UNLIMITED DISTRIBUTION

**National Defence
Research and
Development Branch**

DREA CR/94/410

**STATISTICAL DESCRIPTION OF THE
EAST COAST DIRECTIONAL
WAVE CLIMATE USING
10-PARAMETER SPECTRA**

by
Barbara-Ann Juszko

JUSZKO SCIENTIFIC SERVICES
127 Cliff Drive, R.R.#4
Victoria, British Columbia, Canada
V9B 5T8

Scientific Authority
Ross Graham
March 1994

W7707-3-2618
Contract Number

CONTRACTOR REPORT

Prepared for

**Defence
Research
Establishment
Atlantic**

Canada

ABSTRACT

Ten parameter model spectra were fit by MEDS to the three years of archived ODGP directional wave spectra from 53 locations in the Western North Atlantic. There were acceptable fits for over 93% of the records. Empirical orthogonal function and factor analyses were used to identify six regions, encompassing locations having a high degree of covariability, to allow for an assessment of regional behaviour. A statistical analysis was performed on annual, winter and fall data for the individual, regionally grouped and all combined locations. This consisted of a probability analysis on the model fit parameters, the development of a family of directional spectra, having known confidence levels, and the establishment of predictive relationships between the fit parameters and significant wave height. A set of probability directional wave spectra, chosen to represent a desired spatial scale, can now be generated for any input significant wave height as required by the given application.

RÉSUMÉ

Des spectres modèles à dix paramètres ont été ajustés par MEDS aux spectres directionnels ODGP de vagues archivés. Les spectres archivés comprenaient trois ans de données et venaient de 53 emplacements de la partie ouest de l'Atlantique Nord. Les ajustements étaient acceptables pour plus de 93% des spectres. Des fonctions orthogonales empiriques et des analyses factorielles ont été utilisées pour identifier six régions couvrant des emplacements où la covariabilité était élevée pour permettre une Evaluation des emportements régionaux. Une analyse statistique des données annuelles, hivernales et automnales a été effectuée pour chaque emplacement, pour les emplacement regroupés en régions et pour l'ensemble des emplacements. Elle consistait en une analyse de probabilités des paramétres d'ajustement des modèles, en la mise au point d'une famille de spectres directionnels de niveaux de confiance connus et en l'établissement de relations de prévision entre les paramètres d'ajustement et la hauteur significative des vagues. Un ensemble de spectres directionnels probabilistes de vagues, choisis pour représenter une échelle spatiale souhaitée, peut maintenant être produit pour n'importe quelle hauteur significative des vagues donnée comme Pexigeait l'application en question.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1.0 INTRODUCTION

The results of Juszko, (1989 a,b) show that a 10-parameter model spectrum, based on the Ochi and Hubble (1976) 6-parameter model, can be used to represent both field and hindcast directional wave spectra. With minimum sacrifice in data accuracy, this shorthand description of complex spectra provides significant advantages for data storage and integration of spectral information into complex numerical models. Furthermore, by supplying a fixed set of descriptive parameters, Juszko, (1991) demonstrated that statistical analyses can be performed to categorize large volumes of data, to establish the probability distribution of the spectral parameters and to provide the confidence limits on representative spectra, necessary for risk and operability assessment in ocean engineering.

In this study, three years of Offshore Data Gathering Program (ODGP) hindcast directional wave spectra, from 53 archived locations in the Western North Atlantic, will be analyzed in order to develop a statistical description of the wave climate. The 10-parameter model fits to the ODGP spectra will provide the fundamental data set for the statistical analysis. Annual, winter and fall statistics from individual, regionally grouped and all combined locations will be examined. A family of directional spectra, having known confidence levels, will be generated for each data subset for directed application in ship operability assessment and structural design.

2.0 STUDY BACKGROUND

2.1 Study Objectives

There were four primary objectives of this study. The first objective was to apply the 10-parameter fit software, developed by Juszko (1989b, 1991), to the archived ODGP hindcast wave spectra for the period 1 Oct 1983 through 30 Sep 1986. This was performed by the Marine Environmental Data Service (MEDS) after minor modifications to the software in order to adjust to the VAX operating system. The second objective was to conduct a regional analysis on the data set ill order to determine if an objective grouping of sites was possible based on the available environmental information. The numerical techniques of empirical orthogonal function and factor analyses proved successful in this task with resulting site clusters corresponding well with accepted geographic areas. The third objective was to provide a statistical description of the wave climate based on the 10-fit parameters, for various temporal and spatial data subsets. In particular, a family of directional spectra, having known probability of occurrence, was to be developed. The fourth objective was to assess the contribution of the shape parameters to the observed variability in the probability spectra.

2.2 Data Sources and Processing

Three years of ODGP directional wave spectra, generated every six hours from 1 Oct 1983 through 30 Sep 1986, at 53 grid sites shown in Fig 1 , have been archived at MEDS. As seen in Fig 2 , the time series of spectral information were often discontinuous at northern and nearshore stations with gaps associated with winter, ice-covered periods. Using software developed in Juszko (1991), a total of 220,151 spectra from the 53 sites were fit by the 10-parameter model:

$$M(\omega, \theta) = \frac{\frac{1}{4}\sum_{i=1}^{i=2}(\frac{4\lambda_i+1}{4}\omega_{mi}^4)^{\lambda_i}\,\delta_i^2\exp\left[-(\frac{4\lambda_i+1}{4})(\frac{\omega_{mi}}{\omega})^4\right]A(p_i)\cos^{2p_i}(\frac{\theta-\theta_{mi}}{2})}{\Gamma(\lambda_i)\omega^{4\lambda_i+1}} \qquad (1)$$

Here $M(\omega, \theta)$ is the directional spectrum, is the modal frequency, δ is the variance parameter, λ is the shape parameter, A(p_i) is a normalization factor for the area under a $\cos^{2p}$ curve, p is the directional spread and θ_m is the modal direction. A(p) is expressed as:

