<u>Contrasting Messages: Model-Simulated vs. Observation-Based Data</u>

How attempts to disentangle forced and intrinsic components of climate variability revealed pronounced contrasts between model-simulated and observed data

by

M.G. Wyatt

<u>*Featured Articles*</u>:

Wyatt et al. 2012; Wyatt and Peters 2012; Wyatt and Curry 2014; Kravtsov et al. 2014; Mann et al. 2014

I. Introduction

The observed Northern Hemisphere surface-average temperature trend over the last 100-plus years has increased non-uniformly, its trajectory marked by multiple-decade intervals of strong warming alternating with multiple-decade intervals of stalled warming or slight cooling. Similar behavior emerges in proxy data dating back several centuries (e.g. *Black et al. 1999; Gray et al. 2004*). A question of attribution emerges. What is the source of this non-uniformity? Does the climate system's intrinsic[1] variability influence this low-frequency pattern or are external forcings - both natural and anthropogenic - dominantly in charge?

It is the envisioned relative roles of these two components – intrinsic (I) and forced (F) – that underpin fundamental and divergent views on climate and its sensitivity to external influences. One school-of-thought views multidecadal-scale climate variability as being dominated by external radiative forcing – partly natural, partly anthropogenic - with minimal contribution from the intrinsic component. The other school-of-thought sees a significant role for intrinsic dynamics in multidecadal climate behavior.

For some, these contrasting views regarding forced and intrinsic signals motivate efforts to disentangle the two components so that their relative roles in observed climate variability can be identified. Yet not all research is driven by similar incentive. Some research attempts to separate natural contributions from anthropogenic; while other research targets *timescale* of behavior, and potential relationships among processes exhibiting similar timing. These distinctions in research goals drive selection of strategies.

There is nothing inappropriate about this tendency for one's study focus to influence choice of methodology; it is this way because different techniques emphasize different features and/or dynamics within a data set. Regardless of the analytical strategy used, problems exist. For example, coverage – both temporal and spatial – is far from ideal in long-term climate records. Data sets are more a moving target than a solid representation of past events. There are observational surface-temperature data sets, none of which resemble their original versions. They all have been corrected, modified, and altered numerous times – sometimes for obvious reasons; sometimes for less obvious ones. Controversy surrounds these records[2]. And for many climate indices, records are too short to be informative. One might seek an alternative to counter these limitations.

Modeled data offer potential. But again, this depends upon perspective. If it is assumed that models well represent all climate dynamics, these data may offer an instructive alternative to, or companion to, observational data. Yet which data to use? Each individual computer-model design is tuned with a physics package and subsets of processes and parameterized features that differ slightly from model to model. Input determines output, and the former is based on assumed and/or speculated critical dynamics, estimated parameterizations of processes, feedback responses, and overall climate sensitivity. Output from each individual model design is unique.

For the present, we narrow our focus to isolating F from I.

The points above highlight obstacles inherent in decomposing a climate signal into its intrinsic and forced drivers. Even were we to have access to perfect data, disentangling climate components is steeped in complexity. A simple matter of algebra it is not. If we were to reason that average surface temperature ($T_{average}$) can be described as a combination of a forced (F) component and an intrinsic (I) one, then we are faced with two unknowns and one controversial "known" in a three-part equation: $T_{average} = F + I$, where only $T_{average}$ is "known", and only known according to what data set is chosen. If our goal is to isolate F from I, the dilemma becomes obvious.

A widely held view is that if one could figure out the exact measure of F, the intrinsic component, I, could be ascertained. This assumes the forced component and the intrinsic one imprint contrasting signatures. But what if the frequency of the forced signal is the same as that of the intrinsic one? Such could be the case with frequency entrainment[3] of I by F. How, then, could one tell the temporal difference between them?

II. Overview: Forced versus Intrinsic – A matter of methods and data?

While there are many studies that focus on forced and/or intrinsic components, the studies on which this present memo is based have introduced novel perspectives on the role of internal variability in long-term climate variability, and have identified pronounced spatio-temporal differences between observed and model-simulated data.

III. How to interpret low-frequency climate variability:

Low-frequency behavior is evident in surface temperatures across the Northern Hemisphere, with particular focus on sea-surface temperatures (SSTs) in the North Atlantic. Instrumental and proxy data of the North Atlantic SSTs reflect a multi-decadally repeating pattern. While regularly repeating, it is not strictly periodic (*Vincze and Janosi 2011*). This North Atlantic variability has been termed the Atlantic Multidecadal Oscillation (AMO: *Kerr 2000*). Its timescale of variability is thought to be influenced by the Atlantic sector's Meridional Overturning Circulation (AMOC) (*Knight et al. 2005*) – the ocean's "conveyor belt": a globally encircling, ponderously paced ocean circulation, instrumental in shuttling massive stores of energy along its path, and driven by wind and thermohaline processes. Similar timescales of variability have been identified in different climate patterns across the Northern Hemisphere (*Enfield et al. 2001; Goldenberg et al. 2001; Sutton et al. 2003; Sutton and Hodson 2003, 2005, 2007; Knight et al. 2006*).

