The future role of meteorologists in the age of artificial intelligence

The human/automation relationship: How can we best use AI tools?

Gail Hartfield, NOAA/NWS Meteorologist
Bruce Telfeyan, USAF Meteorologist
Jerry Shields, Emergency Management Meteorologist
Jeffrey Fries, USAF Meteorologist

Working Group Lead: Neil Stuart, NOAA/NWS
Psychology and AI expert consultant: Dr. Robert Hoffman, Institute for Human and Machine Cognition

Overview:
How can humans use AI tools?

- Uses and roles of each will vary based on a number of factors
- Where will increased forecaster involvement result in increased success?
- Current and near-future human/AI interdependences
- What our next steps must be

Automation "versus" humans? It's a partnership, not a contest

- We all need to work together
- Tap into the strengths of each (and reduce the impact of the negatives)
  - Roles of each will vary based on the situation
    - Time scales
    - Spatial scales
    - Weather type
    - Purpose/audience

Automation and humans: Characteristics

[Source: Hoffman et al. (2002)]	Machines
Are constrained in that	**Need people to**
Sensitivity to context is low and is ontology-limited	Keep them aligned to the context
Sensitivity to change is low and recognition of anomaly is ontology-limited	Keep them stable given the variability and change inherent in the world
Adaptability to change is low and is ontology-limited	Repair their ontologies
They are not "aware" of the fact that the model of the world is itself in the world	Keep the model aligned with the world
	People
Are not limited in that	**Yet they create machines to**
Sensitivity to context is high and is knowledge- and attention-driven	Help them stay informed of ongoing events
Sensitivity to change is high and is driven by the recognition of anomaly	Help them align and repair their perceptions because they rely on mediated stimuli
Adaptability to change is high and is goal-driven	Affect positive change following situation change
They are aware of the fact that the model of the world is itself in the world	Computationally instantiate their models of the world

Specific to the weather enterprise…

Automation can process and update rapidly; tunable to needs; consistent; objective; focuses on the task at hand

Automation has the potential for technical issues and error; lacks empathy

Humans can adjust their approach (interpretation and communication) based on user needs; applied intuition; subjectivity (tolerance of ambiguity)

Humans must deal with varying degrees of distractions, time constraints, inconsistency

Automation and humans: Characteristics

A richer view:

Human - Machine Interdependence.

- Things that machines are good at and ways in which they depend on humans in joint activity
- Things that humans are good at and ways in which they depend on the machines.

Automati… tunable to ne… the task at hand

Automation has the potential for technical issues and error; lacks empathy

…tation …pplied intuition; subjectivity (tolerance of ambiguity)

Humans must deal with varying degrees of distractions, time constraints, inconsistency

What tools, procedures, methods will be most useful in different forecasting time frames?

- Near term (warning; 0-24 hrs):
  - Radar, satellite, mesonets, convection-allowing models (CAMs), Warn-on-Forecast output, synthesis of observational data & extrapolation / trends
- Short term (warning/watch; 24-72 hrs):
  - Models (deterministic; calibrated ensemble systems)
- Long term (watch/outlook; beyond 72 hrs):
  - Global models, ensembles, land-sea-atmosphere interactions, large scale flow patterns / applied climatology; probabilistic info / alternate scenarios

Other variations can dictate the best forecast tools and methods to use and type of human involvement

- Spatial scale:
  - Fine scale (neighborhood) = CAMs & CAM interpretation, observations, satellite, local effects
  - Large scale = Models, ensemble systems, ensembles, probabilities, jet phase diagrams, pattern recognition
- Weather type:
  - E.g., winter p-type changes/mixes, tropical cyclones/inland decay, derechos, extreme heat, etc. all necessitate varied approaches
- Forecast purpose/audience:
  - E.g., tornado warning, briefing to county fair on synoptic wind gusts, cloud forecast for military maneuvers, aviation, seasonal hurricane outlooks

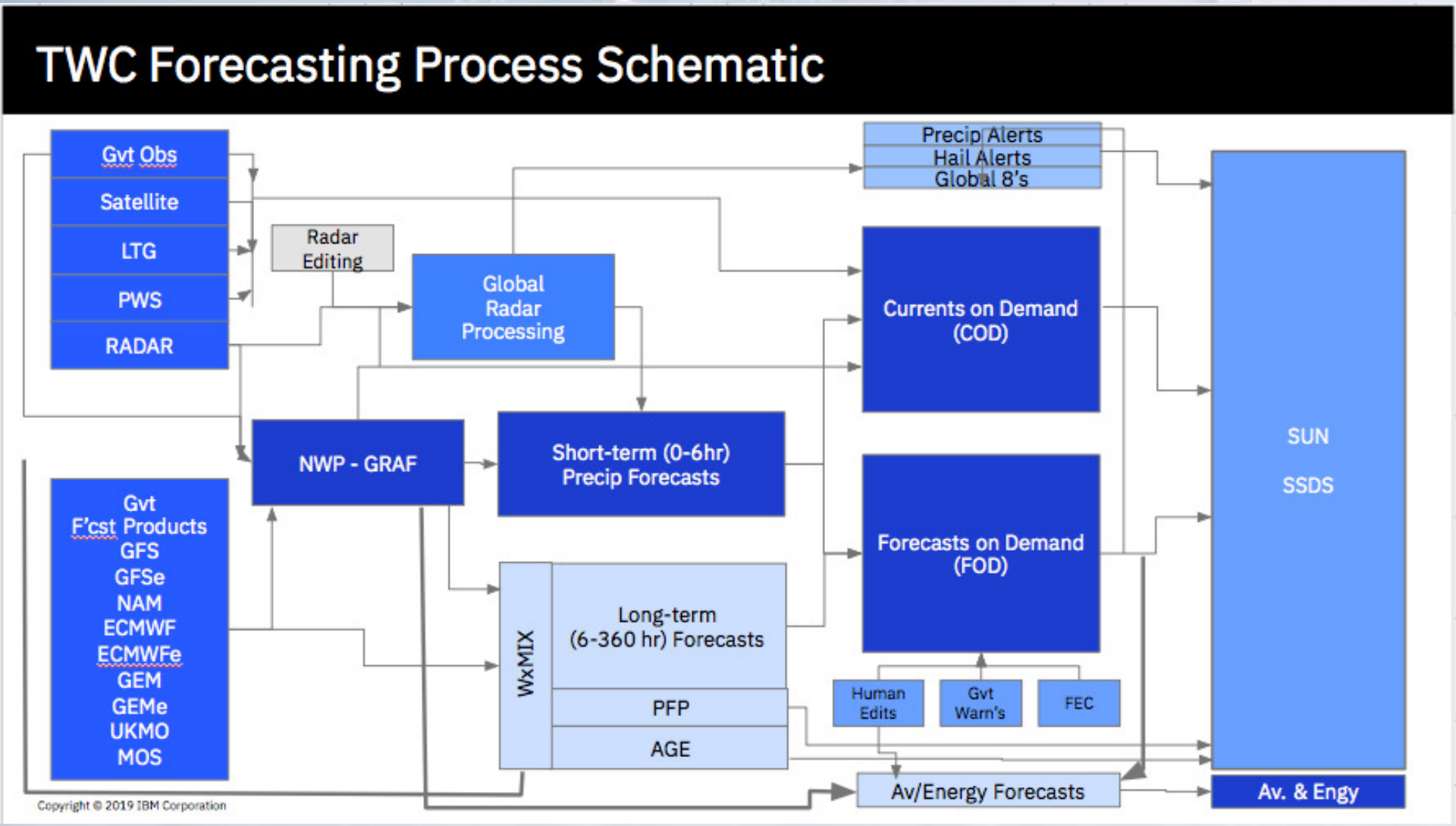
Where can humans improve upon machines? And is it always worth the cost?

Using stats and logic to determine when humans should intervene

- How much human improvement on guidance output is significant?
- What is worth the man-hours?
  - Tornado track precision vs. one-degree max temp improvement
- Involvement (amount and type) will vary by weather enterprise sector
- Techniques like human “over-the-loop” processing can identify opportunities for human involvement

Human Over the Loop: Rose et al. (2015) – The Weather Company/Channel

TWC's forecast system places human involvement in select locations in the process

