Near Median Density Index

An order statistics approach to assessing ensemble forecast uncertainty

Keith F. Brill
I.M. Systems Group, Inc., and NOAA/NWS/NCEP/Weather Prediction Center

Motivational Considerations

1. Forecasters now provide Decision Support Services (DSS) routinely, which can be Impact-based (IDSS).
2. Forecasters must be ready to communicate the degree of forecast uncertainty as part of DSS.
3. Forecasters must identify potential forecast problems quickly (where and when does mentionable uncertainty exist in the guidance).
4. Forecasters still issue (and users still want) deterministic forecasts.

Ensemble spread measures uncertainty, right? Yes, *but* . . .

1. Spread is not a resistant statistic---it is sensitive to outliers.
2. Sometimes spread can be "high"[1], but the ensemble has a strong *central tendency*[2] and there is *probably* no better forecast than the ensemble mean or median. **Call this "typical uncertainty".**
3. Sometimes spread is "high", but the ensemble has a weak *central tendency* and the mean or median is questionable. **Call this "split-decision uncertainty".**

Is it typical or split-decision uncertainty? The forecaster likely needs to know!

[1]"High spread" is defined as spread values exceeding one's tolerance for the forecast error for a parameter. "Low spread" is spread that is not high.

[2]"Central tendency" refers to the degree to which ensemble values for a scalar parameter are clumped into a single mode in their frequency distribution.

The Near Median Density Index (NMDI) defined:

Let *F(x)* be the cumulative distribution function (CDF) for the ensemble order statistics[1] at a point.

$F^{-1}(p)$ is the inverse CDF that returns a value *x* for given cumulative probability, *p.*

Let $\sigma_r = [\ F^{-1}(.96) - F^{-1}(.04)\ ] / (2\sqrt{3})$, a resistant spread estimate.[2] Let $u = F^{-1}(.50)$, the median.

$$NMDI = F(u + \tfrac{1}{4}\sigma_r) - F(u - \tfrac{1}{4}\sigma_r), \text{ given } \sigma_r > 0.$$

From the definition: **0 ≤ NMDI < 1**

High spread and ***low NMDI*** indicate split-decision uncertainty, but how low is “low” for NMDI?

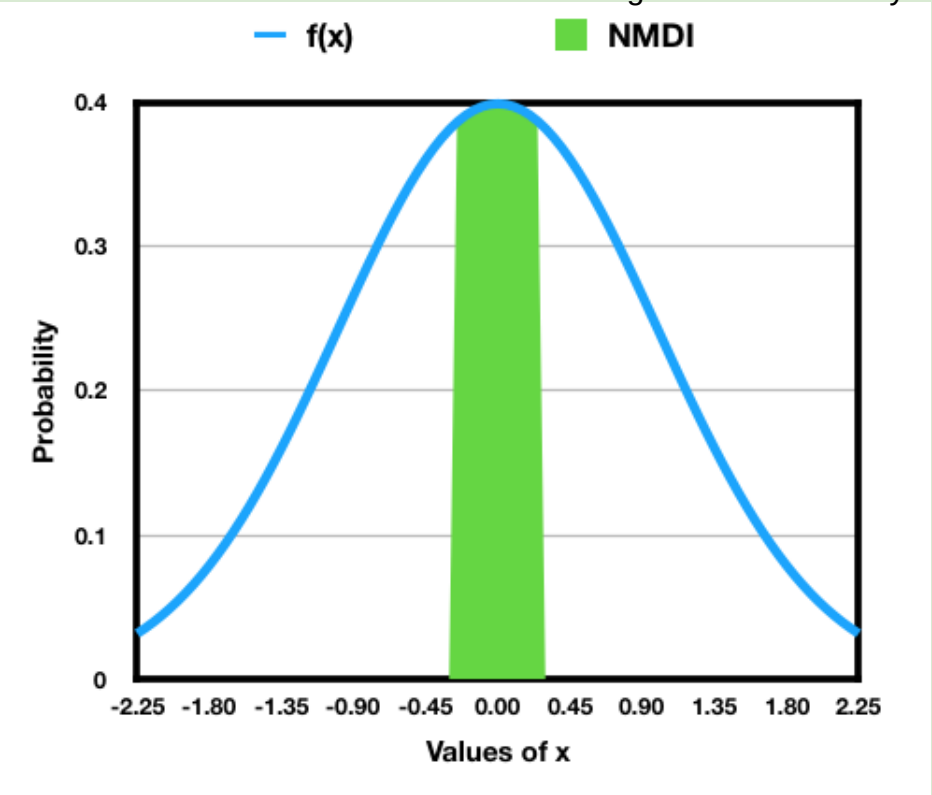
Let’s look at this index for several analytic probability density functions (PDF).

