A comparison of *Ad Hoc Committee Report* (Wegman, Scott, Said) section 2.2, p.15-17
and *Various unattributed sources on noise models (mostly various Wikipedia articles)*

Regular font indicates substantially close wording between the two sources, *italic* represent paraphrased sections, **bold** represents significant departures of Wegman et al from sources, and **bold underline** represent points of outright contradiction. Paragraphs have been reformatted for easy comparison. Within sections of close wording, identical phrases (ID) are highlighted in cyan, trivial changes (TC) with yellow . Changes resulting in issues are underlined.

Issues introduced by changes are classified as follows: E (Error), B (Bias) and C (Change in meaning). Upper-case indicates major problem, lower-case minor.

Wegman et al – Section 2.2, p. 15-17	
Principal Components (Paragraph 1 p. 15) Principal Component Analysis (PCA) is a method for reducing the dimension of a high dimensional data set while preserving most of the information in those data. Dimension is here taken to mean the number of distinct variables (proxies). In the context of paleoclimatology, the proxy variables are the high dimensional data set consisting of several time series that are intended to carry the temperature signal. The proxy data set in general will have a large number of interrelated or correlated variables.	**[No specific antecedent found]**
Principal component analysis tries to reduce the dimensionality of this data set while also trying to explain the variation present as much as possible. To achieve this, the original set of variables is transformed into a new set of variables, called the principal components (PC) that are uncorrelated and arranged in the order of decreasing “explained variance.” It is hoped that the first several PCs explain most of the variation that was present in the many original variables. ***[Issue <C>:** Introducing explained variance goes further than the original.]*	**Jolliffe – Introduction (p. 1-2)** The central idea of principal component analysis (PCA) is to reduce the dimensionality of a data set consisting of a large number of interrelated variables, while retaining as much as possible of the variation present ~~in the data set.~~ This is achieved by transforming to a new set of variables, the principal components (PC~~s~~), which are uncorrelated and which are ordered so that the first few retain most of the variation present in all of the original variables**.** ***[p. 2]*** … *but* it is hoped, *in general, that most of the variation in x will be accounted for by m PCs, where m << p.*
The idea is that if most of the variation is explained by the first several principal components, then the remaining principal components may be ignored for all practical purposes and the dimension of the data set is effectively reduced.	**[No specific antecedent found]**
Noise, White and Red (Paragraph 2 p. 15-6) In a statistical signal processing sense, noise is defined to be unknown external factors or fluctuations in signals. Noise is typically modeled as what are called random innovations, random meaning not predictable. Signals of interest may be actual audio signals, signals in the electromagnetic spectrum, or a known function to which simple statistical noise has been added. In the paleoclimatology context, noise is the unpredictable portion of the climate signal caused by random variations in the factors related to tree ring growth, ice core development, or coral growth. Noise is often thought of in terms of periodic variation.	**[No specific antecedent found]**

Wegman et al – Section 22, p. 15-6 (paragraph 2 cont.)

There are many types of noise with varying frequencies, each classified by a color.

The color names for these different types of sounds are derived from an analogy between the spectrum of frequencies and the equivalent spectrum of light wave frequencies. That is, if the pattern of 'blue noise' were translated into light waves, the resulting light would be blue.

*[**Issue <e>: "**Varying frequencies" is incorrect.]*

Section 2.2, p.16, paragraph 3

White noise, has equal power density across the entire frequency spectrum, that is, it has constant energy at all frequencies.

When this is graphically represented, white noise has a flat power spectral density.

In a practical example, white noise is what is used to refer to that steady, even soothing sound produced when tuning in to an unused radio or TV frequency.

White noise has an equal amount of energy per frequency band in contrast to pink noise, which has an equal amount of energy per octave.

Pink noise has a frequency spectrum that is flat in logarithmic space. The power density of pink noise, compared with white noise, decreases by 3 dB (decibels) per octave.

It is said that pink noise is the most soothing sound to the human ear. Pink noise has the same frequency distribution as falling rain.

.

Wikipedia article – Colors of Noise (April 12, 2006 version)

There are many forms of noise with various frequency characteristics that are classified by "color" ….

The color names for these different types of sounds are derived from an analogy between the spectrum of frequencies ~~of sound wave present in the sound (as shown in the blue diagrams)~~ and the equivalent spectrum of light wave frequencies. That is, if the sound wave pattern of "blue noise" were translated into light waves, the resulting light would be blue~~, and so on~~.