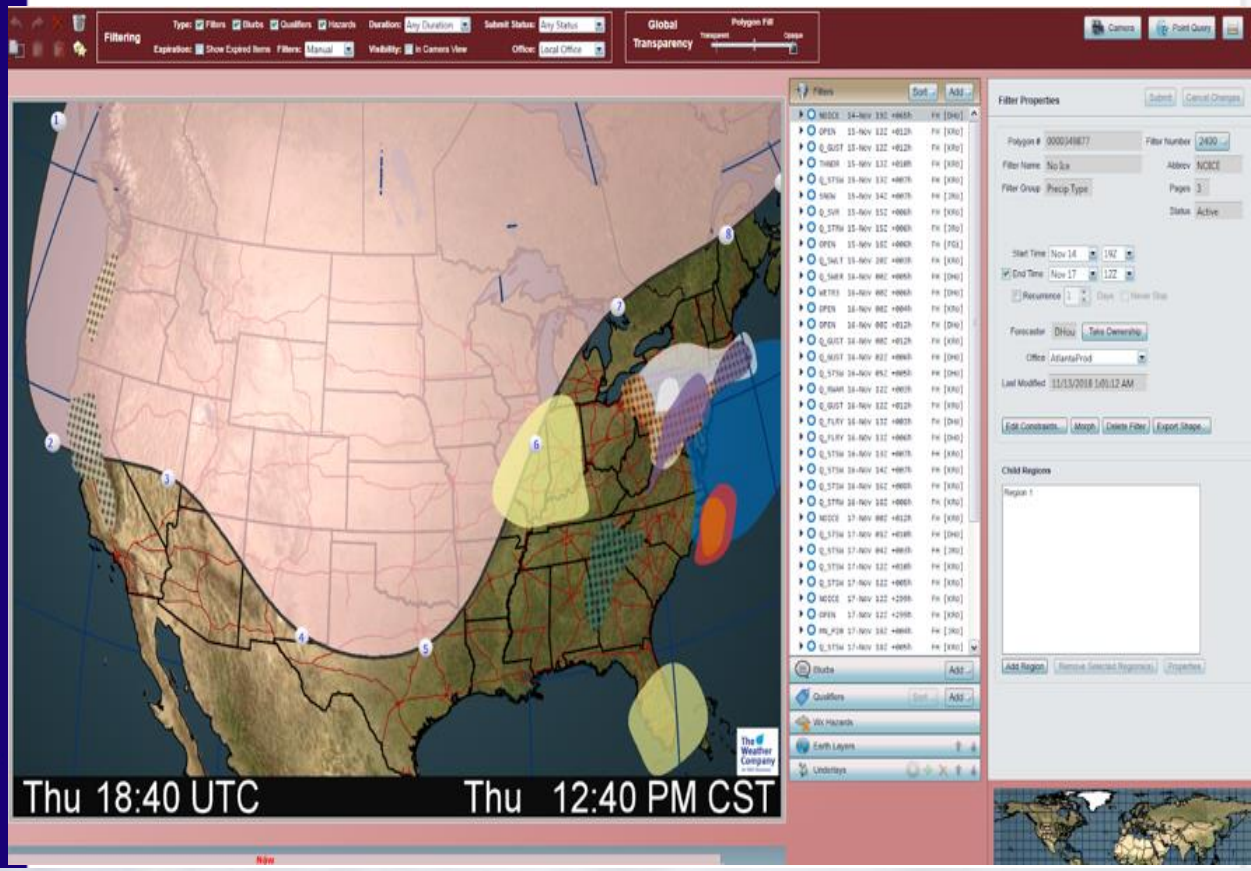


Graphic courtesy of Dr. Peter Neilley and Joseph Koval, The Weather Company/Channel

HOTL (Human over the Loop)

Graphic courtesy of Dr. Peter Neilley and Joseph Koval, The Weather Company/Channel

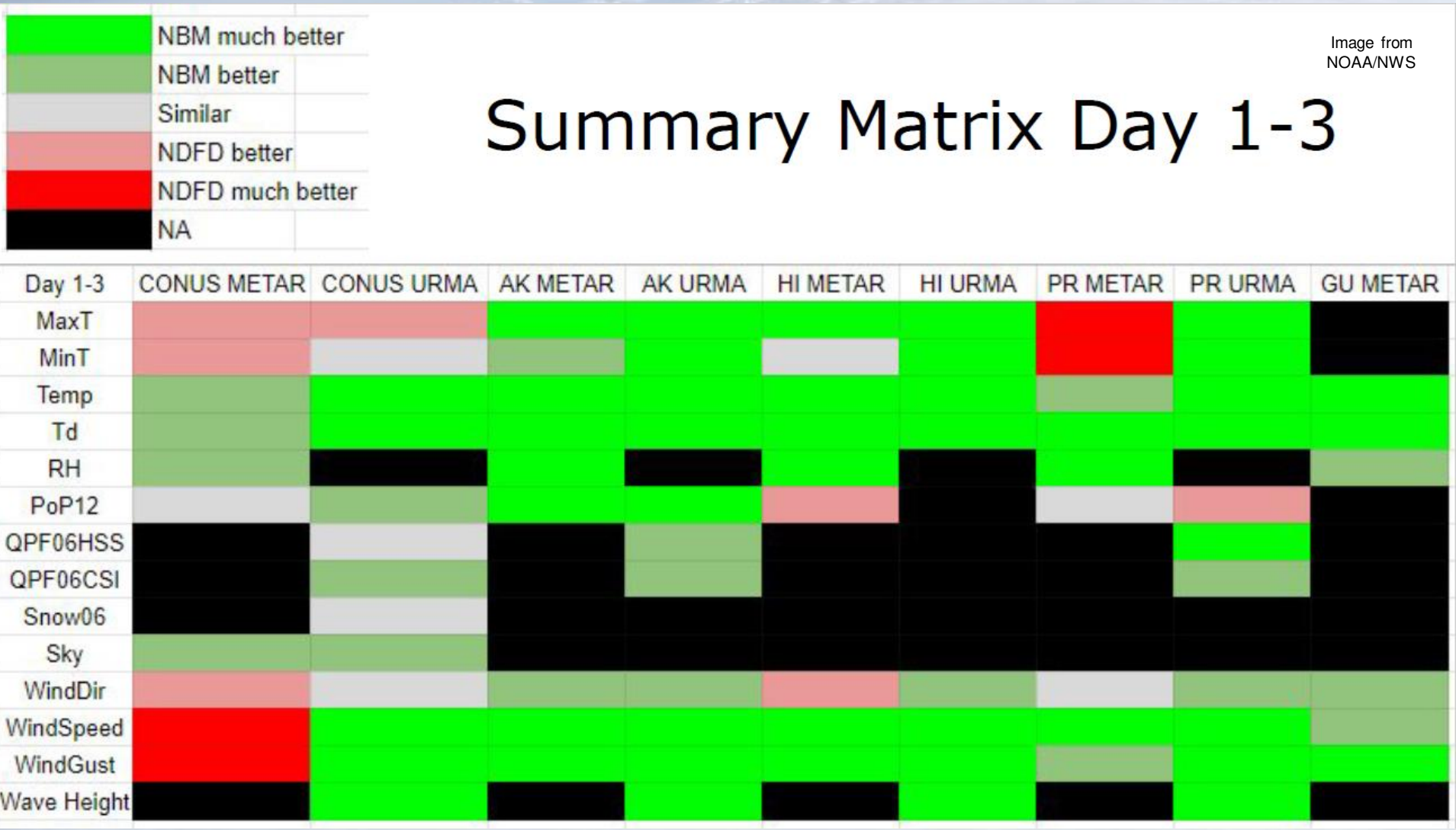
- Human forecaster continues to play a role in our forecasting operation
- But data shows freshest forecasts are the most accurate
- We've moved to an over-the-loop paradigm where model data continues to flow through our system
- Human can influence forecasts through a graphical tool we've developed called HOTL
- Continuous flow of forecast, but forecast is modified at publish time by geo-navigated digital polygons called filters or qualifiers



Copyright © 2019 IBM Corporation

TWC: Humans adjust the forecast at specified locations using filtering tool and text additions without delaying the forecast process

NWS's National Blend of Models (www.weather.gov/mdl/nbm_home)



Model blend often **improves** upon human-adjusted forecasts beyond day 4 **but does not replace humans**

MaxT CONUS URMA

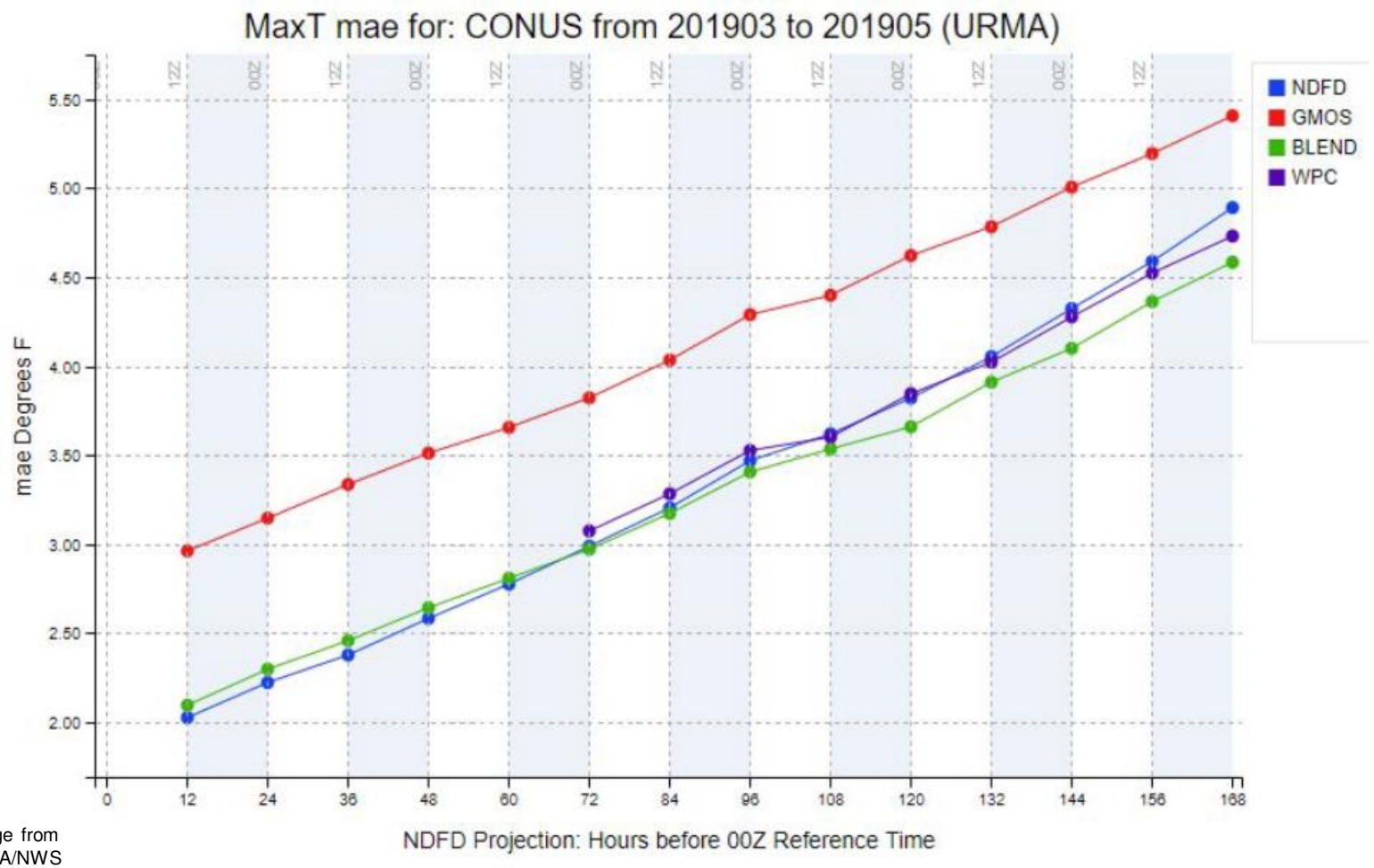


Image from NOAA/NWS

TAFVER

TAF (9685)	Wind (30)	Vis (15)	Cig (15)	ALSTG (10)	Temp (10)	Weather (20)	Tempo (100)	Final (100)	HVA
Forecaster Final	27.97	14.17	12.18	7.35	8.59	15.7	28.36%	85.96%	
TBOW Selected Model TAF	27.92	13.98	11.76	7.91	7.86	14.33	NA	83.76%	13.55%
GALWEM Auto TAF	27.91	14.37	12.07	7.88	7.53	14.97	NA	84.74%	7.99%
NAM Auto TAF	27.92	13.8	11.58	7.92	7.94	14	NA	83.16%	16.63%
GFS Auto TAF	27.33	13.98	11.8	7.5	7.81	13.93	NA	82.35%	20.45%

