DRAFT

Proceedings of the ASME 27th International Conference on Offshore Mechanics and Arctic Engineering
OMAE2008
June 15-20, 2008, Estoril, Portugal

OMAE2008-57461

EMPIRICAL ASSAY OF THE USE OF THE HILBERT-HUANG TRANSFORM FOR THE SPECTRAL ANALYSIS OF STORM WAVES.

Joaquín Ortega
CIMAT, A.C.
Guanajuato, Gto. Mexico

George H. Smith
University of Exeter
U.K.

ABSTRACT

The Hilbert-Huang Transform (HHT) was proposed by Huang et al. [1] as a method for the analysis of non-linear, non-stationary time series. This procedure requires the decomposition of the signal into intrinsic mode functions using a method called empirical mode decomposition. These functions represent the essential oscillatory modes contained in the original signal. Their characteristics ensure that a meaningful instantaneous frequency is obtained through the application of the Hilbert Transform.

The Hilbert Transform is applied to each intrinsic mode function and the amplitude and instantaneous frequency for every time-step is computed. The resulting representation of the energy in terms of time and frequency is defined as the Hilbert Spectrum.

In previous work [6] using the HHT for the analysis of storm waves it has been observed that number of IMFs needed for the decomposition and the amount of energy associated to different IMFs differ from what has been observed for the analysis of waves under 'normal' sea conditions by other authors. In this work we explore in detail the effect that the sampling rate has in the empirical mode decomposition and in the Hilbert Spectrum for storm waves. The results show that the amount of energy associated to different IMFs varies with the sampling rate and also that the number of IMFs needed for the empirical mode decomposition changes with record length.

INTRODUCTION

The Hilbert-Huang Transform (HHT) was proposed by Huang et al. [1, 2, 3] as an adequate method for the spectral analysis of non-stationary, nonlinear processes. Since then it has been used by several authors for the analysis of sea waves under different conditions (Schlumann [7], Veltcheva and Guedes Soares [8], Veltcheva [9], among others).

In previous work the authors used the HHT for the analysis of a North-Sea storm [6] and found some differences with the results previously obtained by other authors. Specifically, the number of Intrinsic Mode Function (IMF) obtained in the Empirical Mode Decomposition and the energy distribution among the different IMFs are different. In this paper we explore the effect of the sampling frequency on the results of the HHT algorithm, as a possible explanation for the differences found.

HHT

We give a brief description of the Hilbert Huang Transform. A detailed presentation can be found in the original articles of Huang et al. [1, 2] as well as in Huang [4, 5].

The Hilbert Huang Transform is based on an empirical algorithm called the Empirical Mode Decomposition (EMD), used to decompose a time series into individual characteristic oscillations known as the Intrinsic Mode Functions (IMF). This technique is based on the assumption that any signal consists of different modes of oscillation based on different time scales, so that each IMF represents one of these embedded oscillatory modes. Each IMF has to satisfy two criteria: 1) The number of local extreme points and of zero-crossings must either be equal or differ at most by one, 2) At any instant, the mean of the envelope defined by the local maxima and the envelope corresponding to the local minima must be zero. These two conditions are required to avoid inconsistencies in the definition of the instantaneous frequency.

Once the signal is decomposed, the Hilbert Transform is applied to each IMF. The Hilbert transform $y(t)$ of a function $x(t)$ is defined as $(1/\pi)$ times the convolution of x with the function $1/t$:

$$y(t) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{1}{t-s} x(s) ds \quad (1)$$

where the integral is taken as the Cauchy principal value. Then, if $z(t)$ is the analytical signal associated to $x(t)$, we have for all t

$$z(t) = x(t) + iy(t) = A(t)\, exp(i\theta(t)) \quad (2)$$

with $A(t) = (x^2(t) + y^2(t))^{1/2}$ and $\theta(t) = arctan(y(t)/x(t))$. The instantaneous frequency is defined now as the derivative of the phase function of the analytical signal $z(t)$:

$$\omega(t) = \frac{d\theta(t)}{dt}. \quad (3)$$

Once the signal has been decomposed into IMFs and the Hilbert transform for each has been obtained, the signal x(t) can be represented as

$$x(t) = \sum_{j=1}^{n} A_j(t) \exp\left(i \int \omega_j(t) dt\right) \quad (4)$$

which is a generalized form of the Fourier expansion for x(t) in which both amplitude and frequency are functions of time.
The time-frequency distribution of the amplitude or the amplitude squared is defined as the Hilbert amplitude spectrum or the Hilbert energy spectrum, respectively.

DATA

Data was recorded from the North Alwyn platform situated in the northern North Sea, about 100 miles east of the Shetland Islands (60º48.5' North and 1º44.17' East) in a water depth of approximately 130 metres. There are three Thorn EMI Infra-red wave height meters mounted on the platform and their heights are between 25 and 35 metres above the water. The data are recorded continuously and simultaneously at 5 Hz and then divided into 20 minute records for which the summary statistics of H_s, T_p and the spectral moments are calculated. For data with H_s > 3m all the surface elevation records are kept. Further details are available in Wolfram et al. [10]. Only data from the North East altimeter are used here.
The data set examined consists of a series of 410 records of 20 minutes, sampled by the altimeter at a rate of 5 Hz., starting on November 16th, 1997. There is a 16 minutes gap of missing data after 32 hours and 20 minutes so that we have two long records, one of 97, 20-minutes records (32h. 20m.) and the other of 313 intervals of 20 minutes (104 h. 20m.).
For the analysis, the data set was divided into 10 intervals of 12 hours, plus two remainders of 8h. 20 min, denoted by P and R, respectively,, in figure 1. This shows the significant wave height (H_s) for the data, calculated for each 20 min. interval, and the divisions of the data into intervals. Also shown is a 7 point moving average of the significant wave height, which gives a regularized view of the evolution of this parameter during the storm.

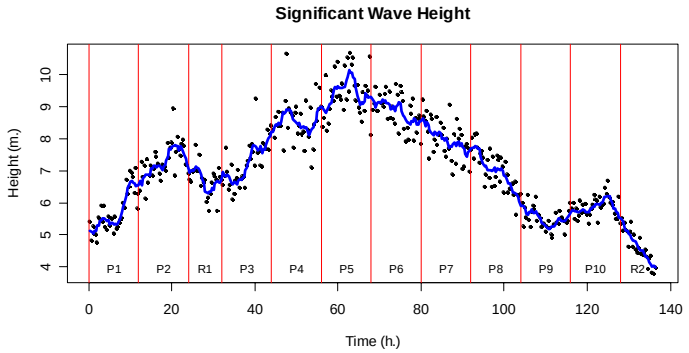

Figure 1. Significant wave height calculated every 20 minutes. The line is a 7 point moving average.

