

# **Towards a Real-time Tidal Analysis and Prediction**

**Hou Tianhang and Petr Vanicek**  
**Department of Surveying Engineering**  
**University of New Brunswick**  
**P.O.Box 4400, Fredericton, N.B.**  
**CANADA**

## **ABSTRACT**

In the practice of tidal analysis and prediction, the number and kind of astronomical tidal components that are to be included in a tidal model depend on the length of available tidal record and the desired accuracy of prediction. Since tidal frequencies, including shallow water constituents, are distributed unequally in a few narrow frequency bands, an inappropriate selection of tidal constituents to be included in the analysis and prediction may cause the normal equations to become ill-conditioned, unstable or even singular, and the prediction to become poor. This investigation shows how to construct lumped tidal frequencies which better characterise ocean tides with diminishing length of observational series.

Further, a sequential tidal analysis model is proposed and an algorithm for its implementation is presented, which can rigorously update a tidal solution when the number of observations increases. The algorithm also bring in automatically additional tidal constituents without a large amount of computation work; the CPU time for this analysis is only about 4 percent of that for the conventional harmonic technique. The sequential algorithm for ocean tidal analysis and prediction has a potential to be used in tide gauge stations for providing continuous up-to-date tidal prediction.

## **Introduction**

Drive towards an increasingly more accurate predictive capability of sea level elevation in coastal zones has been spurred on by concerns relating to navigation, global warming, shore-line engineering and pollutant transport. Traditionally, site specific sea level tidal information is derived from the harmonic analysis of a collected time series which estimates the amplitude and phases of some selected harmonic tidal constituents. Then, these harmonic tidal constituents, which can be either of an astronomic origin, or of shallow water variety [Zetler and Robert, 1967], are used to generate tidal prediction for future dates.

In tidal predictions, the accuracy of predicted values  $\hat{y}(t)$  depends not only on the number of tidal components used in the computations, but also on the accuracy of their estimated amplitudes and phases. Assuming that we order

the components according to their amplitudes, the more tidal components are included in an analysis, the higher the accuracy that can be achieved in tidal predictions. On the other hand, the accuracies of estimated amplitudes and phases of these tidal constituents, of which we would want to select as many as possible, are closely related to the length of the time series used in the estimation with the least squares method. If too many constituents are chosen for the analysis, in other words, if the time period over which observations are taken is too short, then either no solution would ensue, or an unstable solution would be obtained, in which the interference between and among tidal components with similar frequencies would be a detrimental factor. This happens when two or more frequencies are too close together so that they cannot be resolved from the given length of the time series.

As a rule of thumb, two tidal constituents of frequencies  $f_j$  and  $f_k$  can be separated, if their frequencies satisfy the relation [Godin, 1972]:

$$N ( f_j - f_k ) \geq 1, \quad j \neq k . \quad (1)$$

This rule is called the Rayleigh criterion, where,  $N$  represents the number of steps - typically hours - in a continuous sequence of (hourly) observations.

From the above description, it is apparent that the number of tidal components that could be included in a tidal analysis really depends on the length of the available tide gauge record. Since tidal records of 'sufficient' length are not available at all tide gauge stations, it is usually impossible to obtain as many tidal constituents as we would wish to have to make a 'sufficiently' good prediction. The problem then is that when the length of the collected tidal series is short, which and how many tidal constituents should be included in the tidal analysis to give us the best predicted results. Before answering these questions, we should realise that since the tidal frequencies are distributed unequally in a few narrow bands, the selection of the tidal constituents to be included in the analysis and prediction is not a simple matter. Arbitrary selection of tidal constituents may cause either large departures of the prediction from reality, or cause the normal equations to become ill-conditioned (unstable), or even singular. Therefore, the selection of standard tidal analysis and prediction tables of tidal constituents which would fit different lengths of observational series is a critically important step particularly for short tidal series.

As we know, by increasing the time period over which sea-level data are collected, more and more tidal constituents can be separated and thus included in the analysis. Then, the accuracy of tidal prediction will gradually increase. But, with adding new data and new tidal constituents, the normal equations of the harmonic analysis model have to be inverted repeatedly to update all the estimates because no use is made, in the standard approach, of the estimates calculated in the previous stages. Usually, the matrix to be repeatedly inverted is quite large, so updating the estimates takes a good deal of CPU time.

The purpose of this work is to seek a method, that, while adopting the most detailed tidal model possible, would update harmonic results with a minimum computational effort

## Conventional Harmonic Analysis

Let us, for the moment, forget the shallow water constituents and the nontidal effects in the oceans and consider the tide to be composed of astronomical tidal constituents only:

$$\begin{aligned}
 y(t_i) &= Z_0 + \sum_{j=1}^m H_j \cos(\omega_j t + \phi_j + \phi_j) = \\
 &= Z_0 + \sum_{j=1}^m \{H_j \cos \phi_j \cos(\omega_j t + \phi_j) - H_j \sin \phi_j \sin(\omega_j t + \phi_j)\} \\
 &= Z_0 + \sum_{j=1}^m \{C_j \cos(\omega_j t + \phi_j) + S_j \sin(\omega_j t + \phi_j)\}, \quad (2)
 \end{aligned}$$

where

$$C_j = H_j \cos \phi_j \quad S_j = -H_j \sin \phi_j,$$

$t \in \{t_1, t_2, \dots, t_N\}$ ,  $\omega t + \phi = k_1 \tau + k_2 s + k_3 h + k_4 p + k_5 N + k_6 Ps$ , and  $t$  is Greenwich mean time,  $H_j, \phi_j$  are the amplitudes and phase lags. The following relations hold:

$$\forall j=1, m: \quad H_j = \sqrt{C_j^2 + S_j^2}, \quad \phi_j = \arctan\left(-\frac{S_j}{C_j}\right). \quad (3)$$

For the theoretical tide, the arguments  $\omega_j t + \phi_j$  for individual tidal constituents are known from Doodson's harmonic development. For the actual tide we have to estimate the unknown parameters  $Z_0, H_j, \phi_j$  ( $j=1, 2, \dots, m$ ) from the series of measured values  $y(t_i)$  ( $i=1, 2, \dots, N$ ) by the least squares method.

Once the parameters have been determined, the values  $\hat{y}(t_i)$  can be obtained from the above model; this is the tidal "prediction".

