Baltic Astronomy, vol. 12, 253–270, 2003.

MERGING DATA FROM LARGE AND SMALL TELESCOPES – GOOD OR BAD? AND: HOW USEFUL IS THE APPLICATION OF STATISTICAL WEIGHTS TO TIME-SERIES PHOTOMETRIC MEASUREMENTS?

G. Handler[1,2]

[1] *Institut für Astronomie, Universität Wien, Türkenschanzstraße 17, A-1180 Wien, Austria*
[2] *South African Astronomical Observatory, P.O. Box 9, Observatory 7935, South Africa*

Received October 15, 2002

Abstract.

I have investigated the value of the contribution of small telescopes to the success of a whole WET run. To this end, I have applied different data weighting schemes to two extreme WET test data sets. I find that weights proportional to the inverse local scatter in the light curves produce Fourier Transforms of best signal-to-noise. Weighting data stronger than their inverse scatter does not yield optimal results because of the reduction of the effective number of data points.

The contribution of the small telescopes to the combined WET results was found to be very important. They do not only improve the spectral window, but they can reduce the noise in the total FT by more than their light gathering power would imply. Some suggestions for the optimal use of small telescopes in the WET are given.

Key words: techniques: photometric

1. INTRODUCTION

As we all know, the Whole Earth Telescope (still) is a unique astronomical instrument. Although several other global telescope networks exist, the WET regularly achieves the best coverage of the light curves of its targets. This is mainly caused by three typical features of the WET, beginning with the control of a run through the headquarters providing maximum motivation for the observer.

In addition, both large and small telescopes are involved, maximizing the probability of obtaining data. Finally, the observations are pushed to the limits, i.e. measurements are taken at considerably higher air mass than the standard photometrist would accept. In their description of the WET as an instrument, Nather et al. (1990) state, "Poor data, in this context, are far better than no data at all.".

However, there may be cases where data have such poor quality that they may actually compromise the astrophysical analyses to be obtained from the total WET run. An example is shown in Fig. 1.

This figure contains data from WET run XCov12 on the pulsating DB white dwarf PG 1351+489 that was sometimes observed by two telescopes at the same time. The left panels show a part of the light curve, the amplitude spectrum and a prewhitened amplitude spectrum from the larger of the two telescopes, whereas the panels on the right-hand side show the same for the small telescope data. The single lowest panel contains the combined amplitude spectrum of both prewhitened data sets. The prewhitened amplitude spectrum of the large-telescope data shows the presence of a number of additional significant peaks that are lost in the noise of the run from the small telescope. Even when combining the two runs, the additional signals cannot be detected. The data from the small telescope degrade the total result.

The question now is: how can the two runs be combined to give an optimal overall result, or more generally, how can we obtain the maximum possible astrophysical information out of WET measurements? Should small telescope data be disregarded?

I am aware that there is no straightforward answer to these questions, and that many researchers will have controversial opinions on this topic. However, as the organizers of this workshop asked me for a discussion of this subject, I will do my best trying to find a good way to add the contributions of large and small telescopes in the WET network to an optimal result. My aim here is not to satisfy the statistician, but to assist the astrophysicist in obtaining results of the best signal-to-noise (S/N) for the analyses to follow.

The remainder of the paper is organized as follows: in Section 2, I will re-discuss previous work in this direction which will (hopefully) make it clear that the assessment of the relative values of large and small telescope data is inseparable from the problem of weighting data. Section 3 will then briefly introduce different weighting schemes. In Section 4, I describe two test data sets to which these