A detailed account of the HHT analysis of these data is given in [6] where we used, as we also did for this work, the Hilbert-Huang Transform Data Processing System (HHT-DPS), a software developed by NASA. This implementation of the HHT algorithm has several options for the stopping criteria and the endpoint behaviour of the splines. We used Copy Endpoints as the Endpoint Prediction choice with 7 siftings. The procedure for choosing these values is described in detail in [6]. We found that the number of IMFs obtained in the EMD of the data was between 20 and 27 IMFs, depending on the interval considered, which is about twice the amount needed for the HHT analysis reported by other authors under different sea conditions. Also, the energy distribution among the different IMFs is different. Figure 2 shows the contribution of each IMF as a percentage of the total energy of the period.

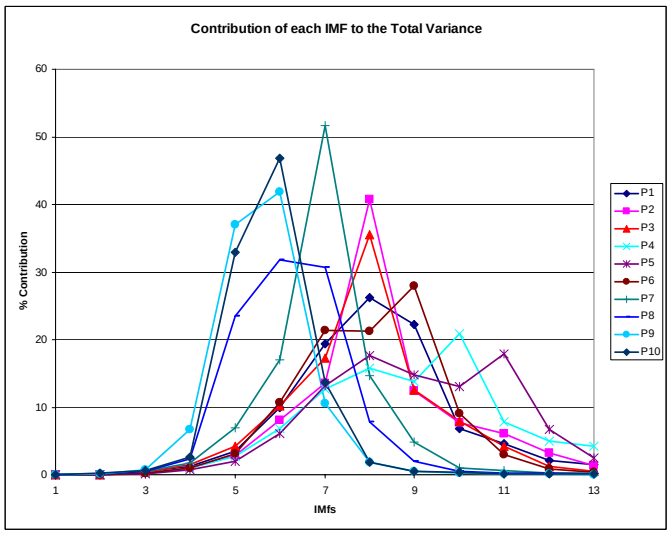

Figure 2: Contribution of each IMF to the Total Variance, all Periods.

As can be seen, the main contribution comes from IMFs 5 to 11, and this is in contrast to what has been reported previously

by Veltcheva and Guedes Soares [8] for various sea conditions off the Portugal coast and Veltcheva [9] for various sea conditions off the coast of Japan, where the first three IMFs are the most energetic ones. For example, in [8], IMF2 was found to be the most energetic followed by IMF3 for two of the data sets, and for the other data set IMF1 is first followed by IMF2. Schlurmann [7] analyses a transient wave recorded in the Sea of Japan and reports IMF2 as having the highest energy.

To explore possible reasons for these differences, we look here at the effect of sampling frequency on the results of the EMD. We consider intervals of one, three and five hours, which are constructed adding 1-hour intervals on each side of the initial one hour interval.

The original data was sampled at a frequency of 5 Hz. To 'change' the sampling frequency we subsample the data we have taking one datum in two, one in three, and so on up to one in ten, so the sampling frequency goes from 5 Hz to 0.5 Hz.

Subsampling Rate	1/1	1/2	1/3	1/4	1/5	1/6	1/7	1/8	1/9	1/10
Sampling Frequency (Hz.)	5	2.5	1.66	1.25	1	0.83	0.714	0.625	0.55	0.5
Sampling Period (s)	.2	.4	.6	.8	1.0	1.2	1.4	1.6	1.8	2.0

Table 1. Subsampling rates.

For each subsampling rate there are as many starting points for the subsampled series as the rate denominator. Whenever the starting point is not explicitly chosen, the same starting point as the original series is used. In this work we focus on the following issues:

- Effect on the number of IMFs.
- Effect on the amount of energy for each IMF.
- Effect on the joint distribution of amplitude and frequency.
- Effect on the EMD.

NUMBER OF IMFs

Although there are some changes, the number of IMFs remains relatively stable for the different sampling rates, and varies from 8 (for one case of the smallest sampling rate) to 12 for the highest sampling rate, for the 1-hour period. For the 3-hour period they go from 11 to 15 IMFs while for the 5-hour period they go from 11 to 16. In Table 2 we give the range of the number of IMFs for each sampling rate (depending on the starting point) and also the mode.

As can be seen the most frequent value is 10 IMFs for the 1 hour period, 12 for the 3-hour period and 13 for 5-hour one. Figure 2 gives the number of IMFs for each value of the sampling period and the regression line for the 3-hour period. Observe that a small noise has been added to the values to avoid superposition.

	Sampling Period	.2	.4	.6	.8	1.0	1.2	1.4	1.6	1.8	2.0
1h	Range	12	11	11	9-11	10-11	9-10	10	9-11	9-10	8-11
	Mode	12	11	11	10	10	9-10	10	10	10	10
3h	Range	15	12-13	13	12-13	11-13	11-13	11-12	11-12	11-12	11-12
	Mode	15	12-13	13	12	12	12	12	12	11	11
5h	Range	16	13-14	13-14	13-14	13-14	13-14	12-13	12-13	12	11-13
	Mode	16	13-14	14	13-14	13	13	12	13	12	13

Table 2. Number of IMFs in the EMD.

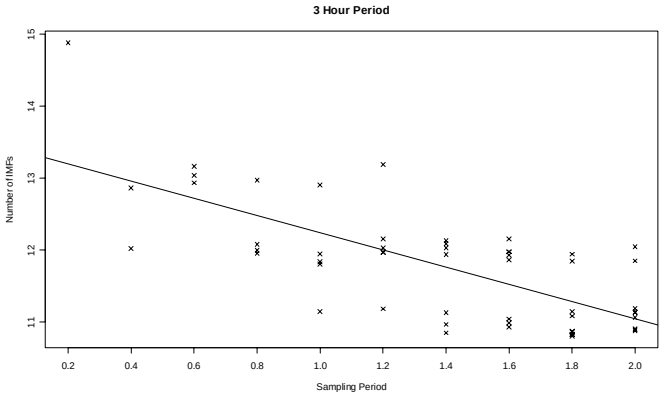

Figure 2. Number of IMFs for each sampling period and regression line, 3-hour period

ENERGY

The energy associated to an IMF clearly changes with the sampling frequency. Table 3 gives the amount of variance (energy) as a percentage of the total variance (energy) for each IMF for the 1-hour period. Figure 3 is the corresponding graph for the first 5 IMFs.

