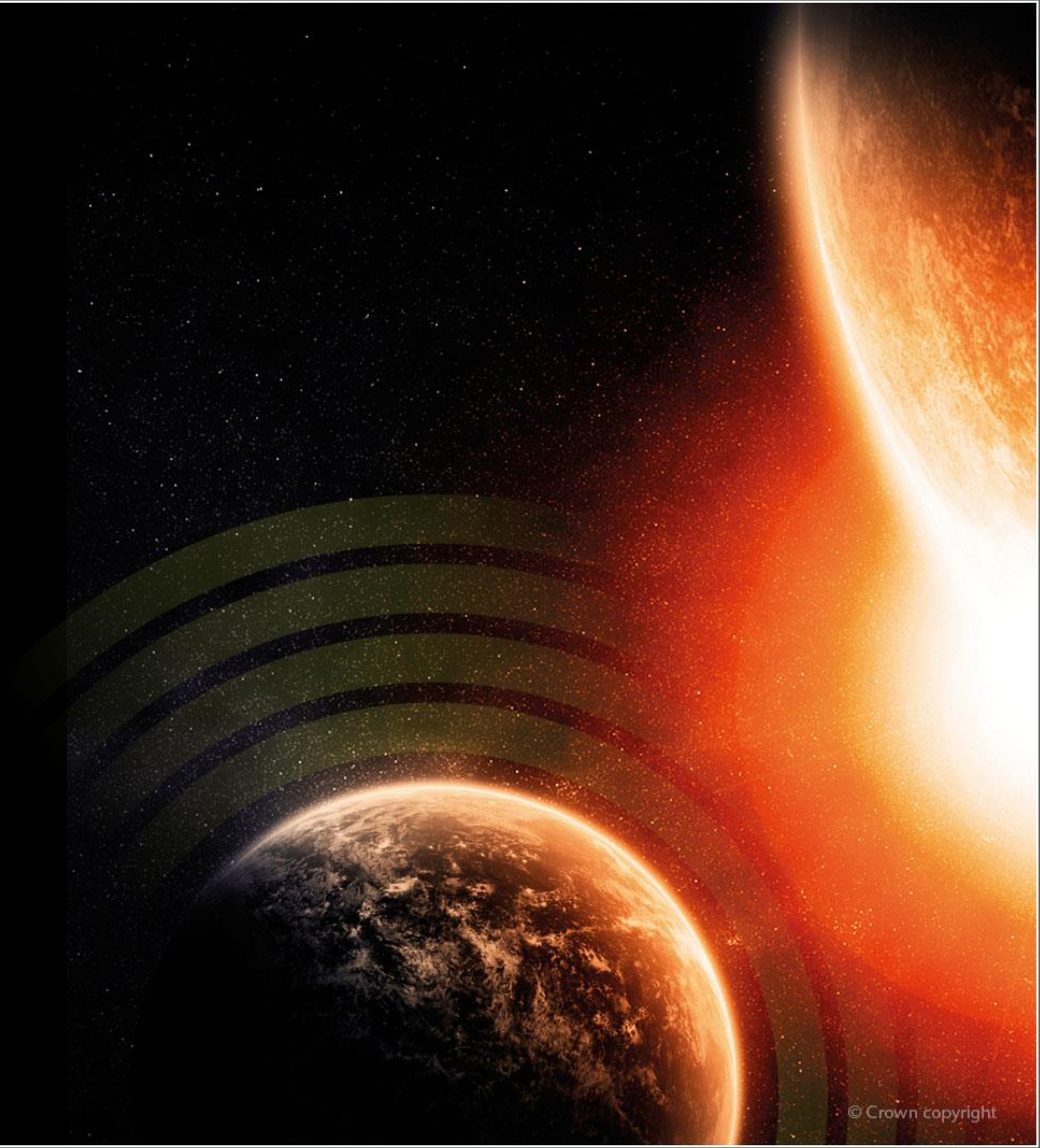
- Assessment of geomagnetic storm forecast skills shows:
    - Difficulty of defining climatology or Markov chain.
    - Markov chain can do better than standard forecast for days 1-2 for low G events.
    - Difficulty in assessing higher G events due to their rarity.
    - More research still needed.

Summary

- Adapting a terrestrial verification system for geomagnetic storms.
  - Used Ranked Probability Skill Score to compare performance of MOSWOC forecasts against climatology.
  - Real time verification system will lead to benefit for MOSWOC forecasters.

- Met Office are involved with ISES, FLARECAST & CCMC Flare Scoreboard.
- Terrestrial weather verification resources are feely available, e.g MET (Model Evaluation Tools).

