Geochemistry Geophysics Geosystems G3
AN ELECTRONIC JOURNAL OF THE EARTH SCIENCES
Published by AGU and the Geochemical Society

Article
Volume 4, Number 7
31 July 2003
1065, doi:10.1029/2002GC000478
ISSN: 1525-2027

Lognormal distribution of precipitable water in Hawaii

James Foster and Michael Bevis
SOEST, University of Hawaii, 1680 East West Road, Honolulu, Hawaii 96822, USA (jfoster@soest.hawaii.edu)

[1] We use four-year time series of precipitable water (PW) and zenith neutral delay (ZND) derived from a GPS network in Hawaii to show that the statistical distributions of these quantities are closely approximated by the lognormal distribution. The long term average and median values of precipitable water decline exponentially with height, or very nearly so. The arithmetic standard deviation of PW declines nearly linearly with height, whereas the geometric standard deviation increases nearly linearly (but relatively weakly). Our finding that PW has, to a good approximation, a lognormal distribution between sea level and 4 km elevation is confirmed by an analysis of radiosonde profiles. Lognormality is a common property of other meteorological quantities such as precipitation, aerosol optical and cloud distributions. This is the first time, as far as we know, that PW has also been shown to have a lognormal distribution.

Components: 3569 words, 6 figures.

Keywords: GPS; water vapor.

Index Terms: 3309 Meteorology and Atmospheric Dynamics: Climatology (1620); 3374 Meteorology and Atmospheric Dynamics: Tropical meteorology; 6904 Radio Science: Atmospheric propagation.

Received 20 November 2002; **Revised** 2 April 2003; **Accepted** 15 May 2003; **Published** 31 July 2003.

Foster, J., and M. Bevis, Lognormal distribution of precipitable water in Hawaii, *Geochem. Geophys. Geosyst.*, *4*(7), 1065, doi:10.1029/2002GC000478, 2003.

1. Introduction

[2] The importance of the roles that atmospheric water vapor plays in the climate and weather systems can hardly be overstated [*Starr and Melfi*, 1991; *Anthes*, 1983]. The vertically integrated water vapor content of the atmosphere, or precipitable water (PW), is also of great interest to space geodesists. Atmospheric water vapor induces a propagation delay in the radio signals from the satellites of the Global Positioning System (GPS) that is highly variable both in space and in time. The statistical properties of PW and the associated 'wet delay' are relevant to geodesists designing algorithms to estimate atmospheric propagation delays from the data collected by networks of geodetic GPS receivers and by related space-geodetic systems such as Very Long Baseline Interferometry [*Hogg et al.*, 1981; *Treuhaft and Lanyi*, 1987; *Lichten and Border*, 1987]. With the advent of 'GPS meteorology' where networks of continuously operating GPS stations are used to estimate the PW history above each station in the network [*Bevis et al.*, 1992; *Duan et al.*, 1996; *Bar-Sever et al.*, 1998; *Emardson et al.*, 1998; *Gutman and Benjamin*, 2001], long time series of PW measurements are now becoming available, and this is contributing to a resurgence of interest in the statistics of PW. Most of this research has focused on the autocorrelation structure and power spectra of PW time series acquired at a given point in space, or on the spatial correlation or cross power spectra of PW time series collected at different locations [e.g., *Williams et al.*, 1998; *Davis*, 2001].

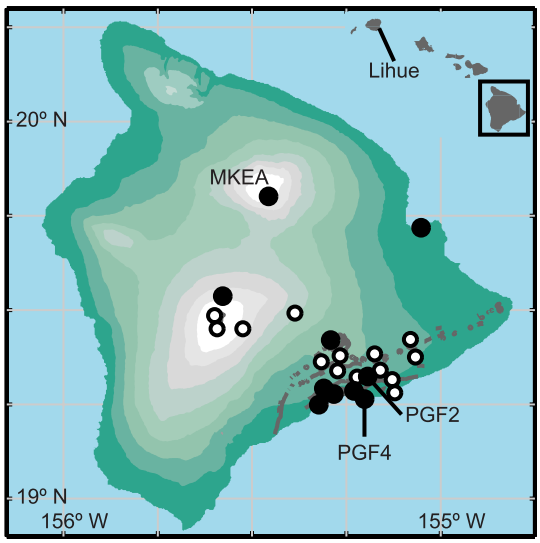

Figure 1. Location of GPS sites on the Big Island of Hawaii. Sites with colocated meteorological instruments are the solid black circles.