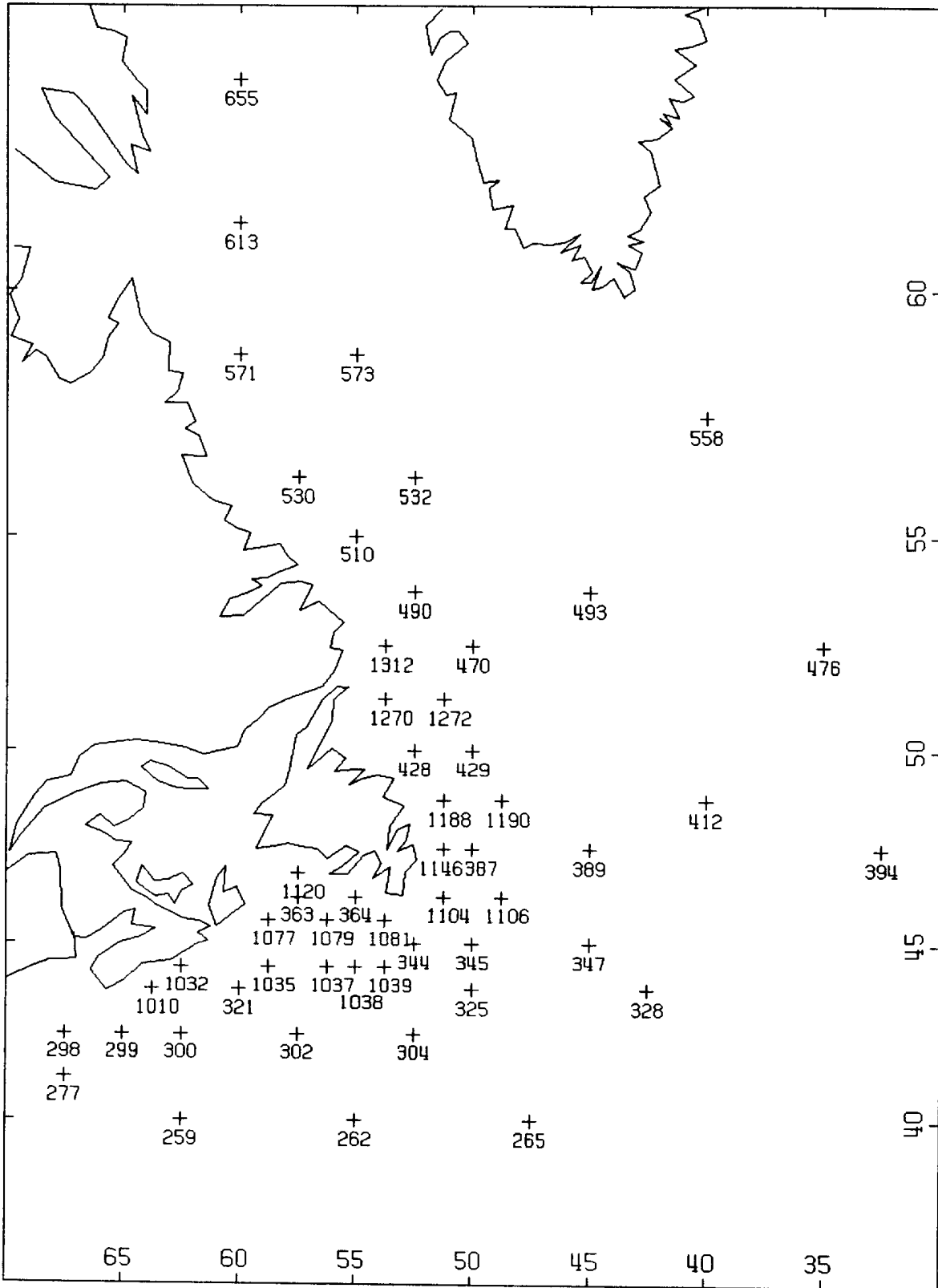$$A(p) = \frac{2^{2p-1}\Gamma^2(p+1)}{\pi\,\Gamma(2p+1)} \qquad (2)$$

Fig.1 Location of GDGP spectra archived according
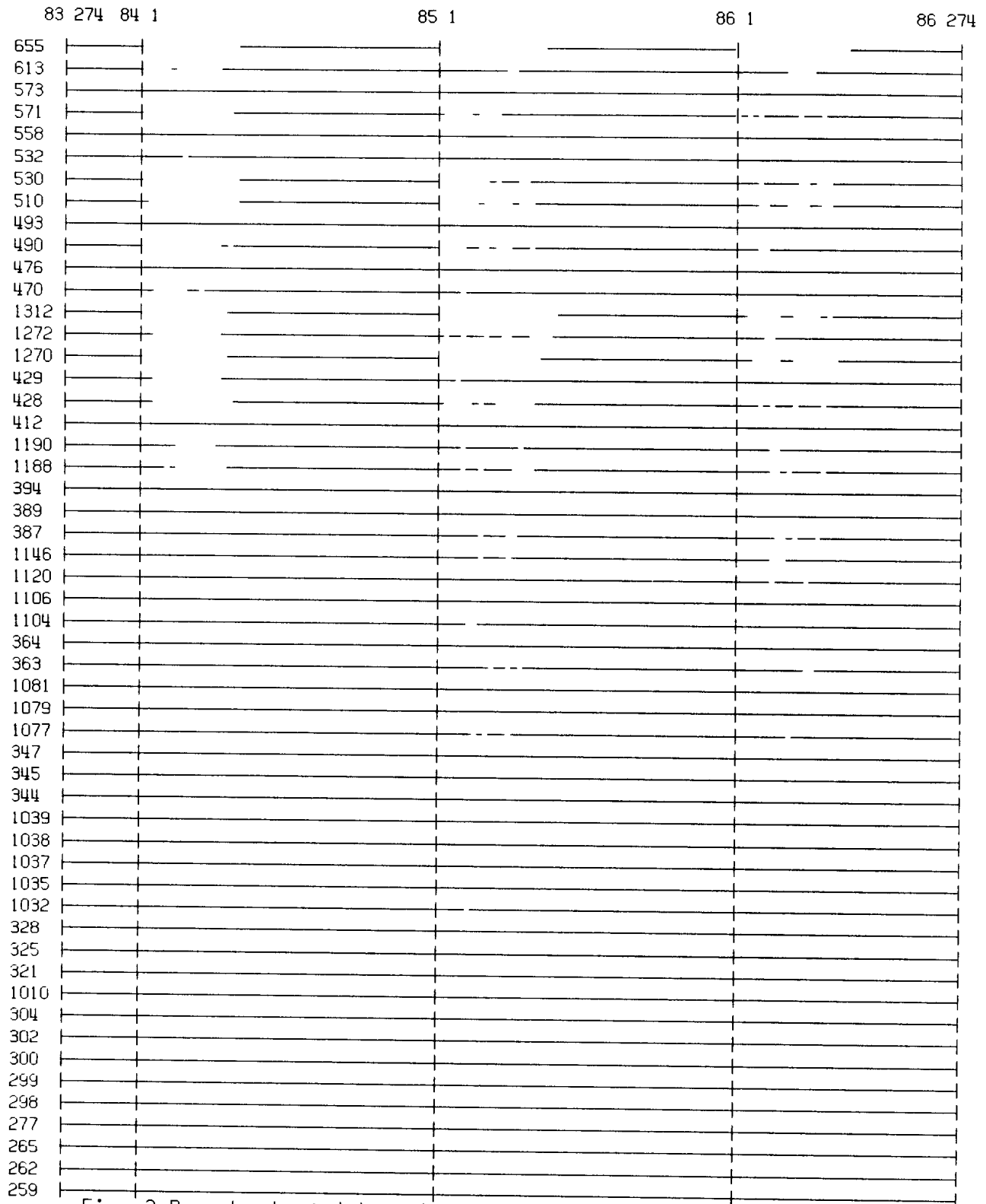to grid point number.

Fig. 2 Bar chart of data return.