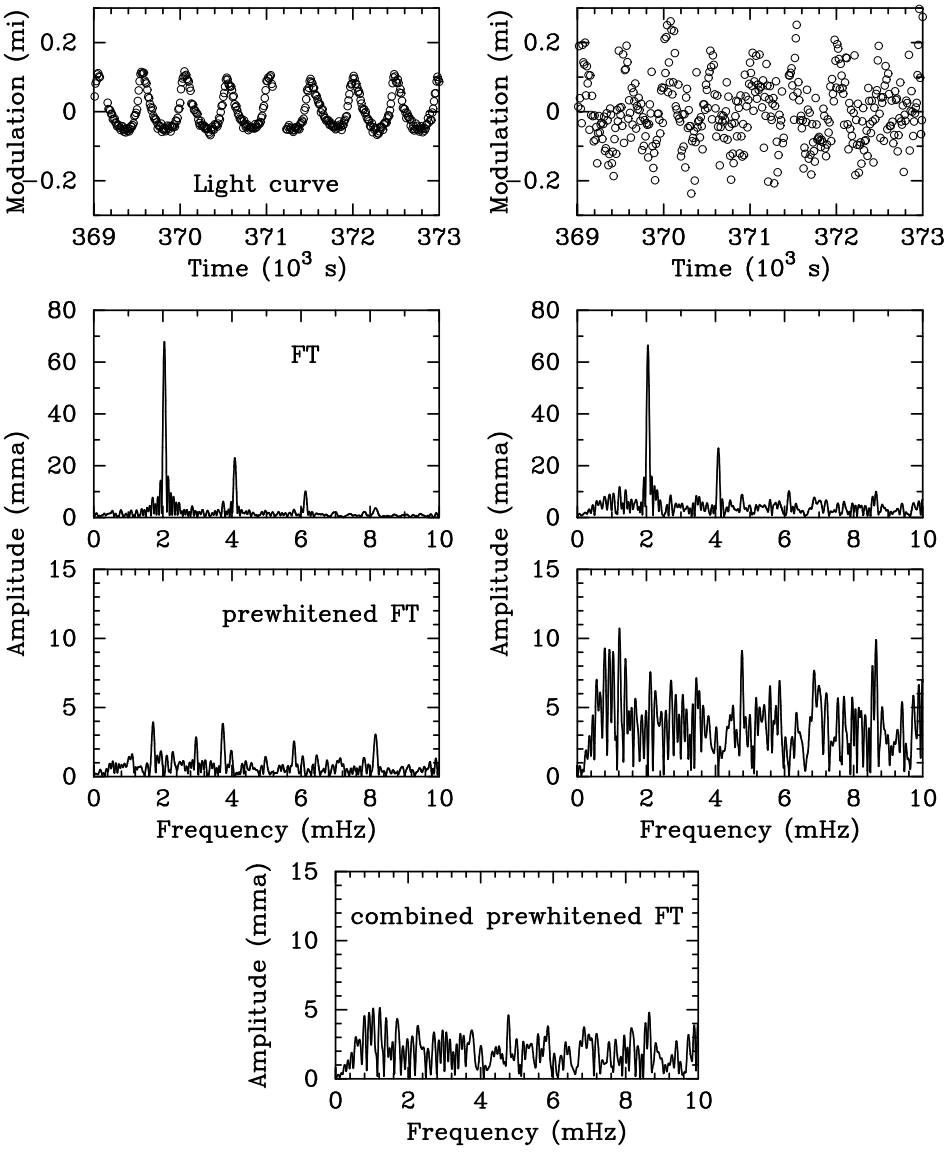

Fig. 1. Upper row: a part of overlapping light curves of PG 1351+489 from a large (left) and a small (right) telescope. Second row: Fourier amplitude spectra of the total overlap, left: large telescope, right: small telescope. Third row: amplitude spectra of the same data prewhitened by the dominant frequency and its first two harmonics. Lowest panel: amplitude spectrum of the combined residuals from both telescopes.

schemes will be applied, and in Section 5 the different schemes will be evaluated. The results will be discussed and be used to assess the value of the contribution of small telescopes to the WET in Section 6. Section 7 finally presents the conclusions of this study.

Finally, I would like to point out that parts of the discussion may be too technical for some readers. Consequently, I have tried to write this paper in a way that will allow the reader to skip over to the next section with a minimum loss of essential information.

2. PREVIOUS WORK/SETTING THE STAGE

The subject of the present paper has already been discussed in the past, although in a more limited context. Moskalik (1993) examined how to add overlapping measurements from two telescopes to optimize the light gathering power of the WET. He examined two simultaneous runs on the DAV GD 154 from the McDonald 2.1 m and 0.9 m telescopes and came to the following conclusions:

- Co-addition of data from two similar-sized telescopes gives the largest increase in S/N.
- Data from a very small telescope added to data from a large telescope do not result in much improvement; in fact, the combined data can be worse than the ones from the large telescope alone (cf. Fig. 1).
- Overlapping data thus need to be weighted.
- The weights should be proportional to the average net count rate for the target star (thus about inversely proportional to telescope aperture squared).
- It follows that non-overlapping runs should also be weighted.

However, some of these conclusions may be challenged. Firstly, it should be noted that Moskalik (1993) only considered photon statistics as a noise source. Although this is the main source of noise in typical WET data, the effects of scintillation, poor observing conditions (clouds), equipment and observer problems etc. can often be significant. Thus the weight of the data should be inversely proportional to the *intrinsic* scatter of the data and be measured directly.

Secondly, the suggested weighting procedure disregards the contribution of sky background to the error budget; the latter is especially important for small telescopes that generally have to use larger

apertures on the sky. Another point (that was outside the scope of Moskalik's work) is that weighting data also affects the spectral window function. I illustrate this in Fig. 2.