In this paper we take a different approach, and investigate the statistical distribution of PW measurements accumulated over time periods of several years. This initial study is focused on Hawaii because the availability of GPS stations over a wide range of elevations provides us with an unusual opportunity to examine how the statistical distribution of PW varies with height.

2. Data and Analysis

[3] The GPS network on the Big Island of Hawaii (Figure 1) consists of 24 sites spanning a height range from sea level to the summits of the two main volcanoes, Mauna Kea and Mauna Loa, at ~4000 m. Of these sites 10 have colocated meteorological instruments to provide direct measurements of pressure and temperature. As GPS-derived PW is typically found to have a random error component of ~1 mm and a bias of up to ~1 mm [*Duan et al.*, 1996; *Fang et al.*, 1998; *Tregoning et al.*, 1998], the PW for these sites is estimated to be accurate to 1–2 mm. In order to be able to incorporate the other sites into our study we extrapolated the pressure and temperature fields to each location, constraining the process by incorporating data interpolated from the NCEP Global Reanalyses. The predicted pressures and temperatures have RMS errors of ~.25 mbar (1 mbar = 1 hPa) and ~1.5°C respectively. This is sufficiently accurate to permit us to include in our analysis PW solutions from all the sites, with the accuracy for those sites using the extrapolated meteorological data estimated as ~2 mm.

[4] All GPS data available from 1997 through 2000 were processed using GAMIT [*King and Bock*, 2000], with 24 atmospheric delay parameters estimated in each 24-hour batch solution (following *Duan et al.* [1996]), giving hourly estimates of zenith neutral delay (ZND) for the entire 4-year period. Surface pressure was converted to zenith hydrostatic delay (ZHD) [*Saastamoinen*, 1972] and the mapping parameter Π was calculated from surface temperature using the seasonal climatology determined by *Ross and Rosenfeld* [1997]. The final transformation into PW is simply given by $PW = \Pi(ZND - ZHD)$.

3. Lognormal Distribution

[5] The two continuous probability distributions most commonly used to describe atmospheric variables are the normal (N) and lognormal (Λ) distributions. Comparison of our PW histograms with many potential statistical distributions [*Evans et al.*, 1993] indicated that only the lognormal family closely matched all the histograms. This is consistent with the prevalence of the lognormal in other investigations of moisture related atmospheric parameters (see for example *Biondini* [1976], *López* [1977], *Soden and Bretherton* [1993], and *Raymond* [2000]). A variate x is Λ-distributed if the variate $z = \log(x)$ is N-distributed. For a detailed discussion of Λ see *Aitchison and Brown* [1957]; here we will simply summarize some of the key details.

[6] The lognormal distribution can be parameterized in a number of ways, and sometimes this causes confusion. We choose to parameterize Λ in terms of the median, M, and the geometric standard deviation, s, of the population, i.e., $\Lambda = \Lambda(x|M, s)$ where $x > 0$. (Note that some authors prefer the parameterization $\Lambda = \Lambda(x|m, s)$ where $m = \log(M)$). The geometric standard deviation (GSD) should not be confused with the more

widely used arithmetic standard deviation (ASD), usually referred to simply as the standard deviation, which is most commonly associated with the N-distribution. The PDF for **X** is given by:

$$\frac{1}{xs\sqrt{2\pi}} \times \exp\left\{\frac{-\left(\log\frac{x}{M}\right)^2}{2s^2}\right\}, 0 < x < \infty \qquad (1)$$

[7] The population mean is $M\ \exp(\frac{s^2}{2})$ and the population variance is $M^2\exp(s^2)(\exp(s^2) - 1)$.

