TOWARDS ECOLOGICAL VALIDITY IN EVALUATING UNCERTAINTY

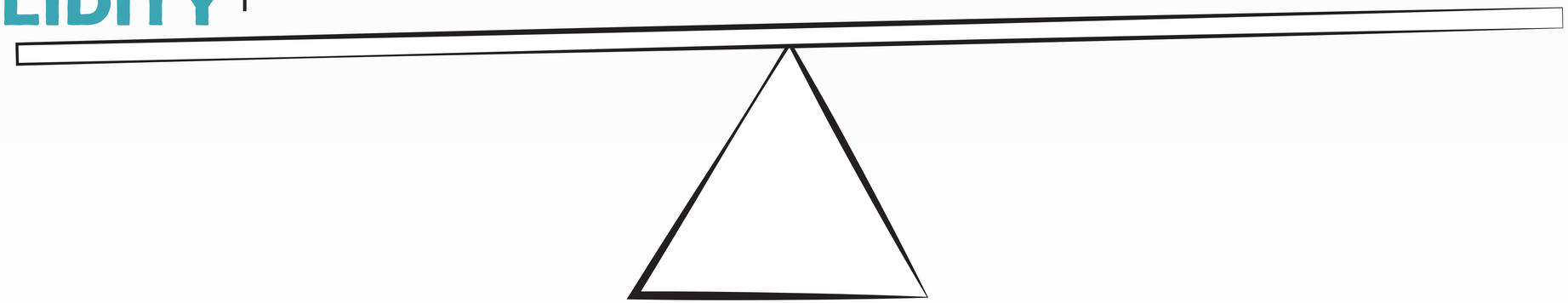
P Samuel Quinan
Lace M. Padilla
Sarah H. Creem-Regehr
Miriah Meyer

ECOLOGICAL VALIDITY [1]

[1] how closely the experimental setting matches the setting in which the results will be applied

ECOLOGICAL VALIDITY [1]

EXPERIMENTAL CONTROL



[1] how closely the experimental setting matches the setting in which the results will be applied

previous **design study** centered around improving the efficacy of visualizations in meteorological decisions

previous **design study** centered around improving the efficacy of visualizations in meteorological decisions

previous **design study** centered around improving the efficacy of visualizations in meteorological decisions

determining how visualizations affect decisions is incredibly difficult

previous **design study** centered around improving the efficacy of visualizations in meteorological decisions

determining how visualizations affect decisions is incredibly difficult

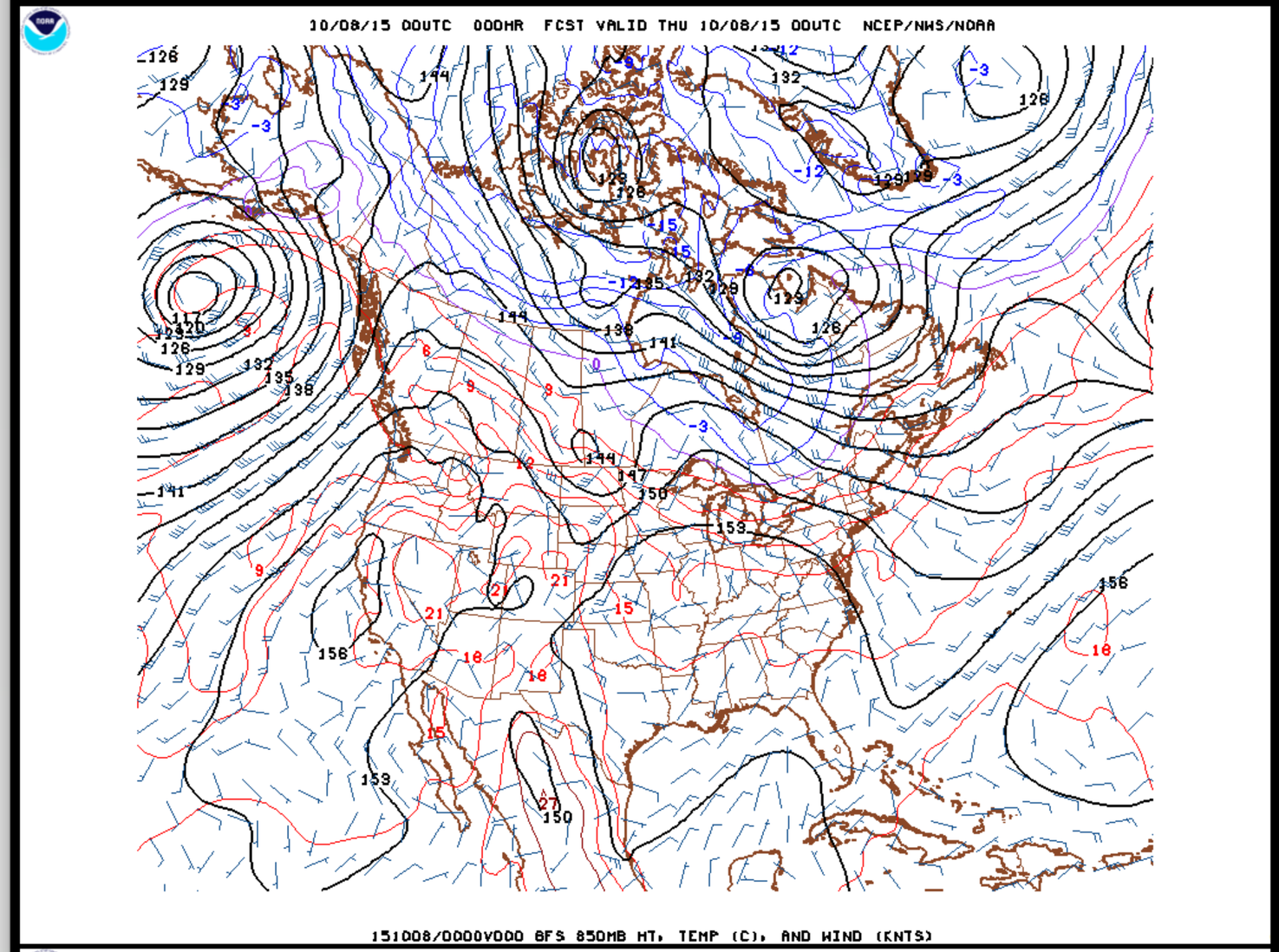
decision-making contexts may require fundamentally different evaluations

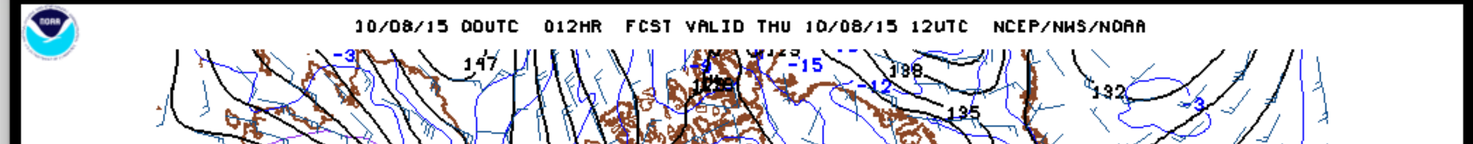
THE FORECASTING PROCESS

ggweather.com/ | eta12hr_850_tm | FNMOC WXMAP | nam_700_4pane | SPC SREF DProg | ECMWF | World | surface_prs_87.g | You

ggweather.com/loops/gfs_00z_850_all.htm

Please be patient while the images load!





ggweather.com/ | eta12hr_850_tm | FNMOC WXMAP | nam_700_4pane | SPC SREF DPro | ECMWF | World | surface_prs_87.(

https://www.fnmoc.navy.mil/wxmap_cgi/cgi-bin/wxmap_single.cgi?area=nvg_namer&dtg=2015100812&prod=500&tau=000&set=All

Models | Product List | Help

Model, Area : NAVGEM, North America, .5 deg
Compare : GFS N Am

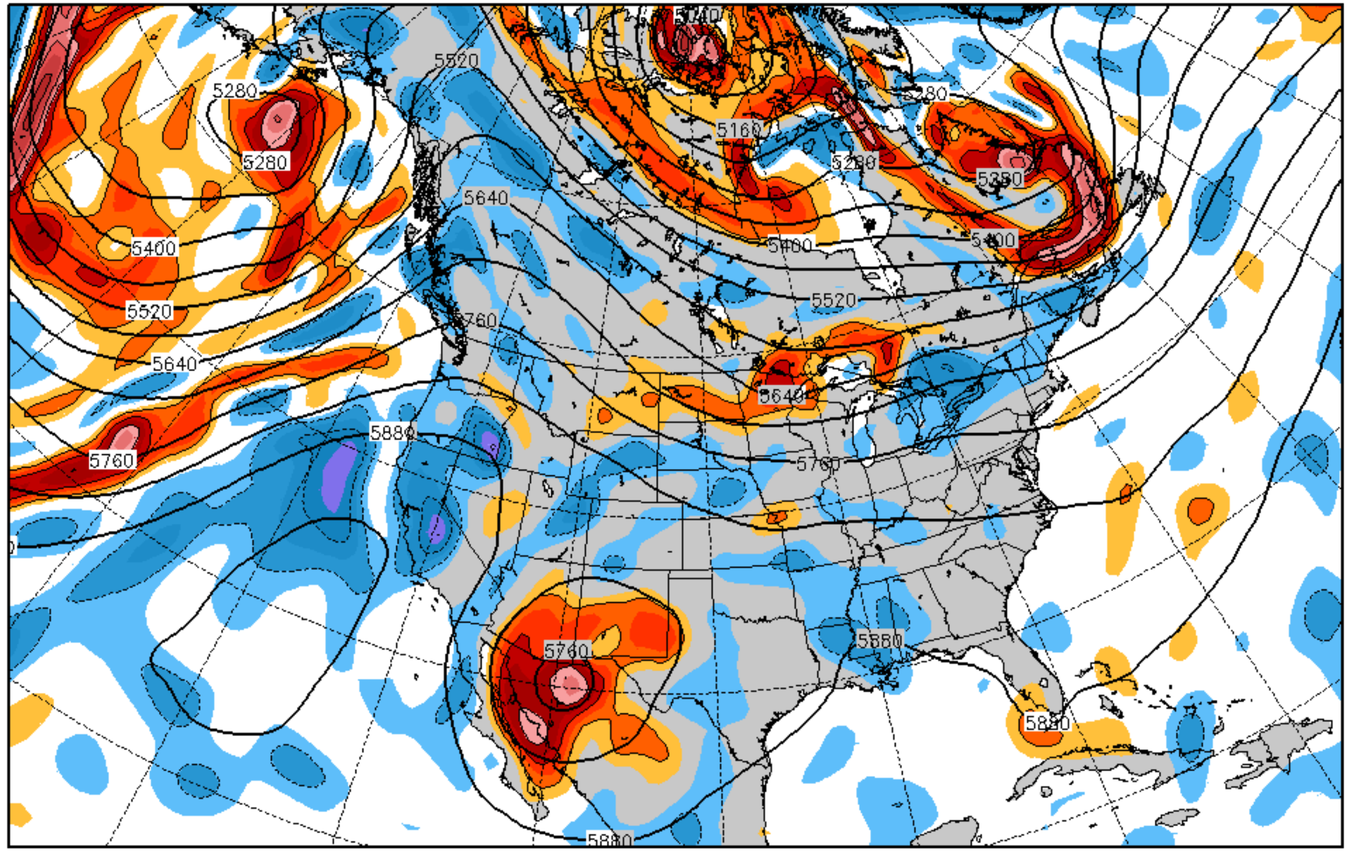
DTG : 2015100812
Run :

Tau 000 006 012 018 024 030 036 042 048 054 060 066 072 078 084 090 096 102 108 114 120 126 132 138 144 150 156 162 168 174 180

Other Products	
10m Winds	loop
2m Temp	loop
300 Hgts/Winds	loop
500 Hgts/Vort	loop
700 Hgts/Vert Vel	loop
850 Temp/winds/RH	loop
1000-500 Thick/SLP	loop
Precip/SLP	loop
Sig Wave/Wind	loop
30-38K Clear Air Turb	loop
22-30K Clear Air Turb	loop
Ditch Heading	loop
Altimeter	loop
Tropopause Press	loop
K-Index	loop

Other Models
NAVGEM N Am
GFS N Am
NAVGEM E Pac
COAMPS NE Pac

UNCLASSIFIED



UNCLASSIFIED

Thu 08 Oct 2015 00UTC @ECMWF Forecast T+0 VT: Thu 08 Oct 2015 00UTC
Mean sea level pressure (MSLP) Ensemble Mean and Normalized Standard Deviation (shaded)

Thu 08 Oct 2015 00UTC @ECMWF Forecast T+0 VT

www.ecmwf.int/en/forecasts/charts/medium/ensemble-mean-and-spread-four-standard-parameters?step=0&relative_archive_date=20151008...

