TOWARDS ECOLOGICAL VALIDITY IN EVALUATING UNCERTAINTY

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

1 how closely the experimental setting matches the setting in which the results will be applied
ECOLOGICAL VALIDITY

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

EXPERIMENTAL CONTROL
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
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
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
control

CASE STUDIES
realism
USER STUDIES
control

CASE STUDIES
realism

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

control

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

**CASE STUDIES**

realism

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

<table>
<thead>
<tr>
<th>WEEK 1</th>
<th>baseline</th>
</tr>
</thead>
<tbody>
<tr>
<td>WEEK 2</td>
<td></td>
</tr>
<tr>
<td>WEEK 3</td>
<td></td>
</tr>
<tr>
<td>WEEK 4</td>
<td></td>
</tr>
<tr>
<td>WEEK 5</td>
<td>baseline</td>
</tr>
</tbody>
</table>
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

**Weeks:**

<table>
<thead>
<tr>
<th>Week</th>
<th>Activity</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>baseline</td>
</tr>
<tr>
<td>2</td>
<td>plume diagrams</td>
</tr>
<tr>
<td>3</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>baseline</td>
</tr>
</tbody>
</table>
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

<table>
<thead>
<tr>
<th>WEEK 1</th>
<th>baseline</th>
</tr>
</thead>
<tbody>
<tr>
<td>WEEK 2</td>
<td>plume diagrams</td>
</tr>
<tr>
<td>WEEK 3</td>
<td>mean + std dev</td>
</tr>
<tr>
<td>WEEK 4</td>
<td>spaghetti plots</td>
</tr>
<tr>
<td>WEEK 5</td>
<td>baseline</td>
</tr>
</tbody>
</table>
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

Ogdan-Hinckley [KOGD] (41.19361, 247.98361)
Forecast: 03Z17AUG2015

SREF 700mb TMP (°C)

PLUME DIAGRAM
DESIGN CHOICE 3  The Uncertainty Visualizations

SREF 700mb TMP – 10 °C

PLUME DIAGRAM

SREF 700mb TMP (C)

SPAGHETTI PLOT

Valid Time: 03Z18AUG2015

Single isovalue across all 21 members (see legend).
DESIGN CHOICE 3 The Uncertainty Visualizations

PLUME DIAGRAM

SPAGHETTI PLOT

MEAN + STD DEV
**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?

<table>
<thead>
<tr>
<th>TIME</th>
<th>SPACE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>
DESIGN CHOICE 3 Which Uncertainty Products?

<table>
<thead>
<tr>
<th>TIME</th>
<th>SPACE</th>
</tr>
</thead>
<tbody>
<tr>
<td>ENSEMBLE</td>
<td></td>
</tr>
</tbody>
</table>
**DESIGN CHOICE 3**  *Which Uncertainty Products?*

<table>
<thead>
<tr>
<th></th>
<th>TIME</th>
<th>SPACE</th>
</tr>
</thead>
<tbody>
<tr>
<td>ENSEMBLE</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CONTEXT</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
DESIGN CHOICE 3  Which Uncertainty Products?

SREF 700mb TMP (C)

PLUME DIAGRAM
DESIGN CHOICE 3 Which Uncertainty Products?

SREF 700mb TMP (C)

PLUME DIAGRAM
DESIGN CHOICE 3  Which Uncertainty Products?

PLUME DIAGRAM

Ogden–Hinkley [KCGD] (41.19381, 247.98361)
Forecast: 03Z17AU2015

SREF 700mb TMP (C)
DESIGN CHOICE 3  The Uncertainty Visualizations

SREF 700mb TMP – 10 °C

Forecast: 03Z17Jun2015

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

Forecast: 03Z 17JUL2015

SREF 700mb TMP (C)

Mean + Std Dev

TIME

SPACE

ENSEMBLE

CONTEXT

Valid Time: 18Z 20JUL2015
Ensemble Mean (contours) and Standard Deviation (fill)
DESIGN CHOICE 3  The Uncertainty Visualizations

SREF 700mb TMP (C)

MEAN + STD DEV

Forecast: 03Z 17 Aug 2015
Valid Time: 18Z 20 Aug 2015
Ensemble Mean (contours) and Standard Deviation (fill)

TIME

SPACE

ENSEMBLE

CONTEXT
DESIGN CHOICE 3  The Uncertainty Visualizations

SREF 700mb TMP (°C)

MEAN + STD DEV

TIME

SPACE

ENSEMBLE

CONTEXT
Design Choice 3: The Uncertainty Visualizations

SREF 700mb TMP (°C)

Mean + Std Dev
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

PROCESS

DECISION
DESIGN CHOICE 3  The Uncertainty Visualizations

**PLUME DIAGRAM**

**MEAN + STD DEV**

**SPAGHETTI PLOT**

**BASELINE**

PROCESS

DECISION

PROCESS

DECISION

PROCESS

DECISION

PROCESS

DECISION
EARLY ANALYSIS & RESULTS
- 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