[8] Note that both the mean and variance are functions of s and M indicating that a change in the variance of the distribution will usually also indicate a change in the mean, and vice versa. This is an important property with consequences for the interpretation of trends in mean water vapor.

[9] A slightly more general version of Λ is the 3-parameter distribution $\Lambda(x|M,\ s,\ t)$. Here an extra term t is included as a "threshold" parameter. The threshold parameter allows the distribution to describe the situation where the variable has a non-zero lower bound. The threshold simply acts to translate the PDF along the x axis. The PDF for the 3-parameter distribution is given by equation 1 above substituting the variate with $x' = x - t$:

$$\frac{1}{(x-t)s\sqrt{2\pi}} \times \exp\left\{\frac{-\left(\log\frac{(x-t)}{M}\right)^2}{2s^2}\right\}, 0 < x - t < \infty \qquad (2)$$

The 2-parameter form is now the special case where $t = 0$. The PDF is now defined for $0 < x - t < \infty$ and the locations of the median and mean are shifted by t.

4. Results

[10] Histograms of PW for three of the sites with colocated meteorological instruments are shown in Figure 2 with the derived Λ curves superimposed. PGF4 is near sea level, PGF2 at mid-elevation and MKEA is near the summit of Mauna Kea. These three sites are typical of all results (though the sites without direct meteorological observations are slightly noisier) and show the progression toward more strongly peaked and skewed distributions with increasing elevation. The typical Λ shape is evident for each site with the histogram rising steeply from a threshold value to the peak and with a long tail for the upper portion.

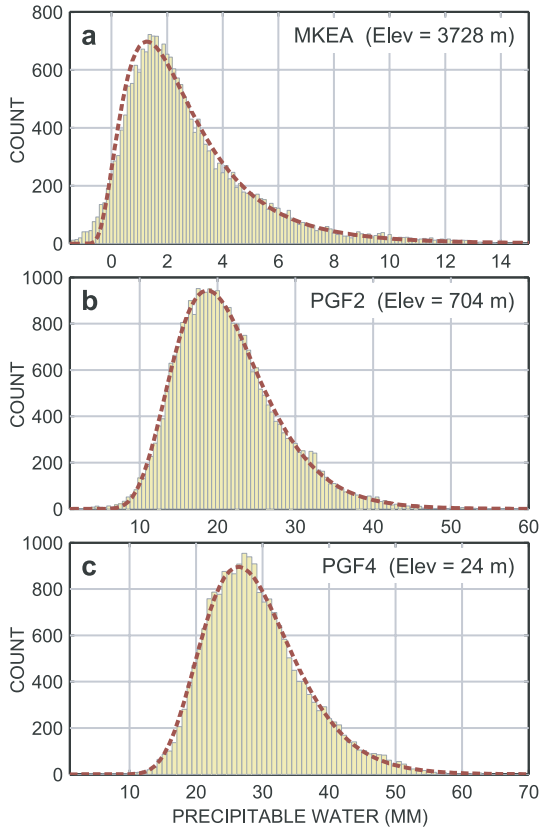

Figure 2. Histograms of precipitable water for MKEA, PGF2, and PGF4 with lognormal PDF curves superimposed. The parameters t, M, and s for each PDF curve are: −1, 2.19, 0.581; 0, 20.45, 0.301; 0, 28.14, 0.256, respectively.

