Defining Events

Tristan Hauser

Affiliation not available

Erin Coughlan de Perez

Affiliation not available

December 26, 2016

1 Set up

The following text is the result of many conversations between the listed authors and many other individuals about 'heat waves'. In situations where limited resources or rapidly shifting social and/or environmental conditions make exacting analyses impossible - what are some ways to use existing tools and publicly available data to be more proactive about responding to the dangers of extreme temperatures? The question seems very worthwhile and straight forward, but it raises another tricky question: When all you have is large scale indicators, what are you really 'on the look out for'? Experiences and circumstances are varied, it's not uncommon for people get sun stroke on ski slopes and hypothermia while out walking in mild weather. What is extreme and/or dangerous can vary from place to place and person to person. Typically such factors are dissected as part of the research and decision making processes. Such nuanced analyses however, require nuanced records of stationary environments. So what can you hope to say when you don't have that luxury? Often there's not much that can be said definitively, or with general applicability. The questions become less focused on "what do we need to know, in order to be sure?" and more on "are we able to understand how much we actually do/don't know?" and "what are we prepared to do with that information, given that what we have is more than nothing, but still opaque at best?" We wanted to provide a work through of some of our efforts and thought process(es), both to illustrate the complications of this sort of analysis, and to provide some starting points for people interested in expanding on these ideas.

2 What do you mean by "heat wave"?

Extreme high temperature events, or "heat waves", have always presented a danger to people's well being. Discussions of climate change have brought about an increasing global concern regarding the dangers of these phenomena [17], which are expected to become more common and severe over the coming years [21, 32, 1, 22, 27]. Heat waves are responsible for large numbers of deaths in vulnerable regions; the Indian heat wave in May 2015 was reported to have caused 2,500 deaths. Since most heat related deaths can be prevented with adequate preparation [10], there is much emphasis and ongoing work in forecasting extreme heat hazards. To identify these hazards, climatologists typically focus on regional weather patterns and base their definitions around long term averages and the probability of exceeding a set metric; e.g., [9] and [13].

However, vulnerability and exposure to 'extreme' heat varies between individuals and regions, and it is often noted that the definition of a heat wave should vary accordingly; e.g., [26] and [25]. Local climates can vary a great deal; 20°C is considered a balmy summer temperature on Fogo Island, NL, but would probably be considered quite 'disappointing' by residents of Cape Town, RSA. Different populations, though, can experience heat waves very differently even within the same city; e.g., [8]. Vulnerability to extreme heat events can be affected by local population demographics, including health and economic status, weather variations, and ability to adapt [16]. 'Vulnerable subpopulations' often include the elderly, chronically ill, those with limited mobility or social contact, individuals living alone or those with place-based risk factors including limited access to public transportation or appropriate shelter [15, 3, 28]. It has also been noted that not just levels of heat tolerance, but notions of comfort and risk acceptance may differ between different societies. What someone would imply from, or describe as, "high" or "low" risk will be influenced by a personal/cultural approach to risk overall [4]. As such, it is pretty well accepted there is no globally applicable definition of a "heat wave" [31].

Due to these complexities, it can be a very involved and local specific task to identify what levels of extreme temperatures (and other factors), when combined with local vulnerability and exposure, will have negative impacts [23]. This becomes even more complicated since adverse effects from extreme heat are often prevented through taking adequate precautions. How do you identify a meteorological event where something bad *would have* happened if someone hadn't prevented it, when, because it was prevented, there's no event to draw your attention to? Health researchers who focus on the physiological impacts of heat waves in terms of instances of heat related illness, morbidity and mortality, typically conclude that specific definitions informed by local experience are more useful for identifying and addressing the needs of vulnerable groups of populations [19]. These varying definitions mean that different time periods and extremities of events are classified as heat waves throughout the published literature. This is in many ways a positive thing, for the reasons just outlined. However, it makes it difficult to do cross-comparisons between studies [2].

Many locations have invested considerable effort into tailored estimates of the risks posed to their communities by heat waves. However, there are many locations for where such studies are not feasible. One factor is limited research resources. Additionaly the type of studies referenced above also require extensive data records. Over many parts of the world these records are unavailable, have been lost, or were never created. As well, changing circumstances can make existing records uninformative; e.g., a community that has lost infrastructure or been displaced by natural or man-made disaster, such as an earthquake site or a refugee camp. In these cases we are starting 'from scratch'. Now the question becomes less how to fine tune our understanding of specific circumstances, but what limitations do we have to accept, and what work is possible within these limitations? What does it mean when there are no local monitoring programs, and so we must depend on low resolution data products? Can existing methodologies to estimate heatwave risk be used as a proxy in these situations? How informative can we expect classifications created for 'similar' regions to be? How can errors from inconsistent reporting be accounted for within an analysis? How can we plan appropriately when we know our analysis is based on incomplete records?