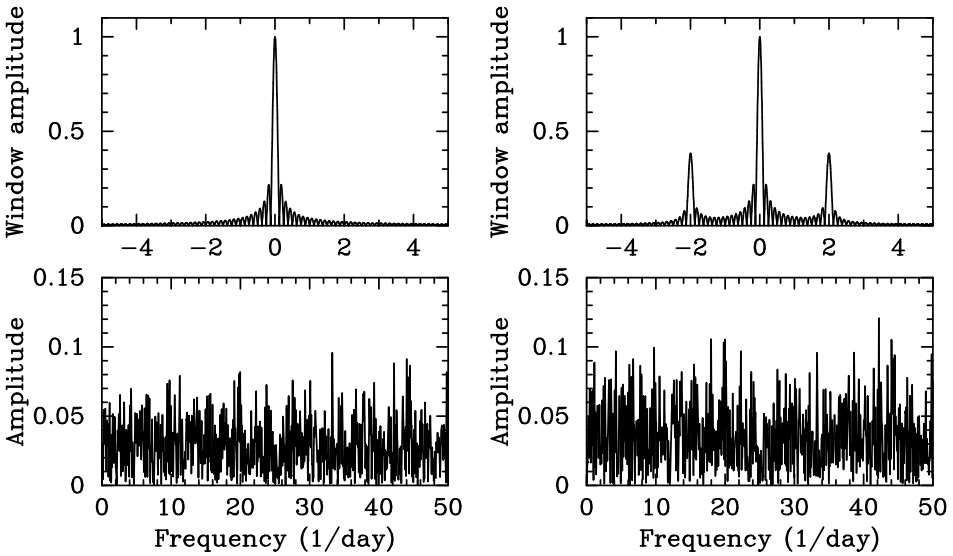

Fig. 2. Upper left: spectral window of a simulated 8-day unweighted data set with 100% coverage, coming from two 2-m telescopes from opposite sides of the globe plus two 1-m telescopes from opposite sides of the globe, 90°away from the larger telescopes; each telescope observes for 6 hours every night. Upper right: spectral window of the same data set weighted according to Moskalik (1993). Lower left: unweighted amplitude spectrum of a data set of random noise with the same temporal distribution. Lower right: weighted amplitude spectrum of the same data set.

We learn several things from Fig. 2. Firstly, if weights are applied to the measurements, the data of better quality begin to dominate the combined data set (desirable because we want to suppress noise), but this modifies the spectral window, usually to the worse. The aliases creep back. Secondly, the white noise level increases in this example[1], because the effective number of data points in the time series is reduced.

Consequently, several effects play against each other if one uses statistical weights for frequency analysis of multisite observations. The decrease in noise is offset by a poorer spectral window function and by the effective removal of data points. Hence the most appropriate weights depend on the structure and distribution of each

[1]Of course, it would decrease as desired if the small telescope data had 4 times higher scatter than the ones from the large telescope.

different campaign's data and can in practice not be theoretically predicted!

At this point it is important to note that a given telescope's contribution to the overall data set can by evaluated by finding the weights that result in the combined FT of highest S/N. The average weight of the data from this telescope in relation to some other parameter (like telescope aperture squared or inverse variance) can then be compared to the same analysis for all the telescopes in the run. Such a comparison will tell if it is useful to have small telescopes in the WET (and up to what point this would be).

2.1. A little excursion: do we want to merge overlapping data?

As we have already seen, weighting affects the spectral window of a given data set. If overlapping data from different sites exist, another weighting effect appears: the overlap has double weight. Obviously, this also modifies the spectral window. A practical example is shown in Fig. 3.

As overlaps mostly occur at longitudes where the multisite coverage is best, even higher weight in the combined FT is given to those data from the corresponding part of the globe. Thus the aliasing problem becomes worse although there is more data! Hence overlapping data should be merged before a frequency analysis, regardless if weighted or unweighted data are used for it.

2.2. Should we use weights at all?

Even despite the example in Fig. 1, where I have demonstrated how simple co-adding of data of good and bad quality may hide some signals present in the light curves, the reader may still wonder if weighting will be useful in practice. A comparison of a weighted and an unweighted amplitude spectrum is therefore shown in Fig. 4.

The weighted amplitude spectrum in the lower panel of this figure represents a dramatic improvement over the unweighted one. The noise level drops to about 40% of that of the unweighted FT and many more signals are now obvious in the data. Hence, weighting is expected to be useful for at least some data sets.

3. WEIGHTING SCHEMES

Having shown that weighting can be of advantage, it is time to look into different possibilities for assigning weights. Besides using the unweighted data as a reference, I have tested three different