[1]The ensemble member values for some forecast parameter *x* sorted from lowest to highest.

[2]Inspired by the standard deviation formula for the uniform distribution.

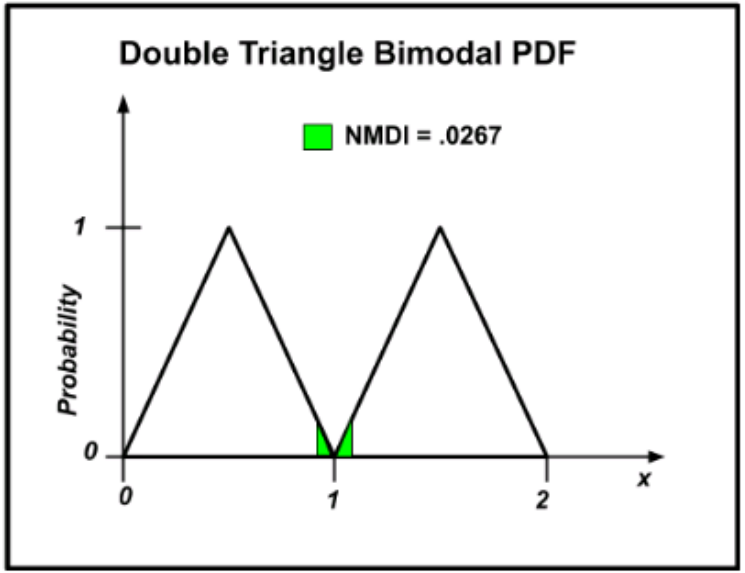
Starting with graphs:

Strong Central Tendency



Normal PDF: NMDI ≈ .2 ⟶ Typical Uncertainty

Weak Central Tendency



Bimodal PDF: NMDI ≈ .03 ⟶ Split-Decision Uncertainty

Now, let's find a threshold value . . .

Table of Analytic PDFs

The exponential PDF has relatively strong central tendency (typical uncertainty) according to NMDI even though σ_r under estimates actual σ.

For normally or near normally distributed data, σ_r gives a good estimate of the actual σ, and NMDI values can be regarded as indicating typical uncertainty.

The uniform PDF is the basis for σ_r and suggests a threshold value (~.15) below which NMDI indicates split-decision uncertainty.

For a bimodal PDF, σ_r *beneficially* under estimates σ driving the NMDI lower to strongly indicate split-decision uncertainty.

PDF name	PDF description	Median υ	Mean μ	σ	σ_r	$\frac{\sigma_r - \sigma}{\sigma}$ X 100%	NMDI
Exponential	Parameter = 1	.6931	1	1	.9174	− 8.26%	.2314
Normal	Standard	0	0	1	1.0108	+ 1.08%	.1995
Binormal (Toth and Szentimrey 1990, *J. Climate*)	Right Skewed Mode = 0 Left $\sigma = .473858$ Right $\sigma = 1.42157$	.6123	.7562	1	.9851	− 1.49%	.1882
Uniform	Interval $[0, 2c\sqrt{3}]$, $c > 0$	$c\sqrt{3}$	$c\sqrt{3}$	c	$.92c$	− 8.00%	.1328
Double Triangle	Interval $[0, 2]$ Two adjacent isosceles Δs of area $\frac{1}{2}$ each	1	1	.5401	.4619	− 14.48%	.0267

