08/22

ARBEIDSNOTAT
WORKING PAPER

The wisdom in crowd sourced weather forecasts

Michael Morreau
Keita Abe
Margrethe Aanesen

SNF

Samfunns- og næringslivsforskning AS
Centre for Applied Research at NHH

SNF
SAMFUNNS- OG NÆRINGSLIVSFORSKNING AS

- er et selskap i NHH-miljøet med oppgave å initiere, organisere og utføre ekstern-finansiert forskning. Norges Handelshøyskole og Stiftelsen SNF er aksjonærer. Virksomheten drives med basis i egen stab og fagmiljøene ved NHH.

SNF er ett av Norges ledende forskningsmiljø innen anvendt økonomisk-administrativ forskning, og har gode samarbeidsrelasjoner til andre forskningsmiljøer i Norge og utlandet. SNF utfører forskning og forskningsbaserte utredninger for sentrale beslutningstakere i privat og offentlig sektor. Forskningen organiseres i program-mer og prosjekter av langsiktig og mer kortsiktig karakter. Alle publikasjoner er offentlig tilgjengelig.

SNF
CENTRE FOR APPLIED RESEARCH AT NHH

- is a company within the NHH group. Its objective is to initiate, organize and conduct externally financed research. The company shareholders are the Norwegian School of Economics (NHH) and the SNF Foundation. Research is carried out by SNF's own staff as well as faculty members at NHH.

SNF is one of Norway's leading research environment within applied economic administrative research. It has excellent working relations with other research environments in Norway as well as abroad. SNF conducts research and prepares research-based reports for major decision-makers both in the private and the public sector. Research is organized in programmes and projects on a long-term as well as a short-term basis. All our publications are publicly available.

SNF Working Paper No. 08/22

The wisdom in crowd sourced weather forecasts

Michael Morreau
Keita Abe
Margrethe Aanesen

SNF Project No. 10068: Climate Futures

The project is financed by the Research Council of Norway

CENTRE FOR APPLIED RESEARCH AT NHH
BERGEN, NOVEMBER 2022
ISSN 2704-0380 (Online)

© This copy has been drawn up by agreement with Kopinor (www.kopinor.no). The production of further copies without agreement and in contravention of the Copyright Act is a punishable offence and may result in liability to compensation.

The wisdom in crowd sourced weather forecasts

Michael Morreau

UiT – The Arctic University of Norway, Tromsø

SNF, Norwegian School of Economics, Bergen

michael.morreau@uit.no

Keita Abe

Musashi University, Tokyo

Margrethe Aanesen

SNF, Norwegian School of Economics, Bergen

Abstract

Weather forecasts were elicited from a representative sample of citizens living in Bergen, Norway. The elicited forecasts exhibit wisdom of the crowd: averages of forecasts by individuals have a greater forecast skill than, on average, these individual forecasts. The forecast skill of the aggregate forecasts compares favorably with that of forecasts published by the Norwegian Meteorological Institute and the Norwegian Broadcasting Corporation (NRK). The individual forecasts do not appear to be copied from the published forecasts nor do they seem to be mere statements of base rates for the relevant kinds of weather.

Keywords: Wisdom of the crowd, weather forecasting

1. Introduction

Floods, landslides, wildfires and other catastrophic events impose great costs on society. Local governments, businesses and private citizens in order to plan and prepare need to know what they might expect, but many emergencies are difficult to predict. One way to build awareness of the risks is to keep track of contributing factors such as heavy rain, high temperatures and strong winds.

National Meteorological and Hydrological Services (NMHSs) provide weather forecasting services to the public. Even with their state-of-the art methods and models though, there scope for improvement in the accuracy and resolution of forecasts. It is of interest therefore to have additional sources of information about weather – especially ones that do not draw too heavily on the limited resources of those in need of it. We investigate here the idea that ordinary citizens can be such a source.

