British Geological Survey
NATURAL ENVIRONMENT RESEARCH COUNCIL

Forecast Evaluation as Applied to Geomagnetic Activity Categories

Ellen Clarke and Alan Thomson

Outline

- The daily 1-3 day ahead geomagnetic forecast
  - WHAT we try to do and WHO we do it for
- Forecast verification against benchmark
  - Year by year comparisons (2000 to 2013)
  - Comparisons between individual forecasters
- Investigation of performance measures – skill scores
  - Important for on-going automated evaluation
  - What is the most appropriate for this type of forecast?
- Future Plans
  - Further comparisons and feedback to forecasters
  - Revision of the service and user perspectives

What is the Geomagnetic Activity Forecast?

- Predictions are of global average geomagnetic activity levels
- Forecasts are issued every weekday before noon
  - Weekends are not included - not a commercially funded service
- Predictions are for 1, 2 and 3 days (intervals) ahead
- Forecast intervals are 24 hours from noon to noon (UT)
  - More likely to capture storms in the local UK night time sector
- Use public domain space weather observations, models, alerts and forecasts
  - Tap into the specific expertise of various groups around the world
- There are four activity levels to choose from (based on *Ap*)
  - MAJOR STORM, MINOR-STORM, ACTIVE or QUIET-UNSETTLED

What is the Geomagnetic Activity Forecast?

- Predictions are of global average geomagnetic activity levels
- Forecasts are issued every weekday before noon
  - Weekends are not included - not a commercially funded service
- Predict ad
- Forecas n (UT)
  - Mor e sector
- Use pul models, alerts and forecasts
  - Tap into the specific expertise of various groups around the world
- There are four activity levels to choose from (based on *Ap*)
  - MAJOR STORM, MINOR-STORM, ACTIVE or QUIET-UNSETTLED

ACTIVITY CLASS	Daily Planetary Activity Level (Ap)
QUIET – UNSETTLED	<=15
ACTIVE	16-29
MINOR STORM	30-49
MAJOR STORM	>=50

The Daily Geomagnetic Activity Forecast

British Geological Survey
NATURAL ENVIRONMENT RESEARCH COUNCIL

MAGIC

nationalgrid

Magnetic Activity | Solar Wind | 3-day Forecast | GIC Estimates | Home

Three-Day Global Geomagnetic Activity Forecast

Forecast Interval (GMT)	Forecast Global Activity Level
Noon 14-NOV-2012 to Noon 15-NOV-2012	QUIET-UNSETTLED
Noon 15-NOV-2012 to Noon 16-NOV-2012	QUIET-UNSETTLED
Noon 16-NOV-2012 to Noon 17-NOV-2012	QUIET-UNSETTLED

ADDITIONAL COMMENTS

Yesterday averaged MINOR-STORM globally with periods of MAJOR-STORM recorded between 00:00-09:00UT. In the UK, MINOR-STORM levels were recorded between 01:00-04:00UT at each observatory. This activity was due to a prolonged period of strongly southward pointing IMF in the solar wind following on from the CME shock arrival on the 12th.

The CMEs assocated with the M-class flares from region 11613 yesterday are not expected to be geoeffective. QUIET-UNSETTLED conditions are expected, with a chance of isolated ACTIVE periods as current activity subsides.

TODAYS FORECASTER: Sarah Reay
TIME OF FORECAST : Wednesday, 14 November 2012 11:37:05 GMT
BGS CONTACT PHONE: 0131 667 1000 (switchboard)

Who gets the Geomagnetic Activity Forecast?

Recipients of the daily forecast (over the years) include:

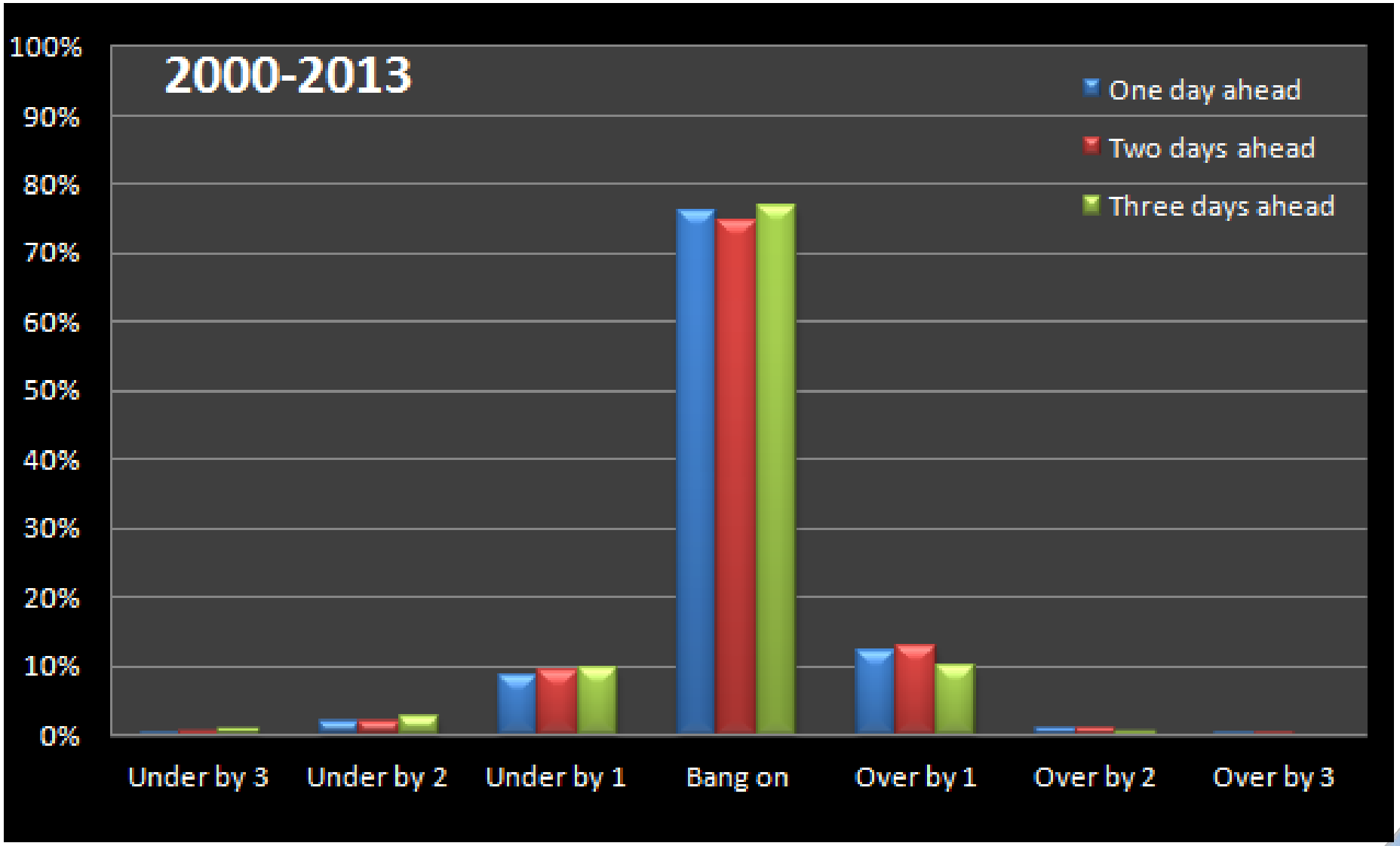
- **Met Office (UK)**
  - Part of the National Hazards Partnerships' Daily Hazard Assessment
  - Informing UK Cabinet Office Civil Contingencies Secretariat
- **Power companies concerned about Geomagnetically Induced Currents**
  - E.g. Scottish Power and National Grid
- **Oil and Gas industry companies using directional drilling techniques**
  - E.g. Halliburton Sperry Drilling and Baker Hughes
- **Geophysical prospecting companies**
- **Organisations working on instrument calibrations**
  - E.g. National Physical Laboratory and Bartington Instruments
- **Geomagnetism colleagues and partners**
  - Planning for field work or absolute observations at observatories

How can we verify the accuracy of our forecasts?

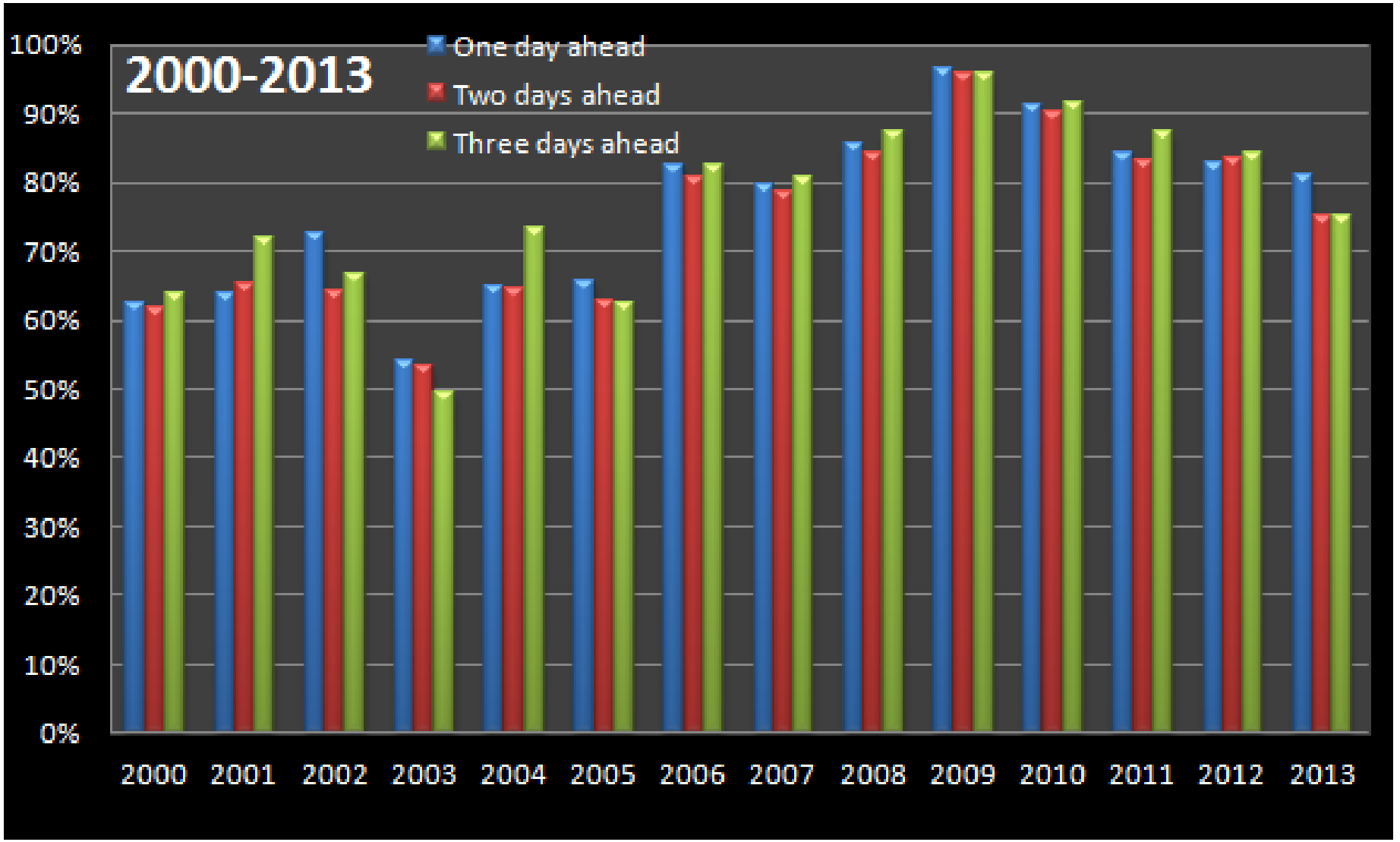
BRITISH GEOLOGICAL SURVEY: NATIONAL GEOMAGNETIC SERVICE
GEOMAGNETIC ACTIVITY FORECAST FOR SPERRY DRILLING

