Wave modelling – The state of the art

The WISE Group
L. Cavaleri a,*, J.-H.G.M. Alves b, F. Ardhuin c, A. Babanin d, M. Banner e, K. Belibassakis f, M. Benoit g, M. Donelan h, J. Groeneweg i, T.H.C. Herbers j, P. Hwang k, P.A.E.M. Janssen l, T. Janssen m, I.V. Lavrenov n, R. Magne o, J. Monbaliu p, M. Onorato q, V. Polnikov r, D. Resio s, W.E. Rogers t, A. Sheremet u, J. McKee Smith s, H.L. Tolman v, G. van Vledder w, J. Wolf x, I. Young d

a Institute of Marine Sciences, Castello 1364/A, 30122 Venice, Italy
b MetOcean Engineers Pty Ltd, 31 Bishop Street, Jolimont, WA 6014, Australia
c SHOM/CNIORED, French Naval Oceanographic Center, 13 rue du Chatellier, 29609 Brest Cedex, France
d Swinburne University of Technology, P.O. Box 218, Hawthorn, Victoria 3122, Australia
e School of Mathematics, UNSW, Sydney 2052, Australia
f Shipbuilding Department, National Technical University of Athens, Agiou Spyridonos, 12210 Egaleo, Greece
g Laboratoire National d'Hydraulique et Environnement (LNHE), 6 quai Watier, BP49, 78400 Chatou Cedex, France
h University of Miami, RSMAS/AMP, 4600 Rickenbacker Causeway, Miami, FL 33149-1098, USA
i WLI/Delft Hydraulics, P.O. Box 177, 2600 MH Delft, The Netherlands
j Department of Oceanography, Mail Code: OCHHe, US Naval Postgraduate School, Monterey, CA 93943, USA
k Naval Research Laboratory, Washington, DC 20375, USA
l ECMWF, Shinfield Park, Reading RG2 9AX, UK
m Delft University of Technology, Fluid Mechanics Section, Faculty of Civil Engineering and Geosciences, P.O. Box 5048, 2600 GA Delft, The Netherlands
n SHOM, 13 rue de Chapellier, 29609 Brest Cedex, France
o Hydraulics Laboratory, Katholieke Universiteit Leuven, Leuven, Kasteelpark Arenberg 40, B-3001 Heverlee, Belgium
p Università di Torino, Dipartimento di Fisica Generale, Via P. Giuria 1, 10125 Torino, Italy
q A.M. Obukhov Institute for Physics of Atmosphere, Pyzhevskii Per. 3, Moscow 119017, Russia
r Coastal and Hydraulics Laboratory, US Army Engineer Research and Development Center, Halls Ferry Road, Vicksburg, MS 39180, USA
s Naval Research Laboratory, Code 7322, Stennis Space Center, MS 39529, USA
t University of Florida, Civil and Coastal Engineering, 757 Weil Hall, Gainesville, FL 32611-6580, USA
u SAIC/OSO at NOAA/NCER, Camp Springs, MD 20746, USA
v Proudman Oceanographic Laboratory, Joseph Proudman Building, 6 Brownlow Street, Liverpool L3 5DA, UK

Received 14 July 2006; received in revised form 16 February 2007; accepted 23 May 2007
Available online 27 September 2007

© Authors of the single sections: For each section, from 2 to 8, the list of contributing authors is given, in alphabetic order but the first one, the latter (e-mail address) to be considered as reference person for the specific subject.
* Corresponding author. Tel.: +39 041 2404746; fax: +39 041 5204126.
E-mail address: luigi.cavaleri@ismar.cnr.it (L. Cavaleri).
Deceased author.

0079-6611/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.pocean.2007.05.005
Abstract

This paper is the product of the wave modelling community and it tries to make a picture of the present situation in this branch of science, exploring the previous and the most recent results and looking ahead towards the solution of the problems we presently face. Both theory and applications are considered.

The many faces of the subject imply separate discussions. This is reflected into the single sections, seven of them, each dealing with a specific topic, the whole providing a broad and solid overview of the present state of the art. After an introduction framing the problem and the approach we followed, we deal in sequence with the following subjects: (Section) 2, generation by wind; 3, nonlinear interactions in deep water; 4, white-capping dissipation; 5, nonlinear interactions in shallow water; 6, dissipation at the sea bottom; 7, wave propagation; 8, numerics. The two final sections, 9 and 10, summarize the present situation from a general point of view and try to look at the future developments.

© 2007 Elsevier Ltd. All rights reserved.

Keywords: Wind waves; Wind–wave generation; Wave–wave interaction; Wave propagation; Wave dissipation; Wave–current interaction; Numerics

Contents

1. Introduction ............................................................. 605
2. Brief review of wind–wave generation ........................................ 607
  2.1. Linear theory ......................................................... 607
  2.2. Nonlinear effects ...................................................... 611
  2.3. Gustiness .............................................................. 612
  2.4. Open issues .......................................................... 613
    2.4.1. Damping of low-frequency swells .......................... 613
    2.4.2. Momentum transfer for high wind speeds ............... 614
    2.4.3. Quality of modelled wind fields .......................... 614
3. Modelling nonlinear four-wave interactions in discrete spectral wave models .................................................. 615
  3.1. Theory ............................................................... 615
  3.2. Solution methods ...................................................... 617
  3.3. Properties ............................................................ 617
  3.4. Development in computational methods .............................. 619
  3.5. Inter-comparison of computational methods ........................... 621
  3.6. Questions and actions ................................................ 621
4. Spectral dissipation in deep water ........................................ 622
  4.1. Theoretical and experimental research of physics of the spectral dissipation ........................................... 623
    4.1.1. Spectral dissipation due to wave breaking .................. 623
    4.1.2. Wave–turbulence interactions .................................. 628
    4.1.3. Wave–wave modulations ....................................... 629
  4.2. Modelling the spectral dissipation function ............................. 630
5. Nonlinear interactions in shallow water waves .................................. 632
  5.1. Nonlinearity in shallow water ........................................ 632
  5.2. Deterministic models: time-domain and spectral-domain ................ 633
  5.3. Stochastic models .................................................. 634
  5.4. Dissipation and wave breaking in shallow water .................. 635
  5.5. Open problems ..................................................... 636
6. Bottom dissipation ..................................................... 636
  6.1. Wave energy dissipation due to bottom friction ...................... 637
    6.1.1. Common formulations for spectral wave models: waves only .................. 637
    6.1.2. Common formulations for spectral wave models: waves and currents ........ 638
    6.1.3. Bottom roughness models for movable beds ................... 638
  6.2. Energy dissipation due to wave–bottom interaction .................. 639
  6.3. Discussion and outstanding problems ................................... 639
1. Introduction

Along the progressive development of the art of wave modelling we have witnessed periods of great advances usually followed by periods of consolidation, when the focus of activity was mainly on the application of the newly developed tools and artifices. Wave modelling is a great art containing two aspects of human knowledge: theory, often touching basic principles from more fundamental sciences, and practical applications. Our ever increasing interaction with the sea has offered endless opportunities to apply to the everyday problems what the theory had just revealed. Granted a certain degree of maturity has been reached, advances are often rapid at the beginning of a science. With a bit of low pass filtering, we can easily recognise in the last 60 years the periods when more fundamental advances in wave modelling have taken place, followed by
periods of application and a proliferation of small scale improvements. Unavoidably, the rate with which we advance tends to decrease. The basic pieces of information, at least within the present perspective, have been brought to light, and we are much closer to providing satisfactory results on a large scale. Somehow, the wave modelling community is asking to itself if and when new basic pieces of knowledge will appear. The alternative would be to carry on with technological and engineeristic improvements, edging our way towards more satisfactory results.

At this stage we feel the need to understand better where we are, and to get a better perspective of the evolution of the problem and of the state of the art of the science we deal with. In this paper, we make a picture of the present situation, when necessary with some historical perspective, and we try to give indications, in some cases hints, of where wave modelling should or it is expected to go in the future. Following a common conceptual model, we have split the discussion into separate subjects. To a good degree of approximation this corresponds to how the problem is presently formulated in its basic equations and physical description. About this point a more extensive comment will be given in the final discussion. We consider progressively the following subjects.

Input by wind is the essential process without which wind waves would not exist. Witnessed by man since the early ages, this elusive process has defeated for a long while human intuition. The theoretical and practical difficulties cannot be overestimated.

Nonlinear interactions are probably the most solid piece of information in wave modelling. Inspired by fundamental physics, and brought to light more than 40 years ago, it is well defined. The problem is practical, in that the necessary computer time for its proper evaluation is not yet available.

White-capping, or dissipation in deep water, is the third basic physical process that governs the evolution of wind waves in the open oceans. It is the least understood part of wave evolution, and, combining some intuition with a pragmatic approach, it has been for a while, and still is, the tuning knob of any wave model.

Once in shallow water, nonlinear interactions become a more active subject of theoretical research. This section provides a summary of the recent advances, with substantial expectations for practical applications.

Bottom dissipation represents the interaction and energy sink of wind waves with/at the sea bottom. It summarises a number of different processes. Although bottom friction represents the most commonly used term, the relevance of each process depends on the local characteristics of the sea floor.

Wave propagation in non-homogeneous media, and in particular wave–current interactions, are the first link between these two more evident characteristics of the sea. The related interests and practical improvements have gone in one with the available knowledge of the distribution of currents at the coasts and in the open oceans.

Finally, numerics represents the practical description and application of the above processes. The discrete description of the sea we use in wave modelling leads to a number of problems whose solution we try to optimise.

Each of the above subjects may, and often does, represent the focus of activity of the single modeller. Hence each section has been written by a different group of persons, with their own style. Although we have applied a minimum of homogenization, there are obvious differences in the way each section is dealt with. In a way, this reflects the multi-dimensional approach to the problem. Granted the constant flow of information to the whole community, each person or subgroup contributes autonomously with his/their own initiative. The joining force of our group is the common interest in waves and the wish to improve our results with a permanent exchange of information.

It has been suggested that a more unified and controlled approach would be more effective. Apart from the obvious financial and institutional difficulties, this could be true in the short term, for a specific problem. With a wider perspective and in the long term, we need the wild horse that comes out with unconventional ideas, one of which may become the seed for further advancements. As human beings, we are far from being a perfect organization, but we are joined by our common desire to understand the essence and beauty of nature.

The paper is organised in the logical sequence outlined above. Sections from 2–8 deal progressively with wind input, nonlinear interactions, white-capping, nonlinear interactions in shallow water, interactions with the sea bottom, motion in non-homogeneous media and wave–current interactions, and numerics. In Section 9 we summarise the situation, pointing out the well established results and, more interestingly from the
scientific point of view, the problems we are still left with. Finally in Section 10 we discuss the challenges and the openings we expect for the future.

Although not up to the level of a paper, each section is self-standing, and it can be easily read autonomously. However, a progressive reading of the various sections will made clearer both the difficulties of the overall problem and how far we have been able to go.

The paper is authored by the whole Group, as we consider any advancement as a collective achievement. The continuous interactions and exchange of information are an essential part of our activity. However, each single section has been written by a definite sub-group, whose components, headed by the reference responsible person, are listed at the beginning of each section.

One final point. Our yearly WISE meeting, where the latest results are regularly presented, is oral only, without any proceedings (this makes the presentations more dynamical and up-to-date). Although usually later published in journals, here and there in this paper we refer to some specific presentations. There is no official reference for these, but the interested reader can refer either to the cited author or to the section coordinator for further details.

2. Brief review of wind–wave generation

Contributing authors: Peter A.E.M. Janssen (peter.janssen@ecmwf.int), Luigi Cavaleri, Donald Resio, Hendrik L. Tolman.

The problem of the growth of ocean waves by wind and the consequent feedback of the ocean waves on the wind has led to quite some controversy and many debates in the literature. Nevertheless, the combination of observations from field campaigns in the 1970’s and the theoretical work on the critical layer mechanism which started in the 1950’s has resulted in parameterizations of the wind-input source function that provide good results in operational wave models. Together with a realistic representation of the high-wave number part of the wave spectrum, these parameterizations of wind-input have the potential to yield realistic estimates of the air–sea momentum transfer. The mutual interaction of ocean waves and the atmosphere has resulted in improved forecast skill for wind and ocean wave height, in particular in documented cases at ECMWF.

In this section, after reviewing the present state of the art of our knowledge on the wind-input source function and the feedback of ocean waves on the wind, we discuss a number of open issues which may need to be addressed in the near future. These concern the problem of high-frequency variability in atmospheric models and the modelling of the extreme cases of large winds and low winds. In particular, it is becoming increasingly clear that the drag coefficient may not be well specified in extreme situations such as hurricanes.

2.1. Linear theory

Understanding the growth of water waves by wind is a very challenging task. On the one hand, from the theoretical point of view it should be realized that one deals with a difficult problem because it involves the modelling of a turbulent airflow over a surface that varies in space and time. On the other hand, from an experimental point of view it should be pointed out that it is not an easy task to measure growth rates of waves by wind in a direct manner. Nevertheless, considerable progress has been made over the past 40 years. The history of the subject of wind–wave generation started in the beginning of the 20th century when Jeffreys (1924, 1925) assumed that air flowing over the ocean surface was sheltered by the waves on their lee side. This would give a pressure difference, so that work could be done by the wind. Subsequent laboratory measurements on solid waves showed that the pressure difference was much too small to account for the observed growth rates. As a consequence, the sheltering hypothesis was abandoned, and one’s everyday experience of the amplification of water waves by wind remained poorly understood. This changed in the mid-1950’s, when Phillips (1957) and Miles (1957) published their contributions to the theory of wave generation by wind. Both theories had in common that waves were generated by a resonance phenomenon: Phillips considered the resonant forcing of surface waves by turbulent pressure fluctuations, while Miles considered the resonant interaction between the wave-induced pressure fluctuations and the free surface waves. Miles’ mechanism looked more promising, because it implied exponential growth, and it is of the order of the density ratio of air and water.
However, there was also a considerable confusion and controversy. One of the main reasons for the controversy was that Miles’ theory oversimplified the problem by following the quasi-laminar approach. This approach assumes that the airflow is inviscid and that air turbulence does not play a role except in maintaining the shear flow. Another reason is that Miles neglected nonlinear effects such as wave-mean flow interaction, which are expected to be important at the height where the wind speed matches the phase speed of the surface waves wind speed (the so-called critical height). Also, early field experiments, in particular by Dobson (1971), gave rates of energy transfer from wind to waves that were an order of magnitude larger than predicted by Miles (1957). More recent field experiments (Snyder, 1974; Snyder et al., 1981; Hasselmann and Bosenberg, 1991) show order of magnitude agreement with Miles’ theory, although the theory still predicts energy transfer rates that are smaller than the measured values, especially for relatively low-frequency waves with a phase speed that is close to the wind speed at 10 m height.

There have been several attempts to overcome these shortcomings by means of numerical modelling of the turbulent boundary layer flow over a moving water surface. With suitable turbulence closure assumptions the interaction of the wave-induced flow with the mean flow and the boundary-layer turbulence can then be simulated explicitly. One such approach (see, for example, Gent and Taylor, 1976; Makin and Chalikov, 1979; Riley et al., 1982; Al-Zanaidi and Hui, 1984; Jacobs, 1987; Chalikov and Makin, 1991; Chalikov and Belevich, 1993) considers the direct effects of small scale turbulence on wave growth. Mixing length modelling or turbulent energy closure is then assumed to calculate the turbulent Reynolds stresses. The resulting diffusion of momentum is then so large that essentially Miles’ critical mechanism becomes ineffective. In addition, in adverse winds or when waves are propagating faster than the wind speed these theories give a considerable wave damping, while in Miles’ theory damping is absent. There are, however, no convincing field observations of wave damping (Snyder et al., 1981; Hasselmann and Bosenberg, 1991), presumably because the actual damping time scales are quite long.

The above turbulence models rely on the analogy with molecular processes. Van Duin and Janssen (1992) pointed out that this approach fails for low-frequency waves. Mixing length modelling assumes that the momentum transport caused by turbulence is the fastest process in the fluid. This is not justified for low-frequency waves which interact with large eddies whose eddy-turnover time may become larger than the period of the waves. In other words, during a wave period there is not sufficient time for the eddies to transport momentum. For these large eddies (which are identified here with gustiness) another approach is needed. Nikolayeva and Tsimring (1986) considered the effect of gustiness on wave growth, and a considerable enhancement of energy transfer was found, especially for long waves with a phase speed comparable to the wind speed at 10 m height.

Belcher and Hunt (1993) have pointed out that mixing length modelling is even inadequate for slowly propagating waves. They argue that far away from the water surface turbulence is slow with respect to the waves so that again large eddies do not have sufficient time to transport momentum. This results then in a severe truncation of the mixing length in the so-called outer layer of the flow. In fact, the greater part of the flow may now be regarded as approximately inviscid and the energy transfer from wind to slow waves only occurs in a thin layer above the surface. Note that the main mechanism for wave growth in the Belcher and Hunt model is the so-called non-separated sheltering: the Reynolds stresses close to the surface cause a thickening of the boundary layer on the leeside of the waves which would result in flow separation when the slope is large enough. This mechanism is akin to Jeffreys’ sheltering hypothesis, which was originally developed for separated flows over moving waves of large slope. The approach of Belcher and Hunt has been further developed by Mastenbroek (1996) in the context of a second-order closure model for air turbulence, confirming the ideas of rapid distortion.

In short, the developments over the past 40 years may be summarized as follows. Miles’ quasi-laminar theory was the first model to give a plausible explanation of the growth of waves by wind. Because of the neglect of turbulence on the wave-induced motion the quasi-laminar model has been criticized as being unrealistic, therefore questioning the relevance of the critical layer mechanism for wind–wave growth. First attempts to describe the effects of turbulence by means of a mixing length model have been criticized as well, however, mainly because the eddies in the outer layer in the air are too slow to transfer a significant amount of momentum on the time scale of the wave motion. But, according to rapid distortion models such as the one of Belcher...
and Hunt (1993) or Mastenbroek (1996), the critical layer mechanism is only relevant for very fast moving ocean waves with a dimensionless phase speed, defined as \( c/u_* \), of the order of 30.

Recently there is evidence that even the rapid distortion approach of Belcher and Hunt overestimates the effects of eddies on the wave-induced flow. Sullivan et al., 2000 studied the growth of waves by wind in the context of an eddy-resolving numerical model. Although the Reynolds number was, compared to nature, too small by an order of magnitude, clear evidence for the existence of a critical layer was found for a wide range of dimensionless phase speeds. As expected from the Miles mechanism, a rapid fall-off of the wave-induced stress was seen at the critical height. Furthermore, nowadays, there is even direct evidence of the existence and relevance of the critical layer mechanism from in-situ observations (Hristov et al., 2003) obtained from FLIP (a FLoating Instrument Platform created by two Scripps scientists some 40 years ago). This is quite a challenge because one has to extract a relatively small wave-coherent signal from a noisy signal. Nevertheless, for the range \( 16 < c/u_* < 40 \), Hristov et al. (2003) did see a pronounced cat’s-eye pattern around the critical height where the wave-induced stress showed a jump. As shown in Fig. 1 there is a good agreement between observed and wave-induced profiles as obtained from the critical layer solution. Note that there is no observational evidence of a critical layer for dimensionless phase speeds less than 16. These conditions can only be observed by means of a wave follower when measurements are taken close enough to the ocean surface, in between the ocean waves.

A reason for the overestimation of the effect of eddies on the wave-induced motion has been discussed in Janssen (2004). Following the rapid-distortion ideas of Belcher and Hunt it is argued that the large eddies are

![Fig. 1. Real and imaginary parts of horizontal and vertical components of the wave-induced velocity as function of phase speed. Full lines: solution of Rayleigh equation; open squares: observations from Hristov et al. (2003).](image-url)
too slow to transport a significant amount of momentum during one wave period. The outer layer is approximately inviscid and only in a ‘thin’ layer above the surface mixing length modelling applies [the so-called ‘inner’ layer].

An appropriate wave time scale is $T_A = 1 / (k U_0(z) - c)$, while Belcher and Hunt take as turbulent time scale: $T_L = \kappa z / u_*$ (with $k$ the wave number and $\kappa$ the von Karman constant). The thickness $z_t$ of inner, turbulent layer then follows from equating the two time scales, $T_A = T_L$ and the mixing length is truncated to the value $k z_t$ (truncated mixing length model).

However, momentum transfer by eddies occurs on a time scale that is larger than the eddy-turnover time. Indications for this follow from observations of flow over a hill (Walmsley and Taylor, 1996), which gives a much thinner layer, and from estimation of the time scale from

$$\tau_U U_0 \propto u_0 \frac{z}{u_*} \left( \frac{u_*}{U_0(z)} \right)^2; \quad (2.1)$$

where for a logarithmic profile $\frac{d U_0}{dz} \approx u_* / k z$ (and not $U_0(z)$).

This estimate gives, compared to the Belcher and Hunt approach, the much longer time scale $T_M = \kappa z / \varepsilon(z) u_*$, since $\varepsilon(z) = u_0 / U_0(z)$ is a small parameter. The time scale $T_M$ gives rise to a much thinner inner layer. The resulting eddy viscosities are so small that the corresponding turbulent momentum transport can be neglected in lowest order. As a consequence, applying the truncated mixing length model with turbulent time scale $T_M$ one rediscovers in lowest significant order Miles critical layer result while in next order the turbulent momentum transport will give small corrections to the growth rate of the surface gravity waves. In particular, the long waves will have a weak damping.

The resulting growth rate becomes the sum of Miles’ critical layer effect and a (small) damping term caused by the inner layer viscosity:

$$\frac{\gamma}{\omega} = s \beta (u_*/c)^2, \quad \beta = \frac{\pi}{\kappa z} \mu \log^4 \left( \frac{\mu}{\kappa z} \right) - 2 \log \left( \frac{\mu}{\kappa z} \right), \quad (2.2)$$

where $s$ is the air–water density ratio, and $\varepsilon = e^{-T/2} = 0.281$ (with $\Gamma$ Euler’s constant). The value of $\varepsilon$ follows from higher order matching (Miles, 1993) and this sets the scaling velocity $V = U_0(z = \kappa/k) \equiv U_0(z = 0.045 \kappa)$. The parameter $\beta$ is plotted as function of the dimensionless phase speed $c/u_*$ in Fig 2.

The analytical form for the critical layer term was checked against the numerical solution of Rayleigh’s equation and with the present choice of $\varepsilon$ the agreement is fair for short waves. For long waves the analytical formula, however, seriously underestimates the numerically obtained growth rate. The observations compiled by Plant (1982) gives for short waves an average value of $\beta$ of about 30, hence the short wave limit of Eq. (4.2) is in fair agreement with observed values of wave growth.

![Fig. 2. Miles parameter $\beta$ versus dimensionless phase speed. Note that the resulting damping rate is very small for waves propagating faster than the wind. For $c/u_* = 50$, $u_* = .2$ spatial damping scale is already 2500 km.](image-url)
2.2. Nonlinear effects

For a given wind profile quasi-laminar theory is fairly successful in predicting growth rates and wave-induced profiles. It ignores, however, a possible change of wind profile while the ocean waves are evolving. The momentum transfer from wind to waves may be so large that the associated wave-induced stress becomes a substantial fraction of the turbulent stress (Snyder, 1974; Snyder et al., 1981). The velocity profile over sea waves is controlled by both turbulent and wave-induced momentum flux. Therefore, deviations from the profile of turbulent airflow over a flat plate are to be expected. In addition, the energy transfer from the air to the waves may be affected by the sea state, so that one expects a strong coupling between the turbulent boundary layer and the surface waves.

Observations confirm this expectation. Measurements by, for example, Donelan (1982), Smith et al. (1992), Drennan et al. (1999a) and Oost et al. (2002) indicate that the drag coefficient depends on the sea state through the wave age. The theory of the interaction of wind and waves was elaborated by Fabrikant (1976) and Janssen (1982). The so-called quasi-linear theory of wind–wave generation keeps track of the slow evolution of the sea state and its effects on the wind profile. At each particular time the wave growth follows from Miles’ theory. It turns out that quasi-linear theory permits an explanation of the observed dependence of the airflow on the sea state. The resulting parameterization of the roughness length in terms of the wave-induced stress shows a fair agreement with observed roughness (Janssen, 1992). Incorporating a wave prediction model in a weather forecasting system, it is possible to determine every time step how much momentum the air flow is transferring to the ocean waves. Extensive research at ECMWF has shown that the sea-state dependent momentum transfer has resulted in improved forecast skill for both wind and waves (Janssen, 2004).

Despite the relative success of quasi-linear theory it still cannot be claimed that the problem of wind–wave generation and the feedback of ocean waves on the wind is well-understood. For example, because the short waves are the fastest growing waves, the wave-induced stress is to a large extent determined by the spectrum of the high-frequency waves (see, e.g. Janssen, 1989; Makin et al., 1995). There is presently hardly any evidence of the wave age dependence of the short wave spectral levels. However, using a wavelet analysis Donelan et al. (1999) did find that the wavenumber spectrum of the short waves depends in a sensitive manner on wave age: ‘young’ windsea shows much steeper short waves than ‘old’ windsea. Nevertheless, the physics behind the wave age dependence of the spectrum is not well-understood presently. Four-wave interactions could play an important role in this issue because the negative lobe of the nonlinear transfer transports energy from the wavenumber region above the peak of the spectrum towards the longer waves beyond the peak of the spectrum. But this probably will not explain the wave age dependence of the spectral levels of the really short waves. On the other hand, it is well known that the dispersion relation of the short waves is affected by the orbital motion of the long waves and/or the Stokes drift. Such a surface drift may have a considerable impact on the spectral levels of the short waves (see, for example, Janssen, 2004), giving an alternative explanation of its sea-state dependence.

