Meteorologische Zeitschrift, Vol. 27, No. 6, 435–440 (published online August 2, 2018)

7IVMW

The WMO Challenge to Develop and Demonstrate the Best New User-Oriented Forecast Verification Metric

Elizabeth Ebert[1*], Barbara Brown[2], Martin Göber[3], Thomas Haiden[4], Marion Mittermaier[5], Pertti Nurmi[6], Laurie Wilson[7], Sarah Jackson[8], Peter Johnston[9] and Dieter Schuster[10]

[1]Bureau of Meteorology, Melbourne, Australia
[2]National Center for Atmospheric Research, Boulder, Colorado, USA
[3]Hans Ertel Center, Deutscher Wetterdienst, Berlin, Germany
[4]European Centre for Medium-Range Weather Forecasts, Reading, UK
[5]United Kingdom Meteorological Office, Exeter, UK
[6]Finnish Meteorological Institute, Helsinki, Finland
[7]Environment and Climate Change Canada, Montreal, Canada, retired
[8]United Kingdom Meteorological Office, London, UK
[9]University of Cape Town, Cape Town, South Africa
[10]EnBW Energie Baden-Württemberg, Karlsruhe, Germany

(Manuscript received November 24, 2017; in revised form April 22, 2018; accepted June 7, 2018)

Abstract

The public, industry, emergency managers and other decision makers can use weather, climate and impact forecasts more effectively in their decision making when the quality of forecasts is measured in terms that are meaningful and comprehensible to them. To encourage the development of user-oriented verification approaches and support the major projects of the World Meteorological Organization (WMO) World Weather Research Program, a challenge was issued to develop and demonstrate new user-oriented forecast verification metrics. Several new forecast verification metrics were proposed to meet the needs of very different user-communities including public safety and amenity, shipping, aviation, defence, agriculture, and water resources. A few general purpose metrics were also proposed. The winner of the inaugural verification challenge proposed a new metric called the Spatial Probability Score for assessing forecasts for the location of a relevant boundary or contour, for example, sea ice edge or extent of flood inundation. We hope and expect that many of the user-oriented forecast verification metrics submitted to the inaugural verification challenge will be taken up by the broader community.

Keywords: forecast verification, user-oriented, challenge, metrics

1 Introduction

Weather and climate forecasts inform many important decisions such as ensuring public safety during dangerous weather, aviation planning and hazard avoidance, farm management, or simply whether to carry an umbrella or wear a coat. As weather and climate forecasts continue to expand into a greater range of products and to improve in accuracy and specificity they are increasingly being used as input to critical decisions, either directly or as input to models for floods, fire spread, water balance, crop growth, renewable energy supply, airport management, and many other applications.

Good forecasts save thousands of lives and hundreds of millions of dollars in avoided damages every year (e.g. Lazo et al., 2009; WMO et al., 2015). Determining the value of forecasts depends on knowing their accuracy (measured through forecast verification which assesses the accuracy through comparison with what actually happened), the degree to which forecasts can potentially lead to a better outcome (not all forecasts are equally useful, for example, daily forecasts of "no rain" during the dry season), and the ability of the user to use the forecast information to make a better decision (Murphy, 1993). When decision makers trust the forecasts they are much more likely to use them in their decision making. Forecast verification is therefore an integral part of the value chain that links weather forecasting to societal benefit (e.g. Morss et al., 2008).

What kind of information can decision makers access in order to develop trust in forecasts? Brown (2006) described a set of levels of user focus for forecast verification. Conventional measures-based approaches can be considered as *Level 0*, comprising summary scores like root mean square error, correlation coefficient, mean bias, and the like. These are typically used for basic administrative verification to monitor performance over time, and to compare whether one forecast performs better than another. *Level 1* verification is also based on scores, but contains diagnostic information that enables deeper investigation of forecast error. It may evaluate

*Corresponding author: Elisabeth Ebert, Bureau of Meteorology, GPO Box 1289, Melbourne VIC 3001, Australia, e-mail: beth.ebert@bom.gov.au

DOI 10.1127/metz/2018/0892
Gebrüder Borntraeger Science Publishers, Stuttgart, www.borntraeger-cramer.com

several variables of interest to the user and offer performance information for particular stratifications (e.g. temperature during clear nights) or thresholds of the data (e.g. rainfall of 25 mm d^{-1} or more). Often diagnostic verification is presented graphically, and ideally it contains information on the robustness of the verification (confidence intervals and *p*-values are examples of such information). Most forecast verification available to users is *Level 0* or *Level 1*.

