*International Hydrographic Review*, Monaco, LVII (1), January 1980.

# AN IMPROVED METHOD FOR SMOOTHING AND INTERPOLATING HOURLY SEA LEVEL DATA

by D. A. KARUNARATNE
Institute of Oceanographic Sciences, Bidston Observatory, Birkenhead, U.K.

## ABSTRACT

A method for smoothing and interpolating observed hourly digital tidal elevations is presented. The method which is based on fitting a small number of harmonic tidal components to the data series has proved to be superior to standard smoothing procedures and can be applied to data from different tidal regimes. The method is relatively simple, economical, and can be executed on standard digital computers with limited storage capacity.

## INTRODUCTION

Very few of the instruments which are used in sea elevation measurement are provided with digitised output, and as a consequence the greatest proportion of hourly sea level time series are obtained by digitisation of the tidal charts from conventional analogue type tide gauges where the height is recorded as a continuous function of time. Irrespective of the attention given by operators in reducing the traces into digital hourly values, with a human error of approximately 1 % there is always a need to check and verify the reduced data for possible errors. Even for automated digital output, verification procedures have to be carried out to ensure against possible faults in the digitising mechanism of the tide gauge.

In the past, the only satisfactory simple method of automatic verification of digitised hourly data was by means of one or two Lagrangian interpolation formulae (LENNON 1964, 1965).

Although this method has proved of great benefit in the routine processing of tidal records, it has shown to be unsatisfactory as an efficient smoothness check for data, resulting in (a) a large volume of pseudo-error output and (b) ignoring some genuine errors, especially for tidal regimes where the basic sinusoidal character of the tide is overshadowed by other harmonics.

A step by step method of estimating the hourly sea levels based on representing a few principal harmonic tidal components is shown to produce considerable improvement over the Lagrangian interpolation methods. The method is simple and economical to program and run on a standard electronic digital computer and can also be extended to fill limited gaps, of up to about twelve hours, in the recorded observations.

The work presented here is part of the procedures being developed at the Bidston Laboratory of IOS in producing a systematic computer-aided scheme for the digitisation and routine processing of sea elevation records, to far more critical standards than hitherto attained.

## DISCUSSION OF SOME OF THE EXISTING METHODS

The most widely used conventional method of checking the smoothness of a tidal time series is by applying two of the following four Lagrangian interpolation formulae:

(i) $\zeta_t^* = 1/6(-\zeta_{t-50} + 4\zeta_{t-25} + 4\zeta_{t+25} - \zeta_{t+50})$

(ii) $\zeta_t^* = 1/6(-\zeta_{t-2} + 4\zeta_{t-1} + 4\zeta_{t+1} - \zeta_{t+2})$

(iii) $\zeta_t^* = -0.0049\,\zeta_{t-7} + 0.0410\,\zeta_{t-5} - 0.1709\,\zeta_{t-3} + 0.6836\,\zeta_{t-1} + 0.5127\,\zeta_{t+1} - 0.0684\,\zeta_{t+3} + 0.0068\,\zeta_{t+5}$

(iv) $\zeta_t^* = 0.05\,\zeta_{t-3} - 0.30\,\zeta_{t-2} + 0.75\,\zeta_{t-1} + 0.75\,\zeta_{t+1} - 0.30\,\zeta_{t+2} + 0.05\,\zeta_{t+3}$

where $\zeta_t$ is the observed elevation and $\zeta_t^*$ is the estimated elevation at time t hours.

LENNON (1964, 1965) suggested the use of (i) followed by (ii) and treated values $\zeta_t$ which failed both tests as suspect, requiring scrutiny. The reason for using two tests rather than a single test was attributed to the shortcomings of (i) and (ii) used independently. The principal shortcoming of (i) was due to the inability of this method to cope with the effects of meteorological phenomena such as storm surges which could generate an appreciable lowering or raising of the astronomical tidal elevation over a period of one to two days. Although this discrepancy could be taken into account by method (ii) due to the shorter span of data considered in the interpolation scheme, method (ii) on its own has a different but equally serious shortcoming in its inability to cope with the conditions at or near high or low water, especially where the symmetry of the basic sinusoidal character of the tide is destroyed due to shallow water effects.

Thus a sequential combination of (i) and (ii) was used, and found to be satisfactory at the time. The criterion of acceptance used is that if any value $\zeta_t$ passed one of the two tests, e.g. $|\zeta_t - \zeta_t^*| <$ some tolerance value, it would be acceptable as smooth and if it failed both tests, it would be treated as a possible error, to be later verified manually in reference to the original recorded tidal trace.

It should be noted that methods (i) and (ii) do have certain advantages over most other schemes. Method (i) is valuable in detecting errors due to misreading sea levels in the digitising process, these being either misplaced by an hour, or more seriously, read as a sequence of several hours from a wrong day. Such errors may escape detection by most other methods which depend upon only a few adjacent data values being available for the interpolation procedure, especially if the displaced values happen to match smoothly onto the surrounding, albeit wrong, data. Method (i) makes such errors easily detectable. Equally the main advantage in method (ii) is the relatively shorter span of data required for the interpolation scheme. This tends to minimise the possibility of masking smaller errors by a single large error within the span of data. However, this becomes relevant only if the errors produced (i.e. $|\zeta_t - \zeta_t^*|$) are in proportion to the actual misread data errors. It is shown later that this criterion of proportionality is not satisfied by the conventional Lagrangian methods used for tidal data smoothing.

