The Bureau of Meteorology

Opportunities for interdisciplinary research to improve the evaluation of S2S forecasts

Beth Ebert and Carla Mooney

Acknowledgements: Caio Coelho

S2S Real Time Pilot Workshop 15-17 Nov 2022

Photo: Riverside Station, Queensland; Bureau of Meteorology

SAGE: Sub-seasonal Applications for aGriculture and Environment

A new 5-year project of WWRP

Enhanced physical science base

- Research on sub-seasonal to seasonal sources of predictability
- Coupled atmosphere-ocean-land assimilation and prediction systems that resolve large-scale and small-scale processes

Focus on users

- Development of products that serve a variety of actors including agriculture, water resource management, renewable energy, etc.
- Probabilistic prediction
- Communication with users and partners including warnings for slow onset disasters

Enhancing and securing the necessities of life – sustainable development goals, food security, energy and well being

Photo: Wayne Twine, Wits Rural Facility

SAGE drivers, research and approaches

Synthesis from August 2022 WWRP Symposium

Barriers to successful S2S forecasting and priorities to address those barriers

1. Limited skill beyond week 4

- Research to understand sources of predictability
- Coupled modelling systems to resolve multi-scale processes
- Knowing where forecasts exhibit skill
- Focus on model biases and skill for monsoons, links to ENSO, extremes, etc.

2. Data not used effectively

- Tailored and co-produced products
- Knowledge of appropriate actions for different futures
- Information allowing users to understand confidence and possible impacts
- Enable greater forecast usefulness (frequency, format, how supplied, etc.)

3. Insufficient user engagement

- Science-user interface and role of private sector
- Enhance communication modes designed with users and producers
- Co-design evaluation metrics with users
- Involve social scientists to assist in user engagement

WWRP has two working groups with expertise in evaluation

Joint Working Group on Forecast Verification Research

- Promote the development and application of improved diagnostic verification methods to assess and enable improvement of the quality of weather forecasts, including forecasts from numerical weather and climate models.
- Engage in the plans and implementation of the verification component of WWRP projects from the outset.

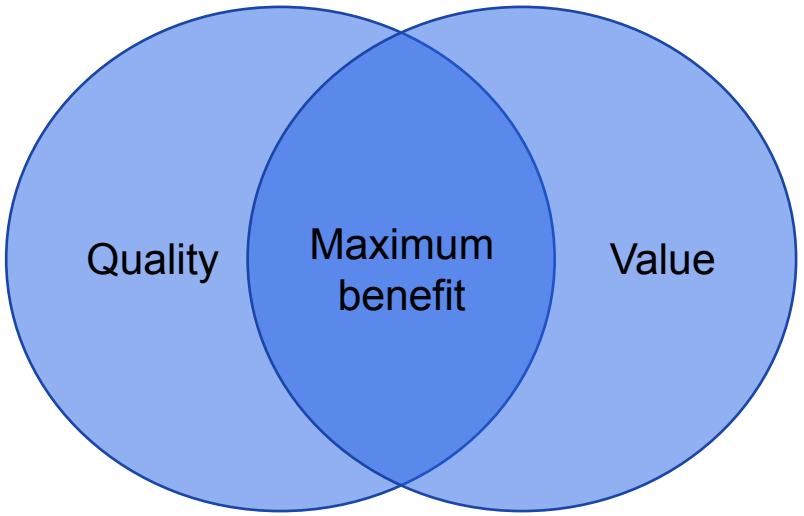
Working Group on Societal and Economic Research Applications

- Advance the science of the social and economic application of weather-related information and services
- Knowledge of how to frame, design, and implement research projects co-designed between physical and social scientists and a range of appropriate actors to achieve more useful information for decision-makers and the public.

Verification and evaluation

Verification perspective (Murphy, 1993)

- **Quality** – corresponds to what actually happened
  - Accurate
  - Unbiased
  - Reliable
  - More skilful than some reference forecast
- Scientific approach



Evaluation perspective

Who is the evaluation for?