In the course of their daily lives, people receive many signals about weather and other factors exposing their societies to risk. Presumably, they are privy to relevant knowledge that is not presently taken into account. For instance, local people might have been able to alert authorities to an abundance of combustible grasses that was a factor in the wildfire on Sotra island off the coast of Norway, in the summer of 2021. Aggregating results obtained using different weather models and weather forecasting methods has long been known to increase the accuracy of weather forecasts [1]. Our hope is that aggregating forecasts provided by ordinary people can similarly increase accuracy — perhaps even to the point that crowd sourcing can become a useful resource for emergency planning and preparedness. Local governments might utilize this resource for instance by using an app running on mobile phones to gather inputs from citizens, whether to improve on the accuracy or resolution of weather forecasts by NMHSs, or to forecast factors other than weather that contribute to catastrophic events.

There is reason to be optimistic about the possibilities. Alteady several centuries ago, the Marquis of Condorcet, a prominent figure in the French Enlightenment, discovered that a majority of voters in an assembly is under certain conditions more likely to be correct than are the individual members [3]. Condorcet's celebrated "jury theorem" tells us that majorities of lay people can have an even greater chance of judging correctly than smaller groups of experts [4]. At the beginning of the twentieth century, Sir Francis Galton, one of the founders of statistics, conducted a famous experiment using different people's estimates of the weight of an ox, obtained from a competition at a county fair. He observed that both the mean and the median of individual estimates were more accurate than, on average, the individual estimates [5]. Meanwhile, this "wisdom of the crowd" is widely recognized to have practical importance for management decisions throughout society [9, 8].

The main question here is whether weather forecasts by ordinary citizens exhibit a useful wisdom of the crowd. It has two aspects. First, are aggregates

of forecasts of ordinary citizens strictly more accurate than the individual forecasts? Second, if they are, just how accurate are they? How does their accuracy compare, say, with that of forecasts obtained with the models and methods currently used by national forecasting organizations?

To answer this question, we conducted an experiment. A representative group of several hundred citizens from the city of Bergen in Norway were asked at the beginning of the week to forecast precipitation and temperatures for the following weekend. Their forecasts were aggregated by averaging them, and forecast skill was assessed both for the aggregate and the individual forecasts. Thus tentative answers could be given to both parts of our research question. We found, in summary, that the citizen forecasts did indeed exhibit wisdom of the crowd. Furthermore, the accuracy of the aggregate citizen forecasts from the survey compared favourably to that of forecasts of the same events published on *Yr* (*www.yr.no*), a popular weather site supported by the Norwegian Meteorological Institute and the Norwegian Broadcasting Corporation (NRK).

This report explains the experiment and critically examines its results. Section 2 briefly describes the experiment itself, the method for aggregating individual forecasts, and the measures of forecast skill used to analyze the data we obtained. Section 3 summarizes the results. These are consistent with our suggestion that citizen forecasts can usefully augment official forecasts. Section 4 introduces and then tentatively dismisses two hypotheses that tend to undermine this suggestion: that respondents have merely copied official forecasts, and that they have merely forecast base rates. Section 5 discusses limitations of the present work. Section 6 suggests promising directions for future work, and appendix 7 has details of the method used to elicit citizen forecasts.

2. Method

During three weeks in July 2022, the polling company Norstat elicited weather forecasts from a representative sample of citizens living in the city of Bergen, in Norway. Each week, on Monday, Tuesday and Wednesday, members of Norstat's citizen panel were asked to forecast precipitation and temperatures for 12:00 noon the following Saturday, in the city center.[1] In the first

[1]The Bergen city center was specified more precisely to be the area, familiar to local people, surrounding the weather station *Bergen-Florida*.

week we received 268 responses to the poll and in the second week 311. In the third week, Norstat provided a final 22 responses to make up the contracted 600 responses (the actual total is 601).

The precipitation forecasts in the poll are probabilistic. That is, people were asked to say how likely they thought it was to rain. Probabilities were elicited in two ways, first using a qualitative Likert scale and then quantitatively, on the scale $0 - 100\%$. Temperature forecasts for the same time and place were single-valued, with responses on the scale $0 - 100°C$. For further details of the elicitation of probabilities in the poll, see the appendix (section 7).

