User decisions – Can utility be used to guide forecast development?

2020 International Verification Method Workshop Online

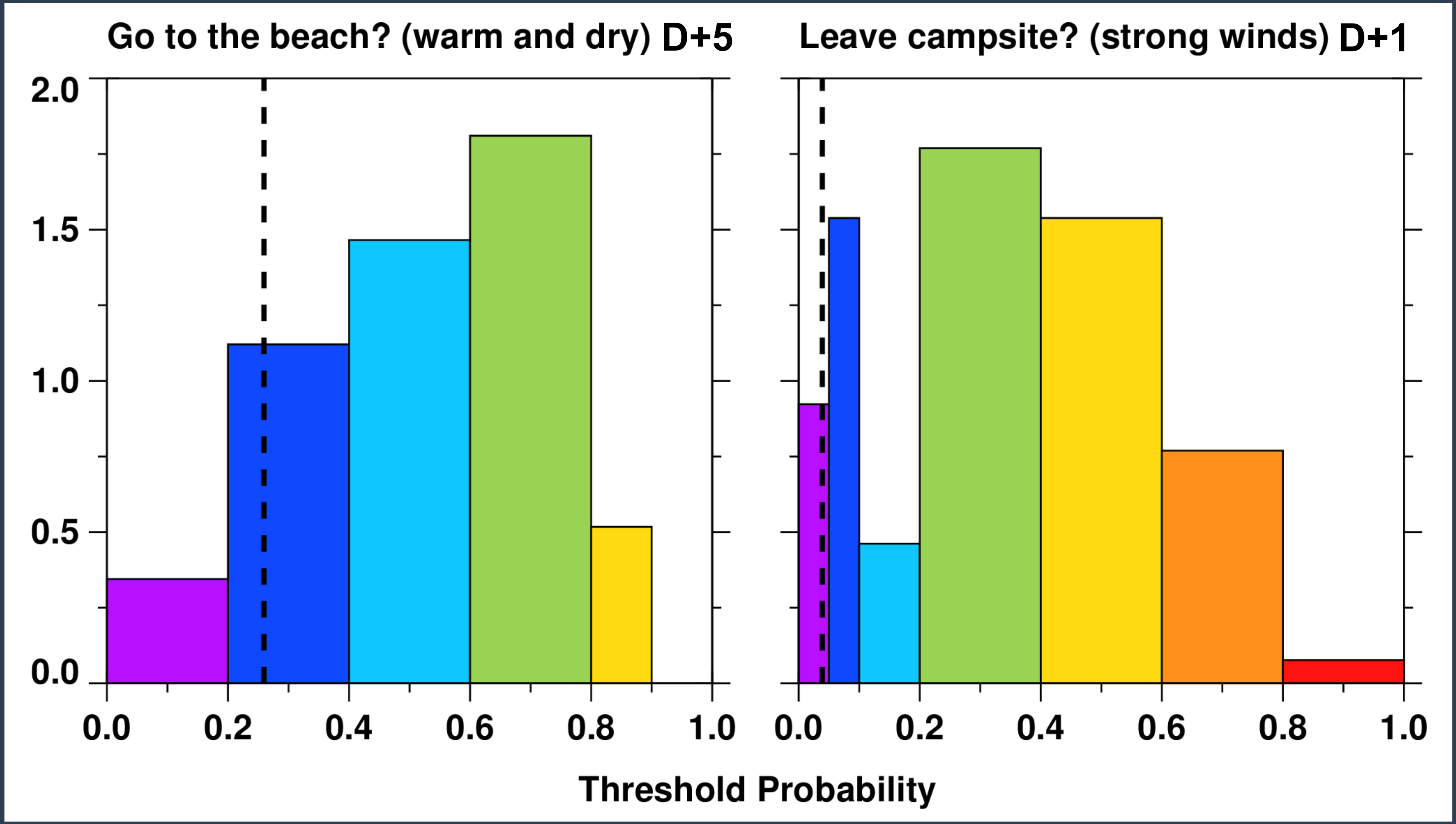
Mark Rodwell, John Hammond, Sara Thornton, David Richardson
https://doi.org/10.1002/qj.3845

ECMWF
mark.rodwell@ecmwf.int

Based on a fun event at the Royal Meteorological Society Christmas meeting (available on Facebook Live)

ECMWF

Live science event: Participants' distributions of threshold probabilities



Participants generally leave the campsite at lower probabilities than they go to the beach

User decisions – Can utility be used to guide forecast development?