Number of TAFS: 9685
Forecaster: All
Reviewer: All
ICAO: All
Region: All
Time: 01 Jan 2019 0Z through 01 Jul 2019 0Z

USAF Forecaster performance relative to model-based first look TAF products in SE CONUS 1st half of 2019

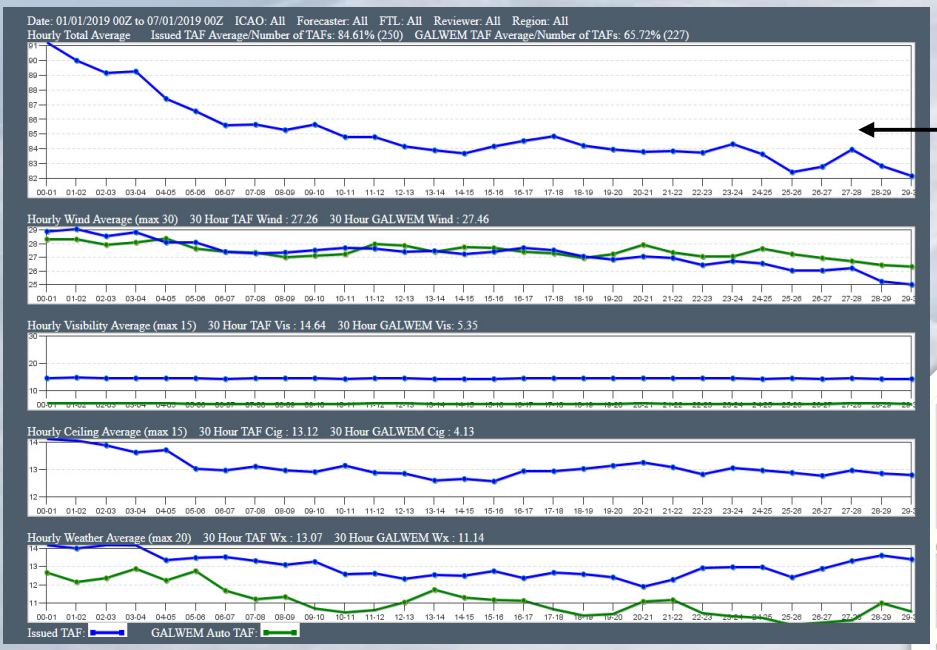
HVA = Error Reduction

TBOW Selected = Forecaster Over the Loop

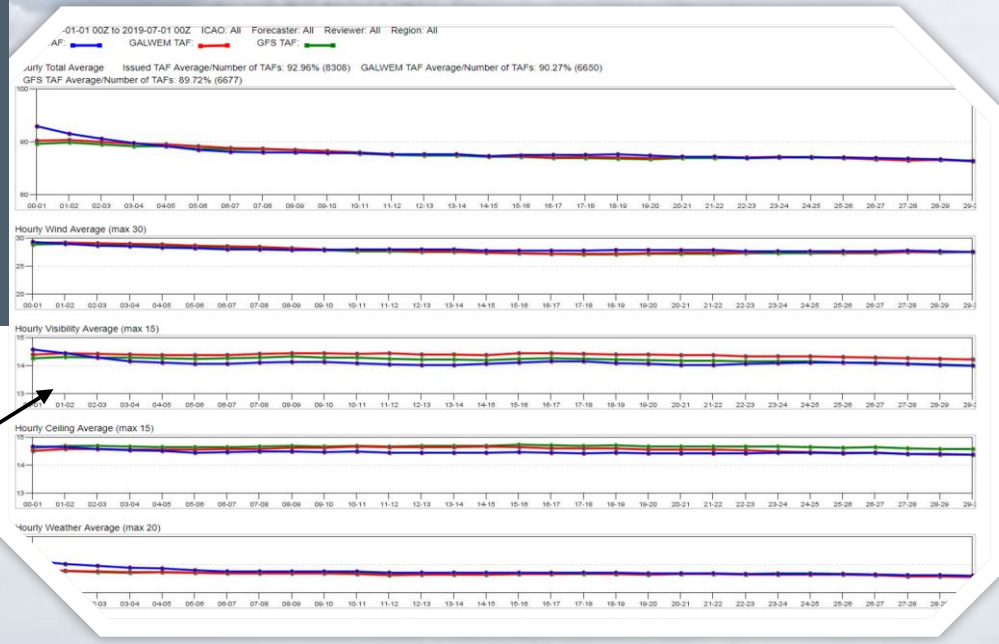
TAFVER

TAF (290)	Wind (30)	Vis (15)	Cig (15)	ALSTG (10)	Temp (10)	Weather (20)	Tempo (100)	Final (100)	HVA
Forecaster Final	27.19	14.59	13.06	7.61	9.09	13.08	20.02%	84.61%	
GALWEM Auto TAF	27.46	5.33	4.11	9.02	8.63	11.16	NA	65.72%	42.94%

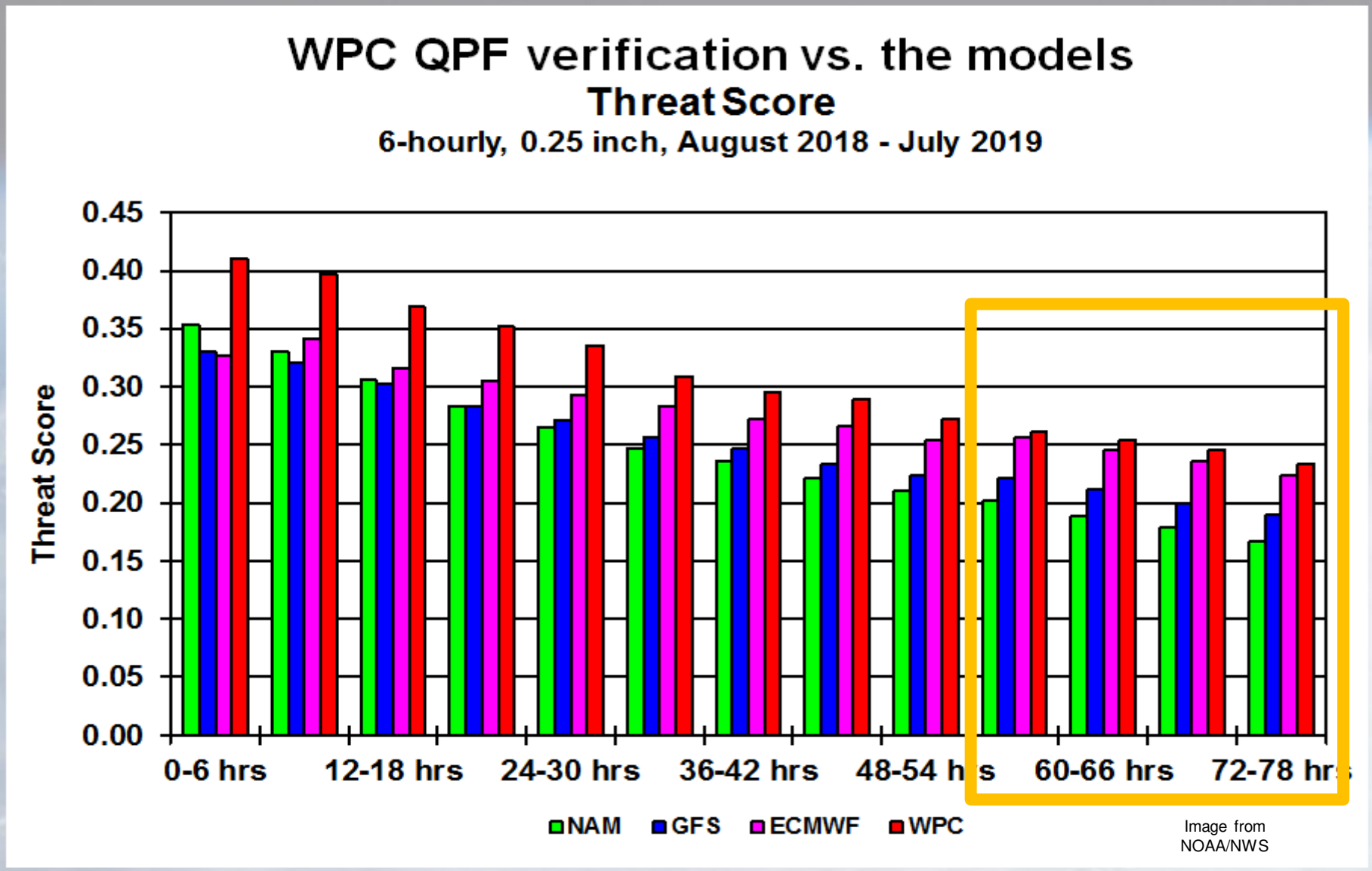
Number of TAFS: 290
Forecaster: All
FTL: All
Reviewer: All
ICAO: All
Region: All
Time: 01 Jan 2019 0Z through 01 Jul 2019 0Z