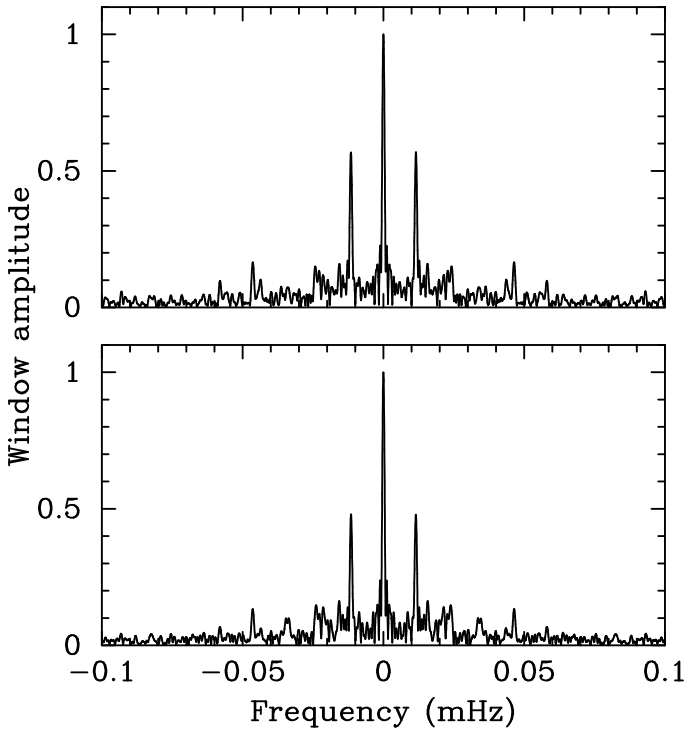

Fig. 3. Upper panel: the spectral window of a subset of the on-line reduced XCov22 data of PG 1456+103. The daily alias has a strength of 57% of the central peak. Lower panel: spectral window of the same data set, but with overlaps merged. The daily alias is reduced to 48% of the amplitude of the central peak.

schemes to determine statistical weights for the data points in a time series. All of them are (as justified before) based on an estimate of the variance in a given light curve. The schemes will be introduced in what follows; all of them have some advantages and problems. There is no perfect weighting scheme.

3.1. Fourier noise weighting

This method uses the white noise level in a Fourier spectrum as a measure of data quality and has been applied to WET data by Dind et al. (2003). The inverse of the average noise level in a frequency region without a stellar signal is computed for each light curve obtained during an observing campaign and is used as the basis for the weights of the corresponding run.

The advantage of this method is its simplicity. Potential shortcomings are that a frequency region without intrinsic FT peaks must

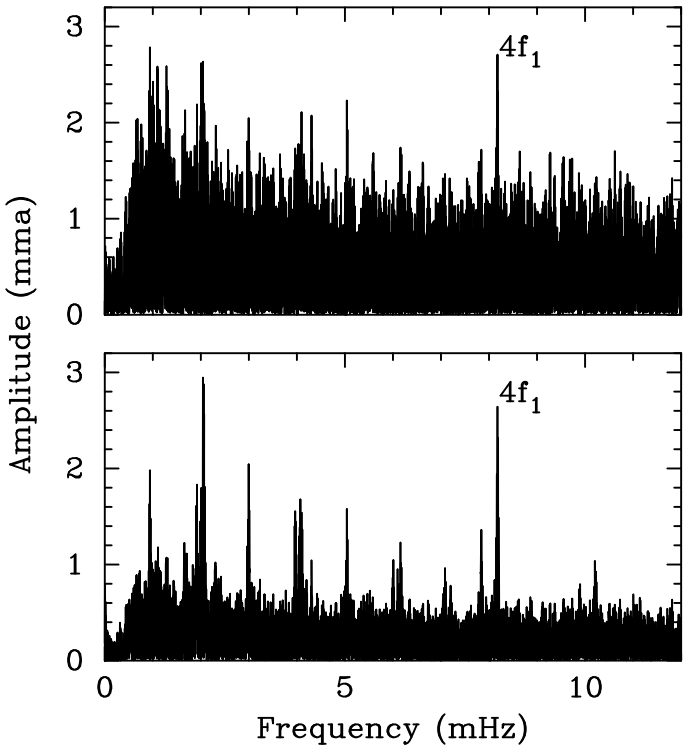

Fig. 4. Upper panel: unweighted residual amplitude spectrum of a subset of the XCov12 data of PG 1351+489. Some of the strongest signals have previously been prewhitened, but the 3rd harmonic of the strongest mode has been left in for reference and is indicated. Lower panel: the amplitude spectrum of the same data set, but weighted with the inverse variance of the individual runs.

be known a priori, that the frequency dependence of the noise is disregarded, that small stretches of poorer data quality (e.g. during a cloud break, twilight, observations at high air mass) will produce lower weights also for the best parts of the same run, and that the method cannot be well applied to very short runs or to runs with large gaps.

It must also be kept in mind that the noise level in an amplitude spectrum depends on the number of measurements. Consequently, the inverse of the FT noise must be divided by the square root of the number of data points in the corresponding run to obtain the correct weight. Therefore, this method is very similar to using the inverse variance of the run-by-run light curve residuals as weights.

