On the guarantees derived from a possibilistic interpretation of ensemble predictions and their operational use

Noémie Le Carrer — noemie.lecarrer@unive.it

June 3, 2021

* Institute for Risk and Uncertainty, Dept. of Engineering, University of Liverpool (UK)
* Università Ca'Foscari Venezia, Dep. of Environmental Sciences, Informatics and Statistics (Italy)

Table of contents

Introduction

Ensemble Prediction Systems (EPS)

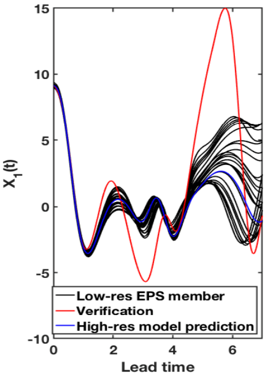
Operational issues for prediction/ predictability:

Atmosphere-Ocean is a chaotic system (high sensitivity to initial cond. -ICs)

Low vertical/horizontal grid resolution (imperfect -ICs)

Model error (closure assumptions, etc)

Solution developed since 2000s:



1. Perturbation of ICs around best estimate (e.g. sample fastest growing perturbations)

2. Forward propagation, at **lower time-resolution**, in a possibly **perturbed model** (stochastic parameterization schemes)

3. Set of M (~10-100) predictions: ensemble members.

The primacy of doubt: Evolution of numerical weather prediction from determinism to probability

Tim Palmer[1]

[1]Department of Physics, University of Oxford, Oxford, UK

Towards the probabilistic Earth-system simulator: a vision for the future of climate and weather prediction†

T. N. Palmer

Limitations

- The ‘ensemble prediction’ paradigm assumes that **model error** is negligible w.r.t. initial condition error [11, 1], which is shown not to be true after a few hours [17, 15].
- High dimensionality: **how to sample randomly**, in a Monte-Carlo-like fashion, with limited computational budget? Hence non-random initial member perturbations (e.g. bred vectors, singular vectors).
- **Extreme events** (EE) generally result from nonlinear interactions at small scale, which makes them hardly detectable in a probabilistic interpretation of ensemble forecasts [14].
- Predictive probabilities are **not actionable** [16], i.e. cannot be used to make profitable bets. They are numbers that bear the name probability but are not really.

Existing solutions

Ensemble forecasts can rarely be interpreted as a random sample from the predictive distribution of future atmospheric states, because of **dispersion errors** and **biases**.

Statistical postprocessing is required to achieve calibration [7].

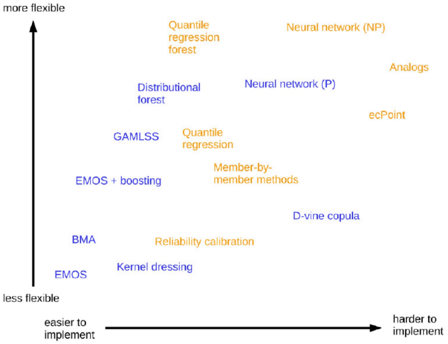


Figure 2: From *Statistical postprocessing for weather forecasts – Review, Challenges and Avenues in a Big Data World*, Vannitsmen et al. (2020) [18].

… and their limitations.

- Generic post-processing methodologies improve the ensemble skills for common events and extend the skillful prediction horizon.
- However they tend to deteriorate performances for **extreme events** [11, 14].
- The latter must be addressed by specific, non-global methodologies (e.g. quantile regression, nonstationary Poisson process) for variable characterised by heavy climatology tails [6].

Research questions

- Given that tailored post-processing steps are necessary to get (more or less) meaningful probabilities, echoing [1], we wonder whether the probability distribution is the **best representation of the valuable information** contained in an **EPS**.
- Rather, the description of **possibility theory** in [3]: *"a weaker theory than probability (...) also relevant in non-probabilistic settings where additivity no longer makes sense and not only as a weak substitute for additive uncertainty measures"* presents new opportunities, in a context where both conceptual and practical limitations restrict the applicability of a density-based (i.e. additive) interpretation of EPS.

We consequently investigated the questions:

1. Can a possibilistic treatment of the EPS provide more information than a probabilistic interpretation?
2. If we get more guarantees, at what cost?

Test bed & information at hand

The Lorenz 96 system

A surrogate model of the atmospheric dynamics, the deterministic *L96* system [12].