USAF Forecaster performance cumulative (top), by forecast hour (bottom) Europe and Africa 1st half 2019. TAF products filtered to display observed TS conditions at any point in 30-hr TAF product



USAF Forecaster performance in data sparse environment (Central Asia) 1st half of 2019 relative to GALWEM and GFS derived first look TAF products

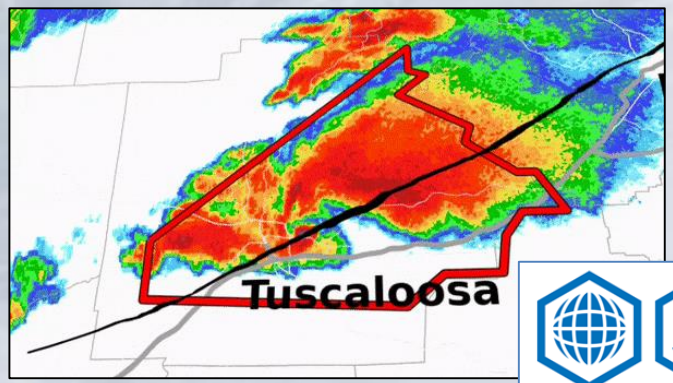


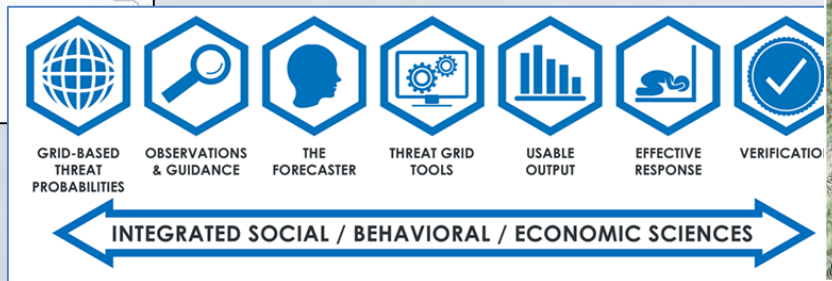
Are we adding enough value behind Day 2 or 3 to justify time spent, or is time better spent on more urgent needs?

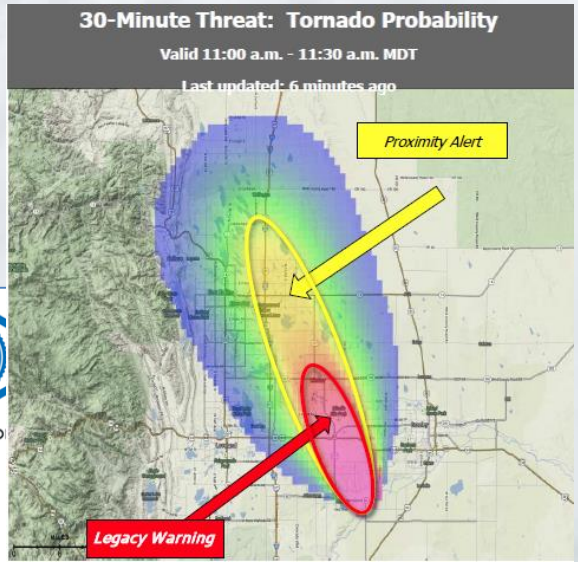
Current and Near-Future Human/Machine Interdependence (Humans Using Machine Tools)

Human/Machine Interdependence: FACETS

- Forecasting a Continuum of Environmental Threats; uses Warn-on-Forecast as a foundation
- Continuous flow of information; info-centric & service-driven
- Refines and adds detail to current watch/warning system
- Auto-updates threat info with satellite, radar, observations (Threats in Motion: warning polygons update every min & move forward in time)
- Provides probabilistic tools to support warning decisions (PHI plumes)
  - Forecasters will need training on how the sausage is made
  - Will forecasters put appropriate reliance on this output?
- Communication methods must be developed and refined for maximum effectiveness



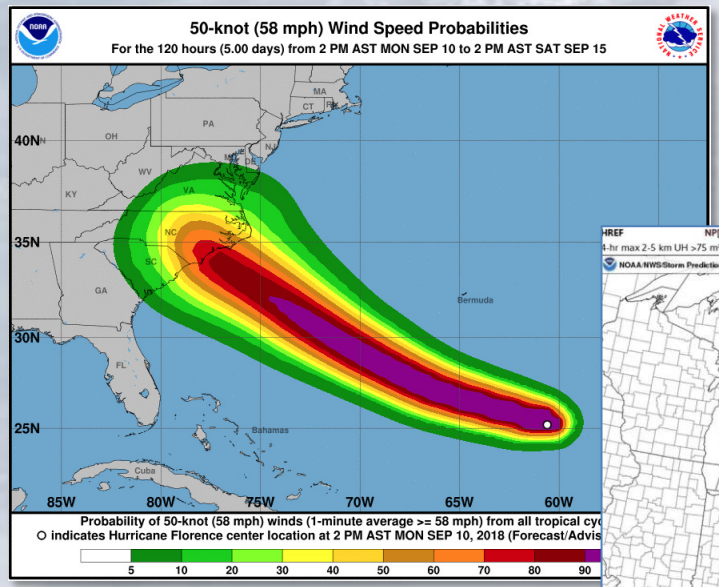


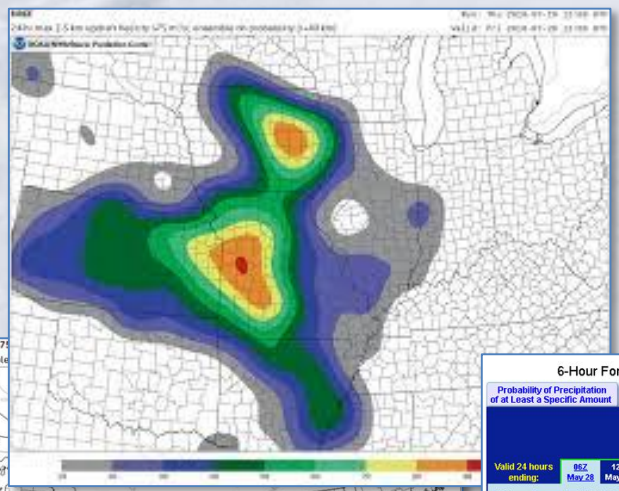


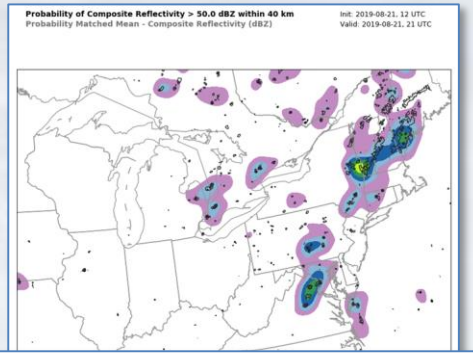
Graphics from NOAA/NSSL

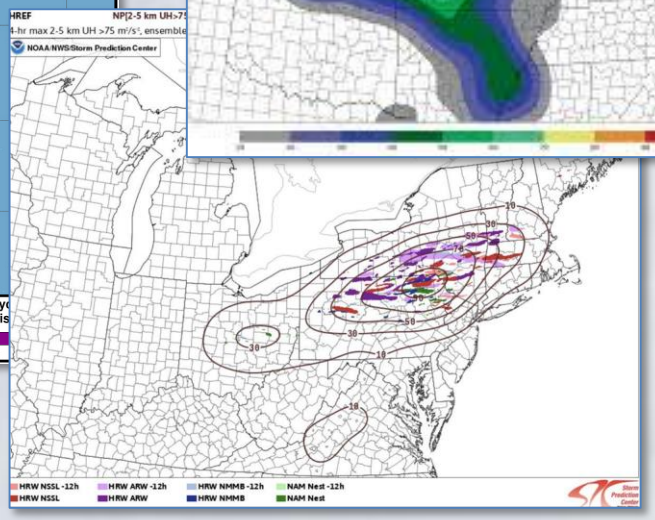
Human/Machine Interdependence: Ensembles and probabilities

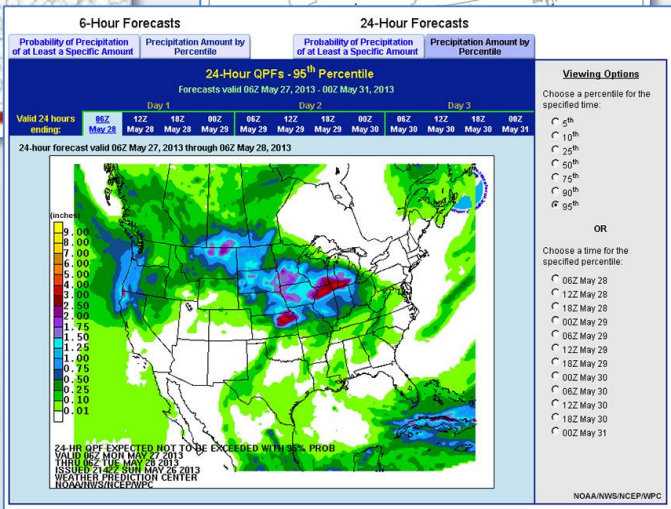
- Nearly every forecast element can be represented by E&P
- Presentation methods vary: Paintball, heat maps, prob matched mean, cone, etc.
- Forecasters MUST fully understand E&P concepts, benefits, and limitations to properly interpret this info for decision-makers and the public