Methods (iii) and (iv) above are also used as alternatives to method (ii) in routine smoothness checking of tidal elevations, where (iv) (CARTWRIGHT, 1968) is essentially the six-parameter member of a family of which (ii) is the four-parameter member.

However over the years IOS Bidston and other laboratories concerned with digitising tidal records have found that even the best considered combinations of (i), (ii), (iii) and (iv) tend to result in either producing a large number of pseudo-errors or ignoring some genuine errors, or both; thus resulting in a tedious manual task of rechecking the original digitised data (GREIG, 1977).

Other methods which are occasionally used for the smoothing of tidal time series can be summarised as follows:

(a) Visual inspection of graphical form of the observed time series on a suitable scale.
(b) Visual inspection of graphical form of the residuals between observed series and a predicted (harmonic) series.
(c) *Formal band-pass filtering of the tidal series and the inspection* of the residuals.

Selecting any one of these methods is subjective and depends on the expertise of the user and on the levels of accuracy required in the final digitised series. It is of some interest to note that at IOS Bidston, where there is a long established involvement in tidal analysis and prediction procedures, the method (b) above is being increasingly adopted as a critical means of aiding the assessment of quality of digitised tidal records.

## DEVELOPMENT OF A REVISED SMOOTHING TECHNIQUE: HARMONIC COMPONENT FIT METHOD (HCFM)

All the computer-aided methods in current use for the smoothing of tidal time series, except for the method of interpreting residuals (using harmonic predictions), are based wholly on applying standard numerical interpolation procedures. Since none of these methods makes use of the basic characteristics of the tide generating force, it would be of value to consider a simple (in computing terms) method, using the harmonic components derived from the tide generating potential.

In terms of the tide generating potential, the height of the observed tide $\zeta$ at time t can be expressed as,

$$\zeta(t) = A_0 + \sum_{n=1}^{N} R_n \cos(\sigma_n t - \phi_n) + S(t) \tag{1}$$

where

$A_o$ is the height of mean sea level above chosen datum.
$N$ is the number of harmonic constituents used.
$R_n$ is the amplitude of the $n^{th}$ constituent.
$\Phi_n$ is the phase of the $n^{th}$ constituent.
$\sigma_n$ is the speed of the $n^{th}$ constituent.
$S(t)$ is the noise due to meteorological disturbances.

This can be written as,

$$\zeta(t) = a_0 + \sum_{n=1}^{N} (a_n \cos \sigma_n t + b_n \sin \sigma_n t) \tag{2}$$

where

$$a_n = R_n \cos \phi_n \quad \text{and} \quad b_n = R_n \sin \phi_n$$

For a set number of selected components, N, we can solve the $2N + 1$ simultaneous equations represented by (2), to obtain $a_o$, $a_n$ and $b_n$ ($n = 1, \ldots, N$), by substituting for $\sigma_n$ and discrete values $\zeta_t$, the digitised representation of $\zeta(t)$ at hourly intervals. (Since only a few values are considered we can assume $S(t)$ in (1) is incorporated in $a_o$.) By selecting $\zeta_t$ such that $t = -(N + 1), -N, \ldots, N$ and $t \neq o$, and solving the system of equations (2) we readily obtain the set of constants $a_o$, $a_n$ and $b_n$ which can be substituted back into the right hand side of equation (2) with $t = 0$, to obtain an estimate for $\zeta_{t=o}$.

Taking four harmonic components, i.e. $N = 4$, the system of equations would be,

$$a_0 + \sum_{n=1}^{4} (a_n \cos \sigma_n t_r + b_n \sin \sigma_n t_r) = \zeta t_r \tag{3}$$

where

$$r = -5, -4, \ldots\ldots\ldots, 4 \; ; \; r \neq 0$$

After solving (3) for $a_0$, $a_n$ and $b_n$ ($n = 1, \ldots, 4$) we can obtain an estimate $\zeta^*$ for $\zeta_{t=0}$,

(4)

Considering a numerical procedure, which can be used in solving the system of equations (2), we can represent the system in matrix form:

$$\mathbf{A}\,\mathbf{X} = \mathbf{Z} \tag{5}$$

where

**X** is a $(2N + 1)$ column vector representing the unknown coefficients.
**A** is a $(2N + 1) \times (2N + 1)$ matrix with cosine and sine terms of the left hand side of the system.
**Z** is a $(2N + 1)$ column vector with $\zeta$ terms.

This system (5) can be solved easily by using standard Gaussian elimination procedure with a scaled partial pivoting strategy. It is of use to note that once the tidal components to be used are chosen the matrix **A** remains unaltered for all the values under investigation, and an example is given in table 4. Such a scheme enables further simplification in devising a solution procedure, particularly in use of a greater number of harmonic component terms.

