Brief Communication: Likelihood of societal preparedness for global change: trend detection

R. M. Vogel[1], A. Rosner[2], and P. H. Kirshen[3]

[1]Department of Civil and Environmental Engineering, Tufts University, Medford, MA 02155, USA
[2]US Geological Survey, Conte Anadromous Fish Research Laboratory, One Migratory Way, P.O. Box 796,
Turners Falls, MA 01376-0796, USA
[3]Environmental Research Group, Civil Engineering Department & Institute for the Study of Earth, Oceans and Space,
University of New Hampshire, Durham, NH, USA

Correspondence to: R. M. Vogel (richard.vogel@tufts.edu)

Abstract. Anthropogenic influences on earth system processes are now pervasive, resulting in trends in river discharge, pollution levels, ocean levels, precipitation, temperature, wind, landslides, bird and plant populations and a myriad of other important natural hazards relating to earth system state variables. Thousands of trend detection studies have been published which report the statistical significance of observed trends. Unfortunately, such studies only concentrate on the null hypothesis of "no trend". Little or no attention is given to the power of such statistical trend tests, which would quantify the likelihood that we might ignore a trend if it really existed. The probability of missing the trend, if it exists, known as the type II error, informs us about the likelihood of whether or not society is prepared to accommodate and respond to such trends. We describe how the power or probability of detecting a trend if it exists, depends critically on our ability to develop improved multivariate deterministic and statistical methods for predicting future trends in earth system processes. Several other research and policy implications for improving our understanding of trend detection and our societal response to those trends are discussed.

1 Introduction

Human impacts on the earth system are now so widespread that it is difficult to find a location that is not impacted by the interaction among human and natural earth system processes (Palmer, 2004; Vörösmarty et al., 2004; Barnosky et al., 2012; Röckström et al., 2009). Human impacts are caused by population growth along with its associated resource consumption, habitat transformation and fragmentation, energy consumption and production and their associated impacts on earth and atmospheric processes (Barnosky et al., 2012). Röckström et al. (2009) define planetary boundaries as the safe operating space for humanity with respect to biophysical processes within the earth system. They argue that human impacts are now so pervasive that at least three of nine planetary boundaries have now been crossed, relating to climate change, biodiversity loss and the nitrogen and phosphorus cycles. Most fields relating to natural hazards and earth system science now have review articles devoted to trends which have been observed in the dozens of key state variables of interest which have been tracked over time. "The only way to figure out what is happening to our planet is to measure it, and this means tracking changes decade after decade, and poring over the records (Keeling, 2008)". Better understanding and prediction of trends in natural hazards and earth system state variables is crucial for helping our society make good decisions which may lead to preparedness in countering trends. For example, understanding trends in demographic, climatic and hydrologic variables is central to enable society to make sensible investments in infrastructure in order to protect against future inland and coastal flood hazards within a local and global earth system which is subject to continuous restructuring and evolution. Many other analogous examples could be given for other earth system state variables.

Studies which seek to identify trends in natural hazards and earth systems signals are now widespread, including river discharge, ocean levels, landslides, bird and plant populations, temperature, snow cover, precipitation, wind, and air, soil and water pollution levels as well as many other important earth system state variables. All of the many previous studies we have reviewed that have sought to determine whether a trend exists in earth system processes have employed a null hypothesis, H_o, of no trend and most have chosen an associated significance level of $\alpha = 0.05$. A significance level of 0.05 implies that if there really is no trend (that is assumption of H_o), we will only (mistakenly) report trends 5 % of the time. The societal consequences of making such a mistake is that we will prepare for a trend even when it does not exist, which we term over-preparedness. Shouldn't society also be interested in the likelihood of under-preparedness? Surely there are situations in which society will regret having been under-prepared for consequences of events that could have been avoided.