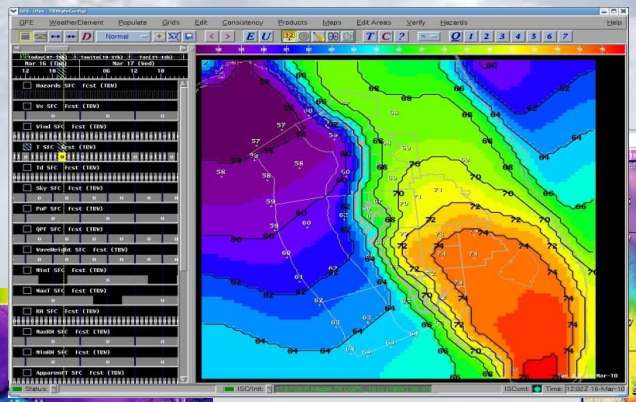


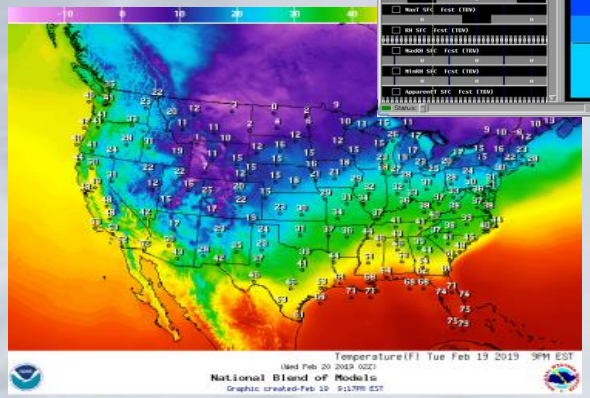


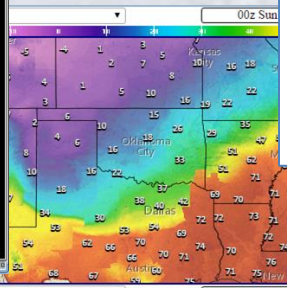
Images from NOAA/NWS

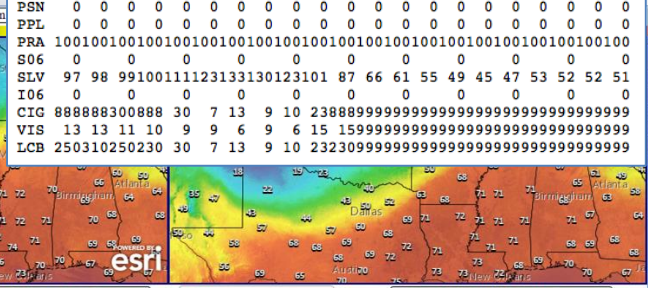
Human/Machine Interdependence: Model Blends/NBM

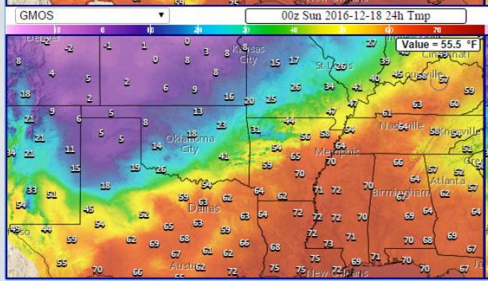
- NWS gridded forecasts often start with a base of model output
- NWS forecasters adjust output based on local knowledge, observations, remote sensing, bias corrections, etc.
- Human-adjusted forecast generally improves over pure model-blended guidance…
- But that guidance is quickly gaining ground

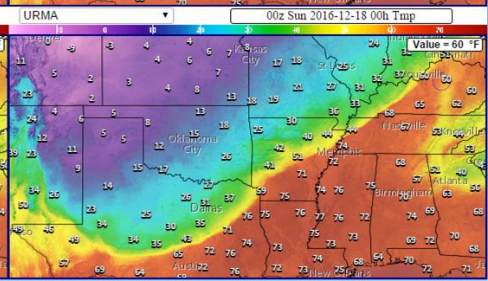










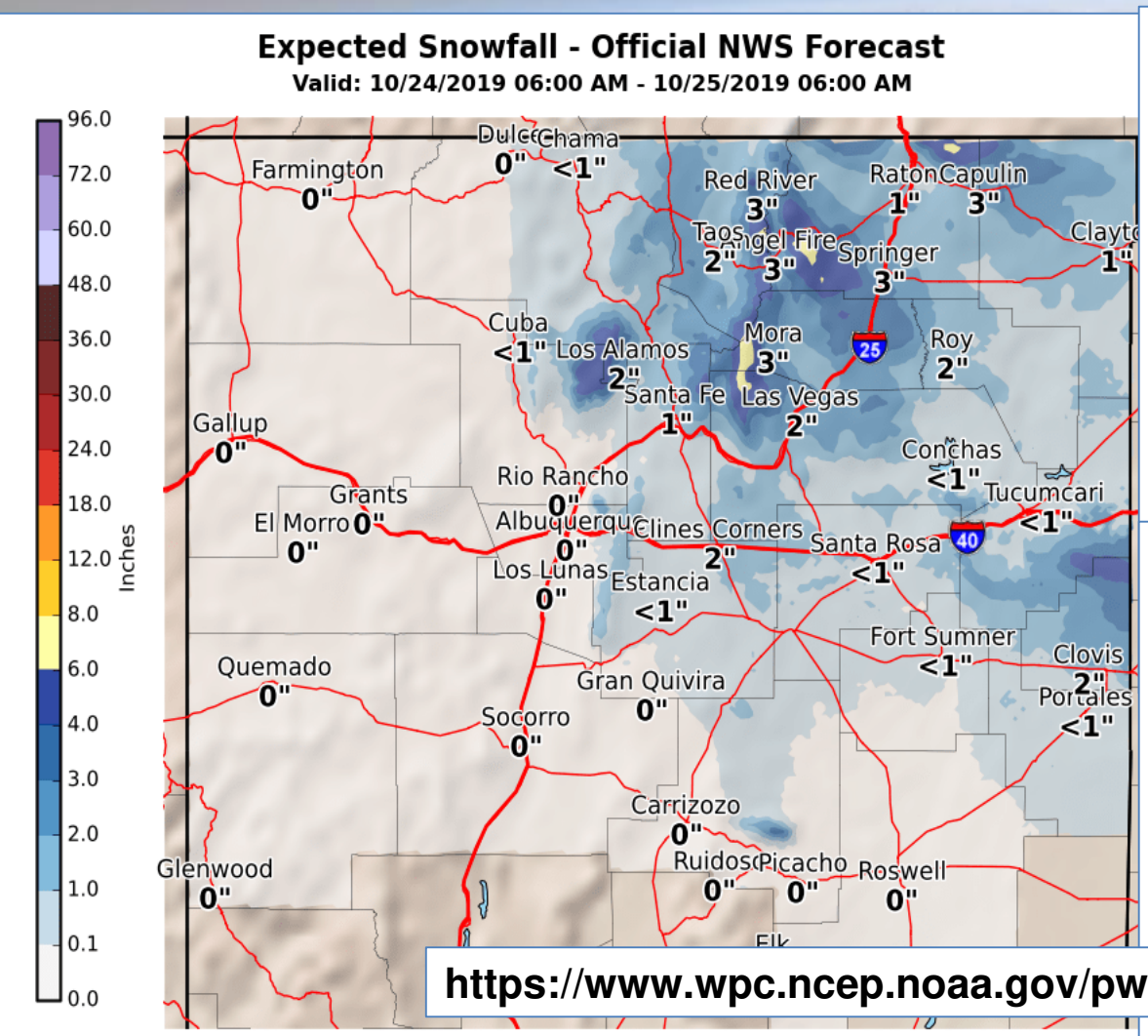


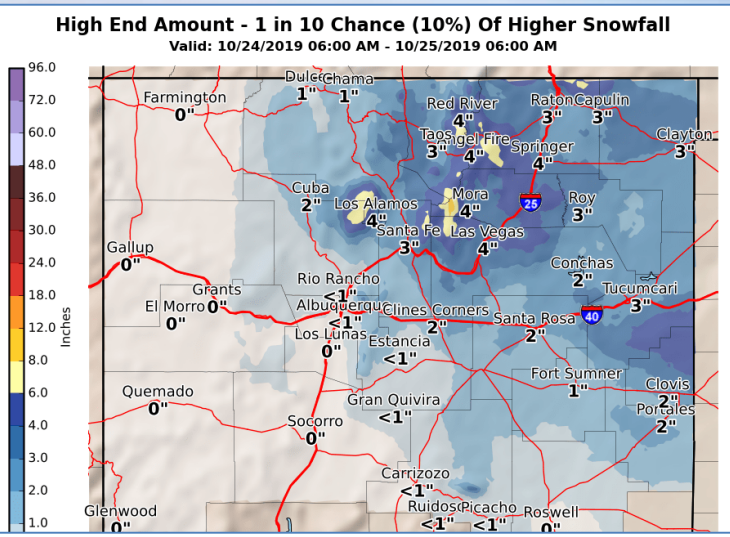

```
KRDU     NBM NBS GUIDANCE    10/21/2019  1200 UTC
DT /OCT  21/OCT  22                /OCT  23                 /
UTC   18 21 00 03 06 09 12 15 18 21 00 03 06 09 12 15 18 21 00 03 06
FHR   06 09 12 15 18 21 24 27 30 33 36 39 42 45 48 51 54 57 60 63 66
N/X      73          58          75          48          67
TMP   70 70 63 60 60 60 61 66 71 73 65 60 55 51 49 60 65 64 54 50 48
DPT   59 60 61 60 60 60 61 65 69 69 65 57 50 46 44 44 41 42 44 43 41
SKY   38 43 78 52 67 84 96 94 81 63 56 49 17  8  4  1  0  0  1  2  3
WDR   00 00 13 13 14 13 13 14 18 23 27 31 31 31 31 32 31 32 00 00 00
WSP    0  0  1  1  1  3  4  5  5  4  3  3  2  2  1  3  3  1  0  0  0
GST    1  1  2  3  4  5  7 15 14 12  9  7  5  5  4  8  6  4  0  0  0
P06    0     0     0    14    95    94     7     0     0     0     0
P12          0          14          95          7           0
Q06    0     0     0     0    35    27     0     0     0     0     0
Q12          0          0           62          0           0
T03    0  0  0  0  0  0  0  2  5 11  5  1  0  0  0  0  0  0  0  0  0
T12          0          0           19          1           0
PZR    0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
PSN    0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
PPL    0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
PRA 10010010010010010010010010010010010010010010010010010010010010
S06    0     0     0     0     0     0     0     0     0     0     0
SLV   97 98 99100111123133130123101 87 66 61 55 49 45 47 53 52 52 51
I06    0     0     0     0     0     0     0     0     0     0     0
CIG 888888300888 30  7 13  9 10 23888999999999999999999999999999999
VIS   13 13 11 10  9  9  6  9  6 15 15999999999999999999999999999999999
LCB 250310250230 30  7 13  9 10 23230999999999999999999999999999999
```