User action $\mathbb{I} = [p > p_T]$

p = forecast probability, p_T = threshold probability

$\mathbb{I} = 0\ (1)$ if expression is false (true)

If we could use simple 'cost-loss' model:

Expense per unit loss L

$$E = \alpha\mathbb{I} + o(1 - \mathbb{I}) = (\alpha - o)\mathbb{I} + o$$

for cost C with $\alpha = C/L$ and outcome $o \in \{0,1\}$

Expected expense $\mathbb{E}(E) = \{\alpha - o(p)\}\,\mathbb{I} + o(p)$

for expected outcome $o(p) \equiv \mathbb{E}(o|p)$

Bayes action $\mathbb{I}_B = [o(p) > \alpha]$

minimises expected expense

Key question: Does $\mathbb{I} = \mathbb{I}_B$?

If we can show* that $p_T = \alpha$ then $\mathbb{E}(E)$ is minimised when $p = o(p)$: so E a proper score of the forecast

*Can user's generalised matrix of feelings $\begin{pmatrix} -\text{Satisfaction} & \text{Regret} \\ \text{Pain} & -\text{Thrill} \end{pmatrix}$ be reduced to Cost-Loss model?

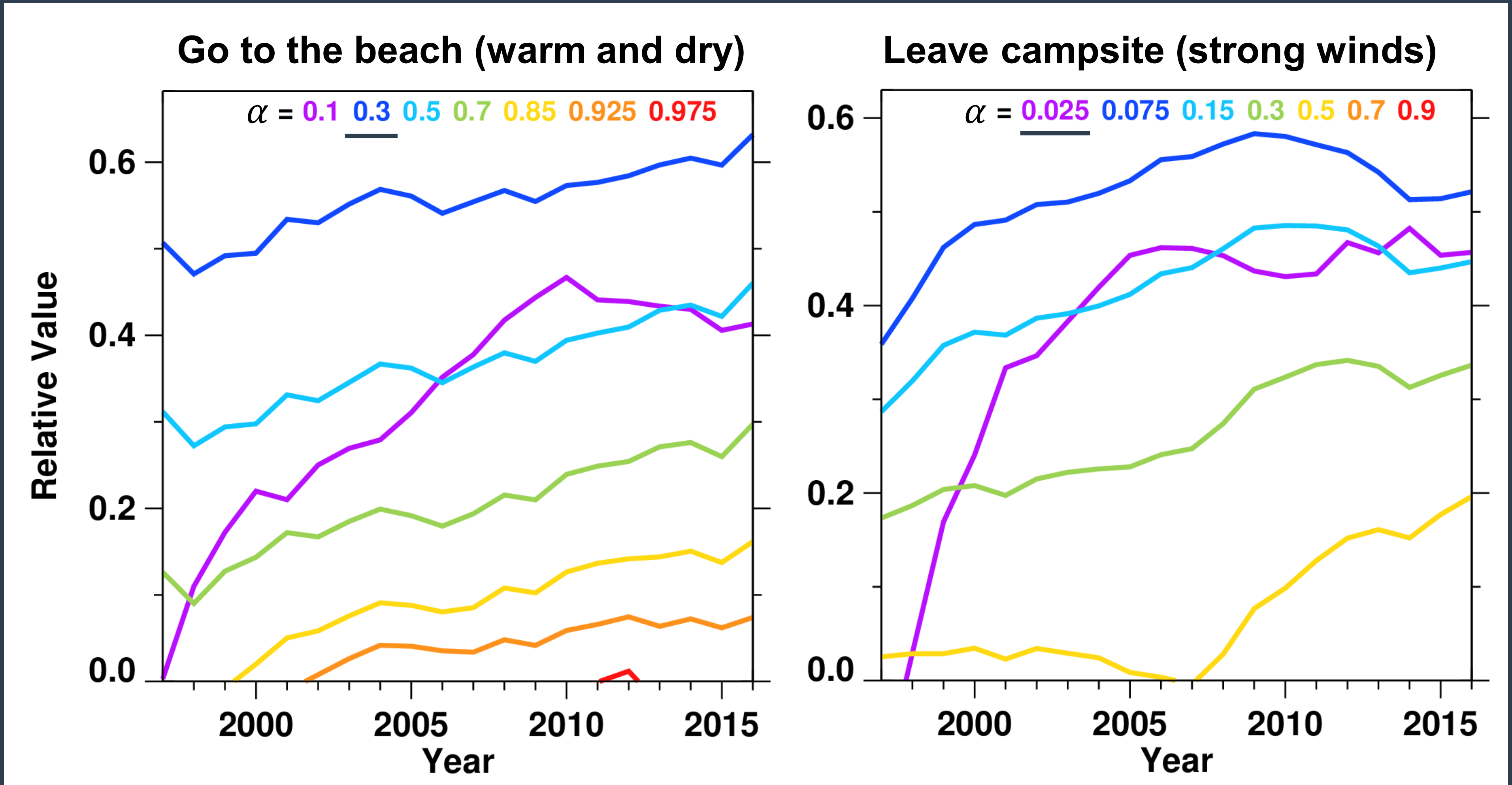
Can threshold probabilities p_T be equated with (subjective) cost/loss ratios α?