		IMF											
		1	2	3	4	5	6	7	8	9	10	11	12
Sampling Period	0.2	0.22	5.05	56.61	30.9	4.83	0.93	0.69	0.40	0.24	0.10	0.02	0.03
	0.4	3.74	64.44	27.10	2.72	0.78	0.51	0.23	0.12	0.12	0.16	0.08	
	0.6	9.87	71.87	15.28	1.28	0.61	0.49	0.27	0.12	0.12	0.07	0.03	
	0.8	26.24	62.54	8.63	1.19	0.53	0.32	0.18	0.08	0.20	0.03	0.07	
	1	38.74	53.61	5.55	0.90	0.46	0.37	0.16	0.10	0.06	0.02		
	1.2	53.75	40.67	3.16	0.92	0.58	0.41	0.25	0.12	0.06	0.07		
	1.4	57.94	36.76	2.97	0.80	0.56	0.42	0.18	0.20	0.06	0.07		
	1.6	66.02	29.30	2.21	0.82	0.61	0.51	0.23	0.15	0.11	0.04		
	1.8	73.73	21.92	2.04	0.91	0.56	0.39	0.19	0.12	0.08	0.04		
	2	75.62	20.02	1.83	1.12	0.63	0.39	0.19	0.10	0.06	0.02		

Table 3. Energy associated to each IMF as a function of sampling rate, 1 hour period

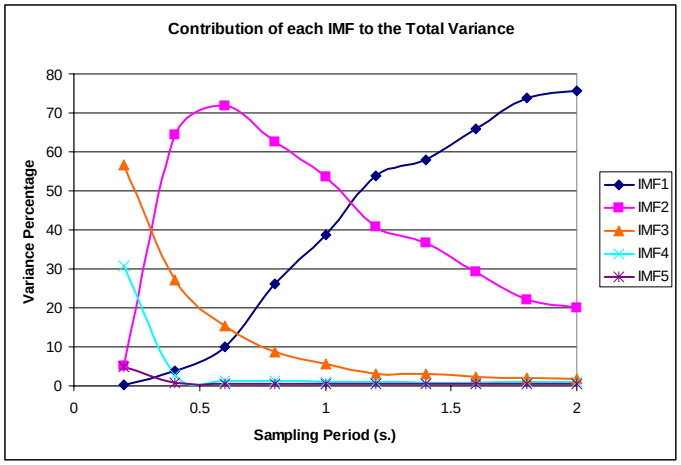

Figure 3. Contribution of IMFs 1-5 to the Total Variance, 1 hour period.

		IMF											
		1	2	3	4	5	6	7	8	9	10	11	12
Sampling Period	0.2	.05	.66	17.08	49.3	27.7	3.57	0.61	0.42	0.31	0.12	0.07	0.10
	0.4	4.93	44.08	41.95	6.24	0.94	0.46	0.30	0.21	0.43	0.07	0.07	0.13
	0.6	13.67	64.08	19.02	1.64	0.66	0.31	0.21	0.17	0.10	0.06	0.03	0.03
	0.8	19.55	64.96	12.87	1.27	0.57	0.32	0.20	0.09	0.06	0.03	0.04	0.02
	1	33.04	57.30	6.96	1.13	0.59	0.45	0.22	0.12	0.05	0.03	0.03	0.05
	1.2	41.89	50.01	5.80	0.97	0.49	0.36	0.24	0.10	0.04	0.05	0.02	0.01
	1.4	51.12	42.69	3.91	0.96	0.52	0.33	0.19	0.15	0.06	0.04	0.02	0.02
	1.6	57.00	36.70	4.00	1.01	0.53	0.34	0.20	0.10	0.05	0.03	0.02	0.02
	1.8	65.99	29.32	2.51	0.97	0.53	0.31	0.19	0.08	0.06	0.02	0.01	0.01
	2	74.21	21.37	2.17	0.98	0.54	0.34	0.19	0.10	0.04	0.03	0.02	0.00

Table 4. Energy associated to each IMF as a function of sampling rate, 3 hour period

		IMF											
		1	2	3	4	5	6	7	8	9	10	11	12
Sampling Period	0.2	0.03	0.15	2.80	37.8	50.3	6.80	1.03	0.44	0.23	0.15	0.11	0.07
	0.4	0.95	16.43	61.01	18.0	2.07	0.61	0.36	0.26	0.12	0.06	0.07	0.03
	0.6	9.73	61.34	24.33	2.85	0.79	0.36	0.25	0.17	0.08	0.06	0.02	0.02
	0.8	12.61	67.07	16.77	2.00	0.65	0.36	0.24	0.13	0.08	0.03	0.03	0.01
	1	19.54	63.67	13.91	1.50	0.60	0.32	0.22	0.11	0.07	0.03	0.02	0.01
	1.2	26.81	59.23	11.00	1.51	0.59	0.35	0.21	0.13	0.07	0.05	0.03	0.02
	1.4	41.86	49.37	6.27	1.09	0.55	0.34	0.25	0.13	0.07	0.04	0.02	0.01
	1.6	53.60	39.54	4.47	1.01	0.55	0.34	0.24	0.11	0.06	0.04	0.02	0.01
	1.8	61.47	32.88	3.34	0.99	0.49	0.33	0.25	0.11	0.06	0.04	0.02	0.01
	2	69.69	25.06	2.87	0.97	0.55	0.35	0.23	0.13	0.07	0.04	0.02	0.01

Table 5. Energy associated to each IMF as a function of sampling rate, 5 hour period

It can be seen that the energy increases with sampling period for IMF1, increases and then decreases for IMF2, and is decreasing (except for some values) for the rest of the IMFs.
For the 3 and 5-hour periods the energy is increasing for IMF1, increasing and then decreasing for IMFs 2 and 3 and decreasing for the rest of the IMFs, as can be seen from tables 4 and 5.

AMPLITUDE AND FREQUENCY

In this section we consider the frequency and amplitude distribution for the different IMFs and the different sampling