Met Office



Thank you

Terrestrial weather verification resources

WMO Working Group under the World Weather Research Program (WWRP) & Working Group on Numerical Experimentation (WGNE)

Activities:

- Verification research
- Training
- Workshops & tutorials
- Publications on ‘best practices’

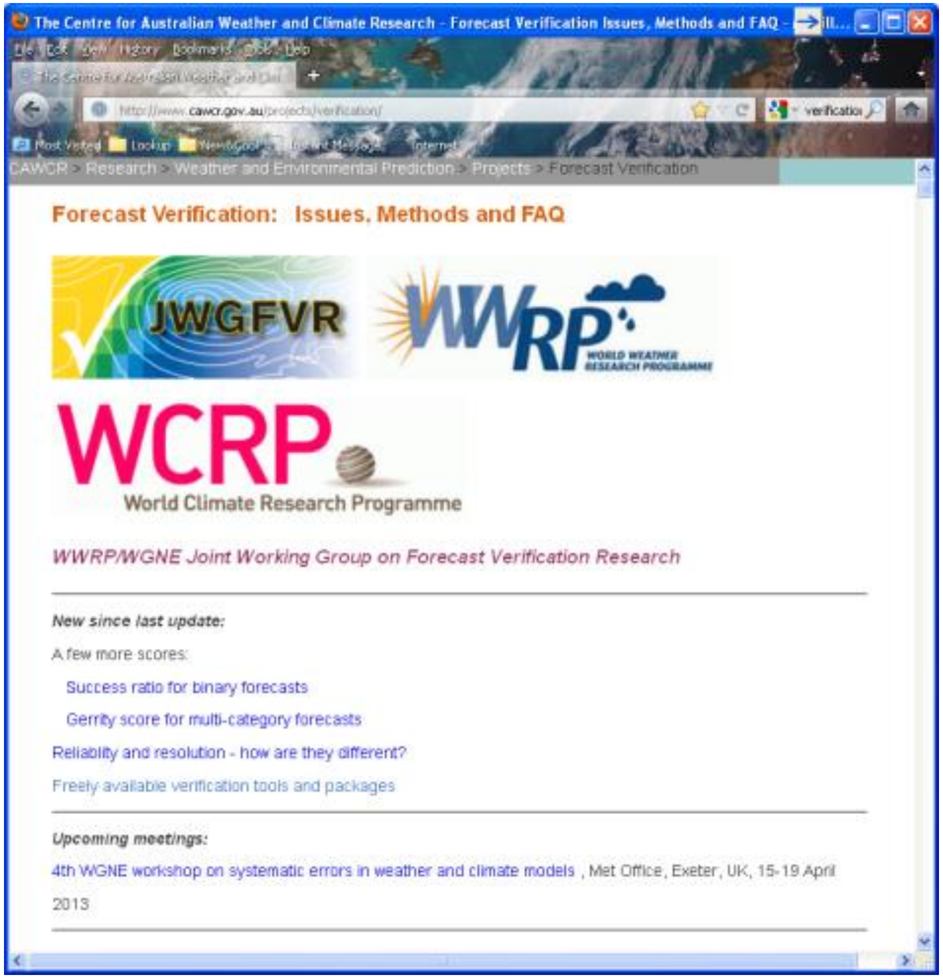
http://www.wmo.int/pages/prog/arep/wwrp/new/Forecast_Verification.html

Info from: B. Brown. ISES Verification Workshop, Apr. 2015

Terrestrial weather verification resources

Website maintained by WMO verification Working Group includes:

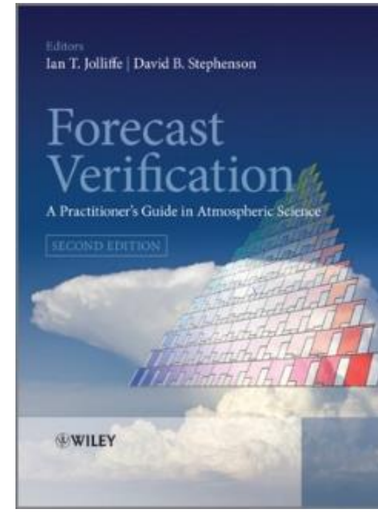
- Methods (brief definitions)
- Verification issues
- FAQs
- Links and references
- Verification discussion group



http://www.cawcr.gov.au/projects/verification/

Info from: B. Brown. ISES Verification Workshop, Apr. 2015

Terrestrial weather verification resources

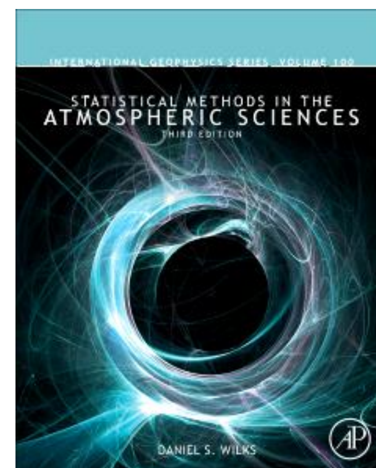


Papers:

- Casati et al. (2008), *Forecast verification: current status and future directions*, Meteorological Applications, **15**, 3-18.
- Ebert et al. (2013), *Progress and challenges in forecast verification,* Meteorological Applications, **20**, 130-139.