- **Researcher** = robust method, published output, impact factor
- **User** = useful, useable, used
  - Relevant
  - Timely
  - Accessible
  - Clear about uncertainty
  - Applies to forecast outcome as well as forecast value chain

SAGE: Sub-seasonal Applications for aGriculture and Environment

A new 5-year project of WWRP

Enhanced physical science base

- Research on sub-seasonal to seasonal sources of predictability
- Coupled atmosphere-ocean-land assimilation and prediction systems that resolve large-scale and small-scale processes

Focus on users

- Development of products that serve a variety of actors including agriculture, water resource management, renewable energy, etc.
- Probabilistic prediction
- Communication with users and partners including warnings for slow onset disasters

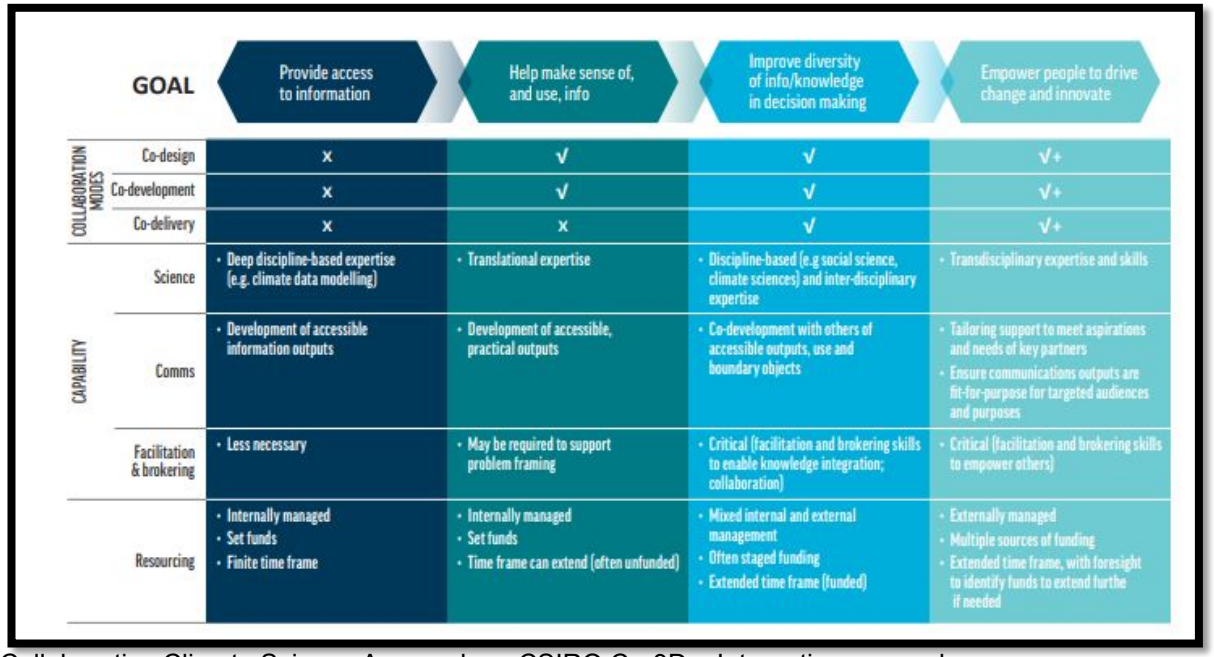
Opportunities for interdisciplinary research to improve the evaluation of S2S forecasts

- Co-develop and co-evaluate impact-based S2S forecasts
- Improve the communication of probabilistic forecasts and their quality and value
- Assess the value chain for S2S services and high impact events

Co-design/co-evaluation

An interdisciplinary opportunity

- Misalignment between researchers' aims and research end – user needs is a major cause of research waste (Slattery et al. 2020).
- Co-design requires commitment, a mix of skills, resources and time.



