Subject: Re: Philosophical Question about NAN
Posted by R.G. Stockwell on Wed, 19 Nov 2008 14:32:43 GMT
View Forum Message <> Reply to Message

"David Fanning" <news@dfanning.com> wrote in message
news:MPG.238b3491ef337cc798a534@news.giganews.com...
> Folks,
>
> I've had a couple of run-ins lately with NANs and I wonder
> why routines like TOTAL and MEAN don't have the NAN keyword
> set to 1 by default. Why does the user have to set it?

My two cents. In spectral analysis, a 'missing point' or NAN has
a profound effect. The FFT _must_ have the behaviour it currently
has (it returns all nans, if there are any nans).

 The position/time of each
measurement is critical in an FFT, and you can't just throw a number away
and translate all the other points.

Also, it is not clear on how to handle a NAN - it is ambiguous.
Does one throw the point away (like in a MEAN() calculation), or
does one interpolate the data (like in an FFT() application)?
Automated interpolation can be a very bad idea (and what happens
if the NAN results in an extrapolation, cause there are no surrounding
points?) Imho, the programmer has to make the decision about how to
handle NANs.

Also, where data is missing can be important even in functions like MEAN().
Suppose you have a year of temperature data taken every hour, but in winter,
you only have daytime data (i.e. NANs at night).
 If you take monthly data, you will get the a very wrong result in terms of
the
annual variation of monthly means, and it will be artificial, and you won't
know about it.

cheers,
bob