Forecast Interval (GMT)	Forecast Global Activity Level
Noon 28-OCT-2003 to Noon 29-OCT-2003	ACTIVE
Noon 29-OCT-2003 to Noon 30-OCT-2003	MINOR-STORM
Noon 30-OCT-2003 to Noon 31-OCT-2003	MINOR-STORM

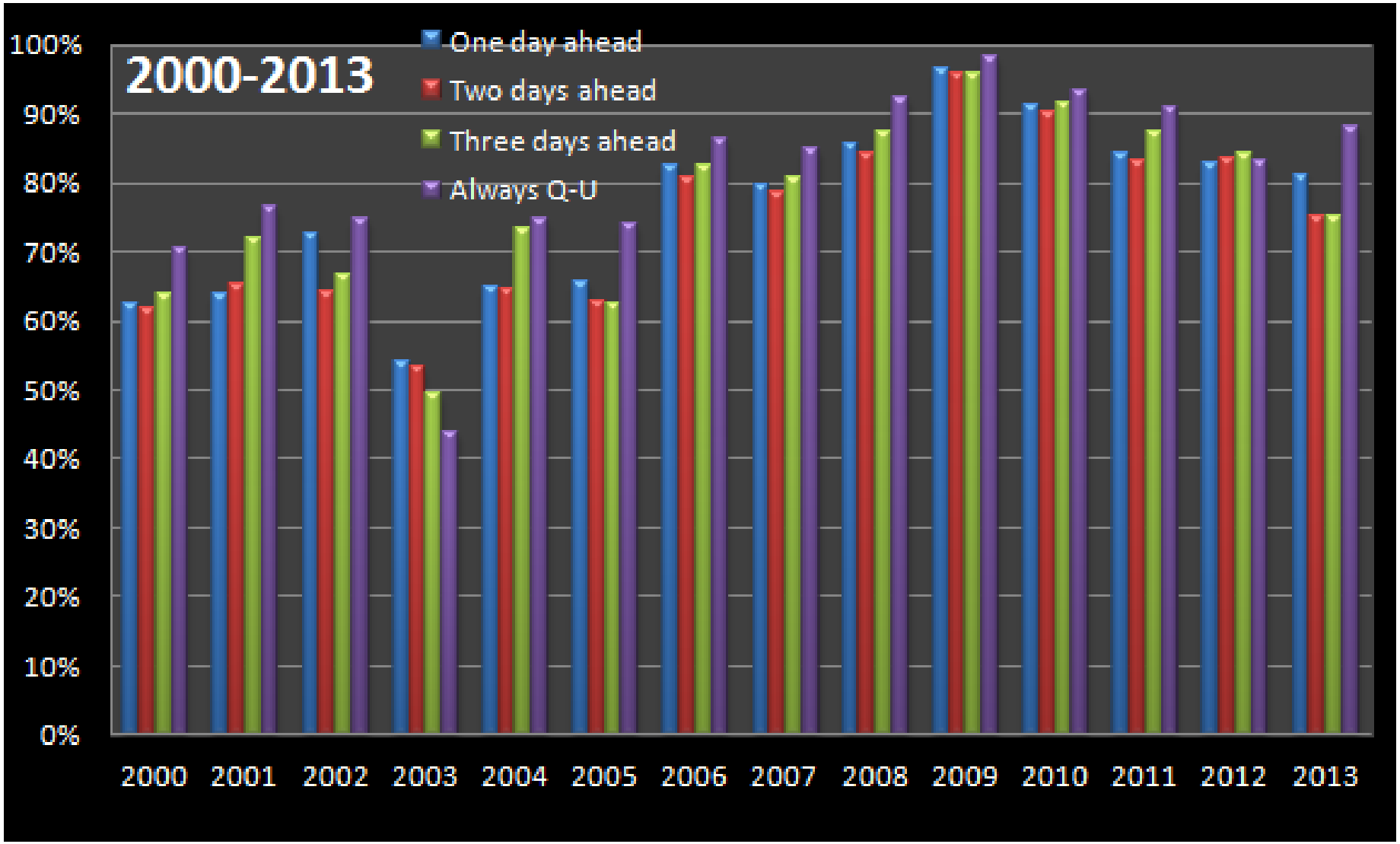
Simple Verification Statistic (% correct)



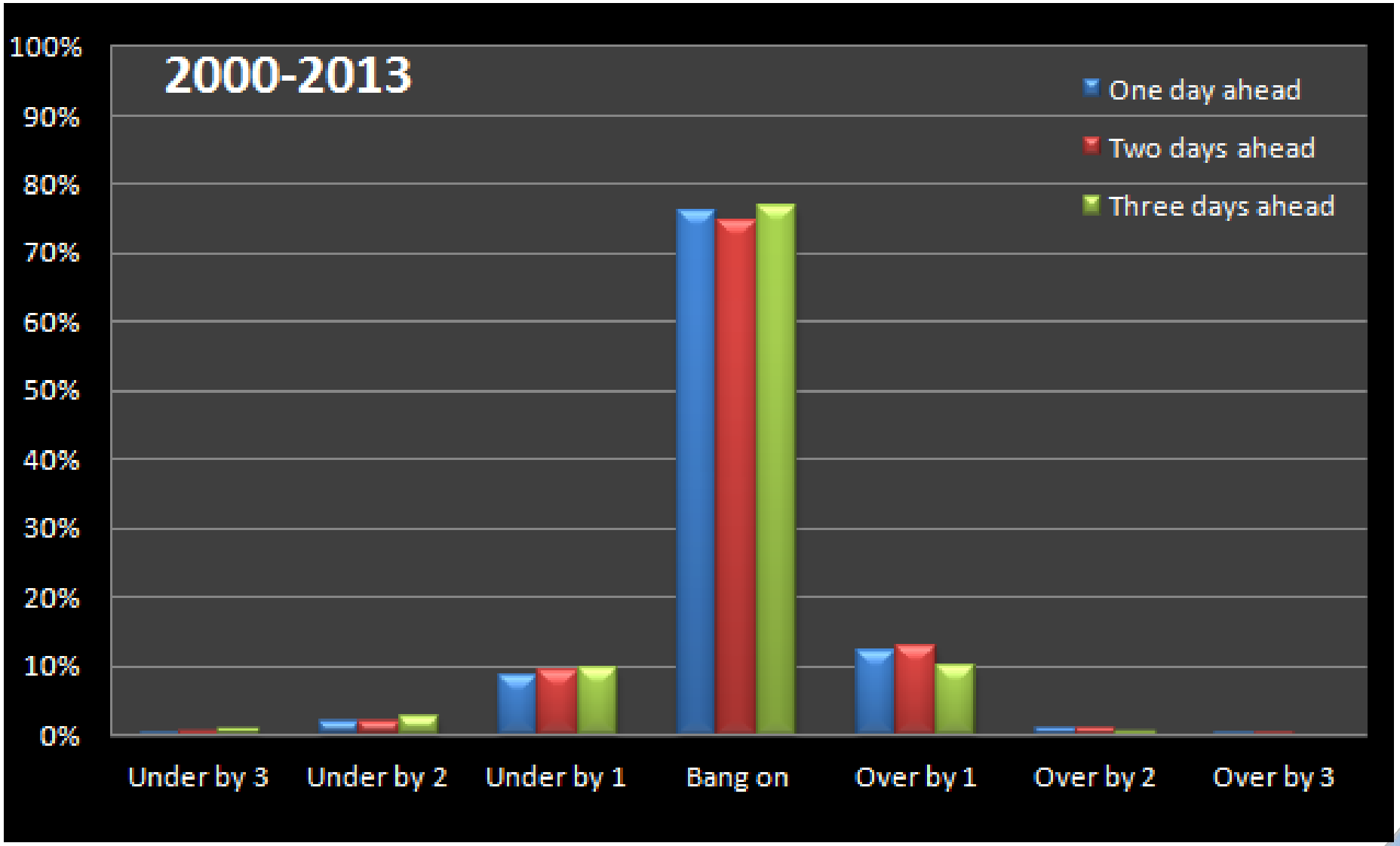
Simple Verification Statistic - % Correct by Year



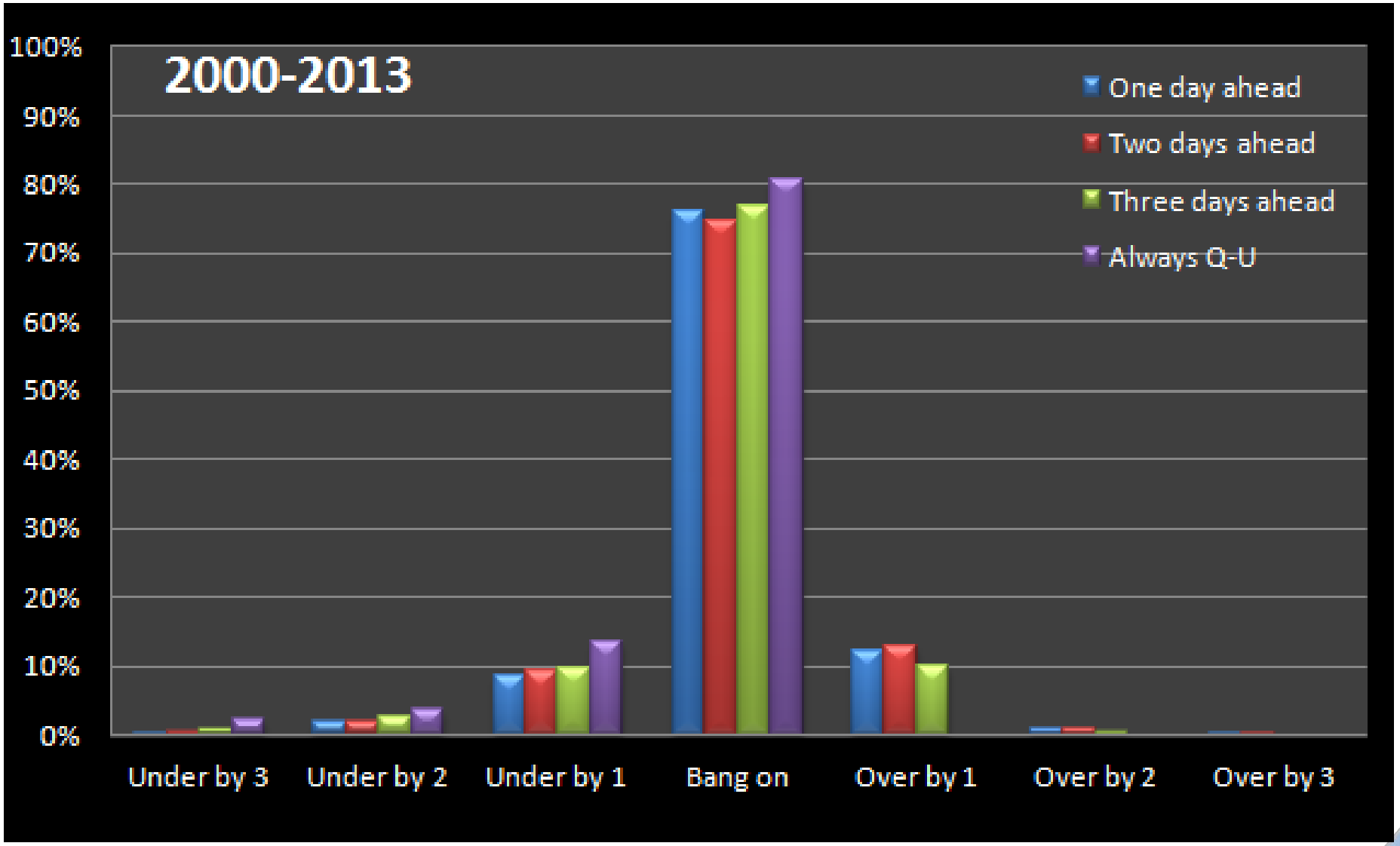
Simple Verification Statistic - % Correct by Year