The accuracy of the 10-parameter fit was evaluated using the residual error statistic:

$$RESD = \frac{\sum\limits_{i=1}^{i=N} \sum\limits_{j=1}^{j=M} [D(\omega_i, \theta_j) - M(\omega_i, \theta_j)]^2 \, WT_i^2}{\sum\limits_{i=1}^{i=N} \sum\limits_{j=1}^{j=M} [D(\omega_i, \theta_j)]^2 \, WT_i^2} \qquad (3)$$

where D represents the ODGP input spectrum, WT is a weighting based on the ODGP spectral bandwidths and the sum is performed over N frequencies and M directions. The input ODGP spectra consisted of N=15 frequencies with nominal values of 0.2545, 0.2792, 0.3142, 0.3491, 0.3840, 0.4189, 0.4538, 0.5062, 0.5760, 0.6458, 0.7331, 0.83777, 0.9948, 1.309 and 1.9373 rps and M=24 directions with a 15 degree resolution. An advantage of the 10-parameter model is that it can be regenerated at any desired frequency – direction combination.

In order to represent both the more energetic wave periods and to account for the winter data gaps at many locations, the data set has been divided into three temporal periods: annual (all records), winter (January, February, March) and fall (October, November, December). From earlier work, a fit acceptance criteria of RESD≤20% was established and the total percent of accepted records per site is shown in Fig 3 . It can be seen in Fig 3   that over 90% of the fit results, at most locations, are accepted justifying the general use of the 10-parameter model as well as verifying programming accuracy. The poorest return was for sites off the east coast of Newfoundland possibly reflecting the reduced number of energetic winter records and single-peaked spectra which are well represented by the 10-parameter model. It will be seen in the regional analysis of Section 2.3   and 3 , that this area could be considered as a separate region based on various other criteria.

The basic data set available for the statistical analyses included the model wind speed and direction, significant wave height, a spectral type code determined for the ODGP spectra, the 10-fit parameters and the RESD error statistic. The spectral type code consisted of a two-digit value where the first digit represents the number of scanned peaks and the second corresponds to one of the following values based on the direction and period separation between the two largest peaks:

0 – single peaks
1 – if $d\theta$ > 45 degrees; $dP \geq 2$ seconds

2 – if $d\theta$ > 45 degrees; dP < 2 seconds
3 – if $d\theta \times \leq$ 45 degrees; $dP \geq$ 2 seconds
4 – if $d\theta \leq$ 45 degrees; dP < 2 seconds

For example, a code of 10 represents a single peaked directional
spectrum (at any frequency or direction), a code of 21 indicates a
bimodal spectrum whose two peaks are separated in both period and
direction, a code of 32 reflects a spectrum having three peaks where
the two largest peaks are separated in direction but close in period,
etc. The spectral type code provides a further data grouping criteria,
if desired. Fig 4   contains maps of the percent occurrence of
spectral type at each grid point. It is important to note that
multiple peak spectra made up over 50% of the records at 48 of the 53
sites justifying the use of a parametric model which allows for at
least two directional peaks.

