ELECTRONOTES
1016 Hanshaw Road
Ithaca, NY 14850

APPLICATION NOTE NO. 401

Dec 22, 2013

YEARLY MOVING AVERAGES AS FIR FILTERS

Many years ago a student from the Physics Department came to me for some quick advice about the way they were processing some data. He apologized for the crudeness of their procedure, and said that they thought they should perhaps be using a digital filter. No surprise to readers here – what he was doing was digital filtering. He had invented something reasonable. Yes – it could be improved. Yet the chicken/egg questions with regard to digital filtering theory/practice is clearly that practice came first. For example, we have frequently introduced digital filters with a notion such as a shopkeeper intuitively using a length-7 moving average to determine long-term trends free of periodic day of the week trends.

Using such simple schemes as a moving average, and forgetting that it is a filter perhaps makes the scheme seem to be just statistics (the mathematics of statistics!), and we may not think of it as a filter with an output, with said output to be thereafter decisively located in time or space. In many digital filter applications, we forget about this as well, and the actual computed output of the filter may be used at our convenience. Indeed, filters usually have time delays or spatial shifts which we take for granted. Of course, time delays may matter. In telephone calls a delay of a second is a disaster, while it is not a problem in a music player (recorded long ago anyway), for two examples.

This relates to the issue of causality. You don't get an output until such time as all the needed input data is available. So do we ever use non-causal filters? Of course we do. As noted, many times we are not worried about the delay, and in the case of spatial filters (such as those for images), we don't even find that time is involved, and it is understood that we process the entire item all at once. Each case is individual, or so it seems.

One special case is where we have a <u>time sequence represented in a spatial manner</u>. That is – a graph of a time function. Is the independent variable here time or space? Does it matter? A graph of a time function could have an arbitrary time reference (famously usually set to an arbitrary zero) <u>but it might have a time reference that is absolute</u>. It might for example be a graph of average household income from 1900 to 2000, or perhaps of temperature in a particular city from 2000 to 2013. In such a case, even though the data are old (and we can use non-causal filters) it makes a difference where we place the output.

Recently I saw a blog posting of monthly temperature data that was smoothed with a moving average of length 13. Really – 13 months in a year – that's asking for trouble. Clearly a 12 month moving average is what you want, if your intention is to pass very low frequency (multi-year) components but block the "seasonal" variations that we quite rightly expect to, and often do, find.

It turned out that the person using the length 13 moving average was aware that this would imperfectly reject the 12 month periodicity, but did not like the notion that a consequence of using length 12 would be a shift of half a month at the output. We are often aware of the need for this sort of "trade-off" but it is also true that we can and should always seek out the possibility of a free lunch. This is easily available in this case, as students of digital filtering are likely to point out.

All one needs to do is use a length 13 filter but make the tap weights on both ends 1/2 (keeping the 11 taps = 1 in the middle) and divide by 12.

This is not magic, and no extensive study of digital filter theory is required. What may be less clear is the exact consequences of using this. Our scheme solves the problem of the imperfect cancellation of the 12 month seasonal cycle, and of the half-month shift. How different is the filter otherwise?

[There in an immense amount of information on digital filter design in ***Electronotes*** such as:

http://electronotes.netfirms.com/EN197.pdf

http://electronotes.netfirms.com/EN198.pdf

http://electronotes.netfirms.com/EN199.pdf

and in our app notes, not to mention many times more in text books and journals. We do not suggest looking at any of this at this point. We just want to point out that the case of FIR (Finite Impulse Response) filters is canonical and is covered in many places, with due attention to the problems of even/odd length, and even/odd symmetry – four classes of "linear phase" FIR filters.]

