

Comments on “An Exceptional Summer during the South Pole Race of 1911/12”

—KRZYSZTOF SIENICKI
Naturally Intelligent Systems
Podkowa Leśna, Poland

Recently, Fogt et al. (2017) reexamined the weather conditions reported by Roald Amundsen’s and Robert F. Scott’s expeditions during their journeys to the South Pole. The authors concluded that “it was therefore likely that the combination of these two climate patterns [Marshall southern annual mode (SAM) and El Niño] gave rise to the overall exceptional summer during the South Pole race.” (p. 2198).

To arrive at this conclusion, the authors used a textbook statistical analysis of pressure and temperature reports recorded by the expeditions along their routes and at base stations and compared these historical data with modern computer-generated data obtained from the European Centre for Medium-Range Weather Forecasts (ECMWF) interim reanalysis (ERA-Interim; Dan et al. 2011).

Fogt et al. (2017) did not provide evidence or analysis that the daily temperatures for the statistical analysis on which they base their conclusions are independent and identically distributed (i.i.d.), and that those are normally (Gaussian) distributed. In the temperature histograms presented in their Fig. 5, the authors themselves noticed that the distribution function is not normal (Gaussian). This lapse critically challenges all of their temperature conclusions based

on the normality assumption of the distribution of temperatures under analysis. Additionally, the authors do not explain how three (sometimes four) temperature measurements by Scott’s party are related to the *true* daily mean temperature, which (if justified) is calculated from the whole range of daily temperatures. The result of a simple division of three temperatures does not yield an average daily temperature (an average continuous function) and is not a sound meteorological conclusion (Sienicki 2016, p. 216; J. E. Gray and A. Vogt 2012, unpublished manuscript).

While taking a position about my criticism that calculating a mathematical average of three data points does not give a true daily average, Fogt et al. (2018) present a plot (their Fig. 1c) of four data points averaged from a set of “(0000, 0600, 1200, and 1800 UTC) and the full 24 hourly observations” (p. 2144) and argue that four data points well approximate a 24-h dataset. By selecting four data points and the data from *only* the South Pole, the authors contradict themselves, contradict historical data, and the simple fact that Scott’s team spent only about 24 h at the Pole. The historical data were measured at different times on many occasions three times per day, but on many occasions only twice or once per day were the times given, and in some cases no measurement time at all was given (e.g., 24 February 1912; Simpson 1923, p. 640). However, even neglecting these historical data recording imperfections and staying with three data points, the authors’ four-data-point analysis is irrelevant to the actual and analyzed case of Scott.

Fogt et al. (2018) offer a clarification that “the skew (0.3365) and kurtosis (−0.3013) for the data in Fig. 5 are relatively small, indicating that the data depicted in Fig. 5 are in fact approximately normally distributed” (p. 2143), which further confirms the validity of my comments that the authors’ conjecture, based on the condition of normally (Gaussian) distributed data, is inappropriate according to textbook statistical analysis. The kurtosis of any univariate normal distribution is 3 and the related skewness is 0, which is followed in the case reported by Fogt et al., where the skew is 0.3365 and the kurtosis is −0.3013. Evidently, the platykurtic character of the plot in Fig. 5 with a kurtosis of −0.3013 instead of 3 confirms the

DOI:10.1175/BAMS-D-17-0282.1
©2018 American Meteorological Society
For information regarding reuse of this content and general copyright information, consult the [AMS Copyright Policy](#).

pronounced *non-Gaussianity* of Fogt et al.'s data, on the contrary to the authors' desire.

However, even without calculating the skew and kurtosis, one readily observes from Fig. 5 that the symmetry of data around the mean value (0°C) expected for a normal distribution is not observed, as the depicted data are unsymmetrical and with right tails, especially Fig. 5a.