Matrix notation can now be used to rewrite the eqn. (2) as:

$$\mathbf{Y} = \mathbf{A}\mathbf{X}, \quad (4)$$

where  $\mathbf{Y} = (y(t_1), y(t_2), \dots, y(t_N))^T$  is the "observation vector",  $\mathbf{X} = (Z_0, C_1, S_1, \dots, C_m, S_m)^T$  is the "solution vector", and the matrix

$$\mathbf{A} = \begin{bmatrix} 1 & \cos(\omega_1 t_1 + \phi_1) & \dots & \sin(\omega_m t_1 + \phi_m) \\ 1 & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ 1 & \cos(\omega_1 t_N + \phi_1) & \dots & \sin(\omega_m t_N + \phi_m) \end{bmatrix} \quad (5)$$

is called the Vandermonde (or design) matrix. The least squares solution of the above system of over determined equations (for  $N > 2m+1$ ) is given by the following normal equations:

$$\hat{\mathbf{X}} = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{Y}. \quad (6)$$

It is known from the theory of the least squares method [Vaníček and Krakiwsky, 1986] that the accuracy of the solution vector  $\hat{\mathbf{X}}$  is estimated by the covariance matrix:

$$\hat{\mathbf{C}}_{\hat{\mathbf{X}}} = \frac{(\mathbf{Y} - \mathbf{A}\hat{\mathbf{X}})^T (\mathbf{Y} - \mathbf{A}\hat{\mathbf{X}}) (\mathbf{A}^T \mathbf{A})^{-1}}{N - (2m+1)} \quad (7)$$

The diagonal elements of  $\hat{\mathbf{C}}_{\hat{\mathbf{X}}}$  represent the variances  $\sigma_{\hat{x}_i}^2$  ( $i=1, 2, \dots, 2m+1$ ) of the resulting parameters and off-diagonal elements are the covariances  $\sigma_{\hat{x}_i \hat{x}_j}$  ( $i, j=1, 2, \dots, 2m+1$ ) between pairs of parameters.

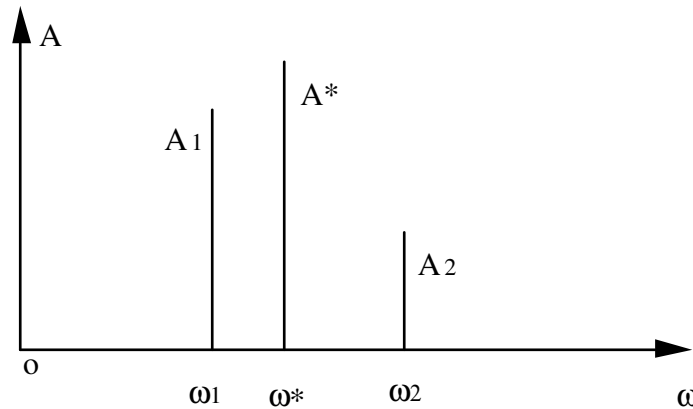
This method is widely used now in tidal operations because it is simple, yet it gives good enough predictions for most purposes.

### Lumped Tidal Constituents

As we have already stated, the number of constituents used in a tidal model depends strongly on the length of the observational series. If the length of observed series is more than 18.6 years, nearly all the harmonic tidal constituents of Doodson's development as well as all shallow water constituents can be separated and thus included in the model. Under this circumstance, it perhaps makes no sense to introduce the sequential approach. But if the duration of observed series is much shorter, sometimes only a few months, weeks or even days, we wish to know which principal tidal constituents should be selected in the model to obtain the best predicted results. In other words, the question arises as to which compositions of principal tidal constituents gives an optimal representation over whole tidal frequency spectrum. Making such a selection in tidal analysis is very difficult. No such composition reflects the tidal energy distribution over all tidal frequency bands rigorously.

For constructing the sequential tidal model, we create lumped tidal constituents from individual astronomical and shallow water constituents by the least squares method. This smaller number of lumped tidal constituents, can be considered to be the best representation of tidal frequency bands. To demonstrate how to form these lumped tidal constituents, let us consider two astronomical tidal constituents with adjacent frequencies  $w_1$  and  $w_2$ ; lumped together the result must have a representative frequency  $w^*$  located

somewhere between  $w_1$  and  $w_2$  and a combined amplitude  $A^*$  related to the amplitudes  $A_1$  and  $A_2$  of the original constituents (see Figure 1).



**Figure 1 A Representative Frequency  $w^*$  for  $w_1$  and  $w_2$**

In our algorithm, for any two tidal constituents to be lumped together, the weighted average  $W^*$  of the frequencies  $W_1$ ,  $W_2$  has been used for the frequency of their lumped representative:

$$\omega^* = \omega_1 + \frac{\omega_2 - \omega_1}{A_1 + A_2} A_2, \quad (8)$$

where,  $A_1$ ,  $A_2$  are the respective theoretical amplitudes of the two constituents. The lumped amplitude  $A^*$  is defined by:

$$A^* = \sqrt{A_1^2 + A_2^2}. \quad (9)$$

These parameters  $w^*$  and  $A^*$  are then used in the next lumping step as the representative values for the original constituents.

### **Method of Construction of Lumped Constituent Tables**

The tidal potential contains about 400 constituents in Doodson's [Doodson, ] harmonic development, about 500 terms in Cartwright's [Cartwright,.....] spectral analysis results and about 1140 terms in Qiwen's [Qiwen, 1987] logical deduction method for precision tidal analysis.

In the construction of our lumped tidal tables, we have selected only 60 principal harmonic constituents whose theoretical relative amplitudes are larger than  $500 \times 10^{-5}$  (neglecting the fact that actual amplitudes may be significantly different, altered by tidal resonance and, of course, by the latitude effect). Since the tidal energy is proportional to the squares of the amplitudes, the 60 main

constituents represent some 99.97% percent of total tidal energy. Beside the 60 principal tidal constituents, an additional 15 shallow water constituents have been also considered in our construction. Amplitudes of shallow water constituents change dramatically from place to place. In our computations, however, for a lack of any better information, we have assumed all their amplitudes, to equal to  $600 \times 10^{-5}$

When building up the design matrix, we re-wrote the observation equation (2) in such a way that the original phase  $\varphi_j$  and the phase lag  $\phi_j$  are added together. Then they become part of the unknown parameters  $C_j^*$  and  $S_j^*$  as follows:

$$\begin{aligned}
 \forall i = 1, N: \quad y(t_i) &= \sum_{j=1}^m H_j \cos(\omega_j t + \varphi_j + \phi_j) = \\
 &= Z_0 + \sum_{j=1}^m [H_j \cos(\varphi_j + \phi_j) \cos \omega_j t - H_j \sin(\varphi_j + \phi_j) \sin \omega_j t] = \\
 &= Z_0 + \sum_{j=1}^m (C_j^* \cos \omega_j t + S_j^* \sin \omega_j t), \quad (10)
 \end{aligned}$$

where:

$$\forall j = 1, m: \quad C_j^* = H_j \cos(\varphi_j + \phi_j); \quad S_j^* = -H_j \sin(\varphi_j + \phi_j). \quad (11)$$

Thus, the elements of the design matrix are expressed as functions of angular velocities  $\omega_j$  and time  $t$ , regardless of the time origin of the analyzed series. It should be kept in mind, however, that the vector of unknown parameters  $C^*$  and  $S^*$  will change according to the choice of origin of time (usually Julian) used in forming the design matrix.