Ensemble me

On this page you
parameters: mean
geopotential heig

VT:

850 mb Heigh

12-hour foreca

<< Previou
Hour
Start
FCST: F08

NAM 24 hour valid 12Z FRI 9 OCT 15

©UNISYS

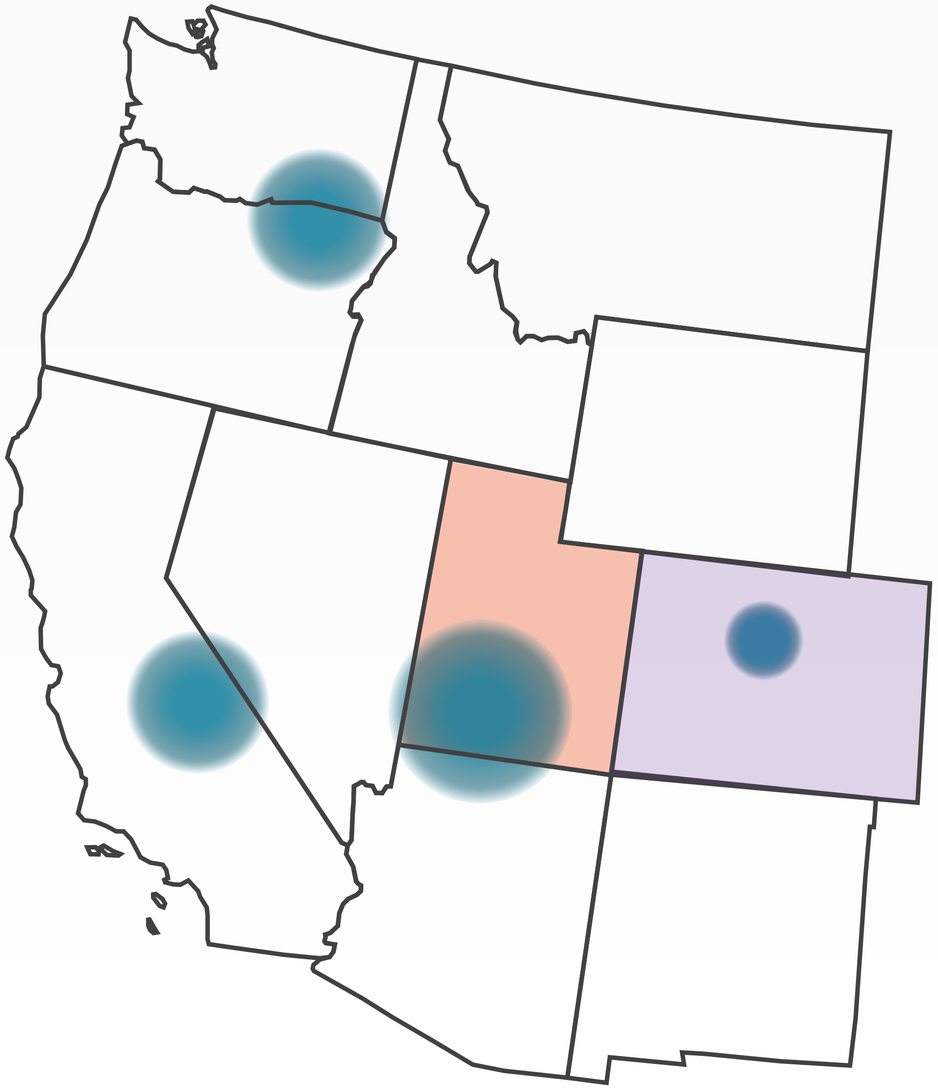
SAT 10 OCT 15

©UNISYS


AGGREGATE
MENTAL MODEL

> *Thus, simply showing a complex visualization, expecting a user to extract the necessary information, and to be finished is an oversimplification of how complex visualizations are used.*

- Trafton and Hoffman, 2007



WILDFIRE POTENTIAL

very uncertain

somewhat uncertain

fairly certain

Q. Should Utah lend firefighting teams to Colorado?

The complexities of real-world decisions are difficult (if not impossible) to model in a controlled lab study.

THE CURRENT STATE OF VISUALIZATION EVALUATION

USER STUDIES

USER STUDIES

CASE STUDIES

USER STUDIES **CASE STUDIES**

control ⟷ realism

USER STUDIES

CASE STUDIES

control ⟷ realism

tell us **why** an effect exists,
but relevance is questionable

USER STUDIES

CASE STUDIES

control ⟷ realism

tell us **why** an effect exists,
but relevance is questionable

demonstrate **that** a relevant effect
exists, but why remains unclear

DESIGNING FOR ECOLOGICAL VALIDITY

PILOT STUDY Summer 2015

GOAL: to evaluate the effect of uncertainty visualizations on forecasting

PILOT STUDY Summer 2015

GOAL: to evaluate the effect of uncertainty visualizations on forecasting

- five week, longitudinal study

PILOT STUDY Summer 2015

GOAL: to evaluate the effect of uncertainty visualizations on forecasting

- five week, longitudinal study
- five **quasi-expert** student forecasters

PILOT STUDY Summer 2015

GOAL: to evaluate the effect of uncertainty visualizations on forecasting

- five week, longitudinal study
- five **quasi-expert** student forecasters
- forecasted daily high temperatures at multiple locations

PILOT STUDY Summer 2015

GOAL: to evaluate the effect of uncertainty visualizations on forecasting

- five week, longitudinal study
- five **quasi-expert** student forecasters
- forecasted daily high temperatures at multiple locations
- added uncertainty visualizations, three of five weeks

PILOT STUDY Summer 2015

GOAL: to evaluate the effect of uncertainty visualizations on forecasting

- five week, longitudinal study
- five **quasi-expert** student forecasters
- forecasted daily high temperatures at multiple locations
- added uncertainty visualizations, three of five weeks
- treated forecasting process like a black-box

PILOT STUDY Summer 2015

GOAL: to evaluate the effect of uncertainty visualizations on forecasting

- five week, longitudinal study
- five **quasi-expert** student forecasters
- forecasted daily high temperatures at multiple locations
- added uncertainty visualizations, three of five weeks
- treated forecasting process like a black-box

WEEK 1	baseline
WEEK 2	
WEEK 3	
WEEK 4	
WEEK 5	baseline

PILOT STUDY Summer 2015

GOAL: to evaluate the effect of uncertainty visualizations on forecasting

- five week, longitudinal study
- five **quasi-expert** student forecasters
- forecasted daily high temperatures at multiple locations
- added uncertainty visualizations, three of five weeks
- treated forecasting process like a black-box

WEEK 1	baseline
WEEK 2	plume diagrams
WEEK 3	
WEEK 4	
WEEK 5	baseline

PILOT STUDY Summer 2015

GOAL: to evaluate the effect of uncertainty visualizations on forecasting

- five week, longitudinal study
- five **quasi-expert** student forecasters
- forecasted daily high temperatures at multiple locations
- added uncertainty visualizations, three of five weeks
- treated forecasting process like a black-box

WEEK 1	baseline
WEEK 2	plume diagrams
WEEK 3	mean + std dev
WEEK 4	
WEEK 5	baseline

PILOT STUDY Summer 2015

GOAL: to evaluate the effect of uncertainty visualizations on forecasting

- five week, longitudinal study
- five **quasi-expert** student forecasters
- forecasted daily high temperatures at multiple locations
- added uncertainty visualizations, three of five weeks
- treated forecasting process like a black-box

WEEK 1	baseline
WEEK 2	plume diagrams
WEEK 3	mean + std dev
WEEK 4	spaghetti plots
WEEK 5	baseline

PILOT STUDY Summer 2015

GOAL: to evaluate the effect of uncertainty visualizations on forecasting

- five week, longitudinal study
- five **quasi-expert** student forecasters
- forecasted daily high temperatures at multiple locations
- added uncertainty visualizations, three of five weeks
- treated forecasting process like a black-box
- accuracy with vs. without additional uncertainty visualizations

WEEK 1	baseline
WEEK 2	plume diagrams
WEEK 3	mean + std dev
WEEK 4	spaghetti plots
WEEK 5	baseline

DESIGN CHOICE 1 *Working with Quasi-Experts*

DESIGN CHOICE 1 *Working with Quasi-Experts*

- senior, self-selected members of undergraduate forecasting club

DESIGN CHOICE 1 *Working with Quasi-Experts*

- senior, self-selected members of undergraduate forecasting club
- publicly release daily 5-day forecasts, which are actually used

DESIGN CHOICE 1 *Working with Quasi-Experts*

- senior, self-selected members of undergraduate forecasting club
- publicly release daily 5-day forecasts, which are actually used
- not yet professionals...

DESIGN CHOICE 1 *Working with Quasi-Experts*

- senior, self-selected members of undergraduate forecasting club
- publicly release daily 5-day forecasts, which are actually used
- not yet professionals...
  ....but also not professionals

DESIGN CHOICE 2 *Process as a Black-Box*

DESIGN CHOICE 2 *Process as a Black-Box*

- no single model for how meteorologists forecast the weather

DESIGN CHOICE 2 *Process as a Black-Box*

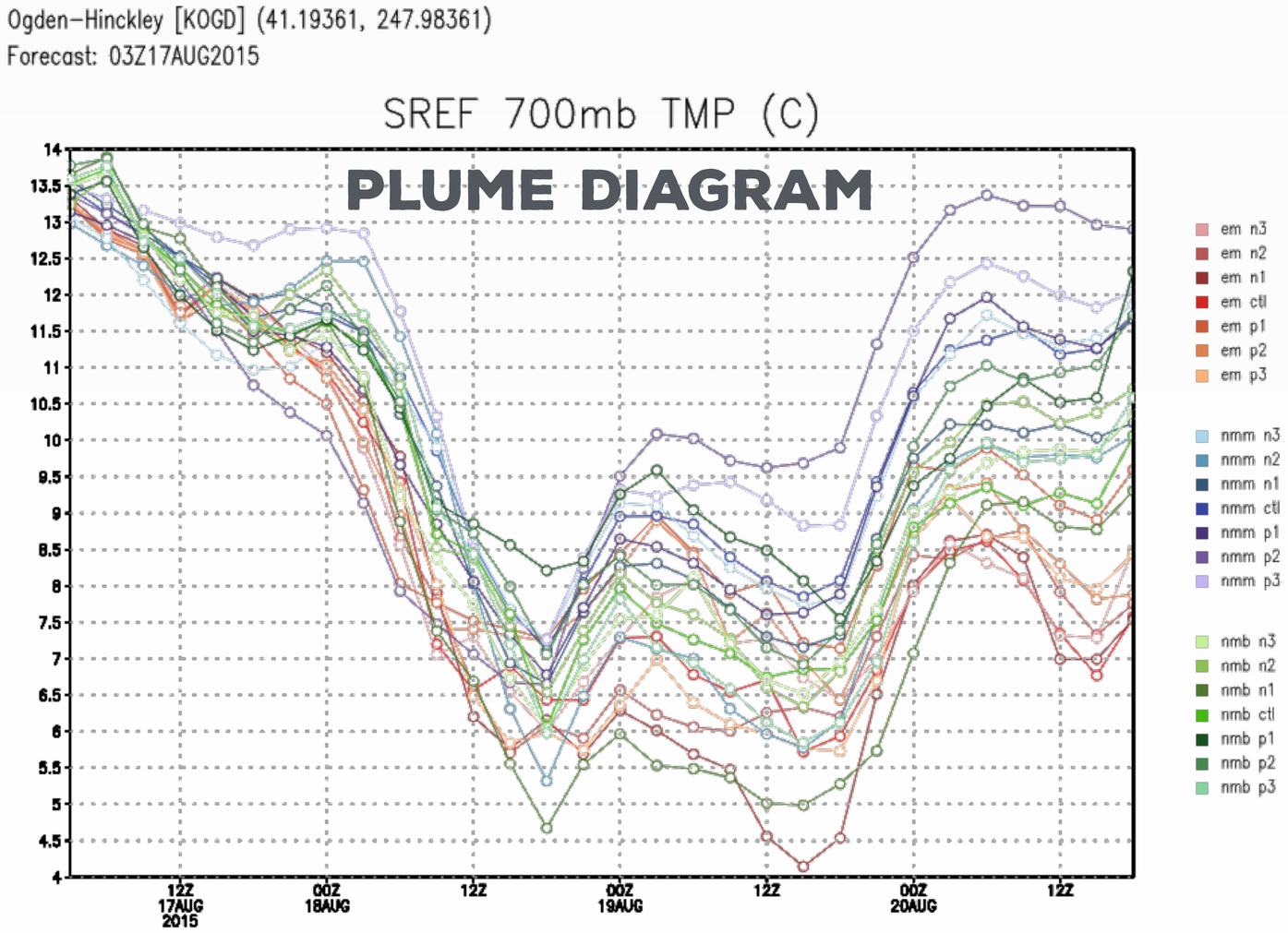
- no single model for how meteorologists forecast the weather
- experimental control changes the decision-making context