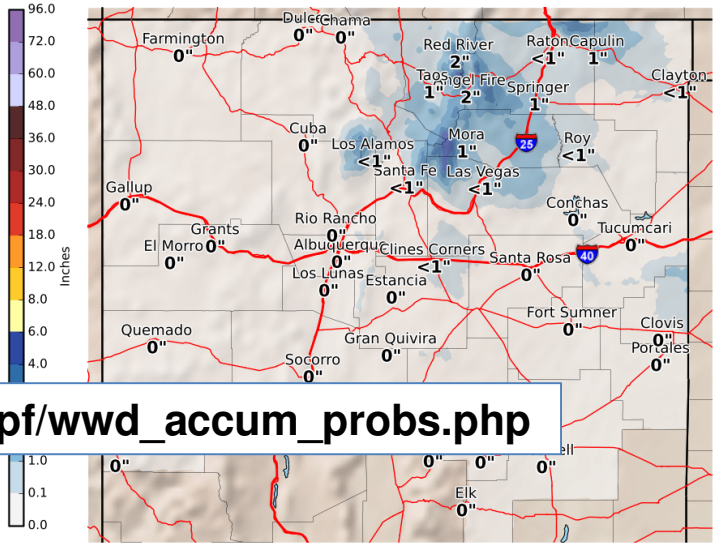
Images from NOAA/NWS

Human/Machine Interdependence: PWPF (Probabilistic Winter Precip Forecast)

Images from NOAA/NWS







https://www.wpc.ncep.noaa.gov/pwpf/wwd_accum_probs.php

National Weather Service
Albuquerque, New Mexico
10/24/2019 03:35 AM MDT

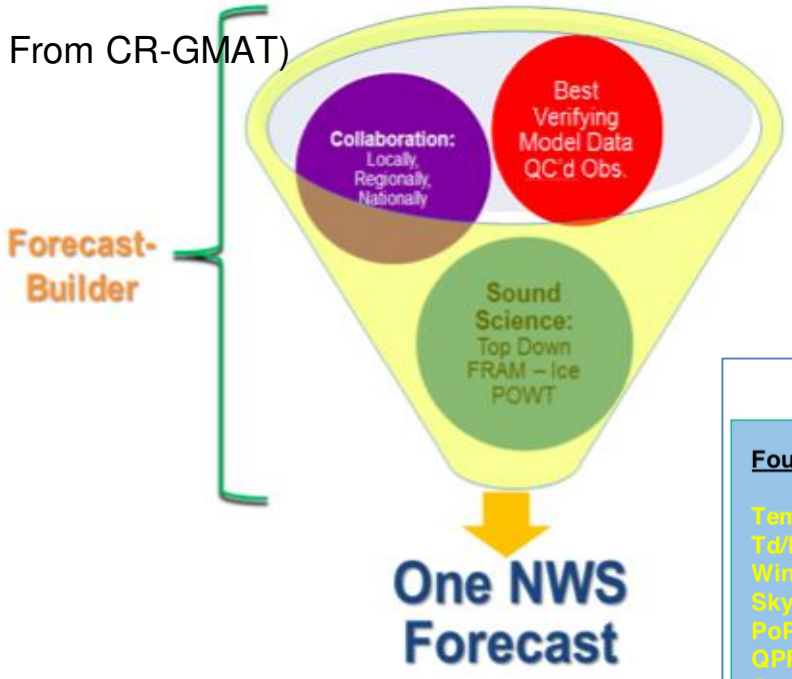
Follow Us:
weather.gov/abq/winter

National Weather Service
Albuquerque, New Mexico
10/24/2019 03:37 AM MDT

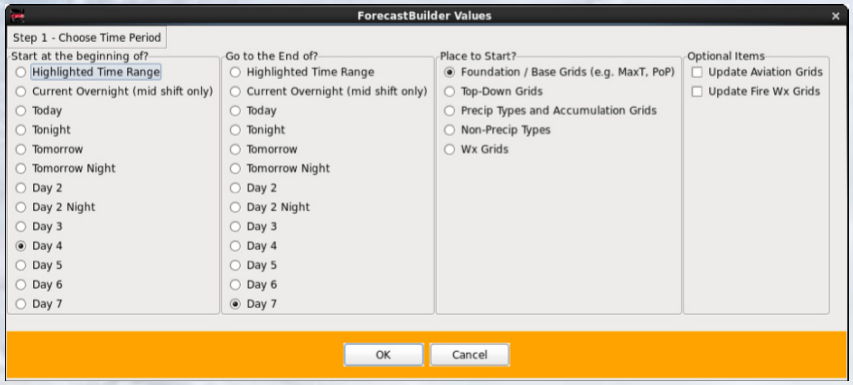
Follow Us:
weather.gov/abq/winter

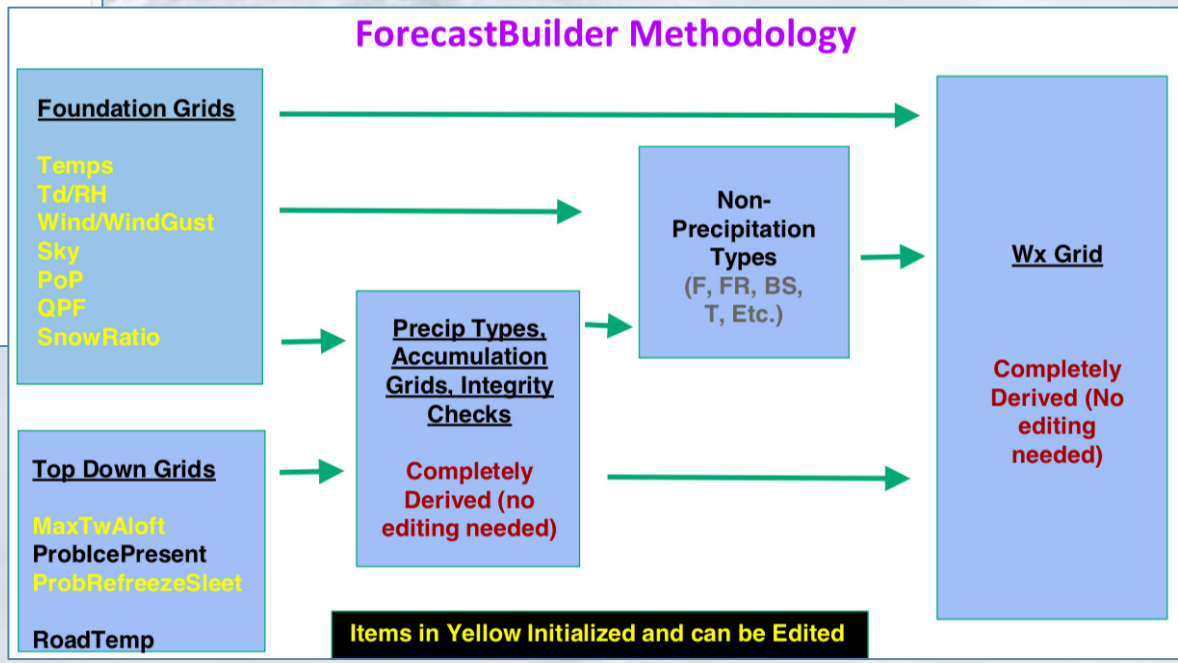
Human/Machine Interdependence: NWS ForecastBuilder

From CR-GMAT)