3 Working example

3.1 Definitions

So, given the situation as just described, how do you monitor an area when you don't have a detailed sense of what to look for? Table 1 lists some examples of a quantitative definitions of heat waves. These quantitative heat wave indices (HIs), are differentiated by:

(i) the metric of heat used, for instance daily mean temperature, daily maximum temperature,

(ii) thresholds of exceedance, either absolute or relative,

(iii) the role of duration in a heat wave identification [29].

This collection represents only a sub-sample of the meteorological conditions referred to as heat waves or warm events in the literature. We have deliberately focused on simpler definitions, that require only temperature data; e.g., as opposed to definitions including humidity, which is also a critical factor in heat distress [18]. This is to make the examples here as reproducible as possible over additional regions and/or smaller spatial scales, where temperature statistics are likely to be the longest running and most dependable observational records.

3.2 Temperature observations

How often, and with what effect, are these thresholds crossed? To check this we need some temperature records. Of course, the more specific your data is to the

Table 1: Summary of meteorological definitions of extreme temperature considered in this study. *In the examples presented here we omit the U.K. Met office definition, so as not to overburden the (freely provided) server space with large data files containing multiple temperature records per day. It is, however, an important example of the types of nuances that can be very useful to consider when monitoring meteorological conditions.*

Label	*Description*	*Locations Used*	*Citation*
AvgRel4	≥ 4 days with average daily temperature $>$ 90th percentile	North Africa	[5]
AvgRel2(mid)	≥ 2 days with average daily temperature $>$ 95th percentile	USA,	[2]
AvgRel3+	period of sequential days $>$ 81st percentile, with ≥ 3 days $>$ 97.5th percentile, with average temperature also $>$ 97.5th percentile	China	[24]
MaxAbs2	≥ 2 days with max daily temperature $> 35°$C, between 15th of June and 15th of September	Shanghai	[30]
MaxAbs2(d/n)	≥ 2 days with daytime max temperature $> 30°$C and nighttime max temperature $> 15°$C	UK	U.K. Met o
AvgRel2(high)	≥ 2 days with average daily temperature $>$ 98th percentile	USA	[14]
AvgRel2(low)	≥ 2 days with average daily temperature $>$ 90th percentile	USA	[14]
MaxAbs3	≥ 3 days with max daily temperature $> 35°$C	Australia	[7]
MaxRel1(low)	$> 5°$C anomaly in daily max temperature, when climatological value is $\leq 40°$C OR $> 4°$C anomaly in daily max temperature, when climatological value is $> 40°$C OR daily max temperature $> 45°$C	India	India Mete
MaxRel1(high)	$> 7°$C anomaly in daily max temperature, when climatological value is $\leq 40°$C OR $> 6°$C anomaly in daily max, when climatological value is $> 40°$C OR daily max temperature $> 45°$C	India	India Mete
MaxRel6	≥ 6 days with max daily temperature $>$ 90th percentile, using base peried of 1960-1990, and 5 day window for calculating percentile	—	[12]

local of interest, the better. However, especially when considering global monitoring and forecasting it becomes necessary to work with more 'broad stroke' estimates. For the examples here we'll consider the global observational records of the NCEP Reanalysis product [11]. Not only does this dataset date back far enough to match as many available disaster records as possible, it is the observational product most closely matched to the NCEP Global Forecasting System. Since contemporary, as well as historical, forecasts are publicly available, it is the product most likely to be used in any practical applications where locally produced options are not available or affordable. Of course, it is necessary to keep in mind that reanalysis products are only proxy indicators of the 'true ground state', especially for derived variables such as surface temperature.

3.3 Disaster reports

When thresholds are crossed, is there any suggestion that humanitarian distress has actually occurred? Again, ideally we would have medical records, and various other data sources to check against. And again, we often don't have access to this sort of detailed information. However, there is the DesInventar Project, an ongoing, publicly available compilation of humanitarian disaster records. These reports are sourced from a variety of of international locations and set in a homogeneous data base format. This is very useful when looking to systematically compare records from different regions. The range of countries for which data is available from this project are limited, however. As well, the reports are often aggregated by 2nd level administrative boundaries (typically the level of provinces/states). Also, typically, data is available only for a subset of regions within the considered country; e.g., Figure [1]. While the dataset as a whole includes documented events from 1926 to 2015 the periods where records exist within individual regions vary considerably, and must be assumed to be incomplete even within these periods. This complicates any analysis of individual temperature thresholds. Any time that a threshold is crossed without there being a reported event, there is the possibility that one did occur and wasn't reported. That means that any count of "false alarms" must be considered to be an 'upper bound' on what the actual count would have been if we had all full reporting. As well, any count of the number of events detected by a certain threshold, must be considered to be a 'lower bound' on what that number might possibly have been had all events been tallied.