DESIGN CHOICE 2 *Process as a Black-Box*

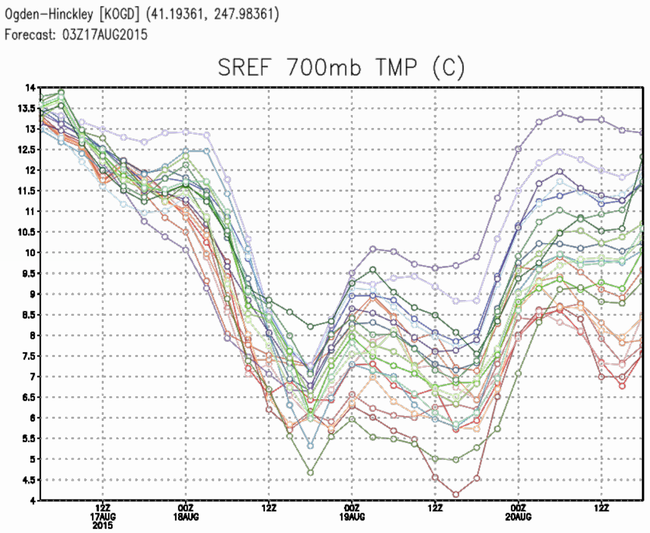
- no single model for how meteorologists forecast the weather
- experimental control changes the decision-making context
- reflects how a single visualization affects the forecasting process

DESIGN CHOICE 3 *The Uncertainty Visualizations*

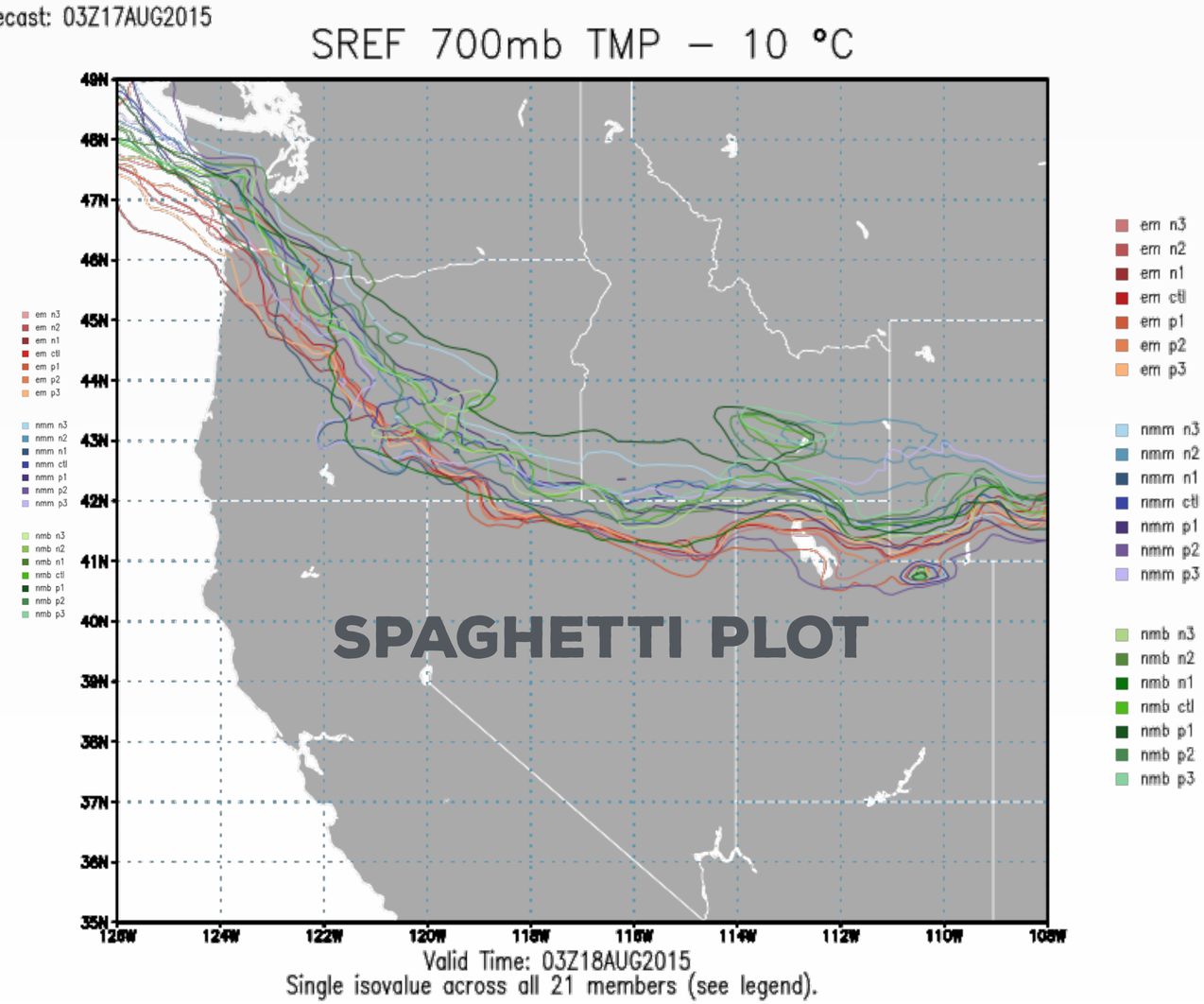
DESIGN CHOICE 3 *The Uncertainty Visualizations*



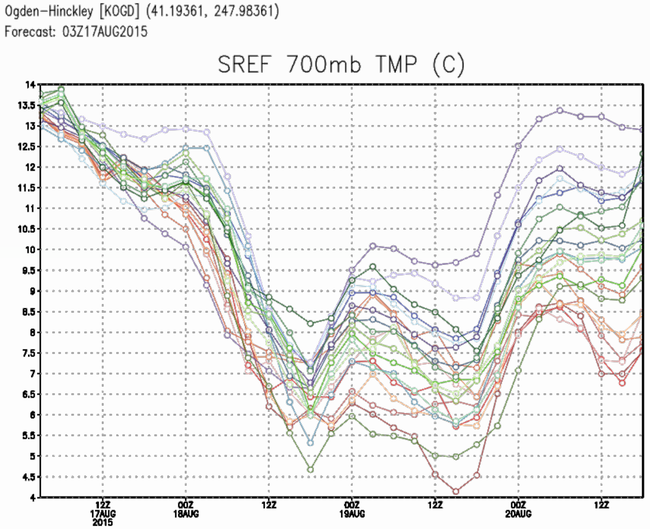
DESIGN CHOICE 3 *The Uncertainty Visualizations*



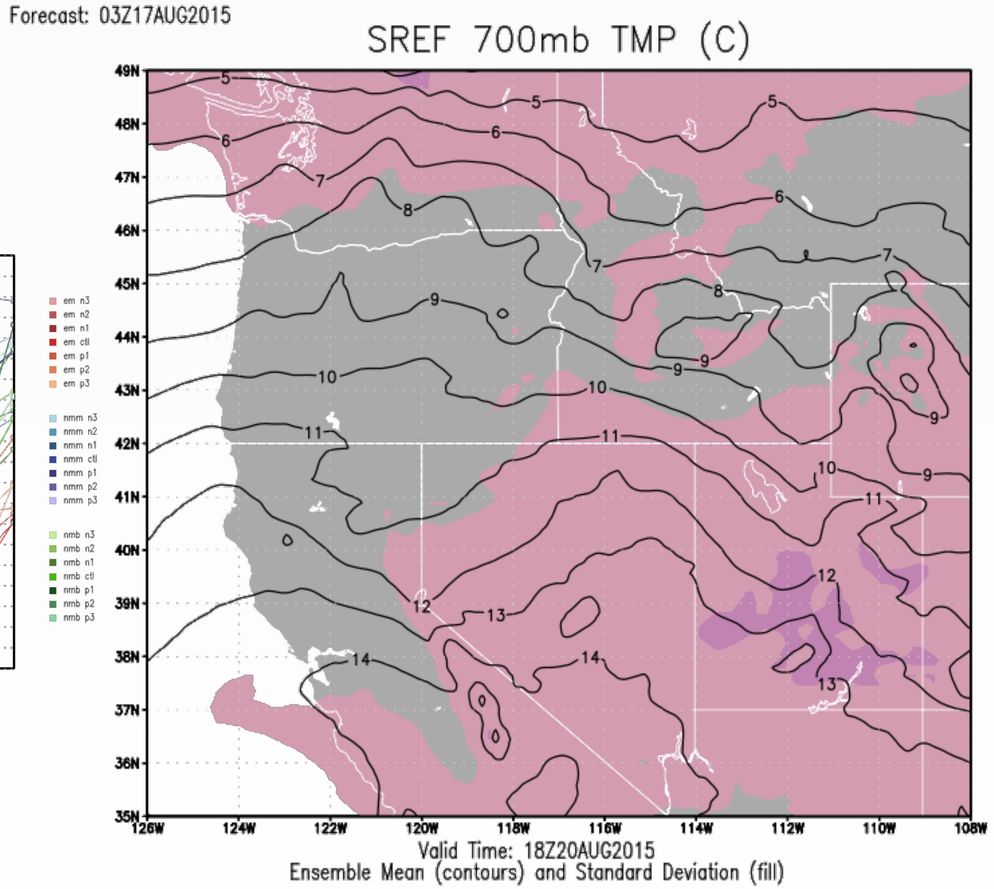
PLUME DIAGRAM



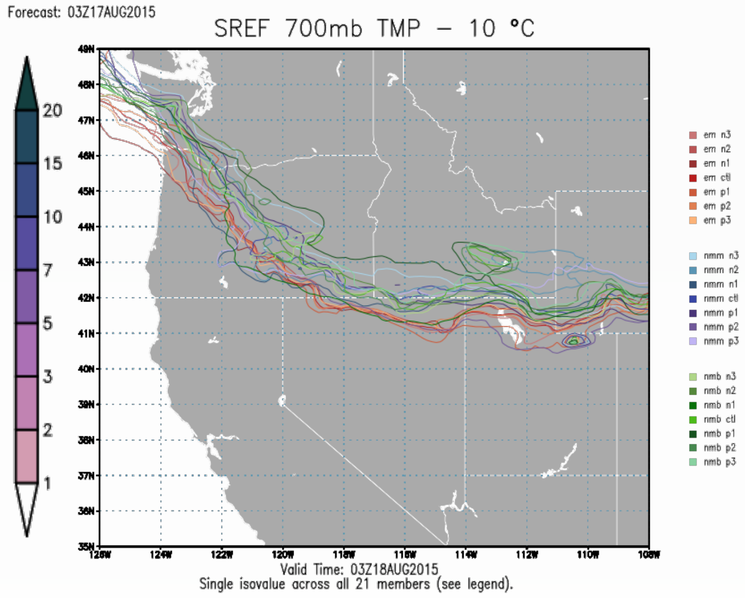
DESIGN CHOICE 3 *The Uncertainty Visualizations*