Images from NOAA/NWS

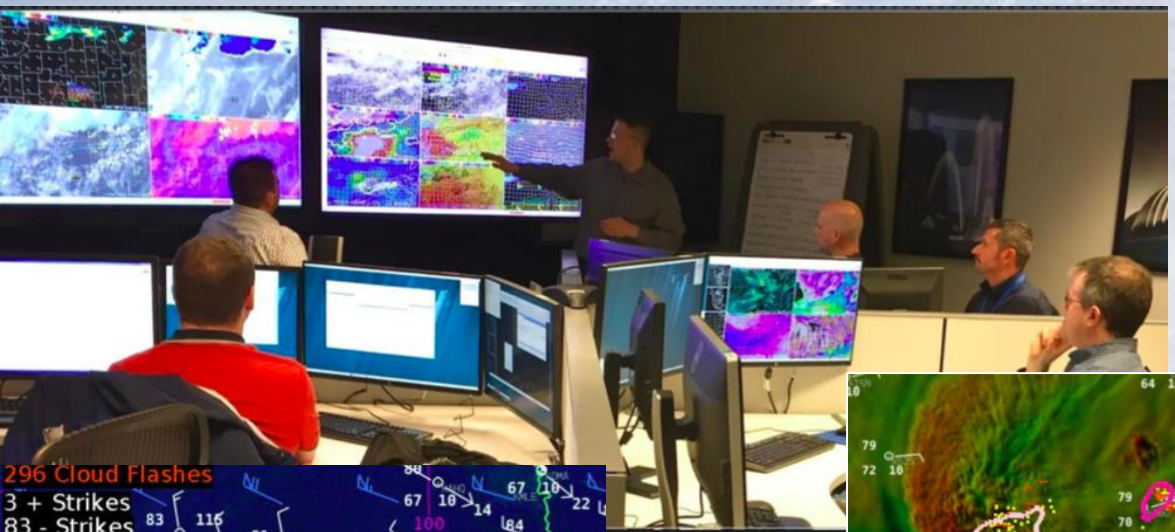


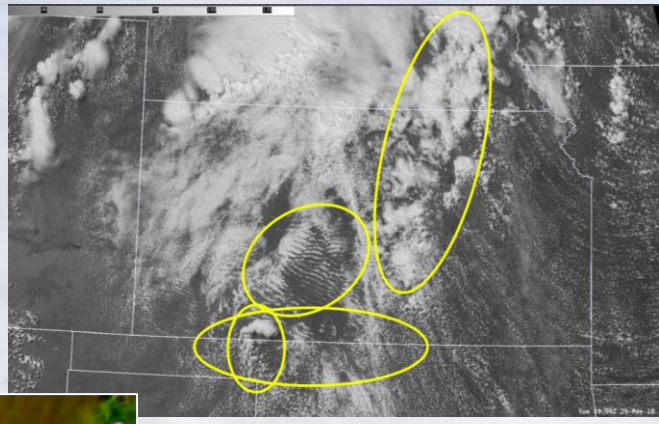


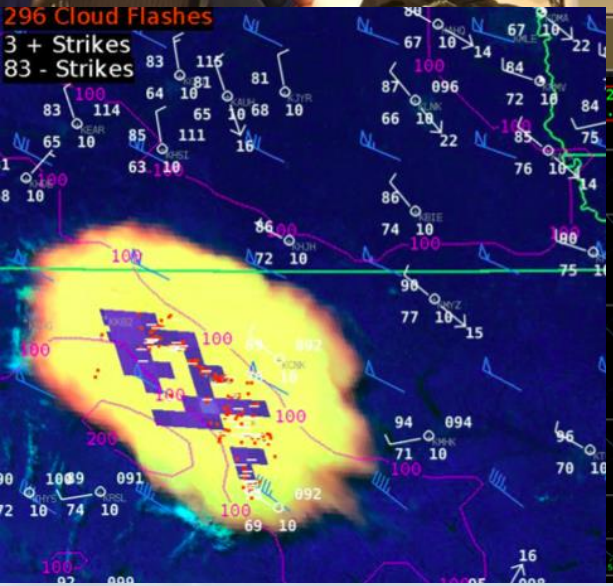
Forecasters are more freed up for IDSS, mesoanalysis, forecast challenges, severe weather monitoring, training, research, etc.

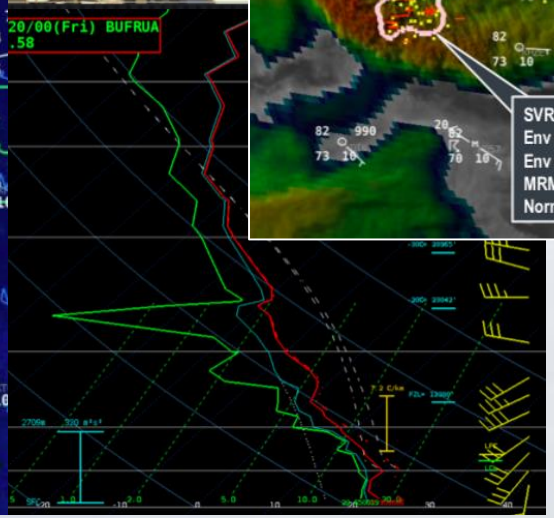
…But forecasters MUST stay involved… can't fall prey to inattention or "good enough"

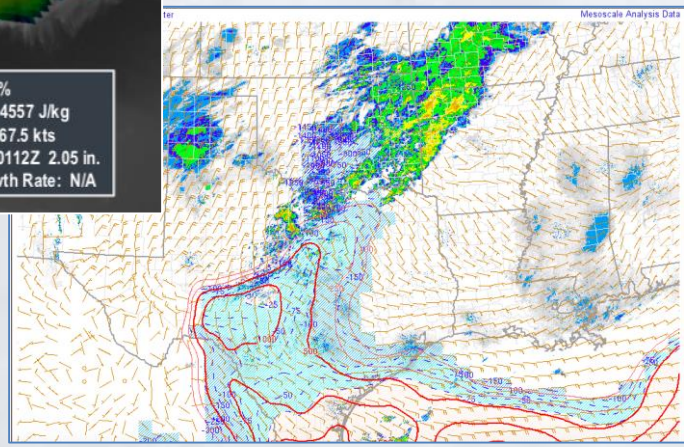
Human/Machine Interdependence: Mesoanalysis











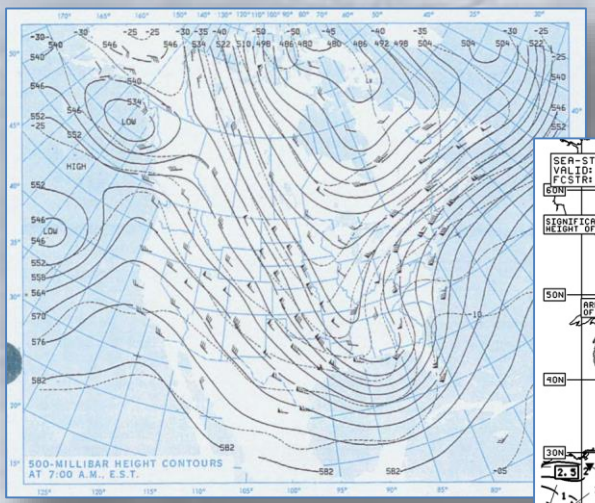
Images from Kim Runk (NWS OPG) and NOAA/SPC

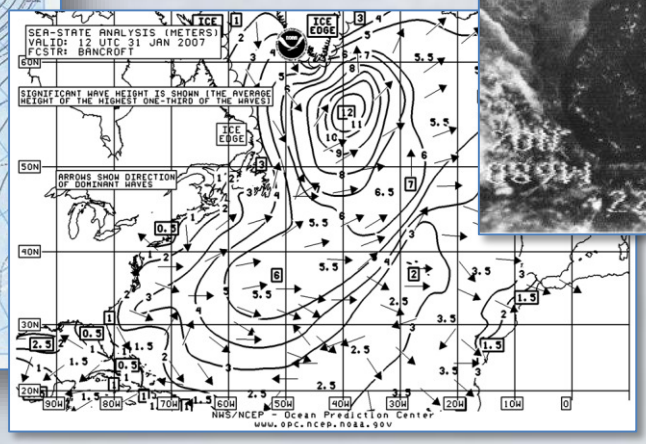
…Human synthesis and interpretation remains vital! (& all must be proficient!)

What Our Next Steps Must Be

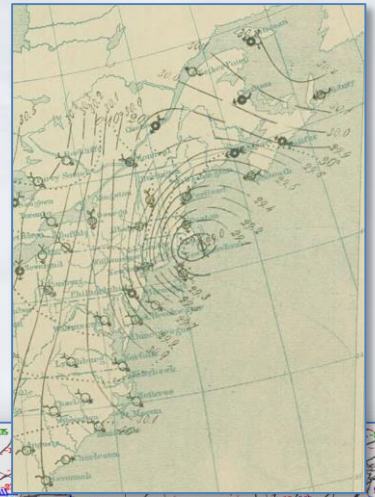
What forecaster involvement used to be…

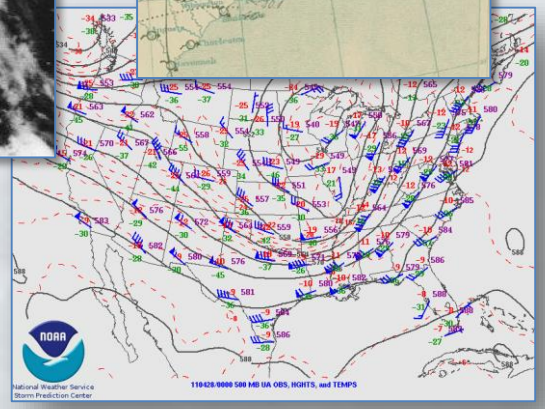
- Pattern recognition and data analysis were (and are) human strengths
- Forecasters needed a deep knowledge of physical and thermal processes
- Scant model output needed to be interpreted (much was inferred)
- Humans determined specifics of the forecast
- Humans create work-arounds when machines fail





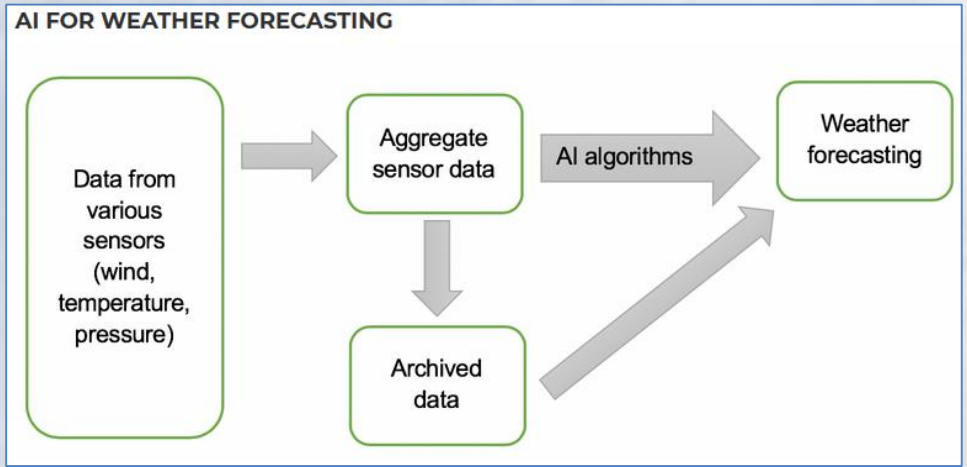
Images from NOAA/NWS and OPG





Forecasters must accept changes and adapt to successfully keep forecasting moving forward