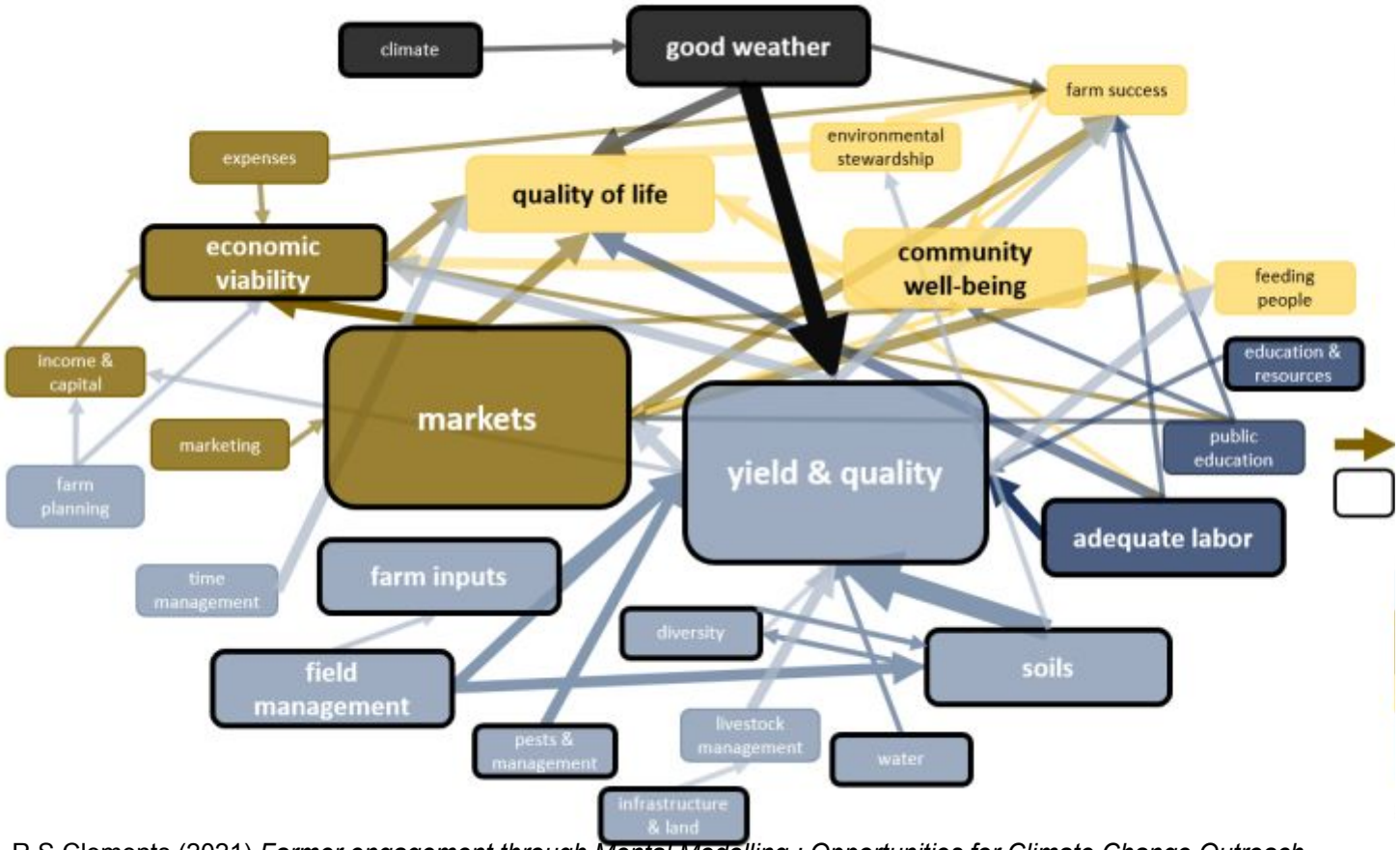
	GOAL	Provide access to information	Help make sense of, and use, info	Improve diversity of info/knowledge in decision making	Empower people to drive change and innovate
COLLABORATION MODES	Co-design	x	√	√	√+
	Co-development	x	√	√	√+
	Co-delivery	x	x	√	√+
CAPABILITY	Science	• Deep discipline-based expertise (e.g. climate data modelling)	• Translational expertise	• Discipline-based (e.g social science, climate sciences) and inter-disciplinary expertise	• Transdisciplinary expertise and skills
	Comms	• Development of accessible information outputs	• Development of accessible, practical outputs	• Co-development with others of accessible outputs, use and boundary objects	• Tailoring support to meet aspirations and needs of key partners • Ensure communications outputs are fit-for-purpose for targeted audiences and purposes
	Facilitation & brokering	• Less necessary	• May be required to support problem framing	• Critical (facilitation and brokering skills to enable knowledge integration; collaboration)	• Critical (facilitation and brokering skills to empower others)
	Resourcing	• Internally managed • Set funds • Finite time frame	• Internally managed • Set funds • Time frame can extend (often unfunded)	• Mixed internal and external management • Often staged funding • Extended time frame (funded)	• Externally managed • Multiple sources of funding • Extended time frame, with foresight to identify funds to extend furthe if needed