Key factors affecting a participant's decision	Threshold probability
"I love being on beaches, whatever the weather"	0%
"I hate sitting on the beach in the rain … and with three kids it's quite an expedition"	70%
"With a low probability, I'd feel responsible for taking away my family's fun. However, as a parent, I wouldn't want to put very young kids at risk of flying branches"	50%
"I don't really go camping … I may as well stay ... A case of making it an adventure with the family pulling together to stop the tent being blown away"	70%
"I never go camping – how should I vote?"	?

Distribution of p_T is a reasonable approximation to distribution of cost/loss ratios α.

Some 'risk-seeking' behaviour due to lack of first-hand experience with rare, yet dangerous, events

Relative value $V(\alpha)$ of ECMWF ensemble forecast (vs point-observations)

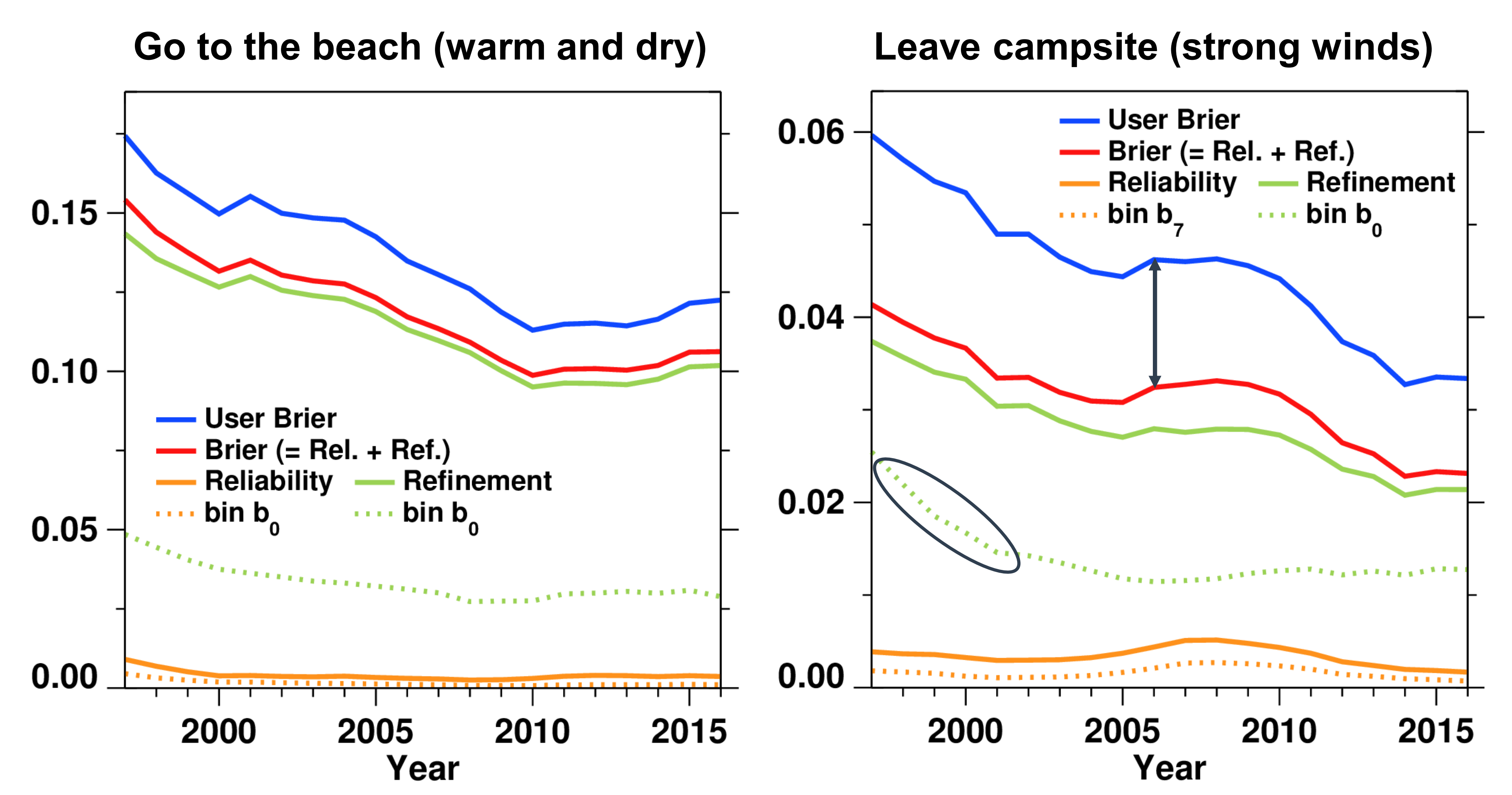


$$V(\alpha) = \frac{\overline{E}_{clim} - \overline{E}_p}{\overline{E}_{clim} - \overline{E}_{best}}$$

Each term is averaged over sample of forecasts

'Complete misses' (no ensemble members capture the event) were/are a major issue – especially for strong winds

'User Brier Score' (UBS) and Brier Score for ECMWF ensemble forecast



New Score

$$UBS = \frac{\tilde{\tilde{E}}_p - \tilde{\tilde{E}}_{best}}{\tilde{\tilde{E}}_{worst} - \tilde{\tilde{E}}_{best}}$$

Each term is averaged over sample of forecasts and integrated over the distribution of cost/loss ratios

UBS is (asymptotically) proper, lies in the range [0,1] and reduces to the Brier Score when the distribution of Cost/Loss ratios is uniform

UBS is larger than BS since users were not interested in high probability thresholds

Users would benefit from better representation of extreme weather in the tails of the ensemble distribution

Taking the experiment a little further

Problems with previous experiments and investigations

- “I never go camping – how should I vote?”
- “Commercial users can be reluctant to reveal their hand when given a particular scenario”

Proposed solution (small step forward)

- New *Live Science* event at the end of this session (please stay at the end)

Live Science Event

Taking the User Decision experiment a little further

Problems with previous experiments and investigations

- “I never go camping – how should I vote?”
- “Commercial users can be reluctant to reveal their hand when given a particular scenario”