3.4 Analysis approach

3.4.1 PODs and FARs

If you're monitoring forecasts and see that one of the above thresholds from Table [1] has been crossed, what do you do? Should you ring someone? Go about your business? If you do phone someone, what are the odds that you've just averted a tragedy, vs the odds that you've just annoyed someone over something that will sort its self out just fine? If you've been phoned by someone

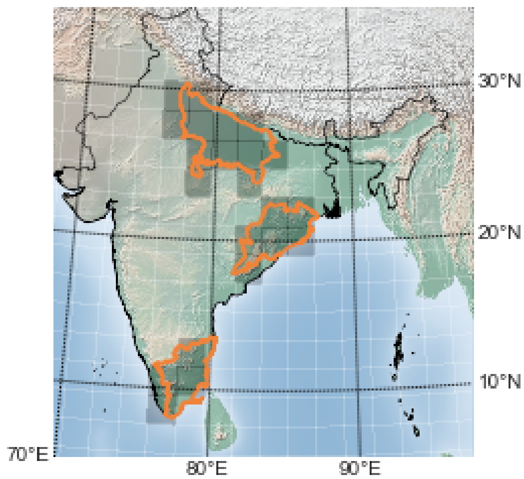

Figure 1: Map of India, outlining regions considered for this study and showing resolution of observational product (minor grid). Darkened squares represent where the reanalysis data is considered to overlap with the areas of interest. [Source code *that generates this image, and performs most of the data preprocessing is available by following the provided link.*]

responding to a forecast, how much of a response are you willing to make, given that you know there is a chance that the thing they are concerned about won't actually create an issue for your area? Given that responding to the call could save lives, how willing are you to dismiss the message?

The answers to these questions will hinge on how effective the different warning signs used as heat wave definitions are at detecting actual adverse situations; i.e., preventing situations where 'we could have done something but didn't', as well as the price tag attached to 'acting in vain'. Failure to detect events may result in needless suffering, and acting on false alarms wastes resources and may erode willingness to head warnings in the future. As such, we'll evaluate these definitions based on Probability of Detection (POD); i.e., the percent of recorded humanitarian events that occur under situations that are identified as dangerous by the meteorological definitions, and False Alarm Rates (FAR); i.e., the percentage of days identified as dangerous by meteorological considerations, where nothing is reported to have happened. It's worth noting that, as illustrated in the contingency table outline [2], that a false alarm is a different situation from a missed event and so the PODs and FARs must be considered separately from each other. A very high POD may come at the cost of a very high FAR (the alarm is always on), and a low FAR may come at the cost of low POD (the alarm never sounds).

Table 2: Template for a standard contingency table.

		Event Observed: Yes	No
Event Forecast:	**Yes**	(detection)	(false alarm)
	No	(miss)	(correct negative)

3.4.2 Statistical nuances

As the number of events vary from location to location and between warm spell definitions, it is necessary to consider how to define when a given POD or FAR percentage is significantly higher or lower than another; i.e., to put error bars on our observations based on uncertainties caused by small sample sizes. For this example we'll use an approach described by [20][1], for modeling binomial[2] data using a Beta likelihood function and a Jefferies prior. Essentially, we start with an *a priori* (hence the name) mathematical description of what chances we would give different percentages of being an appropriate value to assign as the POD or FAR. Then we update this description as new observations come in, based on how likely the likelihood function (hence the name) tells us it would be to 'see what we saw' if our assumptions about what are 'good' POD and FAR values are correct. Here, we choose a prior distribution that is 'uninformative'; i.e., the final estimate of the optimal POD or FAR value is just calculated from the record of detections and false alarms in the data, and the uncertainty on those estimates results from the number of data points that we have. Well, for the most part. Some special consideration is given when we observe a POD or FAR of 0% or 100%. In such situations it might be assumed through 'the law of succession' that in the long term it is very unlikely for any method to perform perfectly (or perfectly terrible), and so that observation will be considered to have a large uncertainty. Alternatively, it is possible to allow a score of 100% or null to be the expected result. The first option seems extreme, and the second seems unlikely given the small sample sizes, and such calculations require a prior that biases the uncertainty estimates for other values. The Jefferies prior that we've decided on offers a compromise solution, with gives results in the middle of those expected from the two extremes.

A brief interlude Wait, what? So we have to decide on the probability of a value being the Probability of Detection? Um, yeah, that's pretty much it. There's rarely any good answer to the question of how much conceptual/mathematical detail to go into when outlining choice of statistical methods. We could start being pedantic at which point some readers will say: "This is old news, they teach Baysian statistics in secondary school now", and some will say: "Enough of this sort of talk, Baysian statistics isn't a valid subject for

[1] "As described by" is academic terminology for: "This is a perfectly normal thing that people do. Really! It wasn't even our idea! Do you really think we'd come up with this on our own?"

[2] Data that gives a record of a binary process; e.g., heads or tails?, meat or veg?, success or failure, yes or no?, etc.

schools or elsewhere", and others will say: "You've described it all wrong, technically you have to say this..." Most likely, either the previous paragraph sounded reasonable, or you don't care. Still, briefly, let's try to elucidate a bit...