A documented imperfect version of the model is used to generate EPSs.

Perfect model

$$\frac{dX_j}{dt} = X_{j-1}(X_{j+1} - X_{j-2}) - X_j + F - \frac{hc}{b}\sum_{k=1}^{K} Y_{j,k}$$

$$\frac{dY_{j,k}}{dt} = cbY_{j,k+1}(Y_{j,k-1} - Y_{j,k+2}) - cY_{j,k} + \frac{hc}{b}X_j$$

Imperfect model

$$0.262 - 1.262X_j + 0.004608X_j^2 + 0.007496X_j^3 - 0.0003226X_j^4$$

Following [Williams 2014, Wilks 2006]:
J= 1, ...J, J=8
K=1, ...K, K=32
h=1
b=c=10
F=20

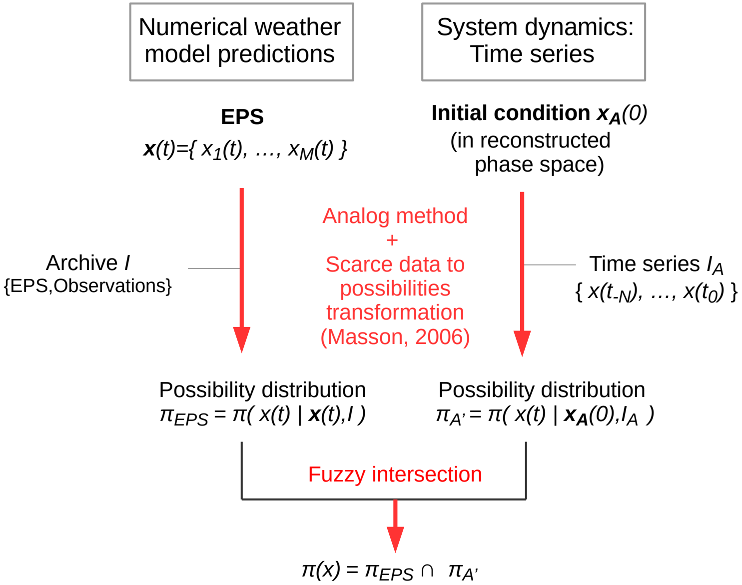
EPS members are issued from M random initial perturbations $N(X_j, 0.1^2)$ evolved forward by means of the imperfect coupled equations (following [Williams 2014]).

Information at hand to make a prediction

- **Monovariate perspective**: we consider independently each dimension of the model space. $x \in \mathbb{R}$ refers to the component of interest of system $\mathcal{S}$.
- We call indistinctly **truth, verification** or **observation** the true state of the system.
- Elements of information at hand for prediction-making:
    1. A set of M **predictions**, the EPS at lead time t: $\tilde{\mathbf{x}}(t) = \{\tilde{x}_1(t), ..., \tilde{x}_M(t)\}$.
    2. An **archive** I containing N pairs **(EPS, verification)** for the lead time of interest
    3. A **time series** I_A of (preferably continuous) past observations of x, containing the initial condition $x(t_0)$ of interest.

Methodological framework

Framework: Global overview



Framework: Making predictions

$$\pi(x) = \pi_{EPS} \cap \pi_{A'}$$

Possibilistic measures for the prediction of a binary event A
$N(A), \Pi(A)$

Predictions

Possibility-Probability equivalence (Dubois, 2004):

$$N(A) \leq P(A) \leq \Pi(A)$$

- Reliability diagram
- Log-likelihood score

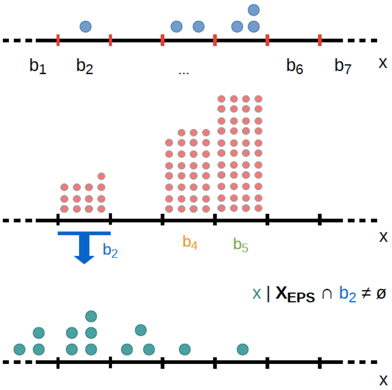
Validation scores; Comparison to probabilistic or deterministic predictive models

Credibility (Liu, 2006):

$$C(A) = 0.5.(N(A)+\Pi(A))$$

- ROC, PRC
- Correlation coefficient, mutual information

Framework: EPS interpretation (a)

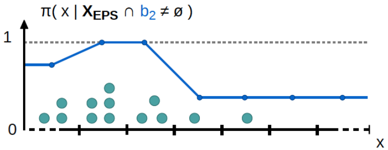


1. The axis is binned and EPS members are placed in the bins.

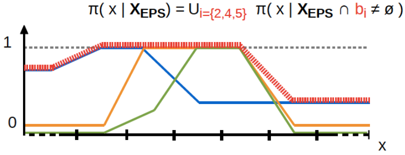
2. For each bin b_i occupied by at least one member of the EPS $\boldsymbol{X_{EPS}}$, we collect the EPS members of the archive I_t that fell in the same bin at that same lead time t.

