Using Machine Learning to improve 1-h Low-level Rotation Forecasts

Montgomery L. Flora[1,3,4], Corey Potvin[1,3,4], Amy McGovern[1,2], Patrick Skinner[3,4]

[1]School of Meteorology, University of Oklahoma, Norman, OK; [2]School of Computer Science, University of Oklahoma, Norman, OK;
[3]Cooperative Institute for Mesoscale Meteorological Studies, Norman, OK ; [4]NOAA/OAR/National Severe Storms Laboratory, Norman, OK

BACKGROUND

NOAA Warn-on-Forecast (WoF): a promising effort to improve warning lead times by providing rapid-update probabilistic guidance to human forecasters (Stensrud et al. 2009, 2013).

Low-level rotation on 3-km grid **may potentially discriminate well between tornadic and non-tornadic** (Wheatley et al. WAF 2015; Sobash et al. WAF 2016; Skinner et al. WAF 2016)

Current storm-scale forecasting techniques rely on uncalibrated **probability of exceedance** (**PoE**; # of ensemble members > threshold / ensemble size)

However, CAMs often contain **large errors in storm intensity, timing, and location**

Machine learning can **leverage ensemble uncertainty, incorporate several model variables, and mitigate forecast bias to produce calibrated probabilistic forecasts** (McGovern et al. BAMS 2017)

DATA SOURCES

NSSL Experimental WoF System for ensembles (NEWS-e) model output
- 18-member ensemble forecast with 3-km grid spacing
- Initialized every 30 minutes starting at 1900 UTC until 0300 UTC
- Generated during the 2016 & 2017 NOAA Hazardous Weather Testbed Spring Forecasting Experiments (total of 30 dates)

NSSL Multi-Radar Multi-Sensor low-level (0-2 km AGL) azimuthal shear
- Azimuthal shear ~ ½ vorticity
- Quality-checked and remapped to the NEWS-e grid

PREDICTORS

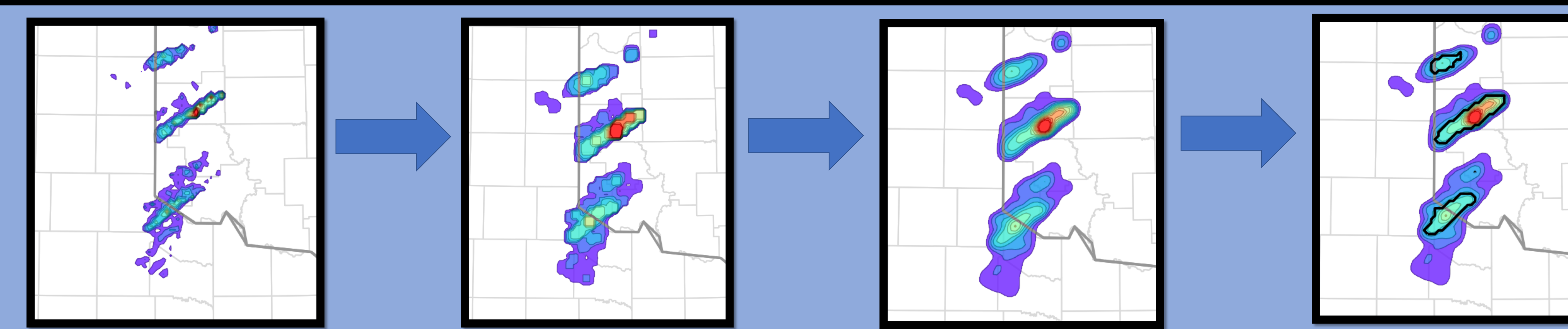
6 ensemble statistics are calculated for 21 variables (see below)
- 0 (min), 25, 50, 75, 100 (max) percentiles and standard deviation
- Both for raw and gradient magnitude fields
- Maximized in 1-h window

Thermodynamics		Kinematics	Storm	
*ML LCL	2 m T_d	0-1 km SRH	0-2 km Vorticity	Max Updraft
*ML CAPE	*ML θ_e	0-3 km SRH	0-2 km UH	Rainfall
*ML CIN	2 m Temp	0-1 km U shear	2-5 km UH	*ML STP
SFC Pressure	2 m Q_v	0-6 km U shear	Composite dBZ	
2 m θ	*ML Q_v	0-1 km V shear	Max Hail	

*ML: 0-75 mb Mixed layer

Funding provided by NOAA – University of Oklahoma Cooperative Agreement NA11OAR4320072, U.S. Department of Commerce.