Collaborative Climate Science Approaches, CSIRO Co-3D – Integration research

Disciplinary mix including science, communications, facilitation, knowledge brokering, evaluation, applied social science

Co-design requires different ways of thinking

Mental models help with understanding how information is used and its relevance



A theory of change (program logic) can help design the intervention to provide relevant/timely/useful forecast information to influence decision making

R.S.Clements (2021) *Farmer engagement through Mental Modelling : Opportunities for Climate Change Outreach*

Barriers to co-evaluation

User engagement

- Too late / tokenistic
- Insufficient interest
- Lack of time
- Evaluation process or results hard for users to understand

Procedural

- Evaluation process not well designed
- Appropriate verification / evaluation methods not used
- Verification / evaluation results not shared
- No baseline for evaluation

Technical

- Data for verification / evaluation doesn’t exist, or is hard to obtain
- Users may have difficulty sharing data with researchers

Commercial

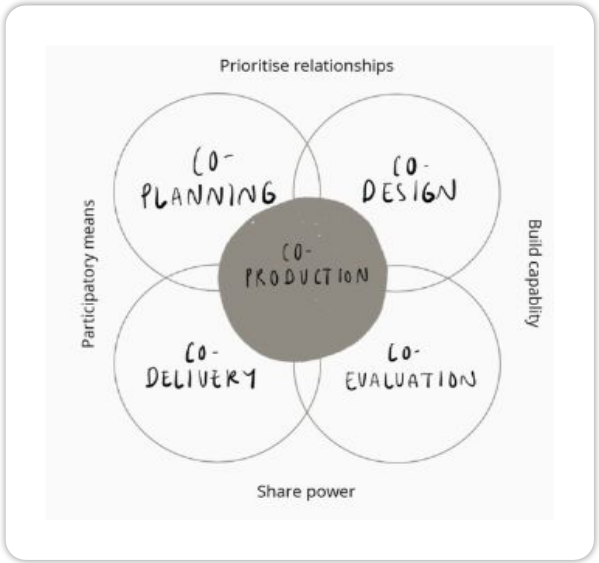
- Commercial users may not be allowed to share data

Project co-design for evaluation

Project design needs to involve users, have clarity about project purpose, program logic, agreed outcomes, measures and relevant evaluation method

Types of evaluation	Description
Process evaluation	How the program was implemented
Outcome evaluation	The effectiveness of the program
Economic evaluation	Costs compared to outcomes
Impact evaluation	Effectiveness in achieving goals
Developmental evaluation	Regular checks to make sure program is on track
Realist evaluation	What worked, for who, how, in what way, when
Utilisation-focused evaluation	Evaluation useful for the user – to enhance utilisation, inform decisions, improve performance

Social Policy Evaluation and Research Unit, Making sense of evaluation (2017) Superu.gov.nz