A statistical analysis of a null hypothesis of no trend, termed Null-Hypothesis Significance Testing (NHST), focuses only on our understanding of conditions of no trend, because all such hypothesis tests were derived under conditions of no trend. Thus the alternative hypothesis, H_A, when trends do exist, is usually ignored along with its probability of occurrence known as the probability of a type II error, termed β. The decision matrix for the general trend detection decision problem is depicted in Fig. 1. In this context statisticians would define the term "power" as the likelihood of detecting a trend when it exists. Of particular concern to us are the type II errors, which are entirely out of our control, because it is only the probability of a type I error α that can be specified in a hypothesis test. Here type II errors (under-preparation) involve significant societal consequences because they imply no societal response is necessary when one is warranted.

Numerous fields including psychology, economics, social sciences, meteorology and medical research have called into question the value of NHST tests due to its focus on its dependence upon a single, often arbitrary, significance level α (Ziliak and McCloskey, 2008; Cohen, 1994; and Nicholls, 2000). Cohn and Lin (2005) stated these concerns succinctly when they said: "Because statistical tests are proof by contradiction, any inconsistency between the null hypothesis and the natural system can itself lead to rejection of the null hypothesis." Concerns over the use of NHST are now widespread, though remarkably, none of those studies we have reviewed dwell on the most important criticism of all, that of ignoring the probability of type II errors, the central theme of this commentary.

Criticisms about NHST are of vital concern to the fields of geophysics, climate science, and water resources engineering, where the trend analysis could have an impact on major infrastructure decisions. It is only very recently and rarely that researchers have raised concern over the importance and

	No Trend H_o	Trend H_A
No Societal Response	$1 - \alpha$	β Type II Error (under-preparedness)
Societal Response	α Type I Error (over-preparedness)	$1 - \beta$ Power

Fig. 1. Decision matrix for the general trend detection decision problem, with null hypothesis H_o and alternate hypothesis H_A shown.

impacts of type II errors in the climate and hydrologic sciences (Cohn and Lins, 2005; Trenberth, 2011; Morin, 2011; Ziegler et al., 2003, 2005). Though those studies discussed the importance of considering type II errors in the analysis of trends, they did not consider the resulting impacts on infrastructure decisions and societal preparedness, as is the focus here. A type II error in the context of an infrastructure decision implies under-preparedness, which is often an error much more costly to society than the type I error (over-preparedness), which the NHST focuses on. Note that type II errors corresponding to under-preparedness are paramount, even in a stationary world as was rigorously shown by Stedinger (1982) for risks posed by floods.

For example, the physical implication of a Type I or over-preparedness error in adaptation decisions for flood management is wasted money on unneeded infrastructure. The physical repercussions of a Type II or under-preparedness error, on the other hand, are major flood damages due to inadequate protection. Decision-makers are poorly served by statistical and/or decision methods that do not carefully consider both sources of error, which is a central point of this commentary.

2 The likelihood of societal preparedness for global change

Societal planning in the context of natural hazards depends critically on our ability to detect change when it exists, thus it is important to understand the likelihood of both under- and over-preparedness. In this section we approximate both the type I and II error probabilities associated with trend tests in an effort to acknowledge the tremendous uncertainty associated with our ability to discern trends from other natural inherent properties of earth system signals, such as persistence (Cohn and Lins, 2005) and complications due to seasonality, censoring, change points and other issues (see Helsel and Hirsch, 2002; Kropp and Schellnhuber, 2011). Helsel and Hirsch (2002) provide a general background on trend tests
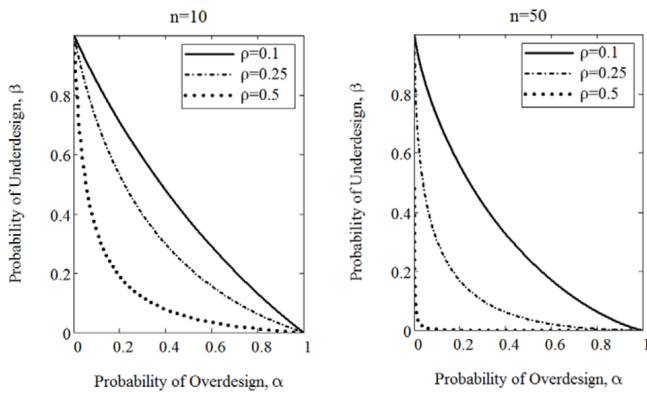
Fig. 2. Relationship between probability of societal under- and over-preparedness, β and α, respectively, as a function of the goodness-of-fit of the trend model ρ, and the length of record n, used to fit the trend model.