Using NMDI with Spread

1. Assume spread meets the *high spread* criterion.*
2. **NMDI ≥ .15** indicates ***typical uncertainty*** and the ensemble mean or median is *probably* the best forecast.
3. **NMDI < .15** indicates ***split-decision uncertainty.*** The lower the value of NMDI below the .15 threshold the less trustworthy the ensemble mean or median.
    a. Suggests importance of consulting other guidance or ensemble clusters
    b. Heightens importance of communicating the uncertainty
    c. Suggests just waiting for the next forecast cycle

***Using the NMDI alone false alarms where spread is low.**

Example Application for a Heat Index Forecast

Using the 20-member NCEP GEFS

Initial time & date: 00 Z 17 July 2019
Valid time & date: 00 Z 22 July 2019

F120 Ensemble Mean Heat Index (HI, deg F) is contoured (brown, labeled, interval=4)

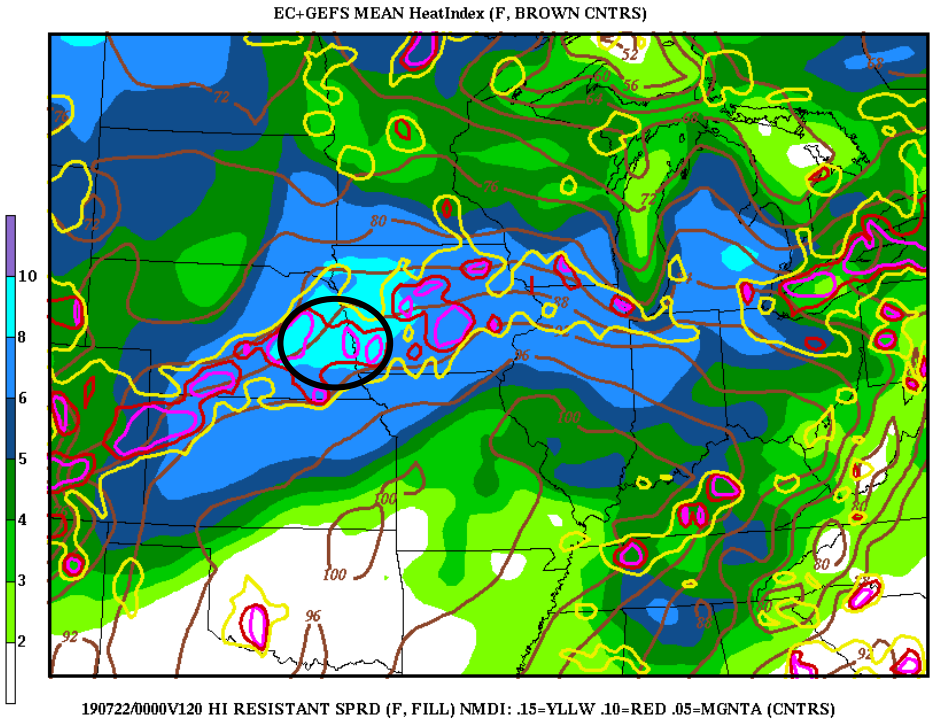
Resistant spread (F) is color filled (see color bar)

Low NMDI areas are outlined by yellow contours enclosing lower red and magenta contours

Highlight:

Western IA and eastern NE are areas of noticeable split-decision uncertainty

Midwest-Plains Heat Wave



Example Application for a Precipitation Forecast

Using the 26-member NCEP SREF

Initial time & date: 03 Z 13 July 2019
Valid time & date: 18 Z 14 July 2019

F39 Ensemble mean 6-h QPF (inch) is color filled (see color bar)