PLUME DIAGRAM

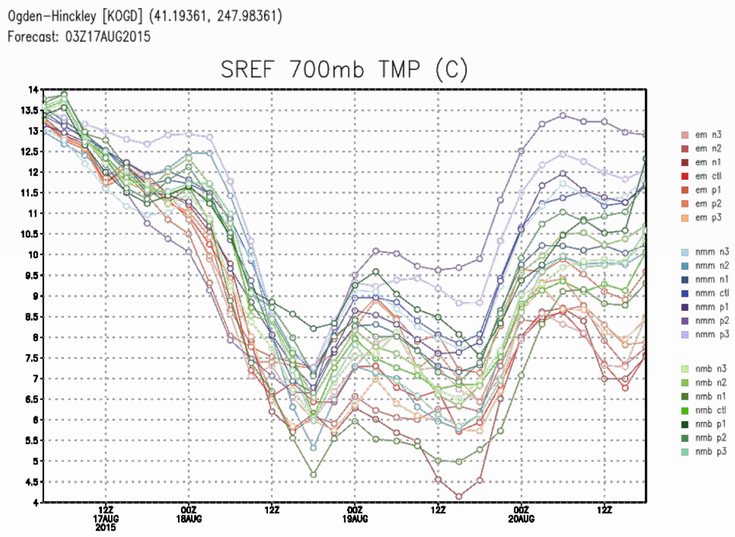


MEAN + STD DEV

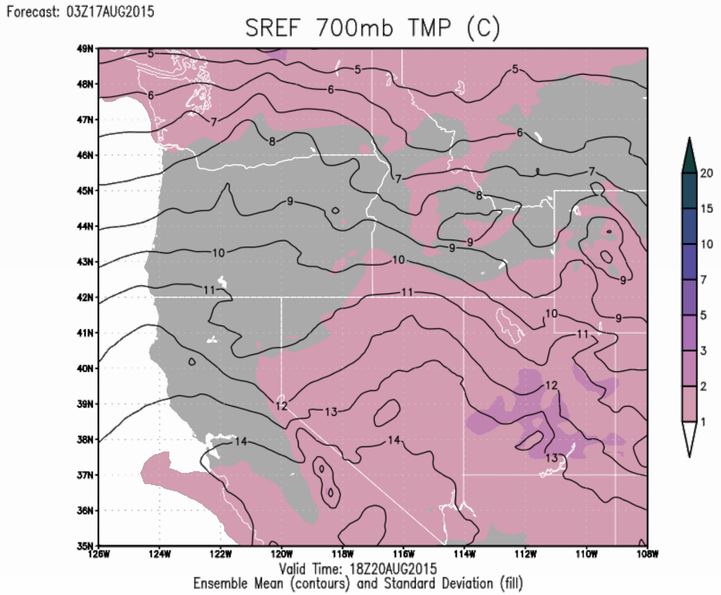


SPAGHETTI PLOT

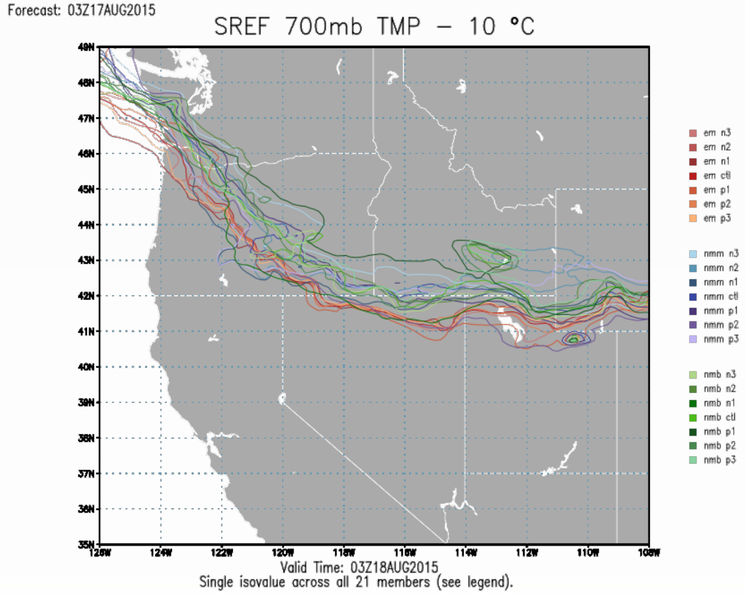
DESIGN CHOICE 3 *The Uncertainty Visualizations*



PLUME DIAGRAM

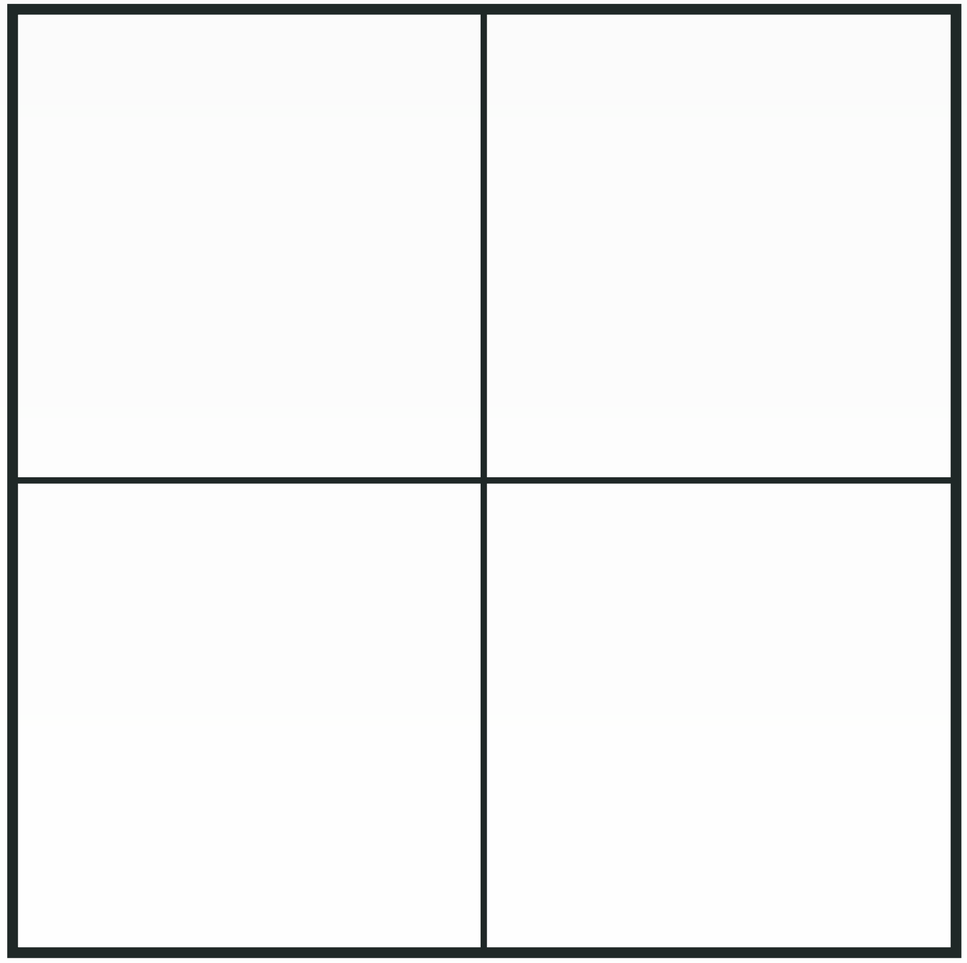


MEAN + STD DEV

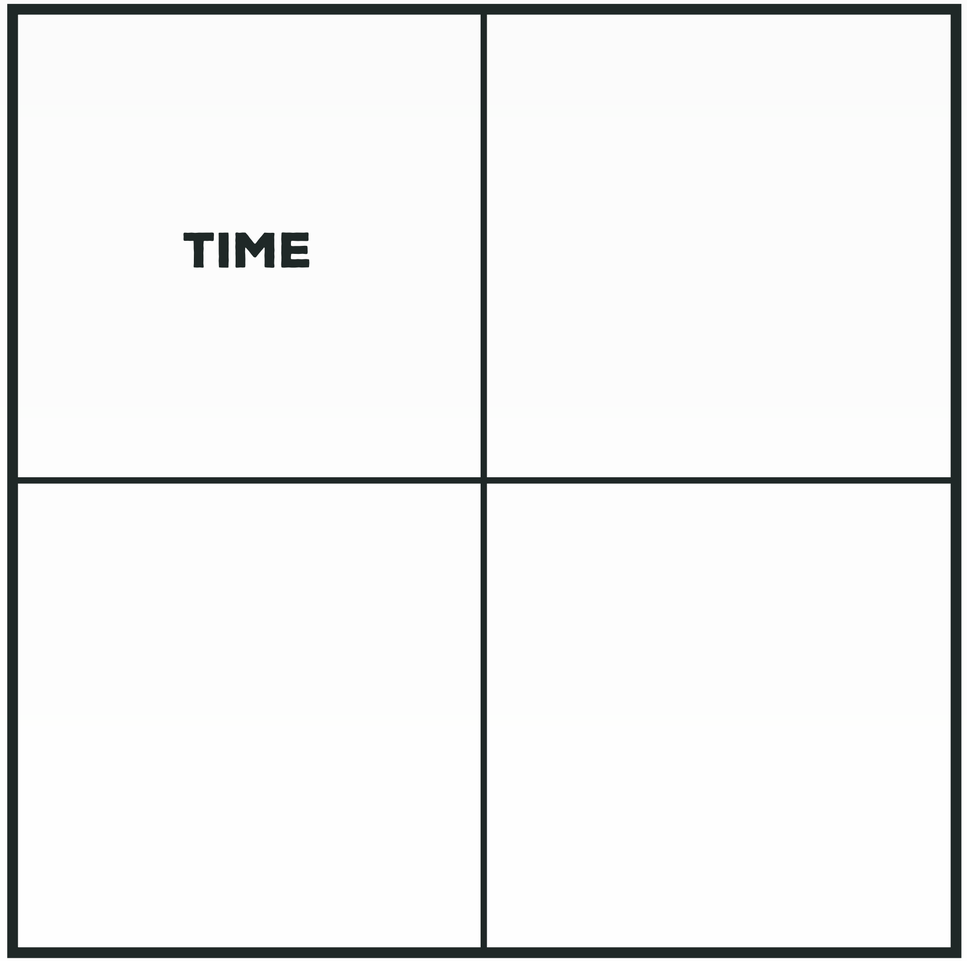


SPAGHETTI PLOT

DESIGN CHOICE 3 *Which Uncertainty Products?*



DESIGN CHOICE 3 *Which Uncertainty Products?*



DESIGN CHOICE 3 *Which Uncertainty Products?*

TIME	SPACE

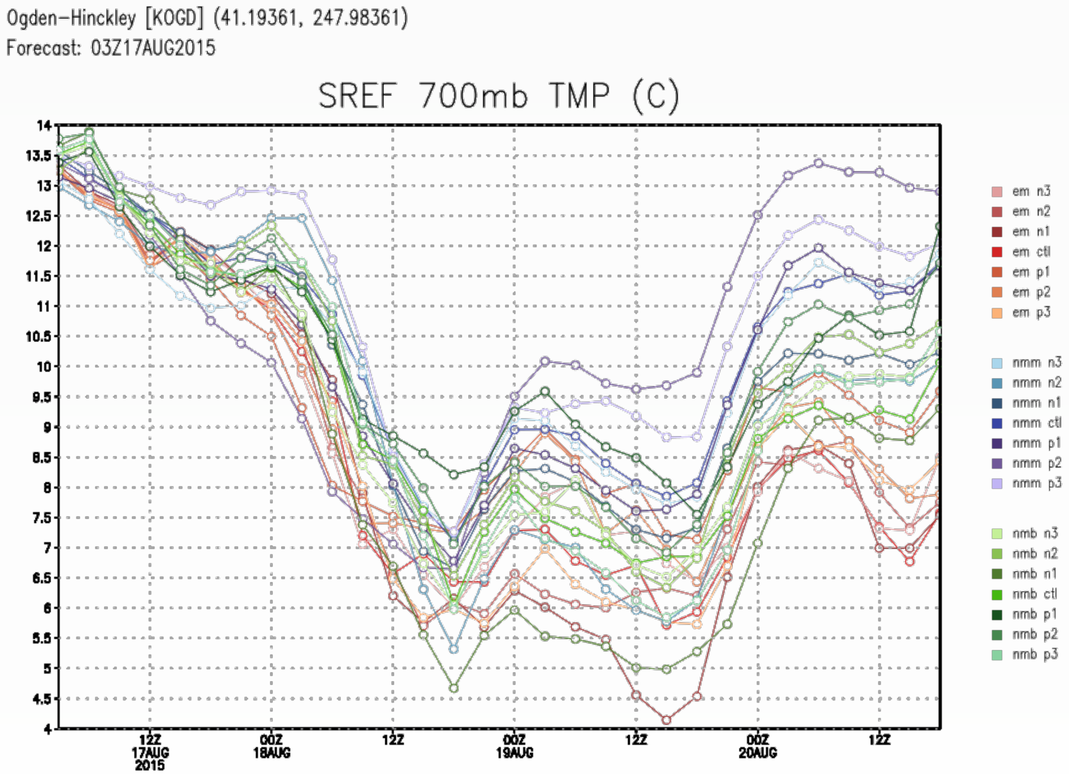
DESIGN CHOICE 3 *Which Uncertainty Products?*