The last two decades have seen the development of more intuitive, features-based and enhanced diagnostic approaches (*Level 2*), which attempt to mimic the way humans subjectively evaluate forecast performance when comparing forecast and analysed weather maps or features in time series, for example. These approaches measure the ability of forecasts to predict multiple attributes such as location, size, and intensity of weather features, or modes of variability, and may offer insight on the sources of forecast error. *Level 2* verification evaluates the variables that were directly predicted, and is especially well suited to "super users" of verification information such as operational forecasters or developers of numerical forecast models.

Level 3 verification tailors performance information to the needs of specific users and sectors. An oft-quoted example is the accuracy of predicting wind power ramp events (sudden changes in wind speed that can damage wind turbines) for the renewable energy industry. To be effective the providers of forecast performance information must interact closely with the users to understand their decision process and jointly determine meaningful approaches and measures. Often specialised datasets are required from the users in order to assess the forecast quantities of interest. When forecast performance information can be linked to actual decisions to quantify the economic or societal benefit, then the highest level of user-focused verification (*Level 4*) is achieved for those stakeholders. This level of verification is often highly bespoke, using proprietary (not necessarily meteorological) observations.

Understanding the users' business takes effort, but the value to both sides in providing the most applicable information is substantial. Clearly when forecast performance is reported in terms that are most meaningful to the users (*Levels 3* and *4*) then those users are equipped with information to make better decisions based on the forecasts.

Developing better user-relevant verification was identified by Ebert et al. (2013) as one of the major challenges in forecast verification. To date few metrics exist specifically to measure forecast quality in user-relevant terms. A good example, developed by the Met Office in the United Kingdom is the Flight Time Error. This measure, developed in consultation with the aviation industry, measures the difference between the actual flight time taken by an aircraft and the time estimated to "fly" the plane along the same path through the predicted wind field (Rickard et al., 2001). Errors in arrival time are meaningful to airlines, which can calculate the costs associated with touching down early or late.

To encourage the meteorological community to think about verification from the user perspective and develop metrics that users would find most helpful, a challenge was issued by the World Meteorological Organization (WMO) Joint Working Group on Forecast Verification Research (JWGFVR)[1] to develop and demonstrate the best new user-oriented forecast verification metric. The goal was that new verification methodologies proposed via the verification challenge ("the Verification Challenge") could be adopted by the global meteorological community, benefitting both forecast providers and forecast users. The outcomes of the Verification Challenge will also support the research and service aims of WMO, in particular the WWRP High Impact Weather, Sub-seasonal to Seasonal Prediction, and Polar Prediction Projects, all of which have major user components.

2 The Verification Challenge

The Verification Challenge was launched in mid-September 2015 and ran for 14 months to allow adequate time for participants to develop and demonstrate their new metrics. As used here a "metric" refers to a system or standard of measurement, and could be a quantitative score or diagnostic (e.g. diagram).

The rules of the Verification Challenge were simple. All applications of meteorological and hydrological forecasts that are relevant to user sectors such as agriculture, energy, emergency management, transport, etc. were eligible. To be considered for the prize the verification metric had to be new, that is, unpublished or only newly published, and certainly not in wide usage. Metrics with a clear statistical foundation and those which could be applied to a broader set of problems were especially encouraged.

The entries were assessed according to six key criteria, each of which was scored on a scale of 1 to 5 (Table 1). The first two criteria were considered more important and hence assigned twice the weight of the other four criteria.

The judging panel consisted of all but one of the authors of this paper, comprising members from the JWGFVR, the WWRP Societal and Economic Research and Applications (SERA) working group, and three user representatives from the aviation, energy, and agriculture sectors.

Participants submitted their entries to the Verification Challenge using a template that requested the following information:

[1]The JWGFVR promotes verification best practice through conducting verification methods workshops and tutorials, facilitating the development and application of improved diagnostic verification methods, encouraging greater awareness of the importance of verification as a vital part of numerical and field experiments, and encouraging verification practitioners to take the needs of users into account to ensure the forecast verification is relevant (JWGFVR, 2017). The JWGFVR also plans and implements the verification component of the World Weather Research Program (WWRP).