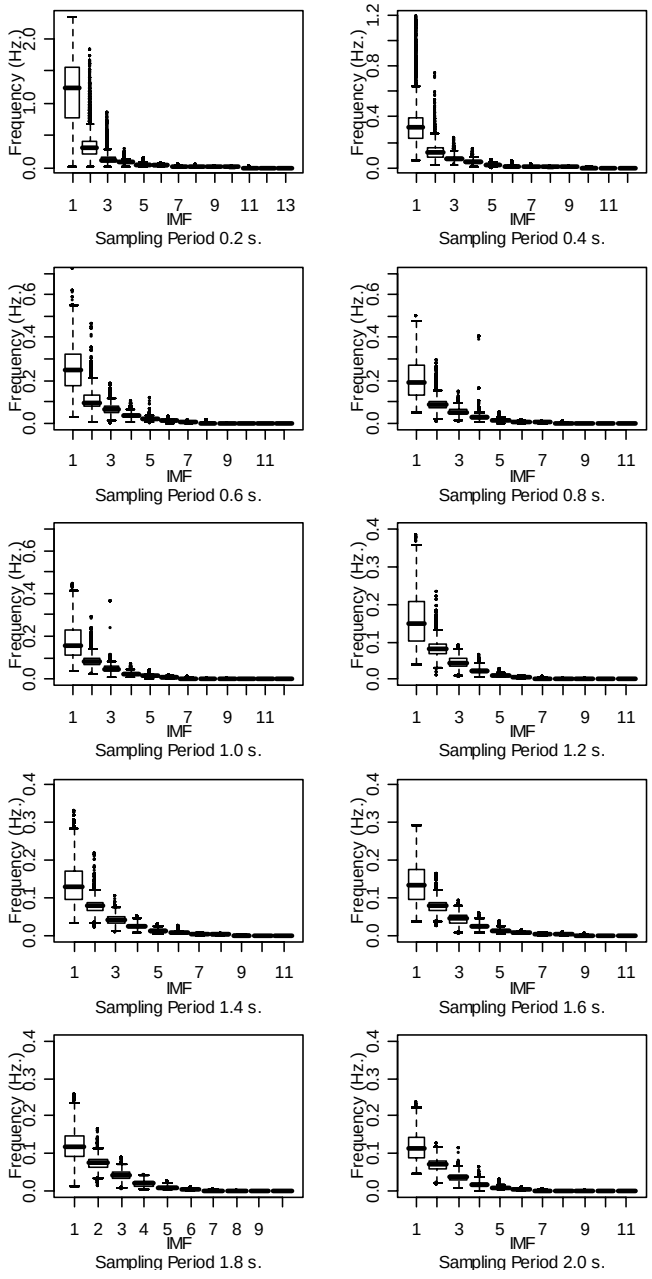

Figure 4. Boxplots for the frequency content of the IMFs, 1 hour period.

periods. Figure 4 gives the boxplots for the frequency distribution. As can be seen the frequency content decreases with IMFs in all cases, but the range of frequencies varies, as one would expect, since a higher sampling rate allows for the detection of higher frequencies in the signal.

Figure 5 shows the amplitude boxplots. They show that the IMF having the largest variation changes from number 3 for sampling period 0.2 s. to number 1 for sampling periods 1.4 s. to 2.0 s. This is consistent with the results presented in the previous section

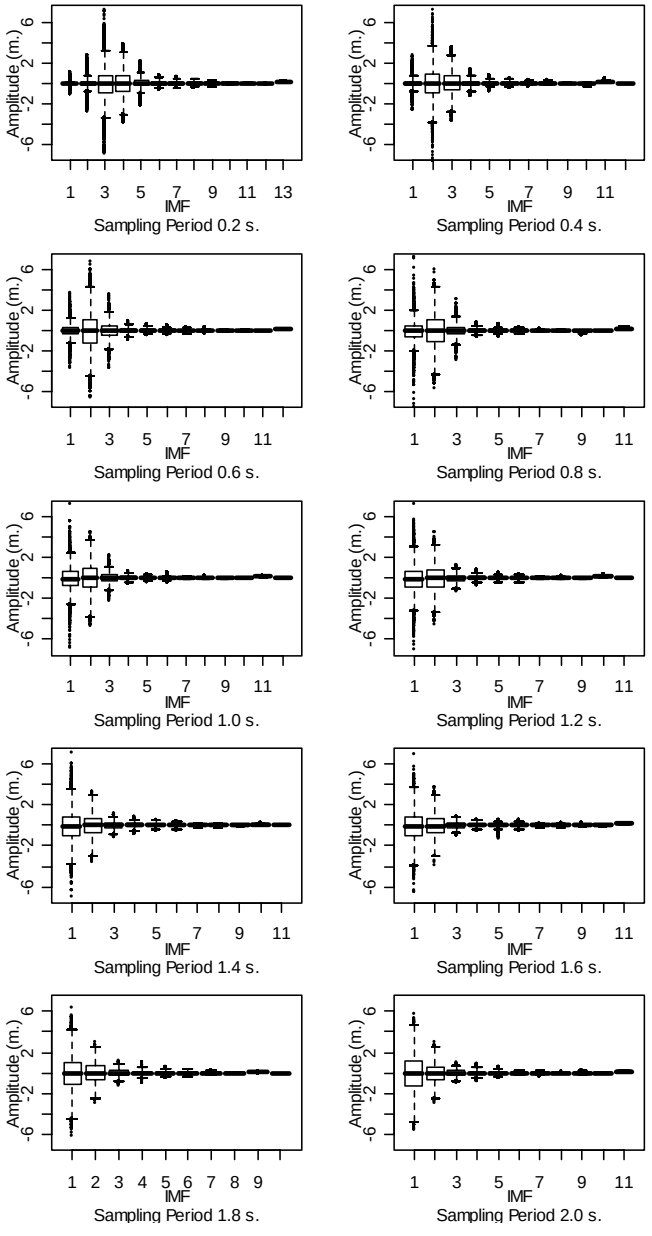

Figure 5. Boxplots for the amplitude content of the IMFs, 1 hour period.

Now we look at the joint distribution for the frequency and amplitude for IMFs 1, 2 and 3 and for the different sampling rates we explore.

IMF1.

There is a very significant change in the distribution as we go from a sampling period of .2 s. to 1.0 s.; this corresponds to the first five graphs in Figure 6, which shows the level curves for the joint density. From sampling period 1.2 s. up to 2.0 s. the joint density is more stable and changes are smaller. Observe that the scale for the x axis is different for the first graph.

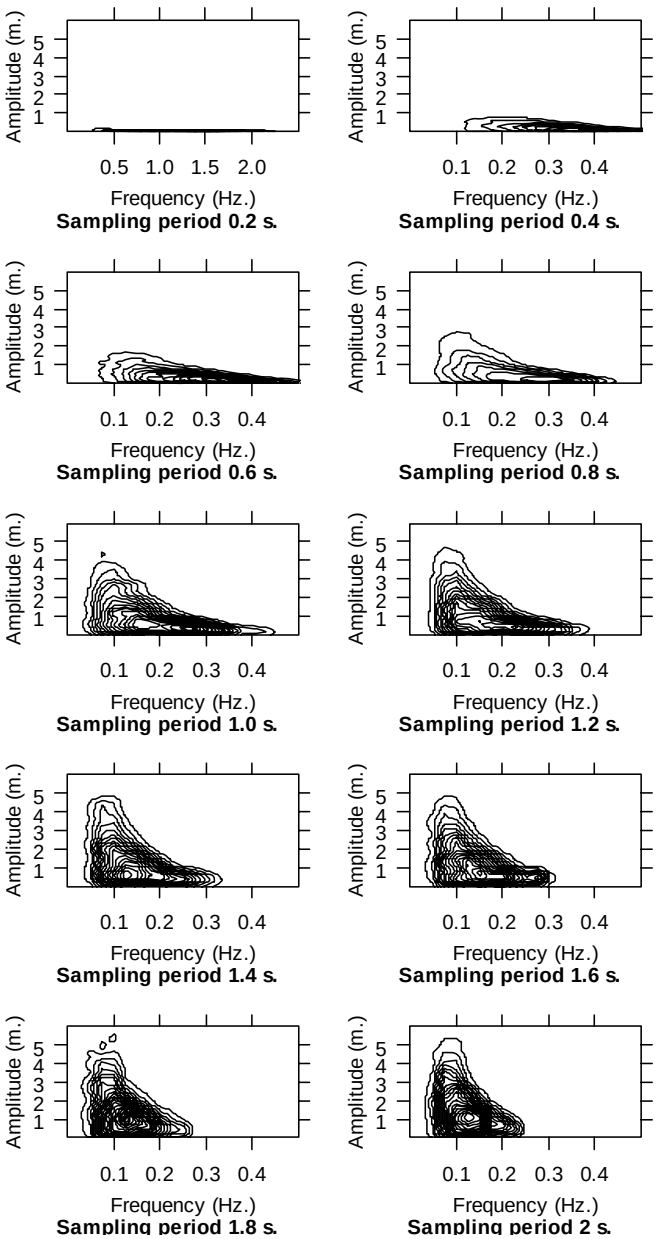

Figure 6. Joint density of frequency and amplitude as a function of sampling frequency for IMF1, 1-hour period.