[11] We have extended our analysis to consider the distribution of the ZND measured by GPS, the ZHD computed from surface pressure, and the ZWD (= ZND − ZHD). The results obtained at PGF2 (Figure 3) are characteristic of the entire network. Not surprisingly the ZWD has the form of Λ. In contrast, the ZHD, like pressure, looks more normally distributed. Although the ZHD is always much larger than the ZWD, the ZWD is far more variable in most parts of the world, including Hawaii. Figure 3 shows that although the ZHD has values over 2000 and the ZWD values less than 300 mm of delay, the total range for the ZHD PDF is only ~30 mm whereas the ZWD has a range of ~300 mm. As a result the ZHD is nearly a constant

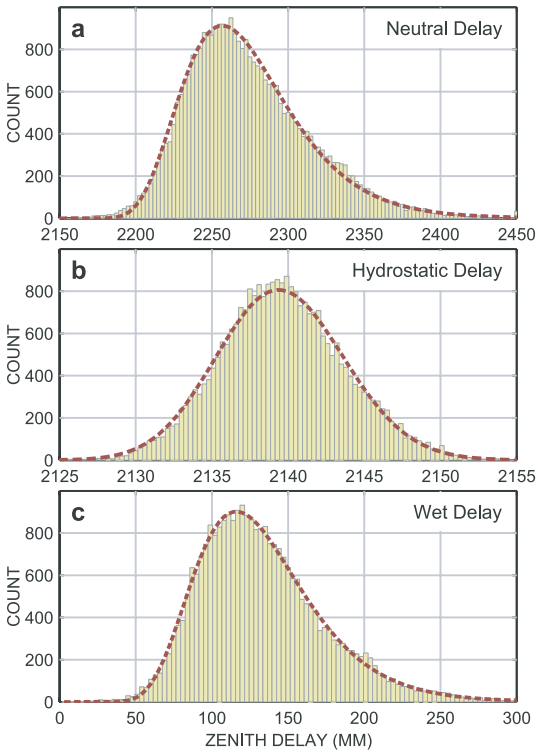

Figure 3. Histograms of the three zenith delays for PGF2 with lognormal PDF curves superimposed. The parameters t, M, and s for each PDF curve are: 2147.5, 2268.1, 0.313; 2000, 2139.5, 0.030; 0, 127.28, 0.304, respectively.

component of the ZND (that could be equated to t) and the ZND has a Λ distribution, or very nearly so.

[12] We now turn our attention to the elevation dependence of the PW. In Figure 4a we plot both mean and median PW against station height. Both mean and median follow the expected exponential form, with PW scale heights of ~1500 m to ~1700 m respectively. These heights need not correspond physically to any real atmospheric layer, but do in fact fall within the normal lower range for the inversion layer in Hawaii. As we would expect, the mean values are consistently higher than the medians due to the positive skew of the distributions. While the exponential form of the average PW with height is expected, it is perhaps more surprising to see that measures of dispersion of the PW, plotted in 4b as both the arithmetic standard deviation (ASD) and geometric standard deviation (GSD), show linear relationships with elevation. While the ASD declines with elevation indicating that the overall range of the PW is decreasing, the increase of GSD with altitude indicates an increase in the skewness (or asymmetry) of the PW distributions. Plotting the median PW on a log elevation scale to highlight the lower elevation data (Figure 4c) reveals that there is a well-defined change in character at ~540 m. This probably represents the mixed layer: the