When the matrix of the normal equations  $\mathbf{N} = \mathbf{A}^T \mathbf{A}$  is created, its elements have one of the following forms:

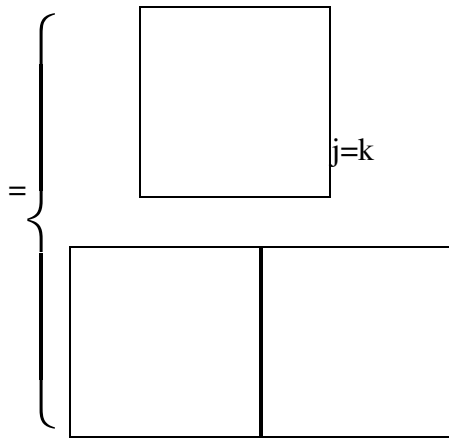
$$\sum_{i=1}^N \cos \omega_j t_i \cos \omega_k t_i, \quad \sum_{i=1}^N \cos \omega_j t_i \sin \omega_k t_i, \quad \sum_{i=1}^N \sin \omega_j t_i \sin \omega_k t_i \quad (12)$$

(with the exception of the first row and first column, where the elements are as follows:

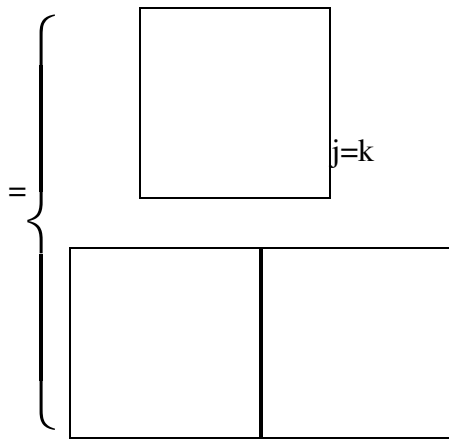
$$\left( \sum_{i=1}^N \cos \omega_j t_i, \quad \sum_{i=1}^N \sin \omega_j t_i \right). \quad (13)$$

These elements can be evaluated much faster from the following expressions [Bronshtein and Semendyayev, 1979]:

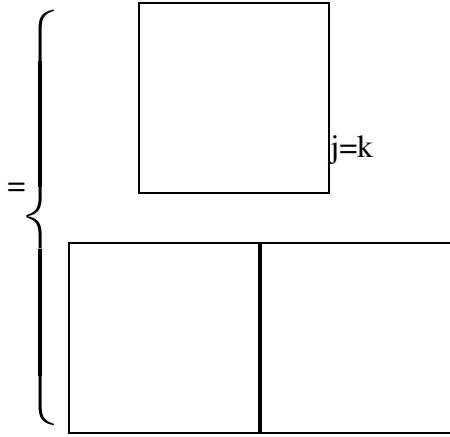
$$\sum_{i=1}^N \cos \omega_j t_i \cos \omega_k t_i = \tag{14}$$



$$\sum_{i=1}^N \cos \omega_j t_i \sin \omega_k t_i = \tag{15}$$



$$\sum_{i=1}^N \sin \omega_j t_i \sin \omega_k t_i = \tag{16}$$



$$\sum_{i=1}^N \cos \omega_j t_i = \frac{\sin [0.5N(\omega_j)] \cos [0.5(N+1)\omega_j]}{\sin \frac{\omega_j}{2}} \quad (17)$$

$$\sum_{i=1}^N \sin \omega_j t_i = \frac{\sin [0.5N(\omega_j)] \sin [0.5(N+1)\omega_j]}{\sin \frac{\omega_j}{2}} \quad (18)$$

Clearly, using the above expressions, the CPU time needed for constructing the matrix of normal equations is independent of the observation series length N. This is very useful in sequential tidal analysis, especially with a very long tidal series.

### **Correlation Criterion for Separability of Constituents**

From the matrix of normal equations, we can estimate approximately which two pairs of columns are likely to interfere with each other, i.e., which pair of tidal constituents is likely to be highly correlated. It is impossible, however, to determine the definite correlation values between any two adjacent constituents. For this purpose, it is necessary to invert the matrix and get the covariance matrix of the estimated coefficients (Eqn. 7), which we will rewrite here as:

$$\widehat{\mathbf{C}}_{\hat{\mathbf{x}}} = \widehat{\sigma}_0^2 (\mathbf{A}^T \mathbf{A})^{-1} = \widehat{\sigma}_0^2 \begin{bmatrix} Q_{11} & \cdot & \cdot & Q_{1L} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ Q_{L1} & \cdot & \cdot & Q_{LL} \end{bmatrix} = \begin{bmatrix} \sigma_1^2 & \cdot & \cdot & \sigma_{1L} \\ \cdot & \sigma_2^2 & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \sigma_{L1} & \cdot & \cdot & \sigma_L^2 \end{bmatrix}$$



(19)

where,  $\widehat{\sigma}_0^2 = (\mathbf{Y}-\mathbf{AX})^T(\mathbf{Y}-\mathbf{AX})/(N-L)$  and  $L=2m+1$ .

The  $r_{ij}$  correlation coefficient of any two estimated coefficients  $X_i, X_j$  (note that here X's stand for the unknowns C's and S's) is given by [Vaníček and Wells, 1972]:

$$\rho_{ij} = \frac{\sigma_{ij}}{\sigma_i \sigma_j}. \quad (20)$$

Clearly, we can also write:

$$\rho_{ij} = \frac{Q_{ij}}{\sqrt{Q_{ii}} \sqrt{Q_{jj}}} = \rho_{ji} \quad (21)$$

and to evaluate the correlation coefficient it is not necessary to know  $\widehat{\sigma}_0^2$ . The following correlation matrix:

$$\mathbf{R} = \begin{bmatrix} 1 & \rho_{12} & \cdot & \rho_{1L} \\ \cdot & 1 & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \rho_{L1} & \cdot & \cdot & 1 \end{bmatrix} \quad (22)$$

can be then calculated without having any observational series. This matrix  $\mathbf{R}$  depends only on the assumed length  $N$  of the time series, and the selected number  $m$  of constituents. Looking at the off-diagonal elements in the matrix  $\mathbf{R}$ , we can make a decision about which two (or more) tidal components are strongly correlated; these two (or more) can then be held responsible for potential ill-conditioning of the normal equations.

To determine which two adjacent tidal constituents are significantly correlated, we first take the 4 by 4 symmetrical submatrix of  $\mathbf{R}$  connected with these two constituents. This submatrix will look as follows:

$$\begin{bmatrix} 1 & \rho_{C_i S_i} & \rho_{C_i C_{i+1}} & \rho_{C_i S_{i+1}} \\ \rho_{S_i C_i} & 1 & \rho_{S_i C_{i+1}} & \rho_{S_i S_{i+1}} \\ \rho_{C_{i+1} C_i} & \rho_{C_{i+1} S_i} & 1 & \rho_{C_{i+1} S_{i+1}} \\ \rho_{S_{i+1} C_i} & \rho_{S_{i+1} C_i} & \rho_{S_{i+1} C_{i+1}} & 1 \end{bmatrix}. \quad (23)$$

From numerical experiments we have established that

$$\rho_{C_i C_{i+1}} \approx \rho_{S_i S_{i+1}}, \quad \rho_{C_i S_{i+1}} \approx \rho_{S_i C_{i+1}} \quad (24)$$

From further numerical experiments we have determined that the quantity

$$\rho_{i, i+1} = \frac{1}{2} \sqrt{(|\rho_{C_i C_{i+1}}| + |\rho_{S_i S_{i+1}}|)^2 + (|\rho_{C_i S_{i+1}}| + |\rho_{S_i C_{i+1}}|)^2} \in \langle 0, 1 \rangle \quad (25)$$

decreases monotonically with increasing  $N$ , and is thus suitable to use as a correlation criterion. Somewhat arbitrarily, we have selected a value of 0.985 to serve as a criterion value for deciding if two adjacent constituents are correlated, i.e., if they are separable or not. For  $\rho_{i, i+1} \leq 0.985$  all the correlation coefficients in the matrix (23) are smaller in absolute value than 0.95.