Books:

- Jolliffe and Stephenson (2012): *Forecast Verification: a practitioner's guide*, Wiley & Sons.
- Stanski, Burrows, Wilson (1989) *Survey of Common Verification Methods in Meteorology* (available at http://www.cawcr.gov.au/projects/verification/)
- Wilks (2011), *Statistical Methods in Atmospheric Science*, Academic press.

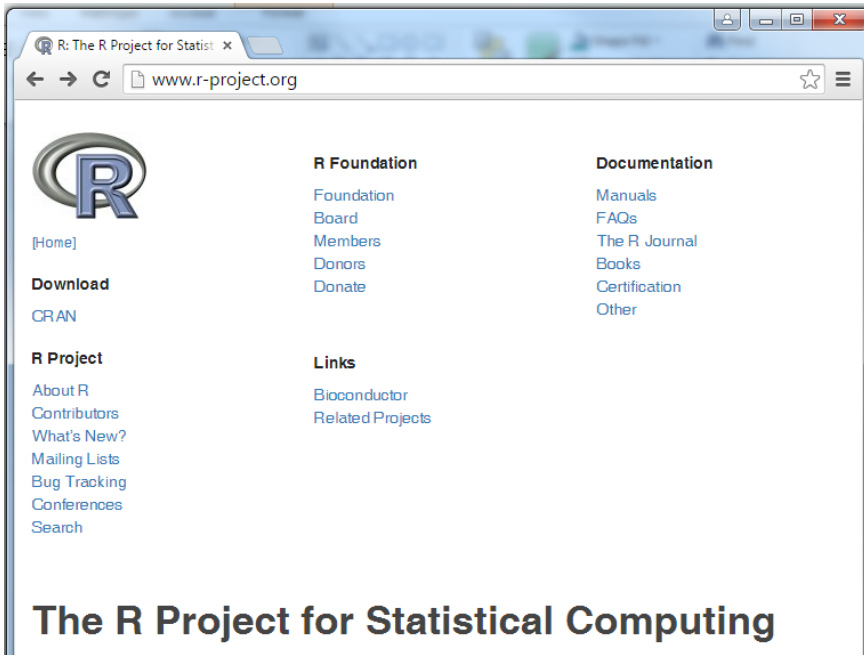


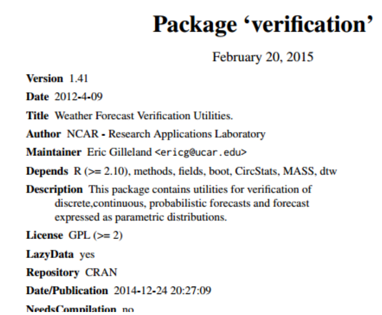
Info from: B. Brown. ISES Verification Workshop, Apr. 2015

Terrestrial weather verification tools: R verification libraries

R verification libraries:

- Freely available statistics packages.
- http://www.r-project.org/
- Maintained & supported by NCAR.





Package 'verification'

February 20, 2015

Version 1.41

Date 2012-4-09

Title Weather Forecast Verification Utilities.

Author NCAR - Research Applications Laboratory

Maintainer Eric Gilleland <ericg@ucar.edu>

Depends R (>= 2.10), methods, fields, boot, CircStats, MASS, dtw

Description This package contains utilities for verification of discrete,continuous, probabilistic forecasts and forecast expressed as parametric distributions.

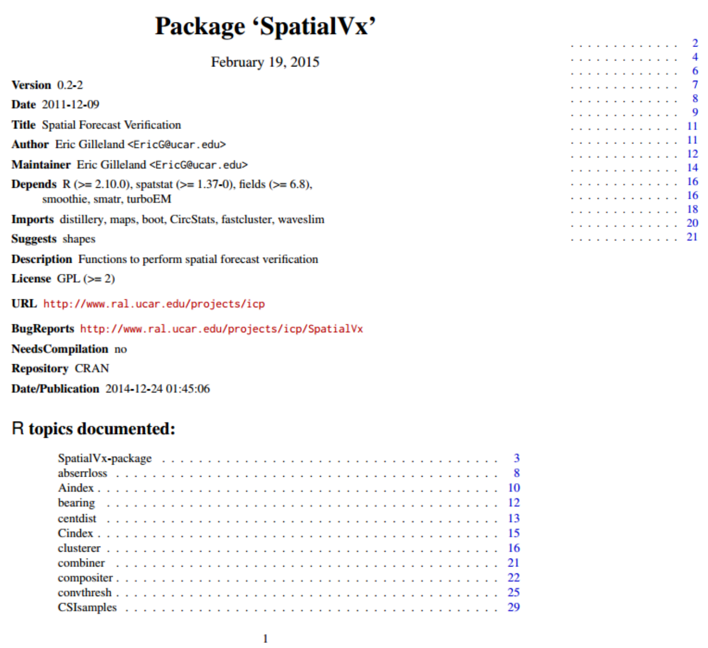
License GPL (>= 2)

