Developing a quick guide on presenting data and uncertainty

Lickiss, M., Mulder, K. J., Black, A., Charlton-Perez, A. J. ORCID: https://orcid.org/0000-0001-8179-6220, McCloy, R. A. ORCID: https://orcid.org/0000-0003-2333-9640 and Chandler, R. (2017) Developing a quick guide on presenting data and uncertainty. Weather, 72 (9). pp. 266-269. ISSN 1477-8696 doi: 10.1002/wea.2998 Available at https://centaur.reading.ac.uk/69829/

Developing a quick guide on presenting data and uncertainty

Weather – September 2017, Vol. 72, No. 9

Matthew Lickiss,[1] Kelsey J. Mulder,[2] Alison Black,[1] Andrew J. Charlton-Perez,[2] Rachel A. McCloy[3] and Richard Chandler[4]

[1]*Department of Typography & Graphic Communication, University of Reading*
[2]*Department of Meteorology, University of Reading*
[3]*Department of Psychology, University of Reading*
[4]*Department of Statistical Science, University College London*

Introduction

As a discipline dealing with large amounts of complex data, meteorology relies on communication through numbers, graphs and charts. While atmospheric scientists are skilled users of these communication modes, many of their readers are not trained to interpret complicated information. Difficulties in interpretation may particularly affect those who are not familiar with the conventions of data presentation, nor with the use of data relating to forecasts subject to uncertainty. Spiegelhalter *et al.* (2011) give an excellent overview of some of the difficulties in communicating uncertainty about the future, and Gigerenzer *et al.* (2005) demonstrate the specific interpretation problems that can occur when percentages are used in rain forecasts. As part of the PURE (Probability, Uncertainty & Risk in the Environment) research programme, funded by the UK Natural Environment Research Council, an interdisciplinary team of meteorologists, designers and psychologists at the University of Reading carried out survey and experimental studies on user interpretations of, and decisions based on, environmental risk information (e.g. Mulder *et al.*, 2017). This collaboration gave rise to more general consideration of environmental data communication and its role in enabling professional and public analysis and decision making. One outcome of this project was the creation of a short leaflet: *Presenting Data and Uncertainty*, the development and testing of which is described below. A full copy of the leaflet is shown in Figure 1.

The context for leaflet development

Current sources of advice

There is current advice, available to atmospheric scientists, on presenting uncertainty in forecast communication. The World Meteorological Organisation (WMO) has published *Guidelines on Communicating Forecast Uncertainty* (WMO, 2008), with a summary guide, *Communicating Forecast Uncertainty,* also available for download from the web (WMO, undated). In addition, the Netherlands Environmental Assessment Agency (PBL) has published two guideline booklets on communicating uncertainty (Petersen *et al.*, 2013; Wardekker *et al.*, 2013). The WMO guide is heavily illustrated with colour examples, *with the central aim of assisting NMHSs* [National Meteorological and Hydrological Services] *to develop strategies and techniques to communicate uncertainty information as part of their services* (WMO, 2008, p. 2). The PBL guides, in contrast, have fewer illustrations and a detailed, primarily text-based approach, with the aim that they help *researchers and communicators to obtain a clear picture of why uncertainty communication is important and to whom it should be addressed, and stimulates them to make well-considered choices when deciding on what, where and how to present this information*. (PBL Netherlands Environmental Assessment Agency, 2013). The length of the main guides from both sources led us to question their suitability for quick reference and potential uptake by their target audience of researchers and professionals who communicate uncertainty.

The substantial research literature on visualising and communicating uncertainty, which is fragmented across a wide range of fields (medicine, physics, chemistry, computer science, business, etc.) as well as GIS (Geographic Information Science) and atmospheric science, is not amenable to quick referencing for guidance on data presentation issues. Even longer guides, such as those of the WMO and PBL, cannot fully represent the scope of relevant research. While some research review articles present a more general approach, such as (Pang *et al.*, 1997; MacEachren *et al.*, 2005; Brodlie *et al.*, 2012), others remain within a specific field of application (e.g. Bostrom *et al.*, 2008 on 'seismic risk'). Experimental studies within the literature often deal with specific situations of communication. Cheong *et al.* (2016), for example, studied the effects of visualisation of risk surrounding wildfire 'burn likelihood' in a static, map based, decision-making task. Necessarily, such studies use controlled variables, but this means it can be difficult to judge to what extent recommendations from one specific situation of communication can be carried over into different systems and contexts.