Simple Verification Statistic (% correct)



Simple Verification Statistic (% correct)



Forecast Verification using Skill Scores

- Binary Events - > Two-dimensional Contingency table
  (E.g. MAGNETIC STORM or NO MAGNETIC STORM)

2x2 contingency table		Magnetic Storm Observed		Marginal Total
		Yes	No	
Magnetic Storm Forecast	Yes	A	B	A + B
	No	C	D	C + D
Marginal Total		A + C	B + D	n (A+B+C+D)

- Many performance measures can be determined using the contingency table entries
- 3 properties we want out of a skill measure for Space Weather are:
  1. Equitability
  2. Discourages hedging
  3. Usefulness for relatively rare events

No single measure designed (so far) that is strong in all three

Forecast Verification using Skill Scores

2x2 contingency table		Magnetic Storm Observed		Marginal Total
		Yes	No	
Magnetic Storm Forecast	Yes	A	B	A + B
	No	C	D	C + D
Marginal Total		A + C	B + D	n (A+B+C+D)

- **Peirce Skill Score** (PSS) also known as True Skill Statistic (TSS)

  PSS= (AD-BC)/(A+C)(B+D)

- **Gilbert Skill Score** (GSS) also known as Equitable Threat Score (ETS)

  GSS= (A-CH)/(A+B+C-CH)

  where CH (chance hit) = (A+B)(A+C)/n

- **Heidke skill score** (HSS)

  HSS= (A+D-E)/(n-E)

  where E (correct random forecast) = [(A+B)(A+C)+(B+D)(C+D)]/n

Forecasters and Skill Scores (Binary : STORM / NO STORM)

Geomag Forecasters		Storm Observed		Marginal Total
		Yes	No	
Storm Forecast	Yes	63	74	137
	No	139	3349	3488
Marginal Total		202	3423	3625

→PSS=0.29

Benchmark		Storm Observed		Marginal Total
		Yes	No	
Storm Forecast	Yes	62	111	173
	No	140	3312	3452
Marginal Total		202	3423	3625

→PSS=0.27

→HSS=0.34

→GSS=0.21

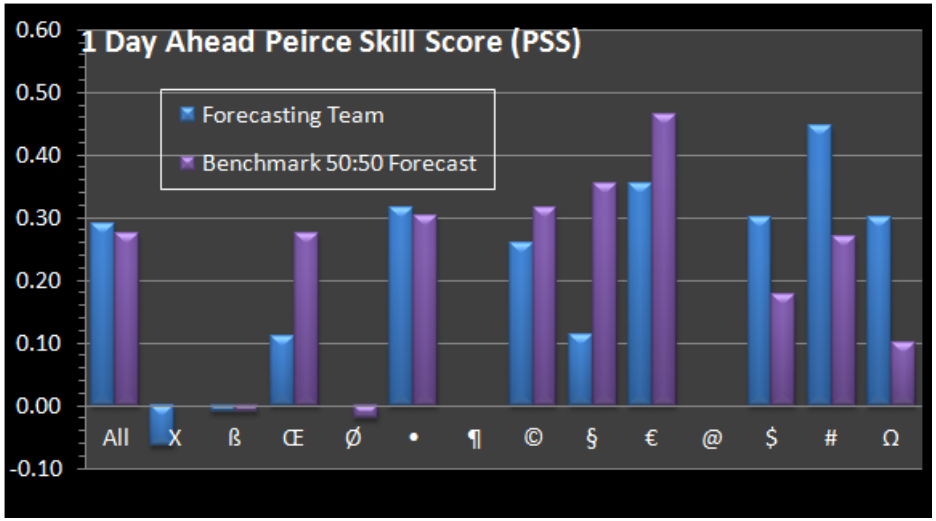
→HSS=0.29

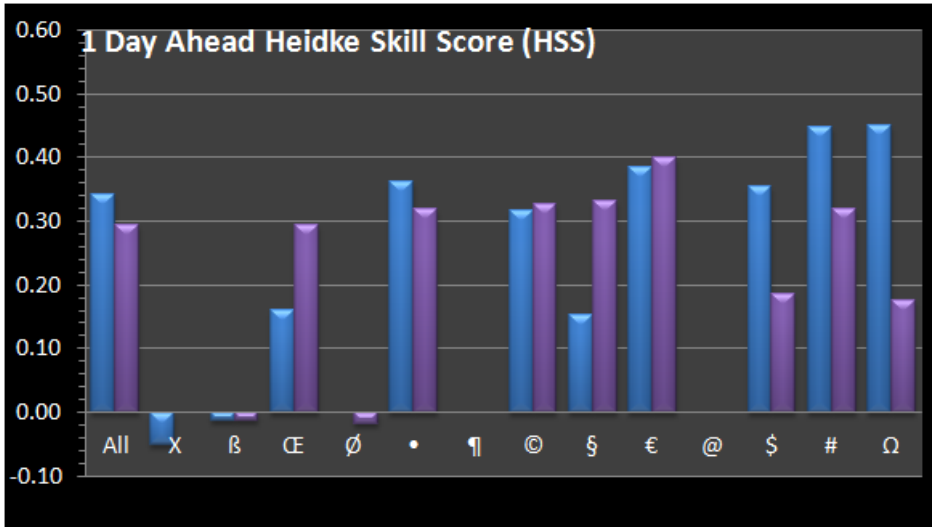
→GSS=0.17

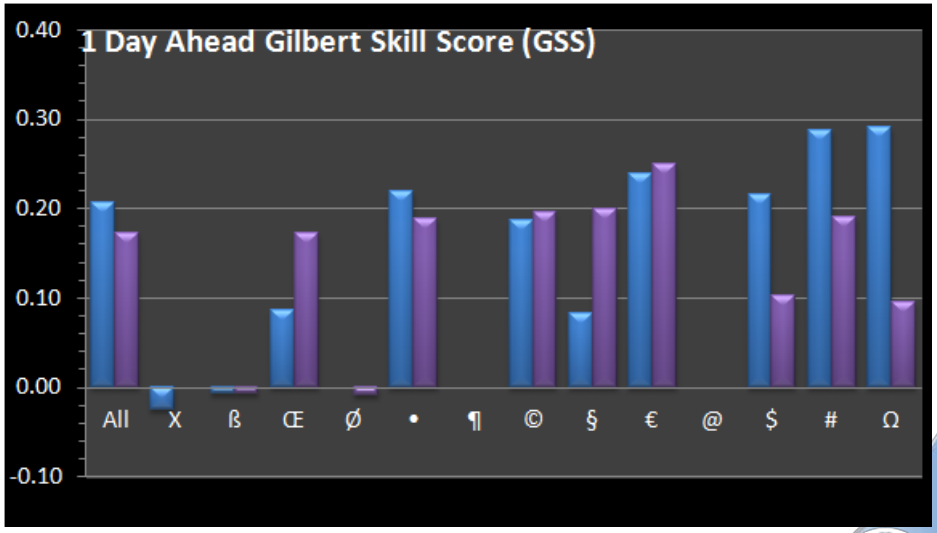
Forecasters and Skill Scores (Binary : STORM / NO STORM)

Geomag Forecasters		Storm Observed		Marginal Total
		Yes	No	
Storm Forecast	Yes	63	74	137
	No	139	3349	3488
Marginal Total		202	3423	3625

Benchmark		Storm Observed		Marginal Total
		Yes	No	
Storm Forecast	Yes	62	111	173
	No	140	3312	3452
Marginal Total		202	3423	3625







Forecast Verification including BIAS

2x2 contingency table		Magnetic Storm Observed		Marginal Total
		Yes	No	
Magnetic Storm Forecast	Yes	A	B	A + B
	No	C	D	C + D
Marginal Total		A + C	B + D	n (A+B+C+D)

Bias score or frequency bias

BIAS=(A+B)/(A+C)

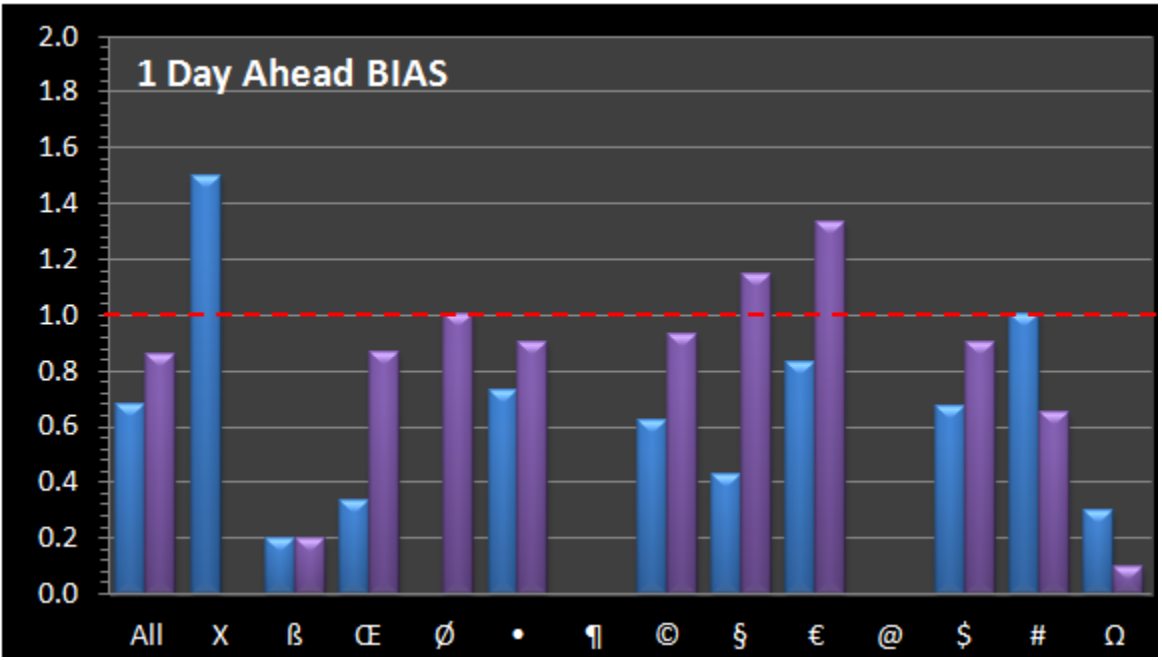
Geomag Forecasters		Storm Observed		Marginal Total
		Yes	No	
Storm Forecast	Yes	63	74	137
	No	139	3349	3488
Marginal Total		202	3423	3625