The tidal harmonic components considered here are a fundamental term ($M_1$) of the diurnal tidal band and its higher harmonics ($M_2$, $M_3$, $M_4, \ldots$) from other tidal bands. These components were chosen since they tend to be the central harmonics of the respective principal species in the tidal spectrum. (Using other constituents with very similar angular speeds, i.e. $S_2$ instead of $M_2$, seems to make little difference to the outcome of the final results.) Although it could be rather tempting to increase the number of components considered in the system in order to obtain better elevation estimates, it is important to ensure the stability of the system against ill-conditioning when solving the matrix equation (5). The requirement to keep the number of data values needed in the smoothing scheme to a minimum in order to minimise the masking effects of small errors due to a single large error (within the span of data), should also be considered, along with the need to keep the method economical (in terms of computation involved) when deciding on the number of components to be used.

Once the estimate $\zeta^*_{t=0}$ is obtained, it is compared, as for conventional methods, with the observed elevation $\zeta_{t=0}$, against a pre-set tolerance level.

## COMPARISON OF CONVENTIONAL LAGRANGIAN INTERPOLATION METHODS WITH THE HCFM

In order to obtain the theoretical limits of performance of these various methods one should test them using the best (in terms of the smoothness) data available. In order to satisfy this condition, one hundred hourly values of a harmonically predicted tidal series at Southend were chosen.

These predicted data are based on an extended harmonic model devised by AMIN (1976). Since harmonic predictions are the limit in smoothness one can achieve when dealing with tidal elevations, one would theoretically expect any method used to investigate the smoothness of the series to produce zero errors.

Initially the tidal component $M_1$ and three of its higher harmonics ($M_2$, $M_3$ and $M_4$) were chosen for the HCFM. The number of components was then increased, one at a time, up to a maximum of 10 ($M_i$, i = 1, 10) and in each test the mean error and its standard deviation were computed in running the HCFM.

The method was also tested with four tidal components $M_1$, $M_2$, $M_4$ and $M_6$, together with a variety of other combined components. Since the simplest HCFM considered uses nine parameters as indicated in (3), one has to compare this method with a Lagrangian interpolation scheme which uses at least nine parameters. Hence the mean error and its standard deviation produced by conventional Lagrangian methods (ii), (iii), and by a new revised Lagrangian nine point interpolation formulae, namely,

$$\text{(v)}\quad \zeta_t^* = 0.0079\,\zeta_{t-5} - 0.0714\,\zeta_{t-4} + 0.2857\,\zeta_{t-3} - 0.6667\,\zeta_{t-2} + \zeta_{t-1} + 0.6667\,\zeta_{t+1} - 0.2857\,\zeta_{t+2} + 0.0714\,\zeta_{t+3} - 0.0079\,\zeta_{t+4}$$

were computed.

Tables 1 and 2 show the mean and the standard deviation of errors produced by each of these methods when tested on the one hundred hourly values of the smooth predicted series at Southend.

**Table 1**

*Mean and standard deviation of errors produced by Lagrangian methods (ii), (iii) and (v), for one hundred harmonically predicted hourly values at Southend.*

| No. of hourly values | Max. range of data in metres | Lagrangian interpolation methods | | | | | |
|---|---|---|---|---|---|---|---|
| | | (ii) | | (iii) | | (v) | |
| | | Mean | SD | Mean | SD | Mean | SD |
| 100 | 5.399 | 0.001 | 0.071 | −0.001 | 0.098 | 0.000 | 0.032 |

On examining Tables 1 and 2, it becomes clear that the revised Lagrangian formula (v) is a considerable improvement on the conventional Lagrangian methods (ii) and (iii), and the HCFM (with any combination of components tested) is superior to any of the Lagrangian methods. The HCFM itself tends to give better estimates as the number of components considered increases, reaching a limit at nine terms. This limit of convergence may be due to the loss of stability of the linear system of equations solved, for more than 9 components. It could also be due to the

**Table 2**

*Mean and standard deviation of errors produced by HCFM (with different combinations of harmonic components* $M_1$–$M_{10}$*), for the one hundred harmonically predicted hourly values at Southend.*

| No. of hourly values | Max. range of data in metres | HCFM | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | $M_1$ to $M_4$ | | $M_1$ to $M_5$ | | $M_1$ to $M_6$ | | $M_1$ to $M_7$ | |
| | | Mean | SD | Mean | SD | Mean | SD | Mean | SD |
| 100 | 5.399 | 0.000 | 0.027 | 0.000 | 0.020 | 0.000 | 0.015 | 0.000 | 0.013 |
| | | $M_1$ to $M_8$ | | $M_1$ to $M_9$ | | $M_1$ to $M_{10}$ | | $M_1, M_2, M_4, M_6$ | |
| | | Mean | SD | Mean | SD | Mean | SD | Mean | SD |
| 100 | 5.399 | 0.000 | 0.012 | 0.000 | 0.011 | 0.000 | 0.012 | 0.000 | 0.022 |

additional data values used at the end points, in determining the unknown parameters, which becomes significant only as the theoretical limit of convergence of the standard error is reached. However a compromise may be made by keeping the number of components considered below nine and taking every other harmonic term of the fundamental $M_1$, i.e. $M_1$, $M_2$, $M_4$, $M_6$, .... instead of every sequential harmonic (M) term. This becomes evident on comparing the standard deviation of the errors produced by taking $M_1$, $M_2$, $M_3$, $M_4$ (0.000 ± 0.027) with those produced by taking $M_1$, $M_2$, $M_4$, $M_6$ (0.000 ± 0.022). However the choice of components can be subjective, and depends on the accuracy required by the individual user.