and how to improve their power, given the tremendous challenges associated with distinguishing between trends, seasonality, and persistence. It can be quite challenging to discern stochastic persistence from a deterministic trend. For example Cohn and Lins (2005) found that when an inappropriate trend test is used, the existence of long-term persistence in a stochastic process can lead one to reject the null hypothesis (conclude that a trend exists) when no trend is present. Similarly, Douglas et al. (2000) document how ignoring cross-correlation among samples can lead one to conclude that many more samples exhibit trends than actually do exhibit trends.

We employ a linear regression model $y = \beta_0 + \beta_1 x + \varepsilon$ to characterize trends in some earth system state variable of interest (or a transformation thereof, e.g. taking logs) as a function of some other explanatory variable(s) x. Such a model would reflect the conditional mean of the dependent variable y as a function of some other measurable system state, variable x, which ideally reflects a physical dependency on the dependent variable over time. The explanatory variable x might be time, carbon dioxide concentration, impervious cover or other surrogates, and without any loss of generality, our discussions and conclusions apply equally to multiple linear regression models, which have several explanatory variables. The trend hypothesis test is based on the theoretical statistical properties of the estimate of the slope term β_1. For example, Vogel et al. (2011) found that a linear model relating the logarithm of instantaneous annual maximum streamflow to its year of occurrence provided an excellent approximation for thousands of river gauges across the continental US. Even for highly nonlinear trends, ordinary least squares (OLS) regression can often provide a good approximation by employing the "ladder of powers" to linearize the relationship. Mosteller and Tukey (1977) provide a guide to selecting appropriate (and possibly different) power transformations of y and x using a plot of y versus x and

their so-called "bulging rule" (also see Fig. 9.5 in Helsel and Hirsch, 2002). Given the power transformation x^θ, and y^θ, going up the "ladder of powers" corresponds to setting $\theta > 1$ (i.e. x^2, x^3, etc), and going down the ladder of powers means setting $\theta < 1$, (i.e. $\ln(x)$, $1/x$, $\sqrt{x}$, etc.).

Interestingly, even though exact analytical expressions exist for computing the power $(1 - \beta)$ of a trend test based on the use of OLS regression estimates of the trend term in a linear model, we found it quite difficult to locate textbooks or primer papers which document such analyses. This is especially surprising given the widespread use of linear regression for performing trend analyses. Lettenmaier (1976) and Dupont and Plummer (1990, 1998) describe an analytical calculation of the type II error probability (β) associated with our estimate of the slope term β_1 for a linear regression. The trend test amounts to a Student's t test on the estimated value of β_1 in a simple linear regression based on a sample of length n. Given the null hypothesis $H_o : \beta_1 = 0$ versus the one-sided alternative hypothesis $H_A : \beta_1 > 0$ one can estimate the probability of a type I error, α, using $P\left[T_{n-2} \geq t\right]$ where T_{n-2} denotes the Student's t random variable with $n - 2$ degrees of freedom and $t = \hat{\beta}_1 / \hat{\sigma}_{\hat{\beta}_1}$ where $\hat{\beta}_1$ is the OLS estimate of the trend slope and $\hat{\sigma}_{\hat{\beta}_1}$ is the standard deviation of that estimate. Similarly, the probability of the type II error β corresponding to a given value of α can be estimated using $\beta = P\left(T_{n-2} \leq \left(t_{1-\alpha,n-2} - \delta\sqrt{n}\right)\right)$ where $\delta = 1 / \sqrt{\frac{1}{\rho^2} - 1}$, where ρ is the Pearson product moment correlation coefficient between x and y and $t_{\alpha,n-2}$ is that value of a Student's t random variable with $n - 2$ degrees of freedom and with an exceedance probability of α. The values of α and β are inversely related to each other as shown in Fig. 2. In fact, the relationship between α and β only depends on the value of the sample size, n, and the correlation coefficient, ρ. Note that the dimensional trend term β is related to the nondimensional correlation ρ via the relation $\beta = \rho\sigma_y / \sigma_x$ where σ_x and σ_y are the standard deviation of x and y, respectively. No correlation implies no trend ($\beta \to 0$ as $\rho \to 0$) and a perfect correlation $\rho \to 1$ implies a trend term equal to $\beta = \sigma_y / \sigma_x$.

