Sweet Promises
(and some Sour Realities)
of Synoptic Full-Disk Vector Magnetogram Data

KD Leka
NorthWest Research Associates

previously or currently licensed to run and/or publish with:
Haleakala Stokes Polarimeter (Stokes II)
Huairou Video SpectroMagnetograph
Advanced Stokes Polarimeter
Imaging Vector Magnetograph
Mitaka Solar Flare Telescope
SoHO/Michelson Doppler Imager
Hinode/SpectroPolarimeter
SOLIS VectorSpectroMagnetograph
SDO/Helioseismic and Magnetic Imager
For one of my early supporters,

(with my daughter Kami In 2002)

Guy “Uncle Fred” Vandiver
1927 – 22 April 2013.
First, how far we've come! One of the first synoptically-run Vector Polarimeters:

6" pinhole, ~2hr to complete.
Now: HMI Vector Field
Data example:
Sweet Promises, I: *Physical Quantities.*

**Full Photospheric Vector:**
- Remove projection effects
  - *no limits as* $f(\mu)$
- Polar Fields are available when visible
- Systematic effects due to *noise* as $f(\mu)$,
  - *no longer due to assumptions* as $f(\mu)$
- Can compare observed to a Potential (lowest-energy) Field
Sweet Promises, I: Physical Quantities.

Full Photospheric Vector:
- Remove projection effects
- *no limits as* \( f(\mu) \)
- Polar Fields are available when visible
- Systematic effects due to *noise* as \( f(\mu) \),
  - *no longer due to assumptions as* \( f(\mu) \)
- Can compare observed to a Potential (lowest-energy) Field
Sweet Promises, I: Physical Quantities.

Full Photospheric Vector:
- Remove projection effects
  - no limits as \( f(\mu) \)
- Polar Fields are available when visible
- Systematic effects due to noise as \( f(\mu) \),
  - no longer due to assumptions as \( f(\mu) \)
- Can compare observed to a Potential (lowest-energy) Field
Sweet Promises, I: Physical Quantities.

Full Photospheric Vector:
- Remove projection effects
  - *no limits as* $f(\mu)$
- Polar Fields are available when visible
- Systematic effects due to *noise* as $f(\mu)$,
  - *no longer due to assumptions as* $f(\mu)$
- Can compare observed to a Potential (lowest-energy) Field

Figure 4 from Is the Polar Region Different from the Quiet Region of the Sun?
Hiroaki Ito et al. 2010
ApJ 719 131
doi:10.1088/0004-637X/719/1/131
Sweet Promises, II:  *No missed “events”*.  

Observer's wisdom: “The... *flare, emergence, eruption, feature-formation of interest*] always happen when it is ...[*cloudy, bad seeing, dark, raining, instruments are broken]*.”

- Fundamental questions still linger.  Example: Magnetic Emergence and Sunspot Formation
  - How do sunspots form?
  - What are the *first* indications of a new region?
  - How/when does the corona react, locally and over distance?
  - First appearance of magnetic flux has minimal obvious pre-cursors.
    - Where to point?
Sweet Promises, II: *No missed “events”*.  

Observer's wisdom: “The... *flare, emergence, eruption, feature-formation of interest*] always happen when it is ...*[cloudy, bad seeing, dark, raining, instruments are broken].”

- Fundamental questions still linger. Example: Magnetic Emergence and Sunspot Formation
  - How do sunspots form?
  - What are the *first* indications of a new region?
  - How/when does the corona react, locally and over distance?
  - First appearance of magnetic flux has minimal obvious pre-cursors.
    - Where to point?
Sweet Promises, II: *No missed “events”.*

Observer's wisdom: “The... *[flare, emergence, eruption, feature-formation of interest]* always happen when it is ...* [cloudy, bad seeing, dark, raining, instruments are broken].”

- Fundamental questions still linger. Example: Magnetic Emergence and Sunspot Formation
  - How do sunspots form?
  - What are the *first* indications of a new region?
  - How/when does the corona react, locally and over distance?
  - First appearance of magnetic flux has minimal obvious pre-cursors.
    - Where to point?
Sweet Promises, II: *No missed “events”*. 

Observer's wisdom: “The... [flare, emergence, eruption, feature-formation of interest] always happen when it is ...[cloudy, bad seeing, dark, raining, instruments are broken].”

- Fundamental questions still linger. Example: Magnetic Emergence and Sunspot Formation
  - How do sunspots form?
  - What are the *first* indications of a new region?
  - How/when does the corona react, locally and over distance?
  - First appearance of magnetic flux has minimal obvious pre-cursors.
    - Where to point?
Sweet Promises, II:  *No missed “events”.*

Observer's wisdom: “The... *[flare, emergence, eruption, feature-formation of interest]* always happen when it is ...*[cloudy, bad seeing, dark, raining, instruments are broken]*.”