In Tables 1 to 3, the frequencies and amplitudes of lumped

TABLES 1 TO 3 TO FIT NEAR HERE

representative constituents, obtained as described above, are given in the individual boxes. These Tables show how the lumping works for diminishing length of observational series, when the series is shortened in successive steps. Looked at from the other perspective, i.e., considering the series as growing in length, these tables show the separability of constituents. Based on these Tables, detailed time schedule for the tidal constituent separability has been designed and included in our computer program for sequential tidal analysis .

### The Sequential Mathematical Model

In applying the sequential technique to tidal analysis, first of all the first estimates of unknown parameters are obtained from an initial tidal harmonic analysis. In the next stage these initial estimates are updated by computing corrections to the earlier results as functions of previous estimates. The information made available for subsequent use consists of the estimated amplitudes, phase lags and their covariance matrix, that serves as the link between subsequent steps

In our sequential tidal model algorithm we distinguish between two different update modes:

- (i) adding only new observations;
- (ii) adding both new observations and new tidal constituents;

Throughout the development of the algorithm (as well as the program based on the algorithm), we restrict ourselves to a rigorous approach, but we will include some discussion concerning approximate approaches for certain situations.

The original mathematical model for tidal harmonic analysis is (cf eqn. 4):

$$\mathbf{A}_1 \mathbf{X}_1 = \mathbf{L}_1, \quad (26)$$

where  $\mathbf{L}_1=(L_1, L_2, \dots, L_N)^T$  is the data vector,  $\mathbf{A}_1$  is the design matrix, and  $\mathbf{X}_1=(X_1, X_2, \dots, X_L)^T$  is the unknown parameter vector. The least squares solution of the above system of over-determined linear equations is given by normal equations as [Vaníček and Krakiwsky, 1986]:

$$\mathbf{X}_1 = \mathbf{N}^{-1} \mathbf{A}_1^T \mathbf{L}_1 = (\mathbf{A}_1^T \mathbf{A}_1)^{-1} \mathbf{A}_1^T \mathbf{L}_1. \quad (27)$$

The sequential updating starts with the acquisition of additional data. Let us assume that a batch of data, consisting of 1 to  $N_1$  new values becomes available. These additional data contain additional information on the analysed tide and now have to be included in the analysis. We note that the size  $N_1$  of the batch (it may be as small as 1!) should be selected beforehand according to what use the results of the analysis are going to be put to.

When adding the new batch of data to the existing series, two or more tidal constituents may become separable. If this is the case, then the separable constituent present in the previous analysis - it could have been a lumped constituent, of course - is replaced by its separate component. This case is referred to as ii) above. We shall first discuss the more simple scenario i), when no new tidal constituents appear in the sequential step.

### **Addition of New Observations**

If only new observations  $\mathbf{L}_2=(L_{N+1}, L_{N+2}, \dots, L_{N+N_1})^T$  are added, the observation equations become:

$$\begin{bmatrix} \mathbf{A}_1 \\ \mathbf{A}_2 \end{bmatrix} (\mathbf{X}_1^{(1)} + d\mathbf{X}_1^{(2)}) = \begin{bmatrix} \mathbf{L}_1 \\ \mathbf{L}_2 \end{bmatrix}, \quad (28)$$

and the new solution is given as:

$$\mathbf{X}_1^{(1)} + d\mathbf{X}_1^{(2)} = \mathbf{N}_{11}^{-1} (\mathbf{A}_1^T \mathbf{L}_1 + \mathbf{A}_2^T \mathbf{L}_2). \quad (29)$$

Here, the matrix  $\mathbf{N}_{11}$  of normal equations is as follows:

$$\mathbf{N}_{11} = \begin{bmatrix} \mathbf{A}_1 \\ \mathbf{A}_2 \end{bmatrix}^T \begin{bmatrix} \mathbf{A}_1 \\ \mathbf{A}_2 \end{bmatrix} = (\mathbf{A}_1^T \mathbf{A}_1 + \mathbf{A}_2^T \mathbf{A}_2) = (\mathbf{N} + \mathbf{DN}) . \quad (30)$$

The inversion of this matrix can be obtained from the following rigorous sequential expression [Morrison, 1969]:

$$\mathbf{N}_{11}^{-1} = (\mathbf{N} + \mathbf{DN})^{-1} = \mathbf{N}^{-1} - \mathbf{N}^{-1} \mathbf{A}_2^T [\mathbf{I} + \mathbf{A}_2 \mathbf{N}^{-1} \mathbf{A}_2^T]^{-1} \mathbf{A}_2 \mathbf{N}^{-1}, \quad (31)$$

Here,  $\mathbf{I}$  is the identity matrix, and

$$\mathbf{DN} = \mathbf{A}_2^T \mathbf{A}_2$$

can be considered the perturbation of the set of the original normal equations due to the added observations. From expression (31), we see that the matrix  $\mathbf{I} + \mathbf{A}_2 \mathbf{N}^{-1} \mathbf{A}_2^T$  has a dimension of  $N_1$  by  $N_1$ , where  $N_1$  is the number of added observations.

If  $N_1$  is large, we still need to invert a large matrix which would eliminate one of the important advantages of the sequential approach. In practice, the number of added observations  $N_1$  should be small. If we let, for example,  $N_1=1$ , meaning that only one new observation is added at a time, the matrix degenerates into a scalar:

$$(\mathbf{I} + \mathbf{A}_2 \mathbf{N}^{-1} \mathbf{A}_2^T) = Q \quad (32)$$

Eqn. (31) is then written as:

$$\mathbf{N}_{11}^{-1} = (\mathbf{N} + \mathbf{DN})^{-1} = \mathbf{N}^{-1} - \frac{1}{Q} \mathbf{N}^{-1} \mathbf{A}_2^T \mathbf{A}_2 \mathbf{N}^{-1} = \mathbf{N}^{-1} - \frac{1}{Q} \mathbf{N}^{-1} \mathbf{DN} \mathbf{N}^{-1}. \quad (33)$$

Obviously, when adding a single new observation at a time, no additional matrix needs be inverted.