Recall from Fig. 1 that the values of α and β may be interpreted as the probability of societal over- and under-preparedness, respectively. As expected, we observe in Fig. 2 that to obtain a very low probability of under-preparedness, one must either accept a fairly high probability of over-preparedness, or, if the value of n and ρ are large, both probabilities can be quite low. This result highlights the fact that the only way to reduce both the under- and over-preparedness probabilities is by either increasing the value of ρ through improvements in our ability to perform trend detection, attribution and prediction or by simply waiting long enough (increasing n).

It must be emphasized that the over- and under-design probability estimates, α and β, respectively, are based on past

observations and any extrapolation of past trends into the future should be accompanied by associated prediction intervals which account for the increasing uncertainty in future trends due to data, model and parameter uncertainty. Furthermore, we recommend the use of physically based models for forecasting future trends with very careful attention paid to the stochastic properties of the error terms (see Vogel, 1999). Nevertheless, the expression for the under-design probability β shown above and illustrated in Fig. 2 depends critically on the goodness-of-fit ρ^2 (or equivalently R^2 for multivariate models), associated with such trend models, thus reductions in β will come from improved trend models.

3 Conclusions

It has long been known (Matalas et al., 1982) that "human activity is inseparable from the natural system." To advance our understanding of and responses to the anticipated trends in natural hazards resulting from such coupled systems, we envision the following research and policy needs:

1. *Develop new statistical hypothesis tests which are responsive to societal needs*
   We have documented in Fig. 1, that the most common null hypothesis concerning trends in earth system state variables is that of no trend. This is probably because this null hypothesis is the one that is most commonly reported in statistics textbooks, and is a good place to begin such analyses. However, as Trenberth (2011) has recently argued, we may already have enough evidence of earth system changes that perhaps it is time to reverse the null hypothesis so that the type I error is defined by the error of most concern to society: under-preparedness. Such an approach involves the derivation of new hypothesis tests which will likely require support and input from the statistical sciences.

2. *Ensure lasting agency commitments to observational programs*
   Public agencies must focus on the continuity of data collection and data management as the essential basis for evaluating change. "Modeling should be used to synthesize observations; it can never replace them. In a nonstationary world, continuity of observations is crucial." (Milly et al. 2008) It is imperative that we increase or at least maintain long-term data collection and use those data to help understand previous earth system changes, thus improving our ability to predict future changes.

3. *Improvements in trend detection, attribution and prediction*
   Importantly, this commentary has shown in Fig. 2 that improvements in societal preparedness against future hazards are likely to come from associated improvements in our ability to predict future changes in earth system state variables. This point was shown quantitatively in Fig. 2, because reductions in the likelihood of both under- and over-preparedness errors can only result from increases in either the goodness of fit of trend models as measured by the correlation coefficient ρ, or by waiting for additional information (increasing sample size, n). Our ability to distinguish stochastic persistence from deterministic trends is in its infancy (Cohn and Lins, 2005). For individual samples, the existence of skewness, serial correlation, periodicities and change points confound our ability to detect, attribute and predict deterministic trends (Khaliq et al., 2009). Spatial correlation further confounds our ability to estimate the overall field significance which results from combining multiple individual hypothesis tests (Douglas et al., 2000). How one actually evaluates the overall field significance associated with multiple hypothesis tests (termed multiple comparison procedures by statisticians) is a topic in need of further attention (see Vogel et al., 2009, Sects. 6.4–6.8). Helsel and Hirsch (2002), Khaliq et al. (2009); Kropp and Schellnhuber (2011) and Sonali and Kumar (2013) provide an overview of recent innovations in both parametric and nonparametric trend detection methods with attention given to most of the above mentioned complications involving detection of trends. Earth systems evolve over space and time, thus new theory and practical algorithms are needed to address long term social and physical drivers and feedbacks. New exploratory and statistical tools are needed to sharpen our insights into the emergent properties of such systems, and to guide modeling and prediction.