Further, in their reply the authors, in order to support that their data are "independent and identically distributed...plot the maximum and minimum values of the lag-1-9 autocorrelation from the full distribution of the 36 years of ERA-Interim daily mean temperatures along the tracks" (Fogt et al. 2018, p. 2144). One has to critically observe that in statistics the autocorrelation (Pearson correlation) concerns random (stochastic) processes and the authors' "daily mean" is not a random variable. The mean of a discrete random variable is a weighted average

of the possible values that the random variable can take. The mean is not variable, but a calculated average of variables. The mean of a random variable provides the long-run average of the variable or the expected average outcome over many observations.¹ Clearly, this is not the case as presented by Fogt et al. (2018) in their Figs. 1a and 1b. While finding the above, the authors unintentionally prove that the maximum and minimum (extreme) values of stochastic processes are not i.i.d. and do not follow the normal (Gaussian) distribution. One has to observe that while making the above statements, the authors are using computer-generated temperature data along with the historical routes of the expeditions. Without providing error analysis or reference and without field verification of computer predictions, the reader is left in justified uncertainty.

While comparing historical in situ data with modern computer-generated (four times per day) temperature data (ERA-Interim), the paper mentions the longitude-latitude resolution to be $0.7^\circ \times 0.7^\circ$ (Dee et al. 2011) but neglects the fact that this resolution is equal to 42 geographical miles (1 geographical mile ≈ 1.88 km), which is equivalent to about 7 marching days' progress for Scott's party during the return leg on the Barrier (Ross Ice Shelf; Sienicki 2016, Fig. 4.12). By taking Scott's (and Fogt et al. 2017) data, one can see (Fig. 1) that the data gradient is about 1°C day^{-1} . Thus, the ERA-Interim resolution introduces a

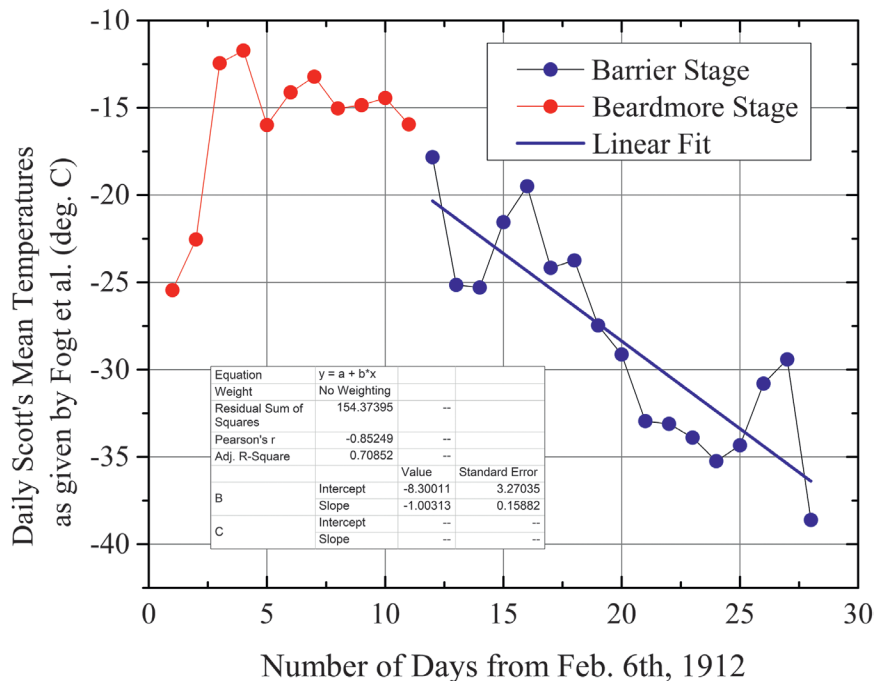


Fig. 1. Scott's daily mean temperatures as given by Fogt et al. (2017) for the Beardmore Glacier through the Barrier stages of the Scott party's return journey. The linear fit slope (blue line) is -1.0 .