Before the rigorous complete sequential solution is given, let us introduce another important approximate formula for matrix inversions which may be useful in some cases where the original matrix is huge, and the number  $N_1$  of observations added is so large that the rigorous inversion would be too time consuming. The approximate expression reads [reference]

$$\mathbf{N}_{11} = (\mathbf{N} + \mathbf{DN})^{-1} \doteq \mathbf{N}^{-1} - \mathbf{N}^{-1} \mathbf{DN} \mathbf{N}^{-1}. \quad (34)$$

The criterion for this expression to satisfy the required accuracy of tidal harmonic analysis reads:

$$\|\mathbf{DN}\| \ll \|\mathbf{N}\|, \quad (35)$$

where  $\| \cdot \|$  denotes a norm [reference]. There are several ways to compute a matrix norm, and we adopt the most commonly used quadratic norm. The formulation is given as:

$$\|\mathbf{T}\| = \{ \sum |T_{ij}|^2 \}^{\frac{1}{2}}, \quad (36)$$

where  $T_{ij}$  are the elements of matrix  $\mathbf{T}$ .

To conclude: if no new tidal constituents are added, the new (sequential) solution is given by:

$$\mathbf{X}_1^{(2)} = \mathbf{X}_1^{(1)} + d\mathbf{X}_1^{(2)}, \quad (37)$$

$$d\mathbf{X}_1^{(2)} = \mathbf{N}^{-1} \mathbf{A}_2^T \mathbf{L}_2 - \mathbf{F} (\mathbf{A}_1^T \mathbf{L}_1 + \mathbf{A}_2^T \mathbf{L}_2), \quad (38)$$

where the matrix  $\mathbf{F}$  is given as:

$$\mathbf{F} = \begin{cases} -\mathbf{N}^{-1} \mathbf{A}_2^T (\mathbf{I} + \mathbf{A}_2 \mathbf{N}^{-1} \mathbf{A}_2^T)^{-1} \mathbf{A}_2 \mathbf{N}^{-1} & \text{all} \\ -\frac{1}{Q} \mathbf{N}^{-1} \mathbf{D} \mathbf{N} \mathbf{N}^{-1} & \mathbf{N}_1=1 \\ -\mathbf{N}^{-1} \mathbf{D} \mathbf{N} \mathbf{N}^{-1} & \|\mathbf{D} \mathbf{N}\| \ll \|\mathbf{N}\|. \end{cases} \quad (39)$$

or

give here the expression for the approximate inversion!

### Addition of New Tidal Constituents

In the case, when not only  $n_2$  new observations  $L_i$  are added, but also  $m_2$  lumped tidal constituents become separable, the situation is somewhat more complicated. Clearly,  $2m_2$  already estimated parameters become redundant, while  $4m_2$  new parameters have to be estimated. Thus, in addition to the new vector  $\mathbf{L}_2$  of observations (of dimension  $n_2$ ) to be added to the previous vector  $\mathbf{L}_1$  (of dimension  $n_1$ ), we have to consider also a new vector  $\mathbf{X}_2$  of unknown parameters (of dimension  $4m_2$ ) that has to be added to the previous vector  $\mathbf{X}_1$  (of dimension  $2m_1 + 1$ ) from which  $2m_2$  elements are first discarded. In the sequel, we shall explain only the necessary manipulations with pertinent matrices without keeping track of the proper dimensions. **(Mr. Hou - the preceding paragraph should replace the first 3 lines in your**

version of this chapter. Please note the notation used for the dimensions of all the vectors. You should change the text in this and other chapters to conform with this notation; what you have been using does not make much sense!

The observation equations can be written here as:

$$\begin{bmatrix} \mathbf{A}_1 & \mathbf{A}_4 \\ \mathbf{A}_2 & \mathbf{A}_3 \end{bmatrix} \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{L}_1 \\ \mathbf{L}_2 \end{bmatrix}, \quad (40)$$

where the new design matrices  $\mathbf{A}_2$ ,  $\mathbf{A}_3$  and  $\mathbf{A}_4$  are added to the original design matrix  $\mathbf{A}_1$ . The new (updated) matrix of normal equations reads:

$$\mathbf{N}^{(2)} = \begin{bmatrix} \mathbf{A}_1 & \mathbf{A}_4 \\ \mathbf{A}_2 & \mathbf{A}_3 \end{bmatrix}^T \begin{bmatrix} \mathbf{A}_1 & \mathbf{A}_4 \\ \mathbf{A}_2 & \mathbf{A}_3 \end{bmatrix} = \begin{bmatrix} \mathbf{A}_1^T \mathbf{A}_1 + \mathbf{A}_2^T \mathbf{A}_2 & \mathbf{A}_1^T \mathbf{A}_4 + \mathbf{A}_2^T \mathbf{A}_3 \\ \mathbf{A}_4^T \mathbf{A}_1 + \mathbf{A}_3^T \mathbf{A}_2 & \mathbf{A}_3^T \mathbf{A}_3 + \mathbf{A}_4^T \mathbf{A}_4 \end{bmatrix}, \quad (41)$$

where the submatrices are denoted as follows:

$$\mathbf{N}_{11} = \mathbf{N} + \mathbf{D}\mathbf{N} = \mathbf{A}_1^T \mathbf{A}_1 + \mathbf{A}_2^T \mathbf{A}_2,$$

$$\mathbf{N}_{12} = \mathbf{N}_{21}^T = \mathbf{A}_1^T \mathbf{A}_4 + \mathbf{A}_2^T \mathbf{A}_3,$$

$$\mathbf{N}_{22} = \mathbf{A}_3^T \mathbf{A}_3 + \mathbf{A}_4^T \mathbf{A}_4.$$

Partitioning the matrix inversion, we then get:

$$\begin{aligned} (\mathbf{N}^{(2)})^{-1} &= \begin{bmatrix} \mathbf{N}_{11} & \mathbf{N}_{12} \\ \mathbf{N}_{21} & \mathbf{N}_{22} \end{bmatrix}^{-1} \\ &= \begin{bmatrix} \mathbf{N}_{11}^{-1} + \mathbf{N}_{11}^{-1} \mathbf{N}_{12} \mathbf{D}^{-1} \mathbf{N}_{21} \mathbf{N}_{11}^{-1} & -\mathbf{N}_{11}^{-1} \mathbf{N}_{12} \mathbf{D}^{-1} \\ -\mathbf{D}^{-1} \mathbf{N}_{21} \mathbf{N}_{11}^{-1} & \mathbf{D}^{-1} \end{bmatrix}, \end{aligned} \quad (42)$$

where,

$$\mathbf{D}^{-1} = (\mathbf{N}_{22} - \mathbf{N}_{21} \mathbf{N}_{11}^{-1} \mathbf{N}_{12})^{-1}. \quad (43)$$

It is seen that the additional matrix inversion for sequential updating is only of the size of the number of added parameters. The matrix inversion  $(\mathbf{I} + \mathbf{A}_2^T \mathbf{N}^{-1} \mathbf{A}_2)$

${}^1\mathbf{A}_2^T)^{-1}$  pertaining to the new observations appears as well, as one would naturally expect.