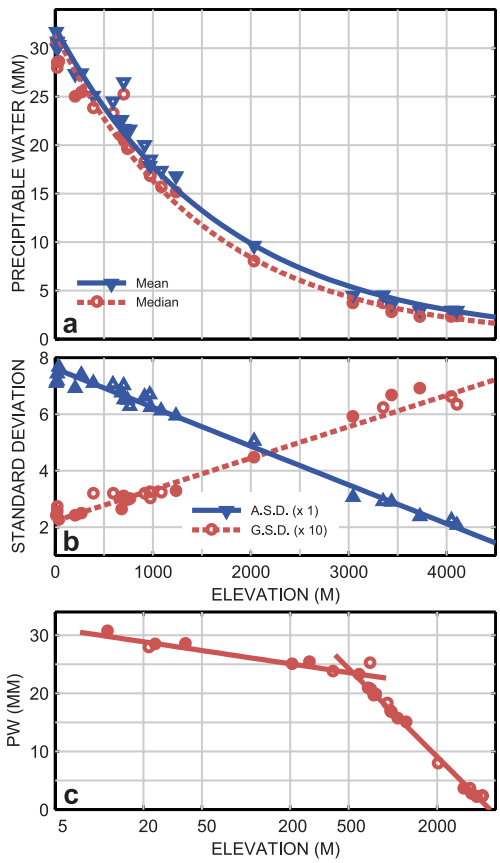

Figure 4. Plots showing the elevation dependence of the PW population averages and dispersions. (a) Mean (triangles) and median (circles) values. Sites with colocated meteorological instruments are shown as solid black. Best fit exponential functions given by a exp($-b$ elev) where a = 32.15, b = .0005911 and a = 31.62, b = .0006613 respectively. The PW scale heights implied are ~1700 m and ~1500 m. (b) Arithmetic (triangles) and geometric (circles) standard deviations. The arithmetic standard deviation (ASD) has units of mm of PW while the GSD is dimensionless. The GSD has been multiplied by 10 to plot on the same range. The best fit linear functions are given by $a + b$ elev where $a = 7.62$, $b = -0.00137$ and $a = 0.222$, and $b = 0.000111$ respectively. (c) Median PW plotted on a log elevation scale. Intersection is at ~544 m.

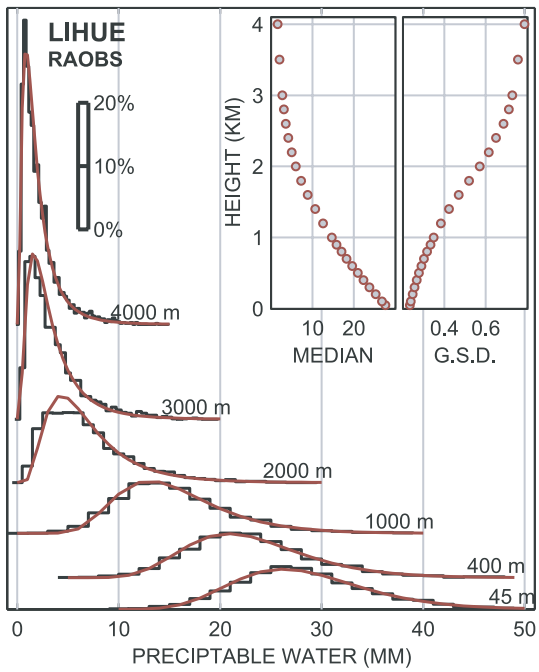

Figure 5. Histograms and lognormal PDF curves for various elevation levels for the RAOBS at Lihue, Kauai. Histograms were generated by interpolating the cumulative PW profiles for launches from 1998 to 2002 to a common set of elevations. Histograms are arbitrarily shifted vertically for clarity. The scale bar indicates the frequency of PW falling within any given 1 mm interval (bar is 1 mm wide). Shown as insets are the distribution medians and geometric standard deviations (GSD) with respect to height.

turbulent, statically unstable layer directly affected by surface heating. Despite the visibility of this low level layer, there is no obvious indication in the data of the inversion layer normally located between 1500 and 2500 m. This is probably a consequence of the inversion layer migrating up and down over this large elevation range. Over the long term, the layer tends be smeared out, producing an average profile with no inversion.

[13] Histograms of PW from the National Weather Service radiosonde site in Lihue, Kauai (Figure 5), offer us the opportunity to confirm our observations with an independent data set. The distributions show the same progression from only slightly skewed from N at the surface toward increasing asymmetry with elevation. As with the GPS results, the medians decline approximately exponentially with elevation, however the abrupt change at ~540 m is less evident and instead there is a more gentle transition. The GSD profile shows an inflection between ~1000 and ~3000 m.

5. Discussion and Conclusions

[14] The lognormal distribution, Λ, appears frequently in meteorology and atmospheric science. Cloud populations [*López*, 1977], precipitation [*Biondini*, 1976] and aerosol optical depth [*O'Neill et al.*, 2000] are all distributed lognormally. *Soden and Bretherton* [1993] note that upper tropospheric relative humidity approximates a lognormal distribution. We have shown that over long periods of time, PW also has a Λ distribution in Hawaii.