We had an opportunity to consult informally with producers and consumers of probabilistic and uncertain information at a workshop attended by academic and industry specialists to consider the forecasting and communication of volcanic ash concentrations (at the Royal Academy of Engineering, London on 22 February 2016). The consultation revealed that participants were not familiar with the WMO or PBL guideline documents. Subsequent discussion with researchers and PhD students in the Department of Meteorology at the University of Reading indicated that they would be unlikely to consult long guides on data presentation, even if aware of them.

Observation of current communications

Our experience with atmospheric science data presented in journals (and wider uncertainty literature) and conferences suggests that shortcomings in data and uncertainty presentation are not unusual. There are often fundamental communication weaknesses, including a lack of clear titling of graphs, little contextualisation for data representations, and indistinct colour use. While these and similar features are not unique to the communication of uncertainty, they can further complicate data interpretation in this already difficult area.

RMetS
Royal Meteorological Society

This guide contains key points to consider when presenting data. It highlights general best practice to make sure your message is understood, with a focus on presenting and framing uncertainty.

Even the best data can be misread or misinterpreted if presented poorly or without adequate context. This is especially true of data relating to uncertainty. Providing information about uncertainty connected to forecasts allows decision makers to make more informed choices relevant to their specific requirements.

Extra attention to clarity and detail is needed when groups with different expertise and interests are communicating with one another. A certain type of chart might be an everyday form of presentation for an atmospheric scientist, but wholly unfamiliar to a government decision maker or journalist. With increasing public access to specialist communications via the web and open access, it pays to consider a potential readership wider than your peers whenever you disseminate data.

> *Without an understanding of uncertainty amongst the public and policymakers alike, scientists will struggle to talk about uncertainty in their research and we will all find it hard to separate evidence from opinion.*
>
> ***– Sense About Science***

further reading

Brewer, C. and Harrower, M., 2009–2013. Color Brewer 2.0 Color Advice for Cartography, URL: colorbrewer2.org

MacEachren, A. M., Roth, R. E., O'Brien, J., Li, B., Swingley, D. and Gahegan, M. 2012. 'Visual Semiotics & Uncertainty Visualization: An Empirical Study', *IEEE Transactions on Visualization & Computer Graphics*, vol.18, no. 12: 2496–2505

Spiegelhalter, D., Pearson, M. and Short, I. 2011. 'Visualizing Uncertainty About the Future', *Science*, 333 (6048): 1393–1400

Tufte, E. 2001. *The Visual Display of Quantitative Information*, 2nd Edition, Graphics Press

Sense About Science. 2013. *Making Sense of Uncertainty*, URL: www.senseaboutscience.org/data/files/resources/127/SAS012_MSU_reprint_compressed.pdf

WMO. 2008. *Guidelines on Communicating Forecast Uncertainty*, (WMO/TD No. 1422, English version), World Meteorological Organization, URL: www.wmo.int/pages/prog/amp/pwsp/publicationsguidelines_en.htm

about

The Robust Assessment and Communication of Environmental Risk (RACER) research team seeks to improve the quantification and communication of uncertainty and risk in natural hazards. We are funded by the Natural Environment Research Council (NERC) under the Probability Uncertainty and Risk in the Environment (PURE) Programme. This leaflet was written as part of an interdisciplinary project at the University of Reading with the Departments of Meteorology, Psychology, and Typography & Graphic Communication.

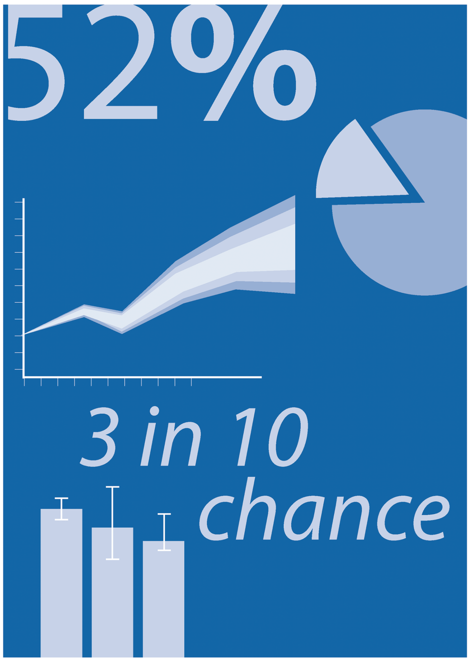
NERC SCIENCE OF THE ENVIRONMENT
University of Reading

Research funded by NERC PURE network
www.nerc.ac.uk/research/funded/programmes/pure

Produced by the University of Reading *www.reading.ac.uk*

presenting data and uncertainty

a quick guide



communication choices

Information can be communicated through words, numbers, diagrams and graphs, each with their own strengths and weaknesses.