The poll included in addition to the precipitation and temperature forecasts also a test question unrelated to weather. People were asked to estimate the area of the upper surface of "den blå steinen", a familiar art work in the Bergen city center. See figure 7. This question was included for diagnostic purposes. Absence of wisdom of the crowd in people's answers to the test question would suggest a problem with the poll as an elicitation method.[2]

2.1. Collective forecasts

Forecasts by individual citizens were aggregated by averaging them. Thus, for any one of the three weeks in which the experiment ran, the aggregate probability forecast for rain at noon on a Saturday is the (arithmetic) mean of all individual probabilities for rain, obtained on the Monday, Tuesday and Wednesday of that week. The aggregate temperature forecast for noon on Saturday similarly is the mean of all individual temperature forecasts. Sometimes we call the aggregate forecasts "collective" or "crowd" forecasts.

2.2. Accuracy

Forecast skill of probabilistic forecasts (precipitation) was assessed as the Brier score [2]. Forecast skill was assessed for single-valued forecasts (temperature) as absolute error. The precipitation and temperature observations used for skill assessment were those posted on *yr.no* for the Bergen Florida weather station at noon on the forecast Saturday.

[2]Estimates of readily observable physical magnitudes such as the sizes and weights of objects generally exhibit wisdom of the crowd.

2.3. The accuracy bench mark

To assess the accuracy of aggregate citizen forecasts, we took as our benchmark precipitation and temperature forecasts published on *yr.no*. Specifically, we used forecasts published at noon of the Wednesday of each week. Recall that the aggregate citizen forecasts are averages for that Wednesday together with the preceding Monday and Tuesday. Thus many of the citizen forecasts were elicited a day or two before the benchmark forecast. We expect that if this difference prejudices the comparison in any way then it tends to give an advantage to the benchmark. This is simply because the accuracy of forecasts tends to increase as the date of the forecast events approaches.

3. Results

Now we answer both parts of our main research question. Section 3.1 considers whether citizen forecasts in the experiment exhibit wisdom of the crowd. Section 3.2 compares the forecast skill of aggregate citizen forecasts with that of benchmark forecasts from *yr.no*.

3.1. Is the crowd weather wise?

The elicited citizen forecasts of both precipitation and temperature exhibit wisdom of the crowd. That is, the aggregate of individual forecasts, formed by averaging them, has a strictly greater forecast skill than, on average, the individual forecasts making up the aggregate.[3]

Figure 1 gives the general picture for the whole experiment, by averaging the forecast skill for aggregate and individual forecasts over all three weeks in which the experiment ran. Figure 2 depicts just the wisdom of the crowd for precipitation forecasts, but for each of the three weeks separately. We see that every week there was wisdom of the crowd in the precipitation forecasts. Figure 3 similarly breaks down wisdom of the crowd in temperature forecasts for the three weeks. There was wisdom of the crowd in the temperature forecasts each week as well.

We now turn to the second aspect of our research question by comparing the accuracy of aggregate citizen forecasts with that of benchmark forecasts.

[3]That the forecast skill of the aggregate is at least as great as that of the individual forecasts is "baked into" the experimental set up, as a direct consequence of Jensen's inequality ([7], §9.1.3), since the Brier score and the absolute error are convex functions. That the forecast skill of the aggregate forecast is *strictly* greater, and by how much it is greater, are on the other hand empirical matters.

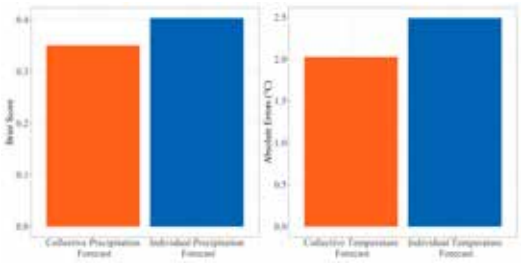

Figure 1: The Brier score for collective precipitation forecasts, averaged over the three weeks, is lower than the mean Brier score of the individual forecasts (left panel). The absolute error for collective temperature forecasts, averaged over the three weeks, is lower than the mean absolute error of the individual forecasts (right panel).