Furthermore, the quasi-linear approach assumes that the short waves are linear, but most likely those waves are fairly steep. Therefore, the nonlinear process of airflow separation, similar to what Jeffreys (1924, 1925) envisaged, may play a role in air–sea momentum transfer. According to Makin and Kudryavtsev (2002) this could provide an alternative explanation of the sea-state dependence of the drag over sea waves. However, this explanation requires that a considerable part of the drag is determined by airflow separation over dominant waves, but it is very unlikely that these large waves are breaking frequently. Even in the absence of flow separation, there may be concern about the basic hypothesis of linearity in generation by wind. Miles’ (1957) theory was derived for unidirectional, monochromatic waves. It has been assumed that the wind–wave interactions are sufficiently linear that the wind input to each spectral component can be considered independently. This topic was investigated by Tsimring (1983) who studied the interaction of two waves and the mean air-flow, which is basically the most simple case of a wave group. The resulting wave growth to one spectral component now depends on the presence of other components. Numerically, the effect is small, however, as it is proportional to the air–sea density ratio time square of the wave spectrum.

Finally, what about evidence in the field for the sea state dependence of the drag coefficient?

It is customary to try to relate the Charnock parameter to a measure of the stage of development of windsea, e.g. the wave age $c_p/u_*$, with $c_p$ the phase velocity of the peak of the spectrum. Here, the Charnock parameter is estimated from observations of $u_*$ and the windspeed at 10 m height, $U_{10}$, through the Charnock...
relation and the logarithmic surface wind profile. As a consequence, the Charnock parameter depends in an exponential manner on the drag coefficient at 10 m height, $C_D(10)$, and is therefore very sensitive to errors in the observations for friction velocity and windspeed. In addition, at a particular measurement site the range of phase velocities is usually limited compared to the range of friction velocities and as a result, based on observations from one measurement site, an empirically obtained relation between the Charnock parameter and the wave age may be spurious because it is in essence a relation between Charnock parameter and the friction velocity. A way to avoid the problem of self-correlation is to combine observations from a number of measurement campaigns so that the range of phase speeds becomes larger (Johnson et al., 1998; Lange et al., 2004). This approach was followed by Hwang (2005). In addition, rather then obtaining a parameterization for the Charnock parameter, which is prone to errors in observed friction velocity, Hwang sought a relation between the drag coefficient and the wave age. The usual reference height for the drag coefficient is 10 m, but Hwang argued that from the wave dynamics point of view (see also Eq. (2.2)) a more meaningful reference height should be proportional to the wavelength $k_p$ of the peak of the wave spectrum. Using wavelength scaling Hwang (2005) found

$$C_D(k_p/2) = A (c_p/u_*)^a$$  

with $A = 1.220 \times 10^{-2}$ and $a = -0.704$, reflecting the notion that the airflow over young windsea is rougher than over old windsea. As shown in Fig. 3 the ECMWF version of the WAM model, the physics of which was developed in the 1980’s, gives, compared to Hwang’s parameterization (2.3), a realistic representation of the drag coefficient at half the wave length. Therefore, for windsea it is possible to find a convincing parameterization of the sea state dependence of the surface stress. The drag coefficient and dynamic roughness under mixed-sea conditions remain difficult to parameterize at this stage.

2.3. Gustiness

In the previous sub-sections the relevance of air turbulence has been discussed as related to the physics of interaction between wind and a wavy surface. Once this physics has been translated into formulas for practical applications in wave modelling, wind is considered constant during each time step and at each grid point of the numerical integration procedure. However, there is wind variability with a time scale longer than wind generated waves, but still below the synoptic scale resolved by the meteorological models, that may have a substantial effect on wave growth.

It is common to assume that the energy transfer from wind to waves is a function of the difference between the nominal wind speed $U$ and the phase speed $c$ of the wave component of interest. If this dependence would

![Fig. 3. Comparison of simulated and parameterized relation of drag coefficient $C_D(\lambda/2)$ versus wave age $c_p/u_*$. Black line: simulation, open circle: Eq. (2.3), and dashed line the case of constant Charnock parameter ($a = 0.01$).](image)
be a linear function then an oscillation of $U$ with respect to its mean value $U_m$ would have on average no effect. However, as is evident from Fig. 2, wave growth depends in a nonlinear manner on $U - c$, in particular when the phase speed is close to the value of $U_m$. For $c > U_m$ there is practically no interaction between the wind and the waves, hence wave growth depends in an almost discontinuous manner on $U - c$. Consider now a wave with phase speed close to $U_m$, which is the case when wind sea is well-developed. For these long waves a positive fluctuation in wind speed will result in enhanced wave growth but a negative fluctuation will not give rise to reduced growth. The growing waves act as a rectifier (Abdalla and Cavaleri, 2002, call it the ‘diode’ effect) and therefore gustiness may have a considerable impact on wave growth. The implications are that, when waves reach a mature stage, they keep growing, although at a progressively reduced rate, well above the limit of a fully developed sea obtained in steady wind conditions. How much the gain in wave height, denoted by $\Delta H_s/H_s$, is depends on the variability $\sigma$ of the wind field (percent r.m.s. deviation from $U_m$). With $\sigma = 10\%$ there is only a small increase of $H_s$. However, this grows rapidly with $\sigma$, and in very unstable conditions, with $\sigma = 30\%$, $\Delta H_s/H_s$ may reach values as large as 0.3.

Apart from the fluctuation level, the gain in wave height also depends on the correlation time scale of the fluctuating wind. If the wind gustiness has a correlation time scale that is shorter than or similar to the integration time step (similar considerations apply in space), then the growth curve for wave height will be smooth. However, if the time scale is longer, then the growth curve will reflect this variability, giving large oscillations around the mean growth curve. This implies that the significant wave height can achieve values larger than expected even from the gusty growth.

In practical applications the diode effect can be taken into account following a procedure described by Janssen (2004), who followed Miles (1997). However, the $H_s$ oscillations due to the coherence in wind variability are not deterministic and are presently not considered in operational models. The same remark applies to the correlated part of the oscillations of the wind speed. This introduces a certain level of randomness in the comparison between observed and modelled $H_s$ values. Together with the common lack of information on the level of gustiness in the input wind fields, this complicates the validation of wave prediction systems. While there are good theoretical and practical reasons to believe that the effect is indeed present, a full quantification of its actual relevance is still missing.

2.4. Open issues

Here, we briefly discuss a number of interesting future developments.

2.4.1. Damping of low-frequency swells

First, the problem of the interaction of low-frequency swells and the atmosphere. This process happens typically in the Tropics in areas of low wind speed, but it concerns also the extra-tropical areas. Swell is an almost permanent feature of the oceans. This is an interesting problem because surface gravity waves may transfer energy and momentum to the atmosphere. In those circumstances the usual Monin–Obukhov similarity theory is not valid (Drennan et al., 1999b). There is, however, some uncertainty regarding the damping rates of the low-frequency swells.

Observations in the field from Snyder et al. (1981) and Hasselmann and Bosenberg (1991) do not support the idea that there is a substantial wave damping for waves propagating faster than the wind. In the lab, however, Donelan (1990) did find evidence for wave damping according to the following empirical formula:

$$\frac{\gamma}{\omega} = sc \left( \frac{U_0(\lambda/2)}{c} - 1 \right) \left| \frac{U_0(\lambda/2)}{c} - 1 \right|$$

which parameterizes growth and damping in terms of the wind speed at height $\lambda/2$. Here $c_p$ equals 0.11 for opposing winds, and 0.28 for following winds. However, when applied to swell cases in the field the damping is far too large: for 15 s waves one finds spatial damping scales of the order of 75 km. These damping scales are so small that swells generated in the extra-tropical storms would never arrive in the Tropics. In one of the earlier versions of the Wavewatch wave prediction model damping rates comparable to Eq. (2.3) were used and the modelled tropical wave climatology seriously underestimated the observed climatology (Tolman et al., 2002). Consequently, damping rates were reduced by an order of magnitude. Thus, for wave damping in
the field there is no real guidance: spatial damping scales are expected to be large, of the order of a few 1000 km. Presumably, laboratory experiments are not representative for what is happening in the field. For example, in the laboratory there may be currents with considerable vertical shear while in the field the vertical shear is much less. Note that straightforward mixing length modelling supports the formulation of wave growth and damping of Eq. (2.3) (Al-Zanaidi and Hui, 1984), but rapid-distortion arguments suggest that such turbulence models overestimate the effects of momentum transport by the eddies. As a consequence, there results an overestimate of wave damping. In contrast, Eq. (2.2) is based on a truncated mixing length model and probably results in a more realistic estimate of wave damping in the field. However, it is emphasized that reliable observations of wave damping in the field are to be preferred.

2.4.2. Momentum transfer for high wind speeds

Another important issue is the understanding of air–sea momentum transfer under high wind speed conditions such as occur for typical hurricanes and typhoons. Not surprisingly, not many observations of wave growth and momentum transfer are available. The recent work by Powell et al. (2003) and Donelan et al. (2004) suggests however that in those extreme circumstances the drag decreases with wind speed or saturates. But, the understanding of the physics of such extreme events is only beginning. What is clear, however, is that because of the strong interaction and interplay between momentum, latent, sensible heat fluxes and spray, each transport process cannot be considered in isolation. In particular, in hurricanes spray production is expected to be an important process which may have some unexpected consequences for the momentum transfer. Following Makin (2005) one may regard spray as suspended particles. In the so-called suspension layer the heaviest particles remain, on average, closer to the surface so that the particle concentration should decrease monotonically with height. Hence, the spray droplets form a very stable boundary layer close to the surface, and such a stable layer may suppress the air turbulence near the ocean surface. In other words, spray production may, in extreme conditions, give rise to a reduction of the drag coefficient for increasing wind speed. Note that Andreas (2004) sketches a somewhat different picture of the impact of spray on the airflow. He argues that when spray droplets enter the airflow they will be accelerated. As a consequence, spray exerts a stress on the airflow which for wind speeds above 30–35 m/s becomes comparable to the interfacial stress. This would result in a sharp increase of the drag with wind speed. Hence, Andreas (2004) proposes that spray has a direct impact on the mean airflow, while Makin (2005) suggests that spray, while forming a stable layer, suppresses the turbulent fluctuations thus inhibiting momentum transfer to the surface. Evidently, more research is required to sort out this delicate issue.

There are other possibilities that could explain that for extreme conditions the drag coefficient is smaller than expected from a straightforward extrapolation of the familiar linear drag law (e.g. Smith, 1980a). Donelan et al. (2004) have suggested a fluid mechanical explanation: for strong winds flow separation may be present. Thus, the outer airflow, unable to follow the wave surface, does not “see” the troughs of the waves and skips from breaking crest to breaking crest. Thus in conditions of continuous breaking of the largest waves the aerodynamic roughness of the surface is limited giving a reduced drag. On the other hand, Andreas (2004) has proposed that when spray returns to the water, short waves will be extinguished. This will no doubt reduce the drag considerably as the short waves carry most of the wave-induced stress. Furthermore, it should be realized that in the most intense part of a hurricane the wind field is strongly curved, hence the effective fetch for wind–wave generation is short and the sea state is extremely young. For extremely young sea states the drag is also reduced quite considerably as explained in Komen et al. (1998).

2.4.3. Quality of modelled wind fields

During the past 10–15 years we have seen a substantial improvement in the quality of the surface wind speed as follows, for example, from the validation of the analysed ECMWF surface wind against Altimeter wind speed observations from ERS-2 (Janssen, 2004).

Despite these impressive improvements it should be pointed out that modelled fields lack a considerable amount of variability in the short scales. This lack of variability is most prominent in the upper layers of the model atmosphere, near the tropopause. Observations of the kinetic energy spectrum obtained from aircraft data (Nastrom and Gage, 1985) show that in the synoptic scales the spectrum shows a $k^{-3}$ power-law behaviour (corresponding to a potential enstrophy cascade) while in the mesoscales (less than about
their action density, where quadruplets are often described by a so-called figure of eight diagram, as illustrated in Fig. 4 for the deep water case.

The spectrum behaves as $k^{-5/3}$, consistent with an energy cascade to even smaller scales (Cho and Lindborg, 2001). Global atmospheric models typically miss the $k^{-5/3}$ power law, presumably because the interpolation in the (semi-Lagrangian) advection scheme acts as a smoother. Also near the surface there is a considerable lack of variability of modelled wind as follows from a comparison with kinetic energy spectra derived from QuikScat scatterometer winds. Because of this lack of variability in modelled surface winds, ECMWF introduced in April 2002 the average effects of gustiness on wave growth. This change had a beneficial impact on the wave height field, in particular its spatial and temporal variability. Note, however, that presently no theory of the atmospheric boundary layer can justify the level of wind variability measured in the field in certain conditions.

3. Modelling nonlinear four-wave interactions in discrete spectral wave models

Contributing authors: Gerbrant van Vledder (vledder@alkyon.nl), Michel Benoit, Igor V. Lavrenov, Miguel Onorato, Vladislav Polnikov, Donald Resio, Hendrik L. Tolman.

It is nowadays widely accepted that resonant weakly nonlinear interactions between sets of four waves play an important role in the evolution of the energy spectrum of free surface gravity waves propagating at the ocean’s surface. This role became clear as a result of the JONSWAP project (Hasselmann et al., 1973). It is described and discussed in e.g. Phillips (1981a), Resio and Perrie (1991), Young and van Vledder (1993), Banner and Young (1994) and Resio et al. (2001).

In this section, we summarize the state-of-the art in the understanding and modelling of nonlinear four-wave interactions. Despite considerable progress, many questions remain. These are summarized at the end of the section, together with suggestions for further research.

3.1. Theory

The basic equation describing these interactions is the Boltzmann integral proposed by Hasselmann (1962) and a couple of years later by Zakharov (1968) who derived it in a form known as the kinetic equation.

Hasselmann (1962, 1963a,b) developed the theoretical framework for nonlinear four wave interactions for homogeneous seas with a constant depth. He formulated an integral expression for the computation of these interactions, which is known as the Boltzmann integral for surface gravity waves.

Hasselmann (1962) found that a set of four waves, called a quadruplet, could exchange energy when the following resonance conditions are satisfied:

$$\vec{k}_1 + \vec{k}_2 = \vec{k}_3 + \vec{k}_4, \quad \omega_1 + \omega_2 = \omega_3 + \omega_4$$

(3.1)

in which $\omega_i$ is the angular frequency and $\vec{k}_i$ the wave number vector ($i = 1, \ldots, 4$). The linear dispersion relation relates the radian frequency $\omega$ and the wave number $k$:

$$\omega^2 = gk \tanh(kh)$$

(3.3)

which reduces to $\omega^2 = gk$ in deep water conditions.

Here, $g$ is the gravitational acceleration and $h$ the water depth. The configurations of interacting quadruplets are often described by a so-called figure of eight diagram, as illustrated in Fig. 4 for the deep water case.

Hasselmann (1962, 1963a,b) describes the nonlinear interactions between wave quadruplets in terms of their action density, where $n(\vec{k}) = E(\vec{k})/\omega$ and $E$ the energy density. The rate of change of action density at a wave number $\vec{k}_1$ due to all quadruplet interactions involving $\vec{k}_1$ is:

$$\frac{\partial n_1}{\partial t} = \int \int G(\vec{k}_1, \vec{k}_2, \vec{k}_3, \vec{k}_4) \times \delta(\vec{k}_1 + \vec{k}_2 - \vec{k}_3 - \vec{k}_4) \times \delta(\omega_1 + \omega_2 - \omega_3 - \omega_4)$$

$$\times [n_1 n_2 (n_3 + n_4) - (n_1 + n_2) n_3 n_4] \, dk_2 \, dk_3 \, dk_4,$$

(3.4)

where $n_i = n(\vec{k}_i)$ is the action density at wave number $\vec{k}_i$ and $G$ is the coupling coefficient. This integral is six-fold in wave number ordinates. The $\delta$-functions in (3.4) ensure that contributions to the integral only occur for
quadruplets satisfying the resonance conditions, and thus formally reduce this expression to a threelfold integral.

It is worth noting that these resonant interactions basically reflect weak nonlinear transfers in the evolution of the wave spectrum for the case of homogeneous conditions. Recent work by Janssen (2003) suggests that quasi-resonant four-wave interactions play a major role in uni-directional wave field, in relation to the development of modulational instabilities and the occurrence of freak waves. Yet unclear is the role of non-resonant interactions in two-dimensional cases.

In (3.4) the $\delta$-functions also ensure conservation of wave energy, wave action and wave momentum. The coupling coefficient is given by

$$G(\vec{k}_1, \vec{k}_2, \vec{k}_3, \vec{k}_4) = \frac{9\pi g^2 D^2(\vec{k}_1, \vec{k}_2, \vec{k}_3, \vec{k}_4)}{4\rho^2 \omega_1 \omega_2 \omega_3 \omega_4}. \quad (3.5)$$

In this expression $D(\vec{k}_1, \vec{k}_2, \vec{k}_3, \vec{k}_4)$ is the interaction coefficient, and $\rho$ is the density of water. The deep-water expression for the interaction coefficient was first given by Hasselmann (1962). Webb (1978) used an algebraic manipulator to simplify the mathematical structure of this coefficient. However, his expression contained some misprints. Corrected expressions are given in Dungey and Hui (1979). Herterich and Hasselmann (1980) derived a finite depth version of the interaction coefficient. Zakharov (1999) re-derived the coupling coefficients for deep and finite depth water, and expressed them in a form similar to those of Webb (1978). Gorman (2003) provides a detailed analysis of the finite depth interaction coefficient and he derived expressions for the treatment of discontinuities.

A remarkable property of (3.4) is that it possesses exact stationary isotropic analytical solutions of the form of power laws that correspond to a constant flux of energy towards high wave numbers and constant flux of wave action to small wave numbers. These solutions have been found by Zakharov and Filonenko (1966). The constant energy flux solution corresponds in the frequency wave spectrum to a power law of the form of $x^{3/4}$, in agreement with experimental observations starting from Toba (1972).

In paper (Lavrenov et al., 2002) a direct numerical simulation of the Hasselmann kinetic equation for gravity waves in water surface confirms basic predictions of the weak-turbulent theory. The kinetic equation for surface gravity waves is investigated numerically taking into account an external generating force and dissipation. An efficient numerical algorithm for simulating nonlinear energy transfer is used to solve the problem. Three stages of wave development are revealed: unstable wave energy growth within a range of external force impact, fast energy spectrum tail formation in high frequency range and establishment of a steady state spectrum. In both isotropic and non-isotropic cases the spectra are found out to be close to the Zakharov–Filonenko spectrum $\omega^{-4}$, in the universal range. Reliable estimations of the Kolmogorov constants are found out as $z_0 = 0.303 \pm 0.033$ in an isotropic case and as $z_1 = 0.239 \pm 0.023$ in a non-isotropic case. Formation of this asymptotic spectrum happens explosively. Accurate estimations of the first and second Kolmogorov
constants are obtained. A good agreement between the Toba experimental data and our results obtained with the help of direct numerical simulation is observed.

In recent numerical simulations of Eq. (3.4), Pushkarev et al. (2003) have shown that nonlinear interactions generate an $\omega^{-4}$ wave spectrum also in anisotropic conditions. Moreover, they have also shown the formation of the bimodal angular distribution of energy, in agreement with field and laboratory experiments.

It should here be mentioned that many properties of the kinetic equation (for example power laws solutions) are consistent with the fully nonlinear water wave equations. In this context, recently a number of direct numerical simulations of those deterministic equations have been performed in order to study the validity and the limitations of the approximations under the kinetic equation (3.4), see Tanaka (2001), Onorato et al. (2002), Dyachenko et al. (2004), Yokoyama (2004).

3.2. Solution methods

The full solution of the Boltzmann integral (3.4) is rather time consuming due to its complexity, in spite of numerical optimization efforts such as e.g. Snyder et al. (1993), Lin and Perrie (1998). It is therefore not yet applicable in operational wave prediction models. To overcome this disadvantage of exact methods, Hasselmann et al. (1985) developed the discrete interaction approximation (DIA). They show that the DIA preserves a few but important characteristics of the full solution, such as the slow downshifting of the peak frequency and shape stabilization during wave growth. The development of the DIA triggered the development of third generation (3G) wave prediction models, like the WAM model (WAMDI Group, 1988), WAVEWATCH (Tolman, 1991, 2002c), TOMAWAC (Benoit et al., 1997), the SWAN model (Booij et al., 1999), and the recently developed CREST model (Ardhuin et al., 2001).

The DIA was initially developed for deep water. The WAM Group (WAMDI Group, 1988) introduced a scaling technique to estimate the nonlinear transfer for an arbitrary water depth. This technique contains a parameterization of the magnitude scaling derived by Herterich and Hasselmann (1980). With this technique the finite-depth source term is simply obtained by multiplying the deep-water source term with a constant factor. This factor is a function of the relative water depth $\frac{k}{h}$, where $k$ is the mean wave number of the wave spectrum:

$$R(y) = 1 + \frac{5.5}{y} \left(1 - \frac{5}{6}y\right) \exp\left(-\frac{5}{4}y\right) \quad \text{with} \quad y = \frac{3}{4} kh. \quad (3.6)$$

This simple modification has however exhibited a number of shortcomings in shallow water conditions. As pointed out by Herterich and Hasselmann this approximation is only applicable for $k_p h \geq 1.0$, which is still relatively deep water for most coastal applications. Also, this approximation retains a stationary $\omega^{-4}$ form independent of depth; whereas observations and theory support the existence of a general $k^{-5/2}$ form in arbitrary-depth water (Resio et al., 2001, 2004). Due to these inherent problems, we recommend that methods be developed which take account of finite water depth effects in a more complete way. For instance Van Vledder (2001a) presents a shallow-water version of the DIA (referred to as the SDIA).

Janssen and Onorato (2007) present a detailed analysis of the application of the Zakharov equation in intermediate water depth. They find that for $kh \approx 1.363$ the magnitude of the nonlinear transfer rate becomes very small. This finding contradicts earlier findings like those of Herterich and Hasselmann (1980) and Resio et al. (2001). In addition Janssen and Onorato (2007) find that a simple shallow water scaling of the deep water nonlinear transfer rate is not correct since the shape of the transfer rate is also affected. As noted by Van Vledder (2006) this holds also for the full Boltzmann equation. A smaller nonlinear transfer rate in intermediate water depth has consequences for the downshifting of the peak frequency. Further studies to these consequences are needed.

3.3. Properties

A summary of the role of nonlinear four-wave interactions is given in Young and van Vledder (1993). The main features of nonlinear four-wave interactions are illustrated on a particular case in deep water from Benoit (2005). In this example we consider the directional wave spectrum corresponding to case 3 of
Hasselmann and Hasselmann (1981). This spectrum combines a JONSWAP frequency spectrum (with Phillips constant $\alpha = 0.01$, peak frequency $f_p = 0.3$ Hz, peak enhancement factor $\gamma = 3.3$) and a (frequency independent) angular spreading function of the form $\cos^4(\theta)$. The input spectrum is plotted in Fig. 5.

For this spectrum the nonlinear transfer term due to four-wave interaction is evaluated “exactly” by the WRT method (see below) and by the standard DIA approximation. The computed frequency-direction nonlinear terms $Q_{nl4}(f, \theta)$ are plotted in Fig. 6. The upper panel (exact evaluation with the WRT method) shows
the typical signature of four-wave interactions: first, there is a positive lobe below the peak frequency in the main wave direction, which corresponds to an increase of wave energy for these frequencies lower than the peak frequency. Then one can see a large negative lobe for the frequencies above the peak still in the main wave direction. In this region of the spectrum the nonlinear interactions pump energy. Finally there are also positive lobes for frequencies higher than the peak but about 45° off the main direction.

The lower panel of Fig. 6 shows that the DIA computation produces a term with some similarities in the general shape, but also significant differences. The first (positive) lobe is lower and shifted about 40° off the main direction. The second (negative) lobe is much higher than the exact one and it is shifted to higher frequencies. Finally the positive lobes at ±45° off the main direction are present, but at lower frequencies and they are clearly higher than the exact ones. The position and magnitude of the positive lobes result for the DIA in a trend to excessively spread the energy over directions, making the spectrum broader than it should be. The frequency nonlinear terms (after integration over wave directions) are plotted in Fig. 7. Again the differences between DIA and exact evaluation (EXACT-NL and WRT) are clear, in particular for the negative lobe, which is twice higher than the exact one and also shifted towards higher frequencies.

3.4. Development in computational methods

The development of the discrete interaction approximation partly resolved the limitations of an exact computation. However, as shown on the above example, experience reveals many deficiencies of the DIA, which hamper the further development of third-generation models. More specifically, deficiencies of the DIA are masked by tuning of the other source terms. Experience with an exact computational method in 1D and 2D applications shows improved prediction of spectral shapes. Thus, we face the dilemma of having a fast but inaccurate DIA and an accurate and time-consuming exact method. Therefore, a need exists for a computational method that would be both operationally feasible and accurate enough for application in operational 3G wave models.

Various attempts have been made to develop such methods. Progress has been made at four fronts.

First, extensions to the DIA have been proposed by adding more interacting wave number configurations. Van Vledder (2001b) describes the general framework for such extensions. Proposals for multiple DIA’s were made by Van Vledder et al. (2000), Hashimoto and Kawaguchi (2001) and more recently by Tolman (2004).
These attempts are promising, but not yet successful in the sense that extensions are generally applicable. The main reason is that each MDIA is developed for a specific set of test spectra.

It is noted that alternative DIA’s have been developed by Abdalla and Özhan (1993). Komatsu (1996) (referred to Hashimoto et al., 2003) developed the SRIAM, which is a multiple DIA based on the exact RIAM method. Polnikov and Farina (2002) proposed a version of the fast DIA, which doubles the speed of calculations without loss of accuracy. Moreover, in Polnikov (2003) it was found some other simple configurations which have lower errors than the original version of DIA; however, a major remaining problem is that these interactions represent only a small subset (in which 2 of the interacting wave number vectors are equal, rather than the more general case of 4 unequal wave number vectors) of the total interactions contributing to the complete integral. For this reason, the DIA will continue to require tuning for different classes of spectra.