What is co-design? — Beyond Sticky Notes

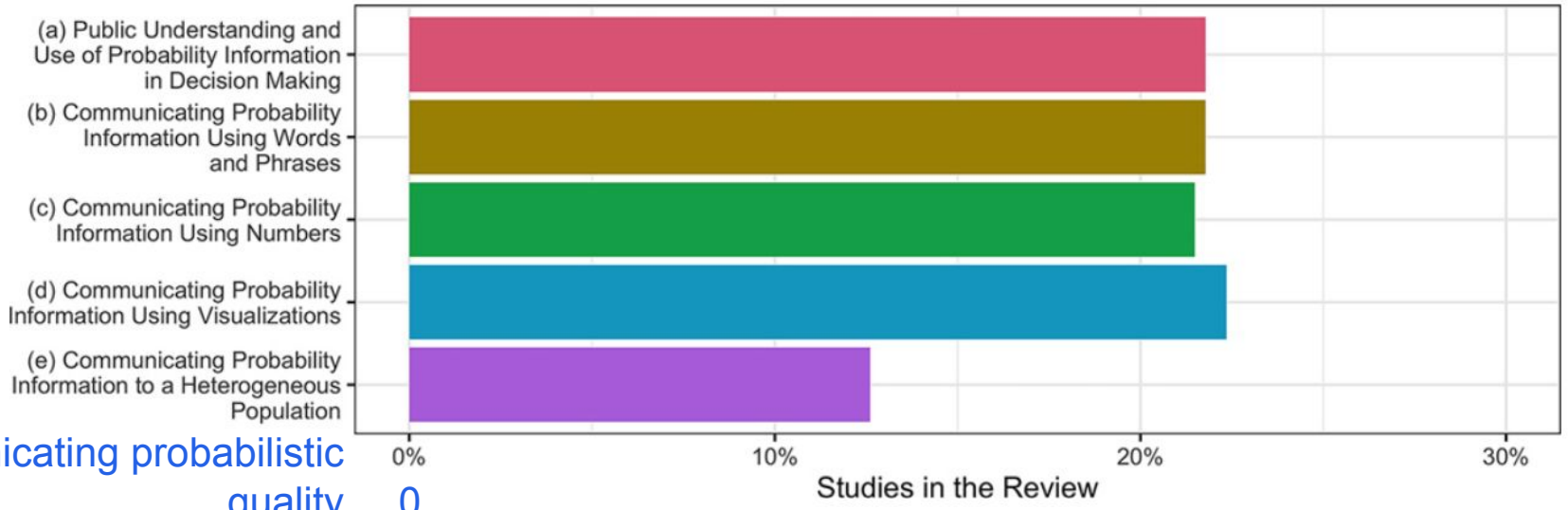
User based evaluation involves the people for whom the system was intended and measures the things of value to them.

Communicating probabilistic forecasts and their quality / value

Communicating Probability Information in Weather Forecasts: Findings and Recommendations from a Living Systematic Review of the Research Literature

JOSEPH RIPBERGER,[a] ANDREW BELL,[a] ANDREW FOX,[a] AARIKA FORNEY,[a] WILLIAM LIVINGSTON,[a] CASSIDY GADDIE,[b] CAROL SILVA,[a] AND HANK JENKINS-SMITH[a]

[a] *University of Oklahoma, Norman, Oklahoma*
[b] *University of Georgia, Athens, Georgia*



Communicating probabilistic quality 0

- Accuracy of probabilistic forecasts is a second-order uncertainty – a huge challenge to communicate!
- Very little work has been done in this space
- Rather than communicating probabilistic *accuracy*, go straight to *value*?

Communicating probabilistic forecasts and their quality / value

User types		Use / concern
	Low stakes users	Losses from over- and under-predicting are similar
	General assessors	Uncertainty bands provide general assessment of likely forecast quality
	Change assessors	Is a change within the bounds of uncertainty or does it warrant action?
	Risk avoiders	Keep the risk of an adverse outcome to an acceptable level – tails of distribution
	Decision theorists	Explicit, quantitative loss function (typically involving money)

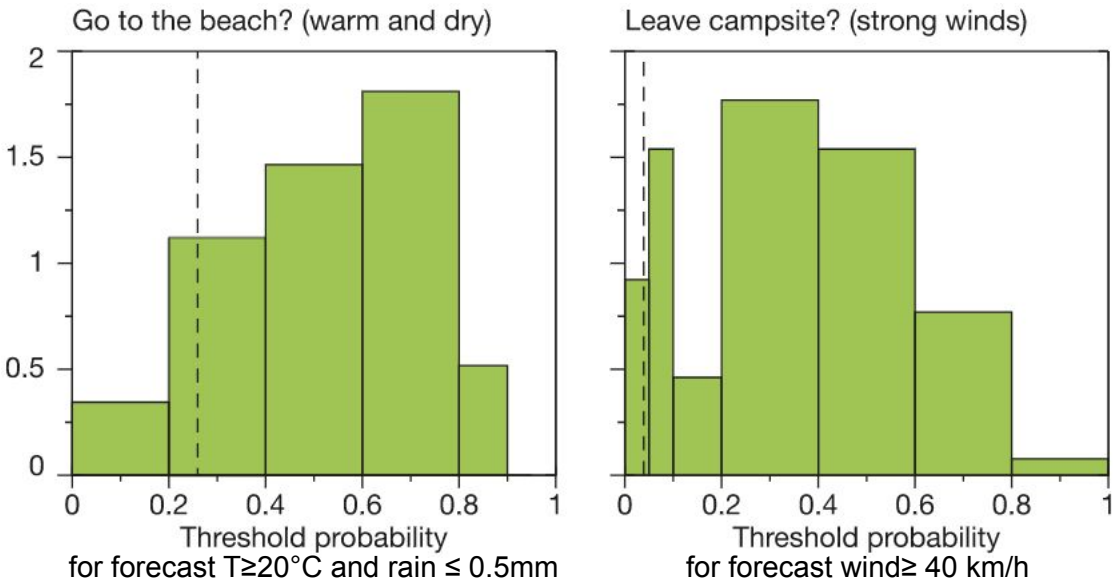
Adapted from Raftery, 2016, *Use and communication of probabilistic forecasts*