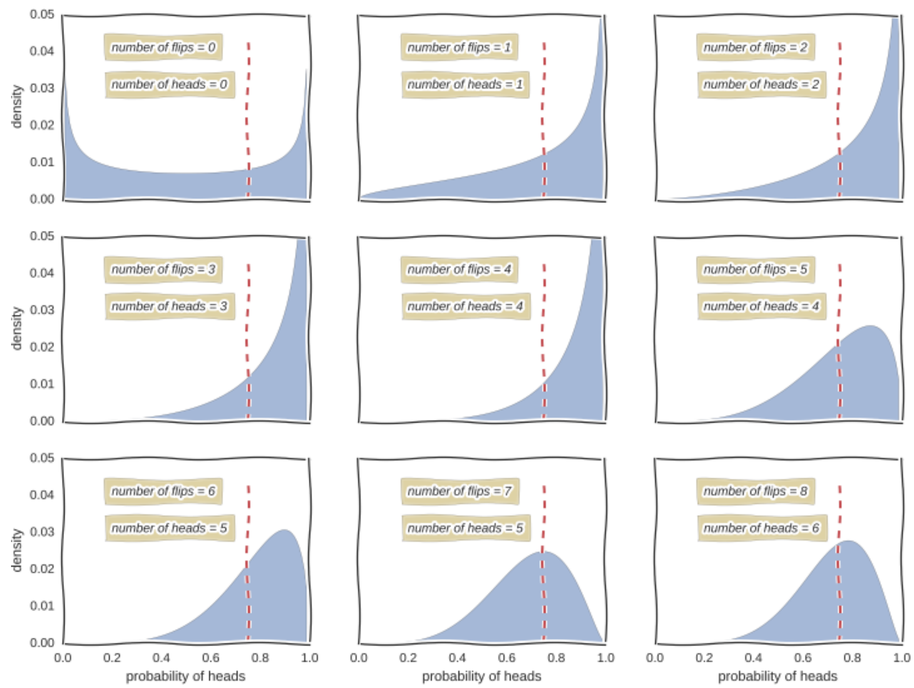

Figure 2: Consider that you have to gamble on a coin toss. Since someone is asking you to gamble on the toss you suspect that it's a weighted coin, but what way has it been doctored? How is your opinion going to change as you watch the coin being flipped? What is the probability that the coin will come up heads $\frac{3}{4}$ of the time? What is the probability that it will come up heads $\frac{1}{2}$ of the time, or only $\frac{1}{3}$ of the time? We might start out saying that you consider that all these scenarios have equal probability. That's the situation shown in the top-left square of the figure. There is no favor shown to different values of the "probability of heads", hence the mostly 'flat' line[4]. Now watch the coin be flipped. As it keeps coming up heads, we become more sure that the coin is weighted strongly towards heads. On the fifth toss we finally get a tails, so we now know that there is zero chance of the coin coming up heads 100% of the time. As we continue to flip, our beliefs on the nature of the coin start to converge around the actual weighting that has been applied (shown by the dashed red line). `[view source]`

In the heat wave example, rather than "the coin landing heads" as in [2] and [3], a 'success' is a meteorological event that has been defined as a heat wave occurring at the same time that a humanitarian crisis is reported. Alternatly, can define "the coin landing heads" as a false alarm; i.e., a meteorological event that has been defined as a heat wave having no recorded corresponding humanitarian crisis. So why do we need to consider the

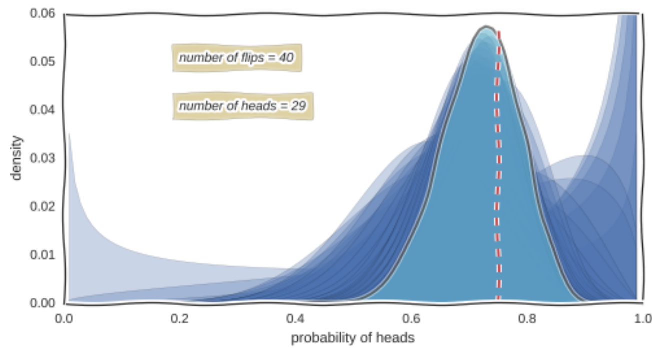

Figure 3: After observing forty coin flips and having heads come up 72.5% of the time, our current idea of plausible values to assign to the "probability of heads" (the pale blue distribution outlined in thick gray) has evolved notably from previous notions (the backing distributions outlined faintly behind the 'current' one). Not only has the distribution moved towards having an optimal value at the 'actual' weighting, as the number of samples increases, it also becomes narrower. That is, we become more confident in a more constrained notion of what are plausible values to assign this statistic. [`view source` - *same notebook as for Figure 2*]

probability of our PODs and FARs being the 'true' values, why can't we just calculate these statistics directly and call it good? The whole procedure described above is a way to describe "credible intervals", a.k.a error bars, on the values we estimate as the POD and FAR for a definition/location combination. Why does this matter? Say that one definition gives a 72% POD for a location while another gives a POD of 77%. Do we say that the second definition is the 'better' one? Maybe within our records we only had a few warm events where we could test our definitions. With this limited sample size the approach outlined above might give the optimal estimates of 72% and 77% but if we consider the width of the distributions around them, in both cases we have to admit that given our limited data we would have to say the plausible PODs that we will expect to see for either definition in the long term are anywhere between 60% and 90%. As such, we have no way to say that we yet have evidence that either definition is the one we should 'stick with'. In the same way that even after observing the coin flip come up heads four times in a row in Figure [2], we did not discount the possibility that the long term behavior might be more varied, and, if we had been gambling, would have been glad we did. Anyways, moving on...