Numbers may appear to be unambiguous, but people do not always interpret numerical information as expected – especially probability and percentages. For non-technical audiences, consider re-wording percentage probabilities, e.g. '30%' as a '3 in 10' chance. Exponential notation may not be familiar to non-technical audiences and should be contextualised or avoided where practical, e.g. 0.004 instead of 4×10^{-3}.

Combination and repetition across different representations can strengthen a message. Numbers might show the core data, a graph might show overall relationships, while words give context or elaboration. Consider which representations and combinations are most appropriate for your message and data, the target audience, and the medium of communication.

Vocabulary may be understood differently across groups, e.g. scientists, industry and the public. Clarify potentially ambiguous words such as 'uncertainty' or 'error' where audiences' interpretations may differ.

Frame risk statements to help people make balanced judgements. E.g. people may respond differently to a 40% chance of an event occurring, compared to a 60% chance of it not occurring (ideally, provide both).

CDF and PDF: Cumulative Distribution Function graphs (CDFs) may be misinterpreted as Probability Density Functions (PDFs) if they show line plots that might be possible for either function. All graphs need clear titles, labels, and contextualisation.

labels, context and colour

Clear and consistent labelling and contextualisation lower the risk of misinterpretation. While this seems fundamental, it is often forgotten.

Support interpretation of graphs, charts and tables with clear titles, labelling of axes (with units) and keys. Many readers will not read a document sequentially, so interpretation of material outside the main text should be possible without the reader needing to look between the two. Annotations can provide point-specific explanations.

Explain uncertainty by showing its sources, options for addressing it and the range of possible outcomes. While this may not be applicable in all cases, consider what contextual information is needed by the reader to make sense of or act upon the data.

Colour interpretation may vary across readers. It is essential to provide a key, but not everyone will refer to it.

- 'Inherent' colour meanings may not be interpreted consistently. For example, readers might interpret blue as representing high values in relation to rainfall, but low values in relation to temperature.
- Too many colours can make it difficult to identify and track data series, especially if the colours used are too close in hue to be distinguished easily. Use clearly distinct colours when colour coding variables.

It is difficult to keep track of too many similar colours.

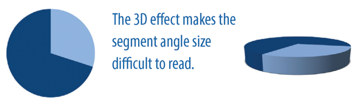
- Be aware that colour blindness reduces the ability to distinguish certain colours (most commonly red and green) and tonal variation. It affects around 7% of the male population.

visual effects and 'chart junk'

A guiding principle should be to keep your presentation clean, simple and informative.

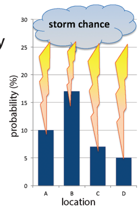
Chart junk (Edward Tufte's term for unnecessary decorative additions) can distract the reader from the data and patterns you want them to observe.

- 3D effects can skew the perception of values shown by bar height or length in histograms and impair perception of angles in pie charts.



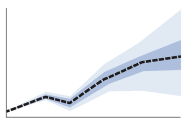
The 3D effect makes the segment angle size difficult to read.

- Illustrations and ornaments added to data presentations may distort or distract attention from your data.



The decorations here distract from accurate reading of the bar heights and might suggest there is more lightning in some locations.

Design and editorial choices can direct the focus of the reader. Consider what you are trying to communicate when showing summary values (e.g. median lines), as readers might fixate on these at the expense of the rest of the data. Whether or not this is desirable will depend on the context.



The median line on this fan chart might be seen as a specific answer in the middle of the uncertainty, potentially distracting attention from the more extreme options.

Figure 1. A reproduction of the leaflet front and back. The final print version was printed on A4 paper and folded in three.

There is a difference between data presentation that allows viewers to draw their own conclusions (where appropriate) and lapses in communication fundamentals that leave space for, or unwittingly encourage, mistaken interpretations and invalid conclusions. For example, a 3D projection of a pie chart with perspective 'effects' applied can make it difficult for readers to accurately assess the angle of a segment or make accurate comparisons between segments.