BIAS = 0.68

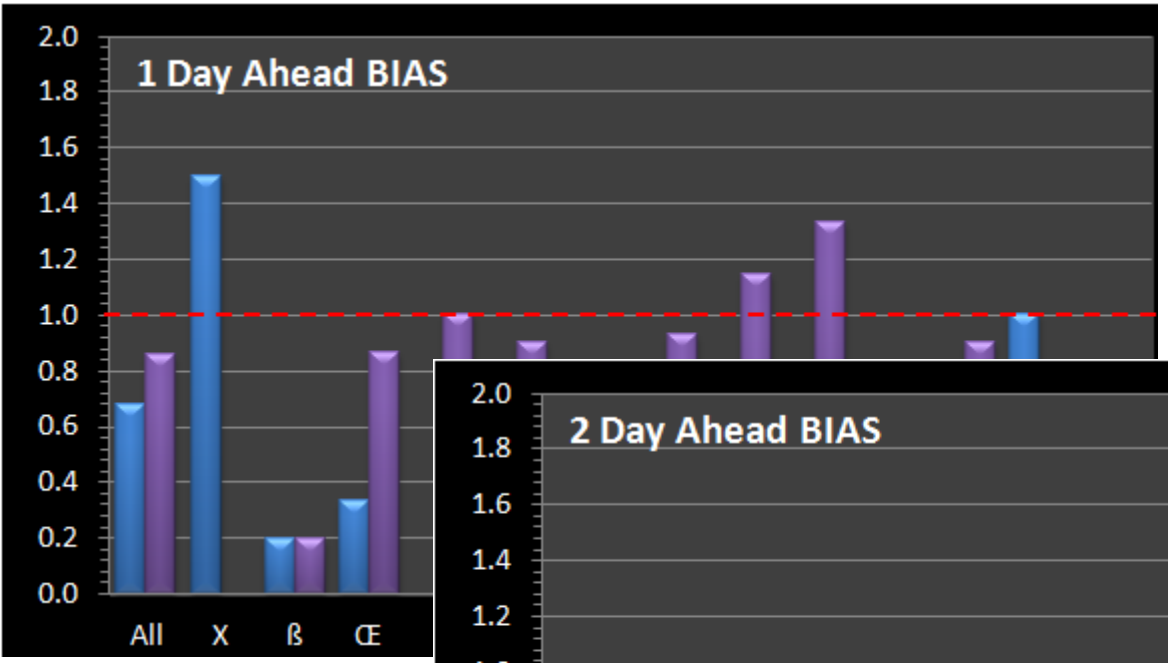
Benchmark		Storm Observed		Marginal Total
		Yes	No	
Storm Forecast	Yes	62	111	173
	No	140	3312	3452
Marginal Total		202	3423	3625