The second line of development consists in starting from exact methods and making some simplifications and/or reductions of the integration space in the evaluation of the Boltzmann integral. Such methods can reduce the workload by a combination of smart integration techniques, coarser interpolation techniques and filtering out unimportant parts of the integration space. These methods differ in the way the delta-functions of the Boltzmann integral have been removed and the final set of equations obtained, and in the treatment of singularities. The following groups of ‘exact’ methods exist:

- **EXACT-NL** (Hasselmann and Hasselmann, 1981, 1985; Van Vledder and Weber, 1988; Van Vledder and Holthuijsen, 1993);
- **Webb** (1978) as implemented by Tracy and Resio (1982), Resio and Perrie (1991) and Van Vledder (2006), referred to as the WRT method. From this computational method Lin and Perrie (1998) developed the reduced interaction approximation (RIA);
- **Masuda** (1980) as extended to finite depth by Hashimoto et al. (1998) or adapted by Polnikov (1997), the RIAM method by Komatsu and Masuda (1996);
- **Lavrenov** (2001), the algorithm is based on a numerical integration method of high precision.

Each of these approaches solves the Boltzmann integral with some method. Differences exist in the transformations applied to the Boltzmann integral to remove the delta-functions, and in the numerical integration technique applied to them. At present it is not clear which of these methods produces the best results in terms of accuracy and computational requirements. Therefore, an objective inter-comparison between the various methods is needed to confirm or reject claims about their performance. Resio and Perrie (2006) introduced a two-scale hybrid method. In this method the spectrum is decomposed in a global part capturing the main spectral shape, for which the nonlinear transfer rate is computed with the WRT method, and a residual part for which an approximate correction term is computed. Preliminary results look promising but further verification is needed.

The third approach is based on neural networks. Tolman and Krasnopolsky (2004) present a method based on a neural network. It appears that the NN approach can result in stable wave growth in model integrations. However, much development work still needs to be done before this approach is suitable for general model applications (Tolman and Krasnopolsky, 2004). A problem that arises in this class of approximation is the difficulty in using any set of functions to linearly represent nonlinear interactions. The cubic dependence of the interactions on the energy/action densities typically produces very strong “cross-interactions” among the different “basis” functions.

The fourth approach comprises diffusion operators. Examples are those presented by Zakharov and Pushkarev (1999), Jenkins and Phillips (2001), and Pushkarev et al. (2004). They developed methods based on a diffusion operator. Some properties of this approximation were revealed in Polnikov (2002), where he found a reasonable correspondence of the diffusion approximation to the exact calculations of the integral. Properly posed, simulations based on this approach can be shown to conserve all constants of motion over time (Pushkarev et al., 2004) and can preserve the basic $\omega^{-4}$ characteristic form during evolution. However, this approximation does not provide a very accurate general approximation to the total integral and must be specifically tuned to fit each different spectral form. Although attractive for its computational simplicity, it is not flexible enough for application in a discrete spectral wave model, since a coefficient of proportionality needs to be determined for the class of spectra under consideration.
3.5. Inter-comparison of computational methods

Up till now no objective comparison has been performed to determine the best (in terms of performance and accuracy) method for computing the nonlinear four-wave interactions in a discrete spectral model. Still, a number of attempts have been made to intercompare different computational methods.

Lavrenov (2001, 2003a) made comparisons between his method and those of Hasselmann, Polnikov, Masuda’s, and Resio. Lavrenov claims that his method produces accurate results with relatively small computational requirements. A comparative study of different approximations for the Boltzmann integral was carried out in a series of papers (Polnikov and Farina, 2002; Polnikov, 2003). Using a certain definition of the error measure, it was shown that the DIA approximation is the best one among some other theoretical approximations: the diffusion approximation (Zakharov and Pushkarev, 1999) and the reduced integration approximation (Lin and Perrie, 1999). Inter-comparisons of results from some of these methods are presented in, e.g., Benoit (2005) for a few wave spectra in deep water.

These claims need further attention and an objective verification under controlled experiments. Most of the comparisons were made for a small set of academic spectra. In the case these spectra are smooth, some numerical integration technique might benefit from this smoothness. However, a good fit for parametric spectra is no guarantee that the computational method will work in an operational model since the interplay with other source terms and numerical procedures is equally important. Thus, the real test for any computational method is to implement it into a wave model and to perform fetch-limited or duration limited growth experiments. The resulting spectra will vary and will often be different from measured or theoretical spectra (cf. JONSWAP). In addition, an objective comparison is often hampered by differences in compilers and computer hardware.

3.6. Questions and actions

Since the derivation of the Boltzmann integral or kinetic equation great progress has been made in the understanding of the role of nonlinear four-wave interaction in wind–wave evolution. Despite considerable advances our knowledge about the application of the Boltzmann integral in numerical wave modelling is still incomplete. For instance, the range of validity of the Boltzmann integral is not precisely known. Also, the role of these interactions in determining the spectral shape in complex situations is not fully known. Of practical interest is the question about the best computational method. Each of these issues is discussed below.

The Boltzmann integral has originally been derived for deep water under the assumption of a homogeneous and stationary sea state and exact resonance between spectral components. Including shallow water effects assuming a flat bottom extended this concept. The validity of the Boltzmann integral can only be tested using numerical simulations of the nonlinear evolution of free surface gravity waves. Various investigations are underway and seem to confirm the validity of the Boltzmann integral for narrow spectra. However, more detailed numerical experiments are needed to assess the validity of the Boltzmann integral for other types of spectra and for shallow water conditions including sloping bottoms. In the modelling practice, the above-mentioned assumptions are often violated, but it is not known to which extent this affects the evolution of the wave field. An example of coping with these uncertainties is the way in which the DIA is applied in shallow water applications. The equations yield that the magnitude of the interactions increases as the water becomes shallower. To avoid a breakdown of the equations, a lower limit of 0.5 for the dimensionless water depth \( kh \) is applied. Recent theoretical developments indicate that also near-resonant interactions may exchange energy between sets of four wave components, both for 1D and 2D situations. There are also theoretical developments that indicate that the magnitude of the nonlinear transfer rate is much smaller for \( kh \approx 1.363 \) than previously thought. Further studies are needed to determine to which extent these findings affect wind–wave evolution in deep and shallow water.

Many field experiments, starting with the famous JONSWAP experiment, confirmed the role of four-wave interactions for fetch-limited wave growth and it shed light on the mechanism behind the downshifting of the spectral peak, shape stabilization and the frequency dependent directional distribution of the wave spectrum. However, in multi-peaked or directionally sheared spectra, the role of the four-wave interactions is far from simple. Examples of such conditions are slanting fetch situations, turning winds and mixed sea states where a wind sea develops on top of a background swell. Of interest is also the role of four-wave interactions in shal-
low water and its relative magnitude compared to other physical processes, such as three-wave interactions and depth-limited wave breaking. Although the magnitude of four-wave interactions is proportional to the frequency to the power 11, yielding that these interactions are very weak for long period swell waves, its role in the propagation of swell over very long distances has not yet been determined. Knowledge about the role of the four-wave interactions in all of the above situations is of importance for the development of optimal computational methods.

Various ‘exact’ numerical techniques have been developed to evaluate the Boltzmann integral to great accuracy. Theoretically these methods should give the same answer, but in practice each method makes some subtle choices in the numerical evaluation of the Boltzmann integral. It is therefore necessary to perform an objective comparison between these methods. In practice the various ‘exact’ methods are computationally demanding such that they cannot be used in operational applications. To overcome this hurdle, various approximations have been developed, of which the DIA is most commonly used. Despite its success in the development of third-generation wave models, the DIA suffers from many shortcomings and more accurate methods need to be developed. The main challenge is to develop computational methods that are sufficiently accurate as well as computationally feasible for inclusion in operational models. The development of such new approximations is not straightforward due to the complicated (nonlinear) nature of the Boltzmann integral. Possible ways to come up with an attractive computational method are to extend the DIA with more generally shaped configurations, neural network techniques, reduced exact integration methods or hybrid methods. It is very likely that each new approximation works best for a certain range of conditions. Therefore, a good understanding of the role of nonlinear interactions in various conditions is essential to make choices in developing an optimal computational method. In practice, the first test for a new approximate method is to compare the results with exact solutions for individual (academic) spectra. However, the real test is to apply new approximations in combination with other source terms in situations where the spectrum evolves, like fetch- or duration limited wave growth or slanting fetch situations. Comparison with ‘exact’ methods then provides a benchmark to assess the operational benefits of approximate methods. In this process the balance between the computational requirements and the need to achieve a certain accuracy in terms of model results should be optimized. This does not necessarily include a good prediction of the shape of the nonlinear transfer rate.

4. Spectral dissipation in deep water

Contributing authors: Alexander Babanin (ababanin@groupwise.swin.edu.au), José-Henrique G.M. Alves, Fabrice Ardhuin, Michael Banner, Mark Donelan, Paul Hwang, Vladislav Polnikov, Ian Young.

Spectral wave energy dissipation represents the least understood part of the physics relevant to wave modelling. There is a general consensus that the major part of this dissipation is supported by the wave breaking, but physics of this breaking process, particularly for the spectral waves, is poorly understood. How much energy is lost due to white-capping and where in the spectrum? What causes waves to break and what causes them to stop breaking? What does the breaking severity depends on? Recent field observations (Banner et al., 2000; Babanin et al., 2001; Banner et al., 2002) have found a threshold-like behaviour of breaking probabilities across the spectrum in terms of spectral steepness parameters, but these results are still to find their way into operational formulations that, today, are often just tuning knobs even in the simplest case of pure wind sea.

Dissipation due to interaction of waves with turbulence is arguably the second most important wave energy sink, certainly most persistent. Whether this is a background turbulence or turbulence generated by wave breaking, it is another source of dissipation that can account for an appreciable fraction of the wave energy loss (Drennan et al., 1997; Ardhuin and Jenkins, 2005, among others). This sink term, however, is still to find a consistent way of parameterization in wave models.

Many more other possible energy sink mechanisms can be formulated for the spectral wind–wave environment. For example, short wave modulation by long waves may also contribute to the dissipation of swell propagating against the wind, by a combination of Longuet–Higgins maser mechanism and Hasselmann’s theory for the exchange of potential energy between short and long waves. All these theories need a modern re-evaluation (e.g. Garrett and Smith, 1976; Ardhuin and Jenkins, 2005).
Also, our general description of dissipation completely ignores the interaction of waves with the vertical structure of the upper layers of the ocean. One step back on the mechanics of wave motion is probably necessary. Indeed, many mechanisms can be proposed for wave dissipation at this level. These include, for example, interaction with internal waves, which can be significant when the orbital motion due to waves is felt well below the thermocline. This leads to an increased mixing of the upper layers. In turn, the latter leads to the attenuation of swell and to consequences presently not considered in wave models, but a sound theoretical basis is available (e.g. Kudryavtsev, 1994).

While a loss for the wave system, whitecapping is a source of momentum and turbulent kinetic energy for the ocean currents or longer waves. Presently this two-step transfer is not considered, and modelled currents are driven directly by the atmospheric stresses. A properly defined body-force representing the momentum flux of waves to the mean flow combined with a surface flux of turbulent kinetic energy apparently leads to reasonable profiles of Eulerian currents, TKE dissipation, and eddy viscosity (e.g. Terray et al., 2000; Sullivan et al., 2004).

Theoretical and experimental knowledge of the spectral wave dissipation is so insufficient that, to fill the gap, spectral models have been used to guess the spectral dissipation function as a residual term of tuning the balance of better known source functions to fit known wave spectrum features. In this section, studies of the physics of the dissipation and the numerical simulations of the spectral dissipation are separated into different subsections. The two approaches target the same objective but should not be confused, as the simulations cannot prove or disprove the physics, and in fact may even disregard the physics and still be successful. Physics, on the other hand, is the ultimate truth. Discovered physical mechanisms certainly exist, but their relative importance with respect to the real waves and therefore their relevance for the models is often not clear.

4.1. Theoretical and experimental research of physics of the spectral dissipation

Physics of the spectral dissipation is an elusive subject, and theoretical and even experimental results in this area are few and often contradictory. Three dissipation sources are considered in this section: those due to wave breaking, wave–turbulence interaction and wave–wave modulation.

4.1.1. Spectral dissipation due to wave breaking

Theories of breaking dissipation, having started with the work of Longuet-Higgins (1969a), underwent some two decades of relatively extensive attention, but have enjoyed very little development in the past 10–15 years. This section is mainly dedicated to recent progresses in the field of wave dynamics, but a brief review of those older analytical theories of the spectral dissipation is necessary to understand where we currently stand. The review provided here uses extensively reviews of Donelan and Yuan (1994) and Young and Babanin (2006), but also accommodates most recent advances in the field.

It is generally assumed that \( S_{ds} \) is a function of the wave spectrum \( E \):

\[
S_{ds} \sim E^n
\]

(4.1)

but there is no agreement on whether the spectral dissipation \( S_{ds} \) is linear in terms of the spectrum \( E \) or not, i.e. whether \( n = 1 \) or \( n > 1 \). Donelan and Yuan (1994) classified theoretical models of the spectral dissipation into three types: whitecap models, quasi-saturated models and probability models. We would add a turbulent model class to this classification (Polnikov, 1993). None of these models, however, deals with the physics of wave breaking which governs the wave energy loss. This physics, to a major extent, is unknown, although relating wave breaking to nonlinear wave group modulations is providing encouraging new insight (Banner et al., 2000; Song and Banner, 2004). Present analytical models for \( S_{ds} \) try to employ either the wave state prior to the breaking or the residual wave and turbulence features after the breaking to derive conclusions on the dissipation due to the breaking.

Of the models which consider the waves prior to the breaking, the first analytical type developed was a probability model suggested by Longuet-Higgins (1969a) and further developed by Yuan et al. (1986) and Hua and Yuan (1992). All of these studies used the Gaussian distribution of surface elevations to predict the appearance of wave heights exceeding the height of the Stokes’ limiting wave or its limiting acceleration \( g/2 \) at the crest (\( g \) is the gravitational acceleration). Such waves were assumed to break until the wave height is
reduced back to a limiting value, and the difference was attributed to the dissipation. The limiting value used varied from the extreme Stokes value (Longuet-Higgins, 1969a; Yuan et al., 1986) to the mean value at a particular frequency derived from the Phillips (1958) equilibrium spectrum. The dissipation was found to be a linear function of the wave spectrum.

More recently, however, it has been shown that the waves do not necessarily have to reach the $g/2$ acceleration limit to break (Holthuijsen and Herbers, 1986; Hwang et al., 1989; Liu and Babanin, 2004). In addition, once they are breaking they do not stop at the Stokes limiting steepness but may keep losing energy until their steepness is well below the Stokes limit and even below the wave mean steepness (Liu and Babanin, 2004). Therefore, even though conceptually attractive, the probability models, as they have been derived, are not quantitatively plausible.

The second type of prior-to-breaking class of models is what Donelan and Yuan (1994) called quasi-saturated models (Phillips, 1985; Donelan and Pierson, 1987). These models rely on the equilibrium range of the wave spectrum, where some sort of saturation exists for the wave spectral density. In this region, the wind input, the wave–wave interactions and the dissipation are assumed to be in balance. Therefore, at each wave scale (wavenumber), any excessive energy contributed by combined wind input and nonlinear interaction fluxes, does not bring about spectral growth but wave breaking and can be interpreted as the spectral dissipation local in wavenumber space. Phillips (1985) found that such dissipation is cubic in terms of the spectral density.

Donelan and Pierson (1987) added consideration of wave directionality to the energy balance of the equilibrium range, arguing that a simple balance between wind input and dissipation is not observed at large angles to the wind. They also separated dispersive (gravity and capillary) waves and non-dispersive (gravity and capillary) waves as the nature of breaking differs for them because of different speeds of propagation relative to wave groups. Donelan and Pierson (1987) obtained a local-in-wavenumber-space dissipation function, similar to that of Phillips (1985) but their exponent $n$ depends on the wave spectrum $E$ and wavenumber $k$. According to them, $n$ can vary significantly: $n = 1–5$. It is essential, however, that $n \sim 5$ in most ranges of interest – both for gravitational and for capillary waves.

This model type has multiple shortcomings. Firstly, the very concept of the quasi-saturated or equilibrium interval is now subject to doubt (Donelan, WISE 2003). Even if it exists, the Phillips saturation level is not constant, but depends on environmental conditions (Babanin and Soloviev, 1998a). And even more importantly, none of the source terms which shape the spectral balance are known explicitly and accurately enough to provide a reliable determination of the dissipation as a residual sink term. Also, a dissipation function based on the breaking of short waves in the equilibrium interval does not account for dissipation due to dominant wave breaking near the spectral peak, which may be more severe and can be quite frequent (Babanin et al., 2001; Young and Babanin, 2006). Finally, there is growing evidence that dominant waves and the breaking of dominant waves affect dissipation at smaller scales (Banner et al., 1989; Meza et al., 2000; Donelan, 2001; Young and Babanin, 2006). If that is true, dissipation in the saturation interval will not be a function local in wavenumber space.

The most mathematically well-advanced and most frequently utilised dissipation model is that due to Hasselmann (1974). This is an after-breaking class model as it relies on the distribution of well-developed whitecaps situated on the forward faces of breaking waves. According to Hasselmann (1974), once there is an established random distribution of the whitecaps, it does not matter what caused the waves to break: the whitecaps on the forward slopes exert downward pressure on upward moving water and therefore conduct negative work on the wave. This model produces a linear dissipation.

Two main assumptions of the model are that the dissipation, even if it is strongly nonlinear locally, is weak in the mean and that the whitecaps and the underlying waves are in geometric similarity. Both assumptions are not always strictly accurate. For example, Babanin et al. (2001) investigated wave fields with over 10% dominant breaking rates, Young and Babanin (2006) examined a 60% dominant breaking case. It is not clear whether the weak-in-the-mean approach is still applicable in such circumstances, which are apparently a regular feature of wind seas.

The geometric similarity is also an approximation for real unsteady breakers. The whitecapping commences at some point on the incipient breaking crest and then spreads laterally and longitudinally (Phillips et al., 2001) and may or may not satisfy the similarity assumption even in the mean. Therefore, both assumptions need experimental verification. We should also point out that, before the distribution of established whitecaps is
formed and they commence the negative work on the wave, some energy is already lost from the wave to form the whitecaps, which is not accounted for by such a model.

Polnikov (1993) suggested another type of an after-breaking model. He argued that, no matter what the cause of the breaking, the result is turbulence in the water. In his approach the rate of wave spectrum dissipation is governed by the effective turbulent viscosity $\nu_T$. Therefore, to describe the wave energy dissipation in a wave spectrum form, it is sufficient to find a link between the wave spectrum and the water turbulence spectrum. To do this, he wrote the dynamic equations, performed an averaging, and introduced a Reynolds stress. Then, the Reynolds stress was expanded into a series with respect to wave velocity components and their spatial derivatives. The Prandtl hypothesis was used to close the turbulent terms in these series. Finally, Polnikov found that the effective viscosity due to turbulence has a form of series with respect to wave spectrum in which the quadratic term should dominate. Therefore, the dissipation should be quadratic in the spectrum.

Again, the idea is attractive, but the theory needs further development. Polnikov (1993) assumes a simplified representation of wave dynamics equations with the efficient stress attenuation that is appropriate for monochromatic waves. But spectral waves of different scales interact, and the turbulent vortexes of particular scales are not only generated as a result of dissipation of counterpart waves, but also as a result of the collapse of larger vortexes (Kolmogorov cascade). Besides, we should point out that application of the eddy viscosity to the wave-induced motion is in contradiction with accepted approaches in this field (see sub-section on wave–turbulence interactions below).

Most importantly, however, generation of the turbulence is not the only outcome of dissipation of wave energy. Melville et al. (1992) showed that 30–50% of energy lost by breaking waves is expended on entraining bubbles into the water against buoyancy forces. This contribution, relative to the turbulence generation, is not constant across the spectrum. For example, microscale breakers do not cause air entrainment and therefore should expend relatively more energy on generating the turbulence.

To summarize this brief overview of existing theories of spectral dissipation, we find several studies which offer four different analytical models. None of the models deals with the dynamics of wave breaking, which is responsible for dissipation. Rather, they suggest hypotheses to interpret either pre-breaking or post-breaking wave field properties. All of the hypotheses lack experimental support or validation. Results vary from the dissipation being a linear function of the wave spectrum to the dissipation being quadratic, cubic or even a function of the spectrum to the fifth power.

Experimental confirmation should be an important element of the development of a theory. There have, however, been few experimental studies of wave dissipation. Thorpe (1993), Melville (1994), Terray et al. (1996), Hanson and Phillips (1999), among others, addressed the total dissipation. Experimental investigations of the spectral dissipation are all very recent: Donelan (2001), Phillips et al. (2001), Melville and Matusov (2002), Hwang and Wang (2004), Babanin and Young (2005), Young and Babanin (2006) have made first attempts to obtain spectral dissipation functions on the basis of field measurements.

Phillips et al. (2001) used high range resolution radar measurements and Melville and Matusov (2002) used aerial imaging to study distributions of the length of breaking wave fronts $A(c)$ where $A(c)dc$ is the average length of breaking crests per unit area of ocean surface travelling at velocities from $c$ to $c + dc$ (Phillips, 1985). They inferred a spectral function for the dissipation in terms of the phase speed $c$ as the spectral parameter. Phillips et al., 2001 obtained it for a single wind speed and Melville and Matusov (2002) included a wind dependence of $U_{10}^3$ into the $A$ function:

$$S_{ds}(c) = \frac{\rho_w}{g} c^5 A(c) \left( \frac{10}{U_{10}} \right)^3,$$

(4.2)

where the wind speed $U_{10}^3$ has to be expressed in m/s. Connection of this dissipation with the wave spectrum was not obtained explicitly and therefore it cannot be directly compared with other dissipation functions below.

Donelan (2001) (as did Phillips (1985) and Donelan and Pierson (1987) in analytical models described above) used the balance of source terms to derive $S_{ds}$. He argued that, for stationary fetch-limited no-current conditions, $S_{in}$ and $S_{ds}$ are more than an order of magnitude larger than the advection and the nonlinear interaction terms in some parts of the wave spectrum. Therefore, there are wavenumbers in the wave spectrum $\Phi(k)$ where the balance is totally dominated by the wind input and the dissipation. If spectra of young fetch-limited waves are considered and an appropriate hypothesis about the form of the dissipation function is used, the
spectral dissipation can be obtained from the spectral wind input function. Using only peak values of his spectra, Donelan (2001) obtained the dissipation as

$$S_{ds} = 36\omega(k)E(k)B(k)^{2.5} \sim E(k)^{3.5},$$

(4.3)

where $B(k) = k^4E(k)$ is termed the saturation spectrum (Phillips, 1984). Here, the dissipation remains local in wavenumber space.

However, once Donelan (2001) applied his function to the measured spectra at wavenumbers above the spectral peak, the $S_{in}$ and $S_{ds}$ balance could not be satisfied. The two energy source functions could only be brought into balance by assuming that the mean square slope $s$ of long waves modifies the dissipation rate at shorter waves. The dissipation function was adjusted accordingly:

$$S_{ds} = 36\omega(k)E(k)[(1 + 500s(k))^2B(k)]^{2.5}.$$

(4.4)

The dissipation (4.4) is not local in wavenumber space, due to the $s$ term, – as the quasi-saturated theories suggested, – but on the contrary, acknowledges the importance of influence of longer waves on the dissipation of short waves.

The influence, according to Donelan (2001), is due to the fact that dissipation rates for the short quasi-saturated waves are modulated by the straining action of longer waves. On the forward faces of longer waves, the short-wave steepness increases causing frequent breaking and correspondingly a net reduction in the energy density.

The factor 500, however, may appear too large. A rather large swell of 2 m significant height and 10 s period gives a dissipation which is greater by a factor 2.3 compared to the case without swell. Such a large dissipation would result in a lower wave growth, which does not seem consistent with the data (e.g. Dobson et al., 1989), although a detailed hindcast would be necessary.

Apart from this mechanism for longer waves affecting dissipation at shorter scales, other mechanisms have also been suggested by experimentalists. The other mechanisms involve effects due to breaking of large waves. Banner et al. (1989) showed that the large scale breaking brings about rapid attenuation of short waves in its wake and therefore may cause the spectral dissipation function to depend on frequency relative to the peak.

Meza et al. (2000), in a laboratory experiment with forced isolated breakers within transient wave trains, showed that large breakers do not cause energy loss from dominant waves – but almost exclusively from wave components well above the spectral peak. An unresolved effect here is whether the loss is predominantly from bound harmonic nonlinearities of the steep dominant waves, or from the shorter free waves.

Hwang and Wang (2004), like Donelan (2001), used the source term balance idea to derive $S_{ds}$. The approach follows closely the discussions of Phillips (1984) who suggested that knowledge of the spectrum dependence on wind speed can be used to understand the behaviour of the dissipation function. They applied the source term balance approach to the spectra of short waves, twice the peak frequency and above, with wavelengths from 2 cm to 6 m, collected in the ocean using a free-drifting measurement technique to mitigate the problems associated with Doppler frequency shift of short-scale waves. A unique feature in their result is a non-monotonic behaviour of the dissipation function, proportional to $E^{2.3}$ for capillary waves, approaching $E^5$ at the other end of the wavelength scale, and reaching up to $E^{10}$ in the middle wavelength range (0.2–1.5 m long). They suggest that the quasi-singular behaviour of the dissipation in the middle wavelength range may be an indication that the important spectral signature of wave breaking has a maximum in the wavenumber domain.

Why would the maximum of spectral restoration occur in the intermediate scale waves with wavelengths between 0.2 and 1.5 m? The approach is based on the assumption of local spectral balance between the wind input and dissipation and, since their spectral wind input is a linear function of the wave spectrum, the sudden rise of the dissipation to being $\sim E^{10}$ apparently reflects the sudden rise of the responsiveness of wave breaking to the wave spectral density disturbances at the respective scales. This enhanced spectral density responsiveness at the middle wavelength range is suggestive that the wind input, which is assumed a monotonic function of wavenumber, is not the only mechanism that generates intermediate scale waves. Detachment of breaking jet, impulsive impact and the waveform deformation due to wave breaking will produce spectral signature in the intermediate wavelengths. Hwang (2005) argues that the excessive generation at these scales of shorter waves is perhaps brought about by breaking of larger dominant waves. Such excessive generation is not
accounted for in our present formulation of the action or energy density conservation equation. As a result, it is compensated artificially by an excessive dissipation function, and subsequently manifests itself via enhanced level of breaking of shorter waves. The correlation between the dominant breaking and the short-scale breaking was observed by means of radar and acoustic sensing of the ocean surface (Hwang, 2005). Thus, if the balance approach remains valid in such circumstances, it is not only the spectral energy dissipation appears to be a function not local in wavenumber space, but the spectral energy input as well. In a way, such mechanism is supportive of the idea of the cumulative term described above.