¹ To illustrate the difficulty in performing the averaging procedure for a small number of data points, one may wonder about the meaning of an average of two data points, say, {2, 100}. The arithmetic mean (average) is 51, which is a formal result. But if one thinks about physical phenomena, then the physical interpretation of averaging, and thus the values obtained by the averaging procedure, must be given. Provided that the variables are i.i.d. (which is not the case for the Fogt et al. data) and provided that the variance of the variable is finite (which was not proved by Fogt et al.), one can calculate the mean (average) value of all variables using the equation $\langle x \rangle = \int_0^\infty x p(x) dx$, where x is a variable and $p(x)$ is the probability of this variable. Provided the specific nature (shape) of $p(x)$, one has to ensure that the integral is convergent $\int_0^\infty x p(x) dx < \infty$. [See Sienicki (2016, p. 57).]

large 7°C uncertainty, which is neglected by the authors. Alongside the daily temperatures, the authors make use of daily minimum and maximum temperatures and use them in the analysis. However, it is well known that the limit distributions do not follow a Gaussian distribution (e.g., Redner and Petersen 2006; Clusel and Bertin 2008; Proistosescu et al. 2016; Sienicki 2011, 2016) and, thus, should be not used to calculate their standard deviation or their average, since its convergence is not assured.

Fogt et al. (2017) suggestion—that comparing temperature data from two weather stations [the authors’ use of Henry (~89.001°S, ~0.391°W) is pointless, as no polar parties ventured northward beyond the pole in a westerly direction] separated by a distance of about 1,400 km (McMurdo and Amundsen–Scott stations) is sufficient to verify computer-generated data—translates into suggesting that about 10 weather stations across the United States are sufficient for precise prediction (and verification) of daily temperatures.

Fogt et al. (2017) thus arrive at their main conclusion, which is contradictory to one of the coauthor’s past conclusions. Namely, the comment that “Despite Scott writing...it was likely not the cold temperatures per se, but rather their persistence, that played a role in their demise, as argued by Solomon (2001)” (p. 2197). Ascribing this conclusion to Solomon is misleading, since she clearly concluded on the contrary that “Scott and his men endured a highly unusual twist of fate...very much colder than normal” (Solomon 2001) and/or “their deaths were, therefore, due, at least in part, to the unusual weather conditions” (Solomon and Stearns 1999, p. 13012).

In the analysis, the authors did not use all of Scott’s temperature data and limited their analysis to up to 3 March, though in Figs. 4b and 4c the depicted data end on 4/5 March. Thus, part of the most crucial set of temperature data recorded in Scott’s journal (Solomon 2001; Sienicki 2016) was omitted (see Fig. 2). One should observe that Fogt et al. (2017) could get daily minimum temperatures from the ERA-Interim data and compare them to Scott’s full range of temperatures, and not limit the analysis to 6 out of 24 data points (Fig. 2).

Based on a partial range of Scott’s data (25 February–19 March, instead of the complete