3.4.3 Expect delays

It is important to note that incomplete reporting means that statistics presented here are not fully representative. As said before, due to unreported and/or prevented events estimated POD values should be considered a lower limit, and FAR values an upper limit, on what would be seen with full documentation of a region's heat wave events. As well, the dating of events is inexact. In many cases there will be a delay between the date of the event and the date that it is reported. To address this we'll experiment with varying delay windows when evaluating performance. If a warm spell occurred within

n days prior to a reported event (with $n < 14$), that event can be considered 'detected' and warm spells within that window can not be considered 'false alarms'. Obviously this does not result in an exact evaluation, but we're not asserting a singular causal relationship between warm spells and humanitarian suffering. Instead we're trying to prioritize evaluating 'performance within an acceptable threshold'. The reporting delay will be considered for the POD estimates; i.e., "how long a window do we need to make sure we've caught all the events we can", rather then the FAR statistics, as these would, by definition, steadily and artificially decrease as the window was extended.

3.4.4 Casting a wide net

As mentioned above, the disaster reports are aggregated by state/province. Some reports have more detailed location information, but this is not consistent across the database. States/provinces are typically large regions; i.e., the daily weather is likely to be inhomogeneous across the municipality. We'll try to mirror this aggregation by considering a region to be within a warm spell if any overlapping observational grid-cell is so classified by a given definition. Overlapping is defined here as a grid-cell whose center is inside or within 50km of a regional boundary. This distance is a conservative choice; as it tends towards considering information from outside the region, rather than excluding information from within, but still keeps a more even evaluation, by not evaluating against grid cells which contain only a minimal amount of the geographical region. Considering that most political boundaries are linked to geographical features/divides, if only a small fraction of the region extends into another grid cell, one would expect its weather to be more correlated with that shown by cells over the rest of the region than that of this other area, as shown in the map of Inida [1]. Admitedly, making comparisons by regions is quite 'broad strokes'. As described above we are unfortunatly limited to what can naïvely be considered the humanitarian relevant information content of large scale temperature monitoring, rather than identifying local/community scale relationships to meteorological variability.

3.5 Pulling it all together

3.5.1 Applying the definitions

So what happens when we actually try to match the example definitions [Table 1] against the regional data that was obtained [1]? We do these calculations and for illustration look at what dates have been flagged for Uttar Pradesh, during the particularly bad year of 1998 [6]. We can easily see that some definitions are more 'trigger happy' than others. As well, they all indicate activity around the period identified by the DesInvetar records. Well, except for MaxAbs2, which, as stated [Table 1], only considers heat waves to occur between mid-June and mid-September.

So with the exception of the definition that was 'switched off' at the time of the event, all the definitions logged a "detection" for this incident. As stated previously, we were even willing to give them some lee-way if they were 'just off'. The other question is: "what counts as a miss?" There are a lot of days that are marked, while only one is recorded as an event. In absolute terms those marked days that don't correspond with an event are misses. There are, however, many possible considerations. For example, the event logs are given absolute dates, but are often representative of the cumulative effect of prolonged conditions. As well, whether a warning was ineffective, can, in practice, depend greatly on the extent that the action taken in response stays 'in

Table 3: Optimal delay thresholds for each Indian region and heatwave definition.

	AveRel4	*AvgRel2(mid)*	*AvgRel3+*	*MaxAbs2*	*MaxAbs3*	*AvgRel2(high)*	*AvgRel2(*
Tamil Nadu	0	5	3	0	4	5	0
Orissa	10	12	10	11	11	12	10
Uttar Pradesh	0	1	1	0	0	5	0

place'. If every warning means spending massive resources on an immediate response, then any false alarm is a costly one. Often though a response can be as simple as restocking medical or water supplies earier in the season than originaly planned.

3.5.2 How wide a net do you need?

What sort of delay thresholds do the data suggest give the most 'benefit of a doubt' to the different definitions. Applying the considerations described above to the Indian regions shows some interesting variety of responses:

It is reassuring that at no point was our preset maximum threshold selected as the optimal value. Also interesting to see the variety in the spread of the optimal values for the different regions.

References

[1] Rais Akhtar. Climate change and health and heat wave mortality in India. *GLOBAL ENVIRONMENTAL RESEARCH-ENGLISH EDITION-*, 11(1):51, 2007.