3.2. Sigma-cutoff weighting

This scheme recently became prominent in the analysis of δ Scuti star observations, e.g. see Breger et al. (2002) or Rodriguez et al. (2003). After the frequency analysis of the unweighted data is completed, a synthetic light curve computed from the detected pulsational signals is subtracted from the data; some people remove remaining trends in addition. The residuals are assumed to consist of noise only and their mean standard deviation $\bar{\sigma}$ is computed. The individual data points are now weighted according to:

$$w_i = 1 \quad \text{if} \quad \sigma_i \leq K\bar{\sigma} \tag{1}$$

$$w_i = (K\bar{\sigma}/\sigma_i)^x \quad \text{if} \quad \sigma_i > K\bar{\sigma} \tag{2}$$

where the w_i are the individual weights, σ_i is the rms residual of the ith data point and K and x are free parameters. These parameters are either chosen ad hoc or according to experience. For instance, Breger et al. (2002) used $K = 1.5, x = 2.0$ and Rodriguez et al. (2003) chose $K = 1.0, x = 2.0$.

Positive aspects of this method are that it has some sensitivity to varying observing conditions, that outlying data points are effectively taken into account and that it can be applied even to very short runs. On the other hand, the method gives full weight to poor data that just happen to lie within the specified limits, the assignment of the weights is somewhat tedious and most importantly it depends on a predetermined solution and it may be affected by intrinsic residual variations if these are not filtered out. The interested reader is referred to Rodriguez et al. (2003) for a more detailed discussion of this method.

3.3. Light curve variance weighting

The last scheme to be tested starts by determining the intrinsic scatter in the light curves by calculating the point-to-point intensity differences of consecutive data points. Such high-pass filtering also removes the intrinsic variability (if the light curve is properly sampled). The time series of the point-to-point scatter is then boxcar-smoothed over a selectable time interval and the inverse of resulting function then gives the weight per point. I thought that this method was my idea until I learned it is merely a small improvement over a scheme by Michael Viskum (Aarhus University) that was applied e.g. by Arentoft et al. (1998).

Changes in the observing conditions are best taken into account with this method. It is simple because it can be used directly on the

observed light curves and it treats inhomogeneous data sets properly. Its main disadvantage is that single outlying data points[2] give longer stretches of the light curve lower weight. In addition, data gaps in which significant light variations have occurred (e.g. even sky measurements!) will affect the weights and very short runs may be given somewhat incorrect weights.

4. TEST DATA SETS

I have used two different data sets on pulsating DB white dwarf stars to examine the effects of weighting, the efficiency of the three weighting schemes, and of the contribution of small telescope data to the overall result.

The first one is a subset of the XCov12 observations of PG 1351+489. The star is rather faint ($B = 16.4$) and was observed with telescopes between 0.6 and 2.5 metres aperture. The second test data set is on EC 20058-5234 from XCov15, a brighter star ($V = 15.5$) observed with telescopes between 1.0 and 1.6 metres aperture. The light curve residuals after subtracting a number of known pulsational signals are shown in Fig. 5 to illustrate data quality and homogeneity.

The reason for the choice of just these two data sets is clear: they are intended to be extreme examples of WET data sets in terms of homogeneity. The scatter of the poorest run on PG 1351 is almost eight times higher than the scatter of the best run; the poorest run on EC 20058 has 2.4 times the scatter of the best run. With these two data sets the applicability of different weighting schemes and the effects of using large and small telescopes can thus be tested comprehensively.

I stress that I refrain from using simulations in such tests. Real data have real noise and a real temporal distribution; all this is irreplaceable here.

5. THE TEST

Before applying the three different weighting schemes to the test data, the basic idea of the experiment needs to be outlined. As stated above, the maximum amount of astrophysical information

[2]However, these are often a product of poor data reduction and should hence be re-examined rather than being "weighted away"!