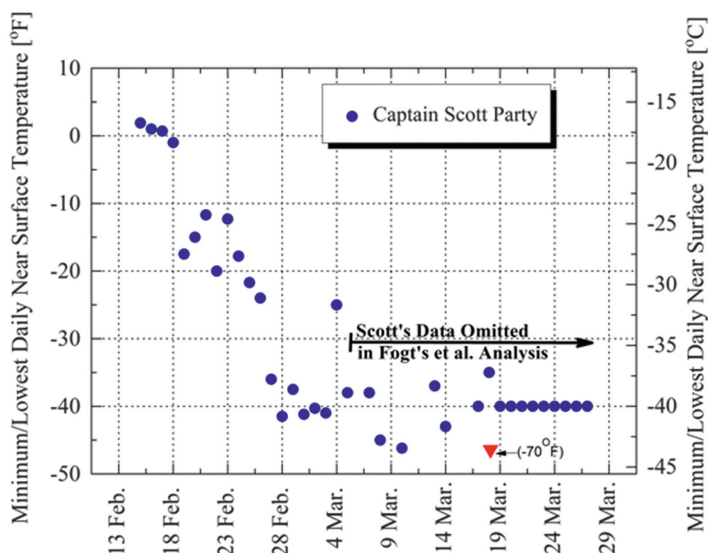


FIG. 2. Historical near-surface minimum daily temperatures (or the lowest) reported by Lt. Bowers and Captain Scott. Temperature readings after 10 Mar are not minimum near-surface temperatures. Additionally, for the following days, as reported by Captain Scott, the temperatures are the lowest reported up until that day: 15–25 Feb, excluding 16, 17, and 20 Feb; 3 and 7 Mar; and 21–27 Mar. The “super extreme cold snap” temperature of -70°F indicated by the red triangle (\blacktriangledown) cannot be placed (drawn) within this figure scale. [Plot adapted from Sienicki (2016).]

range of 25 February–27 March), one of the coauthors in related past research concluded that “Only 1 of the 15 years of modern data (1988) displays... cold daily minima, similar to 1912” (Solomon and Stearns 1999). Fogt et al. (2017), based on a cut-in-half range of Scott’s data (25 February–3 March), conclude that the probability of occurrence of the cold anomaly reported by Scott for a given year is $p < 0.001$ (“exceptionally rare”). The huge difference (about 70 times) between the old result of $p = 1/15 \approx 0.07$ and the new one of $p < 0.001$ remains unaccounted for and unexplained.

Toward the end of Fogt et al. (2017), the authors observe that the “data indicate that during 1979–2015, the temperatures rarely ($p < 0.05$) [where p stands for probability] changed as sharply as Scott and his men experienced” (p. 2198). In terms of numbers, Fogt et al.’s (2017) conclusion means that these rare events occurred about 1.8 times (for simplicity, twice) in 36 years of data. However, their suggestion that two extreme events drawn from a set of random variables (temperatures) have statistical meaning is incorrect. Statistical inference based on the occurrence of two events is faulty, as it does not represent the distribution of the underlying population of events and is the gambler’s fallacy (Farmer

et al. 2017). This fallacy consists of a bias in which individuals make an inference about future (or past) random streaks based on the outcome of previous streaks: that the history of streaks *will affect* future (or past) streaks. Certainly, and on the contrary to the authors' response, the two events are not sufficient to formulate a statistical inference (a null hypothesis) and represent an extreme case of the gambler's fallacy and/or reverse gambler's fallacy. The p values of < 0.05 , and p values themselves, have been extensively criticized for not being rigorous enough or for being arbitrarily applied as a complete solution (Benjamin et al. 2018).

In their analysis, the authors did not use all of Scott's temperature data and limited their analysis up to 3 March, though in Figs. 4b and 4c the depicted data end on 4/5 March. Thus, the most crucial set of temperature data recorded in Scott's journal (Solomon 2001; Sienicki 2016) was omitted (see Fig. 2), and the comparison to a previous publication of one of the contributors is ineffective.

While looking for correlations between temperatures along Scott's route, which stretched for about 749 geographical miles to the pole, the authors make a reference point using respective data simultaneously measured at the base camp at Cape Evans. The sudden drop in temperatures as reported by Scott in late February and early March 1912 at a distance of $<295, 130>$ geographical miles from Cape Evans is attributed by Fogt et al. (2017) to global El Niño (during the year 1912) and SAM. However, no argument nor data are provided to explain why El Niño and SAM showed their signature at the Scott party's location, while the temperature at the globally close location of Cape Evans was not altered as well. Why El Niño and SAM "decided" to strike only Scott's party and not the Cape Evans instruments is inexplicable. Equally mysterious is why during the entire polar journey the air temperatures and pressures reported by Scott and Simpson at Cape Evans were "correlated" (Simpson 1919; Sienicki 2016), with the exception of those in late February and early March, when El Niño and SAM affected only Scott's party for about one month but remained regular at Cape Evans. Neither do the authors explain why this anomaly is caused by El Niño, since the year 1988, which was cited by one of the coauthors (Solomon and Stearns 1999) as being the only year with modern data similar to 1912, was not an El Niño year. The second coldest year to come from modern data was also not an El Niño year (Sienicki 2016). While responding to critical comments about the role of El Niño and SAM in the deaths of the members of Scott's party, the authors

artfully avoid answering a simple question posted in these comments: Why did the global El Niño and SAM "decide" to strike *only* Scott's party and not the Cape Evans instruments?