Young and Babanin (2006), based on Lake George field data, conducted a direct attempt to estimate the spectral distribution of the dissipation due to breaking of dominant waves. A field wave record with approximately 50% dominant breaking rate was analysed. Segments of the record, comprising sequences of breaking waves, were used to obtain the “breaking spectrum”, and segments of non-breaking waves to obtain the “non-breaking spectrum”. The clearly visible difference between the two spectra was attributed to the dissipation due to breaking. This assumption was supported by independent measurements of total dissipation of kinetic energy in the water column at the measurement location.

It was shown that the dominant breaking causes energy dissipation throughout the entire spectrum at scales smaller than the spectral peak waves. The dissipation rate at each frequency appears linear in terms of the wave spectral density at that frequency, less a spectral threshold value, with a correction for the directional spectral width $A(f)$ (Babanin and Soloviev, 1998b). The spectral dissipation source term can be represented by:

$$ S_{ds}(f) = a g f \cdot X((E(f) - E_{th}(f))A(f)) + bg \int_{f_p}^{f} X((E(q) - E_{th}(q))A(q)) \, dq. \quad (4.5) $$

Here, the integral reflects a contribution to the dissipation at each frequency $f$ from waves breaking at frequencies $f_p < f < f_r$, and $X((E(f) - E_{th}(f))A(f))$ is a yet unknown function that controls inherent wave breaking at each frequency (perhaps a function of the form described by (4.1) with $n = 1$, see Babanin and Young, 2005). The experimental coefficients $a$ and $b$ were found to be 0.0065, but these parameters may be also dependent on environmental conditions (only a single record was analysed in the paper).

Thus, the only two experimental dissipation functions available, which cover the entire spectral frequency band, (4.4) and (4.5) exhibit a common feature: cumulative term that puts whitecapping dissipation at smaller spectral scales in dependence on what happens at larger scales. Consistency of this feature has been confirmed by further investigations of the Lake George wave breaking data by independent means (Babanin and Young, 2005; Manasseh et al., 2006) where the two-phase behaviour of the spectral dissipation has also been obtained.

A passive acoustic method of detecting individual bubble-formation events developed by Manasseh et al. (2006) was found promising for obtaining both the rate of occurrence of breaking events at different wave scales and the severity of wave breaking. A combination of the two should lead to direct estimates of the spectral distribution of wave dissipation.

If the wave energy dissipation at each frequency were due to breaking of waves of that frequency only, it should be a function of the excess of the spectral density above a dimensionless threshold spectral level, below which no breaking occurs at this frequency. This was found to be the case around the wave spectral peak. A more complex mechanism appears to be driving the whitecapping dissipation at scales smaller than those of the dominant waves where enhanced breaking frequency and dissipation rates are observed when expressed in terms of the wave spectrum. This signifies a two-phase behaviour: $S_{ds}$ being a simple function of the wave spectrum at the spectral peak and having an additional cumulative term at all frequencies above the peak.

The nature of the induced dissipation above the peak can be due to either enhanced induced wave breaking or additional turbulent eddy viscosity (see the next sub-section on the wave–turbulence interactions) or both. If the latter is true, then the dimensionless spectral threshold below which no dissipation occurs may not be universal (or at least may not have a simple identifiable functional form) across the spectrum.

Young and Babanin (2006) also compared directional spectra of the breaking and non-breaking waves whose difference should be indicative of the directional distribution of the dissipation. They showed that directional dissipation rates at oblique angles are higher than the dissipation in the main wave propagation direction and therefore the breaking tends to make the wave directional spectra narrower. If confirmed, this conclusion may have very significant implications for the directional shape of $S_{ds}$: unlike $S_{in}$, it would be bimo-
dal with respect to the wind direction, and the main wave direction would be characterized by a local minimum of the directional spectrum of dissipation.

Hence, the experimental evidence indicates that the dissipation function is likely to be not local in wave-number space and is rather a functional of the wave spectrum. The experiments do not support any of the suggested theoretical forms for the dissipation as no analytical theories have produced the cumulative dissipation term. There are disagreements between experimental results as well: they offer different conclusions as to the mechanisms by which dominant waves affect smaller-scale dissipation. Banner et al. (1989), Meza et al. (2000), Young and Babanin (2006) attribute the effect to breaking waves, whereas, Donelan (2001) attributes the effect to non-breaking waves. On the other hand, results of Hwang and Wang (2004) indicate that, if the local balance of wind generation and breaking dissipation is true, then the dissipation function exhibits quasi-singular behaviour at the intermediate wave scales.

To conclude the review, we have to summarize that (1) there is no consensus among analytical theories of the spectral dissipation of wave energy due to wave breaking, even with respect to the basic characteristics of the dissipation function, (2) the theoretical dissipation functions strongly disagree with the experiment, and (3) experimental results, even though exhibit some common features, are often in serious disagreement with each other. Such a state of knowledge of physics of the wave breaking losses does not help modelling the wave dissipation which has been drifting in its own way (see Section 4.2).

The review of studies of the dissipation term, however, would be incomplete without mentioning an alternative approach to the description of evolution of wave spectrum which does not require detailed knowledge and, in fact, existence of the spectral dissipation (Zakharov, 1966; Zakharov and Filonenko, 1967; Zakharov, 1968; Zakharov and Smilga, 1981; Zakharov and Zaslavskii, 1982a,b; Zakharov and Zaslavskii, 1983a,b; Kitaigorodskii, 1983; Zakharov, 2002; Zakharov, 2005). In their theory of weak turbulence, Vladimir Zakharov and his colleagues obtain a Kolmogorov spectrum of $E(n) \sim n^{-4}$ as an exact solution of the kinetic equation for gravity waves in the equilibrium interval. This spectrum agrees with many experimental observations (Toba, 1972; Kahma, 1981; Leykin and Rozenberg, 1984; Donelan et al., 1985; Hwang et al., 2000, among others). In addition, Zakharov (2002) was able to reproduce known growth curves of wave integral properties as analytical solutions on the basis of the theory of weak turbulence. This theory relies on the assumption that the whitecap dissipation can be neglected in the frequency range of the spectral peak and the universal region at wavenumbers above the peak. This theoretical assumption, however, is not obvious again, as the dominant waves are known to break, sometimes quite frequently (Babanin et al., 2001; Young and Babanin, 2006) and there are experimental evidences regarding the significant effect that dominant breaking has on wave spectral peak dissipation (Donelan, 2001; Young and Babanin, 2006; Babanin and Young, 2005).

4.1.2. Wave–turbulence interactions

It was recognised very early that viscosity had a negligible effect on waves of periods of about 10 s and longer (Lamb, 1932), so that, once generated, swells were supposed to dissipate slowly due to the action of the wind, as represented by Jeffreys (1925) sheltering theory (Sverdrup and Munk, 1947). These ideas have been gradually abandoned and traded for eddy viscosity analogies (Bowden, 1950; Groen and Dorrestein, 1950) that are used today in some operational wave forecasting models (e.g. Tolman and Chalikov, 1996). Yet there is no evidence that wave-induced velocity profiles are unstable and may become turbulent, except for the surface viscous layer (a few millimeters thick) and the wave bottom boundary layer. Therefore, except in these boundary layers, the local turbulent motions are possibly not related to the wave velocity field and no theory can justify the use of eddy viscosities. Instead, the stretching of turbulent eddies by the wave motion may lead to a different effect, and we should consider also the possible scattering of waves by turbulence. In order to represent the stretching, rapid-distortion theory was applied on the water-side of the surface by Teixeira and Belcher (2002). The theory assumes that the eddy turn-over time is less than the wave period, or, said differently, that the strain rate of the turbulence by the wave motion is more than that of the turbulence by itself, and that the turbulent velocity is much less than the wave-induced velocity. For the wave components that satisfy these conditions, the theory yields the following rate of production of turbulent kinetic energy

$$P_{ws} = \overline{u'^{*} w'^{*}} \cdot \frac{\partial U^{**}}{\partial z}, \quad (4.6)$$

L. Cavaleri et al. / Progress in Oceanography 75 (2007) 603–674
where the Cartesian components of the fluctuating turbulent velocity are $u'_x$ ($x = 1, 2$) and $w'$ in the water, and the (horizontal) components of the Stokes drift are $U_{sx}$. This expression may be considered obvious when compared to the usual production of TKE due to the mean current shear, but it must be kept in mind that the Stokes drift is not a mean current, and has rather funny properties. Among these the Stokes drift is rotational although it is the residual of an exactly irrotational motion.

Eq. (4.6) gives an energy rate of change of the form

$$\frac{dE(\vec{k})}{dt} = S(\vec{k}) = \beta \cos(\vec{k}).$$

Assuming a uniform turbulent flux, for the sake of simplicity, $\beta = \beta_{\text{turb}}$, with

$$\beta_{\text{turb}}(\vec{k}) = -k \frac{\rho_a}{g \rho_w} u_*^2 \cos \tilde{\theta} \cos(2kH) \sinh^2(kH),$$

where $\rho_a$ and $\rho_w$ are the air and water densities, respectively, $u_*$ is the friction velocity of the air flow, $H$ is the water depth, and $\tilde{\theta}$ is the direction of the waves relative to the wind stress direction. Eq. (4.8) takes the following limit for deep water,

$$\beta_{\text{turb}}(\vec{k}) = -k \frac{2 \rho_a}{\rho_w} \frac{u_*^2}{C^2} \cos \tilde{\theta},$$

where $C$ is the phase speed of the wave component of wavenumber $k$.

Ardhuin and Jenkins (2006), arrived at the same expression by using Lagrangian-mean (Andrews and McIntyre, 1978) of the shear production term in the turbulent kinetic equation and assuming that the turbulent flux $u'_x w'$ is uniform, and in particular not correlated with the wave phase, or, at most, weakly modulated. Considering that a large part of the momentum flux may be carried by long-lived and stable Langmuir rolls, the weak modulation of the turbulent flux by the waves is a likely hypothesis. However, turbulence is also to be likely strongest at the peak of wave groups where the Stokes drift is largest. Thus the wave dissipation can easily be larger than that given by (4.9). Employing the latest results on wave breaking relations with wave groups could lead to a better estimate of this effect. It should also be noted that this process is capable of producing turbulence at larger depth compared to that produced by whitecaps. This may resolve some problems faced by models of the ocean mixed layer that fail to predict mixed layer depth deep enough in cases of positive or zero buoyant fluxes, such as in the Southern Ocean summer.

4.1.3. Wave–wave modulations

Phillips (1963) noted that short wave breaking in the presence of long wave modulation was taking some energy from the long waves through the modulation. These ideas were revisited by Longuet-Higgins (1969b) who proposed a “maser mechanism” with the short wave breaking modulation feeding the growth of the long waves. However, Hasselmann (1971) showed how the maser mechanism is largely cancelled by the variation of the short wave potential energy, and found that only the much weaker dissipation remained as proposed by Phillips (1963). When evaluated with reasonable modulation transfer functions, that dissipation is typically slightly larger than the viscous dissipation (Ardhuin and Jenkins, 2005). Yet, Hasselmann (1971) neglected the modulations of the wind stress that can be significant (Garrett and Smith, 1976). One may thus follow Hasselmann’s (1971) and Garrett and Smith’s (1976) derivations and realize that this wind stress modulation, working against the wave orbital velocity, should be added to the long wave energy rate of change,

$$S_{\text{win}} = \langle T_{w}^2 \tilde{u}_w \rangle.$$

This effect dissipates the long waves that propagate against the wind, but may amplify the long waves that propagate with the wind. In all cases, the exchange of energy and momentum takes place between the wind and the long waves through the short waves, and not between the waves and the ocean circulation and the turbulence as in the previous cases. The wind stress modulation was estimated by Kudryavtsev and Makin (2004) and Ardhuin and Jenkins (2006) by means of using the rapid distortion theory in the air. This suggests that the swell dissipation is mostly due to this effect and that dissipation occurs for all directions of waves rel-
ative to the wind, with stronger dissipation for opposing winds. A re-estimation of the mixing length parameterization related to the “inner layer depth” (see Janssen, 2004) changes slightly the magnitude of the results. However, Kudryavtsev and Makin (2004) neglected the modulation of the surface roughness which, as envisaged by Garrett and Smith (1976), may contribute to the growth of the waves in the wind direction. A qualitative estimation of that effect by Ardhuin and Jenkins (2006), based on the modulation transfer functions of Hara et al. (2003), suggests that the roughness modulation should have a weaker effect than the stress modulation. Direct measurements of wind stress modulations is probably a serious challenge, but it should be considered. A better knowledge of the modulation of short wave amplitudes is also needed to improve parameterizations of these effects.

4.2. Modelling the spectral dissipation function

Understanding the physics of wave dissipation from a spectral perspective has been so incomplete that the spectral dissipation rate, unlike the wind input and nonlinear transfer, has been inferred indirectly by modelling the evolution of the wave spectrum rather than by parameterizing known physical features of the dissipation directly. Sometimes, such attempts have been based on trying to fit the dissipation term to an existing analytical model (Komen et al., 1984; Polnikov, 1993), but mostly such terms are tuning knobs that may or may not involve reference to the physics.

As an illustration, it is generally recognised that a major part of the wave dissipation is produced by the wave breaking. Nevertheless many recent experimentally-discovered features of the breaking-induced dissipation have not yet been incorporated in wave models (see Section 4.1). Particular physical mechanisms that have been identified by modellers for inclusion are breaking threshold behaviour based on local spectral saturation (e.g., Alves and Banner, 2003) rather than integral wave steepness (Komen et al., 1984), and additional short wave extinction through cumulative nonlinear interaction with longer waves (e.g., Donelan, WISE 2005), amongst several others.

On another part, there is a growing discussion on what physical features have to be excluded from being damped in the spectral models as a result of artificially tuned dissipations. For example, Lavrenov (2004) showed that, if the dissipation function is not forced to suppress the low-frequency spectral energy, this may result in return energy fluxes from the waves into the atmospheric boundary layer, up to a quarter of the total wind-to-wave flux in magnitude. This considerable additional source of energy for the atmosphere may prove a significant factor in weather and climate forecasts. Another example: at WISE 2005 mentioned above, Donelan and Meza in two separate papers presented dissipation functions responsible for the spectral peak downshift. Such a feature does not appear in dissipation functions presently in use, but is consistent with laboratory experiments of Tulin and Waseda (1999).

However, the relative importance of such mechanisms, identified above in Section 4.1 for real waves, and therefore their relevance for spectral models, is often not clear. Thus, models should not have to shoulder the immediate blame for not conforming to observational physics as soon as the latter is revealed. In any event, progress in modelling the spectral dissipation rests heavily on validation methods that differ intrinsically from those which highlight progress in studies of the physics of the dissipation.

Therefore, in this section we will not concentrate on a detailed analysis of dissipation terms included in wave research and operational models. Given the recent experimental advances, proposed forms for the dissipation rate term are rapidly evolving and are likely to evolve further in coming years. Instead, we shall analyse the progress of methodology for modelling and verification of the dissipation functions and indicate possible future ways for this to progress. In brief, the major historical stages of the methodology of tuning the dissipation term can be summarised as follows: (1) considering the balance of source terms in order to obtain the known integral evolution curves (e.g., Komen et al., 1984); (2) validating the spectral balance evolution to ensure the known spectrum development behaviour is satisfied (e.g., Banner and Young, 1994); (3) uncoupling the dissipation term from the source term balance in an attempt to tune it directly against known wave breaking characteristics (e.g. Banner, Kriezi and Morison, WISE 2004); (4) further tuning the standalone dissipation function against other dissipation-related properties and constraints (next step); (5) employing exact physics, both experimental and theoretical (future).
(1) Up to now, progress on dissipation modelling is seen through the growing ability of the employed dissipation terms $S_{ds}$ to reproduce refined features of spectral wave evolution. As mentioned above, the groundwork was set by Komen et al. (1984) who first demonstrated the possibility of obtaining and tuning a form of the spectral dissipation function by considering the balance of all source terms in the radiative transfer equation. They based their choice of the function form on a rather free interpretation of the Hasselmann’s (1974) analytical model for whitecap dissipation as random pressure pulses, and introduced a set of wave evolution tests to verify the dissipation function. Once the proposed dissipation function was implemented in the evolution runs, the model had to reproduce the experimentally well-known evolution of wave integral properties – variance and peak frequency. A list of more recent dissipation functions falling into this category includes, but is not limited to, Makin et al. (1995), Tolman and Chalikov (1996), Schneggenburger et al. (2000). Unfortunately, all the tests by Komen et al. (1984) were performed for wind sea growth in the absence of swell, which was later found to have a very large spurious effect on the parameterizations (Tolman and Chalikov, 1996; Booij and Holthuijsen, 2002). This problem is inherent to the definition of a mean steepness from the entire spectrum, and leads to overestimations of wind sea growth in the presence of swell by as much as a factor of 2 (Ardhuin et al., 2007), even with the latest modifications to Komen et al.’s formulation by Bidlot et al. (2005).

(2) The next significant step in fine-tuning the dissipation term was achieved by Young and Banner (1992) and Banner and Young (1994) who introduced a requirement for the modelled evolution, based on the use of a chosen dissipation function, to reproduce an experimentally known form of the wavenumber spectrum tail. Obviously, the spectral models need to be able to simulate development of the directional spectrum as well as its integral properties. This additional requirement put the Komen et al. (1984) dissipation term to a serious test and it was concluded that this term can hardly satisfy all the evolution dependences at the same time. Particular difficulties were encountered while attempting to tune this term to reproduce experimentally known directional properties of the wave spectra. Recent dissipation models in this category include Meza et al. (2002), Alves and Banner (2003), Lavrenov (2004), Bidlot et al. (WISE 2005), Donelan (WISE 2005). Among other important conclusions of Banner and Young (1994) was a demonstration of sensitivity of the evolution results to variations of other than the dissipation source terms. Fixing the high-frequency spectrum tail to an $f^{-5}$ dependence, as in Komen et al. (1984), brought about essential changes to the nonlinear term which then had to be compensated by additional alterations of $S_{ds}$. This revealed an ambiguity in verification of the dissipation term on the basis of evolution runs that rely on simultaneous balance of all the sources/sinks.

(3) This ambiguity is being overcome by employing a new series of direct tests in recent attempts to model the dissipation function (Banner, Kriezi and Morison, WISE 2004). This dissipation function is based on a local spectral saturation breaking threshold, refining the approach of Alves and Banner (2003). Banner and his group proposed that, since the major contributor into the spectral dissipation is the wave breaking, the dissipation function should be verified against its ability to reproduce observed spectral distributions of wave breaking, as well as against the evolution dependences for spectral and integral properties. The observed spectral distribution of the length of breaking wave fronts, $\Lambda(c)$, obtained by Melville and Matusov (2002), and more recent results from Gemmrich (2005) have been used for the verification purposes. This work is still in progress.

(4) In the meantime, it is obvious that even though a major part of the wave dissipation is due to the breaking, there are other mechanisms that contribute to the dissipation (see Section 4.1) and a general set of spectral and integral constraints for the stand-alone dissipation function has to include the impact of those mechanisms. For example, in a recent experimental study, Babanin and Young (2005) showed that spectral dissipation rate estimates, when compared to the dissipation rate inferred by Melville and Matusov (2002), indicate that the turbulent viscosity becomes significant at small wave scales, where the cumulative term of the function (4.5) dominates. Therefore, tuning the dissipation function against distributions of $\Lambda(c)$ would require corrections at those short wave scales, as the additional dissipation due to turbulent viscosity does not manifest itself by means of whitecapping. As complex as it might appear in deep water, the physics of wave dissipation in shallow water appears to be dominated by yet more new
physical features (e.g. Song and Banner, 2004). Implementation of these and other constraints, proper mathematical employment of experimentally-observed features of dissipation behaviour, as well as exact physics, still belong to the future of spectral dissipation modelling.

5. Nonlinear interactions in shallow water waves

Contributing authors: Miguel Onorato (onorato@ph.unito.it), Jacco Groeneweg, Thomas H.C. Herbers, Tim Janssen, Alexandru Sheremet, Jane McKee Smith.

In this section, we discuss the role of the nonlinear interactions in shallow water. The section contains the description of the two approaches available for describing waves as they propagate towards the shore. The first approach consists in describing the waves by deterministic equations (simplified models such as for example the Boussinesq equations or even the fully nonlinear equations). The second one deals with stochastic models, i.e., models that are derived from the deterministic ones under a closure hypothesis (usually the random phase approximation is adopted). Limitations of both approaches are elucidated. Some comments on wave breaking and dissipation in shallow water are also included.

5.1. Nonlinearity in shallow water

As waves propagate from deep water into shallow coastal areas, frequency dispersion diminishes and quadratic near-resonances (Bryant, 1973) transform near-symmetrical waves to skewed, pitched-forward shapes as observed on beaches at the onset of wave breaking (see e.g. Elgar and Guza, 1985), and induce radiation of long waves at the ‘beat’ frequencies of the incident wave field, generally referred to as ‘surfbeat’ (e.g., Munk, 1949).

Historically, shallow-water wave models are based on the classical uniform-depth theories of Boussinesq (1872) and Korteweg and deVries (1895), extended to variable depth by Peregrine (1967); these theories assume the Stokes (or Ursell) number O(1) from the outset, i.e., nonlinearity, $a/h$, and dispersion, $(kh)^2$, are assumed to be of the same order. Although the original Boussinesq approximation accounted only for weak dispersion and nonlinearity, limiting its validity to (very) shallow water, recent advances include full nonlinearity (Wei et al., 1995) and high-order dispersion effects (e.g., Madsen et al., 2003), supporting modelling of waves in deep-intermediate water and to very high-nonlinearity (see, e.g., Fuhrman et al., 2004). Reviews of developments in Boussinesq theory are found in e.g. Kirby (1997) and Madsen and Schäffer (1999); more recent advances include Chen et al. (2000, 2003); Watts et al. (2003); Shi and Kirby (2005), and many others.

Hasselmann’s theory for resonant quartet interactions, which forms the basis for most deep water wave prediction models, is restricted to deep and intermediate water depths where the Stokes number $\lesssim 1$ (see Zakharov, 1999). It is well known that the lower order triad interactions are non-resonant in deep-intermediate depth water, forcing second-order bound components that can be important locally but do not contribute to the wave evolution over large distances. However, as ocean surface waves propagate from deep to shallow water, triad interactions approach resonance and assume a dominant role in the dynamics. This transition from quartet to triad interactions is the result of the change in the dispersion relation from a dispersive deep water regime that does not support resonant triad interactions (Phillips, 1960) to a non-dispersive shallow-water regime where all wave components travel with the same speed. Although triad interactions are exactly resonant only for uni-directional waves in the non-dispersive shallow water limit, near-resonant triad interactions can play a dominant role in the evolution of waves in shallow coastal areas. For example, a periodic wave train with frequency $\omega$ and wave number $k$ is accompanied by harmonic components $(2\omega, 2k), (3\omega, 3k)$, etc. that are bound in deep-intermediate water depths where they do not obey the gravity wave dispersion relation, but grow rapidly in shallow water where the mismatch from resonance is weak.

In general triad interactions transfer energy from the incident wave components to higher- (e.g., harmonic) and lower- (e.g., infra-gravity) frequency components (see, e.g., Freilich and Guza, 1984; Elgar and Guza, 1985; Agnon et al., 1993; Herbers et al., 1994; Kaihatu and Kirby, 1995; Agnon and Sheremet, 1997; Herbers and Burton, 1997; Ruessink, 1998; Kaihatu, 2001; Sheremet et al., 2002; Janssen et al., 2003, and many others). These interactions not only broaden the frequency spectrum in shallow water, but also phase-couple
the spectral components, causing the characteristic steepening and pitching forward of near-breaking wave crests.

Shallow water wave propagation models can generally be divided in two major categories (see also Agnon and Sheremet, 2000):

(i) Deterministic (phase resolving) models are usually derived from the Euler equation for potential flows (Laplace equation + boundary conditions) under the hypothesis of weak nonlinearity and in the limit of shallow water, i.e. $kh \to 0$. These models, including both the physical domain Boussinesq models and the complex amplitude evolution models (spectral models), resolve the phases of the individual waves.

(ii) Stochastic (phase-averaged) models are derived from deterministic equations by applying a turbulence-like closure hypothesis to the infinite set of coupled equations governing the evolution of the spectral moments. For any given deterministic wave equation, with a suitable closure hypothesis, a stochastic model can be developed. Since the closure approximation invariably introduces errors, the underlying deterministic model is in principle more accurate than its stochastic counterpart.

As waves approach the shore, additional effects such as bottom friction and depth-induced wave breaking must be considered.

5.2. Deterministic models: time-domain and spectral-domain

Time-domain Boussinesq models are typically applied to domains with spatial scales of the order of 10 wavelengths. Computational demands become prohibitive for larger scale applications (see, e.g., Fuhrman and Bingham, 2004). Moreover, in practice, the required phase-resolving boundary conditions are often not available and the need for wave field statistics (instead of details of a single realization) requires the computation of a multitude of realizations. Given the considerable computational requirements for even a single realization, clearly such repeated simulations are extremely time consuming for two-dimensional applications on domains of appreciable extent. Despite the recent advances in Boussinesq modelling, including the modelling of highly nonlinear waves (Wei et al., 1995) in fairly deep water (Madsen et al., 2003; Fuhrman and Bingham, 2004) and wide ranging capabilities to model refraction, reflections and wave-induced currents, the computational demands for computing wave field statistics for random, directionally spread waves seriously limits the use of such models for operational nearshore wave prediction.