[15] Working from first principles, *Raymond* [2000] finds a solution to the moisture transport equations indicating that the relative humidity field should tend toward a lognormal distribution in most cases. He finds that, providing the temperature and pressure fields are sampled from populations having the same expected value and variance, and that the PDF of perturbations from an equilibrium state tend toward N, the PDF of the relative humidity field should tend toward Λ. PW is clearly strongly connected to relative humidity, and as an island environment, Hawaii is likely to satisfy the former condition. Perturbations from equilibrium conditions are generally found to have N PDFs, and the fact that we do observe a lognormal form for PW suggests that this second condition is indeed met.

[16] Precipitable water is typically reported using the arithmetic mean and standard deviation which are closely associated with the normal distribution (see for example *Ross and Rosenfeld* [1996]) and is often modeled, for example in GPS processing, with an underlying assumption that the stochastic process is sampling a normal distribution. For some cases this might not lead to any serious consequences: the distribution for PGF4 for example, while being clearly lognormal, is not strikingly different from normal and so the arithmetic mean and standard deviation may reasonably describe the distribution. For the higher elevation sites however, as the skewness becomes more and more pronounced, it becomes more important to use the more appropriate statistics.

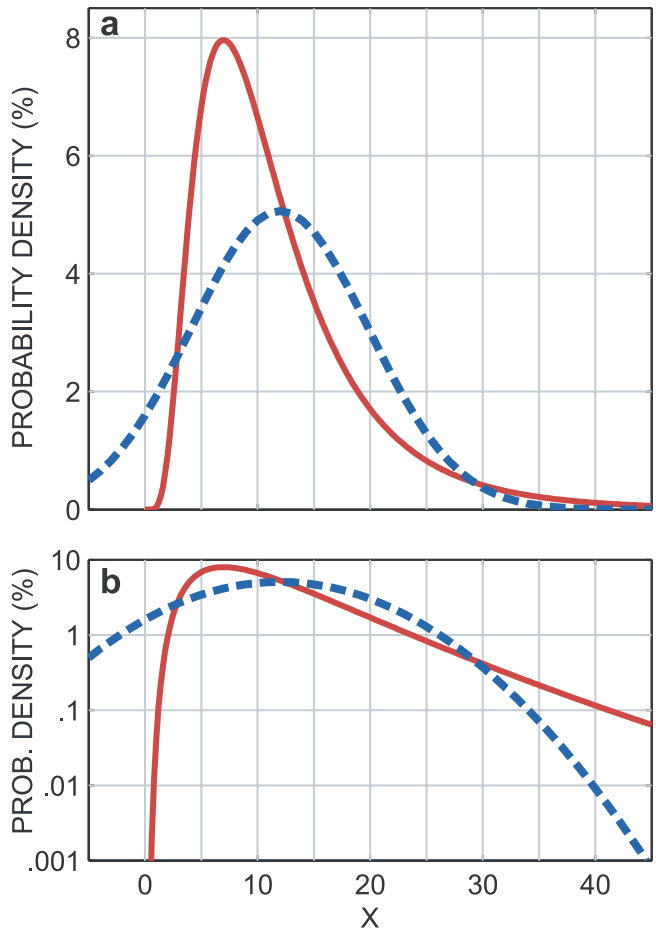

Figure 6. PDF for a Λ variate (solid line) with $M = 10$ and $s = 0.6$. The N PDF derived from these values is shown by the dashed line. Y-axis shows the percent-likelihood of the variate falling with any given 1 mm range. (a) Linear Y-axis. (b) Logarithmic Y-axis to emphasize the low probability values.