LazyData yes

Repository CRAN

Date/Publication 2014-12-24 20:27:09

NeedsCompilation no



Package 'SpatialVx'

February 19, 2015

Version 0.2-2

Date 2011-12-09

Title Spatial Forecast Verification

Author Eric Gilleland <EricG@ucar.edu>

Maintainer Eric Gilleland <EricG@ucar.edu>

Depends R (>= 2.10.0), spatstat (>= 1.37-0), fields (>= 6.8), smoothie, smatr, turboEM

Imports distillery, maps, boot, CircStats, fastcluster, waveslim

Suggests shapes

Description Functions to perform spatial forecast verification

License GPL (>= 2)

URL http://www.ral.ucar.edu/projects/icp

BugReports http://www.ral.ucar.edu/projects/icp/SpatialVx

NeedsCompilation no

Repository CRAN

Date/Publication 2014-12-24 01:45:06

R topics documented:

1

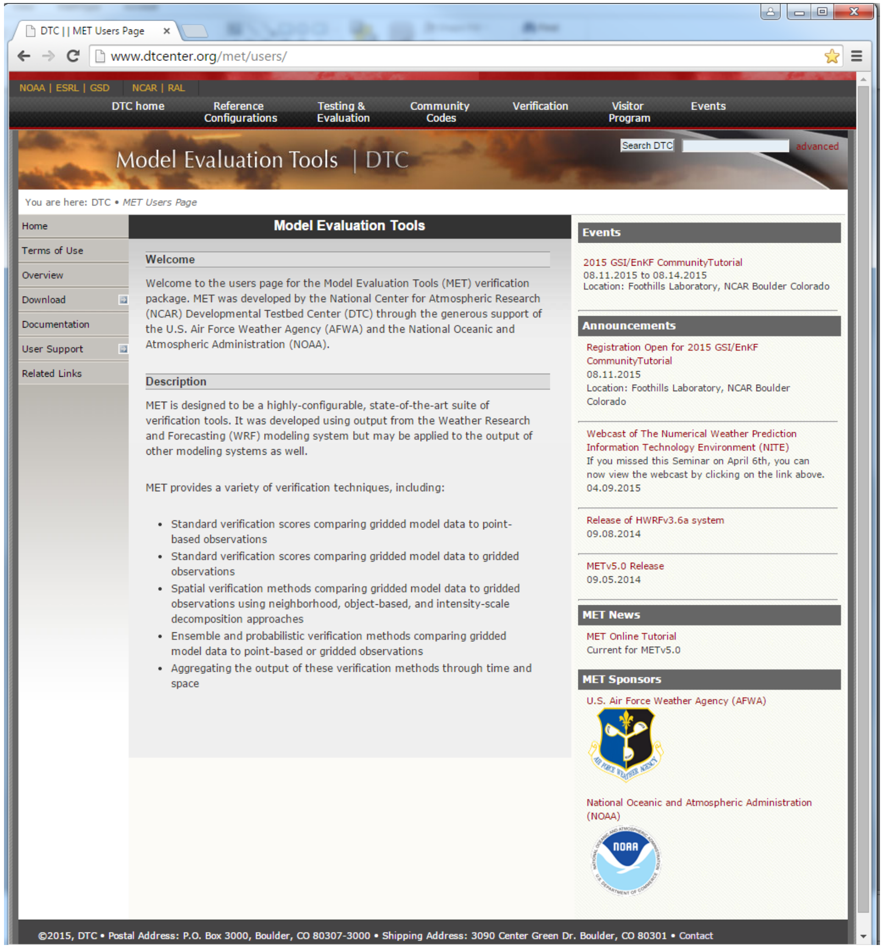
Info from: B. Brown. ISES Verification Workshop, Apr. 2015

Terrestrial weather verification tools: MET

http://www.dtcenter.org/met/users/

Model Evaluation Tools (MET):

- Forecast evaluation tools
- Implemented & supported by the Developmental Testbed Center (DTC) & Joint Numerical Testbed Program at NCAR/RAL
- Includes a suite of standard stats, non-traditional stats (e.g. spatial methods)
- Designed to undertake systematic evaluations
- Has a database & display system for aggregating & plotting data
- Provides a standardized evaluation platform for cross-institution comparisons
- Freely available
- Highly configurable
- Supported via the web & “live” user tutorials



Info from: B. Brown. ISES Verification Workshop, Apr. 2015