Research questions

- How should forecast quality / value be evaluated?
- What metrics are meaningful for the users' applications?
- How should forecast quality / value be communicated?

Toward verification that incorporates user decisions

Understanding user decisions (Rodwell et al. 2020)



- Users identify the threshold probability which optimises their expected feeling about their decision
- Distribution of user thresholds ~ users' "cost/loss ratios of feelings" (thrill, pain, regret)
- User Brier Score (UBS) a more realistic assessment of user's forecast value than the traditional Brier Score

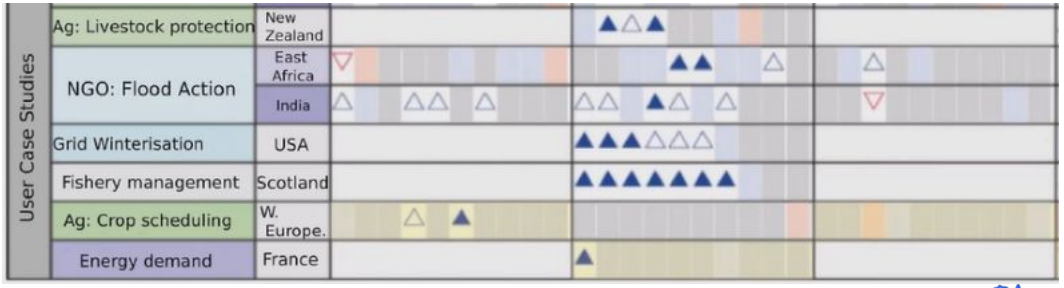
Quarterly Journal of the Royal Meteorological Society RMetS

RESEARCH ARTICLE

Beyond skill scores: exploring sub-seasonal forecast value through a case-study of French month-ahead energy prediction

Joshua Dorrington[1] | Isla Finney[2] | Tim Palmer[1] | Antje Weisheimer[1,3]

- Simplified model of trading strategy to optimise value
- Non-meteorological factors may affect forecast value
- User applications have thresholds, cutoffs, nonlinearities that can make purely meteorological scores misleading
- Augment the model scorecard to include user value