[2] G. Brooke Anderson and Michelle L. Bell. Heat Waves in the United States: Mortality Risk during Heat Waves and Effect Modification by Heat Wave Characteristics in 43 U.S. Communities. *Environ Health Perspect*, 119(2):210–218, nov 2010.

[3] Ulrich; Bittner Martin-Immanuel; Stößel. Perceptions of heatwave risks to health: results of an qualitative interview study with older people and their carers in Freiburg, Germany, 2012.

[4] Mary Douglas and Aaron Wildavsky. How Can We Know the Risks We Face? Why Risk Selection Is a Social Process1. *Risk Analysis*, 2(2):49–58, jun 1982.

[5] Bernard Fontaine, Serge Janicot, and Paul-Arthur Monerie. Recent changes in air temperature heat waves occurrences, and atmospheric circulation in Northern Africa. *Journal of Geophysical Research: Atmospheres*, 118(15):8536–8552, aug 2013.

[6] Handbook. IGI Global, 2016.

[7] Alana Hansen, Peng Bi, Monika Nitschke, Philip Ryan, Dino Pisaniello, and Graeme Tucker. The Effect of Heat Waves on Mental Health in a Temperate Australian City. *Environ Health Perspect*, 116(10):1369–1375, jun 2008.

[8] Sharon L. Harlan, Juan H. Declet-Barreto, William L. Stefanov, and Diana B. Petitti. Neighborhood Effects on Heat Deaths: Social and Environmental Predictors of Vulnerability in Maricopa County, Arizona. *Environmental Health Perspectives*, 121(2):197–204, nov 2012.

[9] B. G. Hunt. A Climatology of Heat Waves from a Multimillennial Simulation. *Journal of Climate*, 20(15):3802–3821, aug 2007.

[10] Hans-Peter Hutter, Hanns Moshammer, Peter Wallner, Barbara Leitner, and Michael Kundi. Heatwaves in Vienna: effects on mortality. *Wiener klinische Wochenschrift*, 119(7-8):223–227, may 2007.

[11] E. Kalnay, M. Kanamitsu, R. Kistler, W. Collins, D. Deaven, L. Gandin, M. Iredell, S. Saha, G. White, J. Woollen, Y. Zhu, A. Leetmaa, R. Reynolds, M. Chelliah, W. Ebisuzaki, W. Higgins, J. Janowiak, K. C. Mo, C. Ropelewski, J. Wang, Roy Jenne, and Dennis Joseph. The NCEP/NCAR 40-Year Reanalysis Project. *Bull. Amer. Meteor. Soc.*, 77(3):437–471, mar 1996.

[12] Thomas R. Karl, Neville Nicholls, and Anver Ghazi. CLIVAR/GCOS/WMO Workshop on Indices and Indicators for Climate Extremes Workshop Summary. In *Weather and Climate Extremes*, pages 3–7. Springer Science and Business Media, 1999.

[13] Shia T. Kent, Leslie A. McClure, Benjamin Zaitchik, Tiffany T. Smith, and Julia M.. Gohlke. Heat Waves and Health Outcomes in Alabama (USA): The Importance of Heat Wave Definition. *Environmental Health Perspectives*, nov 2013.

[14] Shia T. Kent, Leslie A. McClure, Benjamin Zaitchik, Tiffany T. Smith, and Julia M.. Gohlke. Heat Waves and Health Outcomes in Alabama (USA): The Importance of Heat Wave Definition. *Environ Health Perspect*, nov 2013.

[15] Eric Klinenberg. *Heat Wave*. University of Chicago Press, 2002.

[16] Kim Knowlton, Miriam Rotkin-Ellman, Galatea King, Helene G. Margolis, Daniel Smith, Gina Solomon, Roger Trent, and Paul English. The 2006 California Heat Wave: Impacts on Hospitalizations and Emergency Department Visits. *Environ Health Perspect*, 117(1):61–67, aug 2008.

[17] R. Sari Kovats and Shakoor Hajat. Heat Stress and Public Health: A Critical Review. *Annu. Rev. Public. Health.*, 29(1):41–55, apr 2008.

[18] C. Linares, J. Diaz, A. Tobías, R. Carmona, and I. J. Mirón. Impact of heat and cold waves on circulatory-cause and respiratory-cause mortality in Spain: 1975–2008. *Stochastic Environmental Research and Risk Assessment*, 29(8):2037–2046, oct 2014.

[19] Tabea K. Lissner, Anne Holsten, Carsten Walther, and Jürgen P. Kropp. Towards sectoral and standardised vulnerability assessments: the example of heatwave impacts on human health. *Climatic Change*, 112(3-4):687–708, sep 2011.

[20] David MacKay. *Information Theory, Inference, and Learning Algorithms*. Cambridge University Press., 2003.

[21] G. A. Meehl. More Intense, More Frequent, and Longer Lasting Heat Waves in the 21st Century. *Science*, 305(5686):994–997, aug 2004.