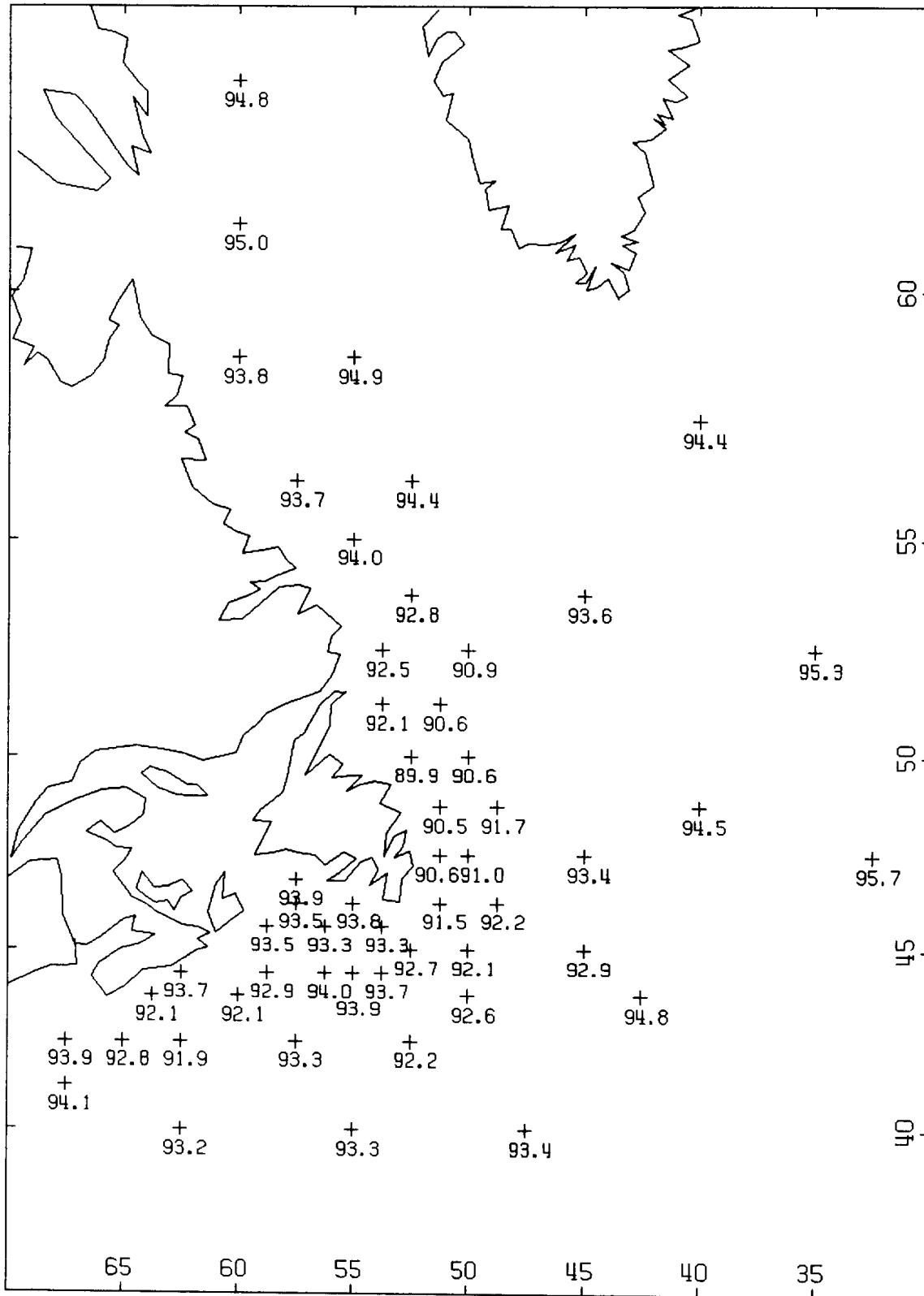Fig. 3 Percentage of acceptable 10-parameter fits
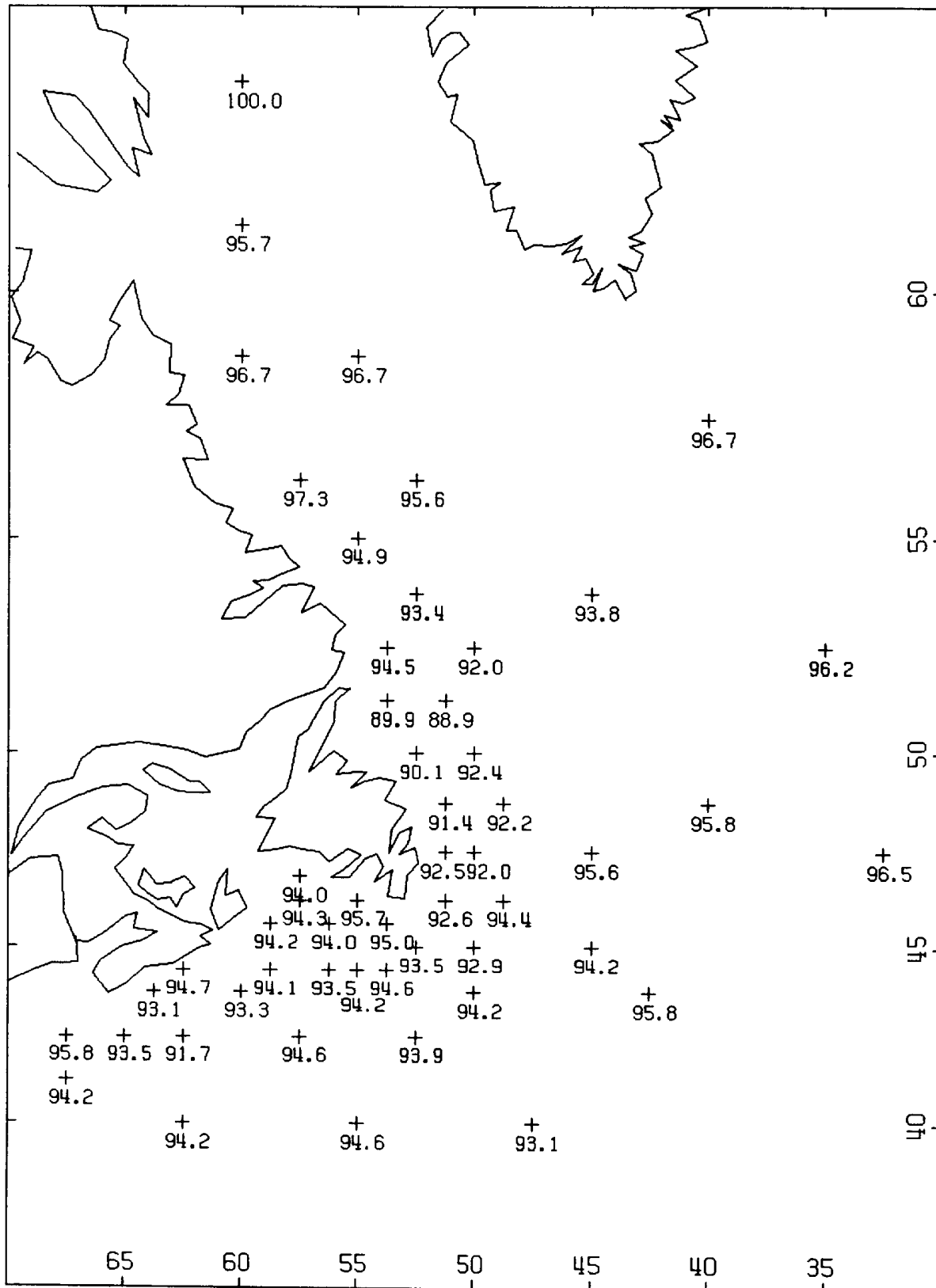   at each site. a) annual data.
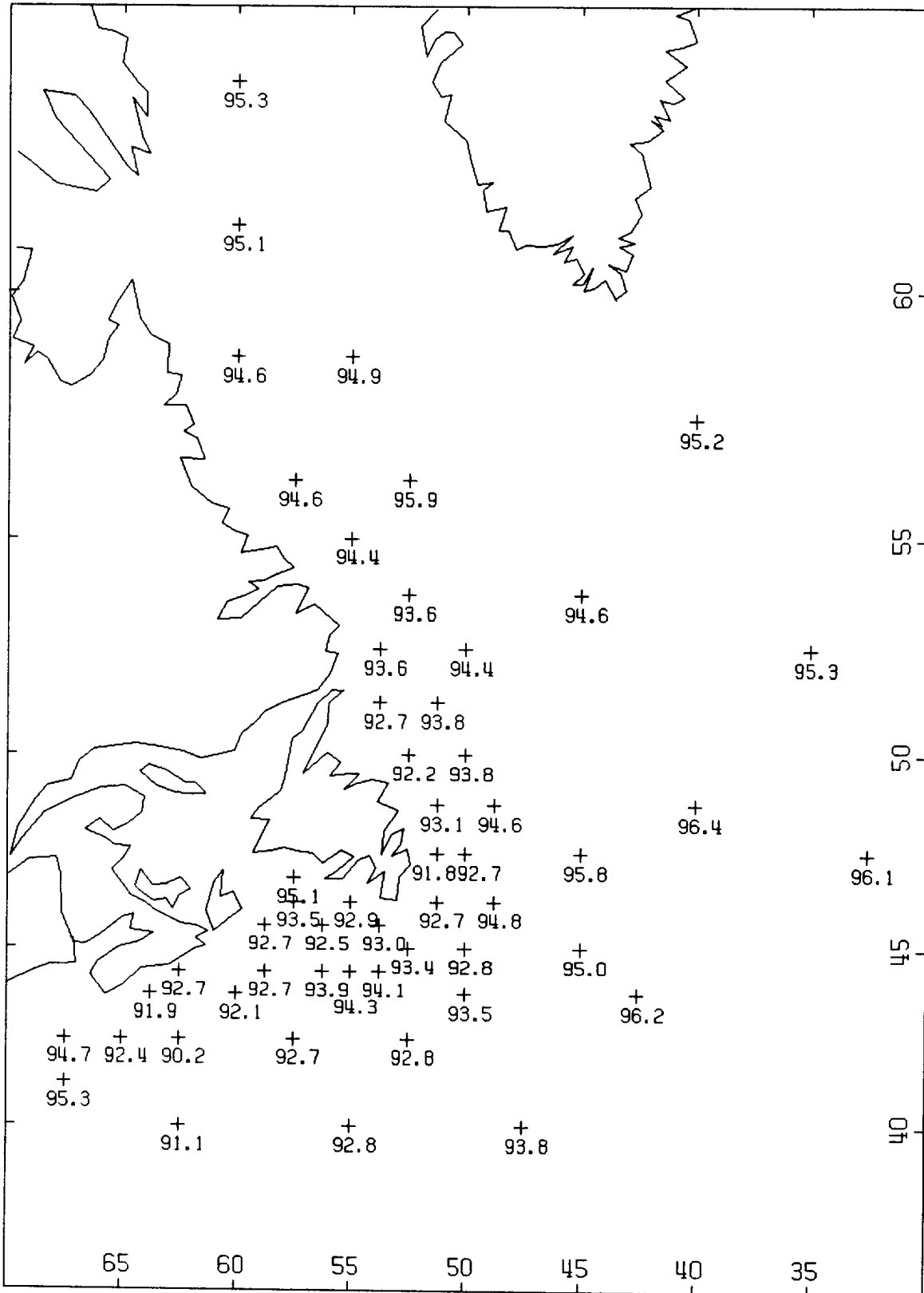
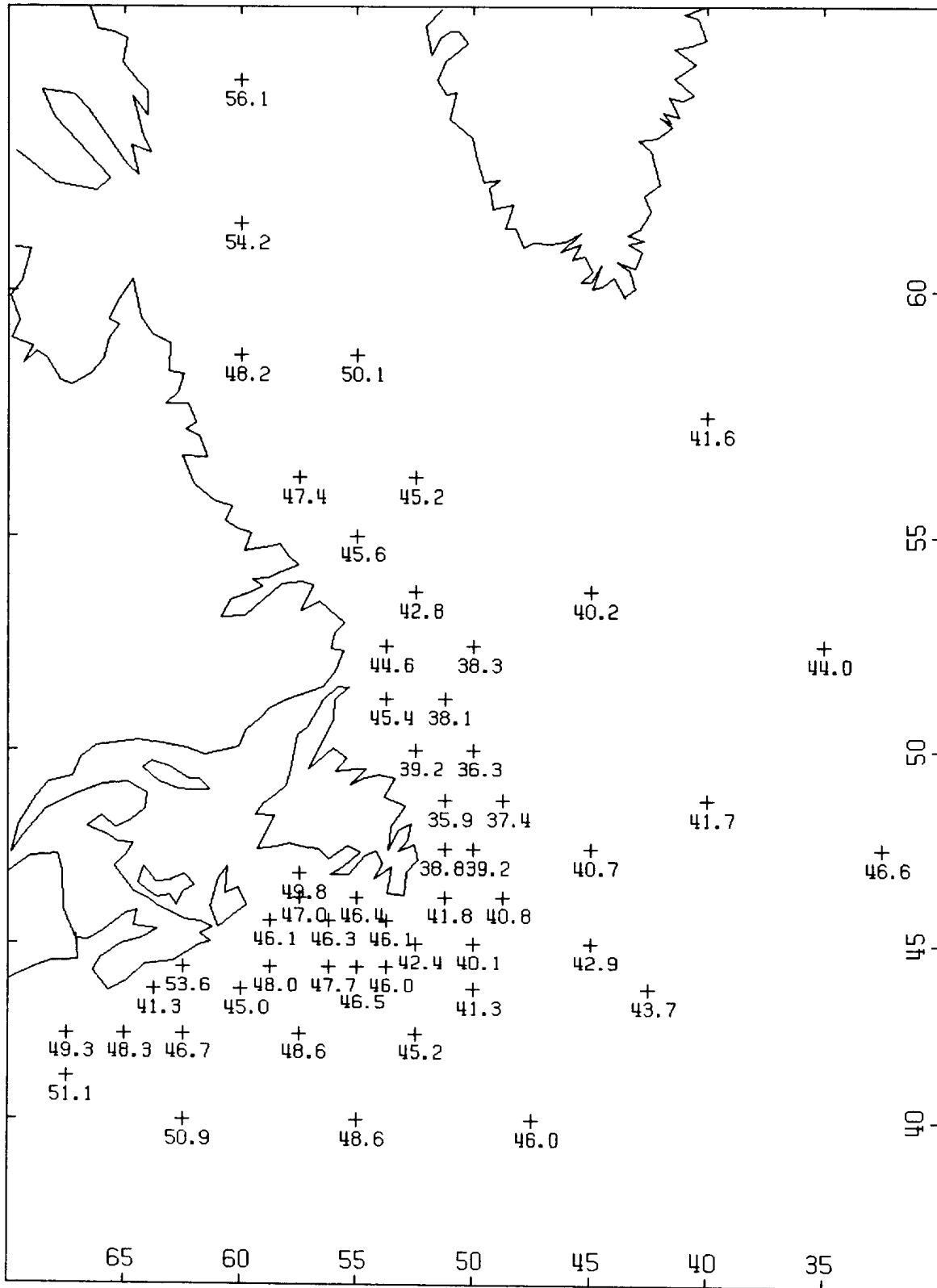Fig. 3 (continued) b) winter data.