IMF2

Again there is significant change in the joint density as the sampling period goes from .2 s. to 1.0 s. but although the range of frequencies is more or less stable thereafter, the joint distribution keeps changing, reducing the amplitude range.

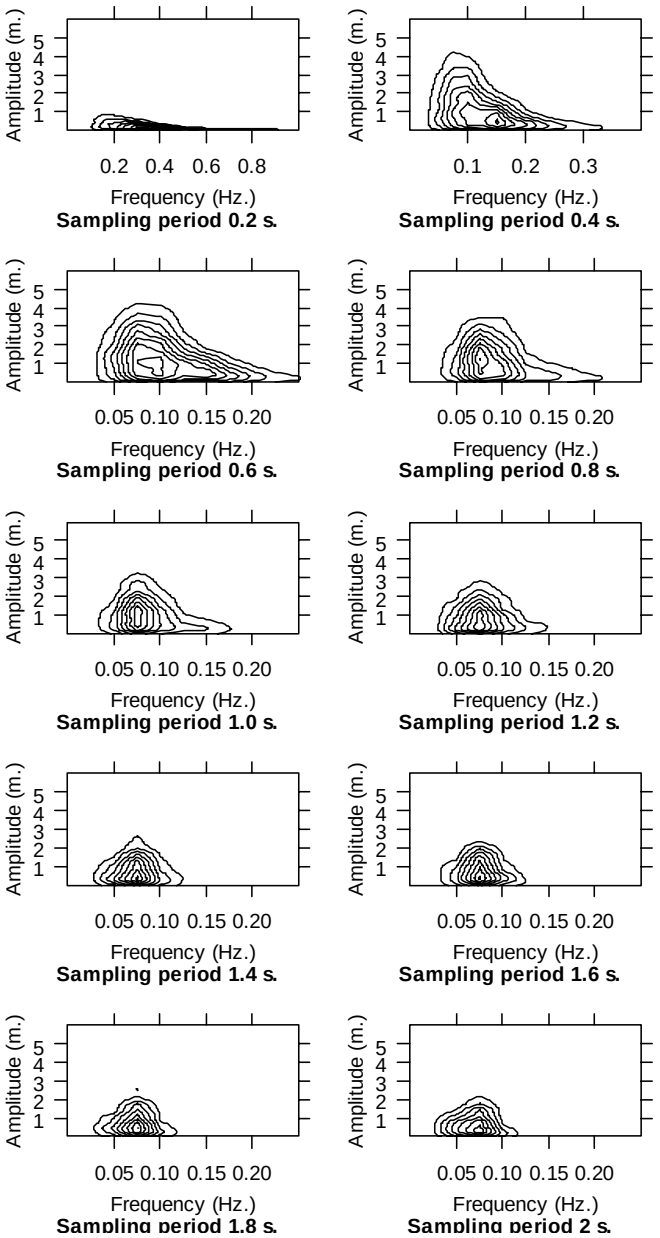

Figure 7. Joint density of frequency and amplitude as a function of sampling frequency. IMF2

IMF3

This case is similar to IMF1: a significant change at first (.2 s. to 1.0 s.) but the distribution is stable thereafter.

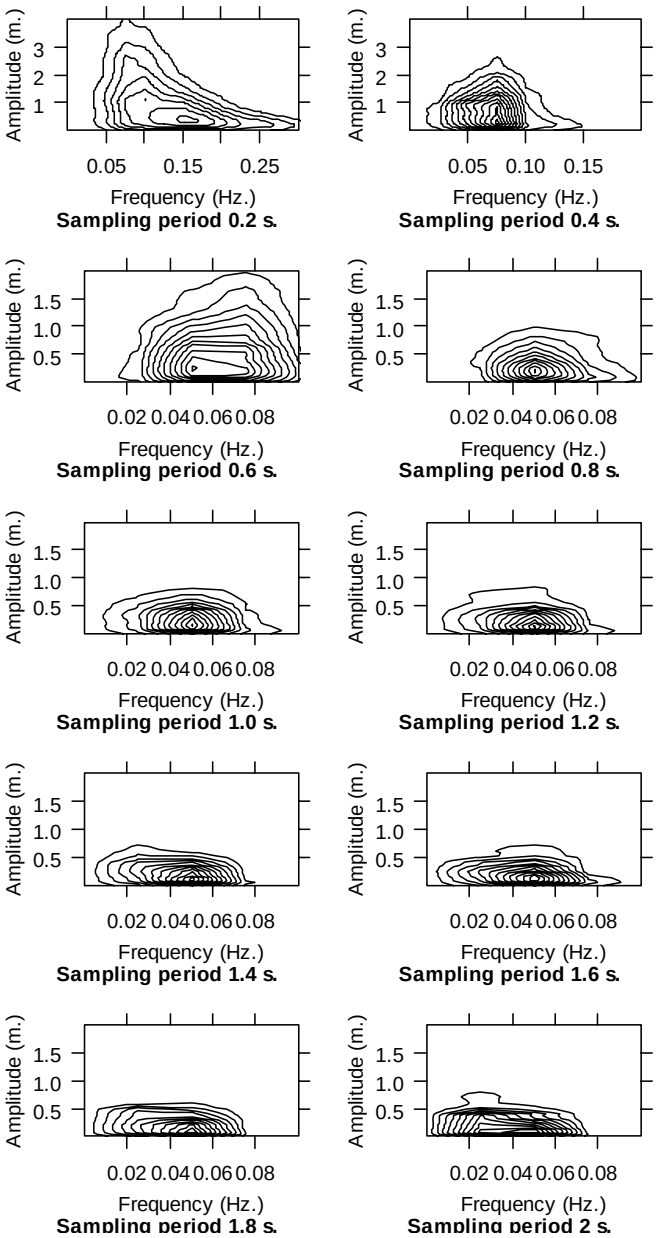

Figure 8. Joint density of frequency and amplitude as a function of sampling frequency. IMF3

EMD

The sampling rate also affects the results of the Empirical Mode Decomposition. Figure 9 shows IMF1 for a 10-minute period in the central one-hour period. As can be seen the shape suffers important changes as the sampling period goes from 0.2 s. to 1.0 s. After that, the shape is more or less stable.