3. For each occupied bin b_i, we collect the verifications associated with the above subset of archived EPS members and place them in the bins over the axis.

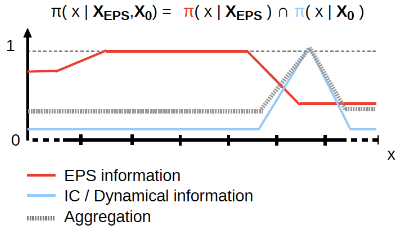
Framework: EPS interpretation (b)



4 We compute from this set of N_s analogs the possibility distribution describing the system state x at lead time t, given that a member of $\boldsymbol{X_{EPS}}$ has fallen in bin b_j.



5 The possibility distribution for the system state at a given lead time, given the EPS, is the union (i.e. envelope) of the possibility distributions associated to each occupied bin.



6 To take into account the initial conditions $\boldsymbol{X_0}$ (IC) and local dynamics of the system, we intersect this possibility distribution with a possibility distribution based only on ICs, possibly expanded through delay embedding if we dispose of a long enough record of the system.

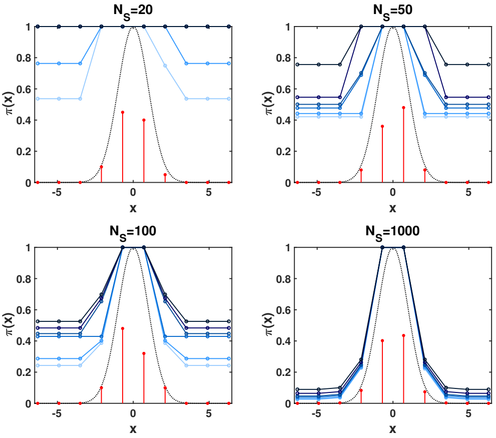
Inferring π from data

Masson & Denoeux (2006) [13] :

- We consider a single variable x defined on universe S, that is binned into n disjoint classes $\{b_i, i = 1, .., n\}$. We assume that the data x has been randomly generated from an unknown probability distribution P.
- The **simultaneous confidence intervals** on the n probabilities $p_i = P(x \in b_i)$ are computed for multinomial proportions by means of the **Goodman formulation** (1965 [8]) at confidence level β.
- From these confidence intervals and considerations about the probability-possibility transformation, eq. (1), the procedure allows to compute a possibility distribution $\{\pi(x \in b_i), i = 1...n\}$, that dominates with confidence β the true probability distribution P (i.e. eq. (1) is verified with confidence β).

In practice

- N_s samples are drawn from $X \sim N(0,1)$ (black-dotted **PDF**, histogram in red)
- $\beta \in \{0.6, 0.75, 0.9, 0.95, 0.99, 1\}$ (the larger the darker blue)
- $n = 10$ bins

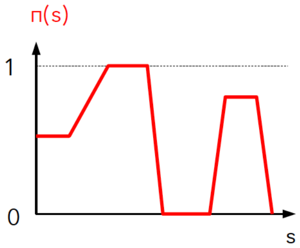


Experiments & Results

Basics: Possibility distribution

- Theory developed from fuzzy set theory to handle **incomplete knowledge** and represent **ignorance** [19, 4, 5, 3].
- Let S be an exhaustive set of the existing states s of the system of interest. The possibility distribution $\pi : S \to [0, 1]$ represents the state of knowledge about the state of the system:

- $\pi(s) = 0$: the state s is rejected (i.e. s impossible)
- $\pi(s) = 1$: the state s totally possible
- Intermediate values are used to represent incomplete knowledge on the system state



Basics: Possibility and Necessity measures

Does the event $A : s \in S_A$, where S_A is a subset of S, occur?