Proposed solution

- Forecasters are not interested in camping either! - Only the value of their forecasts for a given weather event
- Commercial users may reveal their threshold probabilities if they are anonymous participants in a wider analysis of all those interested in a given weather event

Today’s experiment

- You will choose your own scenario associated with a given weather event

Weather event: Tomorrow’s winds >11ms^{-1} (with stronger gusts)

How the event is structured

Choose a scenario affected by this weather event

- Weather might be beneficial or detrimental
- Scenario might be personal or commercial
- Risks might be large or small
- Examples: Sell wind energy, …, Don't hang the washing outside, …

Picture the scenario in your mind

- Who/what is involved?
- What are the costs: financial or subjective?
- What is the decision to be made?

We will start with a probability of 1

- What will be your decision?

We will slowly decrease the probability towards 0

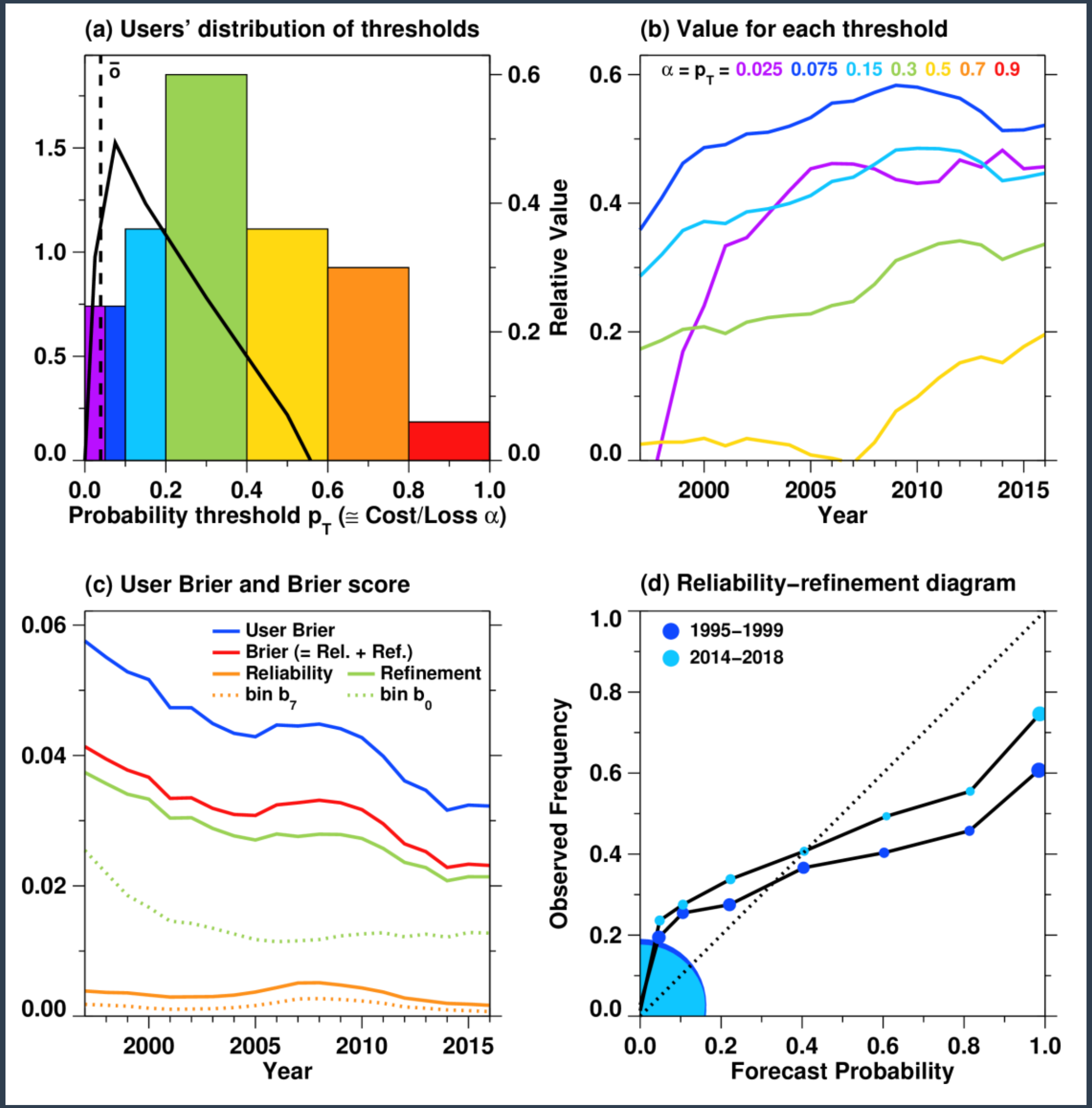
- Raise your hand when you change your decision (<u>and then leave it raised</u>)