These observations have spawned speculation that a North-Atlantic born signature (AMO) is imprinted on the hemispheric (possibly global (*Lee et al. 2011; Feng He 2013*)) climate record. If one assumes a relationship to the AMOC, *and* if one assumes the AMOC to be internally generated, then one might infer the hemispheric (global) temperature signature to contain an intrinsic signal. But how much? Herein lies the debate: Is the observed low-frequency climate behavior of the AMO forced, intrinsic, or both? And if both, in what proportions?

IV. Identifying AMO: **What do we want to know? Behavior or Cause?**

Many methods have been used to identify the multidecadal nature of the AMO: principal component analysis (PCA) (*Parker et al. 2007*), linear detrending (e.g. *Enfield et al. 2001; Knight et al. 2005, 2006, 2009*), and differencing (e.g. *Mann and Emanuel 2006; Trenberth and Shea 2006; Kravtsov and Spannagle 2008; Knight 2009*). If the goal in applying these methods is to highlight the multidecadal nature of AMO by removing the longer-period, secular-scale trend, then linear detrending should serve this purpose. If, on the other hand, the goal is to isolate the intrinsic (I) component from the forced (F), no method is without weakness (*Knight 2009*).

In the case of linear detrending, if the forced signal is, itself, time-varying, then a portion of it will not be removed in the linearly detrended product. Instead, vestiges of it will remain in the residual data, thereby contaminating the residual. Depending on the time-varying structure of the forced signal, the residual's amplitude of variability may be enhanced or dampened. On the other hand, in the case of differencing, there are various versions to consider. In one version, a global signal is subtracted from an Atlantic signal (e.g. *Trenberth and Shea 2006*). In this case, if an Atlantic fingerprint exists within the global signal, the risk of overfitting leads to underestimation of the residual, as a portion of the Atlantic fingerprint is subtracted from itself. And in differencing versions, where modeled terms are used in conjunction with observed data – a semi-empirical approach - (e.g. *Mann and Emanuel 2006; Kravtsov and Spannagle 2008; Knight 2009*), whereby a model-estimated forced signal is generated and subtracted from the observed hemispheric signal, results of the end-product depend upon the modeled forced signal. Forcing profiles, climate sensitivity used to generate them, and successful removal of the model's own internal variability from the simulated data (where applicable) are all subject to assumption and uncertainty. All influence the residual, and it is the interpreted character of the multidecadal component of that residual – the presumed low-frequency intrinsic component - that has ignited debate.

V. Mann et al 2014 – Differencing over detrending, at least for some tasks.

Mann et al. 2014 address the controversy regarding method choice. They make a reasonable case for differencing over linear detrending, if one's goal is component disentanglement.

They make the assumption that the Northern Hemisphere mean surface temperature (NHT) is a hybrid of a forced component (F) and an intrinsic (I) one. Relative roles remain a question. *Mann et al.* suggest that an externally forced component, with both natural and anthropogenic contributions, dominates; while an internally driven one is minor, with limited geographical reach; its source is assumed to be the North Atlantic, i.e. related to the low-frequency component of the AMO.

If the goal is to find support for or against the assumed major role for F, and a complementary role for I, we find ourselves back at the dilemma of disentangling F from I, both components being unknowns. Guided by the perspective of NHT being dominated by the multidecadal timescale of F, *Mann et al.* estimate F using various models that incorporate assumed physics, climate sensitivity, feedbacks, and parameterizations.

To generate these various modeled estimates of the F component projected onto NHT, *Mann et al.* use three approaches; each outcome, slightly different: 1) a simple energy-balance model; 2) an ensemble-average of runs from one individual computer model from the fifth version of the Coupled Model Intercomparison Project (CMIP5); and 3) the multi-model ensemble-average of all runs from all individual computer models of the CMIP5 collection. These variously model-generated estimates are compared to the observed NHT. *Mann et al*. find their model-estimated F versions all are closely correlated to the observed NHT. With this, they move to step two, estimating I - the inferred low-frequency intrinsic component projected onto the NHT.

From the observed average-surface Northern Hemisphere temperatures (NHT), they difference - or subtract - a model-estimated (F) signal. The residual of this differencing procedure is the presumed intrinsic component (I) associated with the particular model-estimated F signal used in the operation. Smoothing the residual with a 50-year low-pass filter generates the low-frequency expression of I. They term this differenced product the "AMO", in reference to the assumed AMOC-driven portion of teleconnected climate variability.

To investigate how methodology affects interpretation of an index's behavior, *Mann et al*. compare their differenced "AMO" to one obtained by linearly detrending the Northern Hemisphere mean temperature data, followed by a 50-year low-pass filter. Contrasts between "AMOs" emerge: The linearly-detrended-NHT version has greater amplitude and 'biased' (shifted) phasing when compared to its differenced-AMO counterpart. *Mann et al* note that this is particularly true for recent years – the 'hiatus' years. In the mid-1990s, the "differenced AMO" began decreasing; while the "detrended AMO" peaked several years to a decade later.

Keep in mind that in the *Mann et al.* study, neither the differenced "AMO", nor the detrended "AMO", are the AMO. In fact, both terms are no more than low-pass-filtered by-products of NHT: the former after differencing; the latter after linearly detrending. "AMO" and AMO are *not* one in the same. AMO refers to variations in a sea-surface-temperature pattern in the North Atlantic. Despite considerable speculation within the climate community regarding AMO mechanism, no conclusion has been made on the exact nature of its driver, be it intrinsic, forced, or both. Yet *Mann et al.* use the term "AMO" interchangeably with the low-frequency intrinsic signal, I. This synonymous use conveys the assumption that low-frequency ocean-circulation patterns in the North Atlantic are intrinsic, the influence of which is teleconnected onto the NHT. The degree of that presumed teleconnected intrinsic influence on NHT is the unknown they seek to quantify.