To answer, we compute the degrees of:

- **Possibility Π**

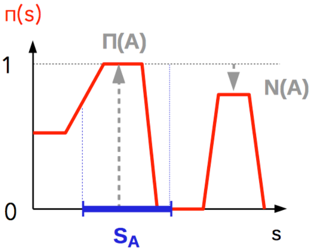
  (to what A is consistent with the knowledge at hand):

  $$\Pi(A) = \sup_{s \in S_A} \pi(s)$$

- **Necessity N**

  (to what extent A is necessarily implied by π):

  $$N(A) = \inf_{s \notin S_A} 1 - \pi(s) = 1 - \Pi(\bar{A})$$



Basics: Axioms and Probabilistic equivalence

1. $\Pi(\mathcal{S}) = 1$ and $\Pi(\varnothing) = 0$
2. $\Pi(A \cup B) = \max\big(\Pi(A), \Pi(B)\big)$
3. $N(A) = 1 \Leftrightarrow \Pi(\bar{A}) = 0$ indicates that A has to happen, it is necessary: $\bar{A}$ is impossible;
4. $0 < N(A) < 1$ is a tentative acceptance of A to a degree $N(A)$;
5. $\big(\Pi(A) = \Pi(\bar{A}) = 1\big) \Leftrightarrow \big(N(A) = N(\bar{A}) = 0\big)$ represents total ignorance: the evidence doesn't allow us to conclude whether A is rather true or false.

Probability measure P and possibility measure Π are **consistent** [2, 3] iff:

$$N(A) \leq P(A) \leq \Pi(A), \forall A \tag{1}$$

Basics: Continuous interpretation of π

From (1), a possibility density π is consistent with the associated probability measure P iff its $\alpha-$cuts $C_\pi^\alpha = \{x,\ \pi(x) \geq \alpha\}$ satisfy:

$$P(x \in C_\pi^\alpha) = P(C_\pi^\alpha) \geq 1 - \alpha \ , \ \forall\, \alpha \in [0, 1]. \tag{2}$$

The possibility distribution satisfying this criterion is **not unique**. The choice is driven by the principle of **maximum specificity** [3]. Maximum specificity w.r.t. the probabilistic information (*a priori* unknown) is achieved when the possibility distribution is probabilistically calibrated :

$$P(C_\pi^\alpha) = 1 - \alpha \ , \ \forall \alpha \in [0, 1]. \tag{3}$$

Each $\alpha-$cut then represents a **frequentist confidence interval** at level $1 - \alpha$ for the variable of interest.

Basics: Deriving confidence intervals from π

- By construction, the individual possibility distributions $\pi(x_t|\tilde{x}_t^m \in b_j)$ verify Equation (1) with a guaranteed confidence level β. π_{EPS} being made of their envelope, it cannot be more specific than any single one of them and consequently the same guarantee applies. In the case of its $\alpha-$cuts, this reads:

$$P\Big(P(x_t \in C_\pi^\alpha) \geq 1-\alpha\Big) \geq \beta. \tag{4}$$

- The data-to-possibility transformation [13] is shown to be rather conservative and provides a possibility distribution that actually dominates the true probability distribution with a rate much higher than the guaranteed β. Even for small sample sizes, the choice of β is not critical and quasi perfect coverage rate is obtained: $\beta \geq 0.8$, ensures that $P\Big(P(x \in C_\pi^\alpha) \geq 1-\alpha\Big) \rightarrow 1$.
- Under this assumption, the $(1-\alpha)$-cuts can be used as candidate confidence intervals of *guaranteed* level α. Ideally, we are looking for $(1-\alpha)$-cuts verifying Equation (3), which ensures optimal specificity of π_{EPS} and thus maximally informative confidence intervals.

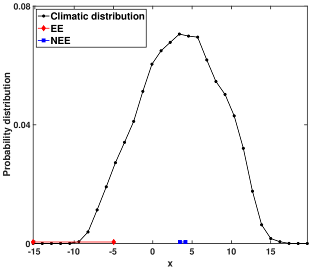
Experimental setting

We compare the performances of the **confidence intervals** I^α at level α, extracted from the methodologies:

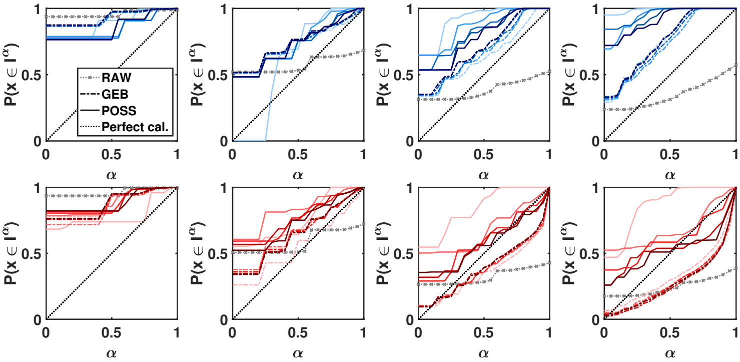
- **POSS** (our analog-based possibilistic interpretation of EPSs) ;
- **GEB** (probability distribution derived from the EPS by means of Gaussian Ensemble Dressing) ;
- **RAW** (probability distribution derived directly from the EPS, without post-processing).

Events of interest:

- The "extreme" event (EE): $x \leq x_5$ (quantile 5 of the climatological distribution) ;
- and the "common" event (NEE): $x_50 \leq x \leq x_55$.

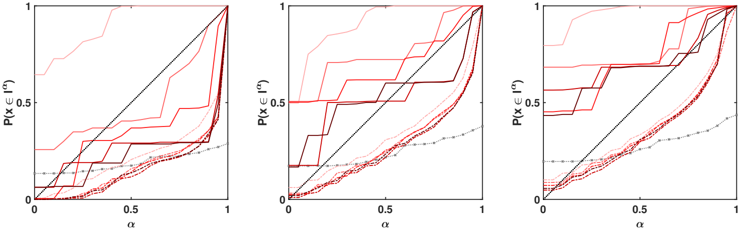


Empirical assessment of formal guarantees: Limitation of very large archives



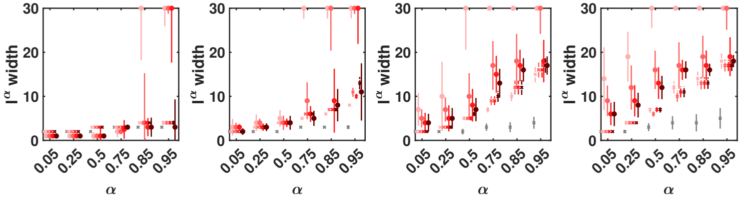
Provided that N_l is not too large (which we assume is always the case for operational archives), these results clearly show the added value of treating the EPS in a possibilistic manner in terms of **guarantees for the EE at large lead times**, or in terms of **reliability for the NEE at very small lead times**.

Empirical assessment of formal guarantees - Case of extreme events



- Coverage probability of the $(1-\alpha)$-cuts of π_{EPS} at lead time $t = 7$ days, in the case of events belonging to a partition of subsets of EE (from left to right: $x \leq q_1$, $q_1 < x \leq q_3$ and $q_3 < x \leq q_5$). The EPS archive size varies: $N_l \in \{156, 1560, 5e3, 15e3, 30e3\}$ (the larger the darker the line).
- Probabilistic calibration for the more extreme parts of EE can be improved by increasing β as long as the archive is not too large.

Interval precision



For EE, the added value of POSS over GEB is observed on two occasions:

1. intervals are as reliable yet narrower for very small lead times and $\alpha < 0.9$, whatever the archive size ;
2. for large lead times and intermediary-sized archives ($N_l \in \{1560, 5e3\}$), possibility-based confidence intervals are both guaranteed, reliable and operational (i.e. not too wide compared to GEB's results, contrary to what $N_l = 156$ produces), while the probabilistic intervals are narrower yet not guaranteed at all.

Conclusion

Conclusions (1)

- We introduced a novel framework to interpret EPSs where a possibility distribution π_{EPS} is derived from the EPS at hand and an archive of $\big($EPS; verification$\big)$. We showed how to use the $(1-\alpha)$-cuts of a continuous interpretation of π_{EPS} to produce confidence intervals at level α about the future value of the variable of interest.
- Our possibility-based confidence intervals come with formal guarantees, and experimental results show that they overpass probability-based ones in two situations:
    1. at very small lead times for both NEE and EE, where they are as reliable yet narrower;
    2. at intermediate and large lead times for EE, where they remain guaranteed and can be brought close to perfect reliability even for particularly rare events, yet at the expense of precision. These results can be reached with operational archive like the $20-30$-year reforecast datasets. The guarantees are retained for smaller archives, which however lead to more conservative intervals and thereby impede operationality.