Table 1: Judging criteria for proposed user-oriented verification metrics.

Criterion	Scoring	Weight
1. Originality	1 (nearly the same as an existing score) to 5 (really great new idea)	×2
2. User relevance	1 (not of interest to user) to 5 (really helps users make decisions)	×2
3. Intuitiveness	1 (hard to understand) to 5 (very easy to understand)	×1
4. Simplicity & ease of computing	1 (really complicated) to 5 (easy to compute)	×1
5. Robustness	1 (poorly behaved, does strange things) to 5 (well behaved, does no strange things). Bonus point for confidence estimates.	×1
6. Resistance to hedging	1 (easy to hedge) to 5 (impossible to hedge)	×1

1. Motivation / problem to be solved
Briefly describe the forecast users and indicate how they use the forecast in their decision making. Identify which aspects of the forecasts are important to predict accurately.

2. Development of the new metric
Describe the metric (including its name) and what it measures. Briefly describe any consultation with users, if done. For new statistical scores please include a mathematical derivation.

3. Demonstration of the new metric
Provide an example of the use of the new metric. Briefly describe the forecasts and observations used and show one or more examples of the verification results. Describe how a user should interpret the results for this case, and more generally for the new metric.

4. Advantages of the new metric
Describe ways in which the new metric is better than other approaches currently used to provide forecast quality information. Refer to the judging criteria (originality, user relevance, intuitiveness, simplicity and ease of computing, robustness, and resistance to hedging[2]) where appropriate.

5. Any other information
Please provide any other information relevant to the use and acceptance of the new metric, e.g. user endorsement.

Entries were limited to five pages to facilitate the judging process.

The advertised prize for the winner of the Verification Challenge was an all-expense paid trip to present a keynote talk on his/her metric at the 7th International Verification Methods Workshop in Berlin in May 2017. All participants were encouraged to present their work at the workshop. This workshop, organized by the JWGFVR every two to three years, is the premier international gathering for scientists conducting research on forecast verification methods, developers of verification systems, and keen verification users in the meteorological community. Therefore the prize conferred to the winner some considerable standing in the verification community.

The Verification Challenge was advertised widely through promotion at conferences, newsletter articles, e-mail posts to participants of past verification workshops and other potentially interested colleagues and universities, and on the WMO and JWGFVR websites.

3 Outcome of the Verification Challenge

In total 17 entries were submitted to the Challenge, originating from 11 countries: Australia, Canada, China, Germany, India, Italy, Norway, Sweden, UK, Ukraine, and the USA. We considered this quite a good response, given that this was the first time a verification challenge had been offered.

The application areas addressed by the new verification methods were many, and included public safety and amenity, shipping, aviation, defence, agriculture, water resources, as well as a few general purpose metrics. Interestingly, a key weather dependent sector that was not addressed was renewable energy. A few entries promoted increased usage of existing metrics ("oldies but goodies"), but the majority of entries proposed new metrics targeted to assist decision making by users in particular sectors.

Each judge rated the entries independently according to the criteria in Table 1. His or her weighted total score (*WTS*) was computed as $WTS = 2S_1 + 2S_2 + S_3 + S_4 + S_5 + S_6$, where S_n is the score for criterion n. To avoid any conflict of interest judges did not give scores to entries coming from their home institutions. The weighted total scores from all judges were averaged and ranked to determine the ultimate winner. One judge from the user community felt able to assess only the user relevance and intuitiveness, so a second approach was taken whereby the average score for each criterion was first computed, then the weighted total scores computed from the average criterion scores and ranked. This produced the same ranking as the first approach. A final check of the judges' top-ranked entries confirmed the results.

Constructive feedback was sent to each entrant in a separate communication following the announcement of

[2]"Hedging" in the context of forecast verification, means forecasting something other than one's true belief in order to get a better verification score (Jolliffe, 2008). Some scores encourage hedging, for example, the probability of detection can be maximised by always forecasting an event.

the Verification Challenge results. As noted earlier, all Verification Challenge participants were encouraged to present their work at the 7th International Verification Methods Workshop. Several have contributed a paper on their work to this special issue of *Meteorologische Zeitschrift* on forecast verification.