Fig. 3 (continued) c) fall data.

Fig. 4 Percentage occurrence of records according to
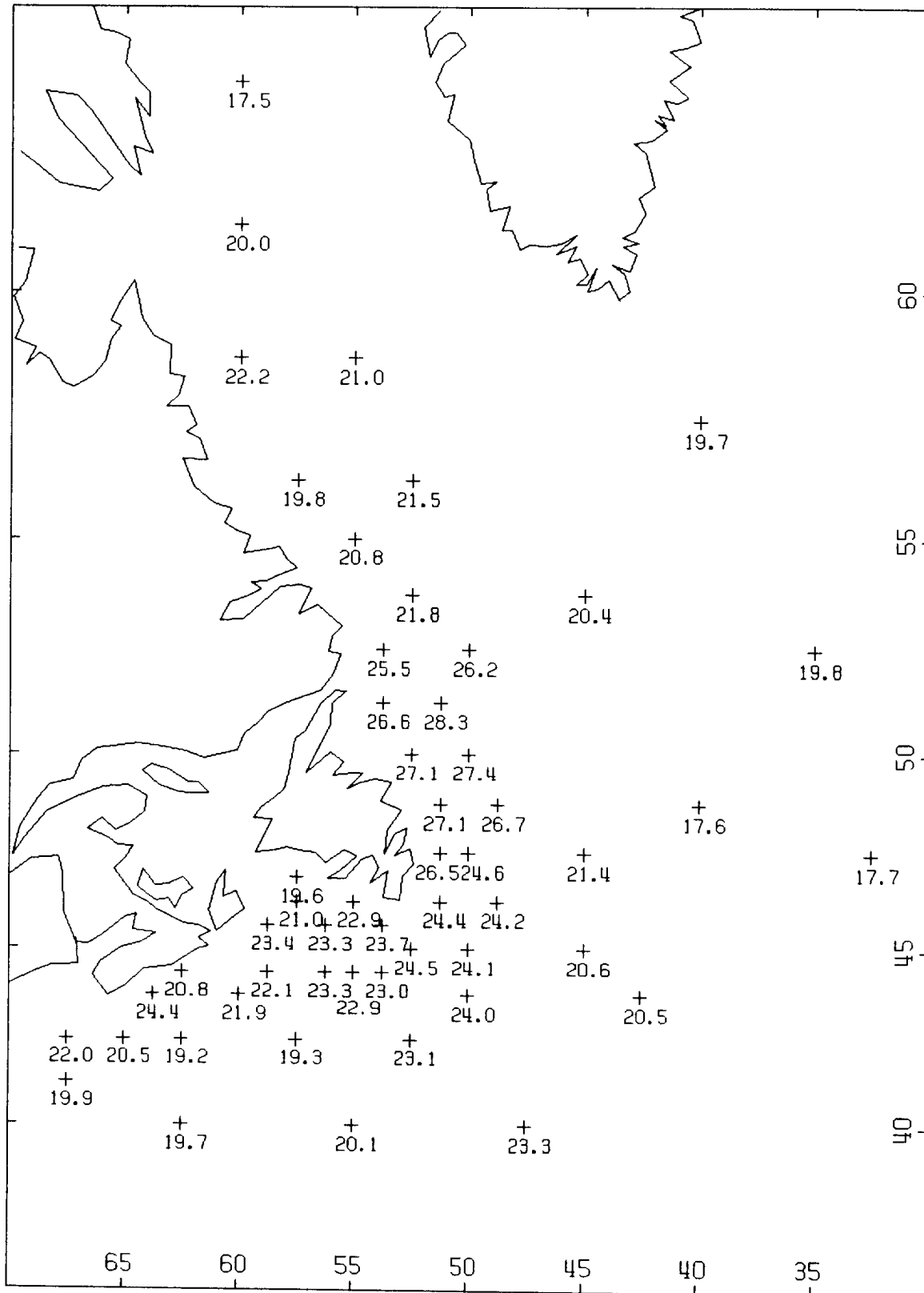spectral type. a) ICODE = 10.