4. *Improve education in statistics in the fields of natural hazards and earth system sciences*
   We have highlighted that unlike the medical sciences (Dupont and Plummer, 1990, 1998), earth system science fields have not focused enough attention on the important concept of power and type II errors when performing trend and other hypothesis tests. Why is this so? Could it be because most earth system scientists, whose focus relates to data analysis have had only one course in statistics at best? The first author has taught a course in environmental statistics for over a decade and noticed in this second course in statistics that the only way for students to truly understand hypothesis tests is for them to derive one themselves and to evaluate the resulting power of the test to discriminate against important alternative hypotheses. Such analyses would be difficult in a first course in statistics. Greater attention should be given to education of earth systems scientists in the discipline of the theory of data, known as statistics. Surely those whose work is devoted to the collection, management and analysis of data should have a deep foundation in the theory of such information, a field known as statistics.

The central task facing earth system scientists whose focus is on natural hazards, is to help inform societal decisions about water, energy, geophysical and ecosystem management that will benefit the economic, social, and spiritual needs of future generations. The study of change is at the very core of our message, just as change must be at the very core of how we approach the resource management challenges of the future (Vogel, 2011).

References

Barnosky, A. D., Hadly, E. A., Basompte, J., Berlow, E. L., Brown, J. H., Fortelius, M., Getz, W. M., Harte, J., Hastings, A., Marquet, P. A., Martinez, N. D., Mooers, A., Roopnarine, P., Vermeij, G., Williams, J. W., Gillespie, R., Kitzes, J. U., Marshall, C., Matzke, N., Mindell, D. P., Revilla, E., and Smith, A. B., Approaching a state shift in Earth's biosphere, Nature, 486, 52–58, doi:10.1038/nature11018, 2012.

Cohen, J.: The earth is round ($p < .05$), American Psychol., 49, 997–1003, 1994.

Cohn, T. A., and Lins, H. F.: Nature's style: Naturally trendy, Geophys. Res. Lett., 32, 1–5, 2005.

Douglas, E. M., Vogel, R. M., and Kroll, C. N.: Trends in Flood and Low Flows in the United States: Impact of Spatial Correlation, J. Hydrol., 1–2, 90–105, 2000.

Dupont, W. D. and Plummer Jr., W. D.: Power and Sample Size Calculations A Review and Computer Program, Controlled Clinical Trials, 11, 116–128, 1990.

Dupont, W. D. and Plummer, W. D.: Power and Sample Size Calculations for Studies Involving Linear Regressionm Controlled Clinical Trials, 19m 589–601, 1998.

Helsel, D. R. and Hirsch, R. M.: Statistical Methods in Water Resources, US Geological Survey, Techniques of Water-Resources Investigations, Book 4, Chapter A3–12, 2002.

Keeling, R. F.: Atmospheric science – recording the Earth's vital signs, Science, 319, 1771–1772, 2008.

Khaliq, M. N., Ouarda, T. B. M. J., Gachon, P., Sushama, L., and St-Hilaire, A.: Identification of hydrological trends in the presence of serial and cross correlations: A review of selected methods and their application to annual flow regimes of Canadian rivers, J. Hydrol., 368, 117–130, 2009.

Kropp, J. P. and Schellnhuber, H. J.: In Extremis: Disruptive Events and Trends in Climate and Hydrology, Springer-Verlag Berlin Heidelberg, 400 pp., 2011.

Lettenmaier, D. P.: Detection of trends in water quality data from records with dependent observations, Water Resour. Res., 12, 1037–1046, 1976.

Matalas, N. C., Landwehr, J. M., and Wolman, M. G.: Prediction in water management, scientific basis of water management, National Academy Press, Washington DC, 118–122, 1982.