The remainder of this section briefly describes the winning entries in the inaugural Verification Challenge.

a. 1st prize – Integrated Ice Edge Error (IIEE) & Spatial Probability Score (SPS)

Users of sea ice forecasts need to understand the quality of predictions for the ice edge in order to plan for safe operations, whether it be naval, fishing, or emergency response. Helge Goessling (Alfred Wegener Institute) developed a new Spatial Probability Score (SPS) that measures the agreement between a probabilistic forecast *field* and a binary observed *field*. In essence this is the continuous ranked probability score (CRPS) applied spatially, where the dimension of the continuous physical quantity has been replaced by one or more spatial dimensions.

The SPS has a number of nice properties: the results are given in meaningful units (e.g. km^2 for two-dimensional spatial forecasts), it can be applied in principle to forecasts with any number of dimensions, and it is resistant to hedging so cannot be "played" to achieve a better score. When the 2D spatial forecast is deterministic rather than probabilistic then the SPS gives the non-overlap area between the forecast and observations. In the case of sea ice this is identical to the Integrated Ice Edge Error (IIEE; Goessling et al., 2016). Probabilistic and deterministic forecasts can therefore be compared cleanly using the SPS. For 2D fields dividing by the mean distance in one dimension gives the mean distance in the other dimension; in the case of sea ice dividing SPS by the ice edge length gives the effective position error of the forecast ice edge.

The SPS can in principle be applied to any probabilistic field for which binary observations on the same grid exist, and it is relatively cheap to compute. The SPS is described in greater detail by Goessling and Jung (2018). The talk on SPS given by Helge Goessling was one of the highlights of the 7th International Verification Methods Workshop.

b. 2nd prize – Spatio-Temporal User-Centric Distance

Outdoor event planners, facility managers, air traffic controllers and civil authorities, as well as the general public, need information on the impending threat of thunderstorms moving over their location, including their spatial extent and timing. Forecasts and nowcasts provide such information, but how trustworthy are they? Dominique Brunet (Environment and Climate Change Canada) proposed a new metric called the Spatio-Temporal User-Centric Distance that compares the distance from a location of interest to a forecast area and the distance from the same location to a set of observations for each time step of the forecast. The forecast and observations need not be of the same type or on the same grid so long as they represent the phenomenon of interest. For example, the forecast could be radar-based thunderstorm nowcasts while the observations could be lightning detections from a lightning network.

The distances are computed using a generalised distance transform (GDT), and provide important information on how close the forecasted and observed events are to the user at each time. The GDT is more resistant to outliers than some other commonly used distance metrics like the Hausdorff distance, and has one parameter that controls the search radius for events. The time lag between the two time series that minimises their mean absolute difference gives an estimate of the timing error of the forecast. The difference between the (temporally corrected) forecast and observed distances to the user can be examined for over- and under-forecasting (akin to false alarms and misses). The distance and time units (kilometres and hours) are completely intuitive, making this a user-friendly metric for experts and non-experts alike. More information on the Spatio-Temporal User-Centric Distance can be found in the paper by Brunet (2018).

c. Equal 3rd prize – Weighted Percent Consistence

Seasonal outlooks for probabilities of rainfall and maximum and minimum temperature above or below median help users in the agriculture, emergency services, tourism, and other sectors plan their activities in coming months. However, describing the skill of such outlooks is problematic as the usual verification metrics for probability forecasts are hard to communicate and difficult for non-experts to interpret. Therefore a simpler metric, percent consistent (PC) is used, with PC $>$ 50 % indicating a "good" forecast. However, by converting the probabilistic forecast to a binary forecast (yes/no to over the median) PC no longer differentiates between forecasts that are significantly good (bad) and those that are only marginally good (bad). William Wang, Andrew Watkins and David Jones (Australian Bureau of Meteorology) modified the PC by weighting the yes/no score (0 or 1) at each point by the relative strength of the anomaly to derive a weighted percent consistence (WPC). The anomaly normalisation can be temporal (i.e. time series at a grid point), spatial (i.e. all grid points in the spatial domain), or both, depending on the type of assessment required.