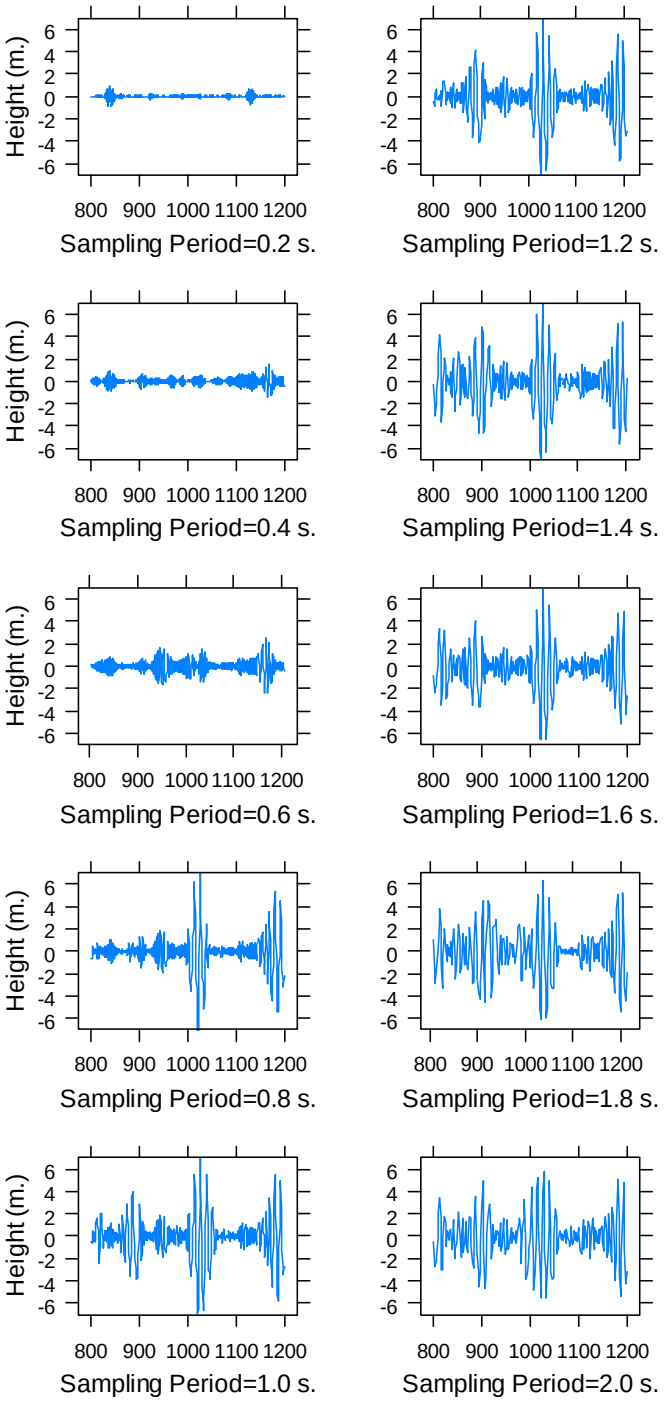

Figure 9. IMF1 for the different sampling periods. 1-hour interval.

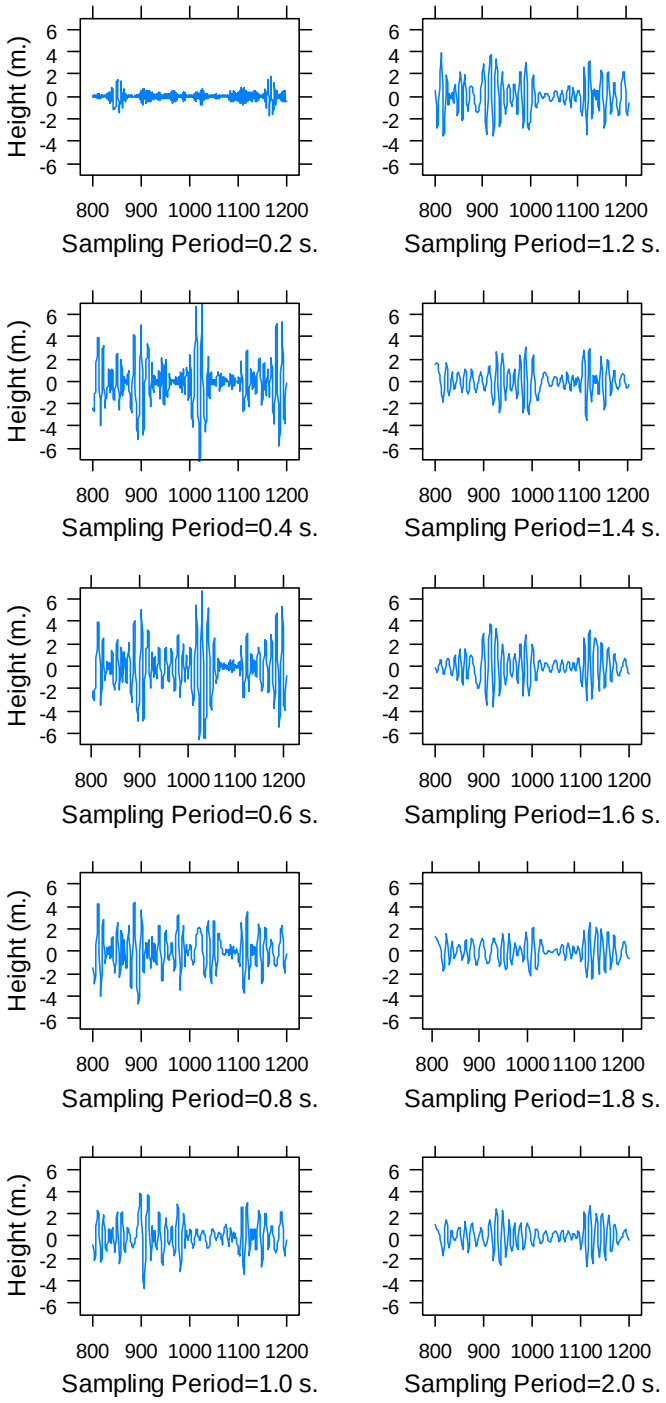

Figure 10. IMF2 for the different sampling periods. 1-hour interval.

Figures 10 and 11 give similar graphs for IMFs 2 and 3. Similar conclusion can be drawn from them.