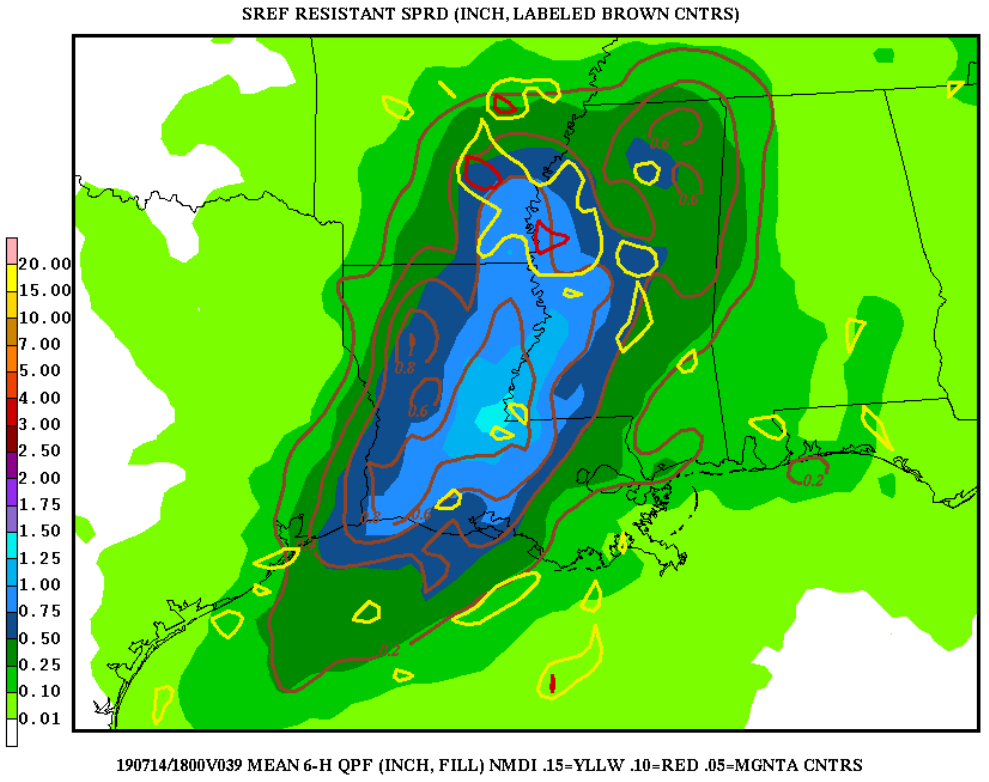
Resistant spread (inch) is shown by labeled brown contours at interval=.2

Low NMDI areas are outlined by yellow contours enclosing lower red contours

Highlights:

- Noticeable region of split-decision uncertainty for northern portion of heavy (>.5 inch) QPF area
- Split-decision uncertainty tends to be on the periphery of QPF areas where placement is often indecisive

Hurricane Barry Case



Summary

- A ***resistant spread*** can be computed from ensemble order statistics.
- High spread with strong central tendency indicates ***typical uncertainty***; so, the ensemble mean or median is *probably* the best forecast.
- High spread with weak central tendency indicates ***split-decision uncertainty*** which may call for human intervention in the forecast process and more attention to conveying the degree of uncertainty for DSS.
- **The *Near Median Density Index (NMDI)* along with ensemble spread provides a way to distinguish between typical and split-decision uncertainty.**

Summary

- A ***resistant spread*** can be computed from ensemble order statistics.
- High spread with strong central tendency indicates ***typical uncertainty***; so, the ensemble mean or median is *probably* the best forecast.
- High spread with weak central tendency indicates ***split-decision uncertainty*** which may call for human intervention in the forecast process and more attention to conveying the degree of uncertainty for DSS.
- **The *Near Median Density Index (NMDI)* along with ensemble spread provides a way to distinguish between typical and split-decision uncertainty.**

THE END
Thank You