The rigorous solution to the combined (old and new) normal equations for all the unknown parameters (old ones without the parameters pertaining to the constituent(s) being separated and the new ones) can be written as:

$$\begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{bmatrix}^{(2)} = \begin{bmatrix} \mathbf{X}_1^{(1)} + d\mathbf{X}_1^{(2)} \\ \mathbf{X}_2^{(2)} \end{bmatrix} = (\mathbf{N}^{(2)})^{-1} \begin{bmatrix} \mathbf{A}_1^T \mathbf{L}_1 + \mathbf{A}_2^T \mathbf{L}_2 \\ \mathbf{A}_4^T \mathbf{L}_1 + \mathbf{A}_3^T \mathbf{L}_2 \end{bmatrix}. \quad (44)$$

Spelling out these results, we have:

$$\begin{aligned} \mathbf{X}_1^{(1)} &= \mathbf{N}^{-1} \mathbf{A}_1^T \mathbf{L}_1, \\ \mathbf{X}_1^{(2)} &= \mathbf{X}_1^{(1)} + d\mathbf{X}_1^{(2)}, \\ d\mathbf{X}_1^{(2)} &= \mathbf{Z}_1 \mathbf{L}_1 + \mathbf{Z}_2 \mathbf{L}_2, \end{aligned} \quad (45)$$

(note that the dimension of  $d\mathbf{X}_1^{(2)}$  is smaller than the dimension of  $\mathbf{X}_1^{(1)}$  by twice the number of constituents that became separable in this sequential step)

$$\mathbf{X}_2^{(2)} = \mathbf{Z}_3 \mathbf{L}_1 + \mathbf{Z}_4 \mathbf{L}_2,$$

where, using eqn.s (39) and (42), we can write:

$$\begin{aligned} \mathbf{Z}_1 &= -\mathbf{N}_{11}^{-1} \mathbf{N}_{12} \mathbf{D}^{-1} (\mathbf{N}_{21} \mathbf{N}_{11}^{-1} \mathbf{A}_1^T - \mathbf{A}_4^T) - \mathbf{F} \mathbf{A}_1^T, \\ \mathbf{Z}_2 &= \mathbf{N}_{11}^{-1} + \mathbf{N}_{11}^{-1} \mathbf{N}_{12} \mathbf{D}^{-1} (\mathbf{N}_{21} \mathbf{N}_{11}^{-1} \mathbf{A}_2^T - \mathbf{A}_3^T), \\ \mathbf{Z}_3 &= \mathbf{D}^{-1} (\mathbf{A}_4^T - \mathbf{N}_{21} \mathbf{N}_{11}^{-1} \mathbf{A}_1^T), \\ \mathbf{Z}_4 &= \mathbf{D}^{-1} (\mathbf{A}_3^T - \mathbf{N}_{21} \mathbf{N}_{11}^{-1} \mathbf{A}_2^T). \end{aligned} \quad (46)$$

In the above mathematical model, the entire covariance matrix  $(\mathbf{N}^{(2)})^{-1}$  in the current step must be available as  $\mathbf{N}^{-1}$  in next step, to obtain again the rigorous sequential solution.

The sequence of iterative solutions for unknown parameters of tidal harmonic analysis looks as follows:

$$\left\{ \begin{array}{l} \mathbf{X}_1^{(2)} = \mathbf{X}_1^{(1)} + d\mathbf{X}_1^{(2)} \\ \mathbf{X}_2^{(2)} \end{array} \right. ,$$

$$\left[ \begin{array}{l} \mathbf{X}_1^{(2)} \\ \mathbf{X}_2^{(2)} \end{array} \right] = \bar{\mathbf{X}}_1^{(2)} ,$$

$$\left\{ \begin{array}{l} \mathbf{X}_1^{(3)} = \boxed{\phantom{\mathbf{X}_1^{(2)}}} \mathbf{X}_1^{(2)} + d\mathbf{X}_1^{(3)} \\ \mathbf{X}_2^{(3)} \end{array} \right. ,$$

⋮

$$\left\{ \begin{array}{l} \mathbf{X}_1^{(n)} = \boxed{\phantom{\mathbf{X}_1^{(n-1)}}} \mathbf{X}_1^{(n-1)} + d\mathbf{X}_1^{(n)} \\ \mathbf{X}_2^{(n)} \end{array} \right. .$$

It can be seen that all the information about the previously estimated parameters is always made available to the subsequent step

### Testing of Accuracy of Fit

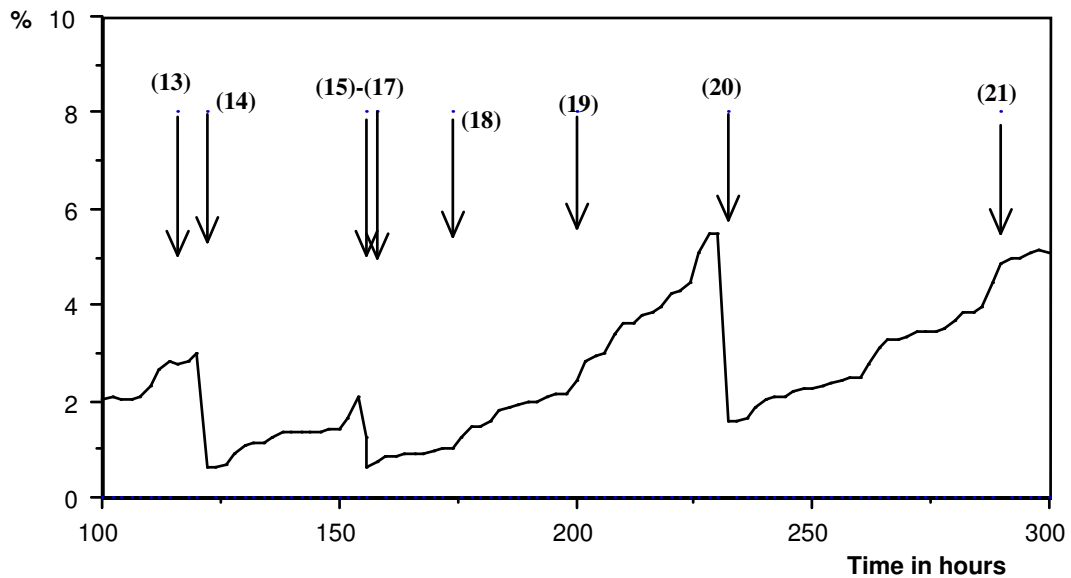
To test the performance of our sequential algorithm, we had generated a synthetical hourly series consisting of the most dominant 60 theoretical (astronomical) constituents. **(Mr.Hou - have you also included any shallow water constituents here? If yes, say so.)** We then analysed this series, starting with the first 100 values and proceeding till 300 hours (i.e.,12.5 days) were reached, using a step of .....hours. From the lumped constituent tables the program selected 12 lumped constituents to be fitted to the first 100-value series and ended up fitting 21 lumped constituents to the whole 300-



value series. At each step we plotted the relative root-mean-square error (RMS)  $\frac{\sigma_i}{|\zeta|}$ , defined as:

$$\sigma_i = \left[ \sum_j^{N_i} (\zeta_j - \widehat{\zeta}_i)^2 / (N-2m-1) \right]^{1/2}, \quad |\zeta| = \left[ \sum_j^N \zeta_j^2 / N \right]^{1/2} \quad (47)$$

These relative RMS are shown in Figure 6.3, where the symbols (13), (14), ..., (21) indicate the number of lumped tidal constituents used at any particular time.