Invoking coupled model simulations (*Knight et al. 2005*) of AMO variability, *Mann et al.* statistically evaluate their two versions of estimated “AMO” and find that of the two (differenced and detrended), the variance of their differenced-“AMO” amplitude more closely matches the variance derived in the *Knight et al.* model-AMO experiments. Furthermore, evidence is inferred from *Mann et al’s* model-estimated F; i.e. similarities between curves of these F model-estimates and those of the detrended-“AMO”. These likenesses suggest to *Mann et al.* that their detrended-“AMO” curve is driven by vestiges of the estimated forced component left behind after detrending, not by an intrinsic component.

These lines of evidence suggest to *Mann et al*. that the differenced-“AMO” is a more apt representation of the “true AMO” – i.e. *Mann et al.’*s I - than is the linearly detrended calculation. This alerts *Mann et al*. to potential biases in studies that use the detrended term: One study, in particular, serves as their example - the “stadium-wave” study.

Mann et al.: Making (stadium) waves:
Mann et al. cite a study – the “stadium wave” - by *Wyatt et al. (2012)* to make their point. *Mann et al*. aver that *Wyatt et al*. detrended raw indices for the purpose of isolating the intrinsic component. While this is not the case – *Wyatt et al.* detrended to remove the secular-scale trend, thereby highlighting multidecadal and shorter oscillations - the *Mann et al*. challenge motivated new research that illuminated some worthwhile points.

But first things first – what is the stadium wave?

Not to be confused with a real “wave”, the “stadium-wave” is a nickname, a reference to a cheer, of sorts, sent round and round a sports stadium filled with fans who alternately raise arms and lower them – passing the “wave” from one section of people to the next. As the term applies to climate, it refers to a hypothesized hemispherically propagating climate signal – one passed along in a chain-like, lead-lag succession through a synchronized network of diverse geophysical indices. This traveling low-frequency signal, or “wave”, first documented in 20^{th} century observational data, is geographically and sequentially communicated via a “passing-the-baton-like” signal through interactions between ocean, ice, and atmospheric patterns. As this signal courses along through the various climate processes, the multidecadal component of NHT evolves with alternating increasing and decreasing trends. It is this hemispheric multidecadal signal-propagation that gives the hypothesis its name – stadium wave.

Dozens of climate and related geophysical indices exhibit this propagation. Its hypothesized mechanism (*Wyatt and Curry 2014*) is well reflected in 20^{th}-century observed data and the signal is evident in at least 150 years of proxy data (*Wyatt 2012*). Yet *Mann et al.* claim the propagation, identified as the stadium-wave signal, is no more than a false-propagation signal, a statistical artifact – a product of flawed methodology – i.e. linear detrending.

Mann et al. take the reader through a seemingly well structured argument, but there are some assumptions underpinning the assertion that invite additional perspective.

In all this talk of the F and I components projected onto NHT, one might be unaware that views diverge on the manifestation of low-frequency intrinsic signals. Views also diverge on the ability to disentangle these components. Quantification remains elusive, as knowledge on F and I is rooted in assumptions.

In the *Mann et al.* view, differences among climate indices arise from regional stochastic interannual-to-interdecadal processes that imprint a "white noise" signature upon a core composition that is common to all indices. That shared common-core composition can be explained by the NHT index, as presented by *Mann et al*. In this construct, NHT is assumed to comprise an in-phase, stationary forced signal and an associated low-frequency intrinsic one related to dynamics of the AMO. *Mann et al.*'s study suggests that a forced signal impacts all climate indices simultaneously, and a teleconnected intrinsic one stamps its low-frequency signal hemispherically with varying impact. In this view, despite the many assumptions involved in estimating F, and assumptions on how I teleconnects its signature, the differencing method is a reasonable choice for disentangling components.

In contrast to *Mann et al.*'s view, *Wyatt et al.* observe that the multidecadal component of NHT evolves in concert with an interconnected network of synchronized indices, which are hypothesized to have self-organized over time through complex interactions, working collectively to re-distribute heat laterally and vertically, from regions of excess to regions of deficit – a climate network functioning as a multidecadally paced heat pump, of sorts. According to this hypothesis, the AMO sets the multidecadal pace. The hemispheric signal-passing, via ocean-ice-atmosphere linkages (*Wyatt and Curry 2014*), does the rest. The pace, or tempo, is linked intimately to the AMO. AMO is likely driven by both intrinsic and forced dynamics. The propagation, itself, likely is intrinsic only. Forced and intrinsic components are not readily quantified or disentangled in this view.

Keeping the two contrasting views of the 'I' component in mind, we return to methods. *Mann et al*. approach the problem by evaluating uncertainties – fluctuations in phasing and phase offsets (in terms of years) between plotted index pairs. This is an apt test for lead-lag relationships and a good choice to test for robustness of signal propagation.