TIME	**SPACE**
ENSEMBLE	

DESIGN CHOICE 3 *Which Uncertainty Products?*

TIME	**SPACE**
ENSEMBLE	**CONTEXT**

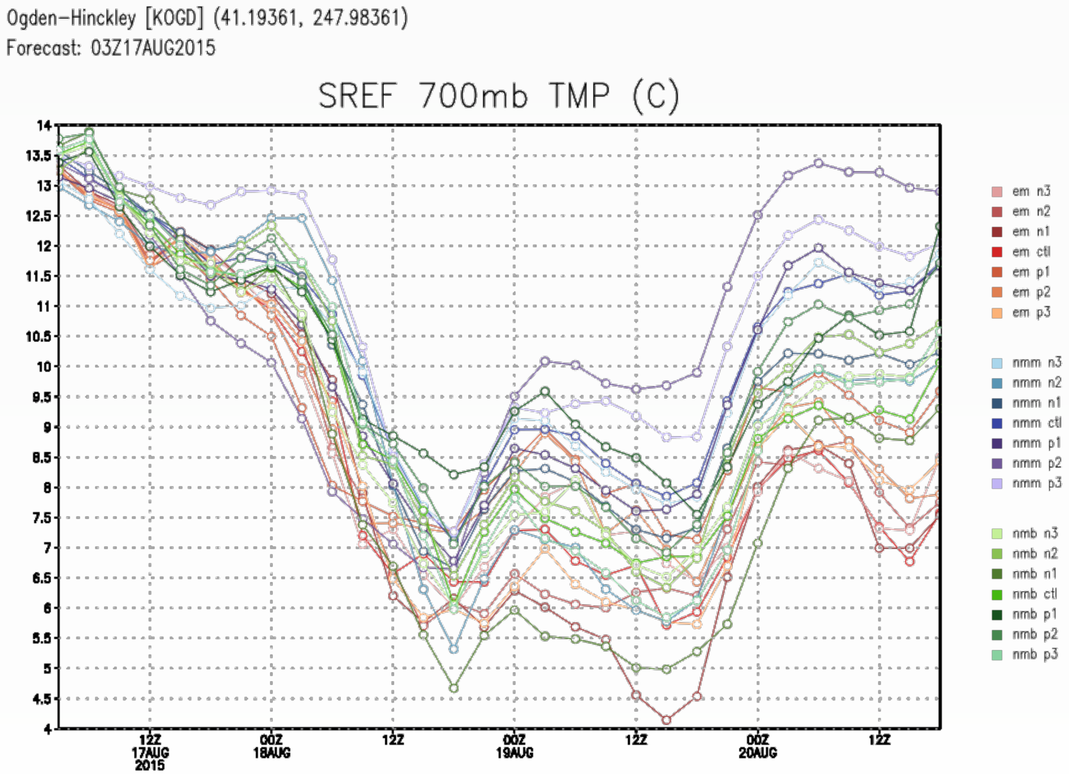
DESIGN CHOICE 3 *Which Uncertainty Products?*



PLUME DIAGRAM

TIME	**SPACE**
ENSEMBLE	**CONTEXT**

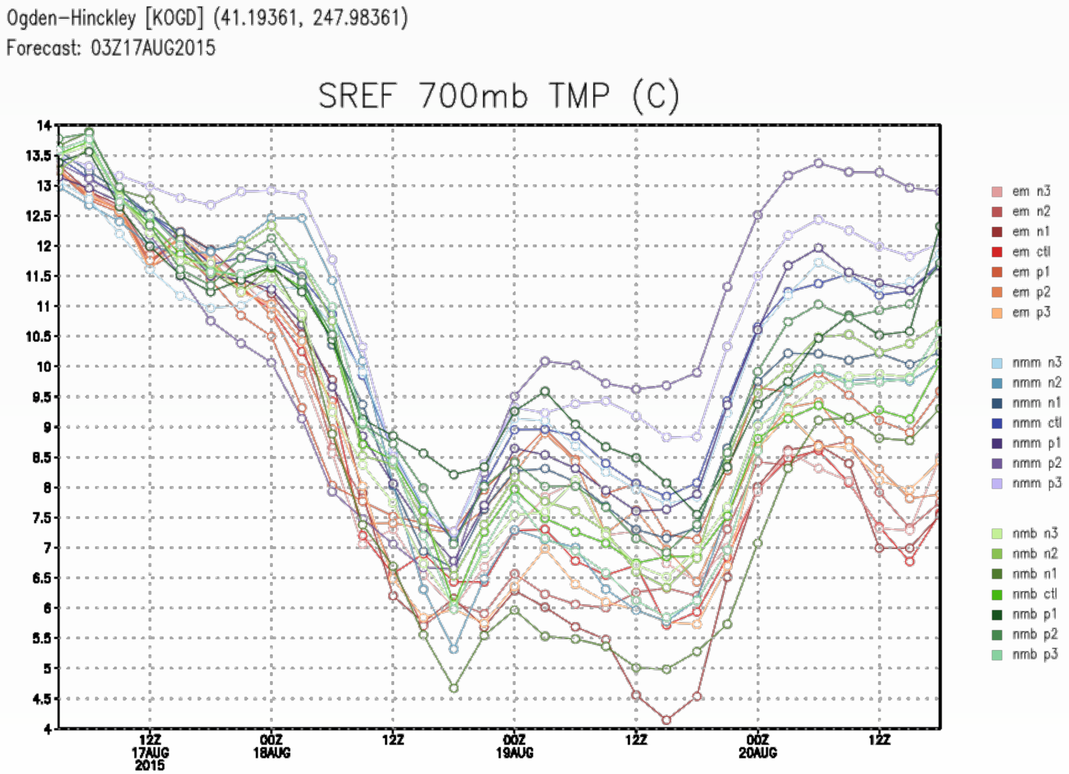
DESIGN CHOICE 3 *Which Uncertainty Products?*



PLUME DIAGRAM

TIME	**SPACE**
ENSEMBLE	**CONTEXT**

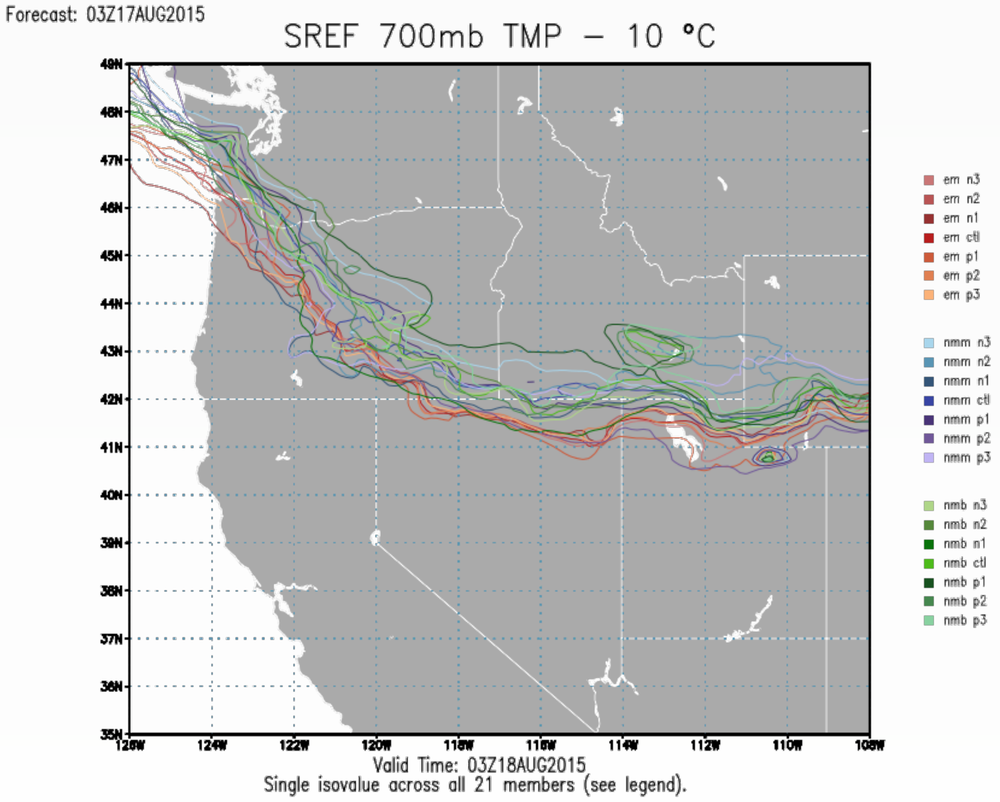
DESIGN CHOICE 3 *Which Uncertainty Products?*



PLUME DIAGRAM

TIME	**SPACE**
ENSEMBLE	**CONTEXT**

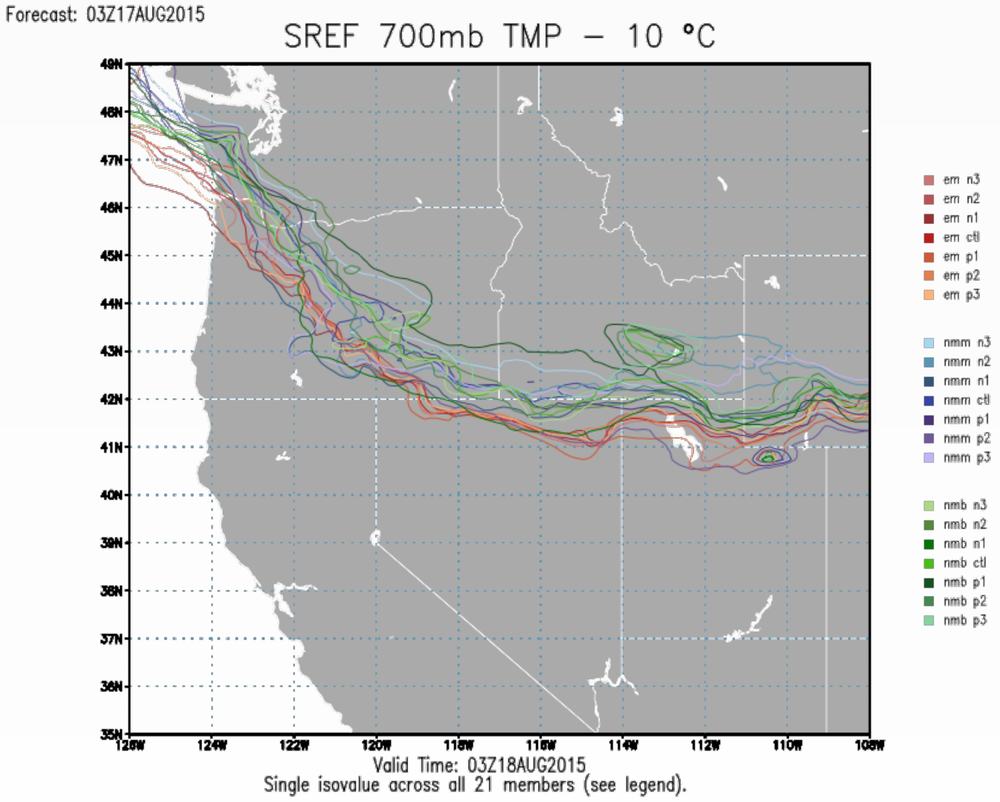
DESIGN CHOICE 3 *The Uncertainty Visualizations*



SPAGHETTI PLOT

TIME	**SPACE**
ENSEMBLE	**CONTEXT**

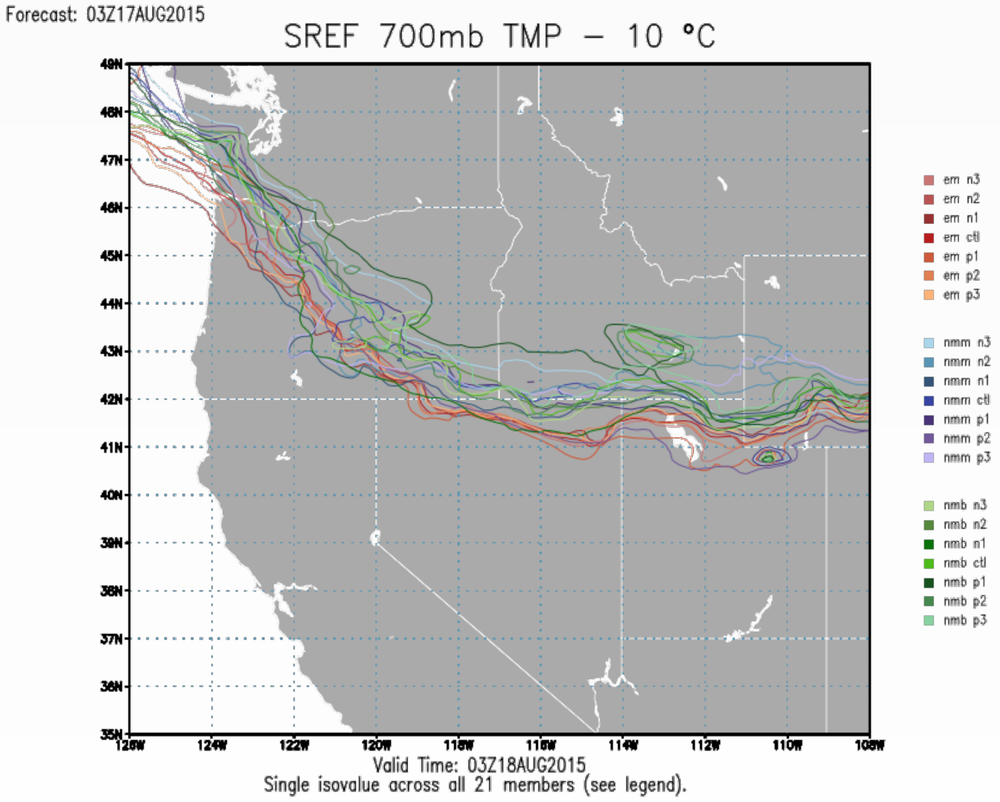
DESIGN CHOICE 3 *The Uncertainty Visualizations*