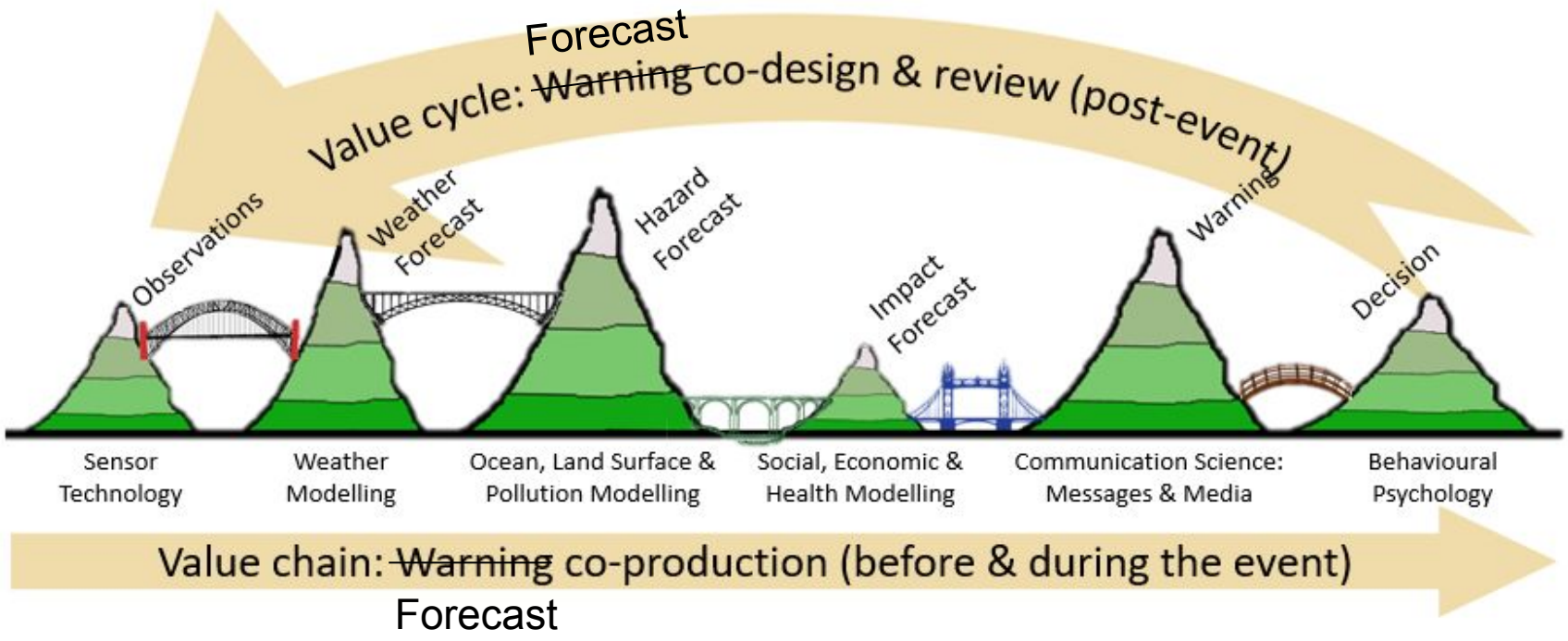


Evaluating the S2S forecast value chain

HIWeather

Research questions

- What does the value chain look like for a particular S2S application?
- How do improvements in the chain flow through to user benefit?
- How did the S2S service perform for a high impact event?



Possible approaches

- Co-create a logic model (flowchart) linking user benefit to improved accuracy and value
- Assess baseline performance using verification and evaluation approaches
- Estimate the benefit of different improvements along the value chain
- Use/adapt HIWeather value chain case study template (hiweather.net) to study high impact events at S2S timescales

Co-design and co-evaluation of impact-based S2S forecasts

- S2S applications are now being used in forecast operations to predict the likelihood of drought, flooding, tropical cyclone occurrence, abnormal cold and heat
- Advanced S2S applications are being developed and used to predict possible impacts on agriculture, energy, water resources, forestry, fishing, public health, shipping, defence, disaster risk, media…

Research questions

- How can the user's decision making process best inform both the design of the application and the evaluation approach?
- How to appropriately measure the use, useability and usefulness of the solution and ultimately its impact on the outcomes of the decision?
- How can data on impacts best be acquired and used in evaluation?
- Co-design a new S2S application from the bottom up
  - Deep user engagement from the start (planning stage)
  - Bottom up verification and evaluation approach centred on the user needs
  - Focussed collaboration on impact data gathering and processing
  - Agreement on the expected outcomes

Toward "what the weather will do…"

Interdisciplinary research – opportunities and challenges

Increasingly regarded as the key to tackle contemporary complex societal challenges and to stimulate scientific innovation (Sun et. al. 2021)

Opportunities of interdisciplinary research	Challenges for interdisciplinary research
• Different disciplines working together in a team • Collaboration to bring different perspectives to solve a problem • Utilisation of methods normally associated with one or more disciples to solve problems in another discipline researchers • Translation of innovative or applied research outcomes from one discipline into another discipline (ARC 2018).	• Time • Disruption • Language • Expectations • Methods • Lower citation rates • Who is your user?

The Bureau of Meteorology

Thank you

Beth Ebert
beth.ebert@bom.gov.au

Carla Mooney
carla.mooney@bom.gov.au