[17] Comparing theoretical PDFs for Λ and N (Figure 6) it is clear that assuming N when the distribution is actually Λ will cause the analyst to underestimate the incidence of values slightly lower than the mean, and overestimate the incidence of values near zero and slightly higher than the mean. Although the incidence of very high values is low for both PDFs, Λ predicts very high values as being many times more likely than is expected from N. For example, in Figure 6 an observed value of 40 is predicted as an order of magnitude more likely if the variate follows Λ rather than N.

[18] For all the lower elevation sites, even though the histograms tend to zero occurrence at some value greater than zero, no threshold value was needed (i.e., $t = 0$) in order to generate the theoretical PDF curves. For example, PGF2 and PGF4 have zero occurrence below ~5 mm and ~10 mm respectively yet the theoretical PDFs with $t = 0$ fit very well. This is aesthetically pleasing as zero is obviously the absolute lower bound for PW so it seems intuitive that zero, rather than some higher value, should be the threshold value for PW in all cases. For MKEA (and the other extremely high elevation sites not shown), however, the Λ PDF fits a little less well than for the lower sites. This is due, at least in part, to the PW estimates drifting slightly negative. These estimates are clearlyincorrect, and may be caused by any combination of a number of possible sources: a pressure bias, a small error in the pressure to hydrostatic delay conversion, problems with the mapping function used to calculate the equivalent zenith delay for each satellite elevation angle. Because PW is low for high elevation sites, the impact of any of these

biases is magnified. We used a threshold value of −1 mm in order to get a reasonable fit for the derived PDF for this site and others at similar elevations. One interpretation of this value is as a bias for the entire series. Alternatively it is possible that the median is not significantly biased, but rather that the effects suggested above are skewing the very dry estimates. The observation that the fit is poorest for that section of the histogram close to, and less than, zero seems to support this interpretation.

[19] The RAOBS data from Lihue are well matched by a family of Λ distributions. The least well modeled histogram is at 2000 m, probably reflecting the influence of the inversion layer which is typically associated with clouds. The smoothing of the abrupt change at ∼540 m in the medians' profile is probably a consequence of interpolating the irregularly reported levels to the common set of elevations.

[20] We have shown that precipitable water in Hawaii is best described by a lognormal probability distribution. We suggest that models that include parameters for PW or ZWD might best be parameterized assuming a Λ distribution. As the logarithm of a Λ variate is an N variate, it may be most convenient to simply parameterize the model using the logarithm of PW. Similarly, we suggest it is appropriate to use, or at least include, the median (geometric mean) and geometric standard-deviation when reporting and comparing PW. Although the differences between these parameters and the more commonly reported arithmetic mean and standard deviation may be relatively small, there is a risk, due to the correlation between the arithmetic mean and standard-deviation in a lognormally distributed variable, that using the wrong statistics may lead to errors in interpretation. In particular, treating PW as an N distribution may lead to an underestimate of the probability of extremely moist events.

Acknowledgments

[21] We would like to dedicate this paper to the late William Raymond who first recognized the theoretical connection between the lognormal distribution and moisture flux. We thank Craig Thompson, John Beavan and an anonymous reviewer for their comments which helped to improve the manuscript. We thank Stanford/HVO-USGS for access to data from their GPS sites. The National Park Service Air Resources Division provided some of the surface meteorological data used and the NCEP Reanalysis data was provided by the NOAA-CIRES Climate Diagnostics Center, Boulder, Colorado, from their Web site at http://www.cdc.noaa.gov/. This is SOEST contribution number 6167.

References

Aitchison, J., and J. A. C. Brown, *The Lognormal Distribution*, Cambridge Univ. Press, New York, 1957.

Anthes, R. A., Regional models of the atmosphere in middle latitudes, *Mon. Weather Rev.*, *111*, 1306–1335, 1983.

Bar-Sever, Y. E., P. M. Kroger, and J. A. Borjesson, Estimating horizontal gradients of tropospheric path delay with a single GPS receiver, *J. Geophys. Res.*, *103*, 5019–5035, 1998.