Milly, P. C. D., Betancourt, J., Falkenmark, M., Hirsch, R. M., Kundzewicz, Z. W., Lettenmaier, D. P., and Stouffer, R. J.: Stationarity is dead: whither water management?, Science, 319, 573–574, 2008.

Morin, E.: To know what we cannot know: Global mapping of minimal detectable absolute trends in annual precipitation, Water Resour. Res., 47, 1–9, 2011.

Mosteller, F. and Tukey, J. W.: Data Analysis and Regression, Addison-Wesley Publishers, Menlo Park, CA, 588 pp., 1977.

Nicholls, N.: The insignificance of significance testing, Bull. Am. Meteorol. Soc., 81, 981–986, 2000.

Palmer, M., Bernhardt, E., Chornesky, E., Collins, S., Dobson, A., Duke, C., Gold, B., Jacobson, R., Kingsland, S., Kranz, R., Mappin, M., Martinez, M. L., Micheli, F., Morse, J., Pace, M., Pascual, M., Palumbi, S., Reichman, O. J., Simons, A., Townsend, A., and Turner, M.: Ecology for a crowded planet, Science, 304, 1251–1252, 2004.

Röckström, J., Steffen, W., Noone, K., Persson, A., Chapin III, F. S., Lambin, E. F., Lenton, T. M., Scheffer, M., Folke, C., Schellnhuber, H. J., Nykvist, B., de Wit, C. A., Hughes, T., van der Leeuw, S., Rodhe, H., Sorlin, S., Snyder, P. K., Costanza, R., Svedin, U., Falkenmark, M., Karlberg, L., Corell, R. W., Fabry, V. J., Hansen, J., Walker, B., Liverman, D., Richardson, K., Crutzen, P., and Foley, J. A.: A safe operating space for humanity, Science, 461, 472–475, 2009.

Sonali, P. and Kumar, D. N.: Review of trend detection methods and their application to detect temperature changes in India, J. Hydrol., 476, 212–227, 2013.

Stedinger, J. R.: Design events with specified flood risk, Water Resour. Res., 19, 511–522, 1982.

Trenberth, K. E.: Attribution of climate variations and trends to human influences and natural variability, WIREs Climate Change, 2, 925–930, 2011.

Vogel, R. M.: Stochastic and Deterministic World Views, J. Water Resour. Plan. Manage., 125, 311–313, 1999.

Vogel, R. M.: Hydromorphology, J. Water Resour. Plan. Manag., Editorial, 137, 147–149, 2011.

Vogel, R. M., Hosking, J. R. M., Elphick, C. S., Roberts, D. L., and Reed, J. M.: Goodness-of-fit of Probability Distributions for Sightings as Species Approach Extinction, Bull. Math. Biol., 71, 701–719, doi:10.1007/s11538-008-9377-3, 2009.

Vogel, R. M., Yaindl, C., and Walter, M.: Nonstationarity: Flood Magnification and Recurrence Reduction Factors in the United States, J. Am. Water Resour. Assoc., 47, 464–474, 2011.

Vörösmarty, C., Lettenmaier, D., Leveque, C., Meybeck, M., Pahl-Wostl, C., Alcamo, J., Cosgrove, W., Grassl, H., Hoff, H., Kabat, P., Lansigan, F., Lawford, R., and Naiman, R.: Humans transforming the global water system, EOS, 85, 509–520, 2004.

Ziegler, A. D., Sheffield, J., Maurer, E. P., Nijssen, B., Wood, E. F., and Lettenmaier, D. P.: Detection of Intensification in Global- and Continental-Scale Hydrological Cycles: Temporal Scale of Evaluation, J. Climate, 16, 535–547, 2003.

Ziegler, A. D., Maurer, E. P., Sheffield, J., Nijssen, B., Wood, E. F., and Lettenmaier, D. P.: Detection time for plausible changes in annual precipitation, evapotranspiration, and streamflow in three Mississippi river sub-basins, Climatic Change, 72, 17–36, 2005.

Ziliak, S. T. and McCloskey, D. N.: The cult of statistical significance: how the standard error costs us jobs, justice, and lives, University of Michigan Pressm Ann Arbor, 2008.