Here we will be comparing three filters: a length 13 moving average, a length 12 moving average, and a modified length 13 in accordance with the sentence in red above. These are:

$$h_{13} = [\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ ] / 13$$

$$h_{12} = [\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ ] / 12$$

$$h_{13mod} = [\ 0.5\ \ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ \ 0.5\ ] / 13$$

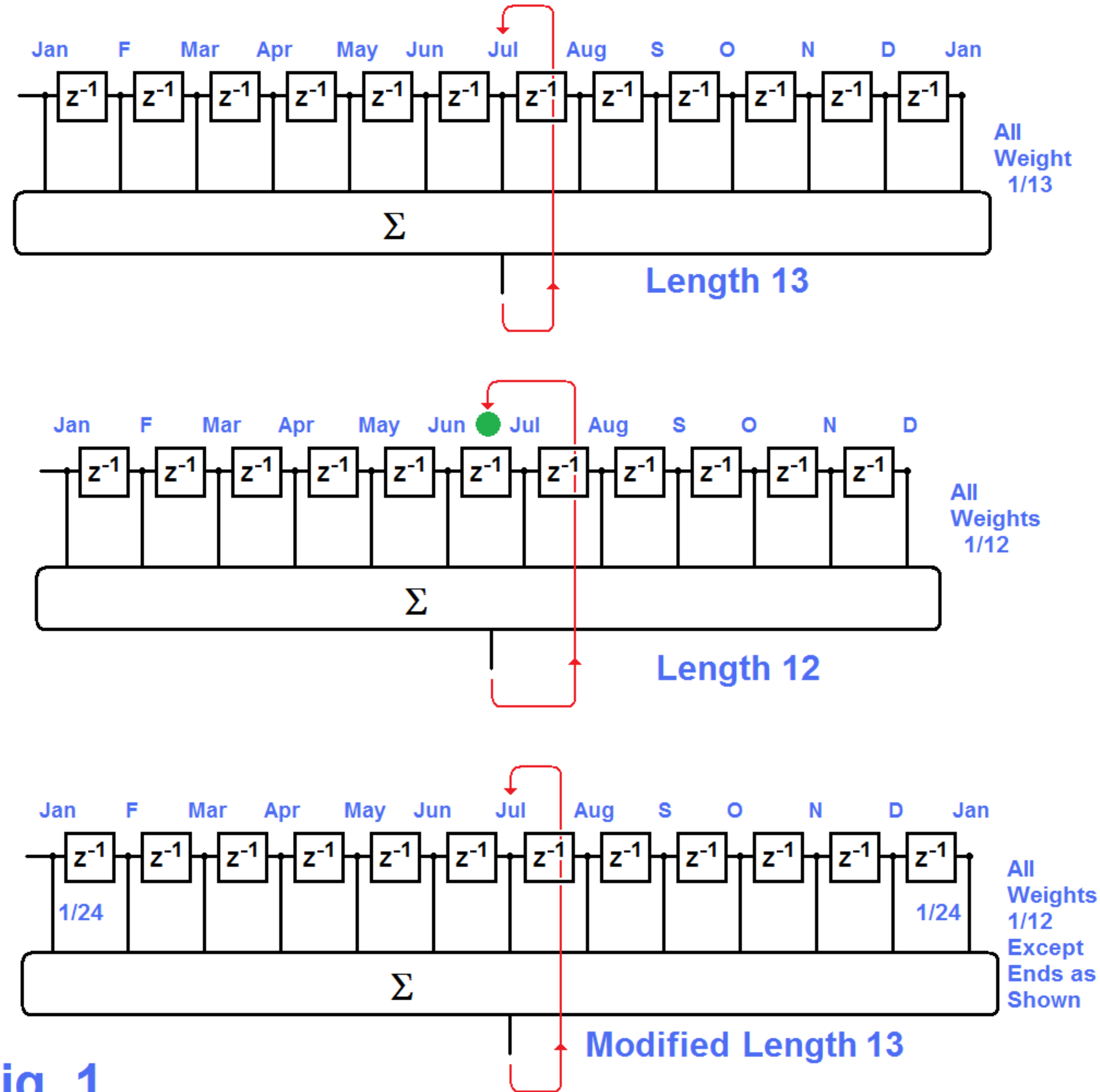
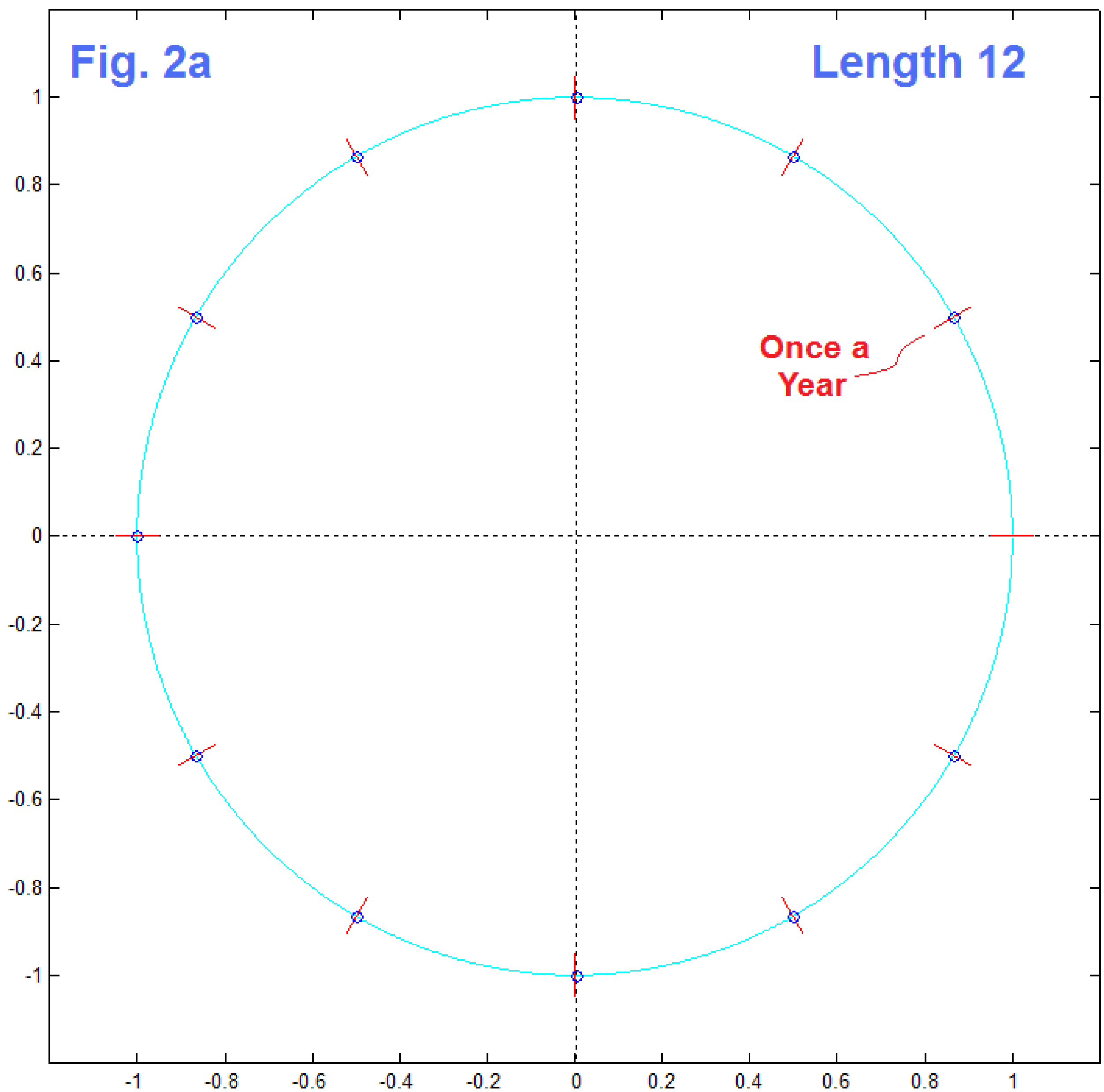