[22] Vimal Mishra, Auroop R Ganguly, Bart Nijssen, and Dennis P Lettenmaier. Changes in observed climate extremes in global urban areas. *Environmental Research Letters*, 10(2):024005, feb 2015.

[23] Paulo Nogueira and Eleonora Paixão. Models for mortality associated with heatwaves: update of the Portuguese heat health warning system. *International Journal of Climatology*, 28(4):545–562, 2008.

[24] Roger D. Peng, Jennifer F. Bobb, Claudia Tebaldi, Larry McDaniel, Michelle L. Bell, and Francesca Dominici. Toward a Quantitative Estimate of Future Heat Wave Mortality under Global Climate Change. *Environ Health Perspect*, 119(5):701–706, dec 2010.

[25] S. E. Perkins and L. V. Alexander. On the Measurement of Heat Waves. *Journal of Climate*, 26(13):4500–4517, jul 2013.

[26] Peter J. Robinson. On the definition of a heat wave. *Journal of Applied Meteorology*, 40(4):762–775, 2001.

[27] E. Roldán, M. Gómez, M. R. Pino, J. Pórtoles, C. Linares, and Julio Díaz. The effect of climate-change-related heat waves on mortality in Spain: uncertainties in health on a local scale. *Stochastic Environmental Research and Risk Assessment*, 30(3):831–839, apr 2015.

[28] Joyce Klein Rosenthal, Patrick L. Kinney, and Kristina B. Metzger. Intra-urban vulnerability to heat-related mortality in New York City 1997–2006. *Health & Place*, 30:45–60, nov 2014.

[29] Tiffany T. Smith, Benjamin F. Zaitchik, and Julia M. Gohlke. Heat waves in the United States: definitions patterns and trends. *Climatic Change*, 118(3-4):811–825, dec 2012.

[30] Jianguo Tan, Youfei Zheng, Guixiang Song, Laurence S. Kalkstein, Adam J. Kalkstein, and Xu Tang. Heat wave impacts on mortality in Shanghai 1998 and 2003. *Int J Biometeorol*, 51(3):193–200, oct 2006.

[31] Shilu Tong, Xiao Yu Wang, and Adrian Gerard Barnett. Assessment of Heat-Related Health Impacts in Brisbane Australia: Comparison of Different Heatwave Definitions. *PLoS ONE*, 5(8):e12155, aug 2010.

[32] Antje Weisheimer and T. N. Palmer. Changing frequency of occurrence of extreme seasonal temperatures under global warming. *Geophysical Research Letters*, 32(20), 2005.