SPAGHETTI PLOT

TIME	**SPACE**
ENSEMBLE	**CONTEXT**

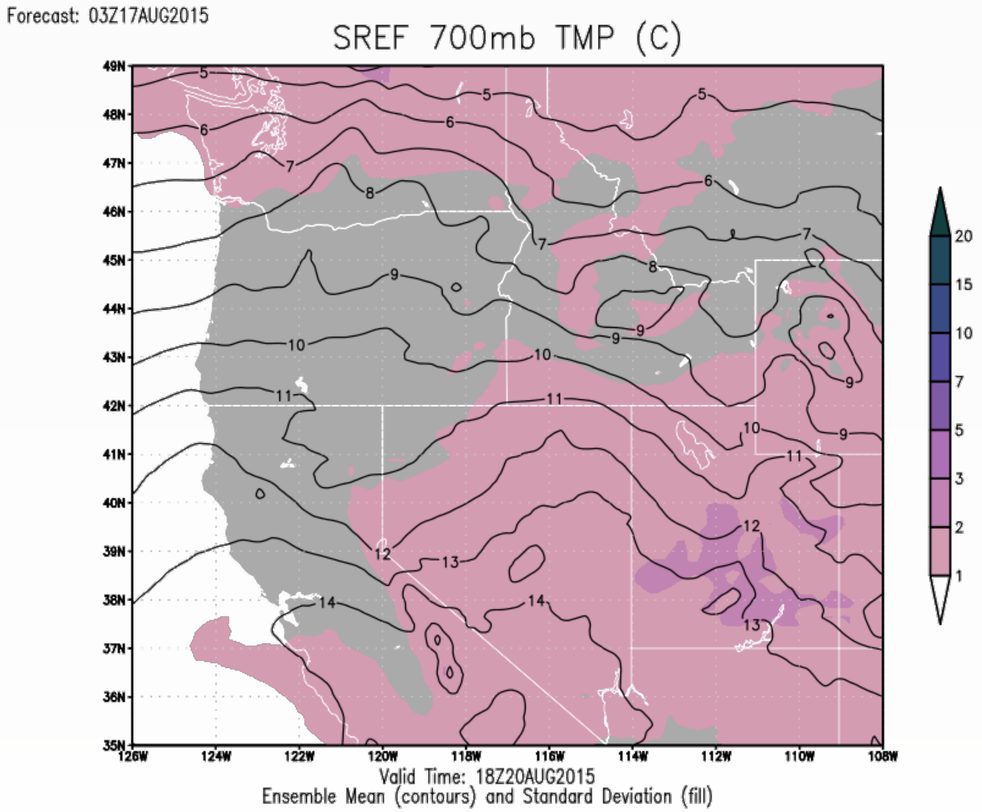
DESIGN CHOICE 3 *The Uncertainty Visualizations*



SPAGHETTI PLOT

TIME	**SPACE**
ENSEMBLE	**CONTEXT**

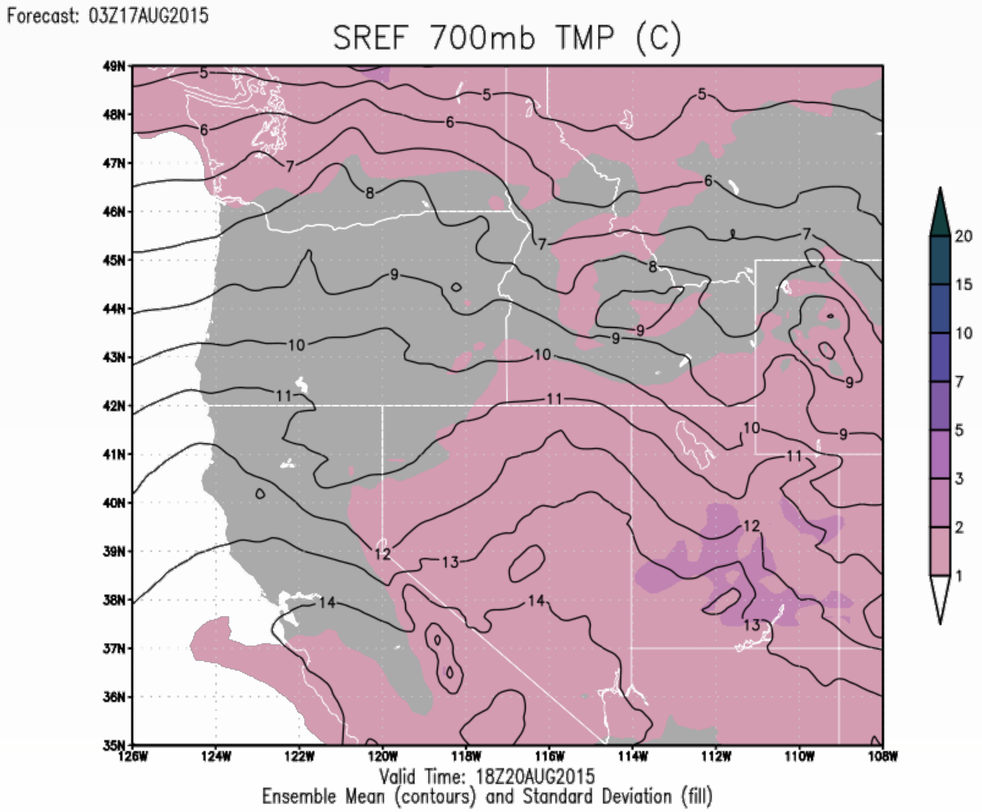
DESIGN CHOICE 3 *The Uncertainty Visualizations*



MEAN + STD DEV

TIME	**SPACE**
ENSEMBLE	**CONTEXT**

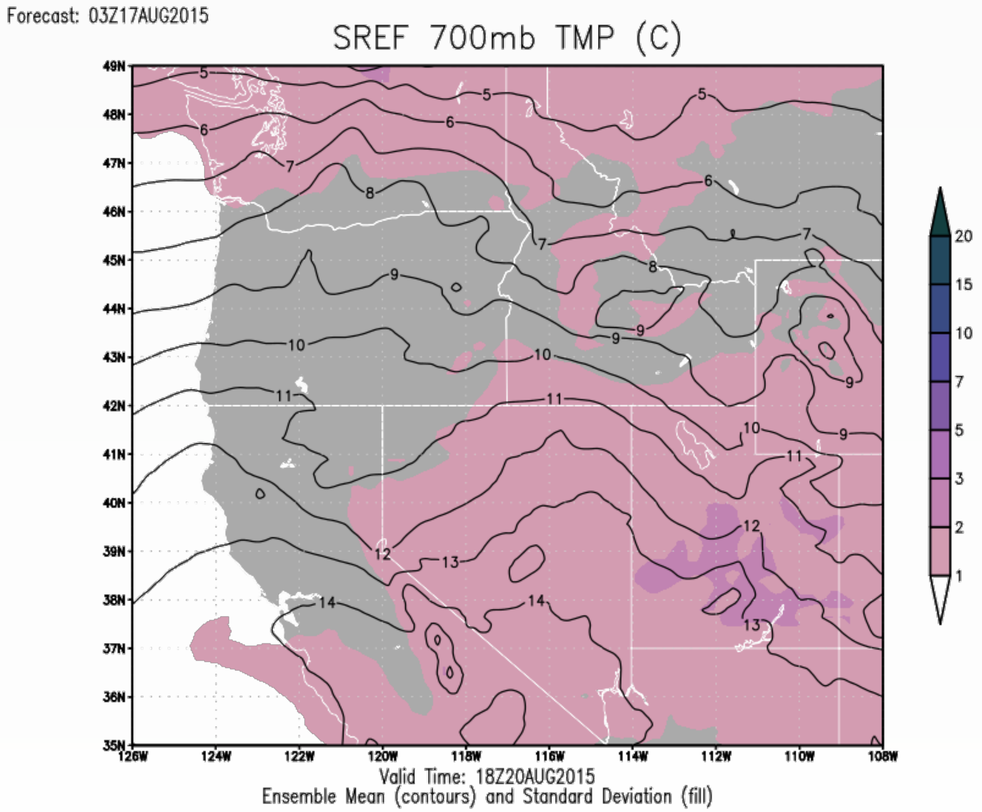
DESIGN CHOICE 3 *The Uncertainty Visualizations*



MEAN + STD DEV

TIME	**SPACE**
ENSEMBLE	**CONTEXT**

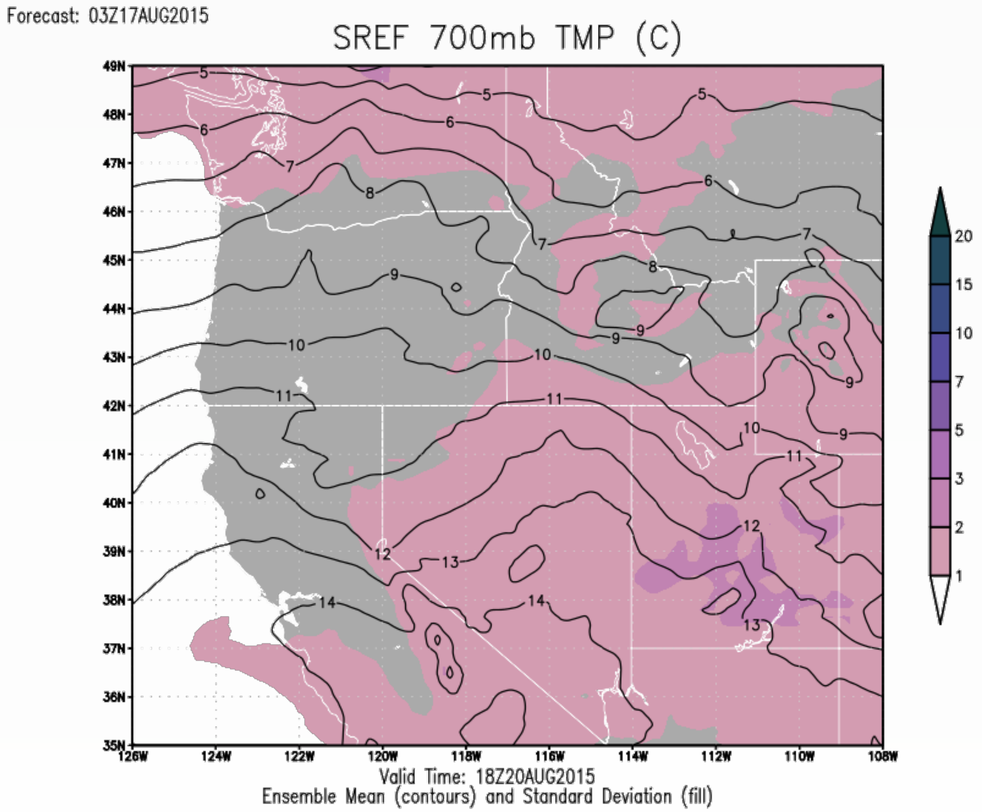
DESIGN CHOICE 3 *The Uncertainty Visualizations*