The last issue concerns the authors' neglect of the research presented in an almost 800-page volume (Sienicki 2016) concerning the central theme of this paper: the Antarctic weather as the main cause of Scott's and his fellow explorers' deaths. Thus, claiming that "our work supports [our] earlier work" (Fogt et al. 2018, p. 2145) and not addressing (neglecting) the opposing research does not align with the scientific method. Neither does omitting from the analysis the data or limiting the data range to avoid not fitting their preassumed conclusions.

ACKNOWLEDGMENTS. The author thanks peer reviewers for comments that improved the presentation. I also thank Kristoffer Nelson-Kilger for comments and proofreading of the manuscript. The author appreciates the support of the Antarctic Meteorological Research Center and the Automatic Weather Station Program for surface meteorological data: NSF Grants ANT-0636873, ARC-0713483, ANT 0838834, and/or ANT-0944018.

REFERENCES

- Benjamin, D. J., and Coauthors, 2018: Redefine statistical significance. *Nat. Human Behav.*, **2**, 6–10, <https://doi.org/10.1038/s41562-017-0189-z>.
- Clusel, M., and E. Bertin, 2008: Global fluctuations in physical systems: A subtle interplay between sum and extreme value statistics. *Int. J. Mod. Phys.*, **22B**, 3311–3323, <https://doi.org/10.1142/S021797920804853X>.
- Dee, D. P., and Coauthors, 2011: The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Quart. J. Roy. Meteor. Soc.*, **137**, 553–597, <https://doi.org/10.1002/qj.828>.
- Farmer, G. D., P. A. Warren, and U. Hahn, 2017: Who "believes" in the gambler's fallacy and why? *J. Exp. Psychol. Gen.*, **146**, 63–76, <https://doi.org/10.1037/xge0000245>.
- Fogt, R. L., M. E. Jones, S. Solomon, J. M. Jones, and C. A. Goergens, 2017: An exceptional summer during the South Pole race of 1911/12. *Bull. Amer. Meteor. Soc.*, **98**, 2189–2199, <https://doi.org/10.1175/BAMS-D-17-0013.1>.
- , —, —, —, and —, 2018: Reply to "Comments on 'An exceptional summer during the South Pole race of 1911/12.'" *Bull. Amer. Meteor. Soc.*, <https://doi.org/10.1175/BAMS-D-18-0088.1>, in press.
- Proistosescu, C., A. Rhines, and P. Huybers, 2016: Identification and interpretation of nonnormality

- in atmospheric time series. *Geophys. Res. Lett.*, **43**, 5425–5434, <https://doi.org/10.1002/2016GL068880>.
- Redner, S., and M. R. Petersen, 2006: Role of global warming on the statistics of record-breaking temperatures. *Phys. Rev. E*, **74**, 061114, <https://doi.org/10.1103/PhysRevE.74.061114>.
- Sienicki, K., 2011: A note on several meteorological topics related to polar regions. *Probl. Klimatologii Polarnej*, **21**, 39–76.
- , 2016: *Captain Scott: Icy Deceits and Untold Realities*. Open Academic Press, 778 pp.
- Simpson, G. C., 1919: *Discussion. Vol. I, British Antarctic Expedition 1910–1913: Meteorology*, Thacker Spink, and Co., 326 pp.
- , 1923: *Tables. Vol. III, British Antarctic Expedition 1910–1913: Meteorology*, Harrison and Sons, 835 pp.
- Solomon, S., 2001: *The Coldest March: Scott's Fatal Antarctic Expedition*. Yale University Press, 416 pp.
- , and C. R. Stearns, 1999: On the role of weather in the deaths of R. F. Scott and his companions. *Proc. Natl. Acad. Sci. USA*, **96**, 13 012–13 016, <https://doi.org/10.1073/pnas.96.23.13012>.