In figure 1, errors produced by Lagrangian formulae (ii), (iii), and (v) are plotted together with those produced by the HCFM (taking $M_1$, $M_2$, $M_4$ and $M_6$). This gives a fair indication of the improvements from the use of the HCFM.

Figure 2 shows the behaviour of the complete error function in Lagrangian methods (iii) and (v), and in the HCFM (with components $M_1$, $M_2$, $M_4$ and $M_6$). This illustrates that the Lagrangian methods are highly oscillatory in between the points taken for interpolation, with the least oscillatory behaviour in the HCFM.

Figures 1 and 2 further illustrate that the errors estimated by the Lagrangian methods bear little or no relationship to the errors found in the series of test data (zero errors expected for smooth predicted data), whereas the HCFM shows a much closer proportionality. Thus for series of real observed data (differing from a theoretical harmonic curve) the HCFM is considered to respond directly to the order of errors found in the data, whereas this is not the case for Lagrangian methods which can have a pronounced random response as suggested even for harmonically predicted test data in figures 1 and 2.

Figure 3 shows the behaviour of the standard error, both in Lagrangian interpolation methods and in HCFM for the one hundred test data values, as the number of parameters considered is increased. It is readily seen that the HCFM converges rapidly to give a minimal standard error with increasing component terms whereas the same is not true of the analogous Lagrangian schemes.

Although a harmonically predicted series has been used to obtain the numerical results presented in this paper (to illustrate the theoretical limits), it should be made clear that, on tests carried out on observed

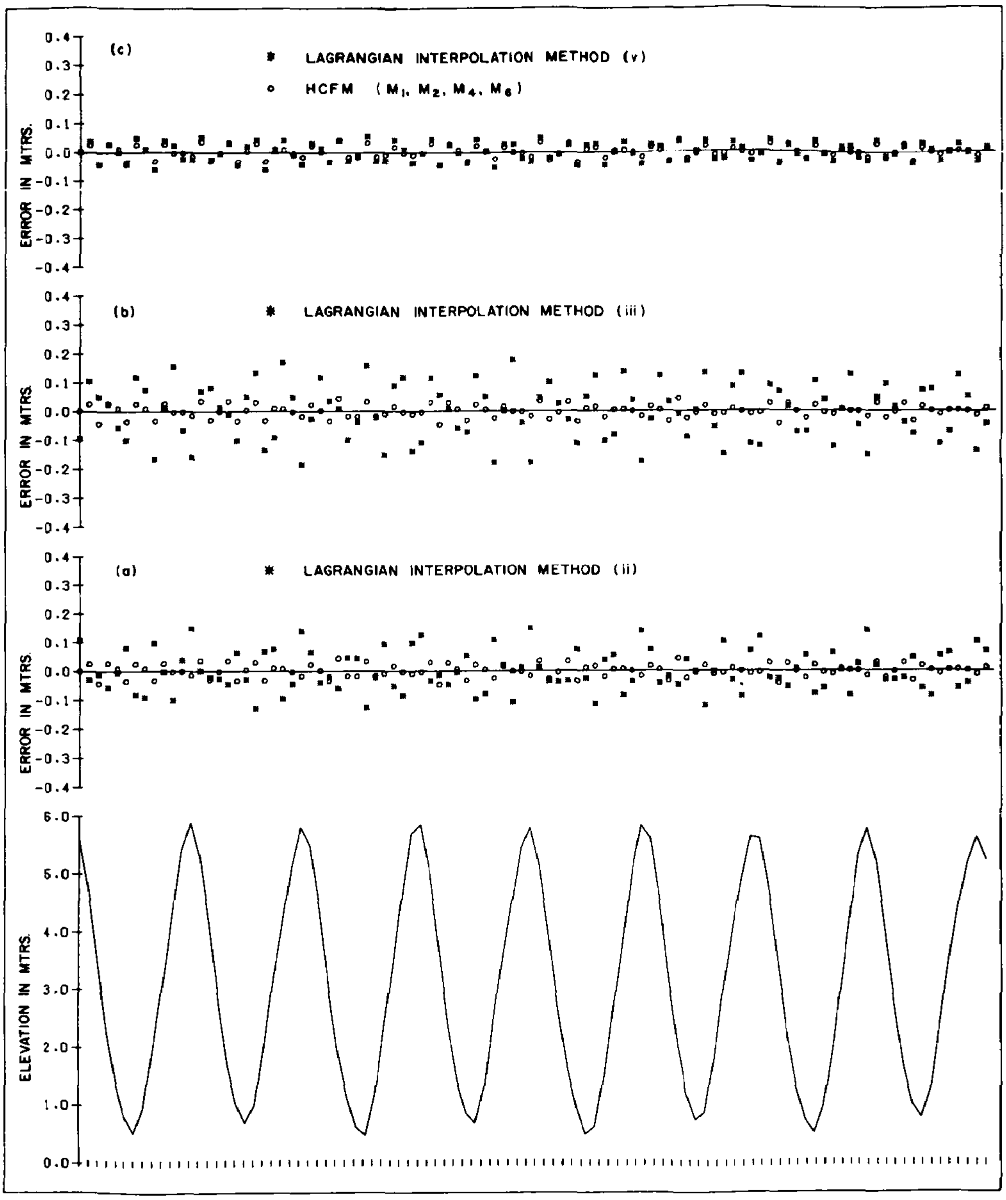


FIG. 1. — (a) Errors produced by Lagrangian interpolation method (ii) and HCFM for one hundred harmonically predicted hourly values at Southend. — (b) As for (a), with Lagrangian interpolation method (iii) and HCFM. — (c) As for (a), with Lagrangian interpolation method (v) and HCFM.