MEAN + STD DEV

TIME	**SPACE**
ENSEMBLE	**CONTEXT**

DESIGN CHOICE 3 *The Uncertainty Visualizations*

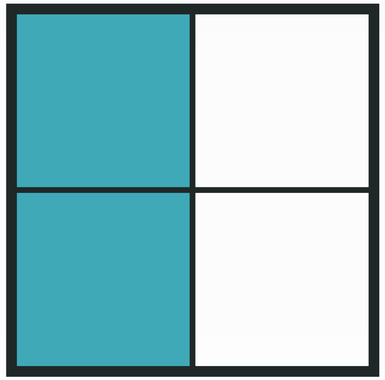


MEAN + STD DEV

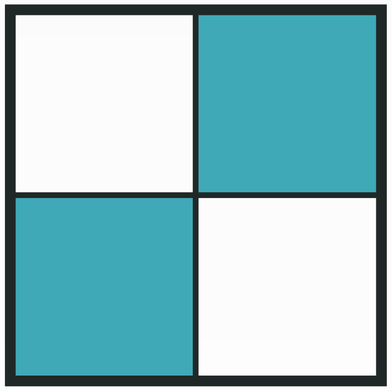
TIME	**SPACE**
ENSEMBLE	**CONTEXT**

DESIGN CHOICE 3 *The Uncertainty Visualizations*

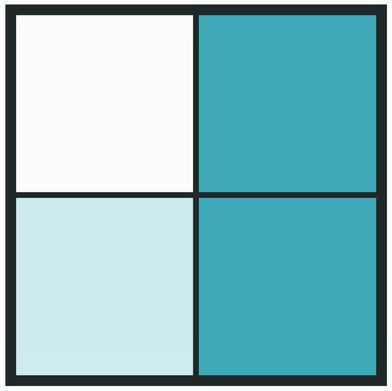
PLUME DIAGRAM



MEAN + STD DEV

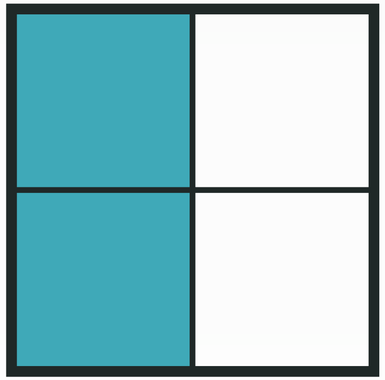


SPAGHETTI PLOT

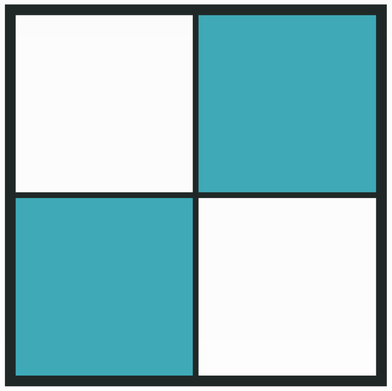


DESIGN CHOICE 3 *The Uncertainty Visualizations*

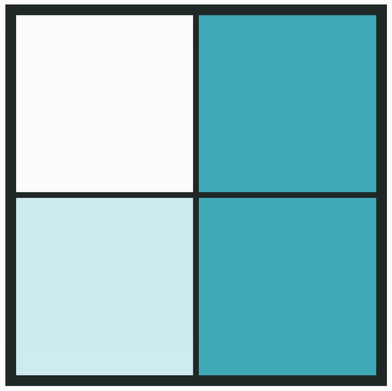
PLUME DIAGRAM



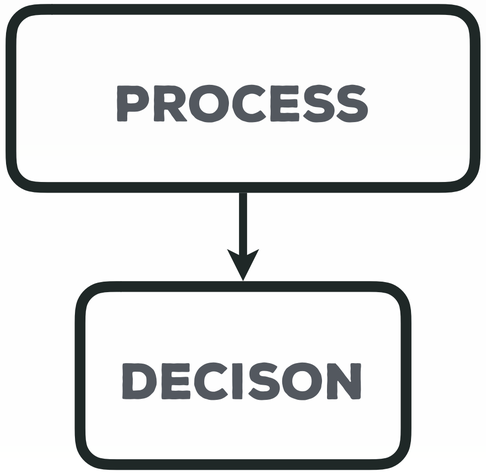
MEAN + STD DEV



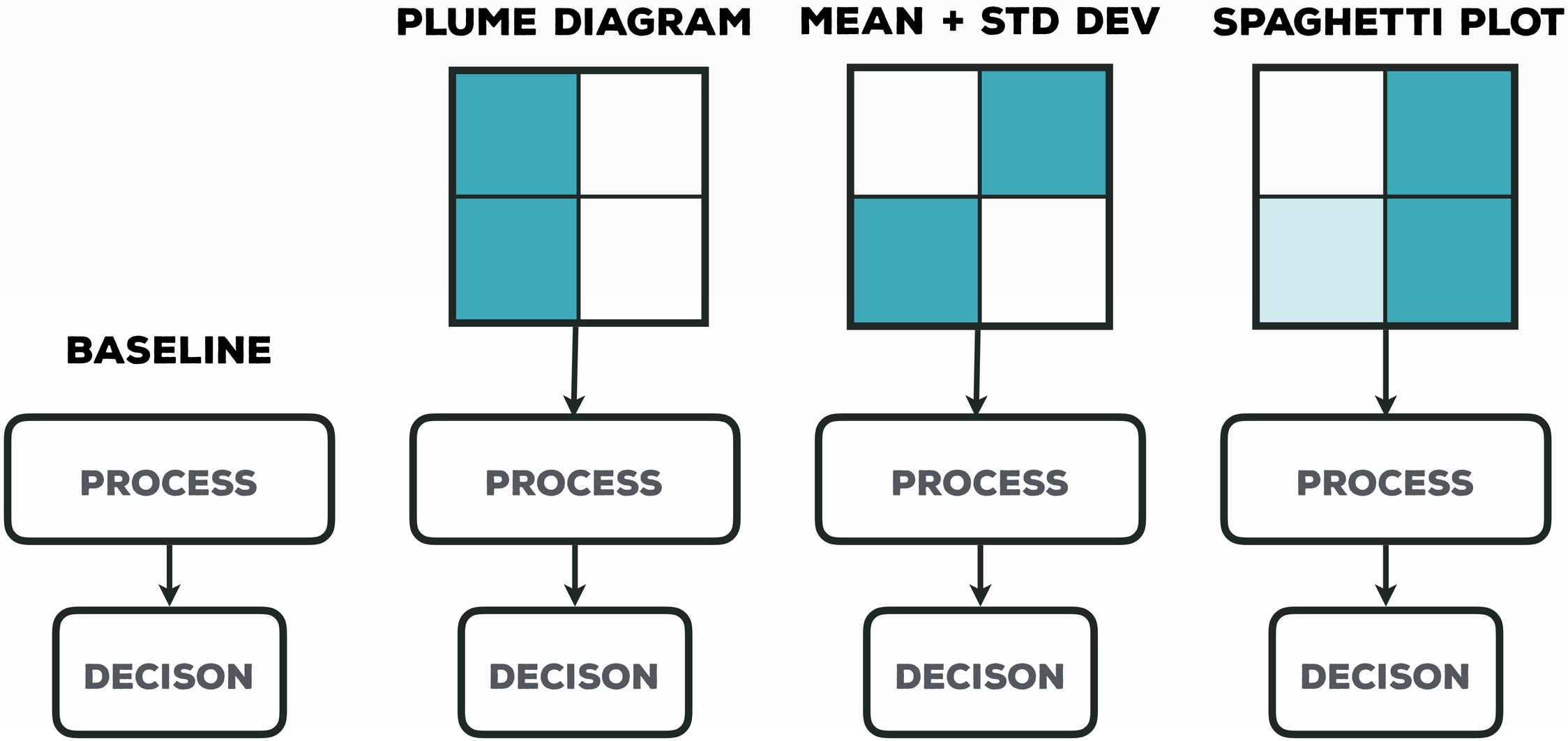
SPAGHETTI PLOT



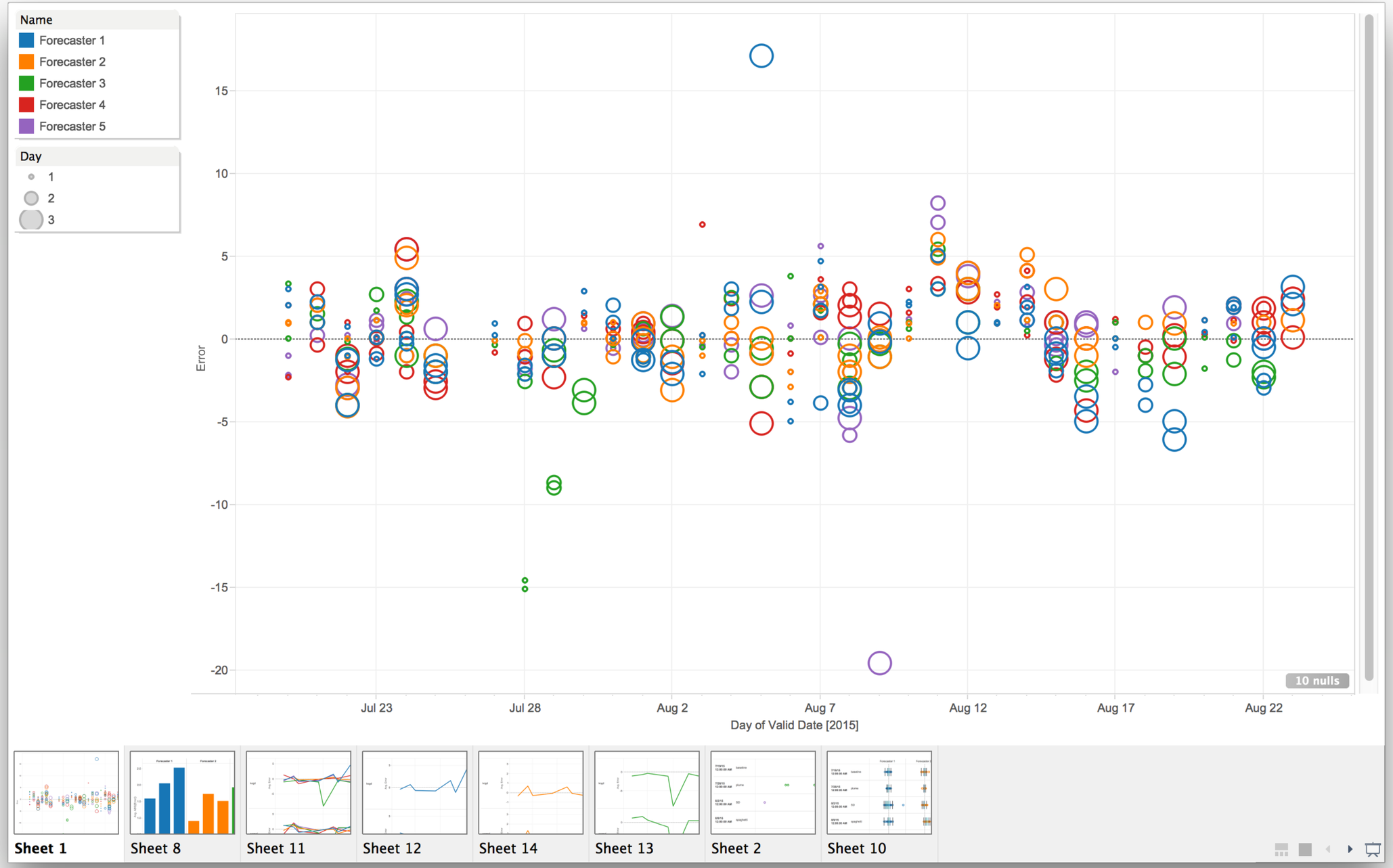
BASELINE

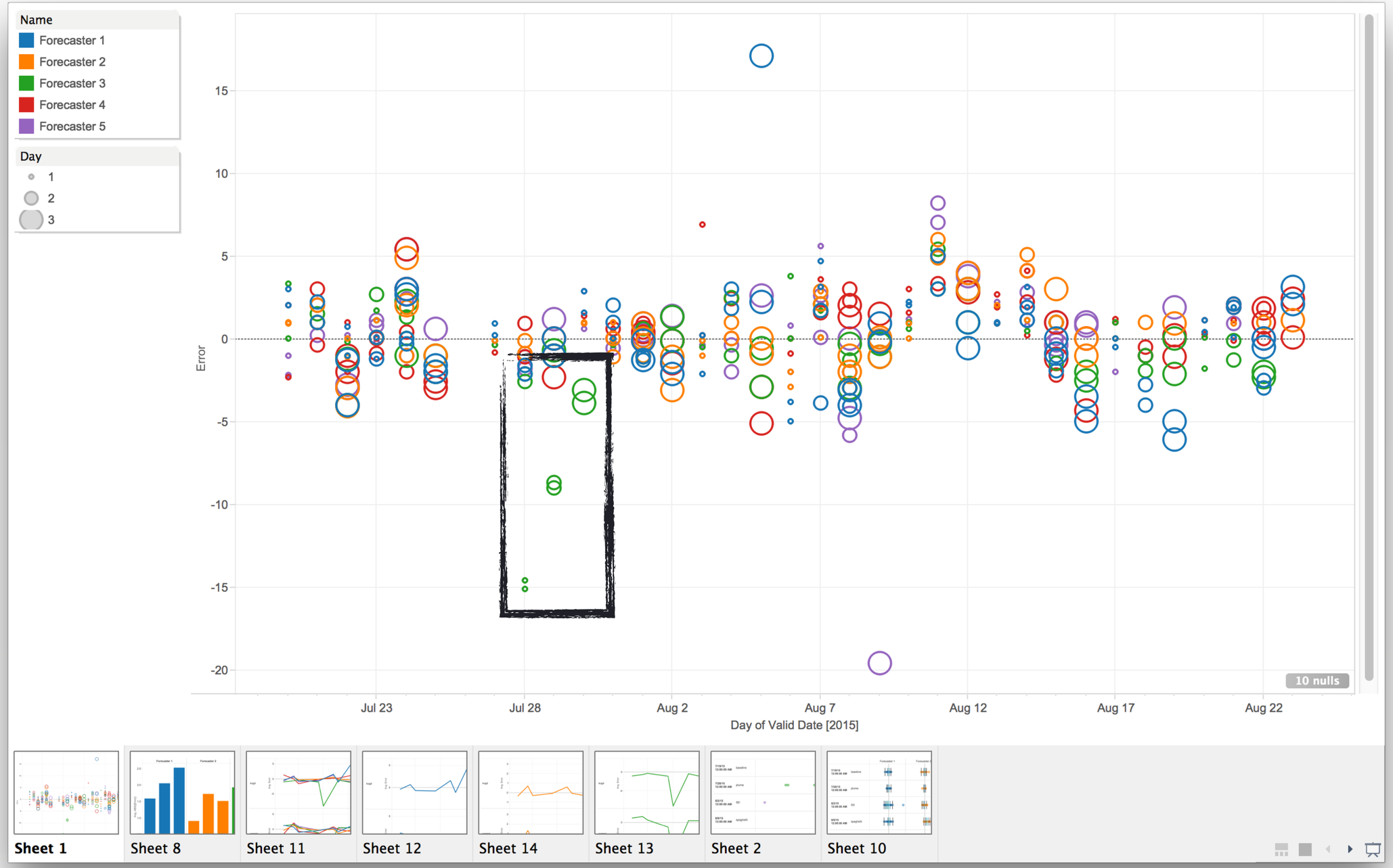


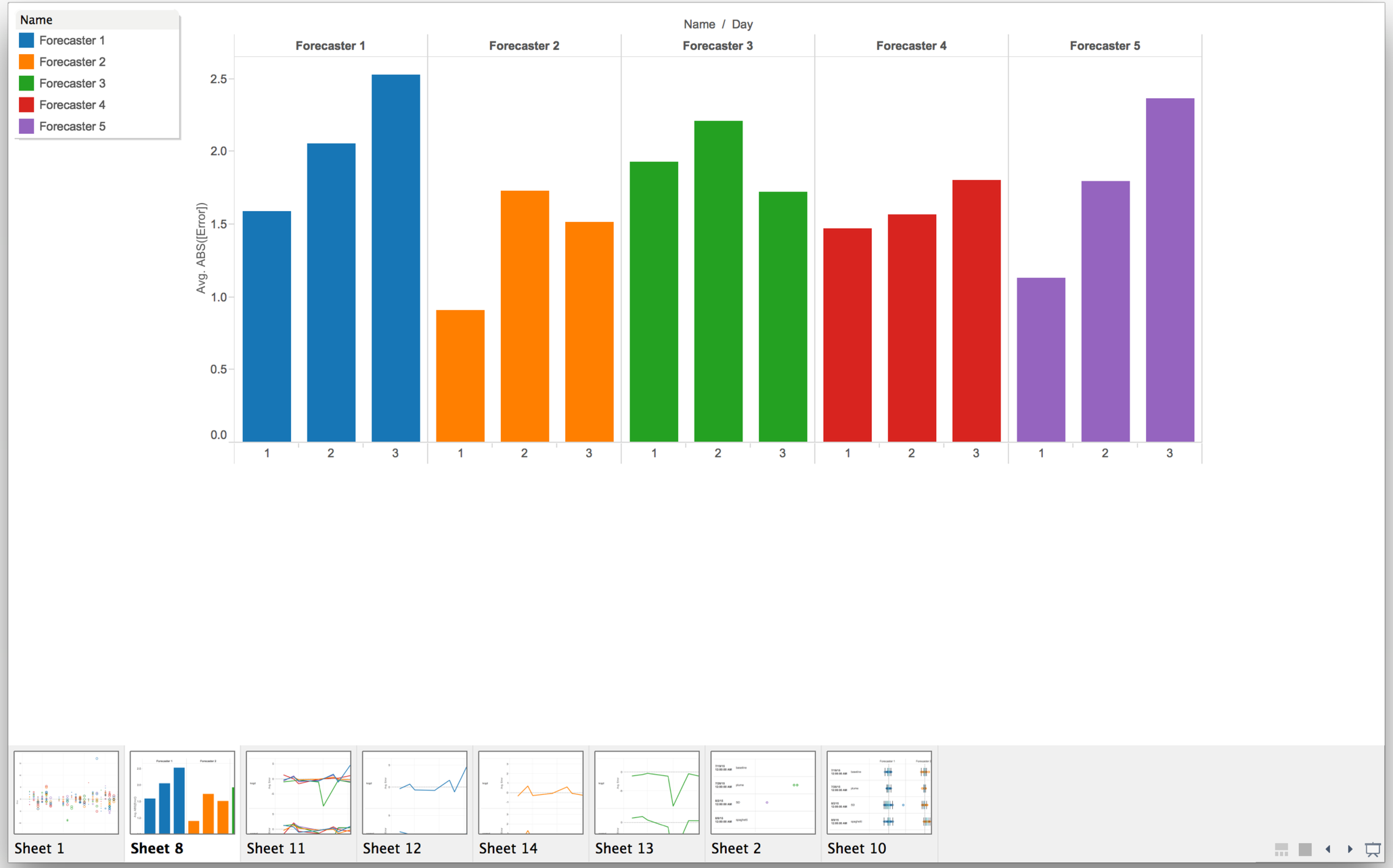
DESIGN CHOICE 3 *The Uncertainty Visualizations*



EARLY ANALYSIS & RESULTS


Name
Forecaster 1
Forecaster 2
Forecaster 3
Forecaster 4
Forecaster 5
Day
1
2
3
15
10
5
0
-5
-10
-15
-20
Error
10 nulls
Jul 23
Jul 28
Aug 2
Aug 7
Aug 12
Aug 17
Aug 22
Day of Valid Date [2015]
Sheet 1
Sheet 8
Sheet 11
Sheet 12
Sheet 14
Sheet 13
Sheet 2
Sheet 10


Name
Forecaster 1
Forecaster 2
Forecaster 3
Forecaster 4
Forecaster 5
Day
1
2
3
15
10
5
0
-5
-10
-15
-20
Error
10 nulls
Jul 23
Jul 28
Aug 2
Aug 7
Aug 12
Aug 17
Aug 22
Day of Valid Date [2015]
Sheet 1
Sheet 8
Sheet 11
Sheet 12
Sheet 14
Sheet 13
Sheet 2
Sheet 10


Name
Forecaster 1
Forecaster 2
Forecaster 3
Forecaster 4
Forecaster 5
Name / Day
Forecaster 1
Forecaster 2
Forecaster 3
Forecaster 4
Forecaster 5
Avg. ABS([Error])
2.5
2.0
1.5
1.0
0.5
0.0
1
2
3
Sheet 1
Sheet 8
Sheet 11
Sheet 12
Sheet 14
Sheet 13
Sheet 2
Sheet 10


Name
Forecaster 1
Forecaster 2
Forecaster 3
Forecaster 4
Forecaster 5
DATETRU..
Product
7/19/15 12:00:00 AM
baseline
7/26/15 12:00:00 AM
plume
8/2/15 12:00:00 AM
SD
8/9/15 12:00:00 AM
spaghetti
8/16/15 12:00:00 AM
baseline
2 nulls
-20 -18 -16 -14 -12 -10 -8 -6 -4 -2 0 2 4 6 8 10 12 14 16 18
Error
Sheet 1
Sheet 8
Sheet 11
Sheet 12
Sheet 14
Sheet 13
Sheet 2
Sheet 10


Name
Forecaster 1
Forecaster 2
Forecaster 3
Forecaster 4
Forecaster 5
DATETRU..
Product
7/19/15 12:00:00 AM
baseline
7/26/15 12:00:00 AM
plume
8/2/15 12:00:00 AM
SD
8/9/15 12:00:00 AM
spaghetti
8/16/15 12:00:00 AM
baseline
2 nulls
-20 -18 -16 -14 -12 -10 -8 -6 -4 -2 0 2 4 6 8 10 12 14 16 18
Error
Sheet 1
Sheet 8
Sheet 11
Sheet 12
Sheet 14
Sheet 13
Sheet 2
Sheet 10


Name
Forecaster 1
Forecaster 2
Forecaster 3
Forecaster 4
Forecaster 5
DATETRU..
Product
7/19/15 12:00:00 AM
baseline
7/26/15 12:00:00 AM
plume
8/2/15 12:00:00 AM
SD
8/9/15 12:00:00 AM
spaghetti
8/16/15 12:00:00 AM
baseline
2 nulls
-20 -18 -16 -14 -12 -10 -8 -6 -4 -2 0 2 4 6 8 10 12 14 16 18
Error
Sheet 1
Sheet 8
Sheet 11