But there is flaw at the outset. *Mann et al*. contend that the *Wyatt et al.* stadium wave is a plot of the intrinsic components of climate indices. This is incorrect. It is this misconception that underpins the ultimate reason that *Mann et al.* and *Wyatt et al*. come up with different results: *Mann et al.* use modeled data and linear detrending to extract contaminated I components, that when plotted, generate false "stadium-wave" propagation. Conversely, *Wyatt et al*. use observed data and multivariate techniques to extract shared patterns of variability among all network indices. Leading modes of co-variability are identified; their reconstructed components then plotted. When plotted, the reconstructed components generate a "real" stadium wave.

Mann et al. have cited linear detrending as the source of purported false propagation. It is significant to note the different use of this detrending method between the two studies. In *Wyatt et al*., detrending of each individual observed climate index is a data-preparatory first-step used to highlight timescales shorter than century-scale, executed *prior* to the principal step involving collective analysis. In *Mann et al.*, linear detrending of each model-simulated index is the principal method in their "stadium wave" analysis.

Mann et al.'s approach:

To test the stadium-wave network of presumed "intrinsic components", *Mann et al*. create a collection of surrogate climate indices. According to their view that all climate indices share NHT at their core, *Mann et al*. first construct five surrogate NHT indices, each consisting of F and I components.

One part of each NHT is the shared forced signal, which is model-estimated. *Mann et al.* use a simple term-adjustable energy-balance model for this estimate. The second part of each NHT is an alternate realization of the low-frequency intrinsic component. This synthetic realization is derived through use of a statistical model, which, itself, is based on the F-associated "real" I.

To each of the five surrogate NHT indices, *Mann et al*. add a unique alternate realization of Gaussian-distributed white noise to represent index contamination by stochastic regional processes. A statistical model is used to generate random realizations. The three-component end-products are the *Mann et al*. surrogate stadium-wave network members. Each surrogate climate index is linearly detrended, smoothed, and plotted. The detrended surrogate indices generate a "stadium wave", but the propagation signal is statistically insignificant, with phasing uncertainties exceeding phase-spreads. Translated, this means that with each Monte-Carlo generated "stadium wave", large phase shifts, varying phase offsets, and even different sequence orders among indices, can emerge. This is false signal-propagation. *Mann et al.*'s surrogate "stadium wave" is, indeed, an illusion. *Mann et al*. claim their finding falsifies the *Wyatt et al.* stadium wave, as well. But does it?

At first glance, it might appear so. *But on second glance…* It seems not. In fact, it appears that the false nature of the *Mann et al*. "stadium wave" is a result of their single-step detrend methodology applied to a model-simulated data set.

Methods and data – worth a closer look.

When the differencing technique is applied to *Mann et al*.'s F + I surrogate climate indices, the F component common to all indices is eliminated, and a collection of low-pass filtered uncorrelated curves, representing the intrinsic component of each index, is generated. A plot of these reveals no hint of propagation – false or otherwise. On the other hand, when the surrogate climate indices are linearly detrended, vestiges of the "wiggly" F component contaminate the residual. The surrogate residuals retain bits and pieces of shared F. Hence, plots of these low-pass-filtered, detrended surrogates will all have a similar shape – the characteristic "wiggle" of the model-estimated F – yet with offsets in phasing. These phase-shifts are the result of random realizations of the "noise", just as *Mann et al*. explain. Indeed, this is how a false stadium-wave is generated. But does the same false propagation emerge from observed data and multivariate statistical analysis?

Wyatt et al.'s approach:

Wyatt et al. did not assume, nor did they intend, that stadium-wave-network indices were intrinsic components. In fact, motivating their research was the quest to understand observed shared behavior among numerous spatially and dynamically diverse geophysical indices, all beating at a similar multidecadal tempo, yet not in unison. The 20th-century stadium-wave indices – all based on instrumental (observational) data – reflected shared timescale of variability. No disentanglement of forced and intrinsic components was sought.

In *Wyatt et al.*'s applied methodology, climate indices are decomposed into temporal components. Linear detrending each climate index individually is a minor initial step in their strategy; but valuable. It is done for two reasons. One, not all geophysical indices in the stadium-wave network have a secular-scale trend. Detrending puts all indices on the same footing. Two, for those indices that do have a linearly increasing or decreasing component, detrending it eliminates the century-scale signal, highlighting the shorter timescales relevant to the stadium-wave study. Linearly detrended time series are analyzed together, as a collection, with Multiple-Channel Singular Spectrum Analysis (M-SSA). In this way, timescales of variability shared by all indices can be identified.

In *Wyatt et al*. and all subsequent stadium-wave studies, M-SSA identified co-variability among indices. In observed data, the two leading modes of co-variability among all indices exhibited behavior at the multidecadal timescale. This dual-pattern temporal component was extracted. An interannual-to-interdecadal residual remained. The multidecadal component showed two equally dominant modes of variability of similar timescales (~64 years[4]), and in-quadrature with one another, meaning they were indicative of an oscillatory pattern at this time scale. These modes, together, are the stadium-wave signal. Time series of these two distinct modes are combined into a "reconstructed component" (RC) for each index. All reconstructed components are plotted. Plotting reveals an ordered, sequential progression of a multidecadally paced signal, with offset phasing creating a lead-lag relationship between indices representing ocean, ice, and atmospheric processes across the Northern Hemisphere. A mechanism explicating each index-to-index coupling is detailed in *Wyatt and Curry 2014*.