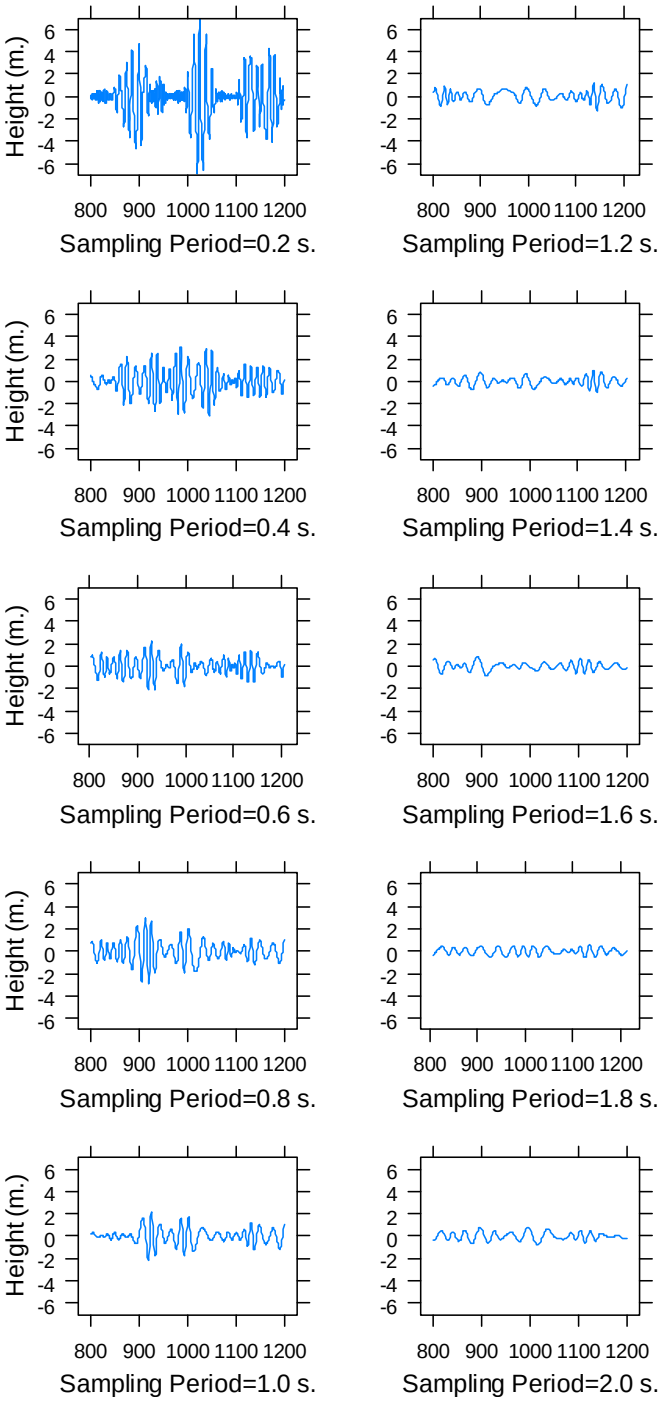

Figure 11. IMF3 for the different sampling periods. 1-hour interval.

For the 3 and 5-hour intervals similar results were obtained, the shape of the IMF changes as the sampling period goes from 0.2 s. to 2.0 s. Another interesting feature can be observed comparing figures 12 and 13, which show IMF1 for the 3 and 5-hour intervals, with figure 9. For a given sampling period and keeping fixed the 10-minute period considered, the shape of IMF 1 changes as the length of record goes from 1 hour to 5 hour. This points to the fact that record length may have an effect on the local shape of the IMFs, and hence on the EMD. This needs further exploration.

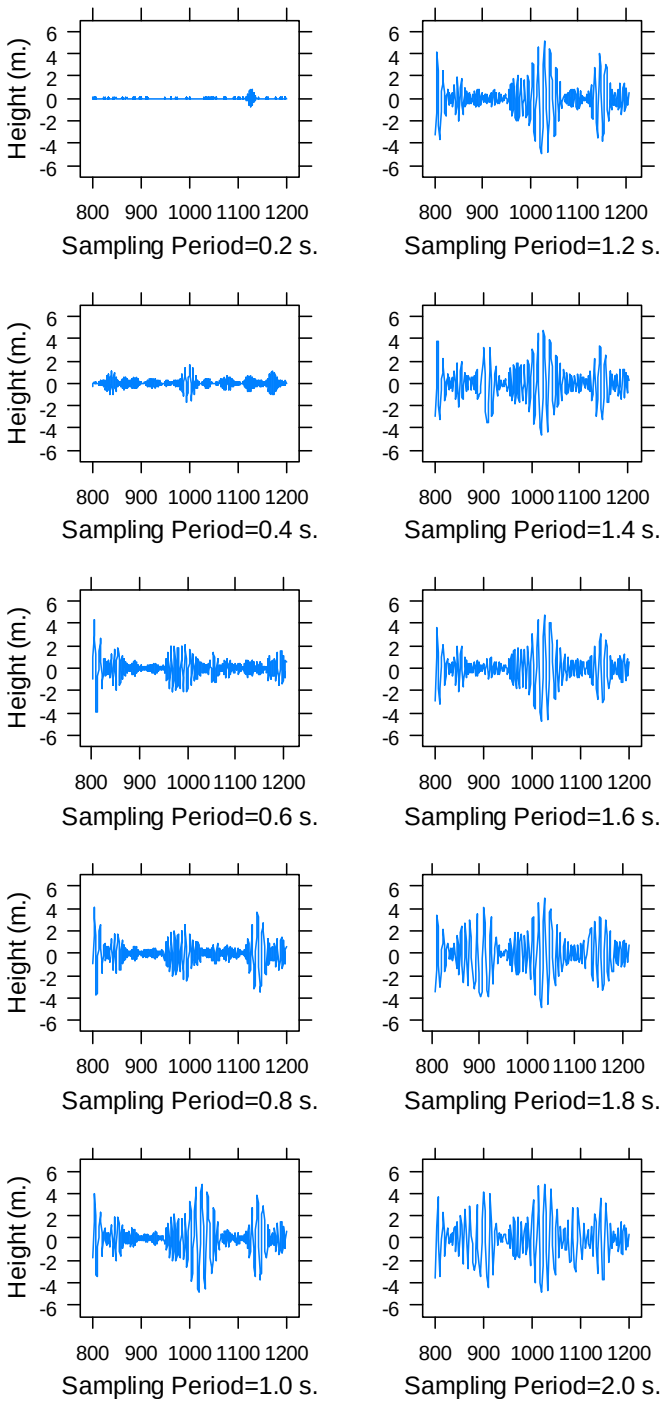

Figure 12. IMF1 for the different sampling periods. 3-hour interval.