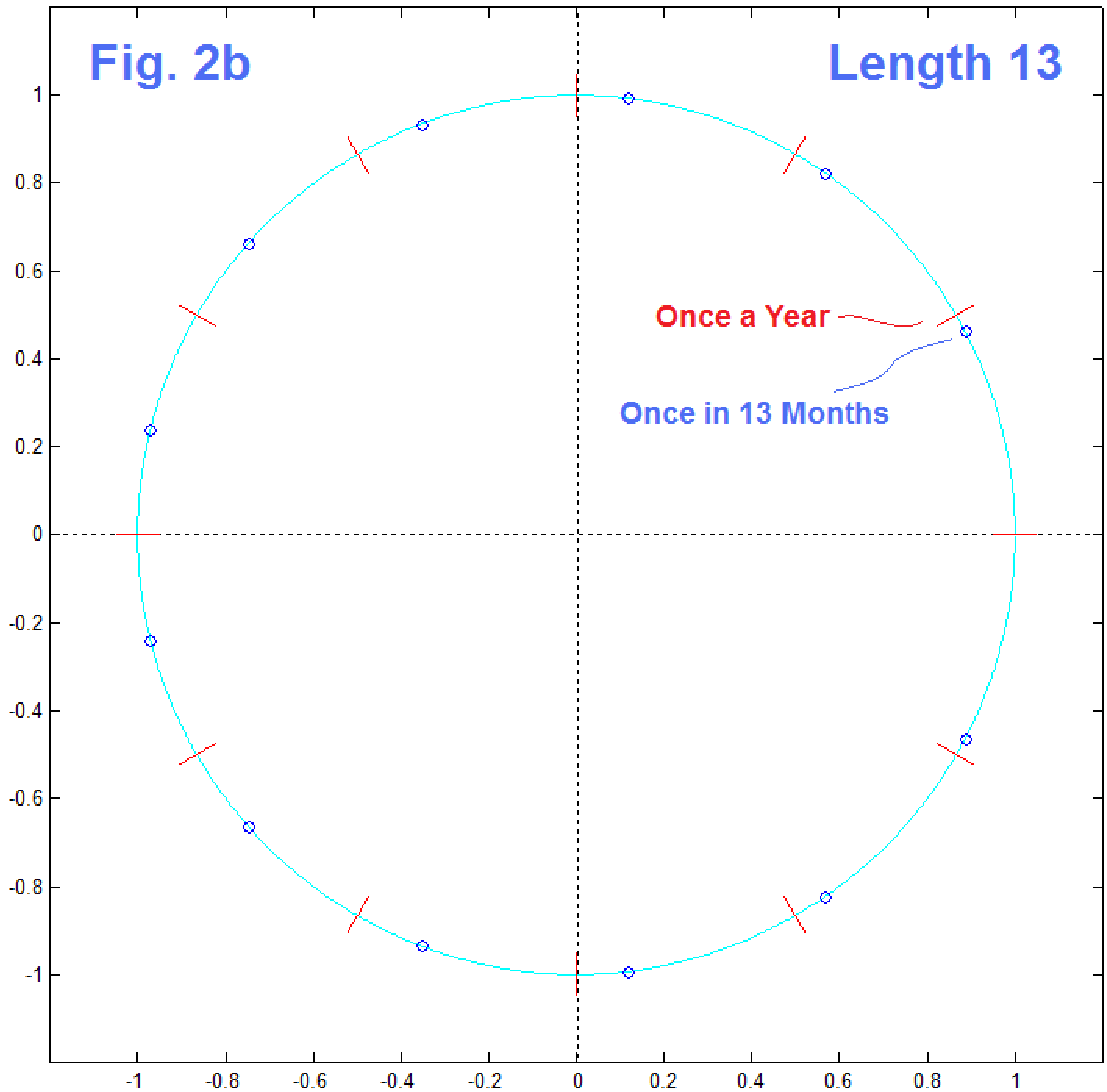
Fig. 1

Fig. 1 shows the structure of these, a delay line with a summer performing the running average, and we show this running over the 12 months of a year (starting in January for the example) and adding the following January in the case of the two length-13 filters. Note that for all three cases the output comes from the summer with no obvious place to go next. Clearly there is no valid output until the full 12 or 13 months are loaded. This would be the "phase" (delay if you prefer) which is a "linear phase" due to the symmetry. While the answer (output) is not available until the rightmost month is loaded, we nonetheless often associate this answer with the center of symmetry as though this were a zero-phase filter. In the sense of a "graphic" this non-causal interpretation is not objectionable, and is probably exactly the path to any insight we were looking for.