- Fundamental questions still linger. Example: Magnetic Emergence and Sunspot Formation
  - How do sunspots form?
  - What are the *first* indications of a new region?
  - How/when does the corona react, locally and over distance?
  - First appearance of magnetic flux has minimal obvious pre-cursors.
    - Where to point?
Sweet Promises, II: *No missed “events”*. 

- Other Features [*flares, bridge formation, sunspot formation/disintegration*] and their evolution:
  - Fundamental aspects of magnetic field evolution are unknown.
  - Limited-FOV instruments: success “sit & stare” observing programs.
  - Evidence of success: Imaging Vector Magnetograph/Mees Solar Observatory
    - Ran every day whenever possible, 1991—2003.
    - First, “survey” then “sit and stare” modes.
    - Limited FOV, but high cadence.
    - Able to catch in-region emergence, pre-flare evolution, flare-related changes, sunspot growth/decay.
  - > 50 discovery papers, including 1st large study of $B$ and flares (*1200+ samples*)
Sweet Promises, III: Statistics

“Only with gobs of data can reliable statistical analysis be performed.”

Distributions:
- With insufficient data, *distributions are not characterizable*.
- If (e.g.) assume a Gaussian Distribution, 2 parameters (3 data points) are required.
  - Many distributions are not Gaussian.
  - We don't know what the distributions *are*.
- *eo ipso*, “gobs of data” are required
  - This is especially true to characterize the tails of distributions.
Sweet Promises, III: Statistics

“Only with gobs of data can reliable statistical analysis be performed.”

Distributions:
- With insufficient data, distributions are not characterizable.
- If (e.g.) assume a Gaussian Distribution, 2 parameters (3 data points) are required.
  - Many distributions are not Gaussian.
  - We don't know what the distributions are.
- eo ipso, “gobs of data” are required
  - This is especially true to characterize the tails of distributions.

Distribution of the length of magnetic neutral line for Flaring (color) and non-flaring (black) active regions.
Sweet Promises, III:  **Statistics**

“Only with gobs of data can reliable statistical analysis be performed.”

**Distributions:**
- With insufficient data, *distributions are not characterizable*.
- If (e.g.) assume a Gaussian Distribution, 2 parameters (3 data points) are required.
  - Many distributions are not Gaussian.
  - We don't know what the distributions *are*.
- *eo ipso*, “gobs of data” are required
  - This is especially true to characterize the tails of distributions.

*Distribution of the length of magnetic neutral line for Flaring (color) and non-flaring (black) active regions.*
“Only with gobs of data can reliable statistical analysis be performed.”

Distributions:
- With insufficient data, distributions are not characterizable.
- If (e.g.) assume a Gaussian Distribution, 2 parameters (3 data points) are required.
  - Many distributions are not Gaussian.
  - We don't know what the distributions are.
- eo ipso, “gobs of data” are required
  - This is especially true to characterize the tails of distributions.

If the length of magnetic neutral line for flaring (color) and non-flaring (black) active regions.
Sweet Promises, III:  Statistics

“Only with gobs of data can reliable statistical analysis be performed.”

Distributions:

- With insufficient data, *distributions are not characterizable*.
- If (e.g.) assume a Gaussian Distribution, 2 parameters (3 data points) are required.
  - Many distributions are not Gaussian.
  - We don't know what the distributions are.
  - *eo ipso*, “gobs of data” are required
  - This is especially true to characterize the tails of distributions.
“Only with gobs of data can reliable statistical analysis be performed.”

**Distributions:**

- With insufficient data, *distributions are not characterizable*.
- If (e.g.) assume a Gaussian Distribution, 2 parameters (3 data points) are required.
  - Many distributions are not Gaussian.
  - We don't know what the distributions are.
- *eo ipso*, “gobs of data” are required
  - This is especially true to characterize the tails of distributions.
“Only with gobs of data can reliable statistical analysis be performed.”

Consider: Very basic Multi-variable Analysis:
- Extend to more than one variable? Lots of data needed.
“Only with gobs of data can reliable statistical analysis be performed.”

Consider: Very basic Multi-variable Analysis:
- Extend to more than one variable? Lots of data needed.
“Only with gobs of data can reliable statistical analysis be performed.”

Consider: Very basic Multi-variable Analysis:
• Extend to more than one variable? Lots of data needed.

see e.g., Silverman 1998
Initial Data Set for Second Flare-Forecasting Comparison Workshop:
- 7 months HMI HARP time-series data: 2011 August – 2012 February.
- Almost 700 distinct HARPs sampled, many over multiple days.
- Over 2,000 time-series.
- 291 flares ≥ C1.0
- 48 flares ≥ M1.0
Initial Data Set for Second Flare-Forecasting Comparison Workshop:
- 7 months HMI HARP time-series data: 2011 August – 2012 February.
- Almost 700 distinct HARPs sampled, many over multiple days.
- Over 2,000 time-series.
- 291 flares $\geq$ C1.0
- 48 flares $\geq$ M1.0

*It's not enough*....
(O. v d L's point: Space Weather/societal Impacts = our bread & butter. T. H.: mentioned HMI space weather quantities.)