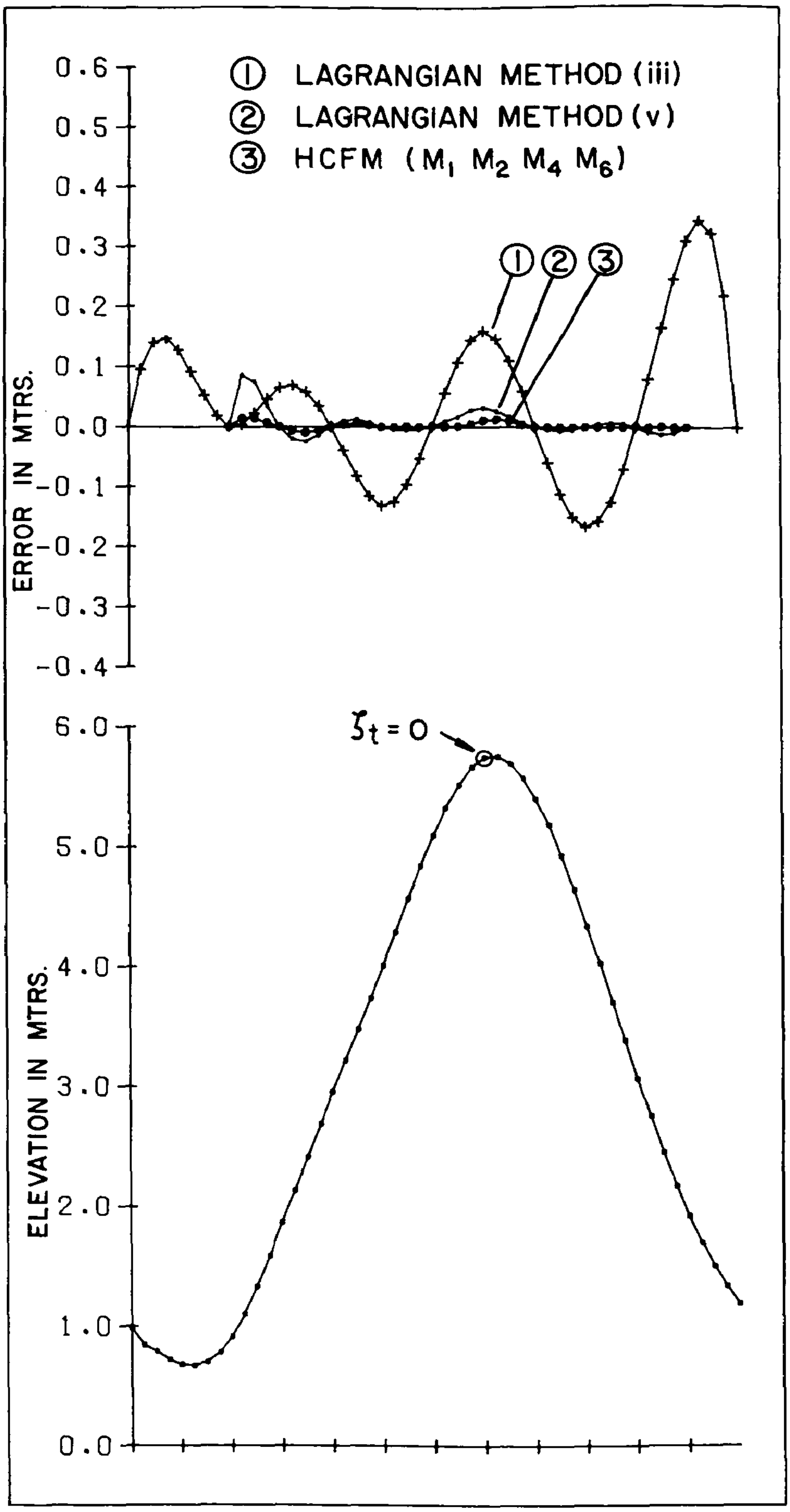


FIG. 2. — Complete error function produced on estimating $\zeta(t = 0)$ (as indicated) by Lagrangian interpolation methods (iii), (v) and HCFM (with $M_1$, $M_2$, $M_4$, $M_6$).