An efficient alternative to time domain models are so-called (complex) amplitude evolution or spectral models. This class of models essentially expresses the wave field as a superposition of plane waves (Fourier modes), and consists of a set of coupled evolution equations for the Fourier amplitudes. The application of the Fourier transform results in a dimensional reduction of the governing equations at the expense of convolution-type forcing terms to account for nonlinear interactions. Fourier models are attractive because they provide a natural continuation of the deep-water approach, and are well suited to handling processes of an intrinsic statistical nature such as dissipation and wind input.

While in deep water the temporal evolution of different wave numbers is usually considered (the Fourier Transform from spatial coordinates $(x, y)$ to wave numbers $(k_x, k_y)$ is taken), a careful treatment is required for the richer family of finite-depth wave-fields (including evanescent, trapped or singular modes, Whitham, 1979), which are also intrinsically inhomogeneous. For variable depth problems, due to the fact that in linear theory the waves preserve their frequency as they shoal, it is more convenient to work in frequency Fourier domain (rather than the wave number domain) and solve for the amplitude evolution in space of waves with different frequencies. Freilich and Guza (1984) developed a frequency domain wave shoaling model based on Peregrine’s (1967) extension of Boussinesq’s theory to varying depth. Many such models have been reported in the literature, either based on Boussinesq theory (Freilich and Guza, 1984; Madsen and Sørensen, 1993; Herbers and Burton, 1997) or fully dispersive theory (e.g. Agnon et al., 1993; Kaihatu and Kirby, 1995; Eldeberky and Madsen, 1999); the latter class of models includes full linear dispersion and has no inherent depth restriction in the linear terms, but is derived assuming quadratic resonances from the outset. The restriction to quadratic near-resonances is removed by Bredmose et al. (2003) for unidirectional wave
propagation; for directional wave propagation over topography a generalized formulation (including off-resonant quadratic components) is derived in Bredmose et al. (2005) and Janssen et al. (2006); the latter model includes also cubic near-resonances, extending the model validity to intermediate water depths.

Similar to the time-domain models, these models are deterministic and in order to obtain wave field statistics, Monte Carlo simulations are required. Since in general, amplitude evolution models are numerically efficient when compared to time-domain models, such simulations (although time consuming) are feasible (see also Freilich and Guza, 1984). Monte Carlo simulations are typically performed by assuming the wave field Gaussian at the seaward boundary. Modal amplitudes are drawn from a Rayleigh distribution with variance derived from the observed (or theoretical) density spectrum, and random phases are added (see Tucker et al., 1984).

Deterministic spectral models are particularly well-suited to derive efficient, stochastic evolution models (see, e.g., Agnon and Sheremet, 1997; Herbers and Burton, 1997).

5.3. Stochastic models

The shoaling evolution of random waves on a beach can also be predicted with stochastic models that solve evolution equations for statistically averaged spectral wave properties. Such equations can be derived by manipulating the deterministic equations and ensemble-averaging (e.g., Benney and Saffman, 1966; Newell and Aucoin, 1971). At the lowest order, the procedure yields an evolution equation for the power spectrum which includes terms involving the third-order cumulant, the bi-spectrum. An evolution equation for the bi-spectrum can be derived at the next order, but this equation depends on the tri-spectrum (fourth-order cumulant), and so on. Thus the system never closes, leading to an infinite set of equations for the spectral moments. Even though it is well known that the probability density function of surface gravity waves can be far from Gaussian in shallow water (especially for large Stokes numbers), a quasi-Gaussian (or quasi-normal) closure is usually introduced.

Most of the stochastic shallow water models consist of two, coupled evolution equations, one for the wave spectrum and the other for the bi-spectrum. Such equations were introduced by Saffman in 1967, starting from the Korteweg de Vries equation. The same methodology has been used later for deriving stochastic models from different deterministic equations (e.g., Agnon and Sheremet, 1997; Herbers and Burton, 1997; Kof­kood-Hansen and Rasmussen, 1998; Eldeberky and Madsen, 1999).

Although even the earlier stochastic models (e.g., Agnon and Sheremet, 1997; Herbers and Burton, 1997) were derived for directional wave fields, apart from the simulations by Becq-Girard et al. (1999a) and recent advances by Herbers et al. (2003), most verification has been done for uni-directional waves. Comparisons of model predictions of wave spectra evolution to observations generally show good agreement at locations well outside the surf zone and for Stokes numbers less than 1.5. Higher-order statistics such as wave skewness and asymmetry are less well predicted, in particular in the surf zone. It is found that these parameters are sensitive to the type of spectral weighting function used in the dissipation source term that accounts for depth-induced wave breaking (Chen et al., 1997).

Stochastic models are efficient in the sense that they compute statistical quantities directly, without the need of repeated simulation; moreover, they can be initialized at the offshore boundary with wave spectra obtained from routine directional wave measurements or regional wave model predictions; the bi-spectrum can usually be initialized with standard second-order theory for uniform depth (Herbers and Burton, 1997). However, inherent to the derivation of such stochastic models is the requirement of some sort of statistical closure. The commonly used quasi-Gaussian closure is not suitable for modelling wave evolution over long distances through regions of strong nonlinearity and dissipation, where it produces an unrealistic divergence from Gaussian statistics, leading to overly strong nonlinear couplings and potentially even negative energies (see Orszag, 1970), which is clearly unrealistic. For larger Stokes numbers and thus at locations close to the surf zone, the fundamental nature of the closure approximation negatively affects the model performance, even to the extent that predictions are physically unrealistic. To alleviate this problem and extend the modelling capability of these stochastic models into the surf zone, Herbers et al. (2003) proposed a heuristic, dissipation-controlled closure approximation, with a relaxation to Gaussian statistics on the scale of the surf zone width. This approach is similar to the relaxation of the quasi-Gaussian closure used in turbulence models (e.g., Salmon,
Generally good agreement between observations and model simulations is found, even at locations well within the surf zone.

Starting from a general three-wave interaction equation for water waves, Zaslavskii and Polnikov (1998) have derived an evolution equation for the wave action spectrum. The approach is basically the same as used in the derivation of the standard kinetic equation in deep water (including the quasi-Gaussian approximation). However, since there are no exact triad resonances, a spread delta function, characterized by a spreading parameter, is retained in place of the usual delta function in frequency. The final set of equations is referred to as a “quasi-kinetic equation”. In the case of one-dimensional propagation, numerical results have been compared with experimental data with some success (see Polnikov, 2000 and Piscopia et al., 2003). Some issues concerning the conservation of energy for the quasi-kinetic equation remain to be resolved (see Becq-Girard et al., 1999a). In the same context, Onorato et al. (2004) have derived a single evolution equation for the evolution of the wave action spectrum including quasi-resonant interactions (a spread delta function of the form of \( \sin(\Delta \omega t)/\Delta \omega \) was derived after analytical integration of the equation for the bi-spectrum), but no comparison with experimental data has been reported.

Earlier models, derived with the purpose of application in operational wave forecasting models (e.g. SWAN), include only self–self interactions (e.g., Eldeberky and Battjes, 1996; Becq-Girard et al., 1999b). These approximations are numerically efficient but rather crude representations of the nonlinear physics. Experiments involving unidirectional wave propagation indicate that, although these models can reproduce the generation of higher harmonics, they do not reproduce the release of such harmonics for increasing water depth; consequently, they usually result in an overestimation of the energy content at harmonic frequency ranges. Difference interactions, forcing low-frequency wave motions, are not accounted for in these simplified models, which further hamper their successful application in a realistic setting.

It is a misconception that stochastic models as described here can be applied to numerical domains with much coarser grids than deterministic models. In order to model quadratic interactions, the resonance mismatch needs to be resolved (e.g., Kofoed-Hansen and Rasmussen, 1998). This implies that grid resolution requirements for these stochastic models are generally similar to that of deterministic (spectral) models, and thus more stringent than those of conventional phase-averaged energy transport equations in deep water.

### 5.4. Dissipation and wave breaking in shallow water

Although nonlinear energy transfers can be predicted with rigorous theories, wave dissipation in the surf zone is not well understood and is modelled heuristically. Schäffer et al. (1993) include a turbulent surface roller in a time-domain Boussinesq model that yields a realistic description of the evolution of wave profiles in the surf zone. Most models for the breaking of random waves are based on the analogy of individual wave crests with turbulent bores (Battjes and Janssen, 1978; Thornton and Guza, 1983). Although these bore models yield robust estimates of bulk dissipation rates in the surf zone, the spectral characteristics of the energy losses are not specified, and somewhat arbitrary quasi-linear spectral forms of the dissipation function are used in Boussinesq models (Mase and Kirby, 1993; Eldeberky and Battjes, 1996). Boussinesq model predictions of wave frequency spectra in the surf zone appear to be insensitive to the precise frequency dependence of the dissipation function, but predictions of wave skewness and asymmetry are considerably more accurate if dissipation is weighted toward high-frequency components of the spectrum (Chen et al., 1997). Estimates of nonlinear energy transfers in the surf zone based on bispectral analysis of near-bottom pressure fluctuations confirm the dominant role of triad interactions in the spectral energy balance, transferring energy from the dominant incident wave frequencies to the dissipative high-frequency tail of the spectrum (Herbers et al., 2000).

The fate of difference three-wave interactions (which are responsible for the generation of low frequency infragravity waves – frequencies of 0.002–0.02 Hz) as the sea-swell propagates through the surf zone has been studied recently using field data (see, for example, Sheremet et al., 2002 and references therein). Observations show that the nonlinear coupling associated with this type of interaction strengthens continuously in the shoaling zone, where three-wave interactions are increasingly active and most of the shape transformation of the waves occurs. In the vicinity of the breaking point, however, the coupling is effectively destroyed, and infragravity waves are released. This process seems to justify the use of “unidirectional” hyperbolic spectral models limited to shoreward propagation, as opposed to more complex elliptic “bi-directional” models (e.g., Mild Slope Equation).
5.5. Open problems

Although great advances have been made in modelling wave propagation in finite-depth, the topic is far from being exhausted. For instance, there is a wide variety of wave–bottom interaction processes that are difficult to fit into a single, complete and effective model. Some aspects, such as bottom friction processes, are discussed elsewhere (reference bottom friction white paper).

Open problems related to nonlinear wave evolution in variable depth (in effect wave–wave–bottom interactions) are abundant. The following short discussion is confined to a few outstanding issues.

Applications of deterministic (phase-resolving) models to random waves are based on a principle similar to Monte-Carlo simulations. Waves enter the domain at a deep-end, where the wave field can be assumed to be Gaussian. An estimate of the deep water energy spectrum is used to generate random modal amplitudes and phases (see Section 5.2) at the domain boundary for each realization. It is unknown how many realizations are needed to obtain statistically reliable predictions of shallow water wave properties. In practice, a balance needs to be struck between a desirable large number (around 50 realizations typically reported) and the required computer time.

Alternatively, stochastic simulations are in principle more efficient since these models compute ensemble-averaged quantities directly, without the need for repeated simulations. However, the quasi-Gaussian statistical closure hypothesis commonly used in these models, can introduce large errors in applications over long distances or through regions of strong nonlinearity and dissipation. If the scope of these models is to be extended so that they can be applied over considerable distances in shallow water and through the surf zone, improvements in the statistical closure are needed (see e.g., Herbers et al., 2003).

It is important to recognize that, successful as the models discussed here have been at reproducing observed beach shoaling conditions, they are far from providing a robust general tool for wave forecasting in water of finite depth. In fact, most models were developed as nonlinear shoaling models, with the implied domain of application a typical sandy beach. Most of the model validation has been conducted on moderate slope beaches (1–5%) with shoaling ranges of the order of 10 characteristic wavelengths and nearly straight and parallel isobaths. Many natural coastlines have complex two-dimensional features such as shoals, banks and reefs where the combined effects of the topography and strong nonlinear interactions transform the wave field. The accuracy of existing models in these environments is not well understood. Additionally, many coastal regions contain wide shallow flats where nonlinear interactions evolve the wave field over hundreds of wavelengths. These large domains obviously strain the numerical resources needed for deterministic model simulations while likely invalidating the closure hypotheses used in stochastic models. Another limitation of most existing deterministic and stochastic evolution-type models is that they assume progressive waves, accounting for the evolution of incident propagating modes, but omitting locally excited evanescent modes and reflections. Consequently, in their present form, they do not predict the reflection of waves from steep shores and the nonlinear dynamics of the associated standing waves, and the excitation of refractively trapped low-frequency modes (e.g., edge waves).

To date, there is no comprehensive model formulation for fully directional wave–wave interactions over two-dimensional bathymetry, applicable to arbitrary scales of propagation and suitable for operational wave forecasting problems.

6. Bottom dissipation

Contributing authors: Jaak Monbaliu (jaak.monbaliu@bwk.kuleuven.be), Fabrice Ardhuin, Judith Wolf.

The dissipation terms in the wave energy equation are the least well-known. The wave energy balance equation explicitly contains a term for white-capping dissipation in deep water. As waves approach shallower water (depth \(< \lambda/2, \text{kh}<3\)) they start to ‘feel the bottom’ i.e. there is a non-negligible wave-induced oscillatory current at the sea-bed and the spectrum adopts a new self-similar shape in which enhanced dissipation is evident, e.g. TMA spectrum (Bouws et al., 1985). The wind input, nonlinear transfer and white-capping terms take different forms in depth-limited conditions and there is evidence for interaction of waves with the bottom. Several bottom-related dissipative processes are known: percolation into a porous bottom, motion of a mobile
bed or dissipation through turbulent bed shear stress with an associated bottom boundary layer (Weber, p. 156, in Komen et al., 1994). Earlier work suggested that the self-similar adjustment of the spectrum might be all that was required to account for, with no need for bottom friction (Resio, 1987), but Weber (1988) showed clearly that this is not the case and dissipation through bottom friction was required to complete the energy balance in shallow water. Most spectral wave models that take into account bottom dissipation as a source term, only model dissipation by bottom friction. Below a short overview is given with reference to the most common literature.

A process that is also worth mentioning is Bragg scattering from bottom irregularities, e.g. sand waves. However, this is not really a dissipative process but a process that in fact redistributes energy. This process is therefore covered elsewhere in this paper.

6.1. Wave energy dissipation due to bottom friction

Bottom friction is responsible for energy dissipation, which may reach a few watts per square meter, which is comparable to the energy input by the wind for moderate winds. Following Mirfenderesk and Young (2003), we can write $S_{bf}(k)$, the time rate of energy density loss at wave number $k$, as:

$$S_{bf}(k) = -\langle \tau_0 u_{h_k} \rangle,$$

where $\tau_0$ is the bottom shear stress and $u_{h_k}$ is the orbital velocity of the wave component with wave number $k$. Much work has been devoted to the detailed study of the bottom boundary layer structure, in order to obtain $\tau_0$, $u_{h_k}$ and other quantities as a function of the wave and current velocities away from the boundary, e.g. Grant and Madsen (1979), Christoffersen and Jonsson (1985), Wiberg (1995), Davies and Villaret (1999), Martin (2004). These models perform well against laboratory measurements, e.g. Jensen et al. (1989), for well-defined conditions (smooth and rippled beds with carefully controlled geometries).

6.1.1. Common formulations for spectral wave models: waves only

Luo and Monbaliu (1994) summarised the work done on the bottom friction term used in spectral wave models (here written in $(\sigma, \theta)$ space, $h$ is the water depth):

$$S_{bf}(\sigma, \theta) = -C_f \frac{k}{\sinh(2kh)} F(\sigma, \theta).$$

The coefficient $C_f$ depends on the closure model used to solve the momentum equations of the bed boundary layer. Of course, flow conditions and bottom roughness (friction factor $f_w$ or equivalent roughness $K_N$) are important parameters.

The following symbols in Table 1 are used: friction coefficient $c$; drag coefficient $C_f$; friction factor $f_w$; bottom roughness height $K_N$; kronecker delta $\delta_{ij}$; ensemble average $\langle \rangle$; bottom velocity components $U_i$ and $U_j$; $U$ is the magnitude of the bottom velocity vector; wave boundary layer friction velocity $u_*$; dimensionless function $T_{\kappa}$ and its complex conjugate $T_{\kappa}^*$, both dependant on the dimensionless argument $\zeta_0$ expressing the ratio between the roughness length and the wave boundary layer thickness.

In the field however a degree of complexity is introduced by the randomness of the wave field, e.g. Zou (2004), but more importantly the bottom is anything but uniform. As a result, a large part of empiricism must be introduced.

So far also only indirect validations have been performed, based on the recordings of wave attenuation between several wave gauges, rather than direct measurements in the wave bottom boundary layer. The val-

<table>
<thead>
<tr>
<th>Formulation</th>
<th>Cited coefficient value</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>JONSWAP $C_j = \frac{c}{\delta}$</td>
<td>$c = 0.038 \text{ m}^2 \text{s}^{-3}$</td>
<td>Hasselmann et al. (1973)</td>
</tr>
<tr>
<td>$C_{DHIC} = 2C_f {\delta_{ij}(U_i) + \langle \frac{U_i U_j}{\zeta_0} \rangle}$</td>
<td>$C_f = 0.015$</td>
<td>Hasselmann and Collins (1968)</td>
</tr>
<tr>
<td>$C_{DC} = 2C_f \langle \zeta_0 U_{ij} \rangle^{0.5} \delta_{ij}$</td>
<td>$C_f = 0.015$</td>
<td>Collins (1972)</td>
</tr>
<tr>
<td>$C_{DM} = \frac{1}{3} f_w \sqrt{2} \langle U_i \rangle^{0.5}$</td>
<td>$f_w$ or $K_N$</td>
<td>Madsen et al. (1988)</td>
</tr>
<tr>
<td>$C_E = u_* [T_{\zeta}^{\frac{1}{2}}(\zeta_0) + T_{\zeta}^{\frac{3}{2}}(\zeta_0)]$</td>
<td>$K_N$</td>
<td>Weber (1991)</td>
</tr>
</tbody>
</table>
idation of the coefficients given above is often discouraging, but many studies do not take into account variable roughness due to e.g. bedforms (see further). Hasselmann et al. (1973) found that $C_J$ inferred from the JONSWAP dataset of swell attenuations varied over two orders of magnitude, and (Young and Gorman, 1995) did not find a clear “winner” in their test of several parameterizations for the bottom friction source term. For the analysis of data from a field experiment of wave decay across the Great Australian Bight in order to determine the spectral decay which can be attributed to bottom friction, Young and Gorman (1995) used the spectral wave model WAM as an analysis tool. It is indeed necessary to take other processes such as atmospheric input, nonlinear interactions, whitecap dissipation, refraction, and shoaling into account since all these source terms are also active. However, this makes the extraction and interpretation of only bottom friction effects challenging.

6.1.2. Common formulations for spectral wave models: waves and currents

The effect of the interaction between waves and currents on the bottom stress is not completely solved. There is still some debate about whether the interaction is weak or strong (Kagan et al., 2005). In the concept of strong interaction both the wave and current bottom stress are enhanced due to nonlinear interactions. The bottom stress under combined waves and currents is larger than the sum of wave and current only. This approach was followed in the formulations of Grant and Madsen (1979) and Christoffersen and Jonsson (1985).

In spite of the importance of these effects on tidal current modelling, there seems to be little field validation of the wave–current theories. There is evidence that waves affect the bottom friction experienced by the mean flow, e.g., Wolf and Prandle (1999), Keen and Glenn (2002), Kagan et al. (2005), but there seems little or no observational nor theoretical evidence that the presence of currents affects wave friction in a substantial way, see e.g. Kagan et al. (2005), Weber (1991), Tolman (1992a). As pointed out by Kagan et al. (2005), it is possible that some effects cancel out.

6.1.3. Bottom roughness models for movable beds

Irrespective of the formulation used, some characterization of the bottom roughness is needed. This is directly evident in the models that require the friction factor $f_w$ or equivalent roughness $K_N$ as an input (see Table 1). In the case of sandy bottoms, bed forms may exist. These bed forms (ripples) are dependent on sand grain size, and on the existing, and the history of, previous hydrodynamic conditions, visible in wave and current ripples and/or relict ripples. There are strong evidences for an important role played by the wave-generated bedforms, so that the bottom roughness, at least over sand, appears more important than the details of the bottom boundary layer (Ardhuin et al., 2003a). In other words an adequate parameterization of the changing bottom roughness is probably more important than a choice between say Madsen et al. (1988) and Weber (1991). In a series of investigations on swell propagation over the shelf (Ardhuin et al., 2001; Ardhuin et al., 2003b), good agreement between modelled and observed wave spectra was shown when taking into account the variability of the bottom friction due to ripples. In contrast, although on average the empirical JONSWAP bottom friction term performed about equally well, it could not reproduce the weak swell decay in low energy conditions (JONSWAP dissipation too strong) nor the strong decay in high energy conditions (JONSWAP dissipation too weak).

There is a considerable amount of literature on friction factors for movable beds. They all relate hydraulic roughness to a combination of skin friction on individual grains and form drag due to bed forms. Bed forms here include ripples formed under oscillatory flow conditions including sheet flow conditions. In principle, once sand grain size is known and the hydrodynamic conditions are known, it is then possible to estimate hydraulic roughness of the bed and consequently also the energy dissipation in the wave field.

Tolman (1994) was probably the first one to investigate the effect of bed forms on bottom friction dissipation in a spectral wave model. He used a roughness predictor based on the sediment grain size and on a characteristic orbital velocity and characteristic orbital amplitude obtained as an integrated parameter from the wave spectrum. Tolman (1995b) even accounted for subgrid variability in sediment parameters in large-scale wind wave models.

In practice however roughness values or energy dissipation factors obtained from different experiments differ often by an order of magnitude or even more, see, e.g., Nielsen (1992). There is the combined effect of
waves and current on bed mobility, bed forms and suspended sediment concentration (Glenn and Grant, 1987). Also the dissipation process might differ depending on the bedforms involved. For example, in the case of large roughness elements or steep ripples in oscillating flow, the momentum transfer in the near-bed layer is dominated by the vortex-shedding process rather than by random turbulence, as pointed out by Nielsen (1992) and Sleath (1991). Detailed process models of the wave boundary layer over ripples that address the vortex shedding that is observed have been developed, e.g. by Malarkey and Davies (2004) and by Davies and Thorne (2005). Possibly these process models may yield new parameterizations for spectral models.

6.2. Energy dissipation due to wave–bottom interaction

Bottom friction is not the only process of importance for the dissipation of wave energy at the water–bottom interface. Shemdin et al. (1978) gave an overview of the different bottom interaction effects and next to friction dissipation discussed above, two other mechanism were discussed:

- damping due to percolation in a permeable bed layer
- absorption of energy in a bottom layer of soft mud

Both mechanisms have been worked out theoretically for more or less idealized cases. In the case of percolation, the dissipation of wave energy is due to the wave-induced pressure field at the bottom which in turns induces a flow in the permeable (sand) layer. The theoretical considerations can be found in Dean and Dalrymple (1984); Shemdin et al. (1978) and literature therein referred to. The wave energy damping rate is proportional to the permeability of the sediment layer and only significant for coarser sediments (grain size > 0.5 mm). For practical applications not only grain size, but also the thickness of the permeable layer needs to be known (thickness larger than 0.3 times the wave length can be considered as infinite according to Shemdin et al., 1978).

In the case of a soft muddy bottom, the energy dissipation is theoretically worked out using a two-layer model. The top layer is the water column and is treated as an inviscid fluid. The free surface wave will induce a wave at the mud–water interface which in turn will induce flow in the mud (lower) layer. The flow in the mud layer is damped rapidly by the high viscosity in the mud layer. The dissipation rate of waves propagating over mud bottoms is considerably higher than over sandy bottoms. For a more detailed description see, e.g., Dean and Dalrymple, 1984, Gade (1958), Hsiao and Shemdin (1980), Dalrymple and Liu (1978), Jiang and Mehta (1996), Ng (2000), and Cavaleri p. 169 in (Komen et al., 1994). Winterwerp et al. (2007) give a formulation for the energy dissipation source term which they used in the SWAN model.

6.3. Discussion and outstanding problems

Qualitatively, the process of energy dissipation due to interaction with the bottom seems relatively well understood. Theories exist for energy dissipation due to friction, percolation and water–mud interaction. Quantitatively however, our understanding is at least incomplete.

First of all, the location of the sea-bottom is in many cases not fixed in time. Fortunately, wave energy dissipation due to bottom friction is not a strong local process. The horizontal movement of a sand bank or shoal over a few hundred meter will not drastically affect the wave energy propagating over a shelf or coastal area. Second, our knowledge of the friction or damping characteristics of the sea bottom is limited. For example not the whole shelf sea has a sandy bottom. Quite often there are areas with sand, areas with rock and areas with mud. Bed forms are changing due to changing hydrodynamic conditions. Changes at subgrid scale are likely both in terms of sediment composition (grain size, percentage sand, etc.) as in terms of bedforms (Tolman, 1995b). Bedforms in mixed mud–sand sediments seem to be inhibited, but this is a poorly known topic. In that respect it is also interesting to quote from pp. 296–297 of Nielsen (1992): “Thus, energy dissipation measurements under waves indicate much greater hydraulic roughness for flat, movable beds than other types of experiments. . . . . Is it possible that the larger roughness indicated by the energy dissipation experiments can be explained in terms of the energy dissipation due to percolation under waves, which is not directly related to the effective bed shear stress. . . . .”. Also, in order to determine the spectral decay which can be attributed to
bottom friction, **Young and Gorman (1995)** analysed the data obtained from a field experiment in which seven wave measurement instruments were placed in a linear transect across the Great Australian Bight. The data available did not allow the relative contributions of bottom friction and percolation to be determined. Or in other words, in some cases we are not sure which dissipation process(es) we need to take into consideration.

Similarly, **Wolf (1999)** points out that it is difficult to make measurements of turbulent shear stresses in combined wave and currents, especially in the field. Laboratory data may not cover the whole range of phenomena, especially so when sediments are involved.

The energy absorption of soft muddy bottoms is not so well known. A standard source term for spectral models is not available. Undoubtedly, there is an effect of enhanced damping by cohesive sediment and vegetation over salt marshes, as has been demonstrated by e.g. **Möller and Spencer (2002)**. However, **Sheremet and Stone (2003)** state that “Contrary to the widely accepted hypothesis that mud-induced wave dissipation is important only for long waves, observations show significant damping of high-frequency, short waves, which interact weakly with the bottom.” The two layer concept with an inviscid fluid at the top and a viscous flow layer underneath is probably too simplistic to characterise a more gradual transition in fluid properties over the water column.