Promise of Synoptic Photospheric Vector magnetic field data for forecast operations:

- Research (using synoptic data) has shown photospheric vector field contains relevant information for forecasting energetic events.
  - And more info than line-of-sight component.
- Automated methods are available for flare/CME forecasts (B_los, Ic)
- R2O (Research to Ops) ready for investment.

Critical aspects:
- Consistency in instrumentation
- Large sample sizes are absolutely required.
Method “B”

- Used all data (7 months, 2062 data sets) for both training and forecasts
- Preliminary Results for a “Free Magnetic Energy Proxy”:

<table>
<thead>
<tr>
<th>Observed (C1.0+)</th>
<th>Predicted</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>flare</td>
<td>129</td>
<td>162</td>
</tr>
<tr>
<td>no flare</td>
<td>62</td>
<td>1709</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Observed (M1.0+)</th>
<th>Predicted</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>flare</td>
<td>12</td>
<td>36</td>
</tr>
<tr>
<td>no flare</td>
<td>7</td>
<td>2007</td>
</tr>
</tbody>
</table>

ACC: 0.89
HSS: 0.48
TSS: 0.41

Biased Skill Scores for Free Energy Proxy

ACC: 0.97
HSS: 0.35
TSS: 0.25

Quiet vs. Flaring with Boundary
Reliability Plots: Sample sizes make a huge difference!

C1.0+ Free Energy Proxy

M1.0+ Free Energy Proxy

ACC: 0.89 UnBiased Skill Scores for Free Energy Proxy
HSS: 0.47 (using Cross Validation)
TSS: 0.40

ACC: 0.98
HSS: 0.33
TSS: 0.24
Sour Realities, I: Data volume.
Balance of spectral, temporal, spatial.

Two aspects:
• Real-time acquisition, processing, archiving.
• Data availability for analysis.

Real-time issues: usually get the focus.
• Store all raw data?
• How/when interrupt for calibration?

Analysis issues: for research, highest “daily” impact
• how to distribute data?
  • Quick-look, Browse, and Massive.
• What level processing to provide?
• how to provide what researchers need? (they’ll all be different)
Sour Realities, II: Algorithm stability
Calibration, Inversion, Ambiguity resolution, *more*...

- Stability is crucial
- Understanding and Documenting changes to instrument, algorithms (long- and short-term).
- Large datasets: processing algorithms need to be fast, but robust against noise, solar conditions, temporal evolution, observing conditions.
  - Larger the data sample, more likely algorithms will fail somewhere (*statistics*).
  - Larger the data sample, more solar conditions that will contradict algorithm assumptions.
- Provocative Statement: we cannot do it yet.
  - Counter Statement: random problems are diluted in large sample sizes.
    - Systematic issues (e.g., angle biases) are worrying.
Data Issues to be aware of:

Sometimes inversions fail.
Sometimes it is “obvious”.

NORTHWEST RESEARCH ASSOCIATES
"unencumbered science"
Data Issues to be aware of:

Sometimes inversions fail.
Sometimes it is “not”.
Data Issues to be aware of:

Sometimes inversions fail.
Sometimes it is “not”.

NORTHWEST RESEARCH ASSOCIATES
"unencumbered science"
Data Issues to be aware of:

Sometimes inversions fail.
The more data, the more pixels,
the more bad pixels/areas will appear.

KD Predicts:
HMI inversion trouble in Big Sunspot visible now (AR 11171.)
Data Issues to be aware of:
Disambiguation: not optimized over time.
- There will be temporal inconsistencies in solutions.
Data Issues to be aware of:
Disambiguation: not optimized over time.
- There will be temporal inconsistencies in solutions.
Data Issues to be aware of:
Disambiguation: not optimized over time.
• There will be temporal inconsistencies in solutions.
Data Issues to be aware of:
Disambiguation: not optimized over time.
- There will be temporal inconsistencies in solutions.
Azimuthal Angle Bias.

Thank your HMI team for finding and eliminating this.

(at least most of it.)
Sour Realities, III: Limitations in the photosphere.
- If solely photosphere, data still limited to single (forced?) boundary.
  - Inconsistent with many scientific queries.
- If data extend beyond photosphere, see SR I (data size realities) and SR II (algorithms), above.

Sour Realities, IV: Limitations of unresolved data.
- Data are discrete and generally unresolved.
- Finite-differences and gradients are causes for caution
  - But hold information anyway?

Both of these are topics for entire separate talks...
Dream List:

No limit on $$ or time:
• HMI-clones (minus the “H”, plus some improvements) at Earth-, Far- and Polar vantage points.
• Hot-spare ready & waiting.

Realistic List:
• GONG-like network of imaging-based vector magnetographs with image-stabilization and deblurring/AO.
• similar temporal and spatial sampling to HMI.

A “must”: Research Support.
• Existing synoptic vector data have not yet been fully exploited.
• Processing/Analysis algorithms need improvement.
• Fundamental questions on activity and evolution are ripe for discovery
  • (or at least substantiation of single detections).