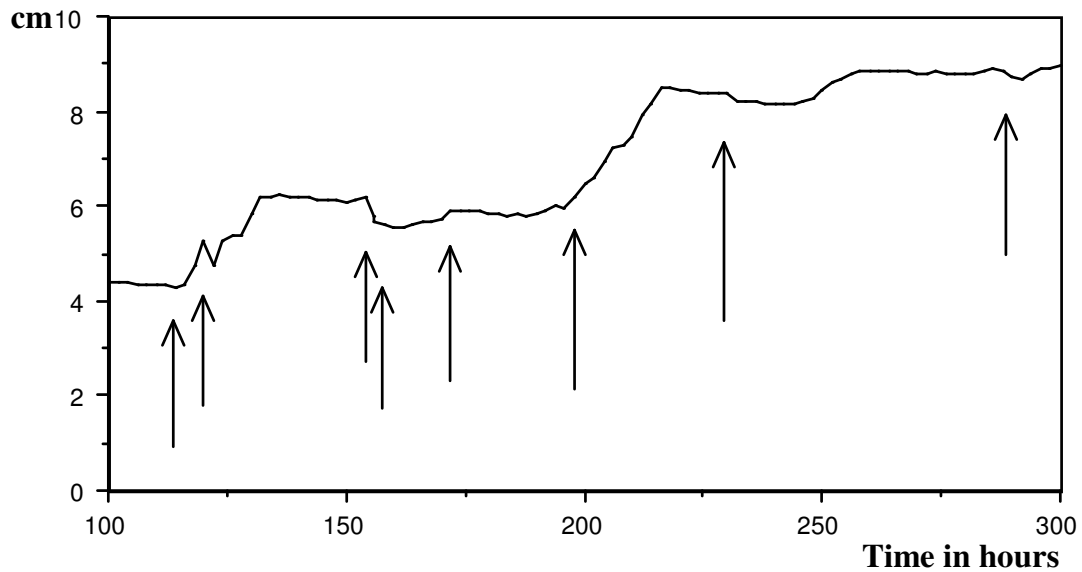


**Figure 2 - Relative RMS of the Sequential Analysis**

The shape of the curve demonstrates that when new tidal constituent are added (really, when used lumped constituents are separated), the relative accuracy of the fit increases. During the time interval when the number of tidal constituents in the model is fixed, the relative accuracy decreases until the next separation of constituents occurs. It implies that before adding the next constituent at a certain stage, the tidal model find it more and more difficult to fit properly the current data series. Intuitively, this behaviour obviously makes a good sense

For comparison, we give the standard deviation curve of real tide-gauge data analysis at Halifax (Fig. 3). It shows that the values of standard deviations of the fit also generally decrease when a new constituent is added to the model. **(Mr.Hou - Here, we have to explain what the standard deviation is and how does it relate to the above used relative RMS.)** The situation in this

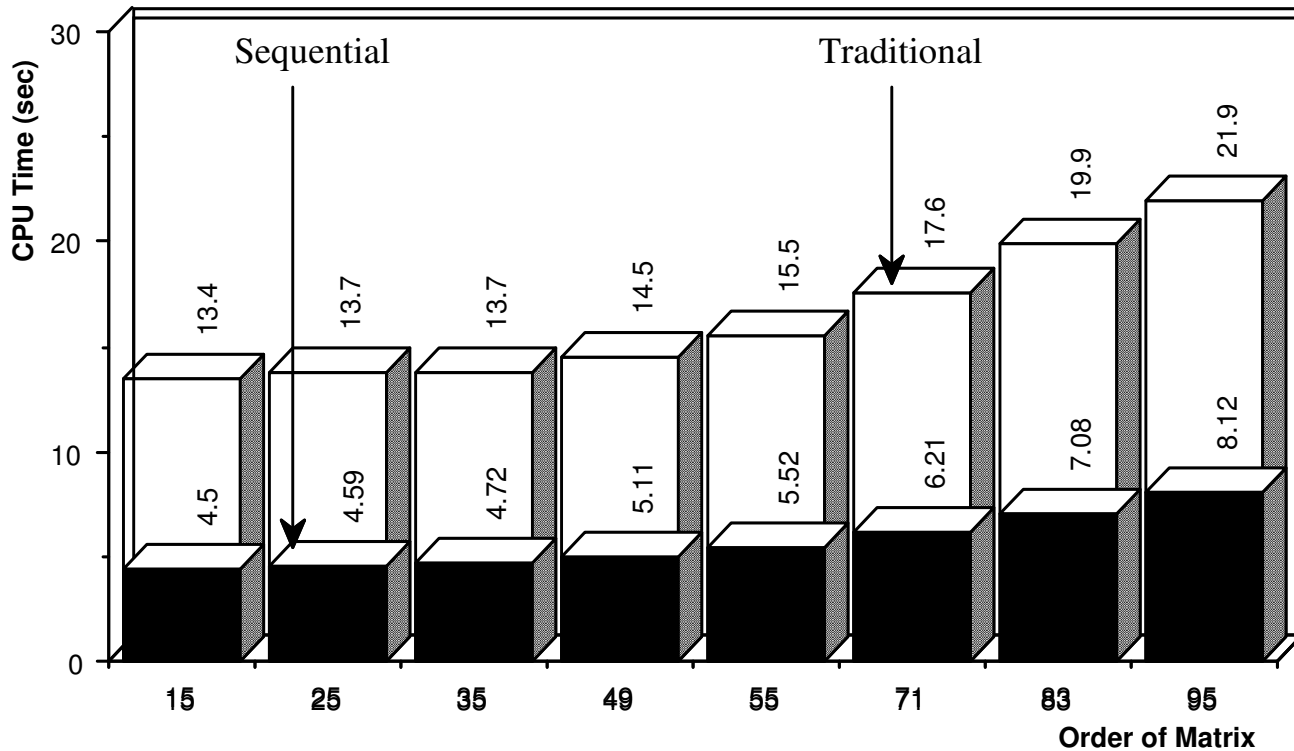
case is more complicated however, because of the presence of non-tidal signals in the data. Thus, Figures 2 and 3 cannot be compared directly.



**Figure 3 - Standard Deviations of the Sequential Analysis at Halifax**

### **Computation Speed Testing**

The tidal harmonic analysis results at permanent tide gauges should be kept up-to-date to maintain the quality of tidal prediction at any given time. To do this, large systems of linear equations have to be solved repeatedly. This will require a lot of CPU time, which, in turn, will increase the cost. It is thus of natural interest, to determine just how much faster the analysis can be performed using the sequential approach. The comparison of the time consumption of the traditional harmonic analysis with that of the sequential approach is given in Fig. 4.