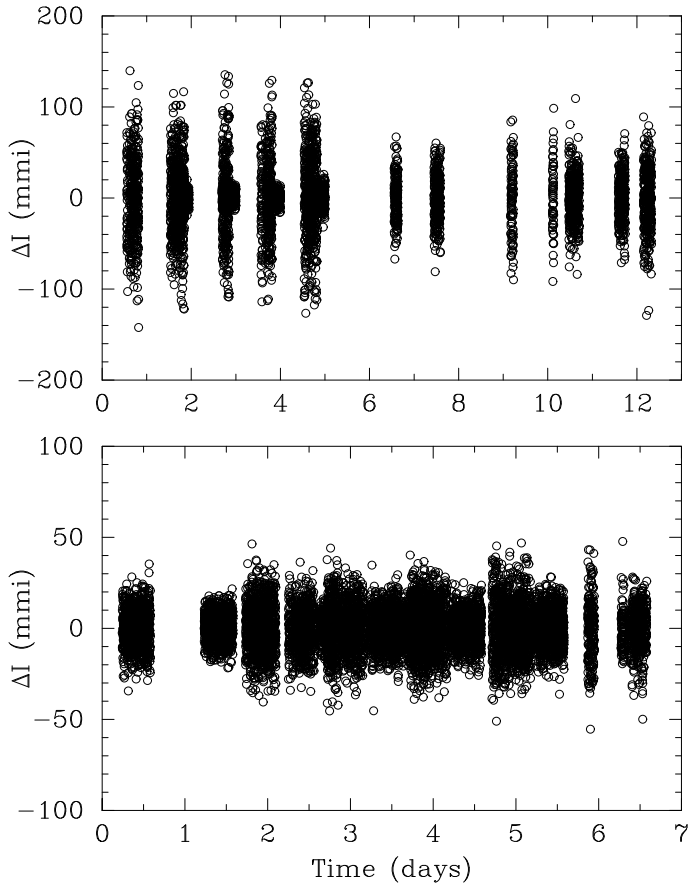

Fig. 5. Upper panel: the residual PG 1351 data. Note the excellent-quality data near days 2–5 just after the light curves with largest scatter. Lower panel: the residual EC 20058 data.

should be extracted from the data. In most cases, this problem translates into detecting the maximum possible number of intrinsic frequencies present in the light curves. This means that the S/N of the intrinsic signals in the FT should be maximized.

To this end, I performed frequency analyses of the test data sets and prewhitened a number of the strongest signals. I did however leave some variations I believe to be intrinsic (because they are combination frequencies) that are close to the limits of detectability in the data. These are the signals that should reach the highest possible S/N by application of the different weighting schemes.

To determine the S/N ratio I have used the method proposed by Breger et al. (1993). It calculates the noise level as the average amplitude in an oversampled ($\Delta f = 1/20\Delta T$, where Δf is the sam-

pling interval in the spectrum and ΔT is the length of the data set) amplitude spectrum centered on the suspected frequency. The ratio of the amplitude of the signal under consideration to the local mean noise level is calculated and defines its S/N.

It should be pointed out that any other false alarm criterion etc. could be used instead. My choice of this criterion is based on my extensive experience with it, its simplicity and because it works directly on signal and noise amplitudes and can therefore be easily visualized. For the reader inexperienced with this criterion it will be useful to know that a noticeable, but not particularly strong noise peak usually has a S/N of about 3, a combination peak to be accepted as intrinsic must satisfy $S/N > 3.5$ and an independent frequency must exceed $S/N = 4.0$ to be judged intrinsic. So far, no signal ever detected with this criterion (if applied properly) had to be rejected later.

In any case, we are now in a position to calculate the S/N of the test signals left in the data and to use them to evaluate the efficiency of the different weighting methods. A frequency interval of 10 c/d was chosen to determine the noise level around the signals of interest, but the actual choice of the size of this interval is not important here.

5.1. Unweighted data

The two test signals left in the PG 1351 light curve residuals reached an average S/N of 3.68 and would thus be marginally acceptable as combination frequency signals. I used four test signals in the EC 20058 data that reached an average S/N of 4.74, a clear detection.

5.2. Fourier noise weighting

I then proceeded with the data weighted according to Section 3.1. However, as it is unclear (because of the interplay of noise decrease, effective removal of data points and spectral window) whether the normalized inverse FT noise itself would result in the optimal weight, I used several different power laws ($w_x = w_i^x, x$ being a free parameter) to look for an optimal result.

The best average S/N for the signals in the PG 1351 data ($(S/N)_{\max} = 4.25$) was obtained for $x = 0.8$. The same test applied to the EC 20058 test data gave a maximum S/N of 4.82 for $x = 0.7$. As expected, the weighting of the data yields a substantial improvement in the results for PG 1351 compared to the unweighted data, whereas the increase in S/N for the EC 20058 data is only

marginal.

5.3. Sigma-cutoff weighting

As demonstrated in Sect. 3.2, this weighting scheme has two free parameters, the bandwidth $K\bar{\sigma}$ containing only data points of unit weight, and the exponent x by which the weight of the data points outside this region decreases. I attempted to find the optimal values for K and x that will result in best S/N for the test signals.

For the PG 1351 data, a maximum average S/N of 4.30 was achieved for $K = 0.3, x = 1.0$. The best average S/N of the four test signals in the EC 20058 light curves was 4.86 for $K = 1.4, x = 0.85$. These results are similar than those for the previous weighting scheme, but give slightly higher significance.

5.4. Light curve variance weighting

This weighting method also has two free parameters to be adjusted. The first is the width W of the boxcar used to smooth the point-to-point scatter of the light curves and the second is again the exponent x specifying the decrease of the weight according to the data variance estimate. If the parameter W approaches the length of a run, this method would be very similar to FT noise weighting. We note that W must not be too small (at least the size of the mean variability period) because it will then give very high weight to only a few data points and will hence modify the spectral window extremely.

Optimal results with this method were: $(S/N)_{\max} = 4.39$ for $W = 15\,\mathrm{min}$ and $x = 0.85$ applied to the PG 1351 light curves, as well as $(S/N)_{\max} = 4.91$ for the EC 20058 data with $W = 4\,\mathrm{min}$ and $x = 0.72$. Again, we see a small improvement over the outcome of the previous weighting strategies.