Afterwards, it would be great to know what your scenario was. Either speak, type in the chat, or email mark.rodwell@ecmwf.int

1. Your scenario
2. Threshold probability
3. Reasoning (key factors)

Results from the *Live Science* event at the workshop

We had 27 active participants, and lots of interesting discussion afterwards. The top left panel shows the participants' distribution of threshold probabilities for changing their decision. 30% is a favourite, but there is no right or wrong answer here – it depends on each participant's own scenario and feelings about this. The range of scenarios and key factors are indicated on the next slide. Some scenarios are 'trivial' everyday decisions while others have major significance (presidential visit etc). Arguably all are important for a user-oriented approach to forecast verification. The reasoning (again) suggests that participants are attempting to make their Bayes Action (optimise utility). When integrating over all the users, the User Brier Score is larger than the Brier Score (less focus on high thresholds). It doesn't suffer so severely from 'complete misses' because the range of scenarios were not so associated with high losses.



Scenario	Threshold probability	Key factors / comments
Don’t cycle to work	30%	Alternative: I take the train The threshold might be higher if the temperature is warm, and lower if the temperature is cold (or if there is precipitation).
Don’t rent a beach umbrella	40%	If there is strong wind, I have to keep the umbrella closed and I pay the rent for nothing. if the wind is not strong and I do not have an umbrella I can get sun burn
Don’t plan a beach volleyball tournament	40%	Knowledge of local climatology was a factor. A lot of people’s feelings (pain, regret etc) to consider
Don’t plan a table tennis tournament	20%	A lot of people to consider
Secure sun canopy for a wedding	10%	Important event with potential for high loss. Blowing dust also a factor
Configure traditional windmill to avoid damage	50%	If I have no wind to mill flour I have to use electric powered stones. If the wind is too much then I may lose the sails. Force 6-7 is slightly higher than my optimum wind speed. High demand for flour during Covid pandemic
Pack a wind-proof jacket	10%	More for comfort than necessary for protection. Low cost decision (easy to pack)
Don’t go cycling	-	A higher windspeed definition would have been more appropriate. Feelings vary smoothly with wind intensity
Don’t go mountain hiking	40%	Can adjust the hike route. Temperature and wind-chill also important. Experienced hiker and danger not a factor
Don’t camp in the garden with the little ones	10%	Easy to do it next weekend instead, so even a small chance of poor weather makes it worth postponing
Go outside for exercise without presence of city smog	40%	A case of winds being beneficial for scenario
Abandon presidential visit (cannot fly surveillance drones)	30%	Event of high importance / sensitivity
Don’t cycle to work	30%	I don’t like wind
Don’t cycle for leisure	40%	-
-	-	A key consideration is frequency of the scenario. More likely to do something despite the weather if the scenario is rare
		A key consideration is whether you have the ability to make or change your decision (if it is planned a long while in advance)
		Responses might be affected by the probability thresholds that are offered
		A probability forecast of 11 m/s gives an implied probability of much stronger winds, which might be part of the consideration