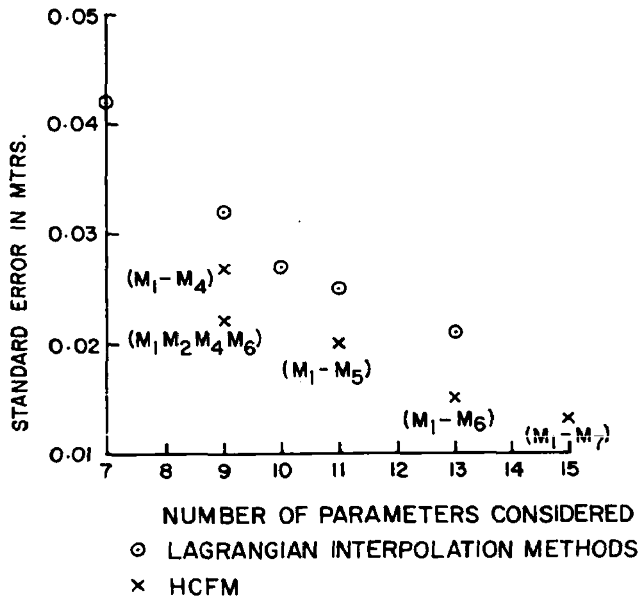


FIG. 3. — Behaviour of the standard error produced for the one hundred test values at Southend, by both Lagrangian interpolation methods and by HCFM with increasing number of parameters considered.

tidal data series (with superimposed meteorological phenomena), the superiority of the HCFM over the Lagrangian methods was very evident.

The HCFM was also tested for harmonically predicted hourly tidal data series from locations in tidal regimes other than the semi-diurnal dominant. They were:

(a) Lucinda—Queensland (marked diurnal inequality);
(b) Rabaul—Pacific (diurnal); and
(c) Anewa Bay—Pacific (mixed).

The results obtained, as given in table 3, show that the method is equally valid and successful for data series from all such diverse regimes.

It should be noted that in table 3, the means and standard deviations of the errors produced in all the different methods are given to four significant figures, only to indicate the consistency of the different methods. (Since the tidal ranges of the three locations considered are fairly small, the errors produced in applying all the methods are significant only in the fourth significant figure.)

## USE OF THE HCFM TO FILL LIMITED GAPS IN TIDAL ELEVATION TIME SERIES

So far the system of linear equations (5) has been solved exactly to estimate any $\zeta_{t=0}$ value. However by extending this to include more data

**Table 3**

*Mean and standard deviation of errors produced by Lagrangian methods* (ii), (iii) (v), *and HCFM, for harmonically predicted hourly values at Anewa Bay, Rabaul and Lucinda.*

| No. of hourly values | Port | Max. range of data in metres | Lagrangian interpolation methods (ii) Mean | SD | (iii) Mean | SD | (v) Mean | SD |
|---|---|---|---|---|---|---|---|---|
| 100 | Anewa Bay | 1.491 | 0.0001 | 0.0037 | – 0.0001 | 0.0015 | 0.0000 | 0.0006 |
| 100 | Rabaul | 0.913 | 0.0000 | 0.0012 | – 0.0001 | 0.0011 | 0.0000 | 0.0004 |
| 100 | Lucinda | 2.949 | –0.0002 | 0.0068 | – 0.0001 | 0.0030 | 0.0000 | 0.0006 |
| | | | HCFM | | | | | |
| | | | $M_1$ to $M_5$ Mean | SD | $M_1$ to $M_9$ Mean | SD | $M_1, M_2, M_4, M_6$ Mean | SD |
| 100 | Anewa Bay | 1.491 | 0.0000 | 0.0006 | 0.0000 | 0.0006 | 0.0000 | 0.0006 |
| 100 | Rabaul | 0.913 | 0.0000 | 0.0004 | 0.0000 | 0.0005 | 0.0000 | 0.0004 |
| 100 | Lucinda | 2.949 | 0.0000 | 0.0006 | 0.0000 | 0.0006 | 0.0000 | 0.0006 |

values than the number of unknowns in the system, thus creating an overdetermined linear system, one can solve for the unknowns in the system in a least squares sense. This strategy can be employed to estimate more than one data value i.e. to estimate several missing values in the time series, by taking M adjacent values on either side of the data gap (keeping $2M > 2N + 1$, where N is the number of tidal components used in the system), taking the first value in the gap as $t = 0$ and solving the system in a least squares sense.

The overdetermined system of equations to be solved (for a gap of j values) would be,

$$a_0 + \sum_{n=1}^{N} (a_n \cos \sigma_n t_r + b_n \sin \sigma_n t_r) = \zeta t_r \tag{6}$$

where

$$r = -M, -(M-1), \ldots\ldots\ldots, (M+j-1) \quad r \neq 0, 1, \ldots\ldots\ldots, (j-1)$$

Writing this in matrix form as before,

$$\mathbf{A}\,\mathbf{X} = \mathbf{Z} \tag{7}$$

where

**X** is a $(2N + 1)$ column vector with the unknown coefficients;

**A** is a $(2M) \times (2N + 1)$ matrix with the cosine and the sine terms of the left hand side of the system;

**Z** is a $(2M)$ column vector with $\zeta$ terms.

Since the system (7) is overdetermined, it will not be possible to satisfy all the equations simultaneously and hence for any particular solution,