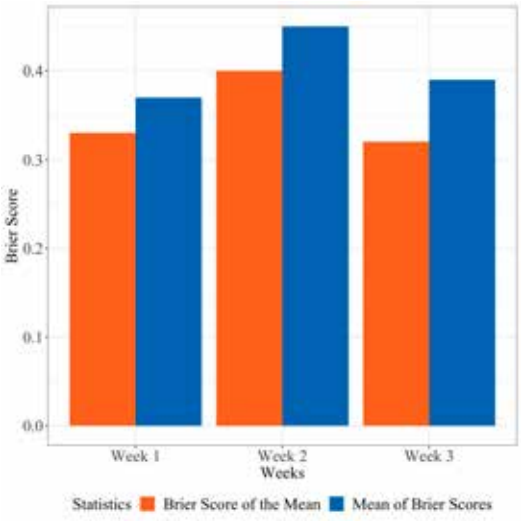

Figure 2: Wisdom of the crowd in precipitation forecasts for each week. The Brier score of the mean is consistently better (lower) than the mean Brier score.

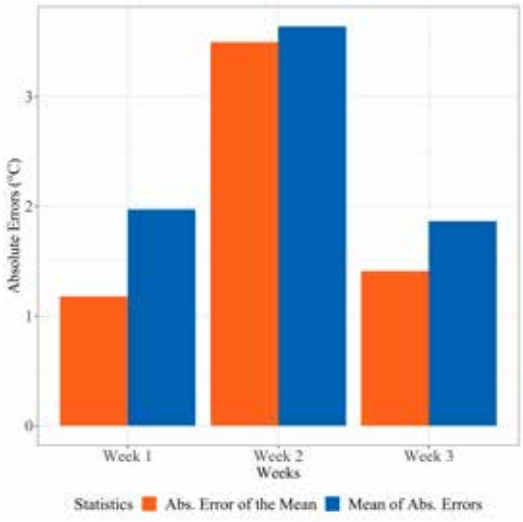

Figure 3: Wisdom of the crowd in temperature forecasts for each week. The absolute error of the mean is in each case strictly lower than the mean absolute error.

3.2. How weather wise is the crowd?

Subsection 3.1 presents our finding that the weather forecasts elicited from citizens of Bergen display wisdom of Crowds: aggregating forecasts from many individuals resulted in increased accuracy. While this might be found interesting in itself, it doesn't tell us whether the crowdsourced forecasts might be at all useful. That depends on, among other things, just how accurate they are. While our study was not large enough to draw any firm conclusions, in this section we will see that our results are compatible with the idea that the accuracy of the aggregated citizen forecasts is comparable to that of benchmark forecasts from a national weather forecasting service.

Figure 4 puts the forecast of skill of aggregate citizen forecasts next to the forecast skill of forecasts from *yr.no*. We do not see in these preliminary results any evidence that the aggregate citizen forecasts are either systematically more accurate or systematically less accurate than the bench mark forecasts. In the first two weeks of the experiment, anyway, the crowd sourced and benchmark precipitation forecasts appear to be about equally good.[4]

[4]In the third week we received only a few tens of responses, so results from that week should be considered less reliable than those from the first two weeks.

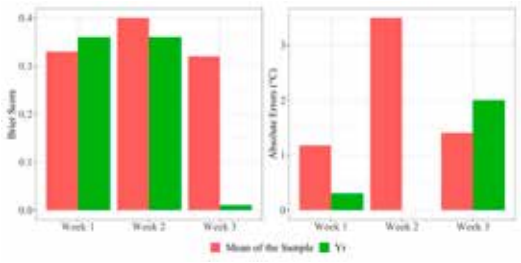

Figure 4: The forecast skill of collective citizen weather forecasts compared with the forecast skill of midweek forecasts from *Yr*.

The crowd's temperature forecasts seem to be quite a bit worse than those of *Yr*. We suggest that this might be due to people's wishful thinking biasing their temperature estimates upwards (in Bergen, and elsewhere in Norway, many people hope for warmer weather).