Could different results be due more to different data than to different methodologies? *Wyatt and Peters 2012* set out to investigate that question.

In 2012, *Wyatt and Peters* incorporated model-simulated data in their study; yet no stadium wave emerged. They used model-generated data from the third version of the Coupled Intercomparison Model Project (CMIP3). From the modeled raw variables

(SSTs, SLP, etc), climate indices that were used in the original stadium-wave study were reconstructed. These simulated indices were then treated exactly like the 'real' indices used in *Wyatt et al. 2012*. It was anticipated that these network-indices, constructed from model-simulated raw data, would reflect a propagating signal. They did not.

Twenty-one of 22 models were represented (one model output was corrupted (*http://pielkeclimatesci.wordpress.com/2011/06/28/comments-by-marcia-wyatt-on-cmip-data/*)). Sixty-six runs were processed – the majority with prescribed "business as usual" CO_2 increase; a few runs were pre-industrial control runs. None of the 66 runs produced a stadium wave. Of the models that produced a low-frequency signal, that signal was stationary and in-phase, with no propagation - reminiscent of an externally radiatively forced signal – a pattern in stark contrast to the stadium-wave signature's quasi-periodic oscillatory pair of multidecadal-scale M-SSA modes.

A question surfaces: If it were true that linear detrending, done in conjunction with multivariate analysis techniques, generated a false propagatory signal in *Wyatt et al's* application to instrumental and proxy data, as *Mann et al.* suggest, then why did the same linear-detrending and M-SSA methodology, applied to indices reconstructed from computer model-simulated data - 20^{th} century data *and* pre-industrial model-simulated data - *not* produce the same false propagation (*Wyatt and Peters 2012*)? It produced no propagation at all.

Linear detrending appears not to be the source of contrasting results. Data sets do.

Mounting evidence of inconsistencies between modeled and observed data motivated *Kravtsov et al. 2014* to investigate further. They adapted the *Mann et al*. approach for generating alternate realizations of climate indices - a step designed to test the statistical significance of the stadium wave. A notable difference distinguishes their strategies: *Kravtsov et al*. retained observational data as the bulk of signal information. *Mann et al.* used modeled data as the bulk of signal information.

The *Kravtsov et al.* approach:

With the goal being to test the role of noise in generating phase uncertainties in the propagation alignment, as *Mann et al*. did; and in the spirit of the *Mann et al* methodology, *Kravtsov et al*. generate random surrogates of residual variability. As this residual component is the component that created the false phase offset among *Mann et al.*'s NHT-based surrogate climate indices, *Kravtsov et al*. add random synthetic realizations of this contaminating component to the observed data to be evaluated with multivariate statistical techniques, followed by assessment of statistical uncertainties.

Details follow.

Kravtsov et al. begin with the original *Wyatt et al*. stadium-wave methodology: They 1) linearly detrend observational data; 2) apply M-SSA to the residual to extract the multidecadal components; and 3) separate out the interannual-to-interdecadal-scale residual. 4) With this residual, they construct a statistical model with which to generate a synthetic realization of this component. This would be analogous to *Mann et al*.'s white-noise component. 5) Then index components are re-combined, this time with the century-scale and multidecadal-scale components in observational data plus the synthetic realization of the "noise" component. 6) Next, the three-component mix of data will again be separated by linear detrending and application of M-SSA to the residual. The linear detrend will differ from the previous one, as the data mix differs from the original. 7) The multidecadal modes of index co-variability are identified and combined into reconstructed components. 8) The RCs are plotted. This is the *Kravtsov et al*. surrogate stadium-wave.

For the purpose of estimating uncertainty of phase offsets among surrogate indices, the above-described procedure is repeated 1000 times, in accord with classical Monte-Carlo testing. In this way, sampling differences resulting from random Monte Carlo selections will generate uncertainties – i.e. variations in phasing (in years) for the surrogate indices. The unknown is; will those uncertainties in phasing be larger than the phase-spread between indices? If the answer is yes, the "wave" would be statistically insignificant, i.e. likely a result of random sampling. In such a situation, indices in the stadium wave likely would *not* retain the same sequence order with each new surrogate stadium wave. This would be a false propagation signal, just as the model-simulated data analysis of *Mann et al*. generated. On the other hand, if uncertainties in index-phasing are smaller than the phase spreads between neighboring indices, then the sequence order would remain constant for surrogate stadium waves and it is likely that the "wave" would be statistically significant. The degree of significance depends on how large or small the uncertainties are with respect to phase spreads.

It turns out that uncertainties in phasings of indices within the *Kravtsov et al*. surrogate stadium wave are small, *considerably* smaller than phase spreads between indices. Through application of the Monte-Carlo significance estimation to phasing of surrogate indices, *Kravtsov et al*. formally argue that the likelihood of false stadium-wave propagation is 5% or less – i.e. the "wave" is robustly statistically significant, most unlikely to be the result of random sampling.

In short, the *Mann et al*. argument against the stadium wave does not hold. Why the different results between studies?