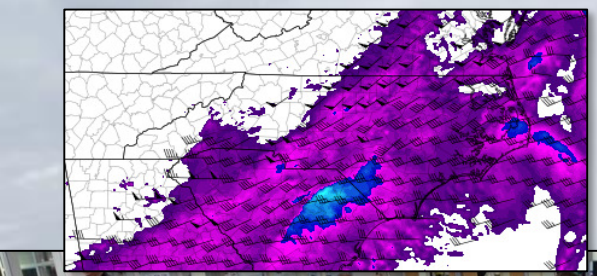
- AI is working to replicate many aspects of human talent, but can never replace the most important human roles in forecasting and communication: **decision making and describing uncertainty**
- Model output becoming more abundant and available, in user-friendly formats, with extremely fine detail *(although precision does not equal accuracy)*
  - **Concern: Is CAM output too realistic for public consumption?**
- Do forecasters need to know the "whys" of AI in order to trust? (may not be realistic for deep learning systems)
- **Bottom line: Need comprehensive training in use (and misuse)**

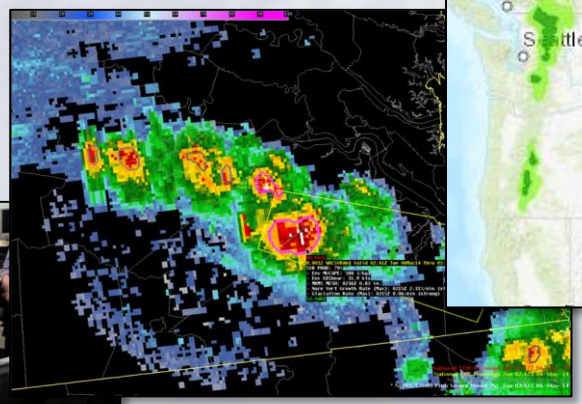


Graphic via Allerin

More experimentation is needed to determine the optimal relationship between humans and machines/automation

- E.g., controlled study of human-only warning vs. AI-assisted human warning vs. AI-only warning
- Need deeper verification (not just max/min temps and QPF but mixing heights, cloud ceiling heights, precip type, regional and local verification, etc.)
- Good visualization of AI information is crucial for good-quality human involvement
- Refine/improve QC'ing techniques for initial conditions



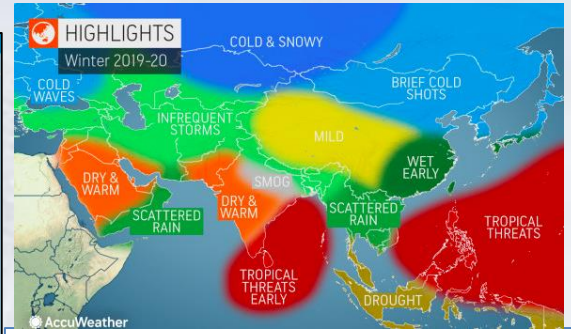


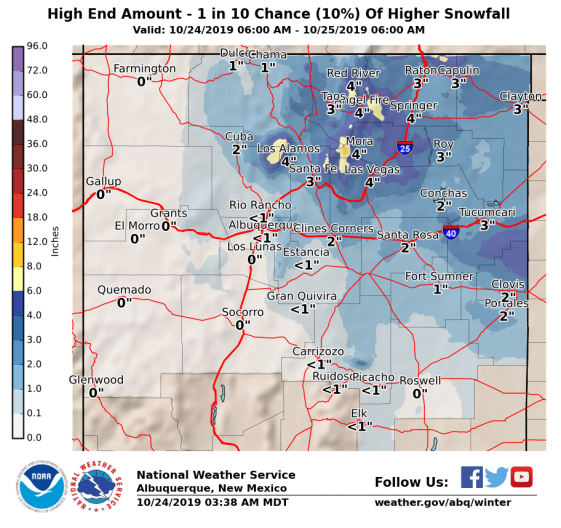
Images from NOAA/NWS and NSSL

Humans excel as interpreters and communicators; this must remain a focus

- Use our empathy, social cues, user feedback
- People still want to talk to people
- Varied audiences require varied communication methods

Images from TWC, NOAA/NWS, Accuweather, Phandroid, James Spann





The Human-Machine Relationship is not a single state!

Think of how you relate to and work using the software on your computer….

Humans ALWAYS Feel some degree of

Justified AND Unjustified *Trust* in the Machine
AND
Justified AND Unjustified *Mistrust* in the Machine

We trust our devices to do certain things well under certain circumstances but not to do other things well.

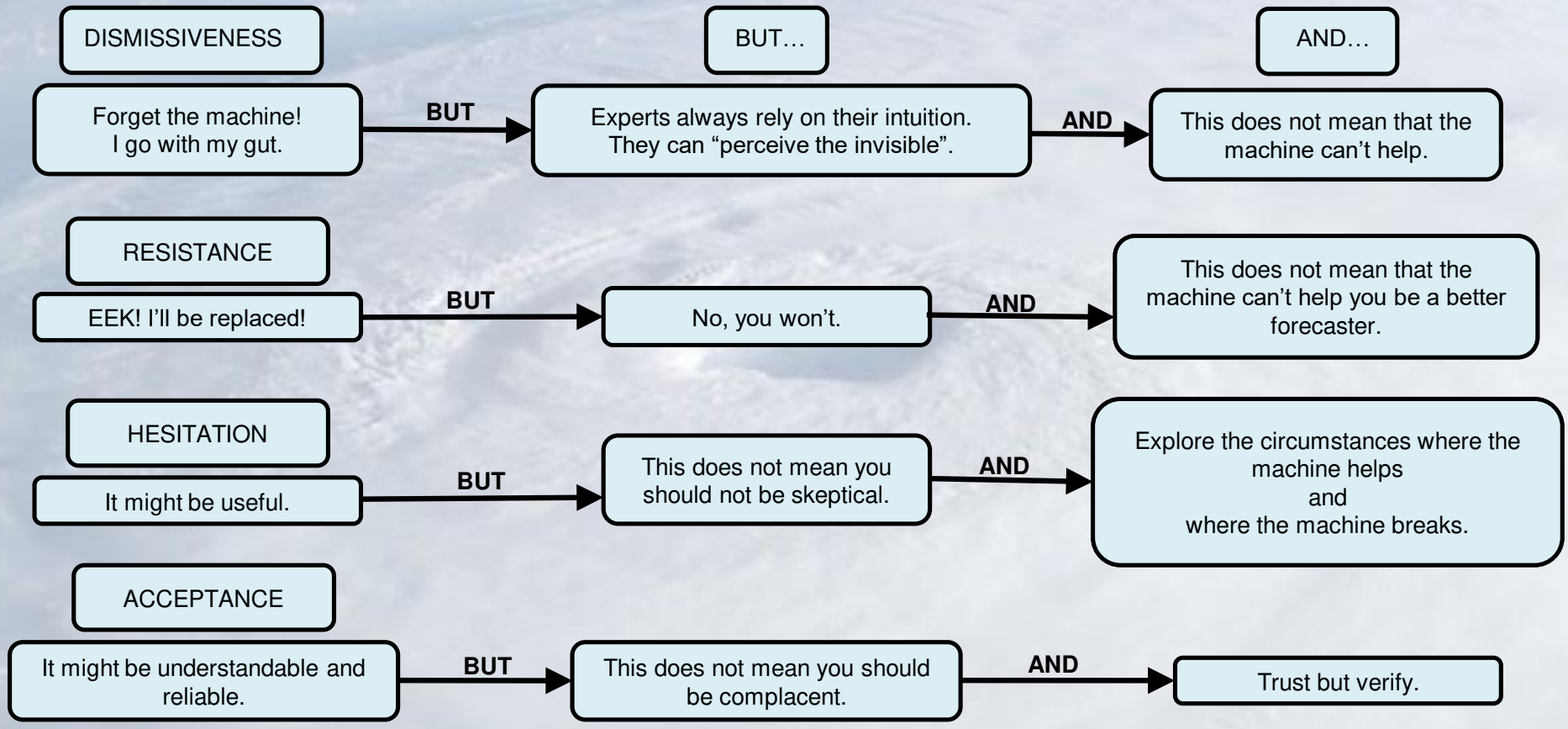
We mistrust our machines under certain circumstances, sometimes even when we should trust them.

Trust in the Machine is not a desirable goal state or target.

Trusting is a complex, dynamic relationship.

The human must explore how the machine operates when it is pushed to the edges of its competence envelope.

What is YOUR relationship to the computer?



Closing Thoughts

- Forecaster involvement will likely be greater in the realms of:
  - Short term
  - High detail
  - Non-standard fields
  - Interpretation for users/customers
- Weather enterprise will need to determine how to best fold increasing AI influence into forecast process on different stages (build the car while we're driving it)
- Build on in-progress collaborations
- Frequent metrics / verification & testbeds
- Train forecasters to be comfortable with new roles
- Applied forecasting and impact communication

Closing Thoughts

- Forecaster involvement will likely be greater in the realms of:
  - Short term
  - High detail
  - Non-standard fields
  - Interpretation for users/customers
- Weather enterprise will need to determine how to best fold increasing AI influence into forecast process on different stages (build the car while we're driving it)
- Build on in-progress collaborations
- Frequent metrics / verification & testbeds
- Train forecasters to be comfortable with new roles
- Applied forecasting and impact communication

Tune in for the next 2 webinars in this series, coming soon!