4. Discussion

Our results appear to be compatible with the idea that aggregation can make citizen forecasts into a useful resource. In this section, we discuss and tentatively set aside two possibilities that cast doubt on this idea: the copying hypothesis and the base-rate hypothesis.

4.1. The copying hypothesis

Suppose citizens responded to the poll by copying down forecasts they have looked up on *Yr* or a similar forecasting service.[5] Suppose also that respondents copy noisily, introducing random errors up and down. This copying hypothesis explains the observed wisdom of the crowd, since averaging the individual forecasts will tend to cancel random errors. It explains also rough parity with benchmark forecasts, of which the crowd sourced forecasts are *ex hypothesi* an imperfect reflection.

Truth of the copying hypothesis drains all interest from the results reported here. If the people's forecasts are just noisy copies of existing weather forecasts then they do not add information that can add to these or improve them.

[5]Both probabilistic precipitation forecasts and single-valued temperature forecasts one week in advance (and longer) are published on *Yr*.

The copying hypothesis, though, seems unlikely. There are two reasons for this. The first is that while single-valued temperature forecasts are easily looked up on *Yr*, accessing probabilistic precipitation forecasts is much more difficult. Having reached the main page for the weather station Bergen Florida, one still has to click through into "details" to find them. It seems unlikely to us that many respondents will have succeeded in this or, indeed, with just a few minutes allocated for the whole survey, even will have tried.

Our second reason to doubt the copying hypothesis is that the citizen forecasts from the poll seem to be much more varied than could be expected on the basis of copying errors. Figure 5 depicts citizen forecasts from the second week of the experiment together with contemporaneous forecasts from *Yr*. Notice that the citizen forecasts are much more variable than the forecasts from *Yr*. Assuming as seems reasonable that people completing the survey at different times have about the same propensity to copying errors, this variability is at odds with the copying hypothesis.

4.2. The base-rate hypothesis

Another possibility that would tend make our results much less interesting is that respondents are just forecasting base rates for precipitation and historical average temperatures. That might result in respectable Brier scores and absolute errors. Respondents in this case would know something important about the weather in Bergen. They would know the relevant base rates and historical averages. Their knowledge would be of little interest for our purpose, though, since it would not entail any responsiveness to the weather from one day or week to the next. People would come with roughly the same forecasts every time.

To test the base-rate hypothesis, we calculated a precipitation base rate using historical data obtained from the website *seklima.met.no*, another service provided by the Norwegian Meteorological Institute. Using observations from 2012-2022, we arrived at the precipitation base rate 0.27.[6] Suppose people had simply forecast this base rate. That is, suppose the individual precipitation forecasts were to distribute noisily around the mean value 0.27. Then the crowd's Brier score for the three weeks of the experiment would be 0.07.[7] In fact, the crowd's score was much worse, consistently well above 0.3 (compare figures 1 and 4). Accordingly, on the basis of our limited experiment, the base-rate hypothesis seems unlikely to us. In the three weeks of the experiment, people would indeed have achieved better Brier scores and compared more favorably with the forecasts from *Yr* if they had simply forecast the base rate instead of doing whatever it was that they actually did.

Another way to approach this matter is to consider the discrimination of the forecasts, which is a measure of the extent to which they depart from historical trends.[8] Other things (measures of forecast skill) being equal, better discrimination means a better forecast.

[6]That is, on 27% of Saturdays at noon during the last ten years it was rainy at noon at the Bergen Florida weather station. In order to arrive at this base rate using tools readily available at *seklima.met.no*, we counted it as rainy if there had been rain at any time in the preceding hour. This method tends to overstate the base rate. Be this as it may, rain in an immediately preceding period is perhaps what ordinary people experience as its being rainy at any given moment. If so, then this way of arriving at a base rate seems suitable for our present purpose.

[7]Since there was no precipitation at noon on any of the three weeks of the experiment, the Brier score obtained by predicting the base rate 0.27 is just $0.07 \approx \frac{1}{3} \times 3 \times (0.27 - 0)^2$.