Epanorma.net web page on Noise Types (January 27, 2006 version)

White noise has equal power density across the entire spectrum (~~per Hz basis~~). White Noise has constant energy at all frequencies.

Wikipedia article – White Noise (May 3, 2006 version)

White noise is a random signal (or process) with a flat power spectral density.

[**No antecedent found**]

Description of GoldLine White/pink noise generators (origin unknown)

~~Very similar in sound~~, pink noise has an equal amount of energy per octave ~~from 20Hz~20KHz~~, while white noise has an equal amount of energy per frequency, ~~in Hz, from 20Hz~20KHz~~.

Wikipedia article – Colors of Noise (April 12, 2006 version)

Pink Noise

The frequency spectrum of pink noise is flat in logarithmic space, … . … The power density, compared with white noise, decreases by 3 dB per octave.

[**No antecedent found** (but appears to be incorrect)]

Section 2.2, p.16, paragraph 3

Red noise is similar to pink noise, but it has relatively more energy at lower frequencies than pink noise. Red noise has a power density that decreases 6 dB per octave as the frequency increases.

Of course, red noise was named after a connection with red light, which is on the low end of the visible light spectrum.

Mathematically speaking, integrating white noise produces red noise.

Red noise in the paleoclimatology context comes from the fact that tree rings have correlation from year to year, that is, if a tree grows well in a given year, it will store carbohydrates and will tend to have a good year of growth the following year as well. Red noise in the paleoclimatology context is modeled by a first-order autoregressive model.

Section 2.2, p.16, paragraph 4
Autoregressive, Moving Average and ARMA Models
Autoregressive, moving averages, and ARMA models are statistical time series models. An autoregressive model of order p means that the present value of the time series depends only on the p most recent past values of the time series. The dependence is taken to be linear. If $p = 1$, then we say that the process is a first order autoregressive process as indicated for the red noise model above. A moving average process of order q is formed by taking a weighted average of q uncorrelated white noise terms, that is, zero mean constant variance terms. The moving average means that the next value of the moving average process drops off the oldest term from the average and adds a new term.

Autoregressive moving average (ARMA) models, sometimes called Box-Jenkins models, are also used to model time series data. These models are used for understanding and predicting future values in the series. There are two parts to the ARMA model, an autoregressive (AR) part and a moving average (MA) part.

Wikipedia article – Colors of Noise (April 12, 2006 version)
Section: Brown (or red) noise
Brown noise is similar to pink noise, but with a power density decrease of 6 dB per octave with increasing frequency…
Wikipedia article – Brownian Noise (May 2, 2006 version)
… it has more energy at lower frequencies, *even more so* than pink noise.

Epanorma.net web page on Noise Types (January 27, 2006 version)
"red noise" got its name after a connection with red light which is on the low end of the visable light spectrum.

Wikipedia article – Brownian Noise (May 2, 2006 version)
Brown noise can be produced by integrating white noise.

[**No antecedent found**]

[**No antecedent found**]

Wikipedia article – Moving Average Model (April 22, 2006 version)
…autoregressive moving average (ARMA) models, sometimes called Box-Jenkins models … are typically applied to time series data.
… the ARMA model is a tool for understanding and, perhaps, predicting future values in this series. The model consists of two parts, an autoregressive (AR) part and a moving average or (MA) part.

Section 2.2, p.17, paragraph 5
Gaussian Noise and Long Memory Processes
Although we have not specified the probability distribution of the random innovations, it is often the case that a normal or Gaussian probability distribution is appropriate to model noise or what we have called random innovations. The basic paleoclimatology model is taken to be a trend with superimposed white or red noise usually with a Gaussian distribution. The Gaussian distribution assumption is a convenient one mathematically.

Random (<u>or</u> stochastic) processes whose autocorrelation function, decaying as a power law, sums to infinity <u>are known as</u> <u>long range correlations</u> or long range dependent processes. Because the decay is slow, as opposed to exponential decay, these processes are said to have long memory. Applications exhibiting long-range dependence include Ethernet traffic, financial time series, geophysical time series such as variation in temperature, and amplitude and frequency variation in EEG signals.

[***Issue (E):*** *The processes discussed are not "long range correlations". They are "processes* ***with*** *long range correlations.]*

Fractional Brownian motion is a self-similar Gaussian process with long memory.

The Box-Jenkins ARMA models described in the previous section are all short-term memory processes.