**Table 4**

*Matrix* **A** *in the system of equations* (6) *with constituents* $M_i$ ($i = 1, 2, \ldots, 9$) *and* $r = -10, -9, \ldots, 9$; $r \neq 0$. *For a system with components* $M_1$, $M_2$, $M_3$ *and* $M_4$, *the relevant matrix* **A** *can be extracted as shown in the box outlined.*

| r | | $M_1$ | | $M_2$ | | $M_3$ | | $M_4$ | | $M_5$ | | $M_6$ | | $M_7$ | | $M_8$ | | $M_9$ | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| -10 | 1.00000 | -0.81653 | -0.57731 | 0.33935 | 0.94066 | 0.27201 | -0.96229 | -0.76968 | 0.63843 | 0.99792 | -0.06442 | -0.86174 | -0.50735 | 0.39347 | 0.91934 | 0.20945 | -0.97782 | -0.73552 | 0.67750 |
| -9 | 1.00000 | -0.64632 | -0.76306 | -0.15896 | 0.98729 | 0.85901 | -0.51196 | -0.94947 | -0.31387 | 0.36364 | 0.93154 | 0.46080 | -0.88751 | -0.97867 | 0.20542 | 0.78928 | 0.61403 | -0.04160 | -0.99913 |
| -8 | 1.00000 | -0.43509 | -0.90039 | -0.61745 | 0.78661 | 0.97581 | 0.21860 | -0.23752 | -0.97138 | -0.77764 | 0.62871 | 0.91076 | 0.41294 | -0.00937 | -0.99996 | -0.89627 | 0.44350 | 0.78928 | 0.61403 |
| -7 | 1.00000 | -0.19624 | -0.98056 | -0.92128 | 0.38890 | 0.55849 | 0.82951 | 0.69751 | -0.71658 | -0.83471 | -0.55069 | -0.36392 | 0.93143 | 0.98235 | 0.18704 | -0.00937 | -0.99996 | -0.97867 | 0.20542 |
| -6 | 1.00000 | 0.05507 | -0.99848 | -0.99434 | -0.10622 | -0.16454 | 0.98637 | 0.97743 | 0.21124 | 0.27201 | -0.96229 | -0.94947 | -0.31387 | -0.37618 | 0.92655 | 0.90443 | 0.42663 | 0.47579 | -0.87956 |
| -5 | 1.00000 | 0.30288 | -0.95303 | -0.81834 | -0.57474 | -0.79750 | 0.60332 | 0.33935 | 0.94066 | 0.99948 | -0.03222 | 0.26293 | -0.96482 | -0.83471 | -0.55069 | -0.77764 | 0.62871 | 0.36365 | 0.93154 |
| -4 | 1.00000 | 0.53147 | -0.84708 | -0.43735 | -0.89929 | -0.99394 | -0.10997 | -0.61745 | 0.78661 | 0.33344 | 0.94277 | 0.97743 | 0.21124 | 0.70379 | -0.71041 | -0.22774 | -0.97372 | -0.94586 | -0.32459 |
| -3 | 1.00000 | 0.72632 | -0.68736 | 0.05319 | -0.99858 | -0.64632 | -0.76306 | -0.99434 | -0.10622 | -0.79750 | 0.60332 | -0.14896 | 0.98729 | 0.55849 | 0.82951 | 0.97581 | 0.21860 | 0.85901 | -0.51196 |
| -2 | 1.00000 | 0.87506 | -0.48401 | 0.53040 | -0.84775 | 0.05507 | -0.99848 | -0.43735 | -0.89929 | -0.81653 | -0.57731 | -0.99434 | -0.10622 | -0.92298 | 0.38485 | -0.62139 | 0.78350 | -0.16454 | 0.98637 |
| -1 | 1.00000 | 0.96826 | -0.24994 | 0.87476 | -0.48456 | 0.72632 | -0.68736 | 0.53040 | -0.84775 | 0.30288 | -0.95303 | 0.05319 | -0.99858 | -0.19624 | -0.98056 | -0.43509 | -0.90039 | -0.64632 | -0.76306 |
| 1 | 1.00000 | 0.96826 | 0.24994 | 0.87476 | 0.48456 | 0.72632 | 0.68736 | 0.53040 | 0.84775 | 0.30288 | 0.95303 | 0.05319 | 0.99858 | -0.19624 | 0.98056 | -0.43509 | 0.90039 | -0.64632 | 0.76306 |
| 2 | 1.00000 | 0.87506 | 0.48401 | 0.53040 | 0.84775 | 0.05507 | 0.99848 | -0.43735 | 0.89929 | -0.81653 | 0.57731 | -0.99434 | 0.10622 | -0.92298 | -0.38485 | -0.62139 | -0.78350 | -0.16454 | -0.98637 |
| 3 | 1.00000 | 0.72632 | 0.68736 | 0.05319 | 0.99858 | -0.64632 | 0.76306 | -0.99434 | 0.10622 | -0.79750 | -0.60332 | -0.15896 | -0.98729 | 0.55849 | -0.82951 | 0.97581 | -0.21860 | 0.85901 | 0.51196 |
| 4 | 1.00000 | 0.53147 | 0.84708 | -0.43735 | 0.89929 | -0.99394 | 0.10997 | -0.61745 | -0.78661 | 0.33344 | -0.94277 | 0.97743 | -0.21124 | 0.70379 | 0.71041 | -0.22774 | 0.97372 | -0.94586 | 0.32459 |
| 5 | 1.00000 | 0.30288 | 0.95303 | -0.81834 | 0.57474 | -0.79750 | -0.60332 | 0.33935 | -0.94066 | 0.99948 | 0.03222 | 0.26293 | 0.96482 | -0.83471 | 0.55069 | -0.77764 | -0.62871 | 0.36365 | -0.93154 |
| 6 | 1.00000 | 0.05507 | 0.99848 | -0.99434 | 0.10622 | -0.16454 | -0.98637 | 0.97743 | -0.21124 | 0.27201 | 0.96229 | -0.94947 | 0.31387 | -0.37618 | -0.92655 | 0.90443 | -0.42663 | 0.47579 | 0.87956 |
| 7 | 1.00000 | -0.19624 | 0.98056 | -0.92128 | -0.38890 | 0.55849 | -0.82951 | 0.69751 | 0.71658 | -0.83471 | 0.55069 | -0.36392 | -0.93143 | 0.98235 | -0.18704 | -0.00937 | 0.99996 | -0.97867 | -0.20542 |
| 8 | 1.00000 | -0.43589 | 0.90039 | -0.61745 | -0.78661 | 0.97581 | -0.21860 | -0.23752 | 0.97138 | -0.77764 | -0.62871 | 0.91076 | -0.41294 | -0.00937 | 0.99996 | -0.89627 | -0.44350 | 0.78928 | -0.61403 |
| 9 | 1.00000 | -0.64632 | 0.76306 | -0.15896 | -0.98729 | 0.84901 | 0.51196 | -0.94947 | 0.31387 | 0.36364 | -0.93154 | 0.46080 | 0.88751 | -0.97867 | -0.20542 | 0.78928 | -0.61403 | -0.04160 | 0.99913 |