Keep in mind, both *Mann et al*. and *Kravtsov et al*. constructed surrogate indices similarly, and both employed linear detrending in their strategies. Differences are key. They lie in two areas: 1) The bulk of data used by *Mann et al*. was model-simulated data; while the bulk of data used by *Kravtsov et al*. was observed data. 2) And while both *Mann et al*. and *Kravtsov et al*. linearly detrended individual index time series, *Mann et al*. did so as their principal analysis step. Their intent was to extract the intrinsic component. In contrast, *Kravtsov et al*. employed the detrend technique as a first step to put all indices on the same footing and to remove the secular-scale trend in order to highlight timescales shorter than century-scale. This was followed by their principal step:

M-SSA, which was applied to all climate indices collectively. The intent was to identify leading modes of *shared* co-variability, not the intrinsic component.

The two studies discussed thus far that consider the stadium wave in context of modeled data – *Wyatt and Peters 2012* and *Kravtsov et al. 2014* – show, via two different approaches, that signals buried in modeled data sets differ from those in observed data sets. This observation prompted *Kravtsov et al*. to probe the spatial signature signals of the data. From what is discussed next, more evidence emerged that the data sets contain different information.

Kravtsov et al. look at spatial signatures:

Kravtsov et al. analyze modeled data generated by the Geophysical Fluid Dynamics Laboratory Coupled Physical Model (GFDL CM3). Results of five historical simulations show an in-phase, stationary collection of "waves" with a period centered on ~75 years. No propagation emerges. The phase and shape of each of the indices in each of the five GFDL CM3 simulated 20th century "stadium waves" are very similar. This similarity of temporal character of individual indices across all five simulations indicates high correlation among individual simulations. In other words, each individual index appears to share a common signal across all model runs. From this, it can be inferred that this signal is forced, not intrinsic. Intrinsic residual variability, defined as being uncorrelated among the five simulations, varies at the interannual-to-decadal time scale.

Furthermore, while two leading modes of multidecadal-scale co-variability are identified in this particular set of model-simulated data, only one mode dominates. This is noteworthy. Recall that in observational data, two leading modes of multidecadal-scale co-variability emerge, yet significantly, and in stark difference to modeled results, each leading multidecadal mode in observed data reflects similar dominance.

Kravtsov et al. evaluate and compare spatial patterns of these two network renditions – modeled and observed. For both the modeled data and observation-based data they define the spatial patterns statistically by regressing the sea-surface-temperature (SST) fields onto normalized sine and cosine predictors. A period of 75 years defines the predictor curves, with the zero-phase set at 1920; the two curves are in-quadrature, with the sine curve well aligned with the 20th century global-temperature wiggle. Roughly, this sine curve also corresponds to the GFDL CM3's stationary, in-phase "wave".

Model-simulated data reveal a "partial fingerprint"…*all sine; no cosine*!

Kravtsov et al. find that the spatial pattern of the linear trend and the sine pattern of the modeled global SST time series are similar to the observational analogues. But the similarities end there. They do not extend to the cosine predictor pattern, which is distinct and pronounced in observations, but not so in the modeled data. Only one pattern emerges in the spatial signature of the modeled data; while two patterns define the spatial character of the multidecadal signal in observed data.

Further breaking down the patterns, *Kravtsov et al.* focus on the fingerprints of individual network-index members. *Kravtsov et al*. show that most indices reconstructed from the GFDL CM3-generated data – e.g. the AMO, the Atlantic SST Dipole, the North Atlantic Oscillation (NAO), the North Pacific Oscillation (NPO), and the Aleutian Low Pressure Index (ALPI) – exhibit variances in the decadal-to-multidecadal range that are far *smaller* than observed variances; in some cases, up to an order of magnitude smaller. This means that these indices play an insignificant, if any, role in the *modeled* multidecadal climate variability. Exceptions to this minor-to-absent participation of modeled indices include the Pacific Decadal Oscillation (PDO) and the NHT. In these two cases, variances are similar to their observed counterparts.

These results suggest that the GFDL CM3's simulated multidecadal signal is dominated by the Pacific sector, and is synchronized to the NHT, with little to no involvement of Atlantic and atmospheric indices. GFDL CM3-modeled multidecadal variability, where the Pacific role dominates, is unlike observed multidecadal behavior, where all discussed indices participate. One mode of variability emerges from simulated multidecadal behavior, with no propagating signal; all simulated indices are in-phase. In contrast, two modes of variability characterize observed data at the multidecadal timescale; observed indices reflect phase shifts – i.e. propagation. The strong similarity of shape and phasing of individual simulated indices across all five realizations suggest a forced response at this timescale.

These fundamental differences in outcome between analyses using observation-based data and analyses using model-simulated data may imply that dynamics operating in the observed multidecadal-scale climate variability are either poorly represented or absent from current model design.