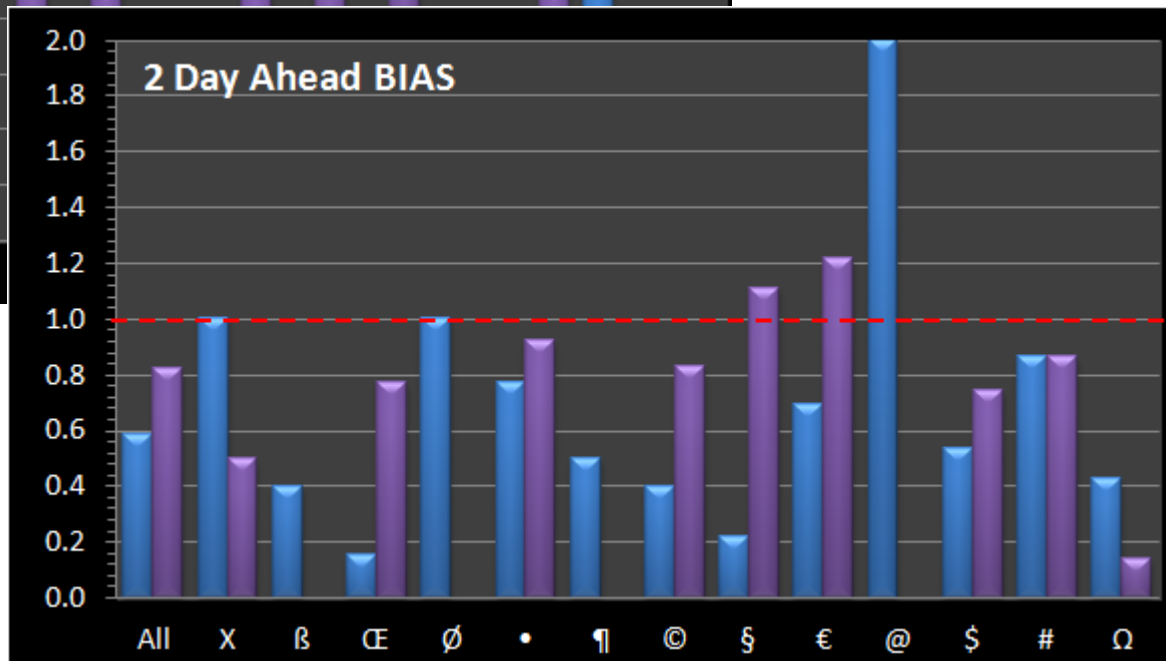
BIAS = 0.86

Individual Forecasters and BIAS

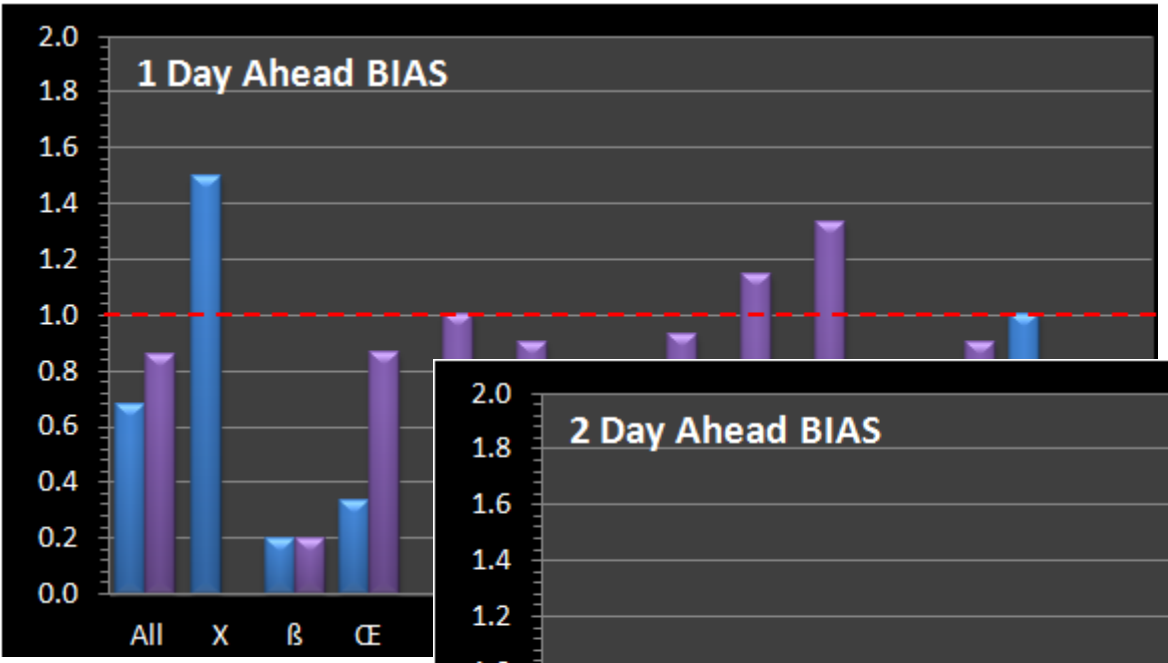


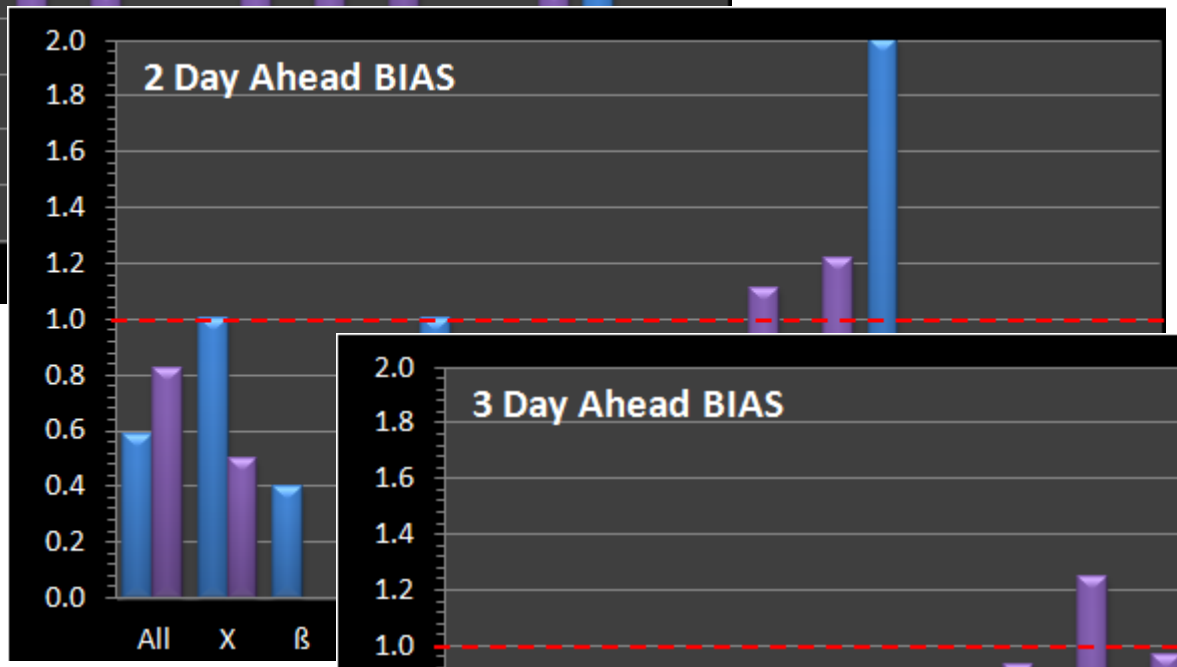
Individual Forecasters and BIAS

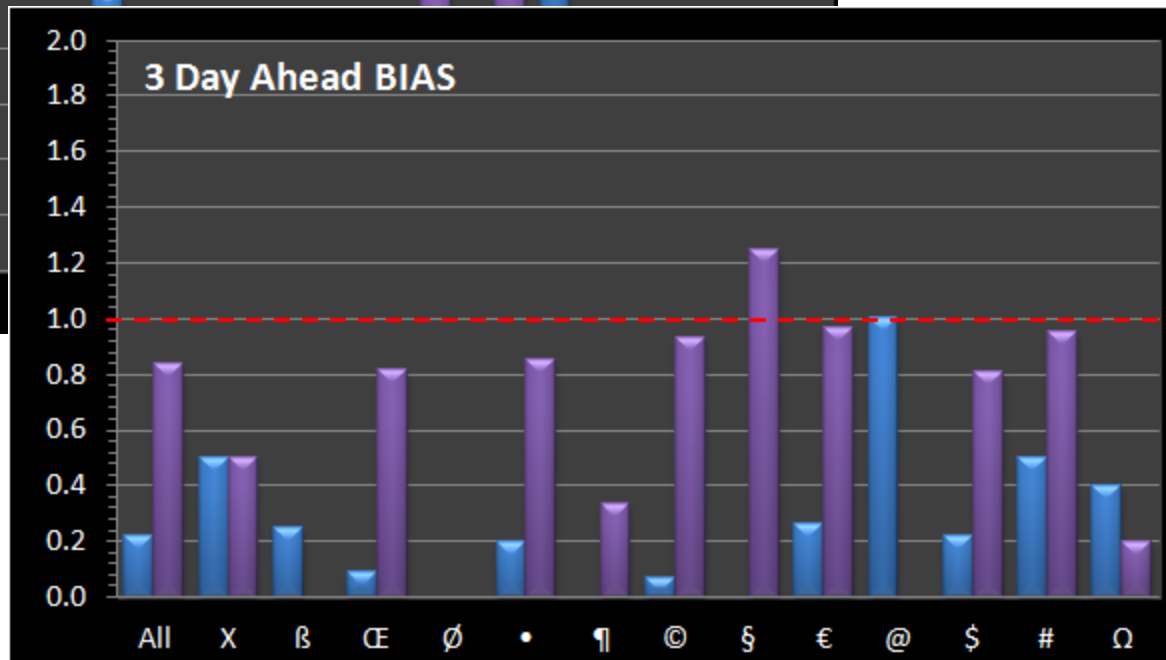




Individual Forecasters and BIAS







Multi-Category Equitable Skill Scores

k x k contingency table with elements A_{ij}		Forecast (F) Category (j)				Marginal Total
		1	2	...	k	
Observed (O) Category (i)	1	$A_{1,1}$	$A_{1,2}$	$A_{1,...}$	$A_{1,k}$	$n(O_1)$
	2	$A_{2,1}$	$A_{2,2}$	$A_{2,..}$	$A_{2,k}$	$n(O_2)$
	...	$A_{...,1}$	$A_{...,2}$	$A_{..,..}$	$A_{...,k}$	$n(O_{..})$
	k	$A_{k,1}$	$A_{k,2}$	$A_{k,...}$	$A_{k,k}$	$n(O_k)$
Marginal Total		$n(F_1)$	$n(F_2)$	$n(F...)$	$n(F_k)$	N

4 x 4 probability matrix (P) where $p_{ij} = A_{ij}/N$		Forecast (F) Category (j)				Climatological probability
		1	2	3	4	
Observed (O) Category (i)	1	$p_{1,1}$	$p_{1,2}$	$p_{1,3}$	$p_{1,4}$	$p(O_1)$
	2	$p_{2,1}$	$p_{2,2}$	$p_{2,3}$	$p_{2,4}$	$p(O_2)$
	3	$p_{3,1}$	$p_{3,2}$	$p_{3,3}$	$p_{3,4}$	$p(O_3)$
	4	$p_{4,1}$	$p_{4,2}$	$p_{4,3}$	$p_{4,4}$	$p(O_4)$
Forecast probability		$p(F_1)$	$p(F_2)$	$p(F_3)$	$p(F_4)$	1

Gandin and Murphy (1992) devised a way of extending equitable skill scores to more than two categories. The general formula is

$$ESS = \sum_{i=1}^{k}\sum_{j=1}^{k} p_{ij}s_{ij}$$

s_{ij} are the elements of a reward-penalty matrix Known as the scoring matrix (S).

Scoring Matrix for Multi-Category Equitable Skill Scores

Gerrity (1992) extended this further and derived formulas for populating the S matrix for >3 categories.

4 x 4 probability matrix (P) where $p_{ij} = A_{ij}/N$		Forecast (F) Category (j)			
		1	2	3	4
Observed (O) Category (i)	1	$p_{1,1}$	$p_{1,2}$	$p_{1,3}$	$p_{1,4}$
	2	$p_{2,1}$	$p_{2,2}$	$p_{2,3}$	$p_{2,4}$
	3	$p_{3,1}$	$p_{3,2}$	$p_{3,3}$	$p_{3,4}$
	4	$p_{4,1}$	$p_{4,2}$	$p_{4,3}$	$p_{4,4}$

$$D(n) \equiv \frac{1 - \sum_{r=1}^{n} p(r)}{\sum_{r=1}^{n} p(r)}$$

$$R(n) = \frac{1}{D(n)}$$

Scoring Matrix for Multi-Category Equitable Skill Scores

Gerrity (1992) extended this further and derived formulas for populating the S matrix for >3 categories.

4 x 4 Scoring Matrix (S)		Forecast (F) Category (j)			
		1	2	3	4
Observed (O) Category (i)	1	$s_{1,1}$	$s_{1,2}$	$s_{1,3}$	$s_{1,4}$
	2	$s_{2,1}$	$s_{2,2}$	$s_{2,3}$	$s_{2,4}$
	3	$s_{3,1}$	$s_{3,2}$	$s_{3,3}$	$s_{3,4}$
	4	$s_{4,1}$	$s_{4,2}$	$s_{4,3}$	$s_{4,4}$

$$D(n) \equiv \frac{1-\sum_{r=1}^{n} p(r)}{\sum_{r=1}^{n} p(r)}$$

$$R(n) = \frac{1}{D(n)}$$

$$s_{m,n} = \frac{1}{k-1}\left[\sum_{r=1}^{m-1} R(r) + \sum_{r=m}^{n-1}(-1) + \sum_{r=n}^{k-1} D(r)\right] \quad ; \quad n = (1,\ldots,k)$$

$$s_{n,n} = \frac{1}{k-1}\left[\sum_{r=1}^{n-1} R(r) + \sum_{r=n}^{k-1} D(r)\right] \quad ; \quad 1 \le m < k, \quad m < n \le k$$

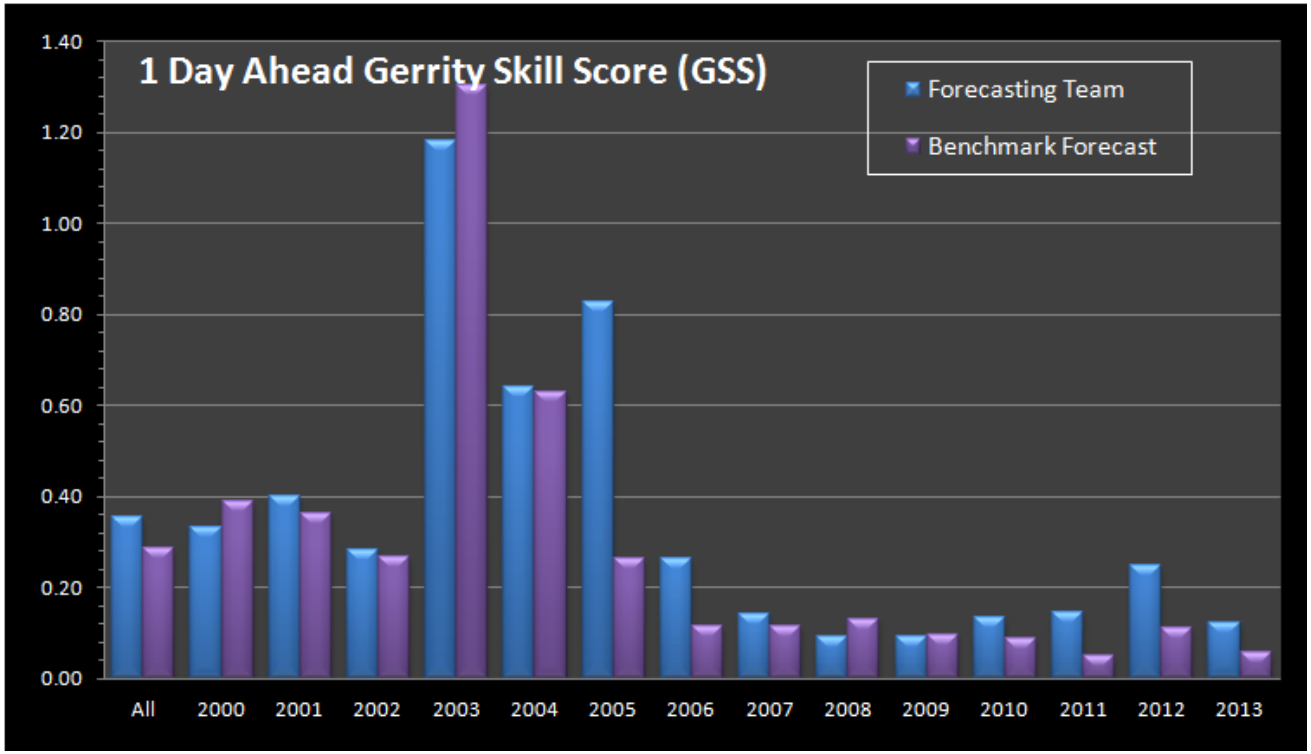
Scoring Matrix for Multi-Category Equitable Skill Scores

4 x 4 contingency matrix		Forecast Category				Marginal Total
		Q-U	ACTIVE	MINOR	MAJOR	
Observed Category	Q-U	3604	422	96	31	4153
	ACTIVE	414	169	46	9	638
	MINOR	96	55	22	9	182
	MAJOR	51	24	11	6	92
Marginal Total		4165	670	175	55	5065