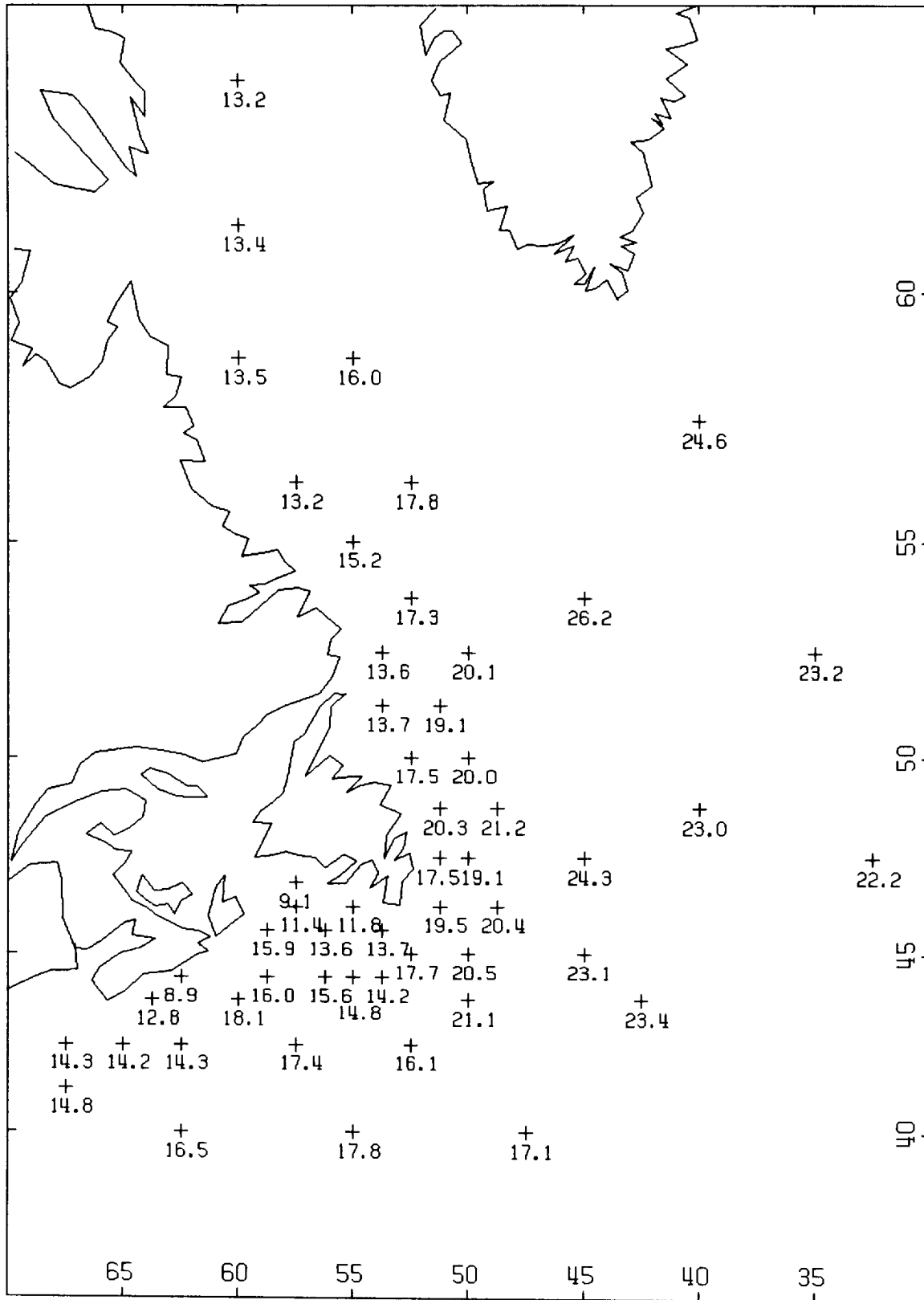Fig. 4 (continued). b) ICODE = 21,31,...
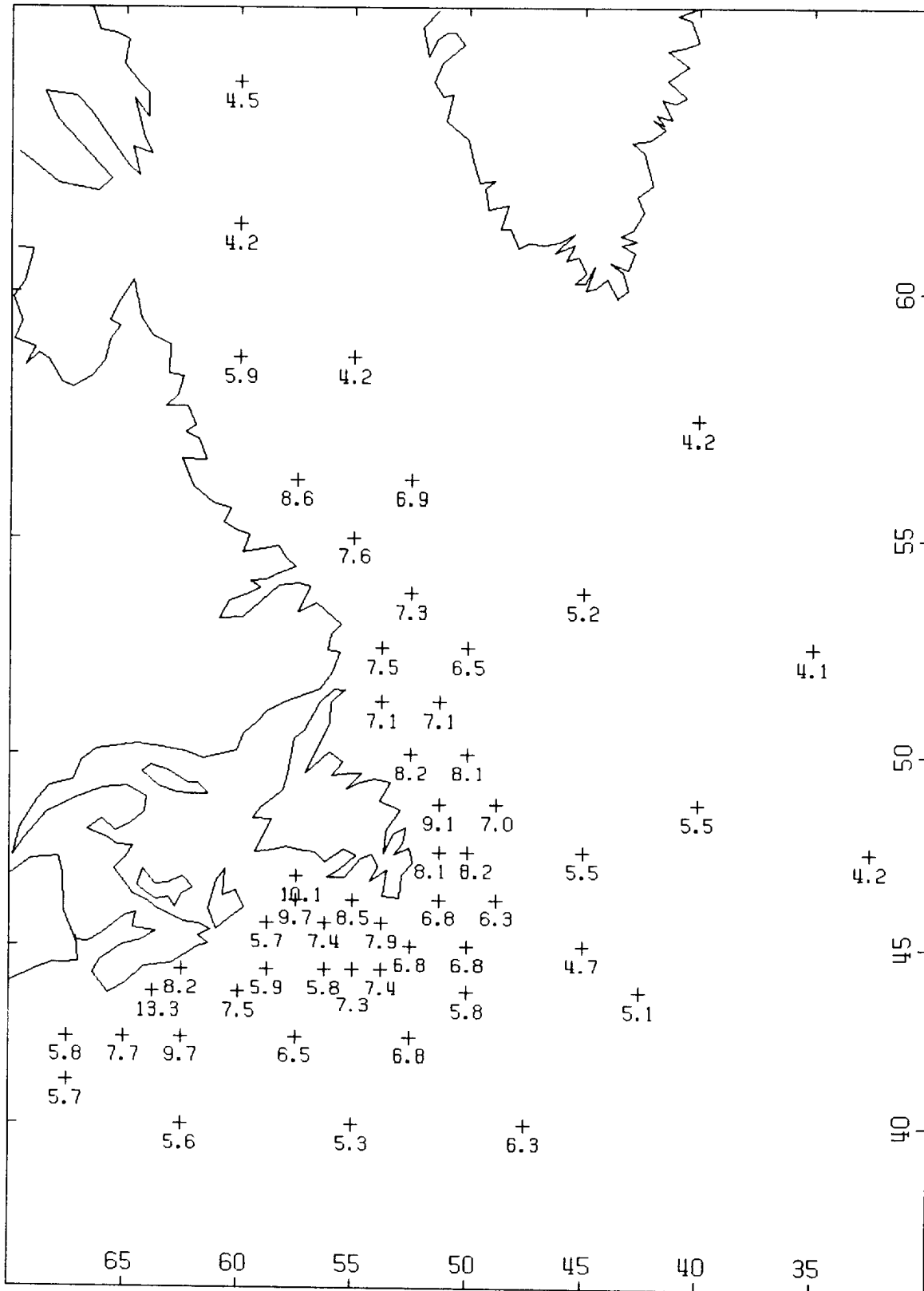
Fig. 4 (continued). c) ICODE = 22,32,...