**Figure 4 - CPU Time Consumption of the Two Methods**

The difference in the CPU time consumption for obtaining 10 solutions (with increasing number of unknown parameters) from the two methods on a mainframe IBM computer is seen very clearly. For instance, if the tidal model contains 7 constituents (15 unknown parameters), solving the pertinent 15 normal equations 10 (**Mr. Hou - Why do we solve the equations 10-times? Can we not make the same point simply by solving the system just once?**) times to get 10 updated solutions, the standard method spends 13.4 seconds of CPU time, and sequential method spends 4.5 seconds - a difference of 8.9 seconds. If the size of the system of equations is increased to 95, the difference of CPU time needed by the two methods increases to 13.8 seconds. With further increases in the number of needed constituents, the CPU time saving increases progressively

### **Comparison Between Using Pure and Lumped Constituents**

The lumped tidal constituent tables discussed above are based on the astronomical tidal constituents and created by the least squares method, in which the covariance matrix of the estimated constituents' amplitudes and phases is inspected by using a specific correlation criterion. With the lumped constituent tables, we establish a standard model that includes as many tidal components as is possible with the limited length of observational series, while assuring that no ill-conditioned normal equation matrix results in any of the steps

of the sequential algorithm. As we have mentioned earlier, the lumped tidal constituents can be considered as a good representation of the pure astronomical and shallow water constituents when the time series is short. For demonstrating the differences between using the two kinds of constituents, pure and lumped, two data series, a synthetic one and the observed data series at Halifax, were analysed. The resulting standard deviations of the respective fits are shown in Table 4.

	<b>N (hour)</b>	<b>Standard Deviation of observation (cm)</b>	<b>Standard Deviation of Equilibrium tide (cm)</b>
<b>Astronomical</b>	<b>40</b>	$\pm$ <b>5.826</b>	$\pm$ <b>1.414</b>
<b>Tidal</b>	<b>70</b>	$\pm$ <b>5.213</b>	$\pm$ <b>1.303</b>
<b>Constituents</b>	<b>100</b>	$\pm$ <b>5.046</b>	$\pm$ <b>0.264</b>
<b>Lumped</b>	<b>40</b>	$\pm$ <b>4.599</b>	$\pm$ <b>0.408</b>
<b>Tidal</b>	<b>70</b>	$\pm$ <b>3.743</b>	$\pm$ <b>0.151</b>
<b>Constituents</b>	<b>100</b>	$\pm$ <b>4.410</b>	$\pm$ <b>0.365</b>

**Table 4 - Standard Deviations of Analyses with Pure and Lumped Constituents**

It can be seen that when the series is not very long, the analysis with lumped constituents yields generally more accurate results than the one with the pure constituents. The only exception is found for the longest analysed stretch of the synthetical data series ( $n = 100$  hours). The reason is that the lumped constituents used in the analysis contain some shallow water contributions. When the length of the series is increased, the effect of these contributions in the lumped constituents become more and more apparent and the misfit of the fitted series to the synthetical series (generated from purely astronomical constituents) becomes more and more obvious. It may be assumed that if the lumped constituent tables were constructed without considering the shallow water effects, the accuracy of the analysis (using these lumped constituents) would also be higher.

As one may expect, prediction with lumped tidal constituents gives also a higher accuracy than that with pure astronomical constituents. This can be seen from numerical results listed in Table 5, constructed for  $N = 100$  and  $N_1 = 30$ .

	Unit Astronomical tide	(cm)
lumped tides		
Stand. deviations for estimates	4.410	5.046
Stand. deviations for predictions	12.881	15.291

### **Table 5 - Standard Deviations of Predictions with Pure and Lumped Constituents**

When the length of the time series is increased, the results by using the two kinds of constituents get closer together, until the difference between them completely disappears. This happens, when the series becomes sufficiently long so that all the lumped constituents can be separated into their component constituents.

### **Conclusions**

The sequential tidal harmonic analysis proposed in this study can be used to provide up-to-date information for ocean tidal predictions in real time. Once new hourly observations (one, two or several hours) become available, updated results (estimated new amplitudes and phase lags, and their standard deviations) can be obtained with very little CPU time expenditure, as the solution time is only weakly dependent on how many tidal constituents are included in the tidal model. If desired, the predicted values  $\hat{y}$  can be naturally computed in each sequential step.

For obtaining accurate results by the sequential algorithm, lumped tidal constituent tables have been constructed for sequential separation of tidal constituents. This has been done by using the correlation matrix for estimated tidal amplitudes and phases without considering any tidal observations and applying a specific criterion for maximum allowable correlation. If the need arises, these tables can be recomputed for a different criterion.

Lumped tidal constituents are a realistic representation of pure astronomical tidal constituents over all tidal frequency bands with observational series of a limited length. The lumped tidal constituents are all separated into the pure astronomical components, when the length of the observational series is greater than 19 years. **(Mr.Hou - The last sentence belongs in the section where we describe the lumping. I would like to do a bit more work on the Conclusions later; they are a little lean!)**

### **REFERENCES**

- Doodson, A. T., (1923) The Harmonic Development of the Tide-generating Potential, Proc. Roy. Soc., London, pp 305-329.
- Godin, G., (1972), The Analysis of Tides, University of Toronto Press, Toronto.

- Grant, S. T., (1988), Simplified Tidal Analysis and Prediction. Lighthouse: Edition 37.
- Melchior, P., (1983) The Tides of the Planets Earth, Press, Great Britain.
- Merry, C. L., (1980), Processing of Tidal Records at Hout Bay Harbour, "International Hydrographic Review", Monaco, LVII(1), pp. 149-162
- Munk, W.H., and Cartwright, D.E., (1966), Tidal Spectroscopy and Prediction, "Philosophical Transactions of the Royal Society of London". Vol. 259, pp. 533-581.
- Thompson, K. R., (1979) , Regression Models for Monthly Mean Sea Level. "Marine Geodesy", pp. 269-290.
- Vanicek, P., (1978), To the Problem of Noise Reduction in Sea-level Records Used in Vertical Crustal Movement Detection, "Physics Of the Earth and Planetary Interiors". 17 pp. 265-280.
- Vanicek, P., and Krakiwsky, E. J., (1986), Geodesy, The Concepts, North Holland.
- Vanicek, P., and Wells, D., (1972), The Least Squares Approximation and Related Topics, Department of Surveying Engineering U.N.B., Fredericton, TR# 22.
- Xi Qin-wen, and Hou Tianhang, (1987) A New Complete Development of the Tide-Generating Potential for the Epoch J2000.0, ACTA Geophysica Sinica Vol.30, No.4, pp 349-362.
- Zetler, B. and Cartwright, D. and et al, (1979), Some Comparisons of Response and Harmonic Tide Predictions. International Hydrographic Review, Monaco, LVI (2).

Note - w's should be replaced with 'omegas'.

I have not checked formulae. Some of them look screwed up. They should all be checked thoroughly after the paper is printed.

Insert systematically two spaces before and after symbols in the text.

Typing of equations should be cleaned up (bold-faced subscripts?) including paragraphing - skip one line before and one line after the equation.

Morrisson missing in citations. Check that all publications listed are actually quoted in the text!

I would like to edit the Tables once they are printed. Ther seems to be some problems there with wording and with placing of headings.