Probably the only way to make further progress in our understanding of wave dissipation due to interaction with a movable or soft bottom, will be through the combined study of the wave field and its effect on sediment motion. These are closely linked. But flow properties and sediment concentration close to the bottom, and in case of mud also visco-elastic properties of the bottom are difficult to measure. This is so in lab experiments, but even more so in field conditions.

### 7. Wave propagation

**Contributing authors:** Fabrice Ardhuin (ardhuin@shom.fr), Kostas Belibassakis, Igor V. Lavrenov, Rudy Magne, Hendrik L. Tolman.

Wave propagation is usually represented by the left-hand side term of the action balance equation, accounting for the well known effects of refraction, shoaling, diffraction and reflection. These effects typically dominate the variation of the wave field over narrow continental shelves, or the evolution of swells over very long distances in deep water. Large spatial scale variations in the depth and current may cause any of these effects, and the time evolution of the depth and current also lead to modifications in the wave field that may be less familiar to the reader but are generally included in spectral wave models. Although such effects are very well verified by observations for varying depths, there is still little validation of wave propagation over horizontally varying currents. Further, the action evolution equation over depth-varying current has not yet been given in the form used for depth-independent current.

In this section we make an overview of the problem, first considering it from a general point of view. Then we analyse in sequence the limitations of geometrical optics and the effects of varying current, both in space and time, with a final look at waves in the real ocean. The feedback of the waves on currents is beyond the scope of the present paper, but it should be noted that waves can generate very strong currents when they break in the nearshore, imparting their momentum flux to the currents. For all water depths, wave breaking is the largest source of upper ocean turbulence, with profound influences on current velocity profiles. The present analysis of wave evolution is thus only one piece of a bigger jigsaw puzzle, with strong interactions with turbulence and mean flows.

#### 7.1. Dispersion, geometrical optics and the wave action equation

Wave propagation at sea has been a subject of scientific interest for centuries. A comprehensive theory for monochromatic linear and nonlinear wave propagation was presented **Airy (1845)** and **Stokes (1847)**, with nonlinear effects specific to shallow water studied by **Boussinesq (1872)** and reviewed in Section 5. A spectral description of wind waves was introduced by **Pierson et al. (1955)** in order to account for the irregularity of waves at sea. In this description, the random wave field is broken into a spectrum of many regular wave components which are distinguished by wavenumber vector $k$, and relative radian frequency $\sigma$. Most wave fore-
casting models in use today consider that the wave components are inherently linear so that the wave properties are attributed to either a component $k = (k_x, k_y)$ or the pair $(\sigma, \theta)$ with $\theta$ the direction normal to the wave crests, which is the direction of the vector $k$. The wavenumber magnitude $k$ and $\sigma$ are related by the dispersion relation for linear waves. Typically, the relative phases of wave components are taken to be random and uniformly distributed so that only the amplitude information of the spectra is retained (dropping phase information). These amplitudes are then translated to a quadratic wave quantity that may be either the spectral density of energy $q_w g E(k)$, momentum $M_w(k) = \rho_w g E(k) / \sigma$ or action $A(k) = g E(k) / \sigma$, where $\rho_w$ is the water density, $g$ is the apparent acceleration of gravity and $E(k)$ is the surface elevation variance spectrum. This definition of the action is not applicable in the case of strong nonlinearity or vertical curvature in the current profile. For these cases one may use the more general definition given by Andrews and McIntyre (1978).

The integral of the spectrum is the local elevation variance $E$, from which the significant height may be evaluated as $H_s = 4E^{1/2}$. Thus the usual maps of $H_s$ produced by phase-averaged wave models (Fig. 8a) correspond to an “average” wave height (trough-to-crest) of a moving surface, such as represented in Fig. 8b.

![Diagram](image)

**Fig. 8.** Different representations of wave propagation over the Scripps-La Jolla submarine canyons. (a) Significant wave height computed for 15 s swells from the West with SWAN (calculation performed by E. Rogers). (b) Solution of the mild slope equation for 1 m amplitude 15 s monochromatic swells, using the model of Athanassoulis and Belibassakis (1999). (c) Wave rays for 15 s with a constant offshore direction from the West and an offshore spacing of 20 m. (d) Backward ray-tracing for the same wave period but for all directions arriving at buoy number 34.
The spectrum \( E(k) \) may be obtained by a Fourier transform of the surface elevation field. These spectral densities generally vary in space and time so that the Fourier transform is generally implicitly replaced by a Fourier transform of the surface elevation autocorrelation function, or a Wigner distribution (Wigner, 1932). It should be noted that the local directional spectrum may have no physical meaning for wave amplitudes that vary slowly in time but rapidly in space. For example, in Fig. 8b it is impossible to define a wave crest in some regions of rapid variation of the topography.

Wave forecasting thus reduces to a quantitative determination of the evolution in space and time of the action spectrum taken as \( N(k)=A(k)/g \) since \( p_w \) and \( g \) are generally regarded as constants. The evolution equation for \( A \) is intrinsically simpler than those for \( E \) or \( M^w \) because the total action conservation is related to the invariance of the physical problem when the wave phases are changed, which is generally the case in intermediate and deep water, in the absence of wind–wave generation, dissipation or triad interactions. This relation is known as Noether’s theorem. In the same way, the conservation of total energy or wave momentum is related to invariances by translations of the medium in time or space, respectively, which is generally not the case (e.g., Andrews and McIntyre, 1978). A general equation for the total action can be derived for any wave field in terms of the wave-induced pressure, velocity and displacement fields (Andrews and McIntyre, 1978). This includes in particular the Earth rotation or current shears that make the wave motion weakly rotational. An explicit approximation may be given for slowly modulated small amplitude waves (Bretherton and Garrett, 1968) with a corresponding equation for the spectral action density (Hayes, 1970; Komen et al., 1994),

\[
dN(k)/dt = \partial N(k)/\partial t + \mathbf{V} \cdot [(C_g + U_A(k))N(k)] + V_k \cdot [C_k N(k)] = S_{\text{tot}}(k),
\]

where \( C_g = k(\partial g/\partial k)/k \) is the intrinsic group velocity, and \( U_A(k) \) is an advection velocity that depends on the current profile (Andrews and McIntyre, 1978), and also on the amplitudes of all wave components (Willebrand, 1975). These latter effects may often be neglected in practice, and \( U_A(k) \) can be approximated to be the surface drift current or, for shallow-water waves, the depth-averaged drift current. The divergence operator \( \mathbf{V} \cdot (\cdot) \) is the classical divergence restricted to the horizontal directions only, and \( V_k \cdot (\cdot) \) is a similar divergence operator in spectral space. The spectral advection velocity \( C_k \) represents the turning of the wave crests (refraction) and change in wave length (shoaling). Explicit expressions for \( C_k \), or their equivalent for the pair \( (C_k, C_\theta) \), can be obtained in terms of the gradients of the water depth and \( U_A(k) \), using the hypothesis of slow modulation (e.g., Keller, 1958; Mei, 1989). This approximation is also called WKB or ‘geometrical optics’ approximation, due to its use in the theory of light refraction. The equations for \( (C_k, C_\theta) \) are usually called ‘ray equations’, and form the basis of the advection part in phase-averaged wave models. Finally, the action of each component is allowed to evolve as it propagates, with a rate of change given by the total source term \( S_{\text{tot}} \).

In the deep ocean, spectral components are indeed shown to propagate according to (7.1). The rotation of the Earth has a negligible influence on wave propagation (Backus, 1962), and individual wave components of the spectrum travel along great circles until they reach shallow water, strong currents, or coast lines. This was powerfully demonstrated by Snodgrass et al. (1966), who followed wave propagation a third of the way around the globe. In shallow water, and in the absence of significant currents, the validity of (7.1) was also demonstrated by Munk and Traylor (1947), and many following studies. As a matter of fact, the propagation of long period swells over a relatively narrow continental shelf with no significant current is generally well described by (7.1), with the right-hand side set to zero outside of the surf zone (e.g., O’Reilly and Guza, 1993; Peak, 2004).

The two forms of the left-hand side of (7.1), the Lagrangian derivative following a wave packet, or the Eulerian derivative plus divergence of action fluxes, are rigorously equivalent. The first form was originally proved for steady conditions without current (Longuet-Higgins, 1957) and is easily used in forward or backward ray-tracing methods of solution (Fig. 8c and d, respectively). Such rays represents the propagation paths of monochromatic wave packets. According to geometrical optics, the energy flux of such monochromatic waves is constant between two rays that are integrated from an area where rays have parallel directions, and where the phase speed is constant (e.g. deep water and uniform current). As a result the spatial energy density varies as the inverse of the distance between the two rays, going to infinity where the rays cross. This location is called a ‘caustic’. In reality the wave height is finite because any given component of the spectrum carries an infinitesimal fraction of the wave energy, and the caustics occur in different places for the different components. Nevertheless the presence of such caustics generally correspond to higher waves for narrow wave spec-
tra. The presence of such caustics makes the backward ray-tracing method more practical (e.g. O’Reilly and Guza, 1993).

The second form of the left-hand side of (7.1) is most easily translated into methods with discretized physical and spectral spaces, and where action transports based on characteristic velocities are considered in each space (see Section 8). The wave action balance can be constructed for any spectral coordinate system, as long as the Jacobean transformation from \( N(k) \) to the alternative description of spectral space is well behaved (Tolman and Booij, 1998, Appendix A). It should be noted that this is generally not the case for spectra described in terms of the absolute frequency and direction, because the corresponding Jacobean has a singularity at the blocking point. Numerical issues that arise in the solution of (7.1) are discussed in Section 8.

Current effects represented in (7.1) are still poorly validated quantitatively, even by laboratory experiments, and some observed effects of currents are not understood.

### 7.2. Limitations of geometrical optics: diffraction, reflection and random scattering

Whenever the water depth \( D \) or the current change on the scale of the wavelength, deviations from geometrical optics are expected. As a result, even a purely monochromatic wave train would have a finite wave height at a caustic predicted by geometrical optics (Fig. 8c). A classical example is the propagation of waves past a semi-infinite and absorbing breakwater with the wave field diffracted behind the breakwater (e.g., Penney and Price, 1952). A general representation of the variation in the wave field at the scale of the wavelength requires a phase resolving model that accounts for the interference patterns, particularly in areas of crossing wave rays (e.g., Berkhoff, 1972; Dalrymple and Kirby, 1988; Athanassoulis and Belibassakis, 1999). However, the representation of the effect of diffraction on scales larger than the wavelength can be included in (7.1) by a proper modification of \( C_0 \) (e.g. Holthuijsen et al., 2003). For natural topographies, the geometrical optics approximation is generally quite robust. This fact was confirmed by the 2003 Near Canyon Experiment, off La Jolla, California, where the bottom slopes reach 3/1 on the walls of Scripps canyon (Peak, 2004). Over such a topography, diffractions from electromagnetic optics are significant only in a small area around the head of Scripps canyon, where the wave height has been found to vary by a factor up to 5 over about half a wavelength (Magne et al., 2007).

Where the water depth goes to zero, on the shoreline, waves are partially reflected. This can be represented by empirical reflection coefficients in phase-resolving models for wave propagation around artificial coastlines, but it may also be important on natural shorelines, including beaches (Elgar et al., 1994) and cliffs (O’Reilly et al., 1999). Such a reflection may be introduced in phase-averaged models as a proper boundary condition for (7.1). Partial wave reflection also occurs over any bottom topography. This is generally negligible, but significant reflection occurs when the depth changes on the scale of the wavelength (Heathershaw, 1982; Elgar et al., 2003). This phenomenon is formally similar to the scattering of long electromagnetic waves over the ocean surface, a phenomenon used for mapping sea surface currents with High-Frequency radars. For linear waves, it can be represented by a Bragg-like bottom scattering source term \( S_{\text{bscat}} \) in the right-hand side of (7.1) (Ardhuin and Herbers, 2002; Ardhuin and Magne, 2007),

\[
S_{\text{bscat}}(k) = \int k'^2 M^2(k,k')F^B(k-k')|N(k') - N(k)|/[\sigma^'(k'C'g + k' : U_A)] \, \mathrm{d}k',
\]

where the primed variables correspond to the component \( k' \) with direction \( \theta' \), such that they satisfy the resonance conditions \( \sigma' = \sigma + l : U \) and \( k = k' + l \). For weak currents the wave–bottom coupling coefficient is \( M = gk : k'/[\cosh(kD) \cosh(k'D)] \). The bottom is represented by its double-sided spectrum \( F^B \), such that its integral over the entire wavenumber plane yields the bottom elevation variance.

This source term accounts for the interaction of triads involving two wave trains and one bottom Fourier component. In the form given by Ardhuin and Magne (2007), it also includes interactions between two waves and one Fourier component of the current or surface elevation that arises from the adjustment of a mean current to the topography. The interaction conserves wave action, but wave momentum is not conserved, which results in a mean recoil force acting on the bottom. This scattering theory over the current fluctuations should be consistent with the theory of Bal and Chou (2002) for the scattering of gravity–capillary waves over depth-uniform and irrotational current fluctuations.
The relative accuracy of reflections coefficient given by $S_{\text{bscat}}$ was found to be proportional to the ratio of the bottom amplitude and water depth, regardless of bottom slope. Reflection coefficients may thus be obtained from any bottom topography of small amplitude, including steps with vertical walls, as can be seen by the correspondence between the Green function method and Fourier transforms (Elter and Molyneux, 1972; Mei and Hancock, 2003; Magne et al., 2005). The source term does not give accurate results, however, when particular phase relationships exist between interacting waves and bottom undulations, e.g. in cases with waves propagating over nearshore sand bars and reflecting over the beach (Yu and Mei, 2000). Except for such conditions, the evolution of wave action over scales larger than the bottom autocorrelation length, is very well predicted by $S_{\text{bscat}}$, in agreement with phase-resolving models for wave propagation in one dimension (Mei, 1985; Kirby, 1988). On natural continental shelves with bottom elevation variances of the order of 1 m² for scales in the range of 0.5–5 times the wavelength of dominant surface gravity waves, this scattering term yields a strong broadening of the directional spectrum over a few kilometers for $kD \approx 1$. This predicted broadening was confirmed by observations of the evolution of narrow offshore directional wave spectra across the North Carolina shelf (Ardhuin et al., 2003a,b), although it accounted for only half of the broadening of relatively broad offshore spectra. For organized bottom topographies such as the sandwave fields found in the southern North Sea (Fig. 9), a strong broadening is expected for narrow swell spectra, with an additional weak reflection. Taking the bottom spectrum shown in Fig. 9, and the mean water depth of $D = 20$ m, a simple calculation was performed with a one-dimensional WAVEWATCH III model, using periodic boundary conditions in the North–South direction. For an incident narrow spectrum (Fig. 10b) the source term shape and magnitude depends on the current strength and direction (Fig. 10c and d). The wave field rapidly evolves to a broader directional distribution (Fig. 10e).

This type of scattering by random media perturbations is quite general (e.g., Ryzhik et al., 1996) and can be extended to other current perturbations that may be rotational and unrelated to the topography. Although Laplace’s equation does not hold in that case, one may use an equation for the pressure (e.g., Kirby and Lee, 1993) or a forced Laplace equation (e.g., McWilliams et al., 2004), or work from the Hamiltonian. With that approach a scattering source term was derived by Rayevskiy (1983) for waves over random current and a corresponding diffusion approximation was derived. This effect is found to be potentially important and was further studied by Fabrikant and Raevsky (1994). For example, these authors found that unidirectional waves of 40 m wavelength in a drift flow of a few centimetres per second evolve into directional waves with a spread of $6^\circ$ in a few kilometers of propagation.
For large bottom amplitudes or steeper waves, higher order interactions are expected to be relevant (Liu and Yue, 1998). Such interactions have been observed for periodic and one-dimensional bottom topographies (Rey et al., 1996). The bottom topography may act as a catalyst, making near-resonant triad wave interactions (see Section 5) exactly resonant. Again, the proper form of the higher order scattering term is yet to be derived for random waves.

7.3. Waves over varying currents, nonlinear wave effects and the advection velocity

A proper description of wave propagation over currents is not only necessary for the forecasting of waves over large-scale currents such as the Gulf Stream or the Agulhas current, or tidal currents on continental shelves. It is also a key element for the interpretation of remote-sensing observations. This applies to microwave radar or radiometers from satellites, used for measuring anything from sea surface height, current and wave heights, to sea surface salinity and winds. In that case the instrument is sensitive to short (few centimeters) waves that are modulated by the orbital velocities of the longer waves, with additional effects of surface slopes and accelerations (see, e.g., Henyey et al., 1988, and Elfouhaily et al., 2001). This also applies to High-Frequency radars, a now popular instrument for mapping coastal currents, or more specifically the phase velocity of a waves of a given wavenumber \( k \). That velocity is the intrinsic phase speed \( k\sigma/k^2 \) plus the advection velocity \( U_A(k) \) given by Kirby and Chen (1989),

\[
U_A(k) = \int U(z) k \cosh[2k(z + H)]/\sinh(2kD) \, dz, \tag{7.3}
\]

where \( -H \) is the mean elevation of the bottom and \( D \) is the mean water depth, which is \( H \) plus the mean surface elevation. Based on the theory of Andrews and McIntyre (1978), the advection velocity for the action has the same expression, but the action may differ slightly from the definition given above. This question needs further research.
The drift current $U^L$ may be approximated by the mean current velocity $\bar{u}$. However, a more general approximation for not-so-small waves is, $U^L = \bar{u} + U_s$, where $U_s$ is the Stokes drift due to the entire wave field. In the case of short wave advection by long waves, with a clear scale separation, Broche et al. (1983) showed that (7.3) is consistent with the theory of Weber and Barrick (1977). In remains to be proved that (7.3), or a more accurate version of it, is also consistent with the known amplitude dispersion of Stokes waves, or other theories for the dispersion of waves in a random wave field (Hayes, 1970; Willebrand, 1975; Huang and Tung, 1976; Masuda et al., 1979), short wave modulation by long waves (Phillips, 1981b), and finite amplitude waves over shear currents (e.g., Dalrymple, 1974; Peregrine, 1976).

In practice wave models and most users of HF-radar assume that $U^L$ is uniform over the depth. This is probably a good approximation for swells propagating over large-scale geostrophic, tidal or wind-driven currents, as the current velocity $\bar{u}$ is generally uniform close the surface due to the strong mixing induced by wave breaking (Santala and Terray, 1992; Terray et al., 2000). However, a differential advection of shorter waves by the sheared Stokes drift is to be expected. For reference, $U_s$ at the surface is typically 1–1.5% of the local 10 m wind speed for fully-developed waves. Further, the advection velocities of short and long waves propagating in stratified estuaries are markedly different due to vertical shears of $\bar{u}$. This has been shown for the advection of the wave phase (Ivonin et al., 2004), but such a validation is still lacking for the wave action advection velocity. The common practice of using the surface velocity is expected to be generally valid.

The wave action advection velocity $U_A$ is known to modify the wave heights by a combination of three effects. We consider monochromatic waves for the sake of simplicity. First of all, the conservation of the wave action flux means that in cases of along-crest uniform conditions, a gradient of $U_A$ in the direction of propagation should result in a change of the local action density in order to keep $(C_g + U_A)A$ constant. Specifically, for waves against an increasingly strong current, $C_g$ is reduced as the wavelength gets shorter and $C_g + U_A$ is made even smaller by the change of $U_A$. Second, the change in surface elevation variance $E = A/\sigma$ is amplified compared to $A$ due to the change in the intrinsic frequency $\sigma$. Third and last, $U_A$ generally varies along the wave crests so that current-induced refraction leads to further local increases of wave heights for waves propagating against a current jet. For weak current shears the current-induced refraction gives a ray curvature radius equal to the ratio of the wave group speed and the current vertical vorticity (Landau and Lifshitz, 1960), giving a scale over which refraction becomes significant. This combination of effects for the wave height and the associated change in wavelength makes current fronts a preferential site of wave breaking. Current jets, from large scales to river mouths are one of the most hazardous areas for navigation (e.g., Gutshabash and Lavrenov, 1986; Masson, 1996).

Practical wave forecasting in which currents are taken into account are, to this day, limited to tidal currents (e.g. at the UK Met Office). Quasi-geostrophic currents are probably not observed or predicted well enough in order to perform these calculations. This may change with the advent of absolute measurements of the ocean dynamic height, using the latest high-resolution measurements of the geoid, and improvements in ocean circulation models. Large benefits are expected for the forecasting of extreme waves.

On smaller scales, when current variations are significant over one wavelength, partial wave reflection occurs. The two cases of current discontinuity (Evans, 1975) and slowly varying current (McKee, 1974) have been well investigated. A Mild Shear Equation analogous to the Mild Slope Equation was derived and extended by McKee (1996). Effects of evanescent modes have also been considered by Belibassakis and Athanassoulis (2004). Partial wave reflection may be relevant for the wave–current interactions that occur in Langmuir circulations (Smith, 1980b; Veron and Melville, 2001), the essential mixing engine in the ocean mixed layer. Indeed, the vortex forces that are generally believed to drive Langmuir circulations only exists as a compensation of the divergence of the wave momentum flux that occurs when waves refract over the current pattern (Garrett, 1976). Analytical solutions suggest that such reflections are generally weak for typical dominant wind waves with periods of a few seconds, except for grazing incidence angles. However in that case the effect is minimal since the reflected and incident wave directions are almost identical.

### 7.4. Waves blocking

Wave blocking occurs where opposing currents are sufficiently strong to stop wave propagation in physical space, i.e. where $C_g + U_A(k) = 0$. In a traditional monochromatic geometric optics approach, a singularity
occurs in the wave energy equation at the blocking point, where the wave action and energy fluxes converge. However, Shyu and Phillips (1990) have shown that a continuous solution exists on both sides of the blocking point. Furthermore, a spectral approach leads to continuous non-crossing characteristics in \( (x - k) \) space, indicating that no singularity exists in a spectral description of wave propagation. Laboratory observations of wave blocking (Lai et al., 1989; Chawla and Kirby, 2002; Suastika and Battjes, 2005) clearly validate the concept of a blocking point. However, the mechanism by which the wave energy is ‘removed’ at the blocking point does not seem to be understood yet.

7.5. Unsteady water depths and currents

Traditionally, waves propagating over stationary currents have been considered. This approach is valid for conditions where the currents are (quasi-) stationary on time scales comparable with the propagation time of waves through the area. This is generally the case for persistent deep-ocean currents like the Gulf Stream, or for current patterns related to bathymetric features such as shoals, headlands and inlets. However, currents on the continental shelf are often largely due to tides. A free travelling tidal wave travels much faster than a wind wave and therefore results in a quasi-homogeneous rather than quasi-stationary current field. Such temporal variations of currents result in Doppler shifts only (Tolman, 1990). In many practical applications, interactions occur due to both spatial and temporal variations of the current field (Barber, 1949; Tolman, 1991).

7.6. Waves in the real ocean

The occurrence of other types of motions (e.g. internal waves) or special boundary conditions (sea ice, surfce films) have significant effects over the wave motion. Although such situations are frequent, they are generally neglected except for the effect of sea ice. Ice is as a powerful attenuator of waves propagating from the open ocean (Wadhams, 1978) and generally prevents any wind–wave generation of significance to the ice-free ocean (Crocker and Wadhams, 1988). Still, 1 m high swells have been observed to break up the ice as far as 500 km from the ice edge, making navigation difficult (Liu and Mollo-Christensen, 1988). As for the other conditions, there is clear evidence of attenuation of waves by oil poured on the sea surface, an ancient technique for ship rescue operations. Theory on surface waves–internal wave interactions lead to possible large changes in the surface wave energy (Kudryavtsev, 1994) with observed significant wave generation by large amplitude internal waves (Osborne and Burch, 1980). Further research on these processes is clearly needed, with an evaluation of their impact in numerical wave models.

8. Numerics and resolution in large-scale wave modelling

Contributing authors: W. Erick Rogers (rogers@nrlssc.navy.mil), Fabrice Ardhuin, Igor V. Lavrenov, Hendrik L. Tolman.

Although most efforts have been devoted to the understanding of the physical processes responsible for the evolution of the wave action spectrum, mathematically represented by Eq. (8.1), it must be recognized that the choice of a numerical method for arriving at the solution may be the source of large errors in the results (Tolman, 1992b). In this section, we provide a description of the basic problem, related to both the finite description of the physical world and to the time step integration. We analyse the related existing solutions. Then we discuss the relative importance of the various sources of error, with a look at the future.

8.1. A description of the problem

Two fundamentally different approaches have been used for solving (8.1). The ray method, using backward ray-tracing to avoid caustics, is convenient and very efficient for steady media, where the rays need to be computed only once (e.g., Cavaleri and Malanotte-Rizzoli, 1981).

The advantages of the grid method are that conservation of action can be enforced rigorously, and that the inclusion of nonlinear source terms is straightforward with a splitting of the integration time step in advection
and source term integration. Problems with grid methods are that high spatial resolution is required for an accurate description of bathymetric and some current effects, and high spectral resolution is required for accurate swell propagation over large distances. It should be noted that where the same governing equation is solved, the different numerical methods should ideally produce the same result, as demonstrated in Holthuijisen and Tolman (1991), for example.

8.1.1. Error due to the numerical scheme for geographic propagation on a grid

For the purpose of discussion, we pose the wave model governing equation in one-dimensional form, with uniform group velocity, no source terms, and only one spectral component (frequency/directional bin) considered:

$$\frac{\partial}{\partial t} N + C_{gx} \frac{\partial N}{\partial x} = 0.$$  (8.1)

When this continuous equation is discretized using finite differencing, numerical error occurs. To give an example, if one uses the explicit, first-order upwind scheme of the WAM model, the numerical error (or truncation error) is the right-hand side of the following equation (from Petit, 2001):

$$\frac{\partial}{\partial t} N + C_{gx} \frac{\partial N}{\partial x} = \frac{1}{2} C_{gx} \Delta x (1 - \mu) \frac{\partial^3 N}{\partial x^3} - \frac{1}{6} C_{gx} \Delta x^2 (1 - \mu)(1 - 2\mu) \frac{\partial^5 N}{\partial x^5}$$

$$+ \frac{1}{24} C_{gx} \Delta x^3 (1 - \mu)(6\mu^2 - 6\mu + 1) \frac{\partial^7 N}{\partial x^7} + O(\Delta x^4),$$  (8.2)

where $\mu$ is the Courant–Friedrichs–Lewy (CFL) number, $\mu = C_{gx} \Delta t / \Delta x$. Thus, we can see that numerical geographic propagation error is dependent on several dimensional quantities:

1. geographic resolution,
2. the time step,
3. the speed of propagation, and
4. the curvature of the field of spectral density, and/or various other spatial derivatives of this field.