Fig. 4 (continued). d) ICODE = 23,33,...
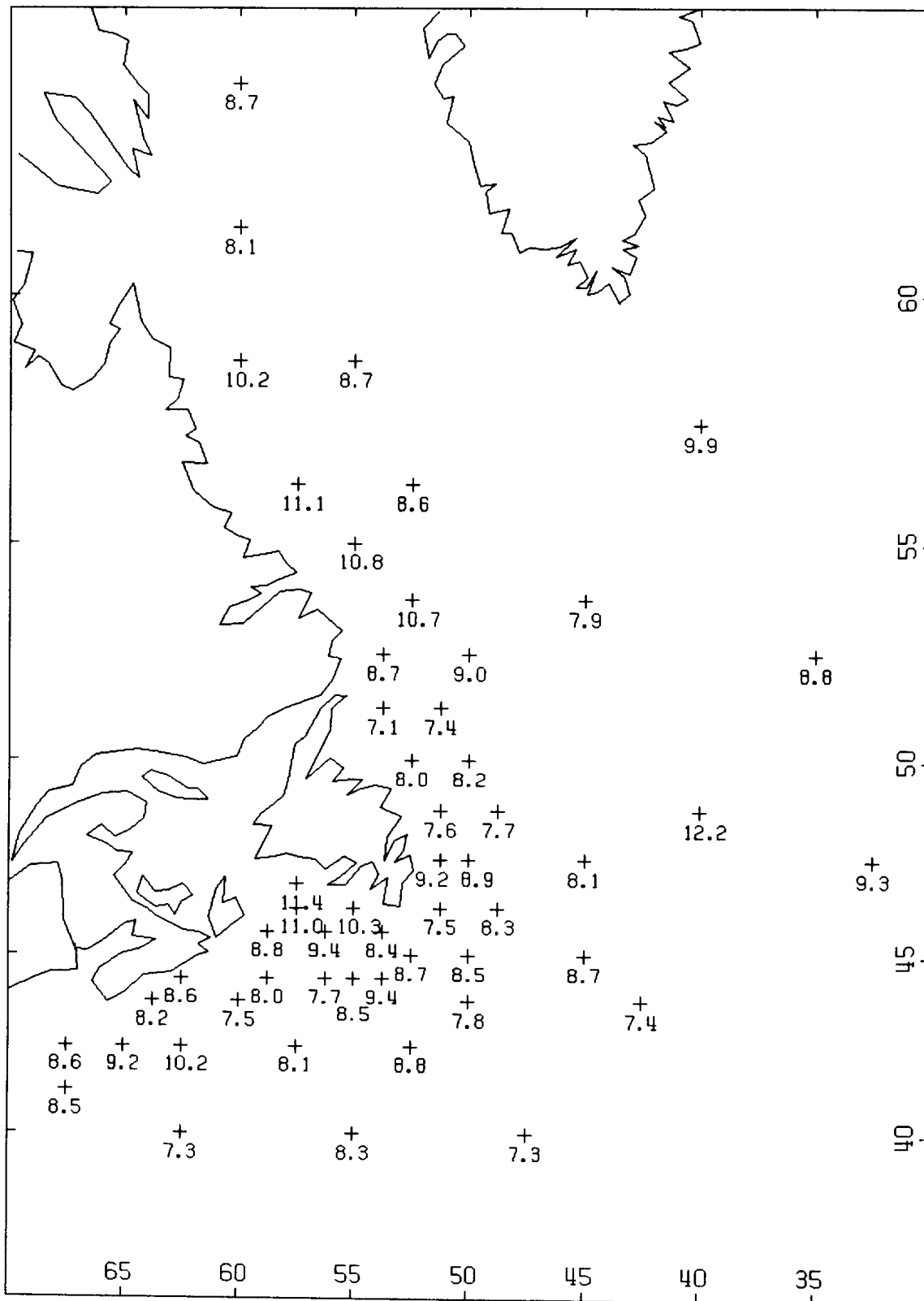
Fig. 4 (continued). e) ICODE = 24,34,...

2.3 Regional Analysis

The 53 grid point locations supply good geographic coverage of the Western North Atlantic. However, they also provide an unwieldy amount of information to assimilate on a location by location basis and could lead to the loss of important regional signals when the volume of data forces the averaging of locations together. This problem has been recognized in other studies as 'representative' grid points for the Grand Banks, Scotian Shelf, etc. have been established. In our application we wish to develop a representative family of probability spectra for a region by averaging over sites which have a high-level of covariability.