6. DISCUSSION

A summary of the results of the previous section is given in Table 1. The best S/N that could be achieved and the parameter x, the exponent of the power law used to weight the poorer-quality data are quoted. The other free parameters for some of the methods tested depend on the individual structures of the data sets and are of no further importance.

Table 1. Comparison of the different weighting schemes.

Data set	PG 1351+489		EC 20058-5234	
	$(S/N)_{\max}$	x	$(S/N)_{\max}$	x
No weights	3.68		4.74	
FT noise weight	4.25	0.8	4.82	0.7
Sigma-cutoff weight	4.30	1.0	4.86	0.85
Light curve scatter weight	4.39	0.85	4.91	0.72

The best results are achieved by weighting according to the inverse local data variance in the light curves; this method is therefore recommended. However, the improvement relative to the other weighting schemes is not particularly dramatic.

One result which is at first glance somewhat surprising are the low values of x which resulted very consistently. They indicate that the weights of the data points should decrease not as steeply as their inverse variance. In particular, the value of $x = 2$ which offhandedly seems a logical choice (also implied by the work of Moskalik 1993) and is often employed in the literature as well (see Sect. 3.2) turns out not to be favourable!

The reason is that high values of x disregard too many data points which would be needed to suppress the noise in the FT. We have investigated this idea by determining the optimal S/N of the two signals for the PG 1351 data for $x = 1.5$ (which would be 4.24). In this case, the optimal K increases from 0.3 (for $x = 1$) to 0.45, i.e. more data points are now needed to have unit weight to give best S/N in the combined amplitude spectrum.

Now we can return to the original question examined in this work: how useful are the contributions of the small telescopes for the astrophysical outcome of a whole WET run? The answer is already indicated in Table 1 and is presented in Fig. 6, where we examine the mean weight of each telescope's contribution to the whole run determined from the best weighting method and its corresponding optimal parameters.

Examining the result for the EC 20058 test data set first, it is found that the smaller telescopes contribute more to the total S/N than their measurement accuracy would imply. These telescopes are therefore quite valuable. One may also suspect that the telescope generating the data of the highest scatter has some problems, but it does not. As a matter of fact, it is just the excellent performance of

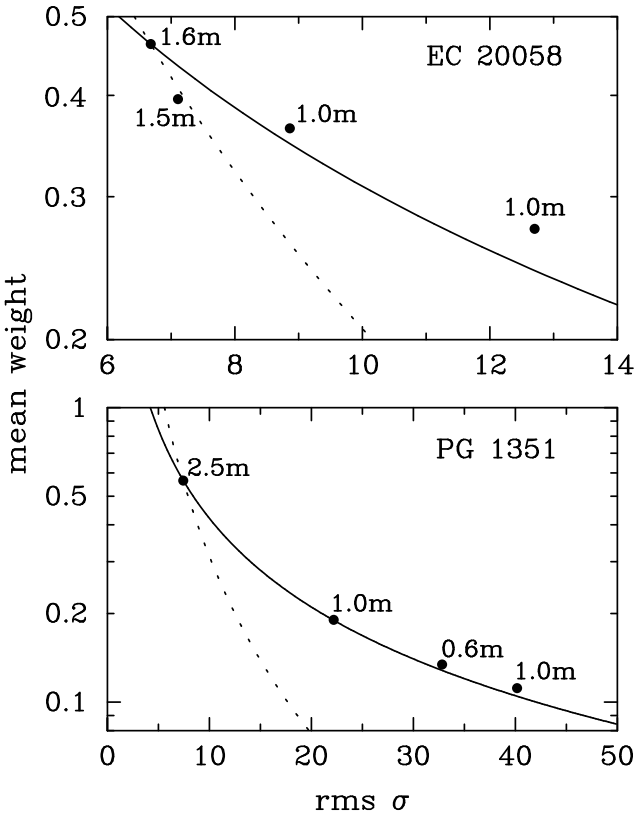

Fig. 6. Upper panel: mean weight of the data from the different telescopes used to observe EC 20058 versus rms scatter of the light curve. The full line corresponds to a weight inversely proportional to the rms scatter of the data; the dotted line indicates weighting with the sqaure of the inverse variance. Lower panel: same as above, but for PG 1351.

the other 1.0 m telescope in the campaign that creates this impression.

Turning to PG 1351, it is clear that this data set is dominated by the by far largest telescope. But even here, the smaller telescopes contribute more to the total result than their apertures would suggest. There is one notable exception, however: one of the two 1.0-m telescopes actually produced worse data than the 0.6 m telescope[3]. We summarize that in both cases the contribution of the smaller telescopes to the overall result was more valuable than expected.

For reasons of completeness, we finally examine how the spectral windows of the two test data sets are changed by the weighting

[3]The probable reason is a photometer aperture that was too large.

procedure. We show a comparison of the unweighted window and that of optimally weighted data in Fig. 7.