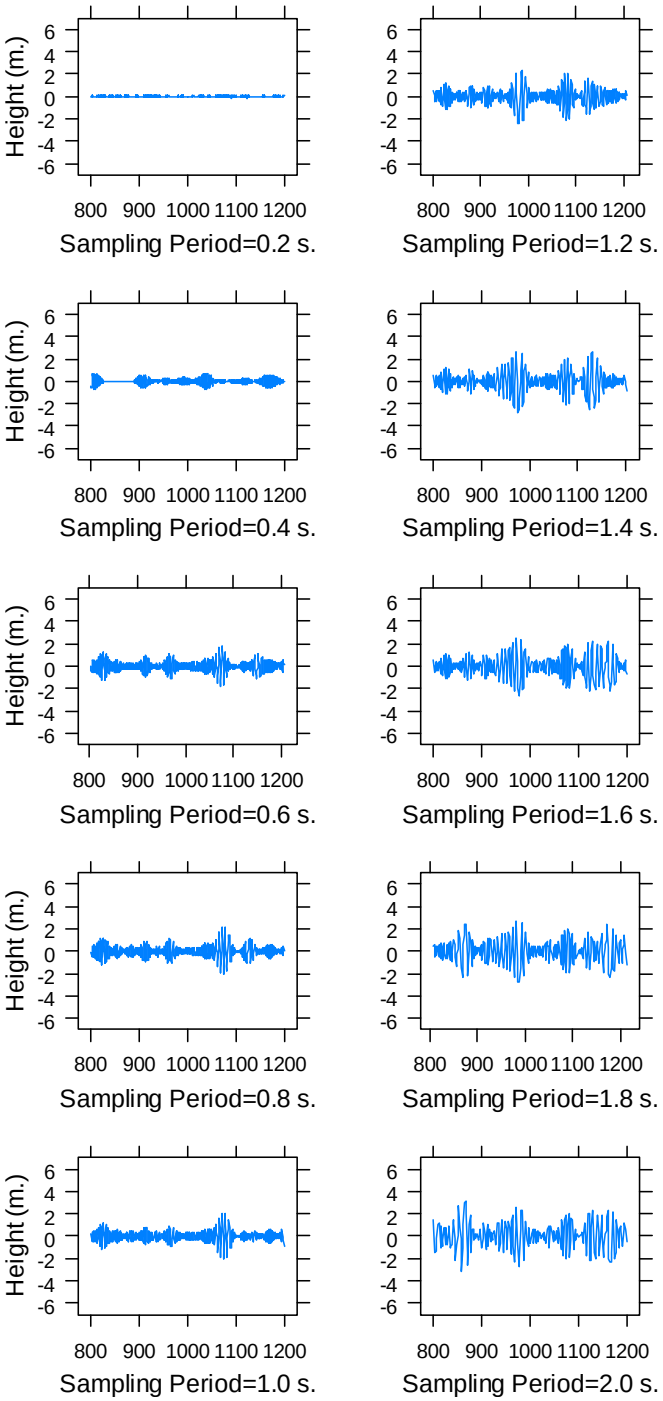

Figure 13. IMF1 for the different sampling periods. 5-hour interval.

CONCLUSIONS

- Sampling frequency were found to have a small effect on the number of IMFs needed for the decomposition of a given signal. As the sampling period increases the number of IMFs decreases. Also, the number of IMFs seems to increase with record length, as table 2 shows.
- The energy distribution is clearly dependent on the sampling frequency. The higher the sampling frequency, the higher the IMF having highest energy contribution to the total energy of the wave record (see tables 3, 4 and 5). Also, record length seems to have an effect on the energy distribution: as the record length increases, the 'main' IMF also increases.
- The joint distribution of amplitude and frequency also changes with the sampling frequency, although the effect is not uniform: the change is more pronounced as the sampling period goes from 0.2 s. to 1 s.
- Finally, the EMD also changes with sampling frequency. Figures 9 to 11 show the changes for a given IMF as the sampling period increases. On the other hand, figures 9, 12 and 13 show the shape of IMF1 for a 10-minute period as a function of the sampling period for different record length. As can be seen the shape changes not only with sampling period but also with record length, which is somewhat unexpected.

These results support the conclusion that the differences observed between the analysis of storm waves and the analysis of waves reported by other authors under different sea condition is partly due to the difference in sampling frequency. There is also evidence that record length may have an effect on the HHT analysis, but this requires a more thorough exploration.

ACKNOWLEDGMENTS

The authors would like to thank Total E&P UK for the wave data from the Alwyn North platform.
The Hilbert-Huang Transform - Data Processing System is copyright United States Government as represented by the Administrator of the National Aeronautics and Space Administration and was used with permission.
This work was partially supported by CONACYT, Mexico, Proyecto Análisis Estadístico de Olas Marinas.

REFERENCES

[1] Huang, N.E., Shen Z., Long S.R., Wu M.C., Shin H.S., Zheng Q., Yuen Y., Tung C.C., Liu H.H. "The empirical mode decomposition and Hilbert spectrum for nonlinear and non-stationary time series analysis," Proc R Soc Lond, 1998; 454: 903–95.
[2] Huang, N.E., Shen Z., Long S.R. "A new view of nonlinear water waves: the Hilbert spectrum," Ann Rev Fluid Mech 1999; 31: 417–57.
[3] Huang, N.E., Wu M.L., Long S.R., Shen S.P., Per W.Q., Gloersen P., Fan K.L. "A confidence limit for the empirical mode decomposition and Hilbert spectral analysis," Proc R Soc Lond 2003; 459: 2317–45.
[4] Huang, N.E. "Introduction to Hilbert-Huang Transform and some recent developments," In: Huang N., Attoh-

Okine N.O., editors. The Hilbert-Huang transform in Engineering. CRC Press, 2005; 1-23.

[5] Huang, N.E. “Introduction to Hilbert-Huang Transform and its related mathematical problems,” In: Huang N.E., Shen S.S.P., editors. Hilbert-Huang Transform and Its Applications. World Scientific, 2005; 1-26.

[6] Ortega, J., Smith, G. H. Hilbert-Huang Transform Analysis of Storm Waves. (In preparation).

[7] Schlurmann, T. “The empirical mode decomposition and the Hilbert spectra to analyse embedded characteristic oscillations of extreme waves,” In: Rogue Waves, Edition Infremer, ISBN: 2-84433-063-0, 2000; 157-165.

[8] Veltcheva, A.D., Guedes Soares, C. “Identification of the components of wave spectra by the Hilbert Huang transform method,” Applied Ocean Res. 2004; 26: 1-12.

[9] Veltcheva, A.D. “An application of HHT method to the nearshore sea waves,” In: Huang N., Attoh-Okine N.O., editors. The Hilbert-Huang transform in Engineering. CRC Press, 2005; 97-119.

[10] Wolfram, J., Feld, G., Allen, J. “A new approach to estimating environmental loading using joint probabilities,” 7th Int. Conf. on Behaviour of Offshore Structures, Pergamon, Boston, 1994; Vol. 2: 701-713.