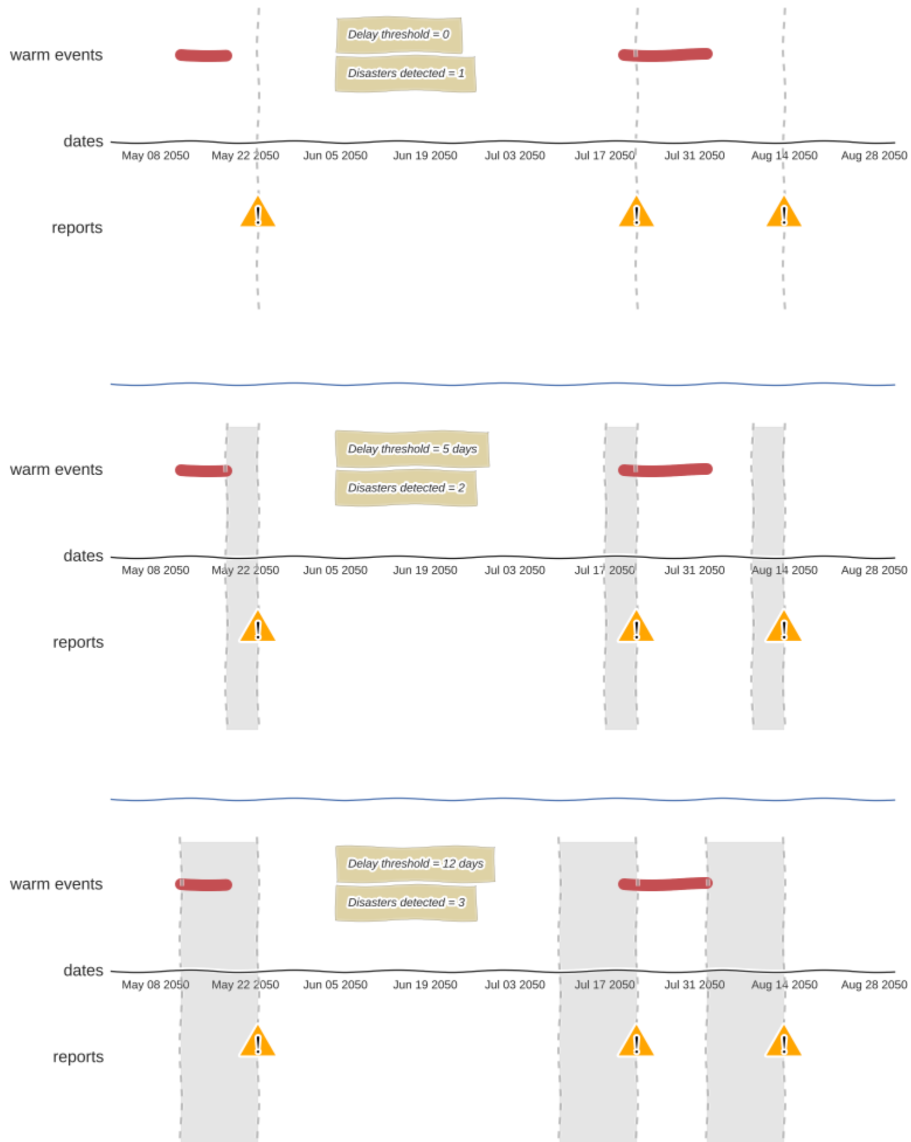

Figure 4: A quick, fictitious (note the dates are decades into the future) illustration of some of the implications of using a delay threshold. With no delay threshold only one of the reported events happens during a marked 'warm event'. It is quite plausible that the earliest reported event was a result of the previous warm spell, but was not documented or fully experienced until a few days afterwards. Using a threshold within which the event still counts as detected means that the meteorological definition that identified the warm event still gets 'credit' for having warned of possible consequences. Further extending the threshold allows for the claim that the final report was due to the second identified warm spell. That one warm event is considered responsible for two disasters isn't considered an issue. The meteorological states and reports are aggregated from over a large spatial region, and different areas could experience and/or report the events at different times. Alternatly, the second report occurs early in second warm event, and perhaps the third report resulted from the continuation of the warm event. Again, there's no attempt here at an analysis of what causes what, just "If we were monitoring X would we have had reason to expect something?" It is also possible that the twelve day threshold is too generous, and that the final report was a result of a new event that was not captured by the definition of 'warm event' used in this illustration. What might

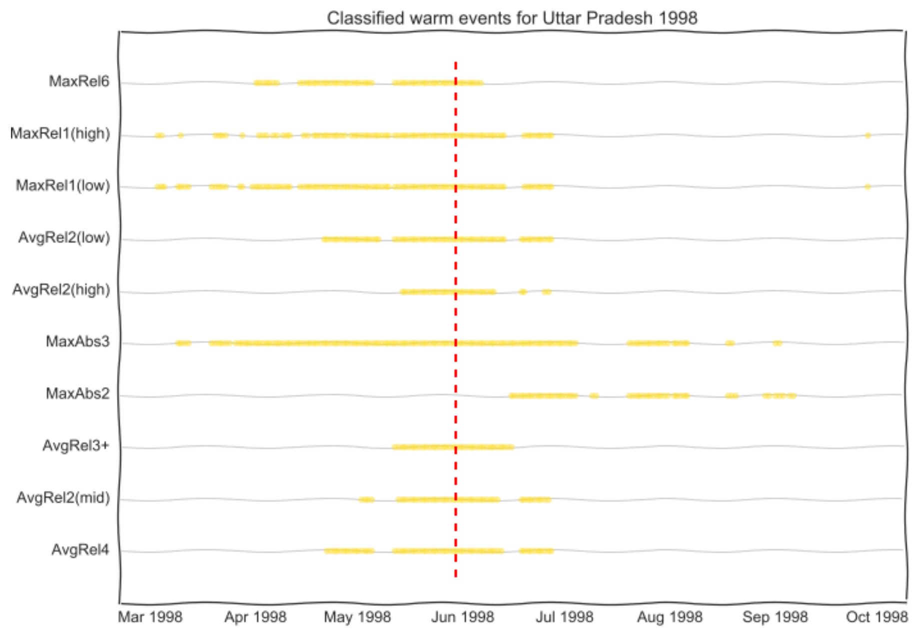

Figure 5: Days flagged by the various heat wave definitions are marked in yellow. The date of the DesInvetar reported heat wave event is marked with a red dashed line. [view source]

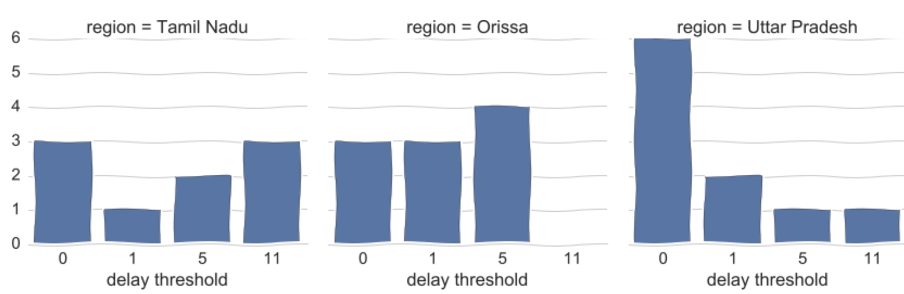

Figure 6: Counts of each optimal delay threshold value for each Indian region.