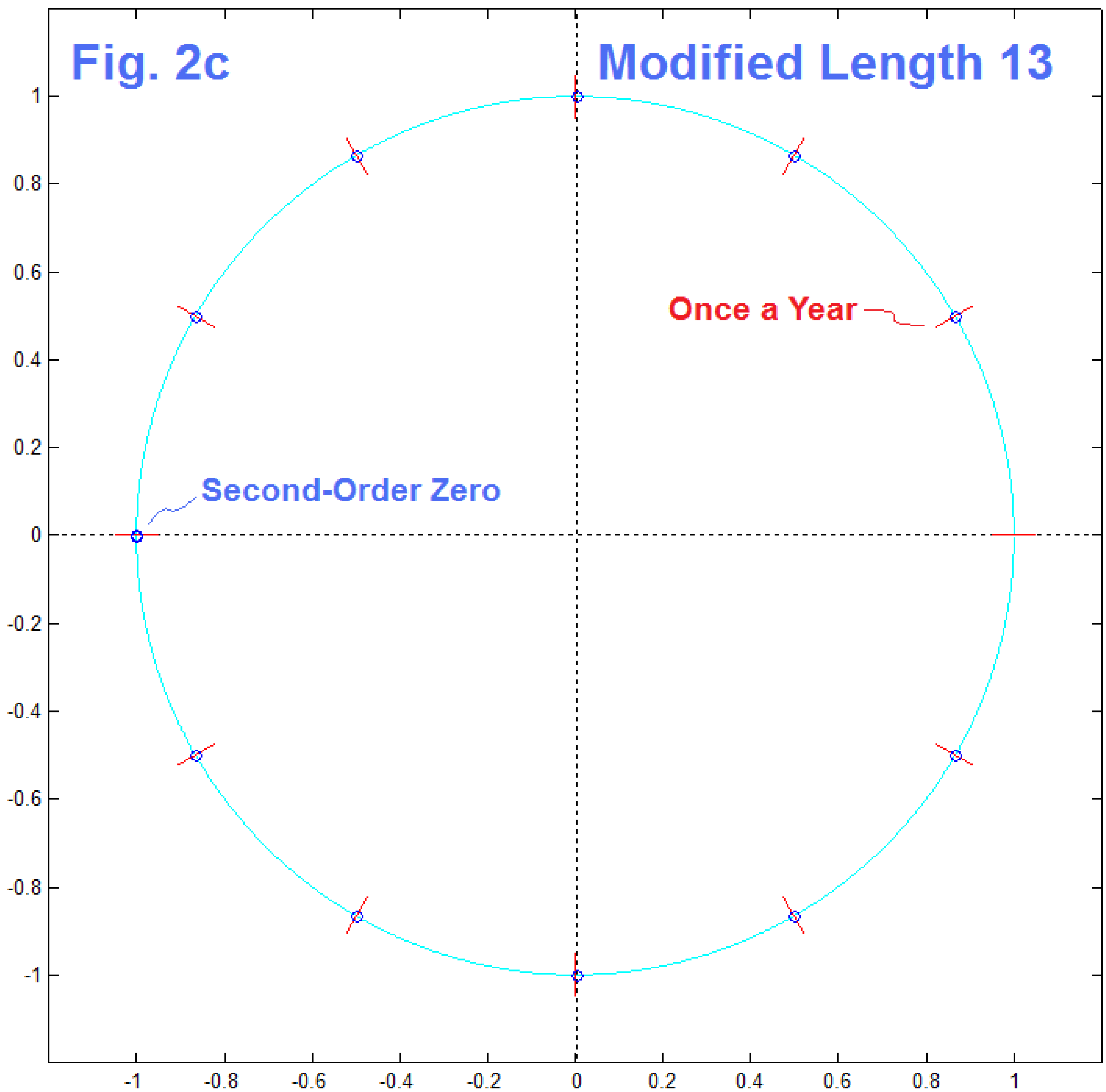
Accordingly, in Fig. 1 we show red paths which return the answer to a position that is the center of symmetry, which is "July" in the case of the two length-13 filters, and the "green dot" in the case of the length-12. But what does "July" mean? Is it July 1, July 31, July 15, or what? None of these – just "July". And the green dot is correspondingly just a newly defined month of "Green", which might be something like June 15 to July 15. So we see the discomfort if we are plotting the original monthly data <u>along with</u> the "smoothed" yearly average and trying to associate time in the original data sequence with time in the output data sequence.