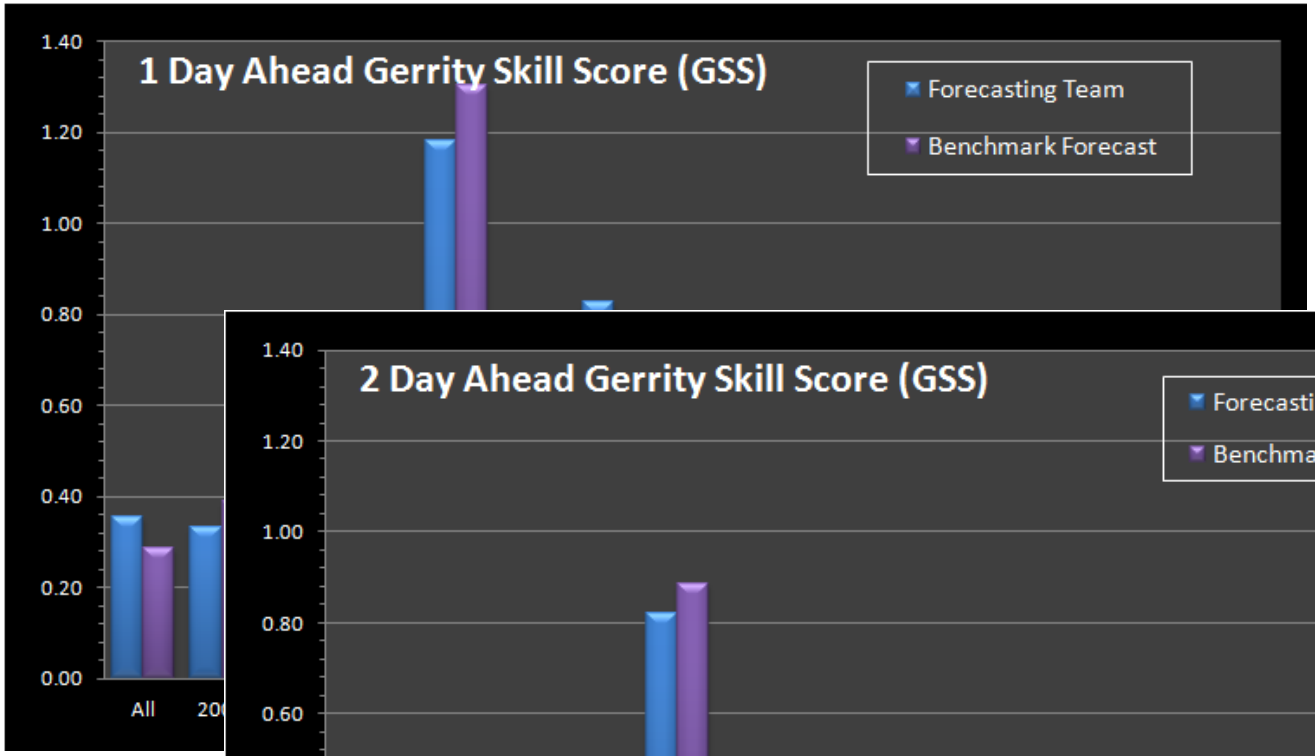
4 x 4 Probability matrix (P)		Forecast Category				Climatological probability
		Q-U	ACTIVE	MINOR	MAJOR	
Observed Category	Q-U	0.7115	0.0833	0.0190	0.0061	0.8199
	ACTIVE	0.0817	0.0334	0.0091	0.0018	0.1260
	MINOR	0.0190	0.0109	0.0043	0.0018	0.0359
	MAJOR	0.0101	0.0047	0.0022	0.0012	0.0182
Forecast probability		0.8223	0.1323	0.0346	0.0109	1.0000

4 x 4 reward-penalty or scoring matrix (S) as per Gerrity (1992)		Forecast Category			
		Q-U	ACTIVE	MINOR	MAJOR
Observed Category	Q-U	0.0984	-0.3081	-0.6605	-1.0000
	ACTIVE	-0.3081	1.5431	1.1907	0.8512
	MINOR	-0.6605	1.1907	7.3525	7.0130
	MAJOR	-1.0000	0.8512	7.0130	25.3645

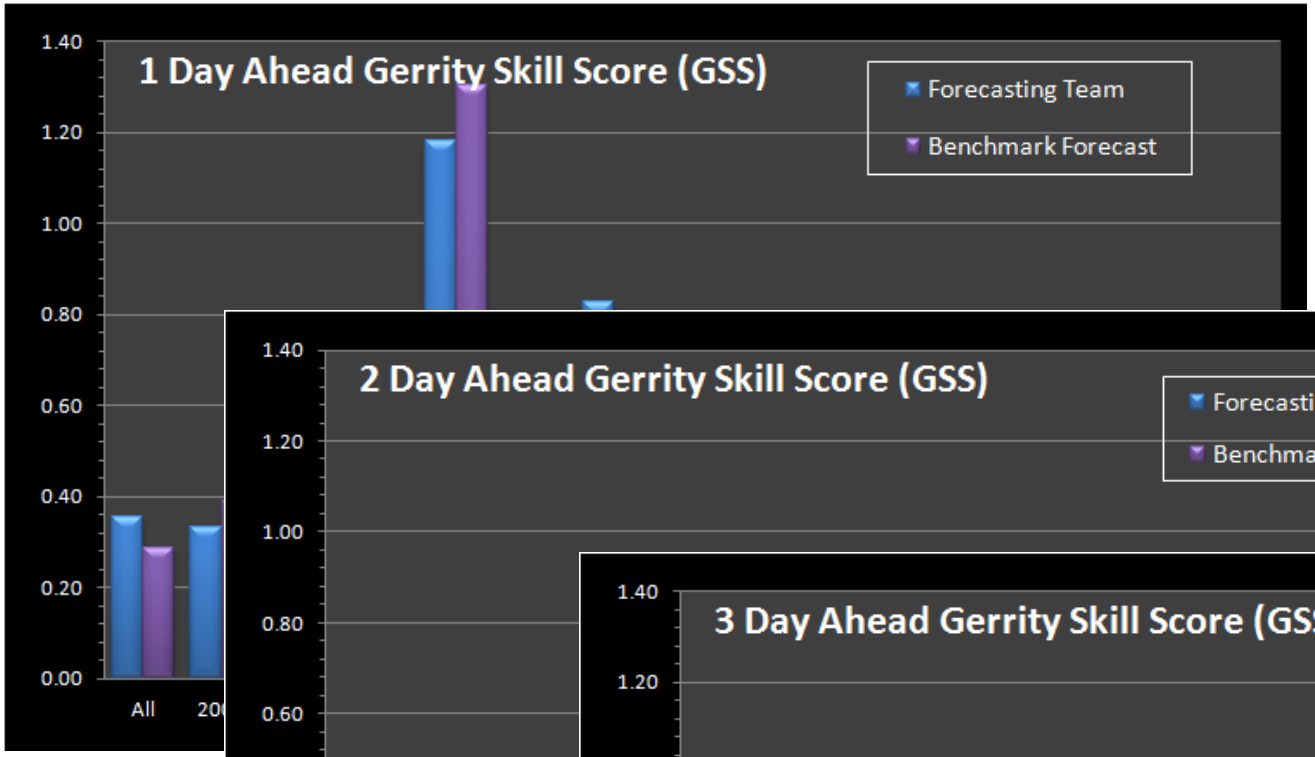
Yearly Forecast Verification using ESS (Gerrity)

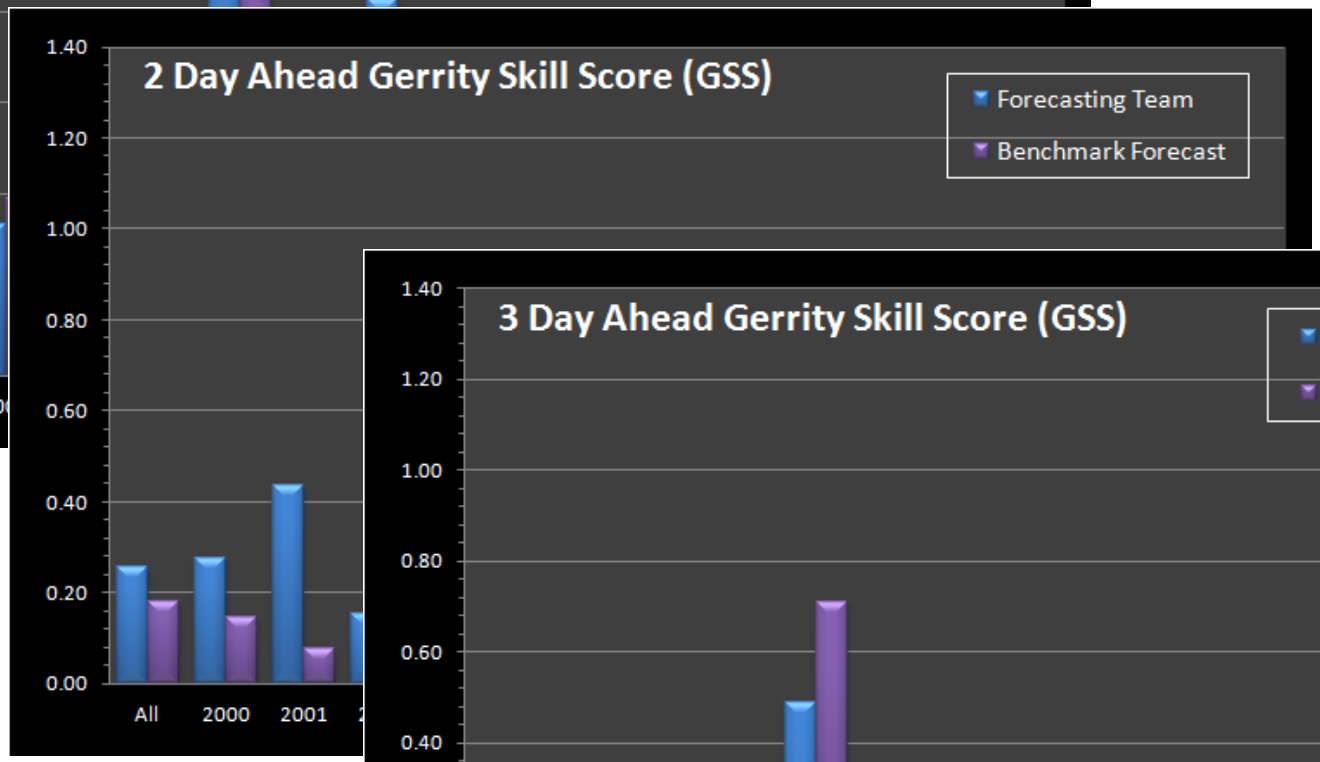


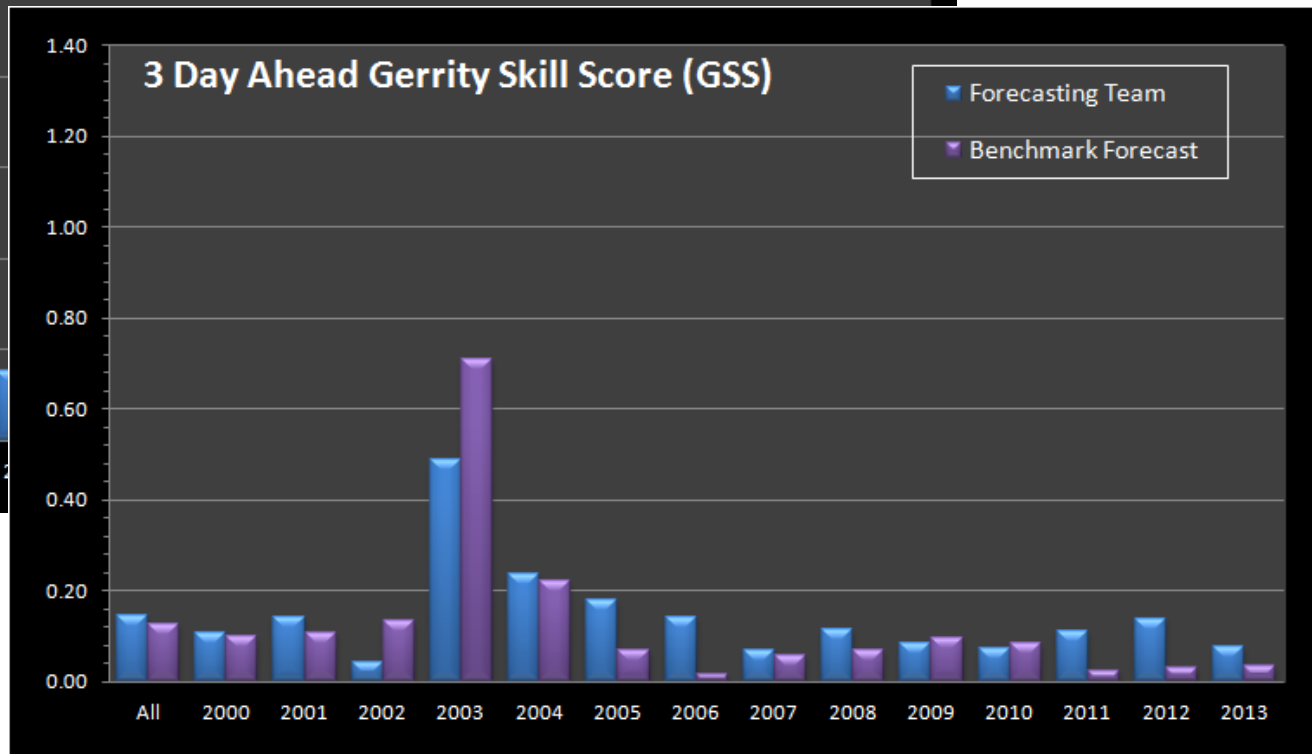
Yearly Forecast Verification using ESS (Gerrity)