[**No antecedents found**]

Rangajaran & Ding (ed.), p. vi
Processes with long range correlations (also called long range dependent processes) ~~occur ubiquitously in nature~~. They are defined as random stochastic processes whose autocorrelation function, decaying as a power law ~~in the lag variable for large lag values~~, sums to infinity. Because of this slow decay (as opposed to an exponential decay), these processes are also said to have long memory. … A partial list of problems involving long range dependence include: ~~Anomalous diffusion, potential energy fluctuations in small atomic clusters~~, Ethernet traffic, geophysical time series such as variation in temperature ~~and rainfall records~~, financial time series, ~~electronic device noises in field effect and bipolar transistors~~, and amplitude and frequency variation in ~~music,~~ EEG signals ~~etc~~. …

Fractional Brownian motion is a ~~prototypical~~ self similar Gaussian process with long memory.

[**No antecedents found**]

Section 2.2, p.17, paragraph 6
In reality, temperature records and hence data derived from proxies are not modeled accurately by a trend with superimposed noise that is either red or white. There are complex feedback mechanisms and nonlinear effects that almost certainly cannot be modeled in any detail by a simple trend plus noise. These underlying process structures appear to have not been seriously investigated in the paleoclimate temperature reconstruction literature. Cohn and Lin (2005) make the case that much of natural time series, in their case hydrological time series, might be modeled more accurately by a long memory process. Long memory processes are stationary processes, but the corresponding time series often make extended sojourns away from the stationary mean value and, hence, mimic trends such as the perceived hockey stick phenomena.

Section 2.2, p.17, paragraph 7
One type of such long memory processes is a process driven by fractional Gaussian noise (fractional Brownian motion).

An object with self-similarity is exactly or approximately similar to a part of itself. For example, many coastlines in the real world are self-similar since parts of them show the same properties at many scales. Self-similarity is a common property of many fractals,

as is the case with fractional Brownian motion.

A serious effort to model even the present instrumented temperature record with sophisticated process models does not appear to have taken place.

[**No antecedents found**]

[**No antecedent found**]

Wikipedia article – Self-similarity (Mar. 20, 2006 version)
A self-similar object is exactly or approximately similar to a part of itself.. … Many ~~objects~~ in the real world~~, such as~~ coastlines, are ~~statistically~~ self-similar: parts of them show the same ~~statistical~~ properties at many scales. Self-similarity is a typical property of fractals.

[**No antecedent found**]

Statistical summary:

Total words: 1200. Striking similarity (SS) 32%, Trivial Changes/identical (TC+ID): 28%, Identical (ID): 26%.
Issues: 4

References

1. Edward J. Wegman, David W. Scott and Yasmin H. Said; *Ad Hoc Committee Report on the "Hockey Stick" Reconstruction* A Report to Chairman Barton, House Committee on Energy and Commerce and to Chairman Whitfield, House Subcommittee on Oversight and Investigations, 2006.
2. Ian T. Jolliffe, *Principal Component Analysis* (Springer, 2nd ed. 2002)
3. Wikipedia article – Color of Noise (April 12, 2006 version) - Available online at: http://en.wikipedia.org/w/index.php?title=Colors_of_noise&oldid=48074859
4. Wikipedia article – White Noise (May 2, 2006 version) - Available online at: http://en.wikipedia.org/w/index.php?title=Colors_of_noise&oldid=48074859
5. Epanorma.net web page on Noise Types (January 27, 2006 version) http://web.archive.org/web/20060127084954/http://www.epanorama.net/documents/audio/noisetypes.html
6. Description of GoldLine White/pink noise generators (origin unknown). http://www.google.ca/#q=%22White+noise+has+an+equal+amount+of+energy+per+frequency%22+%22Pink+noise+has+an+equal+amount+of+energy+per+octave%22&hl=en&filter=0&fp=9ceddb59471f0aa1
7. Wikipedia article – Brownian Noise (May 2, 2006 version) - Available online at: http://en.wikipedia.org/w/index.php?title=Brownian_noise&oldid=51190134
8. Wikipedia article – Moving Average Model (April 22, 2006 version) - Available online at: http://en.wikipedia.org/w/index.php?title=Autoregressive_moving_average_model&oldid=49741677
9. Govindan Rangarajan, Mingzhou Ding (ed..), *Processes with long-range correlations: theory and applications* (Springer, 2003)
10. Wikipedia article – Self-similarity (Mar. 20, 2006 version) – Available online at: http://en.wikipedia.org/w/index.php?title=Self-similarity&oldid=44580086