Sheet 12
Sheet 14
Sheet 13
Sheet 2
Sheet 10


Name
Forecaster 1
Forecaster 2
Forecaster 3
Forecaster 4
Forecaster 5
Name
DATETRU..
Product
Forecaster 1
Forecaster 2
Forecaster 3
Forecaster 4
Forecaster 5
7/19/15 12:00:00 AM
baseline
7/26/15 12:00:00 AM
plume
8/2/15 12:00:00 AM
SD
8/9/15 12:00:00 AM
spaghetti
8/16/15 12:00:00 AM
baseline
2 nulls
-20
-10
0
10
Error
Sheet 1
Sheet 8
Sheet 11
Sheet 12
Sheet 14
Sheet 13
Sheet 2
Sheet 10


Name
Forecaster 1
Forecaster 2
Forecaster 3
Forecaster 4
Forecaster 5
DATETRU..
Product
7/19/15 12:00:00 AM
baseline
7/26/15 12:00:00 AM
plume
8/2/15 12:00:00 AM
SD
8/9/15 12:00:00 AM
spaghetti
8/16/15 12:00:00 AM
baseline
2 nulls
-20 -18 -16 -14 -12 -10 -8 -6 -4 -2 0 2 4 6 8 10 12 14 16 18
Error
Sheet 1
Sheet 8
Sheet 11
Sheet 12
Sheet 14
Sheet 13
Sheet 2
Sheet 10


Name
Forecaster 1
Forecaster 2
Forecaster 3
Forecaster 4
Forecaster 5
DATETRU..
Product
Forecaster 1
Forecaster 2
Forecaster 3
Forecaster 4
Forecaster 5
7/19/15 12:00:00 AM
baseline
7/26/15 12:00:00 AM
plume
8/2/15 12:00:00 AM
SD
8/9/15 12:00:00 AM
spaghetti
8/16/15 12:00:00 AM
baseline
2 nulls
-20 -10 0 10
Error
Sheet 1
Sheet 8
Sheet 11
Sheet 12
Sheet 14
Sheet 13
Sheet 2
Sheet 10

- plume diagrams had statistically significant improvement over baseline

- plume diagrams had statistically significant improvement over baseline
- are there more appropriate analysis methods?

- plume diagrams had statistically significant improvement over baseline
- are there more appropriate analysis methods?
- to what extent is this effect task dependent?

DISCUSSION POINTS

Are our results convincing? Are any existing results?

Are our results convincing? Are any existing results?

Does increased realism benefit other decision contexts?

Are our results convincing? Are any existing results?

Does increased realism benefit other decision contexts?

Alternative experimental designs?

Are our results convincing? Are any existing results?

Does increased realism benefit other decision contexts?

Alternative experimental designs?

Do we give up and go home?

Are our results convincing? Are any existing results?

Does increased realism benefit other decision contexts?

Alternative experimental designs?

Do we give up and go home?

Are there other ways to evaluate?

Are our results convincing? Are any existing results?

Does increased realism benefit other decision contexts?

Alternative experimental designs?

Do we give up and go home?

Are there other ways to evaluate?

For More Information

http://goo.gl/S5ona9

Related Design Study

Visually Comparing Weather Features In Forecasts

WED 8:30 AM, GRAND

Special thanks to our Ute Weather participants and to Jim Steenburgh for his insightful feedback

samquinan@sci.utah.edu

THE UNIVERSITY OF UTAH