one or more of the equations are likely to be in error. If $e_k$ is the error in the $k^{th}$ equation of the system, one can represent the system in matrix form as,

$$\mathbf{A}\,\mathbf{X} = \mathbf{Z} + \mathbf{E} \tag{8}$$

where **A**, **X** and **Z** are defined as before and **E** is a (2 M) column vector with the error terms.

Now applying the standard procedure of least squares method to minimise this error, one can obtain:

$$\mathbf{X} = (\mathbf{A}'\mathbf{A})^{-1}\,\mathbf{A}'\mathbf{Z} \tag{9}$$

Once the coefficients $a_o$, $a_n$ and $b_n$ are obtained from (9) one can substitute back in (6) with $r = 0, 1, \ldots, (j-1)$ to estimate the j values in the gap $(t_o - t_{j-1})$ as:

$$\zeta^*_{t_r} = a_0 + \sum_{n=1}^{N} (a_n \cos \sigma_n t_r + b_n \sin \sigma_n t_r), \text{ with } r = 0, 1, \ldots\ldots\ldots, (j-1) \tag{10}$$

On testing this method on artificially created gaps in the harmonically predicted time series at Southend (with tidal components $M_1$, $M_2$, $M_4$ and $M_6$ and $M = 25$), it is found to be reliable for estimating gaps of up to twelve hourly values (which includes a tidal turning point) with a mean error of about 8–10 cm, which is around 2 % of the complete range of the tide at Southend.

However when the method was tested on observed data, at the same location and associated time period, the mean error increased sharply up to about 40 cm. This sharp increase in the inaccuracy was later found to be due to a high storm surge activity during that particular period. When tested on time series from the same, and different locations, with very little or no significant storm surge activity, the mean error decreased once again to around 2 %–5 % of the respective tidal range.

## CONCLUSIONS

Although the conventional Lagrangian interpolation methods have, up to now served the purpose of automatically checking the smoothness of hourly tidal time series, the superiority of the new harmonic component fit method (HCFM) in producing better estimates is self evident. In conventional Lagrangian methods, the data were treated mainly as numerical quantities with little or no information input regarding the characteristics of the tide itself. The success of the new method may be largely due to the additional information (regarding the basic character of the tide itself) used, namely treating the time series as a Fourier series with a fundamental frequency component and its higher harmonics.

In practice, when this new method is used for smoothing tidal data, care should be taken in choosing the components to be used, to obtain the maximum accuracy required, bearing in mind the need to keep the system stable. When used to fill gaps in the data series, the data should be free of any substantial meteorological disturbances if results with a maximum accuracy are to be expected.

The method is easy to program and economical to run on a standard electronic digital computer.

## ACKNOWLEDGEMENTS

The work was funded by a Consortium consisting of the Natural Environment Research Council, the Ministry of Agriculture, Fisheries and Food, and the Departments of Energy and Industry. I am grateful to Mr. M. AMIN and Mr. J. GRAFF for their useful suggestions and remarks concerning the work presented in this paper.

## REFERENCES

AMIN, M. (1976) : The fine resolution of tidal harmonics. *Geophysical Journal of the Royal Astronomical Society*, Vol. 44, No. 2.

CARTWRIGHT, D.E. (1968) : A unified analysis of the tides and surges round North and East Britain. *Philosophical Transactions of the Royal Society*, A, 263, pp. 1-55.

GREIG, M.A. (1977) : Digitising of Torres Strait tidal records. CSIRO Division of Fisheries and Oceanography, Australia. Report No. 74.

LENNON, G.W. (1964) : Some computer techniques for the analytical treatment of tidal data. Communication de l'Observatoire royal de Belgique, 236.

LENNON, G.W. (1965) : The treatment of hourly elevations of the tide using an IBM 1620. *Intern. Hydrog. Review*, Vol. XLII (2), pp. 125-148.