[8]The *discrimination* of probabilistic forecasts is:

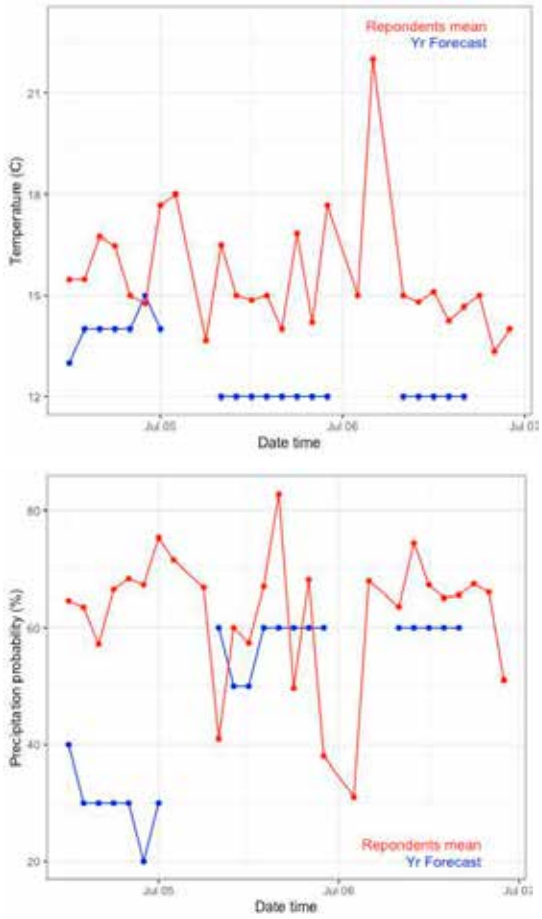

Figure 5: Average citizen forecasts for precipitation and temperature in the second week of the experiment and the benchmark forecasts from *Yr*. Citizen forecasts appear to vary more from one time to the next than do the benchmark forecasts. This suggests that they are not simply the results of copying.

The discrimination of the crowd's precipitation forecasts using our base rate is 0.10. This is comfortably greater than 0, the discrimination score that would be achieved by forecasting the base rate. Since the maximum achievable discrimination score is 0.53 (obtained by predicting precipitation with probability 1) the observed figure for the crowd seems to be at least comparable to the discrimination of the forecasts from *Yr* for the same three weeks, which we reckon to be 0.07. Our data do not suggest that the people are simply forecasting a base rate.

We stress that the discussion of this section is very tentative, in that the computations are made on the basis just three forecasts by the crowd and three taken from *Yr* (one for each week in which the experiment ran).

5. Limitations

The main limitation of this study is due to its small size. In each of the three weeks for which our experiment ran, the crowd makes a single precipitation forecast and a single temperature forecast. The two crowd forecasts from the third week of the experiment are based on a much smaller number of citizen forecasts than those from the first and second weeks, and might therefore be regarded as a less faithful reflection of collective opinion. Setting aside these two, we just have four crowd forecasts to go by, run in consecutive weeks when weather conditions were much the same. As described in section 3, these four aggregate forecasts display a forecast skill that is comparable to that of *Yr*. With so little data to go by, though, we cannot confidently attribute this to the collective forecasting ability of citizens. For all we know at present, it might be down to chance.

A further important limitation of our study is that we have considered just two kinds of weather: precipitation and temperature. Other weather variables such as wind strength and direction are implicated in fires, floods and other catastrophic events. Even assuming that our results concerning precipitation and temperature forecasts hold up in future studies, it remains

---

$$\frac{1}{N}\sum_{t=1}^{T} n_t(c_t - c)^2 \tag{1}$$

Here, $c_1, \ldots c_T$ are the forecast probabilities for some event, n_t is the number of forecasts with probability c_t, and c is the base rate for this event.

to be seen whether the crowd is any good at forecasting other important weather events.