Summary of Kravtsov et al. findings:
In summary, *Kravtsov et al*. use a variety of strategies to examine robustness of the stadium-wave signal. They address *Mann et al*.'s challenge regarding signal propagation and then go further, comparing observed multidecadal behavior with modeled analogues, mapping their hemispheric spatial signatures. They find and/or add support to the following: 1) Propagation alignment of the stadium wave indices does not appear to be a statistical artifact of linear detrending and its documented hemispheric propagation is highly unlikely to be due to random sampling associated with higher-frequency noise; 2) Two leading distinct modes are required to explain the observed multidecadal variability and rationalize the observed stadium-wave propagation. One mode follows the general trend of the multidecadal component of NHT; the other mode evolves in-quadrature with it. These modes are well expressed spatially, with climate indices across the Northern Hemisphere all reflecting the multidecadal signature. 3) This stands in contrast to signals within modeled data. Only one dominant mode is identified in model-generated data; hence, no propagation among indices. This principal mode strongly reflects a stationary, in-phase signal in all indices. And due to strong temporal similarity among each "stadium-wave" index in each of five GFDL CM3 simulations, this high correlation among signals in all modeled realizations is inferred to describe a forced climate response. This response dominates the model's multidecadal behavior. The spatial

signature of model-generated data is confined mostly to the North Pacific, projected onto the NHT, with minimal expression of an intrinsic multidecadal influence, particularly in atmospheric and Atlantic-centered indices.

VI. Discussion:
Failure of modeled data to produce a stadium wave (*Wyatt and Peters* 2012; *Kravtsov et al. 2014*); potential mechanisms underpinning stadium-wave propagation through coupled ocean, ice, and atmospheric indices (*Wyatt and Curry 2014*); and rejection of *Mann et al's* null hypothesis (*Kravtsov et al. 2014*) combine to provide strong support for the stadium-wave propagation signal in generating the multidecadal component of the Northern Hemisphere's observed 20th century climate variability.

Fundamental distinctions in the outcomes of analyses between modeled and observed data sets, as illustrated through the *Wyatt and Peters 2012* and *Kravtsov et al. 2014* studies, may imply deficiencies in climate-model design, potentially reflective of their omission of, or poor representation of, dynamics critical to the generation of climate variability on multidecadal timescales.

Beyond focus on the stadium wave, general arguments put forth by *Mann et al.* regarding methodology used to isolate forced and unforced components of climate variability serve only to highlight the debate over disentanglement of components, while incautiously promoting the tacit assumption that all studies using linear detrending use it to isolate the intrinsic component of NHT. In the case of the stadium-wave, this assumption is false.

There is value to *Mann et al's* method of differencing. It is one way to attempt to isolate an intrinsic component, if that is the goal. Others have used differencing with alternate terms being subtracted (*Trenberth and Shea 2006; Mann and Emanuel 2006; Kravtsov and Spannagle 2008; Knight 2009*), some with results divergent from *Mann et al's*, in fact, some with results that are not too dissimilar from those derived from linear detrending (*Kravtsov and Spannagle 2008; Knight 2009*).

VII. Conclusion:
This memo addresses arguments set forth in *Mann et al. 2014* regarding the statistical method of linearly detrending indices, and this method's purported role in generating a statistical artifact resembling the hypothesized 'stadium wave'. The methodology of *Mann et al.* is described, their arguments outlined and addressed, and the debate put into perspective. Two conclusions stand out: 1) Additional support for the "stadium-wave" in observed data emerged via application of a classical Monte-Carlo approach for statistical-significance testing; and 2) Spatio-temporal signatures differ markedly between modeled and observed data sets. Taken together, these two lines of evidence suggest that intrinsic dynamics in nature either are absent or poorly represented in current model designs.

Papers and Supplementary Materials:
See: http://people.uwm.edu/kravtsov/publications/
And: http://www.wyattonearth.net/publicationsstadiumwave.html

Footnotes from text:

1) Intrinsic variability (internally generated) is not devoid of external forcing. To exhibit oscillatory behavior, the system must have a source of energy. Internal variability emerges with a constant force applied. In addition, an inconstant external forcing can influence the temporal character of an intrinsic system. *See footnote 3*.

2) See :http://www.wyattonearth.net/images/Uncertainty_in_Climate_Science_NotesPages_.pdf pps.20-40.

3) Frequency entrainment: An inconstant external force can influence tempo of an intrinsic system, if that intrinsic system varies as a self-sustained oscillator, and if the frequency of the inconstant externally forced signal is similar to that of the self-sustained oscillator. Ex: Solar variability may indirectly influence timescale of climate variability via frequency entrainment (*Wyatt (dissertation) 2012*).

4) Time series of observed (instrumental) data are much too short to test for statistical significance of a multidecadal-scale periodic signal; yet no periodicity is suggested in the stadium-wave oscillatory behavior. Instead, over the last 150 years-plus, the stadium-wave network has regularly repeated in quasi-periodic fashion, with a timeframe between ~ 55 to 70 years, centering on ~64 years average since late 1800s.

Note: Three more papers authored by S. Kravtsov and colleagues came out in 2017. They all deal with modeled data. The CMIP5 collection of models is analyzed, as are some other models not in the collection. Discussion of results from these three papers will be the subject of the next essay to be posted.

References:

Black D, Peterson LC, Overpeck JT, Kaplan A, Evans MN, Kashgarian M (1999) Eight Centuries of North Atlantic Ocean Atmosphere Variability. Science 286: 1709-1713. doi: 10.1126/science.286.5445.1709

Enfield, D. B., et al. (2001), The Atlantic multidecadal oscillation and its relation to rainfall and river flows in the continental US, Geophys. Res. Lett., 28, 2077–2080, doi: 10.1029/2000GL012745

He F, Shakun JD, Clark PU, Carlson AE, Liu A, Otto-Bliesner BL, Kutzbach JE (2013), Northern Hemisphere forcing of Southern Hemisphere climate during the last deglaciation, Nature 494, 81-85, doi:10.1038/nature11822.