Using either of the length-13 filters means that the uncomfortable half-month shift is not apparent. The value for "July" at the output is found from "July" at the input, six months prior to July, and six months that followed July. Yes, we had to wait for the six months following July. If we used the length-12, we would have had to associate the month of "Green" with either June or July (an unforced choice!). This likely would not even show on a graph of tens of years, but could show on graphs of just a few years. Hence a choice of length-13 with regard to the "phase" issue. What about the frequency response?



We want to see the frequency response of our choice of filter. We are probably interested in this for general frequencies, but are particularly concerned with it at two frequencies: 0 and 1-per-year. Here note that the "sampling frequency" is 12-per-year. In our digital filter studies, we use once around the circle in the z-plane to represent the sampling frequency. A frequency of 1-per-year (or 1-per-12 months) is thus 1/12 of a circle or 30° (as in Figs. 2a, 2b, and 2c). To pass zero frequency (dc, or a slowly varying "trend") we want no zero ("no notch") at 0°. To reject a frequency of 1/12, we want a zero there at 30°. That is our first design concern. For other frequencies we will look at the full frequency response as in Fig. 3 to come.



From the z-plane plots of Figs. 2a, 2b, and 2c we see that the length-12 (Fig. 2a) and the modified length-13 (Fig. 2c) meet or requirements on the placement of zeros, while the length-13 (Fig. 2b) fails. If we were choosing between length-12 and the modified length-13, note the entertaining fact that length 13 should have 12 zeros, although only 11 zeros appear in Fig. 2a and (apparently) in Fig. 2c. In fact, as noted, there is a second-order zero at z=-1 in Fig. 2c. [Matlab's built-in root finder, like many root finders has a tiny difficulty with multiple order roots such that the plot of the zero at z=-1 in Fig. 2c is slightly blurred and appears as a heavier plot.] This additional zero is clearly the result of h_{13m} being the convolution of h_{12} with a sequence [0.5 0.5]. So this is "cute" but also means that the modified length-13 h_{13m} is a "better" low-pass than h_{12}, in case that is an issue.

The "weights" listed at the bottom of page 2 are the "impulse responses" of the filters, and the Discrete-Time Fourier Transforms (DTFTs) of these are the frequency response. These are not difficult to compute, and programs such as Matlab make this easy using a Fast Fourier Transform (FFT) or the function ***freqz***. It is however fairly intuitive that summing 12 consecutive samples of a sinusoidal waveform of period 12 gives a zero. For example, 12 such samples of the sine would be:

0 0.5 0.866 1 0.866 0.5 0 -0.5 -0.866 -1 -0.866 -0.5

and we can cancel these in pairs. In this case, a 13^{th} sample would be zero, and not change the sum, but suppose the samples were offset in phase (such as by 0.2 radians), we would have the 13 samples:

0.1987 0.6621 0.9481 0.9801 0.7494 0.3180

-0.1987 -0.6621 -0.9481 -0.9801 -0.7494 -0.3180 0.1987

and the last sample does not cancel. If however the 1^{st} and 13^{th} samples were 0.1987/2, they would cancel with the 7^{th} sample, -0.1987. Hence the call for the modified length 13.

[A detailed discussion of the conditions relating to moving averages of sinusoidal sequences is found in our app note AN-375 "Moving Averages and Single Sinewave Cycles", November 2011 located at

http://electronotes.netfirms.com/AN375.pdf]

An even more direct demonstration is seen in Fig. 1 (top) where we see that a moving average centered on July would include January twice! This means that the yearly cycle "leaks through" the length-13 to an extent of 1/13 or about 7.7% (see Fig. 3 below).