Weighting the score by the magnitude of the anomaly has the effect of rewarding the good high impact forecasts more, but also penalising the bad high impact forecasts more. Integrated over space and/or time, the WPC better differentiates the truly good forecasts and truly bad forecasts compared to PC. When applied to seasonal forecasts of above-median rainfall probability forecasts in Australia, one industry user commented, "Yes – we always knew that something like WPC was needed – it's the divergence from the 'normal' season that is more important … [and] drives decision making." Wang et al. (2018) provide more information on the WPC score.

d. Equal 3rd prize – Rain-Free Window Accuracy

For some applications it is important to verify predictions of good weather rather than bad, when a task requires a certain amount or period of time to complete and would be seriously hampered by adverse conditions. For example, people living in rainy climates who cycle to and from work often consult radar-based nowcasts to find a rain-free window during which to travel. Knowing the accuracy of "good weather" forecasts would assist people to place appropriate reliance on them. Thomas Nipen and Ivar Seierstad (Norwegian Meteorological Institute) developed a verification approach that focuses on window lengths and can be easily understood by the general public.

The approach has three steps: (1) Remove from the verification sample all cases where no rain was observed in the 3 hours leading up to the forecast start time (to focus the verification on cases when people would actually consult the forecasts), (2) compute the forecast and observed rain-free window lengths, (3) compute the threat score for the exceedance of different window lengths and plot/tabulate performance as a function of window length. Knowing the threat score of the forecasts for the rain-free window required to conduct the desired activity (e.g. 30 minutes to ride home) tells the user how much to trust those specific forecasts. Forecasts for short rain-free windows are typically more accurate than forecasts for longer rain-free windows. If the user has a particular error tolerance (e.g. wrong no more than 20 % of the time) then the maximum rain-free window length that still meets that criterion can be determined. The method can be applied not only to rain, but to any variable seen as a threat. See the paper by Nipen and Seierstad (2018) for further details and results for short-term rain-free window forecasts in Norway.

4 Discussion

The World Meteorological Organization and a variety of end users have emphasised the need for more user-relevant metrics, particularly as weather information is propagated into user impact information through coupled modelling approaches (e.g. Ebert et al., 2015). The Verification Challenge has had the desired effect of raising the profile of user-oriented verification in the meteorological community. In fact, several of the entries were submitted by verification practitioners in the climate, ocean and water communities, reflecting the need for user-oriented verification in all forecasting disciplines as well as the growing linkages between these related communities. The encouraging number of entries (17) gives us confidence that user-focused verification is a topic of significant interest, and the quality of the entries highlights that there is considerable creativity in the community to develop better user-oriented forecast verification. We hope and expect that many of the user-oriented forecast verification metrics submitted to the Verification Challenge will be taken up by the broader community, moving us closer to the desired *Level 3* and *Level 4* user-focused verification that most effectively helps end users take forecast performance into account in their decision making.

This was the first time a challenge to develop user-focused verification methods has been issued. The JWGFVR has decided to sponsor another Verification Challenge in the lead-up to the next International Verification Methods Workshop in 2020. While keeping the focus on user-relevant verification, some changes may be desirable.

It is difficult to conceive of how else to assess the "best" metric except subjectively according to agreed criteria. The JWGFVR agreed the criteria in Table 1 when the Verification Challenge was formulated, but it is worth reconsidering them in light of our experience. In most cases we found it difficult to assess the "hedgeability" of the metrics. However, as this is an important property the next Verification Challenge should explain this concept more clearly in the call for entries and ask the entrants to support their claim about it. There may be other aspects of metrics that should be rewarded. For example, users may value a metric's ability to directly measure the spatial and/or time scales at which forecasts have an acceptable level of skill (the skilful scale associated with the Fractions Skill Score (Roberts and Lean, 2008) is one example). Other aspects of user relevance could be explored such as the value of improved decision making supported by the verification, where "user testimonies" could add weight to the entry.

The composition of the judging panel should include both verification experts and representatives from the user community. Finding users who would like to participate is challenging because by definition they come from specific sectors and often do not feel qualified to assess the applicability of metrics intended for other sectors. It is not possible to know in advance which sectors the proposed metrics will address, nor likely that user representatives from all of those sectors could (or should) be recruited. It would be possible to recruit the judging panel after the entries have been submitted, and ask the entrants to clearly state their target user community or communities. For the Verification Challenge reported here we believe that backgrounds of the "user judges" did not significantly bias the selection of the winning entries.

User judges did not find it easy to judge criteria such as robustness and resilience to hedging, and whether a metric was in fact new. The next Verification Challenge should consider whether user judges should be required to assess all criteria or rather focus on user relevance and intuitiveness. If the decision is that they should judge all criteria, then JWGFVR could provide some relevant training.