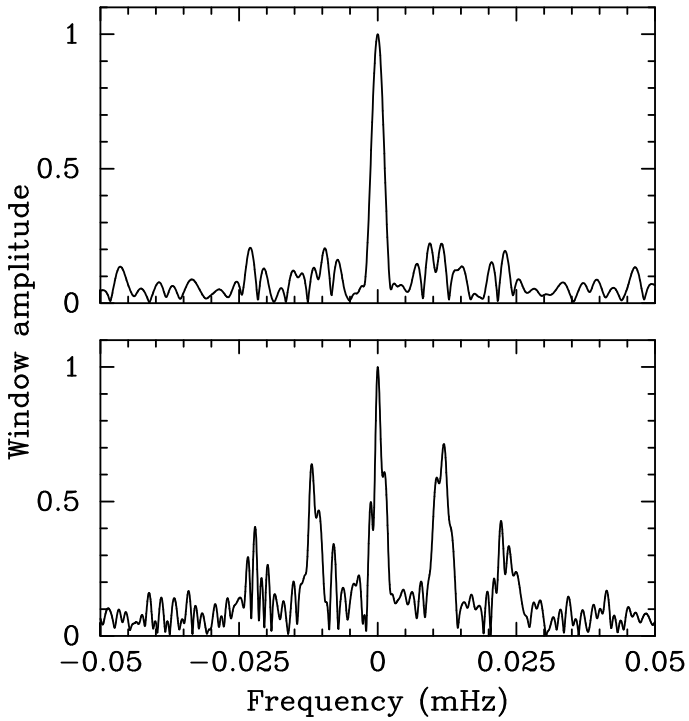

Fig. 7. Upper panel: spectral window of the EC 20058 test data set. The unweighted window is shown at negative frequencies, whereas the weighted window extends over the positive frequencies. Lower panel: same for the PG 1351 test data.

This figure has no surprises to offer. The spectral window for EC 20058 remains almost unaffected except for somewhat stronger sidelobe structures. The window for PG 1351 shows more notable differences between weighted and unweighted data. Besides the increased amplitude of the daily alias, the window peaks also become wider as the data set is weighted in favour of the four nights of very good data from the large telescope. Consequently, weighting may not only affect the alias structure, but also our ability to resolve closely spaced frequencies.

7. CONCLUSIONS

To evaluate the contribution of small telescopes to the astrophysical results to be obtained by the WET, their weight in the overall solution was examined. To this end, different weighting methods were applied to two WET data sets and tested at the same time. Unsurprisingly, weighting most efficiently reduces the noise for qualitatively inhomogeneous data sets. It appears that weighting proportional to the inverse local scatter in the data gives best results. Interestingly, weighting by a function steeper than the inverse variance was not found to be optimal because it reduces the effective number of data points.

Consequently, the small telescopes were found to be very important for the success of a WET run. Not only do they improve the spectral window, they also contribute more to the noise reduction in the combined FT than naively expected. On the other hand, there is not much point in filling a few gaps with noisy data. To be more specific, measurements with a scatter about 3–4 times higher than that of the best data in a given run will no longer reduce the noise level in a combined FT considerably. Such telescope time may be better spent on an alternate target.

We consider the small telescopes an asset for the WET. It must also be kept in mind that in most cases longer allocations of time are possible on small telescopes, i.e. they make up for their generally poorer data quality by providing larger data quantities.

These conclusions may also be useful for the planning of future WET runs, in particular with the increased use of CCDs. Telescope size then becomes less important for data quality, but the latter should (in an ideal world) be well known before asking for telescope time and a particular instrument. Permanent monitoring of data quality (e.g. by simple estimates of the point-to-point scatter) at headquarters will also help in scheduling a run on the fly. Ideally, a WET run should produce data of the most homogeneous quality possible. This would also make procedures like data weighting superfluous.

ACKNOWLEDGMENTS. I thank the organizers of this WET workshop for triggering interesting discussions on “hot” (sometimes even unpleasantly hot) topics for the benefit of the whole network of collaborators.

REFERENCES

Arentoft T., Kjeldsen H., Nuspl J., Bedding T., Frontó A., Viskum M., Frandsen S., Belomnte J. A. 1998, A&A, 338, 909

Breger M. et al. 1993, A&A, 271, 482

Breger M. et al. 2002, MNRAS, 329, 531

Dind Z. E., Bedding T., O'Toole S. J., Kawaler S. D. 2003, in *Asteroseismology Across the HR Diagram*, in press

Moskalik P. 1993, Baltic Astronomy, 2, 485

Nather R. E., Winget D. E., Clemens J. C., Hansen C. J., Hine B. P. 1990, ApJ, 361, 309

Rodriguez E., Costa V., Handler G., Garcia J. M. 2003, A&A, submitted