6. Future work

Bearing in mind the small size of the present study, a first priority is to run an experiment like this one but for a longer period, and at different times of the year. In this way, we can hope to arrive at a more confident assessment of the forecasting skill of aggregate citizen forecasts. A second promising direction for future work is to consider other aggregation methods. In the present experiment, aggregate forecasts are arrived at by averaging individual forecasts. Other aggregation methods including trimmed and Winsorized means could be used instead [11]. There is reason to think that using them instead of averaging might boost the forecasting skill of the crowd [6].

As explained in the appendix (section 7), our weather poll elicits probabilistic precipitation forecasts in two ways: on a scale from 0 to 100% and on a five-point Likert scale. In the present analysis of data collected from the poll, collective forecasts are arrived at using only the quantitative inputs. No use at all is made of the qualitative Likert-scale forecasts we collected. Now, ordinary citizens might be more happy to provide qualitative inputs couched in a natural language than quantitative inputs in the form of numerical probabilities. Likert-scale inputs from individual forecasters are used to good effect by some well-established forecasting services, such as the national avalanche warning service in Switzerland [10].[9] Accordingly, future work can assess the extent of wisdom of the crowd in qualitative citizen forecasts.

Some contributing factors in extreme events such as floods, fires, avalanches and mudslides are currently forecast by official meteorological services. Others are not. An example is the presence of combustible grass and shrubs, which can contribute to wildfires. Accordingly, a promising direction for future research is to identify causal factors that are not currently forecast but about which ordinary citizens might be expected to provide useful inputs.

Acknowledgement

The authors gratefully acknowledge comments and suggestions from Alex Lenkoski, George Loewenstein, Anders Doksæter Sivle and Arjan Wardekker.

[9] These forecasts are provided by professional forecasters, not by laypeople.

Remaining errors are ours alone.

7. Appendix: the weather poll

The weather poll was administered by the commercial polling company Norstat, using its citizen panel. Norstat gathered responses from six hundred inhabitants of the Bergen municipal area, representative of the general population with respect to age and gender. The poll included two questions eliciting weather forecasts. The first question elicited a probability forecast:

> How likely do you think it is that it will rain next Saturday in Bergen (Florida) in the middle of the day (12 o'clock)?[10]

Respondents were asked to answer in qualitative terms, by selecting one of five natural-language probability expressions:

> Extremely likely, Quite likely, Likely, Not very likely, Extremely unlikely.[11]

They were asked also to give a quantitative probability forecast, by moving a virtual slider on a scale from 0% to 100%.

An effort was made to help respondents to think in terms of quantitative probabilities. Having indicated a probability of rain using the slider, this probability was depicted in an accompanying figure by the proportion of blue tiles among one hundred tiles, some blue and the rest yellow. For instance, if a respondent moved the slider to indicate a 71% probability of rain, they would see this input depicted as 71 blue tiles and 29 yellow tiles. Compare figure 6. Steps were taken to engage implicit knowledge. Thus respondents were encouraged in an accompanying subtext to think of any conditions that might make it especially likely to rain on the relevant day, or especially unlikely.[12]

[10] In the original Norwegian, the question was *Hvor sannsynlig tror du det er at det regner i Bergen (Florida) midt på dagen (klokka 12) på lørdag i denne uken?*

[11] In Norwegian, *Svært sannsynlig, Ganske sannsynlig, Sannsynlig, Ikke så sannsynlig, Svært lite sannsynlig.*

[12] In the original Norwegian, they were encouraged to *tenk på forhold som gjør det svært sannsynlig at det vil regne i løpet av denne dagen, tenk på forhold som gjør det lite sannsynlig at det vil regne denne dagen.*