This may also be posed as a dependence on two dimensionless quantities:

1. the CFL number, which quantifies the number of grid spaces traversed by a packet of energy in one time step, and
2. the geographic resolution relative to the scale of the feature in the wave field which is being propagated. (In (8.1), “scale” would be in the $x$-space.) All else being equal, a larger-scale feature will have smaller gradients; if these gradients are small, the numerical error will tend to be small.

8.1.2. Diffusion

The term “numerical diffusion” is used in this paper to describe the unintended spreading or smearing of wave energy during propagation due to discretization of a continuous problem. More specifically, it is due to even-ordered truncation error terms in the governing equation finite differencing associated with propagation, e.g. (8.2). The behaviors of individual schemes are rather unique. With any proper numerical scheme, diffusion becomes small at very high resolutions, but does not necessarily do so in a monotonic fashion. Dependence on CFL is even more varied. For example with increasing CFL, diffusion of the implicit first order upwind scheme (for most resolutions) will increase, while diffusion of the explicit first order upwind scheme will decrease until it becomes zero at $\mu = 1.0$. In the two-dimensional case, some schemes are more sensitive to propagation direction than others; the first order scheme of WAM is especially notorious for this (see Fig. 1 of WAMDI Group, 1988).

In some computational fluid dynamics literature, this diffusion is referred to as “dissipation”. We do not use the term “dissipation”, since that would improperly imply to most wave modelers a loss of energy. Diffusion does not directly cause a loss of energy: the numerical schemes of widely-used 3G wave models (WAM, WW3, SWAN) are energy-conserving.
8.1.3. Numerical dispersion

Numerical dispersion is the practical effect of the odd-ordered truncation error terms in the governing equation finite differencing associated with propagation, e.g. (8.2) (In the context of a wave model, one should specify that this is numerical dispersion, to avoid confusion with physical dispersion.) Like diffusion, numerical dispersion is dependent on CFL number and relative resolution. Consider a geographic feature in the wave spectral density field as a “signal” being propagated. Due to the discrete representation of a finite difference model, the celerity of different Fourier components of this signal deviate from the proper celerity of the signal (which is the group velocity calculated by the wave model); thus the Fourier components “disperse” as this feature is propagated in the model.

8.1.4. Combined effect of diffusion and dispersion

The error in celerity tends to be greater for the shorter Fourier components. As numerical dispersion occurs, two things can happen to the shorter components: they can either be smoothed by numerical diffusion (merging with the longer components), or become visible in the solution. In the latter case, the components are referred to as numerical oscillations or “wiggles”. The wiggles do not indicate model instability, but they do have an entirely unnatural appearance, and should therefore be prevented. The most straightforward way to do this is to employ a numerical scheme which tends to produce dispersion and diffusion in roughly equal portions; another method is to intentionally add diffusion as a separate term in the governing equation (denoted below “controlled diffusion” to distinguish it from “numerical diffusion”, which is a type of error).

8.1.5. Error due to the numerical scheme for spectral propagation

Like propagation in geographic space, propagation in spectral space is treated with finite differencing methods in all widely-used 3G wave models. As such, it is subject to the same types of numerical error (diffusion and dispersion).

8.1.6. Error due to coarse geographic resolution

We have already mentioned that geographic resolution has a strong influence on numerical error (diffusion and dispersion). It can also affect model accuracy in a manner not directly related to numerics. This tends to be most noticeable in shelf-scale and nearshore applications, but can also be apparent in large-scale models. In the latter case, if an island or peninsula is not well represented by the computational grid, then the blocking and scattering of wave energy by this land mass will not be well represented. Present-day global wave models are computed at 0.5–1.5° resolution; at these resolutions, some island groups will not be represented at all in the computational grid, which will lead to a persistent underprediction of the blocking/scattering of energy.

In cases where high resolution (finer than 1°) ocean-scale wind forcing is available, there may be some benefit to running the wave model at comparable resolution. This, of course, depends on the scale of meteorological features and the wave model sensitivity to these features.

8.1.7. Error due to coarse spectral resolution

When spectral (frequency/directional) discretization is too coarse for the scale of propagation, non-physical discontinuities manifest in the wave field as natural dispersion occurs. In the extreme case, as a propagating swell field propagates, it disintegrates into discrete geographic features, with each feature corresponding to a frequency/directional bin in the model’s computational grid. This is known as the “garden sprinkler effect” (GSE) (e.g., SWAMP Group, 1985).

With higher order propagation schemes, the GSE is unfortunately more apparent. Numerical diffusion, though it is an error, has the positive quality of tending to counteract the garden sprinkler effect, smoothing these discrete features together. Note, however, that numerical diffusion in existing models is unrelated to physical dispersion and is not controlled, so it does not properly mimic the natural dispersion of continuous spectra. In fact, the GSE can be clearly observed in WAM predictions of very old swell fields, despite the diffusive first-order scheme of WAM.

The GSE is not limited to propagation of swells across great distances: the directional GSE (i.e. the part of GSE related to directional discretization of the wave spectrum) is sometimes seen in the lee of islands: in these cases, the gradation between “illuminated” and “shadowed” areas is stepwise, rather than smooth.
Tolman (1995a) demonstrates that the conventionally used frequency resolution may be inadequate insofar as the spectral peak is not well represented during the growth stage, leading to incorrect dispersion of resulting swell (this is not the GSE, so it is not addressed by GSE-correcting methods described in Section 8.2).

8.1.8. Errors in source term integration

The integration in time of the source terms is usually performed in a separate ‘fractional step’ of a wave model. In this step, the following equation is solved

$$\frac{\partial F}{\partial t} = S.$$  \hspace{1cm} (8.3)

The simplest way to solve this equation is a simple first order Euler approach

$$F_{n+1} = F_n + S \Delta t,$$  \hspace{1cm} (8.4)

where \(n\) is the discrete time counter and \(\Delta t\) is the discrete time step. The main difficulty with integrating the source terms in time is the inherently small time scales involved with this process, particularly at higher frequencies. When the simple Euler approach is used, the attainable time step is at the best of the order of minutes. For early third generation wave models, this was unacceptable, and methods were developed to be able to integrate the source terms with time steps of about 20 min. The WAM group (WAMDI Group, 1988) solved this problem in two ways. First, the Euler approach of (8.4) was replaced by a semi-implicit method

$$F_{n+1} = F_n + \frac{S}{1 - \alpha D\Delta t} \Delta t,$$  \hspace{1cm} (8.5)

where \(D\) represents the diagonal contributions of the partial derivative of \(S\) with respect to \(F\). The parameter \(\alpha\) represents the centricity of the scheme. Originally, it was set to \(\alpha = 0.5\), making this scheme central in time. More recently \(\alpha = 1\) is favored. This makes the scheme lower in order, but increases the stability of the integration. Particularly, oscillations are avoided at higher frequencies, as the scheme more properly represents a root finder for the quasi-steady solution that dominates the source term integration in the equilibrium range of the spectrum (e.g., Hargreaves and Annan, 2001). Introducing the semi-implicit scheme is not sufficient to allow for time steps of the order of 20 min. The WAM model therefore used a so-called limiter, which sets a maximum allowable (absolute) change \(\Delta F\) per time step \(\Delta t\). The combination of the semi-implicit scheme and the limiter resulted in stable model integration with large time steps. The limiter of WAM Cycles 1–3 had the favorable characteristic of not affecting the solution for small time steps (in other words, as the time step size approaches zero, the solution will “converge” to the solution of the model without a limiter). However, since the limiter was not prescribed as a function of time step size, the effect of the limiter was shown to be rather sensitive to the time step size, particularly for initial wave growth (Tolman, 1992b).

8.2. Existing solutions

8.2.1. Improved numerical schemes for propagation on a grid

Both WAM and WW3 have, or have been used with, higher order schemes which can be employed instead of the explicit, first order upwind scheme. The higher order scheme used in WAM is the second order leapfrog scheme (which has zero numerical diffusion). The higher order scheme of WW3 is the “ULTIMATE QUICK-EST” scheme and limiter (see Tolman, 1995a; Leonard, 1979, 1991; Davis and Moore, 1982). The QUICK-EST scheme is third order when solved in one dimension. In the case of WAM, the higher order scheme is not part of the present official version and is, to our knowledge, rarely used. In the case of WW3, it is the suggested numerical scheme (used for both geographic and spectral propagation). So-called “total variance diminishing” limiters can be used to control wiggles (the reader is referred to Leonard (1991) for a description of the scheme used in WW3 and Fletcher (1988) for a general overview).

Numerical schemes are sometimes presented in the literature in one-dimensional form. For application in a wave model, they must obviously be extended to two geographic dimensions. There exist more than one method for doing this, and the efficacy of the extension method will depend on the scheme being extended. One method is to solve for propagation in both dimensions simultaneously, with each finite difference term...
being equivalent to the one-dimensional form. A second method – the fractional step method – is to propagate each dimension in sequence, with each operation being identical to the one-dimensional operation (see e.g. Yanenko, 1971). In either of these methods, the order of accuracy of the scheme should be expected to decrease compared to the one-dimensional equivalent. A third method is the “Product Generalization” method of Petit (2001) which preserves the order of accuracy of the one-dimensional scheme; this extension method may be prohibitively expensive in many cases.

The SWAN model uses an implicit propagation scheme which is second order when solved in two dimensions; this model is less efficient than WAM or WW3 applied at oceanic scales. Other schemes have been proposed, such as semi-Lagrangian schemes (e.g., Lavrenov and Onvlee, 1995), using analytical ray-tracing solutions to search the grid for the position of wave packets at the previous time-step.

### 8.2.2. Alternatives to the finite difference schemes on a grid

The oldest wave forecasting models have been based on the propagation of energy or action along wave rays, which are geodesic lines in deep water. The ray method is ideally suited for addressing long-distance swell propagation, for which the source terms on the right-hand side of (8.1) are small, with minimal numerical accuracy issues. Among such models, The Navy Swell Model (Hsu et al., 2004) is a ray-tracing model; propagation within this model contains virtually no error associated with numerics and resolution. This model contains no source/sink terms (e.g. swell attenuation is not represented). This model is initialized using spectral density fields from WAM simulations. The initialization fields thus are affected by numerical error of the input model. Nevertheless, it is useful for creating long-range swell forecasts and as a research tool.

Ray-based advection also disconnects the advection along rays, which are different for each component, from the grid where the wave spectra are put together. This allows an easy use of unstructured grids (e.g., Benoit et al., 1997), or no grid at all; some models are used for forecasts at a single point (e.g., the Navy Swell Model, Hsu et al., 2004). The main advantage of ray-based advection is that it does the spectral and spatial advection in a single step, with virtually no numerical diffusion. Some numerical diffusion is still introduced if the same rays are not followed all the way to the model boundaries (Ardhuin and Herbers, 2005).

A potential drawback of ray-based advection is the greater difficulty of ensuring a conservation of the total wave action over a given area, although there is no evidence of a lesser accuracy on modelled wave parameter. However, the largest challenge is the integration of source terms if the rays are followed over more than one time step in order to benefit from the low diffusion. Several levels of complexity have been tested. Cavaleri and Malanotte-Rizzoli (1981) have thus restrained their representation of source terms to parameterization that are local in the spectrum, i.e. \( S(k) \) is a function of \( A(k) \) and external parameters only (see also Lavrenov, 2003b). In order to be able to use generic parameterization, the rays must be linked to the grids where the spectra are assembled. This was done by Ardhuin et al. (2001) for application to bottom friction and that approach has been used to hindcast fetch-limited growth with both the DIA parameterization and the Webb–Resio–Tracy method for estimating nonlinear interactions, with very good agreement with finite-difference models (Ardhuin et al., 2007).

Since high-spectral resolution is mostly needed for swells, and swell attenuation may be described as a fully linear process (e.g., Kudryavtsev and Makin, 2004), there should be some benefit in computing swell and wind sea evolutions with different methods, even on a global scale. Such benefits for swell hindcasting in the coastal ocean were demonstrated by Ardhuin et al. (2003b) and Ardhuin and Herbers (2002). A careful comparison of ray-based advection with finite difference schemes would probably provide useful guidance. Further, alternative methods using unstructured grids are also possible. These are not necessarily less diffusive but provide an efficient use of a variable grid resolution when details are needed close to the coast. Hsu et al. (2005) proposed a Taylor–Galerkin FEM method to solve the spatial advection discretized to second order.

### 8.2.3. Addressing error due to coarse geographic resolution

The approach of Fleet Numerical Meteorology and Oceanography Center has been to simply increase resolution (the global WW3 implementation, from 1° to 1/2° resolution). In this case, the computation time of the propagation routines of the model is increased by a factor of eight (\( 2^3 \), for two geographic dimensions and for the increased temporal resolution, to satisfy the stability criterion).
The approach of Hardy et al. (2000) is to represent the blocking/scattering of wave energy by topography using sub-grid approximations. With this method, a transparency matrix that is dependent on wave direction and geographic location is specified within the model. Similar methods have been adopted with the WAVEWATCH-III model (Tolman, 2003) and with the WAM model at ECMWF (Bidlot et al., 2005; Janssen et al., 2005).

Geographic resolution is of obvious importance to shelf-scale and nearshore applications. In that case, nesting methods are available in all 3G models. For example, Lahoz and Albiach (1997) (also Albiach et al., 2000) use two-way nesting with the WAM model, with step-wise increases in resolution.

8.2.4. Garden sprinkler effect correction methods

For oceanic-scale models, no operational center has yet taken the approach of simply increasing spectral (frequency-directional) resolution, due to the computational cost. WAM, WW3, and SWAN all have the option of adding controllable diffusion to deal with the garden sprinkler effect. In the case of WAM, the controllable diffusion, like the leapfrog scheme, is rarely used. In the case of WW3 and SWAN, the controllable diffusion is specified in the manner of Booij and Holthuijsen (1987). WW3 also includes the option of simple grid point averaging, in lieu of the Booij and Holthuijsen method (Tolman, 2002a). In either case, the scheme requires a tuning parameter to control the degree of smoothing. Another type of procedure is used by Lavrenov and Onvlee (1995), in which an angular diffusive operator is included with the advection scheme numerical realization, spreading energy in directional space.

8.2.5. Errors in source term integration

Three solutions have been applied in operational models to remove the sensitivity of the results to the time step size described in the previous sub-section.

Tolman (1992b, 2002b) dynamically adjusts the time step, using the limiter to compute the maximum allowed time step. This results in a numerically accurate solution of (8.3). For large-scale model applications, this method was found to be very economical; average global model time steps of up to 40 min could be attained. For small scale applications, where rapid wind and wave changes occur over the entire domain at once, this approach can still become fairly expensive, because of the small time steps involved.

In WAM Cycle 4, the limiter was made proportional to the time step size (see Hersbach and Janssen, 1999). This modification greatly reduces the dependence on the time step size, but also prevents convergence of the solution: for very small time steps, the model does not converge to the solution of the model with no limiter (Hersbach and Janssen, 1999, Fig. 4). The disadvantage with non-convergent limiters is that the limiter becomes part of the solution, and appears to result in significant impacts on the spectral shape, even if wave heights are well represented (see Tolman, 2002b). This is especially noticeable in cases of fast wave growth (short-fetch applications).

Hersbach and Janssen (1999) reformulated the limiter of WAM4 to remove the time step dependence from the solutions. This scheme is still non-convergent, but does appear to be much closer to convergence than the earlier WAM4 limiter (Hersbach and Janssen, 1999, compare their Figs. 4 and 5). The primary advantage of this method is that cheap and robust model results are obtained, without notable time step dependencies.

An alternative for the time source term integration can be based on a spreading numerical method (Lavrenov and Kozhevnikov, 2003; Lavrenov, 2003b). It uses the semi-analytical solution for integration source term which includes the wind wave input, dissipation term, and exact nonlinear energy transfer function. The authors present an idealized test case in which reliable and stable results are achieved for time steps as large as three hours without limiters.

8.3. Relative importance of problem

In this section, we discuss the relative importance of errors of numerics and resolution. This discussion is deliberately separated from the previous descriptions, since it contains some subjective statements (or, at least, statements that are not possible to prove here). Further warning to the reader: be wary of generalizations on the subject of error in wave modeling; they are rarely, if ever, universal.
8.3.1. Error due to the numerical scheme for geographic propagation

Is it worthwhile to use higher order propagation schemes, or is a first order scheme sufficient? There is some controversy to this, so we present two points of view here (both are essentially correct). Note that there are implications for future directions: if the accuracy of a model with a first order propagation scheme is similar to that of a model with a third order scheme, this further implies that the benefit of upgrade (e.g. from third order to fifth order) will be trivial.

8.3.2. Argument

In a majority of papers that treat the subject of numerical error, there is at least one presentation of a spike of wave energy, which is supposed to represent a swell field, propagated with a first order scheme. The signal is, of course, greatly diffused. This naturally leads readers to believe that swell predictions with a first order scheme will not bear even the faintest resemblance to nature. This conclusion is incorrect for two reasons. Firstly, the curvature of the wave field – and higher order derivatives – are very rarely this extreme in nature, so the demonstrated level of diffusion is extreme. Secondly, only one spectral component is represented in this simple case. In an actual model, at any given location, numerical error of all spectral components will rarely be of the same sign; the effect of numerical geographic propagation error on wave height (i.e. the integrated wave spectrum) will tend to be relatively smaller than its effect on individual spectral components. It was shown by Rogers (2002) that the difference in error statistics (root-mean-square error and bias) between two models (one with a first order scheme, the other with more accurate propagation) can be trivial, even if only very old swells are considered.

8.3.3. Counter-argument

If methods are available to compute propagation more accurately without a large increase in computation time, then these methods should be used. Wave modelers should not be satisfied with continued reliance on cancellation of errors via spectral integration, since this hinders further model development and leaves significant errors in the spectral distribution of energy. Further, even if error statistics are not particularly sensitive to the accuracy of propagation, a model with more accurate propagation will produce images of geographic distributions of swell fields that are much more realistic in appearance than would be produced with first-order numerics. The difference, though aesthetic, is important to operational forecasters. Further, a diffusive propagation scheme makes it much more difficult to identify individual swell fields in a time series (Wingeart et al., 2001). Also, error in spectral distribution due to diffusion and dispersion make it more difficult to calculate the origins of swell energy. Lastly, even if model wave height bias is not sensitive to numerical accuracy in the open ocean, it has been demonstrated that it can be very sensitive in cases where strong gradients exist (e.g., in the lee of islands in shelf-scale applications, Rogers et al., 2002).

8.3.4. Error due to the numerical scheme for spectral propagation

This subject has received attention in the literature only in limited cases, e.g. Tolman (1991, p. 791), where it is shown that in cases of significant propagation (i.e. refraction by bathymetry and currents), a first order scheme for propagation in directional space leads to broader directional distributions. Thus far, there has been little to suggest that it should be a concern. Implementation of surface current input for wave models (such as the Gulf Stream) will make this numerical error more important.

8.3.5. Geographic resolution

On the subject of blocking and scattering by unresolved topography, the practical effect follows common sense: there is a significant positive bias near island groups, which tends to vanish in the far-field. The method of Tolman (2003) is an effective way to address the problem.

On the subject of resolving O(1°) variations in surface wind forcing: all else being equal, we expect that more variable wind fields will produce greater wave energy, analogous to the treatment of gustiness, see Section 2. Once the level of variability within the integration time step is known, this can be taken into account using the procedure presently in use at ECMWF (Janssen, 2004). The time variability at the single points also implicitly represents the spatial variability. Longer period variability, i.e. of the order of a few time or grid steps or more, is in principle automatically taken into account. However, as discussed in Section 2, while
one hand the meteorological models tend to underestimate the wind variability, the ECMWF procedure only considers its average effect. Therefore, the apparently random oscillations we see in the recorded $H_s$ time series around an otherwise smooth, e.g., growth curve are not represented or strongly smoothed in the model results.

8.3.6. Spectral resolution

The practical effect of the garden sprinkler effect (GSE) is expected to be similar to that of the propagation scheme (small impact on error metrics, significant impact on aesthetics). In time series of swell fields, the garden sprinkler effect associated with directional resolution ($15^\circ$ in most operational models) is almost always more apparent than that associated with frequency resolution (logarithmic spacing factor of 1.1 in most models). The problem with representation of the spectral peak demonstrated by Tolman (1995a) does, however, suggest that the factor 1.1 in frequency resolution is insufficient for accurate dispersion in global-scale applications. Yet, sensitivity of source/sink terms must also be considered; choosing a too fine spectral resolution may cause unphysical behavior of spectral evolution (Van Vledder et al., 2000).

8.3.7. Source term integration

With the three mainstream 3G wave models (SWAN, WAM4, WW3) each using very different solution methods, the relative importance of this numerical error is also different. As mentioned above, in the context of WW3, the primary impact of the limiter is on computation time, rather than on error. SWAN uses the WAM Cycles 1–3 limiter, so growth rates are very sensitive to the time step size: for example, even with a time step size of five minutes, the growth rate of that model is considerably slower than that with a five second time step.

The WAM4 limiter with improvements by Hersbach and Janssen (1999) appears to be much more accurate than the previous WAM limiters, though still lagging behind the “no limiter” growth rate at early stages of growth; this occurs even for very small time step sizes, so it is clearly symptomatic of non-convergence. From practical experience it appears that the limiter of Hersbach and Janssen (1999) is a good solution for engineering problems, where the goal is to estimate wave conditions accurately and economically. However, it can be argued that this non-convergent limiter is less suitable for scientific research of source term parameterizations, because the effects of the limiter on the final solution are difficult to assess, short of disabling the limiter.

8.4. Future solutions

8.4.1. The numerical scheme for geographic propagation

The objective of future development should be toward greater computational efficiency while maintaining or improving the accuracy of existing numerical schemes. In this regard, semi-Lagrangian schemes (Lavrenov and Onvlee, 1995; Ardhuin et al., 2001; Petit, 2001; Rogers and O’Reilly, 2002) are an attractive alternative to traditional Eulerian schemes. These can be simultaneously accurate, efficient, and unconditionally stable. There are two difficulties, however:

1. For the general case where propagation speed is not uniform, ray-tracing must be performed for the Lagrangian stage of the schemes, which requires some extra work. Assuring mass-conservation is generally less straightforward than with an Eulerian scheme.

2. The primary benefit of these schemes is that a parcel of energy can be propagated a long distance in a single time step, as opposed too many small steps. Thus, less error accumulates, and the higher the CFL number, the more accurate the propagation. Unfortunately, source/sink terms must be applied along the ray at the Lagrangian stage (otherwise, a parcel of wave energy might skip past a storm without receiving energy from it). Doing this in a computationally efficient manner is a challenge.

8.4.2. Geographic resolution

For shelf-scale applications, unstructured grid methods are expected to become more prevalent, since scales of variation tend to be small nearer to the shoreline, while at the same time the offshore wave field only varies on the scale larger than that of the wind field. Thus high-resolution away from the coast is generally useless, and even in hurricane conditions a resolution of a few kilometers is probably adequate. Unstructured grids are
already used now in TOMAWAC (Benoit et al., 1997), MIKE21, and have been implemented as a non-standard version of the SWAN model (Hsu et al., 2005). The present version of SWAN is also able to apply curvilinear grids allowing for finer resolution near the coast. However, the variation of the aspect ratio of the grid cells may not vary too much.

8.4.3. Spectral resolution

The existing operational methods for dealing with the garden sprinkler effect require a tuning parameter ostensibly related to wave age, but applied as a constant since the actual age of wave energy is not known to the model. Thus, there is apparent room for improvement. Tolman (2002a) proposes a new, more correct, technique using “divergent advection” but that method is still too expensive to apply. With more efficient propagation methods and/or more powerful computers, it will be feasible to increase spectral resolution (the most direct method of addressing GSE). Increasing frequency resolution may be troublesome, since the nonlinear interaction computations are sensitive to this. Thus, a reasonable approach would be to let the source/sink terms dictate frequency resolution, and gradually increase directional resolution as computational resources allow. At the same time there is little study on the improvements provided by higher directional resolution, although it is expected that better than 15° is probably necessary in coastal areas with headlands and islands in order to properly define the shadow areas. At present a frequency resolution of 10% is recommended. This choice seems to be related to the shape parameter of \( \lambda = 0.25 \) of the DIA, but the motivation for this choice is not clear. Applying other parameterizations of the nonlinear four-wave interactions will possibly lead to other optimal frequency resolutions. This implies that numerics and physics are coupled through some parameterizations of physical processes.

8.4.4. Errors in source term integration

Alternative non-convergent limiters have been proposed by Luo and Sclavo (1997), Hargreaves and Annan (1998), and Monbaliu et al. (2000). A prototype for a convergent limiter with reduced time step dependencies is proposed by Tolman (2002b).

8.5. Numerics and resolution: problems particular to finite depth and high resolution applications

In shallow water the higher resolution and stronger refraction require smaller time steps when conditionally stable Eulerian advection schemes (based on finite differences) are used (as with WAM and WW3). Even with unconditionally stable advection schemes, such as that used by SWAN for geographic propagation, accuracy decreases with larger Courant numbers. The traditional solution is to avoid the problem by switching to a stationary mode of computation at these smaller scales. This mode of computation inherently assumes that wave energy propagates across the domain instantaneously, and – in the case of models that include wave growth – that the wave field responds instantaneously to changes in the local wind field; both assumptions are reasonable at smaller scales. SWAN allows this infinite-duration mode of computation, and many nearshore models use it exclusively. At this scale, stationary models often have significant numerical challenges (e.g., Zijlema and van der Westhuysen, 2005), but since these problems, limitations, and solution methods are often unique to each model, we cannot discuss them in detail here.