The requirement that a metric be "new" in order to win the Verification Challenge should be revisited. Existing scores that are well known and demonstrably useful can be applied in new ways to provide benefit for users, and have the advantage that their properties are already known. The next Verification Challenge might invite the application of existing verification to new or transformed variables of relevance to the user. In this case the mathematical link connecting the new metric to the established score should be provided.

We welcome feedback from participants and interested readers on additional ways to improve the Verification Challenge in the future.

Acknowledgements

The authors would like to thank all of the participants of the Verification Challenge, and the WMO for their support of the challenge including publicizing it to the community.

References

Brown, B., 2006: User-relevant verification. – North American THORPEX Societal and Economic Research and Applications (NAT SERA) Workshop, 14–16 August 2006, Boulder, Colorado, USA.

Brunet, D., 2018: A Spatial-Temporal User-Centric Distance. – Meteorol. Z., this issue.

Ebert, E., L. Wilson, A. Weigel, M. Mittermaier, P. Nurmi, P. Gill, M. Göber, S. Joslyn, B. Brown, T. Fowler, A. Watkins, 2013: Progress and challenges in forecast verification. – Meteor. Appl. **20**, 129–255. DOI:10.1002/met.1392.

Ebert, E., B. Brown, J. Chen, C. Coelho, M. Dorninger, M. Goeber, T. Haiden, M. Mittermaier, P. Nurmi, L. Wilson, Y. Zhu, 2015: Numerical prediction of the Earth System: Cross-cutting research on verification techniques. – In Brunet, G., S. Jones and P.M. Ruti (eds), Seamless Prediction of the Earth System: from Minutes to Months. WMO-No. **1156**, (ISBN 978-92-63-11156-2), Geneva.

Goessling, H., T. Jung, 2018: A probabilistic verification score for contours: Methodology and application to Arctic ice-edge forecasts. – Quart. J. Royal Meteor. Soc., published online. https://doi.org/10.1002/qj.3242.

Goessling, H.F., S. Tietsche, J.J. Day, E. Hawkins, T. Jung, 2016. Predictability of the Arctic sea ice edge. – Geophys. Res. Lett. **43**, 1642–1650.

Jolliffe, I.T., 2008: The impenetrable hedge: A note on propriety, equitability and consistency. Meteor. Appl., **15**, 25–29.

JWGFVR, 2017: Forecast Verification Research. – http://www.wmo.int/pages/prog/arep/wwrp/new/Forecast_Verification.html.

Lazo, J.K., R.E. Morss, J.L. Demuth, 2009: Sources, perceptions, uses and values of weather forecasts. – Bull. Amer. Meteor. Soc. **90**, 785–798, DOI:10.1175/2008BAMS2604.1.

Morss, R.E., J.K. Lazo, B.G. Brown, H.E. Brooks, P.T. Ganderton, B.N. Mills, 2008: Societal and economic research and applications for weather forecasts: Priorities for the North American THORPEX program. – Bull. Amer. Meteor. Soc. **89**, 335–346.

Murphy, A.H., 1993: What is a good forecast? An essay on the nature of goodness in weather forecasting. – Wea. Forecast. **8**, 281–293.

Nipen, T., I. Seierstad, 2018: Rain-free window accuracy. – Meteorol. Z., this issue.

Rickard, G.J., R.W. Lunnon, J. Tenenbaum, 2001: The Met Office upper air winds: Prediction and verification in the context of commercial aviation data. – Meteor. Appl. **8**, 351–360.

Roberts, N.M., H.W. Lean, 2008: Scale-selective verification of rainfall accumulations from high-resolution forecasts of convective events. – Mon. Wea. Rev. **136**, 78–97.

Wang, W., A. Watkins, D. Jones, 2018: Weighted percent consistence. – Meteorol. Z., this issue.

WMO, WBG, GFDRR, USAID, 2015: Valuing Weather and Climate: Economic Assessment of Meteorological and Hydrological Services. – World Meteorological Organization, World Bank Group, Global Facility for Disaster Reduction and Recovery, and United States Agency for International Development, WMO No. **1153**, Geneva, Switzerland. Available online: https://www.gfdrr.org/sites/default/files/publication/SEB%20HYDROMET.pdf.