Conclusions (2)

- Possibility theory is a promising tool for the prediction of extreme events, given a limited and imperfect amount of information on the system's dynamics.
- Beyond the results presented in this article [9], further developments [10] show how π_{EPS} can be combined with additional possibility distributions constructed from alternative sources of information such as the IC or dynamical information (see step 6 of Figure 13).
- Therein, the concept of ignorance is developed and presented as an interesting tool for risk communication in weather forecasting.

Questions?

J. Bröcker and L. A. Smith.
From ensemble forecasts to predictive distribution functions.
Tellus A: Dynamic Meteorology and Oceanography, 60(4):663–678, 2008.

M. Delgado and S. Moral.
On the concept of possibility-probability consistency.
Fuzzy Sets and Systems, 21(3):311–318, 1987.

D. Dubois, L. Foulloy, G. Mauris, and H. Prade.
Probability-Possibility Transformations, Triangular Fuzzy Sets, and Probabilistic Inequalities.
Reliable computing, 10(4):273–297, 2004.

D. Dubois and H. Prade.
On several representations of uncertain body of evidence.
Fuzzy Information and Decision Processes, pages 167–181, 1982.

D. Dubois, H. Prade, and S. Sandri.
On Possibility/Probability Transformations.
In R. Lowen and M. Roubens, editors, *Fuzzy Logic: State of the Art*, pages 103–112. Springer Netherlands, Dordrecht, 1993.

P. Friederichs, S. Wahl, and S. Buschow.
Postprocessing for Extreme Events.
In S. Vannitsem, D. S. Wilks, and J. W. Messner, editors, *Statistical Postprocessing of Ensemble Forecasts*, pages 127–154. Elsevier, 2018.

T. Gneiting.
Probabilistic forecasting.
Journal of the Royal Statistical Society. Series A (Statistics in Society), pages 319–321, 2008.

L. A. Goodman.
On simultaneous confidence intervals for multinomial proportions.
Technometrics, 7(2):247–254, 1965.

N. Le Carrer.
Possibly extreme, probably not: Is possibility theory the route for risk-averse decision-making?
Atmospheric Science Letters, page e01030, 2021.

N. Le Carrer and S. Ferson.
Beyond probabilities: A possibilistic framework to interpret ensemble predictions and fuse imperfect sources of information, 2020.
Under review.

T. P. Legg and K. R. Mylne.
Early Warnings of Severe Weather from Ensemble Forecast Information.
Weather and Forecasting, 19(5):891–906, 2004.

E. N. Lorenz.
Predictability: A problem partly solved.
In *Proc. Seminar on predictability*, volume 1, 1996.

M.-H. Masson and T. Denœux.
Inferring a possibility distribution from empirical data.
Fuzzy Sets and Systems, 157(3):319–340, 2006.

K. Mylne, C. Woolcock, J. Denholm-Price, and R. Darvell.
Operational calibrated probability forecasts from the ECMWF ensemble prediction system: implementation and verification.
In *Preprints of the Symposium on Observations, Data Asimmilation and Probabilistic Prediction*, pages 113–118, 2002.

D. Orrell.
Ensemble Forecasting in a System with Model Error.
Journal of the Atmospheric Sciences, 62(5):1652–1659, 2005.

L. A. Smith.
Integrating Information, Misinformation and Desire: Improved Weather-Risk Management for the Energy Sector.
In P. J. Aston, A. J. Mulholland, and K. M. Tant, editors, *UK Success Stories in Industrial Mathematics*, pages 289–296. Springer International Publishing, Cham, 2016.

Z. Toth and E. Kalnay.
Ensemble Forecasting at NCEP and the Breeding Method.
Monthly Weather Review, 125(12):3297–3319, 1997.

S. Vannitsem, J. B. Bremnes, J. Demaeyer, G. R. Evans, J. Flowerdew, S. Hemri, S. Lerch, N. Roberts, S. Theis, A. Atencia, et al.
Statistical postprocessing for weather forecasts–review, challenges and avenues in a big data world.
arXiv preprint arXiv:2004.06582, 2020.

L. Zadeh.
Fuzzy sets as a basis for a theory of possibility.
Fuzzy Sets and Systems, 1(1):3–28, 1978.