Bevis, M., S. Businger, T. A. Herring, C. Rocken, R. A. Anthes, and R. H. Ware, GPS meteorology: Remote sensing of atmospheric water vapor using the Global Positioning System, *J. Geophys. Res.*, *97*, 15,787–15,801, 1992.

Biondini, R., Cloud motion and rainfall statistics, *J. Appl. Meteorol.*, *15*, 205–224, 1976.

Davis, J. L., Atmospheric water-vapor signals in GPS data: Synergies, correlations, signals, and errors, *Phys. Chem. Earth*, *26*, 513–522, 2001.

Duan, J., et al., GPS meteorology: Direct estimation of the absolute value of precipitable water, *J. Appl. Meteorol.*, *35*, 830–838, 1996.

Emardson, T. R., G. Elgered, and J. M. Johansson, Three months of continuous monitoring of atmospheric water vapor with a network of Global Positioning System receivers, *J. Geophys. Res.*, *107*, 1807–1820, 1998.

Evans, M., N. Hastings, and B. Peacock, *Statistical Distributions*, 170 pp., John Wiley, New York, 1993.

Fang, P., M. Bevis, Y. Bock, S. Gutman, and D. Wolfe, GPS meteorology: Reducing systematic errors in geodetic estimates for zenith delay, *Geophys. Res. Lett.*, *25*, 3583–3586, 1998.

Gutman, S. I., and S. G. Benjamin, The role of ground-based GPS meteorological observations in numerical weather prediction, *GPS Solutions*, *4*, 16–24, 2001.

Hogg, D. C., F. O. Guiraud, and M. T. Decker, Measurement of excess transmission length on earth-space paths, *Astron. Astrophys.*, *95*, 304–307, 1981.

King, R. W., and Y. Bock, Documentation for the GAMIT GPS analysis software version 10.03, Mass. Inst. Technol., Cambridge, 2000.

Lichten, S. M., and J. S. Border, Strategies for high precision Global Positioning System orbit determination, *J. Geophys. Res.*, *92*, 12,751–12,762, 1987.

López, R. E., The lognormal distribution and cumulus cloud populations, *Mon. Weather Rev.*, *105*, 865–872, 1977.

O'Neill, N. T., A. Ignatov, B. N. Holben, and T. F. Eck, The lognormal distribution as a reference for reporting aerosol optical depth statistics: Empirical tests using multi-year, multi-site AERONET sunphotometer data, *Geophys. Res. Lett.*, *27*, 3333–3336, 2000.

Raymond, W. H., Moisture advection using relative humidity, *J. Appl. Meteorol.*, *39*, 2397–2408, 2000.

Ross, R. J., and S. Rosenfeld, Estimating mean weighted temperature of the atmosphere for Global Positioning System applications, *J. Geophys. Res.*, *102*, 21,719–21,730, 1997.

Saastamoinen, J., Atmospheric correction for the troposphere and stratosphere in radio ranging of satellites, *The Use of Artificial Satellites for Geodesy*, *Geoph. Monogr. Ser.*, *15*, 247–251, 1972.

Soden, B. J., and F. P. Bretherton, Upper tropospheric humidity from the GOES 6.7-μm channel: Method and climatology for July 1987, *J. Geophys. Res.*, *98*, 16,669–16,688, 1993.

Starr, D. O., and S. H. Melfi (Eds.), The role of water vapor in climate, *NASA Publ. 3120*, 50 pp., NASA, Easton, Md., 1991.

Tregoning, P., R. Boers, D. O'Brien, and M. Hendy, Accuracy of absolute precipitable water vapor estimates from GPS observations, *J. Geophys. Res.*, *103*, 28,701–28,710, 1998.

Treuhaft, R. N., and G. E. Lanyi, The effect of the dynamic wet troposphere on radio interferometric measurements, *Radio Sci.*, *22*, 251–265, 1987.

Williams, S., Y. Bock, and P. Fang, Integrated satellite interferometry: Tropospheric noise, GPS estimates and implications for interferometric synthetic aperture radar products, *J. Geophys. Res.*, *103*, 27,051–27,067, 1998.