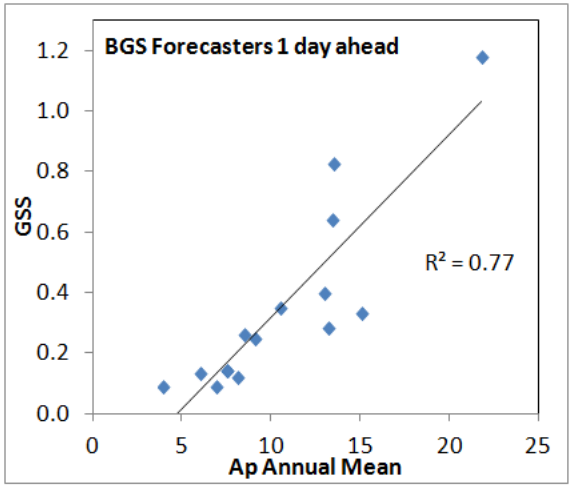
Yearly Forecast Verification using ESS (Gerrity)

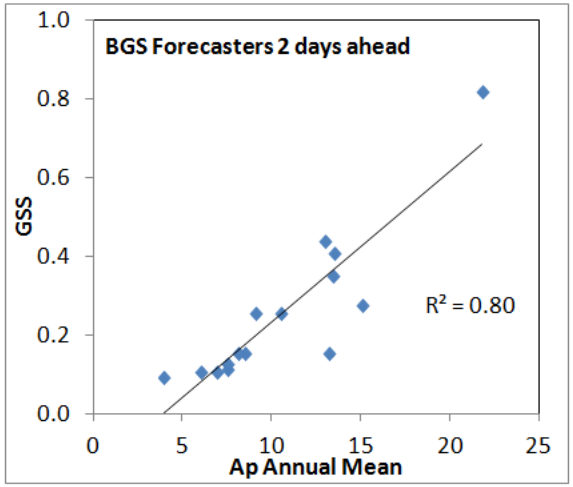


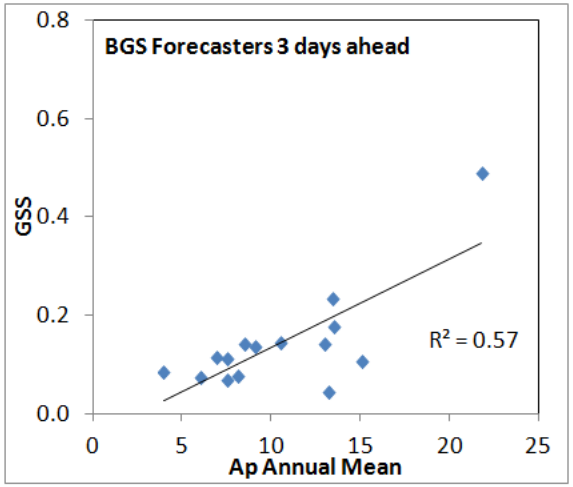


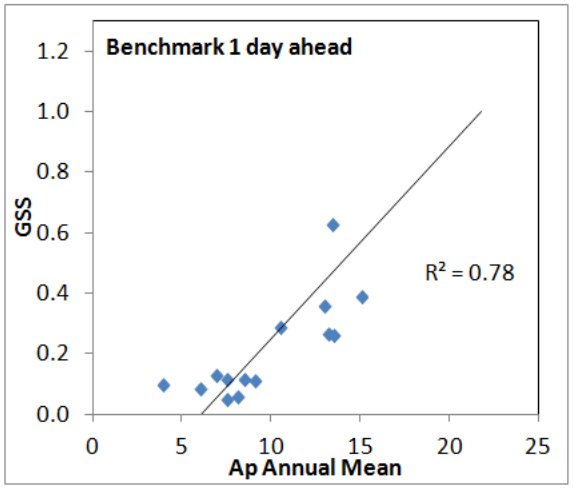


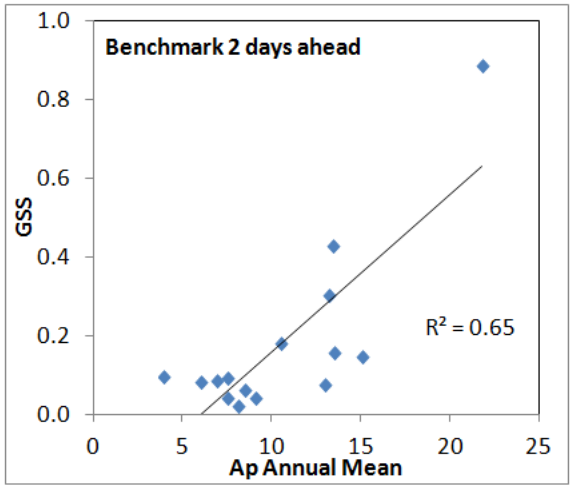
Correlation of Annual Mean A_p (AMV) and GSS