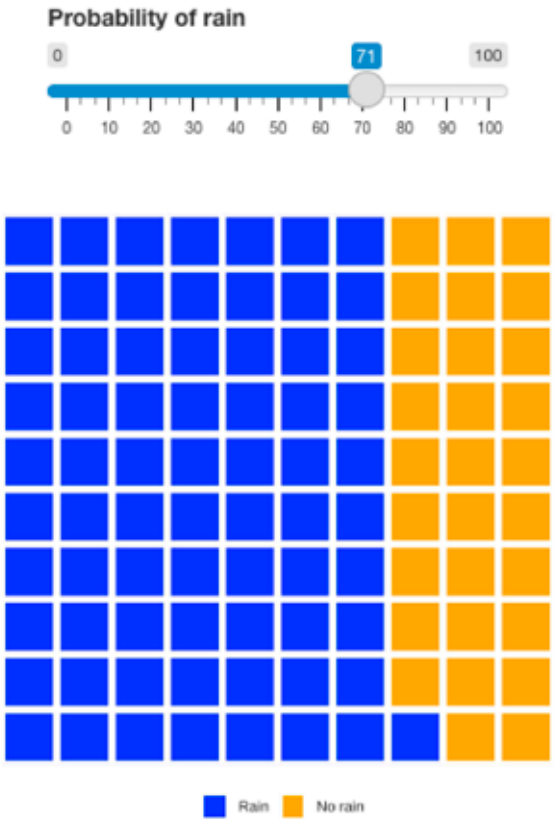

Figure 6: The forecast probability is depicted visually, using blue and yellow tiles.

The poll included a single question unrelated to the weather. Respondents were asked to estimate the surface area of *Blåsteinen*, a familiar sculpture in the center of Bergen. This question was intended as a test of the poll as an elicitation method. Wisdom of the crowd has been demonstrated very often in regard to the physical dimensions of objects. Were no wisdom of the crowd to have been observed in answers to the question about Blåsteinen, this would likely have been down to a failure of the poll to elicit meaningful inputs from citizens.

Figure 7: Blåsteinen is a familiar landmark in the center of the city of Bergen, often used by residents as a meeting place.

References

[1] Bates, J. M. and C. W. J. Granger (1969). The combination of forecasts. *Journal of the Operational Research Society 20*(4), 451–468.

[2] Brier, G. W. (1950). Verification of forecasts expressed in terms of probability. *Monthly Weather Review 78*, 1–3.

[3] Condorcet, J.-A.-N. d. C. (1785). *Essai sur l'application de l'analyse a la probabilite des decisions rendues a la pluralite des voix [microform] / par M. le Marquis de Condorcet.* Imprimerie royale Paris.

[4] Dietrich, F. and K. Spiekermann (2022). Jury Theorems. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2022 ed.). Metaphysics Research Lab, Stanford University.

[5] Galton, F. (1907). Vox Populi. *Nature 75*, 450–1.

[6] Jose, V. R. and R. Winkler (2008). Simple robust averages of forecasts: Some empirical results. *International Journal of Forecasting 24*, 163–169.

[7] Kranz, S. G. (1999). *Handbook of Complex Variables.* Boston, MA: Birkhäuser.

[8] Page, S. (2008). *The Difference.* Princeton University Press.

[9] Surowiecki, J. (2004). *The Wisdom of Crowds*. Doubleday.

[10] Techel, F., S. Mayer, C. Pérez-Guillén, G. Schmudlach, and K. Winkler (2022). On the correlation between a sub-level qualifier refining the danger level with observations and models relating to the contributing factors of avalanche danger. *Natural Hazards and Earth System Sciences 22*(6), 1911–1930.

[11] Wilcox, R. R. and H. J. Keselman (2003). Modern robust data analysis methods: measures of central tendency. *Psychological methods 8 3*, 254–74.

Weather forecasts were elicited from a representative sample of citizens living in Bergen, Norway. The elicited forecasts exhibit wisdom of the crowd: averages of forecasts by individuals have a greater forecast skill than, on average, these individual forecasts. The forecast skill of the aggregate forecasts compares favorably with that of forecasts published by the Norwegian Meteorological Institute and the Norwegian Broadcasting Corporation (NRK). The individual forecasts do not appear to be copied from the published forecasts nor do they seem to be mere statements of base rates for the relevant kinds of weather.

SNF

Samfunns- og næringslivsforskning AS
Centre for Applied Research at NHH

Helleveien 30
NO-5045 Bergen
Norway

P +47 55 95 95 00
E snf@snf.no
W snf.no

Trykk: Allkopi Bergen