Despite the solution of using stationary computations, there is recently some impetus to push exclusively nonstationary models such as WAM and WW3 closer to shore, since this avoids learning, maintaining, and running multiple wave models at a given operational center. Ray-tracing can be very efficient to avoid the cost of very small time steps, and in coastal areas source terms may often be completely neglected (e.g., O’Reilly and Guza, 1993; Peak, 2004). In general source terms may be important and the general problem is more the relationship between the advection and source term integration time step. Indeed, a few minutes or less is often needed for high resolution applications, but the source terms do not evolve on this scale and remain virtually unchanged over tens of minutes. The separation of these time steps allows great gains in CPU time in WAVEWATCH-III for example. Yet, even in the case when source terms are strong, the separation of advection and source term integration requires an update of the spectrum after the advection step, which is usually performed by recomputing the source terms. Efficient solutions may be obtained by applying the diagonal part
of the previously computed source term to the new but almost identical spectrum, or considering the evolution
of the wave field as a series of steady state conditions, as discussed above.

9. Where we are

In the previous sections we have described the present situation in the various branches that, all together,
compose the art of spectral wave modelling. We can look at this overview with two different approaches. On
one hand we can be pleased with what has been achieved. After all, the bias and scatter index of an operational
global wave model, e.g. at the European Centre for Medium-Range Weather Forecasts (Reading, UK), are an
impressive 4% and 0.11 (statistics of the first four months of 2006), or even lower once the error of the instru-
ment we compare with, in the above case the altimeter, is taken into account (P.A.E.M. Janssen, personal
communication). Also, better results are occasionally achieved by local scale modelling. Indeed, on the back-
ground of these results stand the substantial improvements in the definition of the surface wind fields.
Nevertheless, if for a moment we detach ourselves from our daily habit, it is a sort of a marvel that we can
anticipate the wave conditions in any part of the globe a few days in advance. However, as scientists we like
and must also be critical with our results and look always forward to the next steps ahead. If we do so, we
realise that there is still plenty to do. Although we are able to evaluate with good accuracy the integral prop-
ties of the sea (significant wave height, period and direction), our results are definitely less impressive once
we look at the shape of the one- and, more so, two-dimensional spectra. Peaks and extreme conditions are
frequently not well reproduced, and not only because in these cases the meteorological input is not good
enough. The point is that in such conditions the validity of the physical assumptions we have more or less
consciously absorbed in our theories are often stretched to their limits. Imbedded in our models there is still
a substantial degree of empiricism, that unavoidably is due to fail at a more or less large degree once we act out
of the usual range of conditions. Clearly a critical review is required, and this is what we have tried to achieve
with this paper. It is worthwhile to summarise where we stand in the single subjects we have described.

The generation by wind is an extremely complex process. We deal with the highly nonlinear interaction of
two fluids whose densities differ by three orders of magnitude. This implies a multi-phenomenological behav-
iour at the interface, more or less complex as the difference of speeds in the two layers increases. Also, direct
visual observation is of little help, providing evidence of the integral results rather than of the mechanism by
which energy is transferred from one fluid to the other one. Nevertheless, using some simplifying assumptions,
quite a bit of physical intuition and devoted measurements we have been able to formulate some basic theory
that indeed, once applied to the models, provides rather good results.

On the other hand the very fact that two of the most popular models, WAM and WAVEWATCH, oper-
ational at two of the most prominent meteorological centres, use different approaches to the problem is in
itself an indication that a single “best” solution has not yet been accepted.

In the present theories the very hypothesis of linearity, i.e. to consider the sea as a superposition of sinu-
soidal components, should at least be open to doubts. An immediate example is the skewness of a stormy sur-
face, by definition not considered in the standard spectral wave models. This is likely to have an effect on
generation, whose process, like white-capping, is not so smooth in space and time as the theories imply. Who-
ever has been at sea in the middle of a storm is led to question the hypothesis of linearity. Of course this can be
said for all the processes where the hypothesis has been used, but generation by wind is the only one where at
present, safe for the dependence of $u_*$ on the overall sea conditions, the modelled energy input at certain fre-
quency and direction does not depend on the contemporary situation at the other components.

Of course this makes even more noticeable the results achieved till now, and it is a good proof of the inge-
nuity and brilliant hypotheses that stand at the base of the present theories. Indeed the very fact that with a
theory based on the linear hypothesis we manage to achieve good results should in itself be a valuable piece of
information.

For nonlinear interactions in deep water the basic problem seems to be the practical implementation of an
already well established theory. The struggle between the sheer volume of calculations implied by the theory
and the practical possibilities of the present computers has been dominating the stage for a long while. The
capability of routinely carrying out full exact computations is still far away. The present efforts aim at devel-
oping new methods (MDIA, neural, diffusion), while exploiting the ever increasing computer power, reducing
the necessary time within manageable limits. These calculations are always compromises, and usually this appears as undesirable characteristics of the final results. Each one of the newly proposed methods has its own limitations, often still to be explored.

Notwithstanding its sound theoretical definition more than 40 years ago, the subject is still characterised by an active development. The full properties of the kinetic equation are still to be explored. One first brilliant example of recent developments is the evaluation of the probability of freak waves starting from the modelled spectra, in so doing correcting as a following step the limitations on the skewness of the sea surface we pointed out above. Another similarly valuable example has been to show that the nonlinear interactions lead to a bimodal spectrum also in anisotropic conditions. In particular, the considerations of the quasi-resonant interactions seem to be a promising field of research.

The number of different approaches and proposed new solutions to the calculation of nonlinear interactions suggests that an intercomparison exercise, both in idealised and practical conditions, is required. This will help to define in a comparative way the characteristics and the capabilities of the single approaches.

The dissipation of wind waves in deep water is by definition the source term we know less. There is hardly any agreement neither on the basic physics of the process nor on the best way, although empirical, to model it.

We find worthwhile to repeat here two paragraphs of Section 4 related to the physical knowledge of the process:

“To summarize this brief overview of existing theories of spectral dissipation, we find several studies which offer four different analytical models. None of the models deals with the dynamics of wave breaking, which is responsible for dissipation. Rather, they suggest hypotheses to interpret either pre-breaking or post-breaking wave field properties. All of the hypotheses lack experimental support or validation. Results vary from the dissipation being a linear function of the wave spectrum to the dissipation being quadratic, cubic or even a function of the spectrum to the fifth power.”

“To conclude this review, we have to summarize that (1) there is no consensus among analytical theories of the spectral dissipation of wave energy due to wave breaking, even with respect to the basic characteristics of the dissipation function, (2) the theoretical dissipation functions strongly disagree with the experiment, and (3) experimental results, even though they exhibit some common features, are often in serious disagreement with each other. Such a state of knowledge of physics of the wave breaking losses does not help modelling the wave dissipation which has been drifting in its own way.”

This could be a rather discouraging situation and shows how much there is still to be done in this subject. On the other hand this has stimulated quite a bit of basic research in the recent years. However, the results of this research have still to find their way into the operational models that, as just quoted, given the theoretical situation have been mostly drifting in their own way. Indeed, given the relative level of knowledge, spectral dissipation has been for a long while, and still is, the tuning knob of the numerical wave models to make them fit at least the wave integral properties (significant wave height, period, direction).

Attempts to reproduce more integral properties of the wave field, e.g. the characteristics of the spectra, have recently led to various lines of research. In particular, it has been made clear that any pre-assumption of the spectral shape, like the power law of the high frequency tail, is bound to make sooner or later the solution diverge from the truth. This has led to more fundamental approaches that have yet to find their way into operational models.

Nonlinear interactions in shallow water are characterised by the relevance of the third-order ones. Dealing with interactions, not only in resonant, but also in near-resonant conditions, is today an active field of research, and the associated wave modelling activity has different lines of attack in this respect.

While the stochastic (i.e. spectral energy) approach is the undebated approach in the open oceans, close to shore, where changes can take place at a high rate and the degree of nonlinearity may jump at high levels, the deterministic approach (i.e. complex amplitude or surface elevation) would appear to be the natural solution. For the time being the obvious limit of the required computer power makes this approach suitable for short distances only (a limited number of wavelengths). However, a practical problem is also the connection with the offshore, spectrally modelled, wave conditions, from which different realizations of the boundary conditions would need to be modelled if suitable statistics are to be derived. This is presently off-limits, even at the level of devoted experiments. However, an efficient alternative would be provided by spectral, complex-amplitude
evolution models, with high spatial resolution where nonlinearities are significant, and low resolution where they are not, run in frequency space and on unstructured geographic grids.

A third alternative is offered by the stochastic approach, derived from deterministic equations and ensemble averaging. Most models limit the derived hierarchy of equations to two coupled equations for spectrum and bi-spectrum. This solution is attractive, because it allows the direct computation of statistical quantities without the need for Montecarlo simulations. The model can be initialised with standard spectra (buoys or offshore spectral models), while the bi-spectrum is derived from second order theory.

While there is a tendency to push the operational large scale spectral models towards the shore, it is necessary to point out that some of the solutions present in these models are still rather crude, especially when compared to the phase resolving and complex amplitude models. A strong obstacle is given by the lack of sound physical approaches on how to handle dissipation, particularly the depth induced one, so relevant in shallow water. We still do not know how to distribute the energy loss throughout the spectrum. Also, we should not forget that most of the calculations with the nonlinear models mentioned above have been carried out on very simplified, regular bathymetries. Any operational application in real conditions is much more problematic.

Dissipation associated to the interaction of waves with the bottom is another subject where we still have a lot to learn. The problem is associated with two basic characteristics of what is going on: the number of contemporary and alternative bottom mechanisms that can be active to dissipate the wave energy, and the difficulty of analysing and measuring a process while it is active. As a matter of fact practically all the data we have concern the measurements of wave characteristics at different progressive locations, in so doing providing information only on the integrated effect of the process, rather than on the physics and its details. Somehow we can also think to be more sensitive in our observations, hence more speculative, to surface breaking, simply because of its visibility, while of bottom dissipation we have only a perception of its consequences. In general we can say we have a fair idea of the physics involved, but we lack a solid quantification of the energy lost in the process. Related model data, estimated to be off by an order of magnitude, are not unknown.

There are practical difficulties. On one hand also the integrated characteristics of the surface are not always purely indicative of the bottom dissipation processes, simply because there are often other, not necessarily bottom, processes at work, e.g. generation by wind and white-capping. On the other hand the true characteristics of the bottom are mostly unknown (dimensions of the ripples, sheet flow, etc.) or, at best, modelled only with large approximations, and they can easily change the estimate of the derived energy loss of an order of magnitude.

Also the physics of the influence of a current on bottom dissipation is not fully understood. The intuition suggests that, when contemporarily present, both losses, those due to waves and to current, should be enhanced. However, the evidence is not clear, notwithstanding the relevance of the subject for storm surge modelling and the evaluation of wave and current conditions in tidal inlets.

For practical and operational applications a serious problem is given by the sub-grid variability. Particularly close to shore this can be quite high, and average conditions over one grid step are not granted to provide the correct integral over its extent.

Notwithstanding this rather pessimistic panorama, bottom dissipation, mostly represented by the bottom friction process, is regularly considered in shallow water modelling. The point is that, with the exception of particular conditions as the Southern North Sea or a long swell on oceanic coastlines with an extended continental shelf, bottom friction is rarely the dominant process for the proper evaluation of the wave conditions at a given location. Of course this does not cancel the need for a deepening of the subject.

Although non-dissipative, bottom scattering, discussed also in Section 7, has a more positive situation, at least from the theoretical point of view. The interaction of the surface spectrum with the geometrical characteristics of the bottom is relatively well understood. In recent times also the effects of single perturbations of the bottom, like a single step, have been dealt with mathematically. While the laboratory results support these approaches, confirmation from the field seems more difficult to obtain.

Wave propagation addresses the problem of waves propagating on an uneven bottom or across a non-uniform and time varying current. Most of the present models rely on the validity of the linear theory, using the classical linear dispersion relationship to relate frequency and wavenumber. Also the Earth rotation has a lim-
ited influence, and waves propagate with very good approximation along great circles that, on limited distances, coincide at all the effects with straight lines.

If depth and current change over distances much larger than the considered wavelength, the usual geometrical optical approximation is quite robust. Expectably complications arise when the changes take place over distances comparable with the wavelength. In phase-averaged models these discontinuities are usually treated introducing frequency dependent reflection coefficients at the proper locations and directions, providing quite reasonable results. More in general the interaction of the surface waves and the bottom elevation spectra implies a conservative scattering of the surface waves. This process is still not yet included in most wave models because a proper theory has only been given recently, and practical methods for its calculation are still to be defined when only little information is available on the bottom spectra.

The same approach used for wave–bottom conservative interactions is usable also for currents. Here too the level of interaction depends on the amplitude and the spatial scale of the current variations. The modifications of waves when interacting with current are not interesting only on themselves, but also for remote sensing, both from space and from coastal water. A strong limitation to the operational implementation of the extensive theory available is the lack of sufficiently accurate description of the current field in the open sea. While improvements are expected in the near future, in practice for the time being the only interactions with currents that receive sufficient attention in operational models are the ones with tidal currents.

In any case all these approaches are generally applied with the current assumed to be uniform on the vertical. This is not always the case, but the implications are not considered in standard wave modelling. While the problem is probably limited for large scale currents, the Stokes drift is expected to have a non-negligible impact on shorter waves.

With respects to the other subjects, numerics has the big advantage of being perfectly defined, and suitable for an analysis of the practical results with respect to the ones expected from theory. This does not make the problem simpler, but at least we can have a clear idea of where we are. Of course the problem is associated to the discretization with which we describe an otherwise continuous nature. This implies some approximations, as for instance in the description of the peak of the spectrum (frequency resolution) or in the characterization of the bottom profile (spatial resolution). More seriously, it implies a modification of the signal while it propagates, theoretically undisturbed, across the grid. The approaches to this problem are different, depending if we deal with advection, both in lat-lon and in spectral space, or with the description of the spectra and the geography of the area.

In the case of advection, the problem is well understood and a whole hierarchy of approaches has been proposed. Indeed it is remarkable that a “best solution” is not universally adopted. Clearly this points to the fact that in a certain environment any practical solution, besides being linked to historical reasons, is always a compromise between several requirements. One peculiar fact of these compromises is the apparent compensation introduced by the signal diffusion for the patchy distribution due to the garden sprinkler effect. Although criticised, the solution has certainly served its purpose. Higher order advection schemes, paralleled by a controlled diffusion algorithm, are presently available, although the opinions on which one is preferable are certainly not uniform.

A correct geography is just a matter of resolution, and implicitly of computer power, because the system of differential equations must be solved at each grid point at each time step. The combination of these two needs makes the overall computer power to grow as $1/A^3$, where $A$ is the geographical resolution. A substantial problem are the sub-grid characteristics of the area of interest, typically small islands not represented in the computational grid. In this case the solution is a transparency coefficient, calculated from a much higher resolution bottom topography, for each point of the grid and for each component of the spectrum.

The natural solution to the general problem is to use a variable grid resolution, typically more coarse in the large ocean spaces, and highly defined close to the coasts. This can be achieved either with nested modelling or with unstructured grids. This latter solution has never been very popular in wave modelling, but it is rapidly gaining ground, particularly for dealing with an optimised resolution with coastal and inner areas with a complicated bathymetry.

The discretization in space is reflected also in time, and the step integration of the set of equations at the base of a model has its implications. With the traditional Eulerian approach the time step is upper limited by the grid step size due to either stability or accuracy. In some models the introduction of a semi-Lagrangian
advection has somehow relaxed this condition, but attention must be given to the physics of the processes. In particular the use of the same $\Delta t$ for all the frequencies can be questionable, leading to the use of suitable, but artificial, limits to the changes during each integration step. Notable progresses have been made in this respect in recent times.

This compact summary, and more in general the material presented in the previous sections, points to the extensive effort that is still going on in wave modelling. This is due to two characteristics not easily found in other subjects. On one hand we deal with a very complex physical process where physics, from fundamental principles till very practical problems, plays a dominant role. On the other hand the subject is highly in demand for its very wide applications, with a continuous push by the market forces to improve the quality of the results.

Since the first order approximation of the historical SMB method (Sverdrup and Munk, 1946), we have well achieved the next step, with much reduced bias and r.m.s. errors of the integral parameters, particularly off the coasts. What is next? We expect to decrease further the above errors. This can be achieved refining the formulation of the single processes following the various approaches described in the various sections, improving the numerics, and, still critical, although not so much as in the past, improving the input wind fields. However, the real task is to ameliorate the quality of the spectra. Although not yet strongly required by the market, their use in practical applications is growing and the present limitations of spectral wave modelling in this respect are beginning to be felt. It is not only a matter of users. To improve the quality of the spectra will allow a better description of some physical processes that depend so much on their details.

A substantial question concerns the high frequency tail of the spectrum, presently parameterised in a not yet agreed way, notwithstanding its relevance in the overall physics and for practical applications, e.g. remote sensing and coupling with meteorological models.

Notwithstanding the good average results of a wave model, at least as integral parameters, a still missing point is the physics, hence modelling, of extreme conditions. We still are not sure of the processes that are taking place and of the resolution required for their representation. The difficulties cannot be underestimated, also because the corresponding laboratory results provide only limited replies. However, the recent events and our growing interaction with the sea are clearly pushing towards a better understanding of what is going on in these conditions.

Clearly an area where action is required is the interaction between waves and currents. At the simplest level of a vertically uniform current field, improvements are expected in a relatively short while from global circulation models. However, this will concern the general features of a field. Somehow this is similar to the argument on the tail of a wave spectrum mentioned above. Both because of a lack of information and of the present limits of the circulation models, the representation of the details of the fields is rather approximate. However, this is still a scale capable to affect the wave fields at an appreciable level.

The difficulty of the problem steps up once we consider the currents as three-dimensional. Particularly, but not only, in coastal areas this can indeed be the case. Even assuming we know the details of the current field, the know-how of how to deal with this problem is not yet a granted background of the wave modelling community.

So our feeling is a mixture of satisfaction for the results achieved so far and of realization of our present limitations and the need to go further. Some of the areas where to act are quite clear, other ones are more foggy. The next and final section will deal, although also in a speculative way, with this last point.

**10. Where to go**

Having stated where we are and the obvious problems to face in the immediate future, we need to think in longer terms and argue about the strategy for the future. A forecast in a still partly unknown territory is always a hard bet, but it is worthwhile to try, at least to quantify the problem. Cavaleri (2006) argues about the far future of wave modelling. Here we concern ourselves with more immediate developments.

In the introduction we had mentioned that, just because we are acting at the far front of research, our opinions are often not uniform. We had also pointed out that this is a necessary and favourable condition to go further, simply because we do not know in advance which will be the winning strategy. Expectably, the spectrum of opinions widens in a nonlinear way the further we speculate on the future. Therefore, this section re-
resents in some cases some obvious requirements and expectations, in other cases ideas floating around with a different level of agreement. Perhaps this is the most exciting part of our work.

Clearly our field is highly variegated. We have different branches where we act with different levels of confidence, and where the physics and the possible paths for the future are known with a similarly variable degree of uncertainty. Some of the problems are technical, other ones are physical, so it is not possible to give a single general statement. Rather, we can touch several points in sequence.

We begin with the generation by wind. All the present approaches stand on the spectral hypothesis, i.e. the sea is conceived as a superposition of sinusoids, and we estimate the input to each component on the base of the, although modified, Miles’s theory. This approach has been very successful, but the view of a stormy sea hardly suggests this idea. Already 30 years ago Banner and Melville (1976) have shown that the input by wind to waves is not the smooth continuous process implies by the Miles’s approach. Rather, it is highly discontinuous, with strong bursts of momentum and energy transfer. The point is that we do not know how to deal with such a process. However, this should not make us hide the fact that our present approach, albeit successful, is not a faithful representation of what is going on the sea. How to deviate from our present path is an open question, but sooner or later something will have to be done.

The work by Banner and Melville (1976) has shown the clear link in an active young sea between generation and white-capping dissipation. While for the time being they are independently evaluated, it is a real possibility that at some stage they will have to be considered as a single process. However, this is not for the near future. For the time being a more physical description of white-capping is highly in demand. There are indications that the careful analysis of the available experimental data is opening doors in this direction. In any case the move must clearly be from empirism towards the physics.

Remaining in the physical realm, the bottom dissipation processes are a real challenge, perhaps not so much for their physics that, at least in the first approximation, is relatively understood. The problem is the availability of the information (the characteristics of the bottom) required for their correct evaluation. Within the relevance of the process for the evaluation of the wave conditions at a certain location, a detailed knowledge of the bottom characteristics of the area is a mandatory condition. This will also help to decide which processes can be locally relevant and it is therefore worthwhile, if not all of them, to consider. However, granted this information, the correct quantification of the energy involved in the processes is still a problem, as their physics itself implies that small changes of the wave conditions can lead to an order of magnitude difference of the involved energy budget. How to deal with this problem is still an open question.

Also, quite a bit of physics is still to be clarified. Although limited to some special areas, the anelastic motion of a viscous muddy bottom is not properly understood, especially in connection with the dissipation of also relatively high frequencies. The relevance for hurricane affected areas as the Gulf of Mexico or the Bay of Bengal is evident. This requires some devoted measurements and physical intuition.

It can be surprising, but, at a second thought, instructive, that nonlinear interactions, the most purely physical process we deal with, is theoretically the best known. The sheer complexity has its revenge in the present practical impossibility of routinely evaluating the exact result. Of course in the long term we can expect the computer power to keep growing, although perhaps not so rapidly as during the last thirty years. However, this will not be enough, and, as already discussed, compromise solutions need to and will be found. The question is how accurate we need to be to guide the evolution of the spectrum towards the correct results. Somehow this needs to be quantified through the already proposed intercomparison exercise.

In shallow water there seems to be more ground for not-only-numerical developments. Somehow the exploration of this area of research has begun in more recent times, and further developments are needed and expected. The substantial gap of computer power required by phase resolving and phase averaging models leaves ample ground for intermediate solutions. Considering spectra and bispectra is just an example in this direction, but quite a bit of activity is expected for the future along this or similar lines of activity. Whichever the solution adopted, it is clear that a higher resolution is required close to the coasts. The tendency for having a single model for the whole area of interest stresses the need for variable resolutions, with an expected increased use of unstructured grids.

A stronger interaction between the wave and the circulation modelling community is a must and an expected development. It is not only a problem of operational applications, but also of physics of both
the models. For applications, we have the mentioned need of a better description of the current fields to properly evaluate their effect on the wave field. Conversely, there is also the effect of waves on the current. Similarly for what done for the coupling with the atmosphere, we need a two-way coupling between wave and circulation models. There are various aspects open to findings. Perhaps the most macroscopic one is the driving of circulation by wind. While this is presently done using the wind stresses, the flow of energy and momentum wind → waves → breaking → circulation needs to be considered as the real driving mechanism.

The increased coastal resolution mentioned above needs to be considered also as regards the propagation on an uneven bottom. Apart from the technical aspects with which the irregularities of the bottom profile can be dealt with, clearly these features need to be resolved. While intuitively we associate an increasing resolution to the approach to the coastline, we can certainly think of using it also on required isolated spots.

Finally, concerning numerics, improvements are expected in two directions. On one hand we need more efficient and accurate algorithms, both for advection and for time integration. Some improvements are expected, although apparently the clear definition and limits of the problem leaves a limited ground for manoeuvre. Possibly a stronger improvement will come from the combined use of Eulerian and Lagrangian advection techniques, both in open and coastal waters.

Having discussed how to improve the modelling of the single processes and what we expect for the near future, we need to ask ourselves a basic question. Even assuming that all the representations of the single processes are improved with respect to their present state of the art, can we assume that this will produce better overall results with respect to the present performances? The point we should not hide is that in the present models, although they are declared as purely physical, there is quite a bit of tuning and artifices to make them fit the measured data. This happens at different parts of the models and with different strategies, but it is there. We have mentioned that white-capping, just because it is the least known process, is often used as a tuning knob to best-fit the results and measurements. Given this situation, what can we expect once each process is independently described at its best, even improved, level of knowledge? Most likely, if not certainly, the results will be worse than the present ones, at least at the beginning. With progressive improvements we will move ahead of the present performance. However, most likely also in the longer term new “optimised” models will continuously branch out of the main line of development, improving for the time being the overall performance.

Should we refrain from acting in this direction? Certainly not, because we must keep in mind the duty of practical applications. While we develop our models towards the best and most physical solution, we have also to provide continuously the best possible results to the users. So somehow we have to live with this dichotomy that we recognise also in the present large scale operational models, where we find different levels of pragmatism depending on where one institution puts the focus for its results.

How to conclude? Many doors are open, and work will be done in many directions. Some are known or expected, in other cases we look for new ones. However, one general idea is clear. Whatever we do, we have to move towards a more fundamental coupling between the sea and the atmosphere. The meteorological models must interact continuously with the ocean circulation models, not through empirical formulations, but through the physically sound interface of a wave model, acting as the element that determines how the exchanges take place and their extent. This is how nature works, and this is how we have to represent it if we aim at a better understanding and modelling of the thin layer of fluid that surrounds our beautiful planet.

Acknowledgement

The knowledge of a branch of science by a specific group of researchers unavoidably derives also from the many persons with whom, in a direct or non-direct way, they have interacted. In this respect, although written by us, this paper represents in our intention the ideas and the long time efforts of the wave modelling community. We are pleased to acknowledge the contribution of whoever has added his piece of knowledge to the pyramid of wave modelling. Jane Frankenfield has done an appreciated editing of the final version of the paper, carefully revising all its details and the (long) reference list. With her usual care of details Luciana Bertotti has then cross-checked all the references in the text and in the list.
References


Tracy, B.A., Resio, D.T., 1982. Theory and calculation of the nonlinear energy transfer between sea waves in deep water. WIS Technical Report 11, US Army Engineer Waterways Experiment Station, Vicksburg, MS, USA (pagine?).