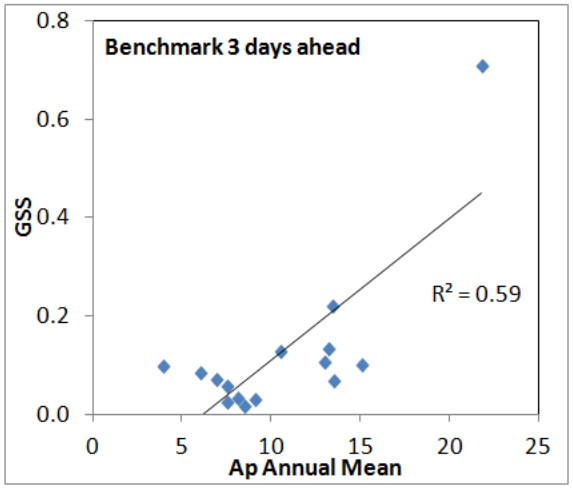




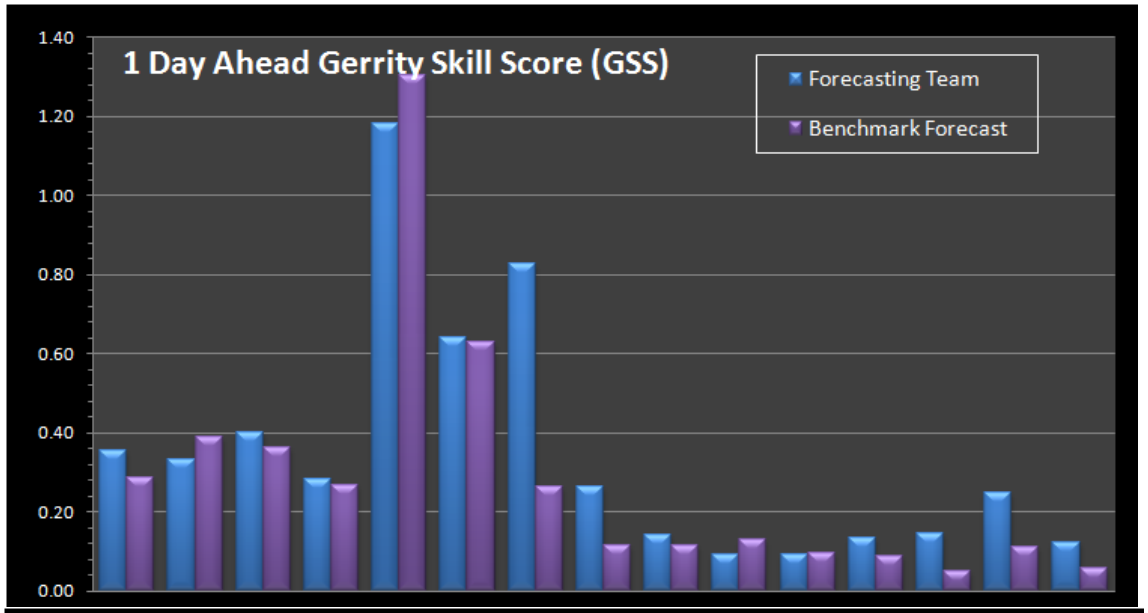


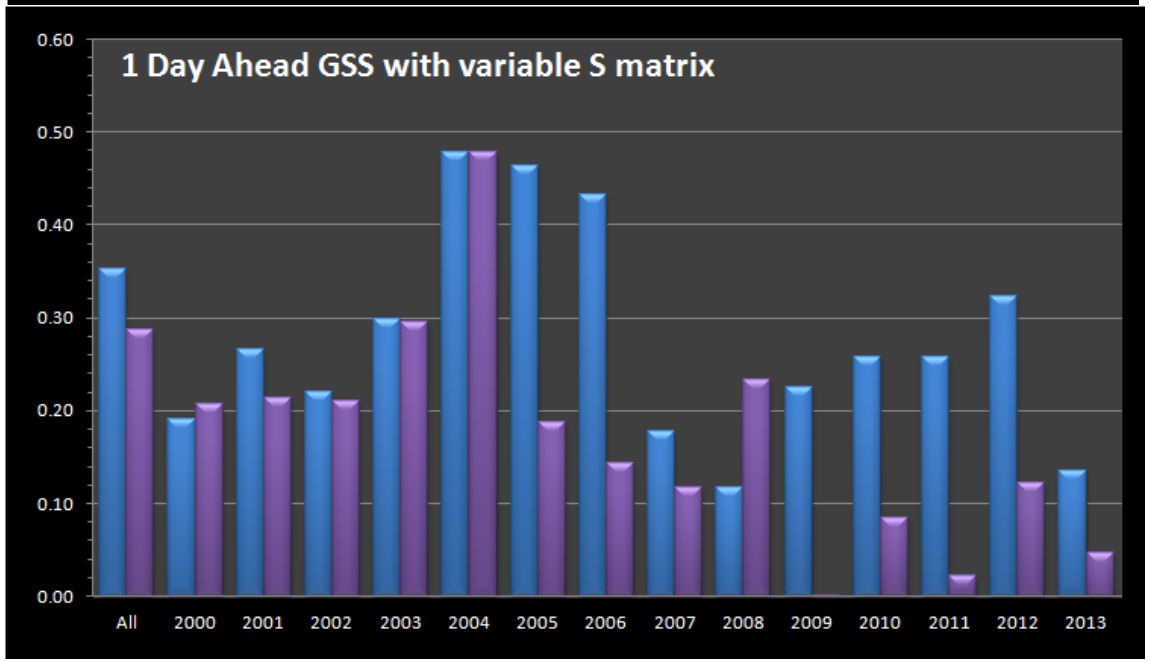




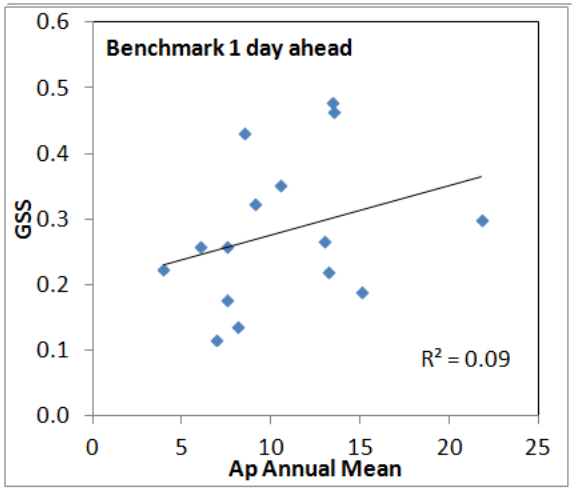


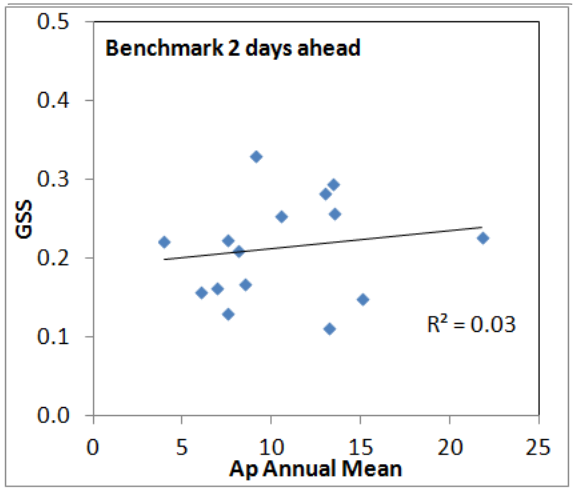
Yearly Forecast Verification using ESS (Gerrity)

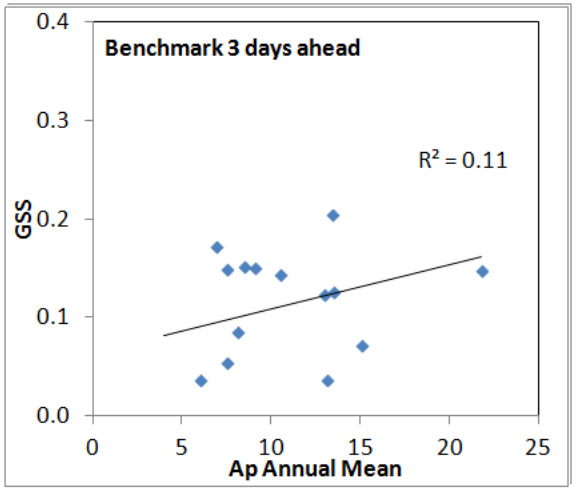


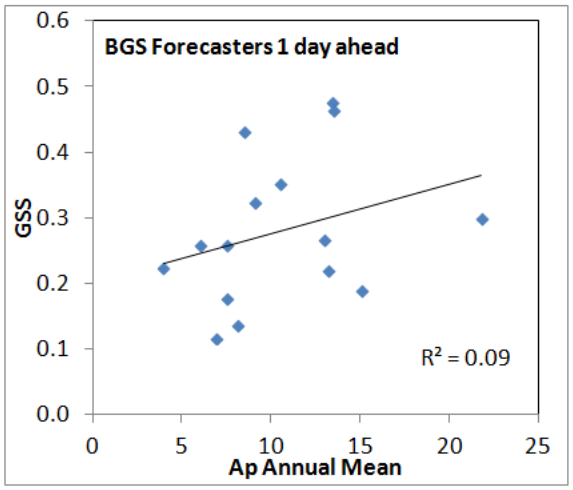


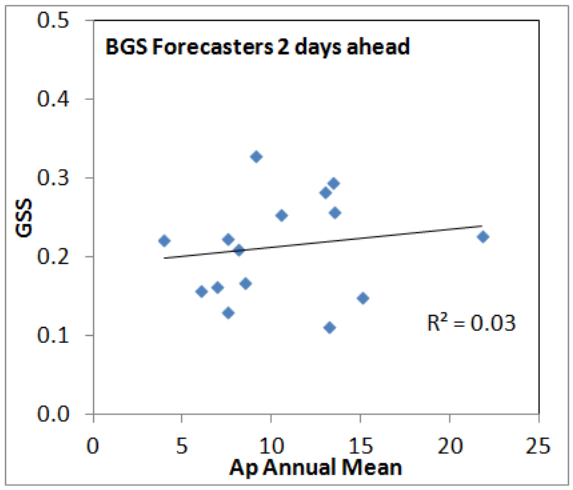
Correlation of Annual Mean A_p (AMV) and GSS

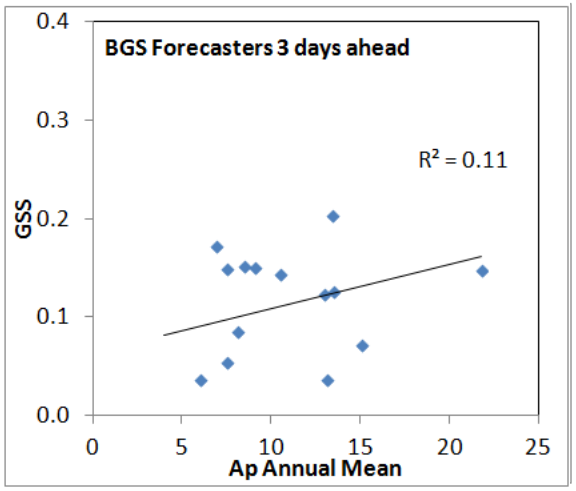




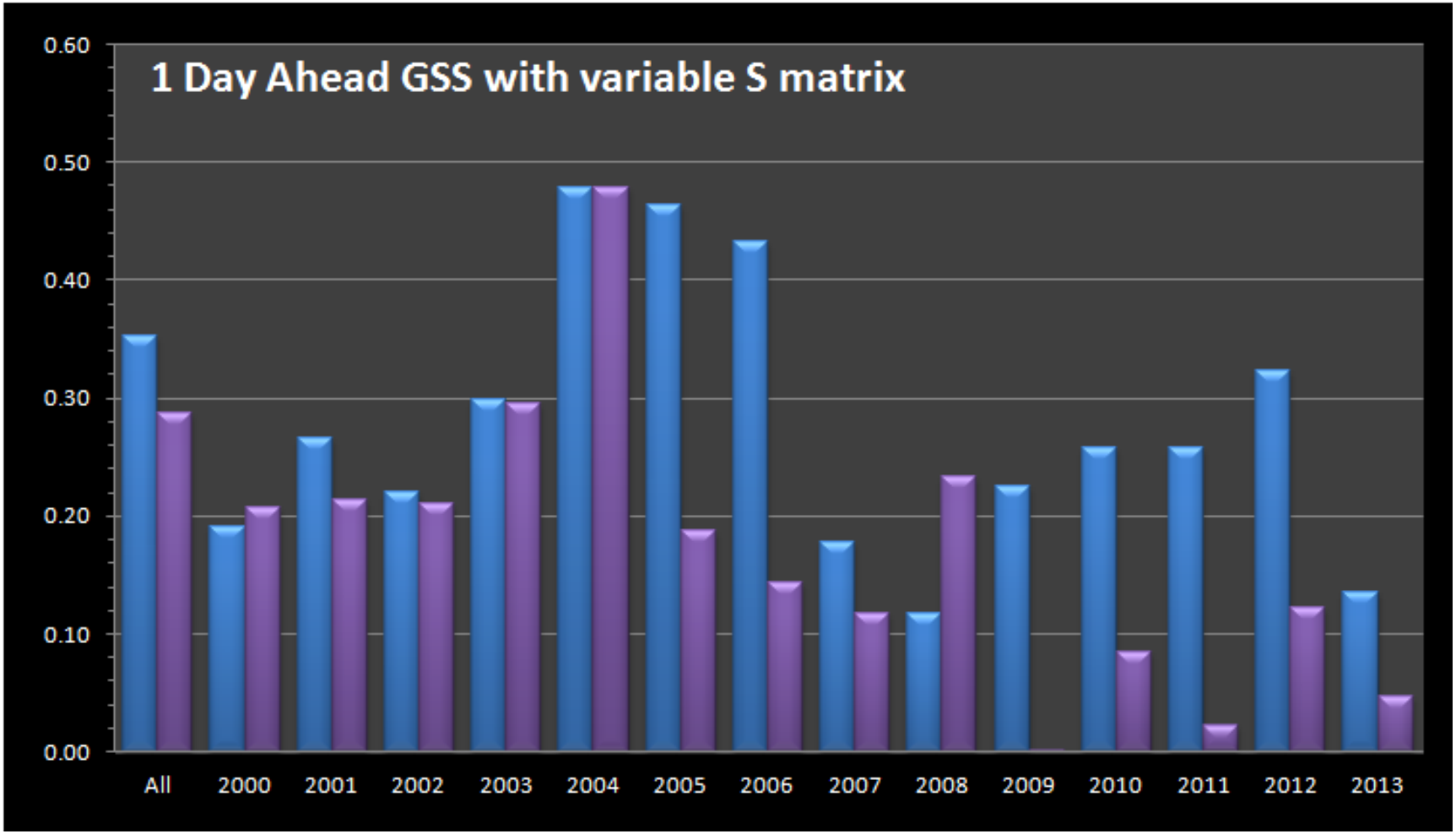








Yearly Forecast Verification using ESS (Gerrity)



Summary and Future Plans

- Analysed the performance of 1-3 day ahead forecasts, 2000 to 2013
- Suitable performance measures for on-going verification have been found
- Measures of both skill and bias are required to fully evaluate performance
- Results show an overall bias indicating tendency to under-forecast storms
- In most cases forecasters have higher skill scores for predicting storms than a simple forecast model of persistence and recurrence

Plans

- Adapt the forecast to predict maximum activity levels at any point in the day, as well as the current daily average
- Alter the categories to match the NOAA/SWPC G scale
- Devise suitable means of providing (motivatory) feedback to forecasters
- Obtain feedback from users and if required adjust verification measures
- Measure forecast performance against other BGS models (E.g. NeuralNet)
- Investigate the possible use of Extreme Dependency Scores
- Include error bars/confidence limits on the skill and bias scores

Acknowledgement: This work was partly funded under the EURISGIC consortium project, having received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement no 260330