Weaknesses in contextualisation can occur at several different levels. In the most basic examples contextualisation might be limited by missing labels or units on axes, or titles that are either absent or lacking in detail. Mistakes at this basic level can leave readers struggling to work out what a graph is trying to show at all. On a wider scale, poor captioning and integration of the chart/graph figure into the main text can cause confusion, especially if the text does not directly support the figure. In some cases, poor communication results from inconsistencies or clear errors in presentation, but in others there may be a system of internal logic that is not transparent to the reader. That is, when taken in isolation a specific visualisation (for example) has a reasoning and sense behind its organisation, but this internal sense is then at odds with the wider context or conventions of interpretation and understanding. Similar effects can be seen when a discipline or group has developed its own conventions (use of colour, terminology, assumptions) that appear immediately comprehensible to its own members, but opaque to other readers. Public facing weather forecasts, for example, are a strongly conventionalised area, with approaches to presentation and public interpretation varying by region and nationality. Gigerenzer *et al*. (2005) demonstrated this variation, showing differences in interpretation and assumed context for a basic forecast, *30% chance of rain tomorrow*, in New York compared to four European cities.

Responding to context

Our investigation (albeit informal) suggested a need for a quick reference guide on data and uncertainty presentation, with a focus on awareness of good fundamentals and a 'light', easily retrievable, approach to increase the likelihood of potential users assimilating the key messages. Guidance would need to be relatively general, as specific details vary with the context, even within atmospheric science communication. We therefore aimed to produce a short leaflet that would:

– *raise awareness of the needs of some audiences to be supported in understanding data presented to them, particularly in the presence of uncertainty, and*

– *highlight approaches to presentation that would improve data transparency and, conversely, those that can hinder interpretation.*

The small format of a leaflet, while useful for quick reference, imposed a hard limit on how much could be communicated. The WMO two-page summary guide concentrates on *why* it is useful to communicate uncertainty information, followed by some examples of uncertainty data/graphics. We aimed to give more direct advice.

The content of the leaflet is focused around a list of core topics. The list was based on the research and teaching experience of our meteorologist team members and insights from project researchers in psychology and design. This experience includes teaching at undergraduate and postgraduate levels in the three disciplines, and, outside of meteorology, expertise in perceptual psychology, cognition and user-centred information design. Copy and illustrations to cover the topic list were devised and compiled by our designer team members into a draft leaflet for feedback.

Survey feedback on leaflet draft

A short survey was conducted ahead of full publication to check if the leaflet (in draft form) was relevant and useful for its intended purpose. The survey was distributed via Survey Monkey to PURE Network members (including representatives from insurance companies, government agencies, energy companies, first responders, and the aviation industry) and academic staff and postgraduate researchers in the Department of Meteorology at the University of Reading. Twenty-nine responses were returned, with responses from both academic- and industry-based recipients.

As the leaflet dealt with both general issues in data presentation and specific issues relating to the presentation of uncertainty, most of the survey questions had two parts, addressing 'data' and 'uncertainty' presentation as distinct points. Overall, respondents did not feel that the leaflet introduced them to *new* aspects of presenting data or uncertainty (only 7 and 9 respondents, respectively, agreeing that it introduced new aspects). However, a majority (21 of the 29 respondents for both data and uncertainty questions) felt that the leaflet *reminded* them of aspects of which they were already aware, but did not always implement in their own work. The benefits of the guide in collating reminders of basic good practice were indicated by 13 respondents, who agreed that as a result of the leaflet they would make changes to the way they presented their own data. Similarly, 13 respondents agreed that they would change the way they reviewed the data presentation of others.

In response to a question regarding the potential audience for such a leaflet, 25 of the 29 respondents thought it would be useful for trainee research scientists and undergraduates, and 20 for experienced research scientists. Recommendations for audiences outside of academia dropped to around half of respondents in other key areas (14 recommending it for civil servants, and journalists and broadcasters, 15 for industry partners). This bias towards potential for academic users might reflect the research background of many of the respondents, or that the respondents felt that academic users were more likely than others to create data presentations.

An open-ended question (*any other comments*) yielded detailed suggestions for reduction of some topics (specifically on colour use) and augmentation of others (for example on the use of captions, labels and accompanying text to provide the full context for data presentation), to which we responded in a revision of the draft leaflet. Two respondents thought that the leaflet content was too basic to be helpful. Nonetheless, although the leaflet does not engage in detailed or advanced specifics, our initial research suggested that basics are often forgotten, as indicated by responses to the survey. Across a group of volunteer respondents, we would expect a range of perspectives and approaches to data presentation.