Ghil M., et al. (2002), Advanced spectral methods for climatic time series, Rev. Geophys., 40(1), 3.1–3.41, doi:10.1029/2000GR000092

Goldenberg SB, Landsea CW, Mestas-Nunez AM, Gray WM (2001) The Recent Increase in Atlantic Hurrican Activity: Causes and Implications, Science 293, 474-479, doi:10.1126/science.1060040

Gray ST, Graumlich LJ, Betancourt JL, Pederson GT (2004) A tree-ring based reconstruction of the Atlantic Multidecadal Oscillation since 1567 A.D. Geophys Res Lett 31. L12205: doi:10.1029/2004GL019932

Kerr, R. A. (2000), A North Atlantic climate pacemaker for the centuries, Science, 288, 1984–1985, doi: 10.1126/science.288.5473.198

Knight, J. R., et al. (2005), A signature of persistent natural thermohaline circulation cycles in observed climate, Geophys. Res. Lett., 32, L20708, doi: 10.1029/2005GL02423

Knight JR, Folland CK, Scaife AA (2006), Climate impacts of the Atlantic Multidecadal Oscillation, Geophy Res Let 33, L17706, doi:10.1029/2006GL026242

Knight J (2009), The Atlantic Multidecadal Oscillation Inferred from the Forced Climate Response in Coupled General Circulation Models, J of Clim, 22, 1610-1625, doi:10.1175/2008JCLI2628.1

Kravtsov S, Spannagle C (2008), Multidecadal Climate Variability in Observed and Modeled Surface Temperatures, J of Clim 21, 1104-1121, doi:10.1175/2007JCLI1874.1

Kravtsov S, Wyatt MG, Curry JA, Tsonis AA (2014), Two contrasting views of multidecadal climate variability in the twentieth century, Geophys. Res. Lett., 41, 6881-6888, doi: 10.1002/2014GL061416.

Lee SY, Chiang JCH, Matsumoto K, and Tokos KS (2011), Southern Ocean wind response to North Atlantic cooling and the rise in atmospheric CO2 : Modeling perspective and paleoceanographic implications, Paleoceanography 26, PA1214, doi:10.1029/2010PA002004

Mann, M. E., and K. A. Emanuel (2006), Atlantic hurricane trends linked to climate change, Eos, 87(24), 233–241, doi:10.1029/2006EO240001.

Mann, M. E., et al. (2014), On forced temperature changes, internal variability and the AMO, Geophys. Res. Lett., in press, doi: 10.1002/2014GL059233.

Moron, V., et al. (1998), Trends, interdecadal and interannual oscillations in global sea-surface temperatures, Climate Dyn., 14, 545–569, doi: 10.1007/s003820050241.

Parker D, Folland C, Scaife A, Knight J, Colman A, Baines P, and Dong B (2007), Decadal to multidecadal variability and the climate change background, J of Geophy Res, 112, doi:10.1029/2007JD008411.

Sutton RT, Hodson DLR, Mathieu P (2003) The Role of the Atlantic Ocean in Climate Forecasting, Proceedings of the ECMWF Workshop on the Role of the Upper Ocean in Medium and Extend Range Forecasting, ECMWF, Reading, UK

Sutton RT, Hodson DLR (2003) Influence of the Ocean on North Atlantic Climate Variability 1871-1999. J Climate 16: 3296-3313. doi: 10.1175/1520-0442

Sutton RT, Hodson DLR (2005) Atlantic Ocean forcing of North American and European summer climate. Science 309: 115-118. doi: 10.1126/science.1109496

Sutton RT, Hodson DLR (2007) Climate response to basin-scale warming and cooling of the North Atlantic Ocean. J Climate 20: 891-907. doi: 10.1175/JCLI4038.1

Trenberth KE, Shea DJ (2006), Atlantic hurricanes and natural variability in 2005, Geophy Res Let, 33, L12704, doi:10.1029/2006GL026894.

Vincze M, Janosi (2011), Is the Atlantic Multidecadal Oscillation (AMO) a statistical phantom? Nonlinear Processes in Geophysics 18, 469-475, doi:10.5194/npg-18-469-2011

Wyatt MG (2012) A Multidecadal Climate Signal Propagating Across the Northern Hemisphere through Indices of a Synchronized Network. Ph.D. Dissertation. University of Colorado, Boulder, CO, 220 pp. [Available from UMI ProQuest. Publication #UMI3527373 Ann Arbor, MI.]

Wyatt MG, Kravtsov S, Tsonis AA (2012) Atlantic Multidecadal Oscillation and Northern Hemisphere's climate variability. Clim Dyn 38 (5-6): 929-949 DOI: 10.1007/s00382-011-1071-8

Wyatt, MG, Peters J M (2012), A secularly varying hemispheric climate-signal propagation previously detected in instrumental and proxy data not detected in CMIP3 data base, SpringerPlus, 1–68, doi:10.1186/2193-1801-1-68

Wyatt MG, Curry JA (2014), Role for Eurasian Arctic shelf sea ice in a secularly varying hemispheric climate signal during the 20th century, Climate Dyn., doi: 10.1007/s00382-013-1950-2