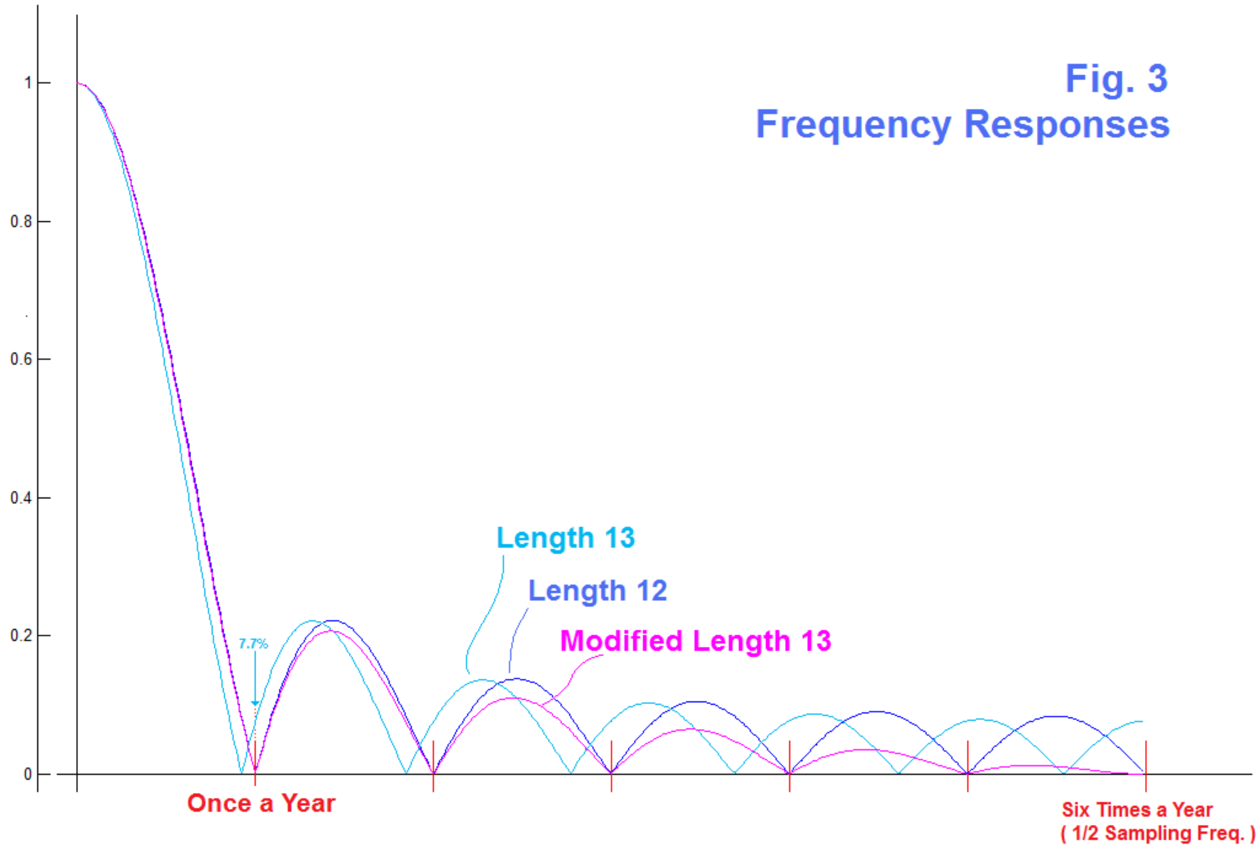


Fig. 3 shows the computed frequency response curves (using ***freqz***) corresponding to the three filters under consideration. Here we find a preference for the length-12 and the modified length-13 with regard to their nulls at a frequency of 1/year. Note that the "sampling frequency" here, being monthly, is 12/year. Because of the sampling, we only show the response for a range from 0 to half the sampling frequency (the aliasing issue). As mentioned, at the frequency 1 (Once a Year) the length-13 does not null, but passes about 7.7% of the response level we had at 0 frequency.

The above in total seems to suggest a preference for the modified length-13.