Implementation and access

The aim of the leaflet was to reinforce good practice at a fundamental level, such as the need for clear and consistent labelling and contextualisation, in order to support some audiences in interpreting atmospheric science data. Our survey indicated that although many people have been exposed to the underlying principles of presenting data and uncertainty, there is value in providing a concise and easily digestible summary that will serve as an *aide memoire* for day-to-day use.

The fundamentals of clear communication and data presentation are of renewed importance at a time when digital technology offers almost unlimited options for the presentation of information. With the wealth of options available, it can be easy to overlook basic factors that still have a central impact on comprehension. In some cases, it may be necessary or beneficial to further support understanding of data presented graphically or in text with discussion and face-to-face communication. While this leaflet does not focus on spoken communication, the same principals of clarity and contextualisation are relevant. The leaflet provides quick-reference reminders

RMetS
Royal Meteorological Society

of these fundamentals and some specifics for the more specialist area of uncertainty communication.

The leaflet is available to download from www.reading.ac.uk/web/files/infodesign/presenting-data-and-uncertainty-quick-guide.pdf.

Acknowledgements

This research was funded as part of the PURE (Probability, Uncertainty & Risk in the Environment) research programme, funded by the UK Natural Environment Research Council, grant number NE/J017221/1.

References

Bostrom A, Anselin L, Farris J. 2008. Visualizing seismic risk and uncertainty, a review of related research. *Ann. N.Y. Acad. Sci.* **1128**: 29–40.

Brodlie K, Osorio RA, Lopes A. 2012. A review of uncertainty in data visualization, in *Expanding the Frontiers of Visual Analytics and Visualization*. Dill J, Earnshaw R, Kasik D *et al.* (eds). Springer: London, pp 81–109.

Cheong L, Bleisch S, Kealy A *et al.* 2016. Evaluating the impact of visualization of wildfire hazard upon decision-making under uncertainty. *Int. J. Geographic Inf. Sci.* **30**(7): 1377–1404.

Gigerenzer G, Hertwig R, van den Broek E *et al.* 2005. "A 30% chance of rain tomorrow": how does the public understand probabilistic weather forecasts? *Risk Anal.* **25**(3): 623–629.

MacEachren AM, Robinson A, Hopper S *et al.* 2005. Visualizing geospatial information uncertainty: what we know and what we need to know. *Cartogr. Geographic Inf. Sci.* **32**(3): 139–160.

Mulder KJ, Lickiss M, Harvey NJ *et al.* 2017. Visualizing volcanic ash forecasts: scientist and stakeholder decisions using different graphical representations and conflicting forecasts. *Weather, Clim. Soc.* **9**: 333–348.

Pang AT, Wittenbrink CM, Lodha SK. 1997. Approaches to uncertainty visualization. *Visual Comput.* **13**(8): 370–390.

PBL Netherlands Environmental Assessment Agency. 2013. *Guide for Uncertainty Communication*. PBL Netherlands Environment Agency: The Hague, the Netherlands. URL: www.pbl.nl/en/publications/guide-for-uncertainty-communication.

Petersen AC, Janssen PHM, van der Sluijs JP *et al.* 2013. *Guidance for Uncertainty Assessment and Communication*, 2nd Edition. PBL Netherlands Environment Agency: The Hague, the Netherlands. URL: www.pbl.nl/en/publications/guidance-for-uncertainty-assessment-and-communication.

Spiegelhalter D, Pearson M, Short I. 2011. Visualizing uncertainty about the future. *Science* **333**(6048): 1393–1400.

Wardekker JA, Kloprogge P, Petersen AC *et al.* 2013. *Guide for Uncertainty Communication*. PBL Netherlands Environment Agency: The Hague, the Netherlands. URL: http://www.pbl.nl/sites/default/files/cms/publicaties/PBL_2013_Guide-for-uncertainty-communication_1339.pdf.

WMO. 2008. *Guidelines on Communicating Forecast Uncertainty* (WMO/TD No. 1422, English version). World Meteorological Organization: Geneva, Switzerland. URL: www.wmo.int/pages/prog/amp/pwsp/publicationsguidelines_en.htm.

WMO. 2010. *Communicating Forecast Uncertainty* (WMO/PWS–SG 1). World Meteorological Organization: Geneva, Switzerland. URL: www.wmo.int/pages/prog/amp/pwsp/publications-guidelines_en.htm.

Correspondence to: Matthew Lickiss

jj907016@reading.ac.uk

doi:10.1002/wea.2998