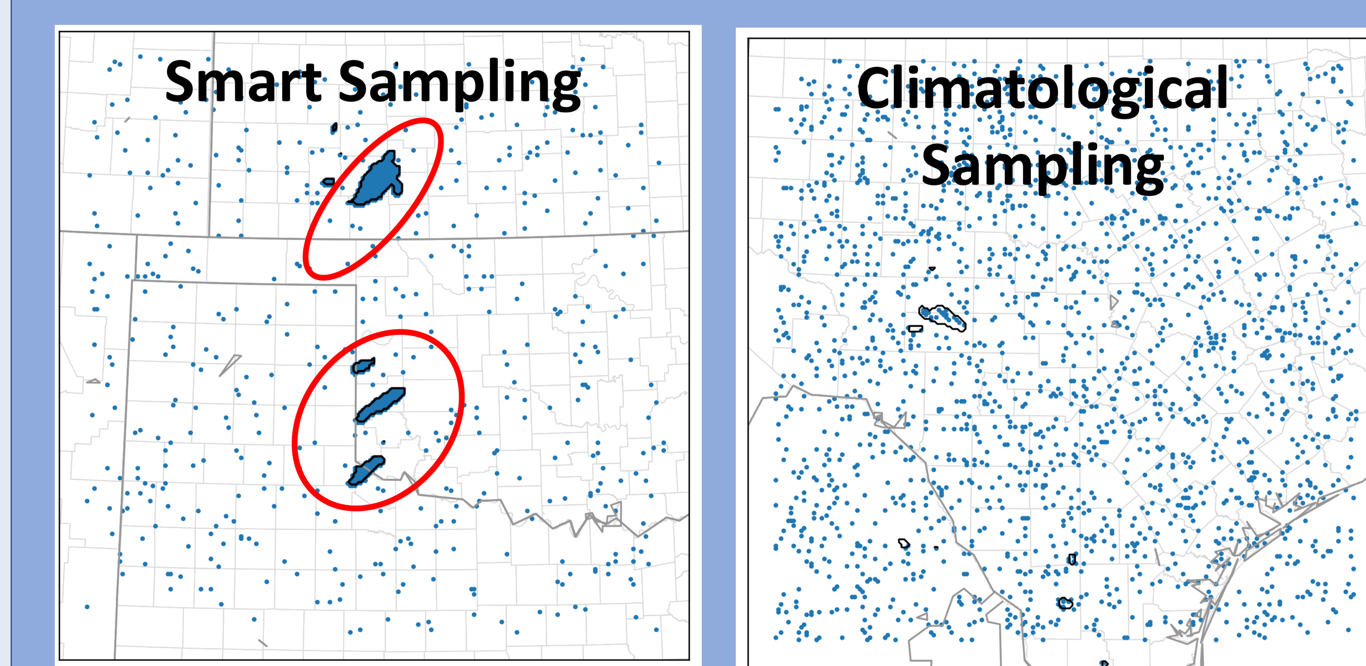
DEFINING SIGNIFICANT LOW-LEVEL ROTATION



Maximize 2*Azimuthal Shear in 1-h window → Spatial Maximum in 3 grid point radius → Convolve with Gaussian Kernel (2 grid point radius) → Significant low-level rotation > 0.01 s^{-1}

Smoothing Process

DATA EXTRACTION



Smart Sampling
- Balanced training datasets improve model performance
- For each date and initialization time, extract all data from regions within the significant low-level rotation
- Randomly extract from outside in equal portion

Climatological Sampling
- Needed for calibrating the probabilities
- Randomly sample 3% of domain with 3.5% of points coming from within the low-level rotation regions

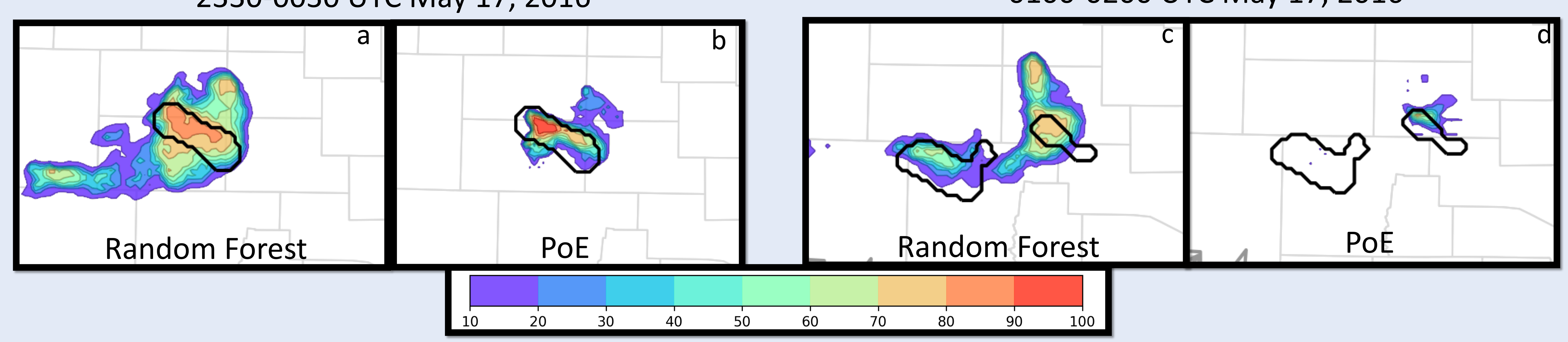


Fig. 1 Random forest probability of double the observed azimuthal shear > 0.01 s^{-1} (a,c). Uncalibrated probability of 0-2 km AGL UH > 14 m^2s^{-2} within a 3 grid point radius. The 14 m^2s^{-2} threshold has been associated with the 0.01 s^{-1} double azimuthal shear threshold following the methodology of Skinner et al. WAF 2016.

Early conclusion: Machine learning may fail to improve upon uncalibrated PoE in some cases (cf. Fig. 1a,b), but may also reduce missed events (cf. Fig. 1c, d)

NEXT STEPS…

- Employing an object-matching forecast verification approach
- Employing additional algorithms such as logistic regression and gradient-boosted trees
- Removing storm predictors to evaluate the discrimination ability of environment variables