The demarcation of regions is often based on an assessment of geographic features such as water depth and fetch limiting land masses, subjective selection upon examination of mapped measured physical signals, etc. One approach is the examination of the maps of spectral type in Fig 4 and of maximum significant wave height (HSIG) in Fig 5 . For example, the locations where single peak spectra make up less than 40% of the records (Fig 4a) while multi peaked spectra, with peaks separated in both period and direction, make up over 25% (Fig 4b) corresponds to a region off the east coast of Newfoundland. Offshore vs inshore regions can be differentiated by the occurrence of multi peaked spectra, with peaks separated in direction but close in period, (Fig 4c) being less than or greater than 22% of the records. A separation between the Scotian Shelf – Western Grand Banks and Eastern Grand Banks can be made based on the number of occurrences of single peaked spectra (Fig 4a) being less than or greater than 45% and by the observations of maximum HSIG (Fig 5) greater than 10.5m on the Eastern Grand Banks. This approach however is limited as the range in the mapped information is small and possibly sensitive to slight changes in the data sets, the selection process is subjective based on pre-conceived expectations of the variability (e.g. with no preconceptions, when examining Fig 4a , one may decide to group all sites with occurrences >50% which would result in an averaging of selected southern sites with the physically independent northern ones), and all temporal information has been lost so that an understanding of the covariability cannot be made.

An objective numerical method, which utilizes the covariance information inherent in the spatially distributed time series and whose results can be critically assessed, is desired. A method based on an initial eigenvector analysis of either the spatial covariance (empirical orthogonal function technique) or correlation matrix (principal component technique) is specifically designed to reduce large volumes of spatial data into a manageable format. It results in

the partitioning of the total spatial variance structure into sets of
mutually independent behavioral Modes. The eigenvectors can then be
rotated through a statistical technique known as a factor analysis to
determine regional grouping.
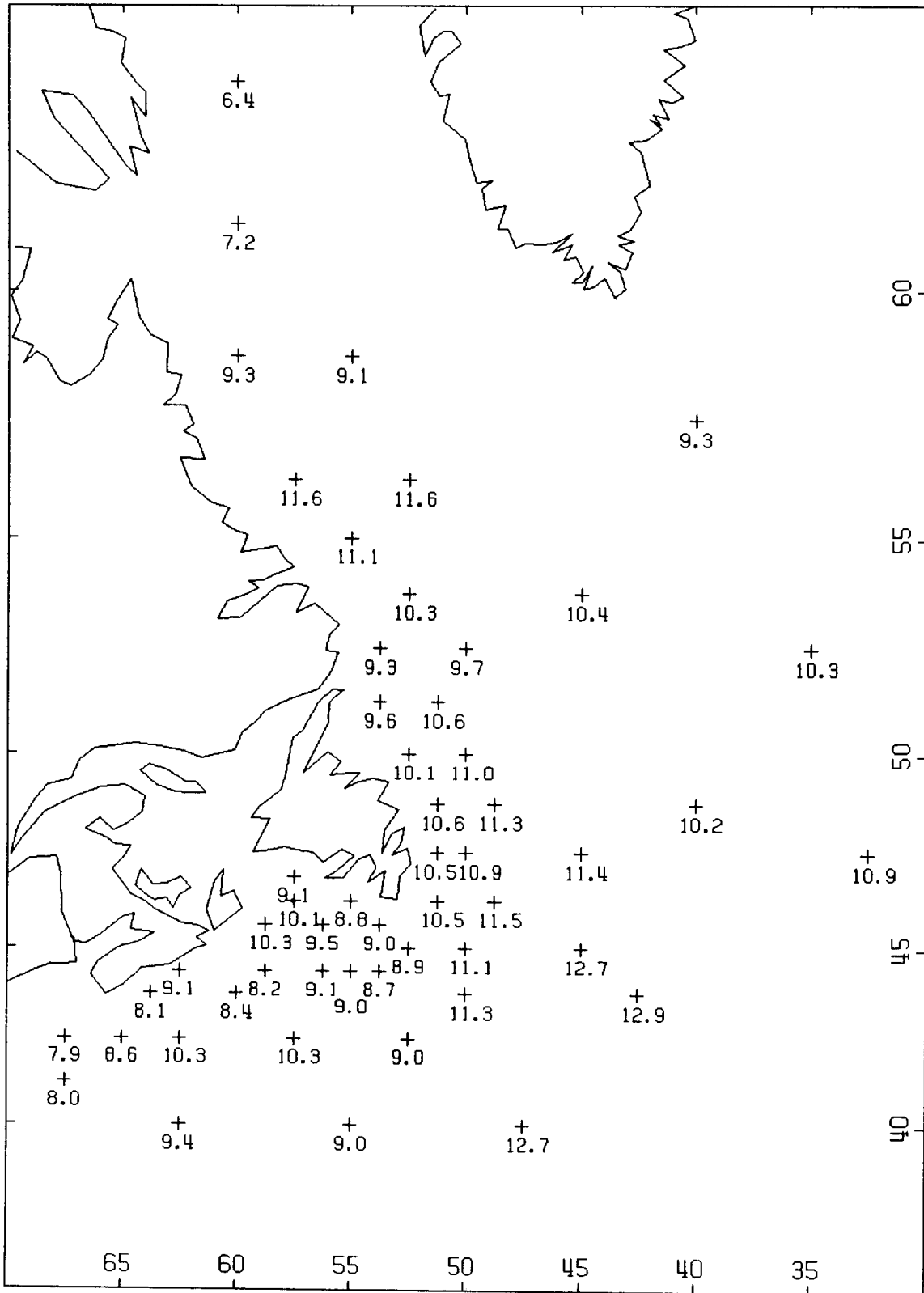
Fig. 5 Measured maximum HSIG at each grid point.

The procedure consists of four steps. First, the eigenvectors and eigenvalues are determined for the M x M covariance matrix obtained from the M x N matrix of M locations at which N simultaneous samples are made. Second, various test are applied to assess which P(P ≤ M) eigenvectors are significant and the data set reduced from the original M site locations to P "eigen locations". The remaining (M – P) vectors cannot be differentiated from random noise. Third, a principal component analysis is performed which consists of a further normalization of the covariance matrix by the sample standard deviation and of the eigenvectors by their respective eigenvalues. And fourth, a factor analysis, using the VARIMIAX approach, is conducted on the P principal components which allows for regional grouping of covarying stations.

Step 1. Theory of eigenvalues and eigenvectors

Consider a grid of points at M locations not necessarily equally spaced, but sampled simultaneously in time over N samples. The sampled quantity (e.g. Significant wave height) will be denoted by $x'_{i,j}, i = 1,$ M; j=1, N. At each grid point we calculate the mean in time ie:

$$\underline{x'}_i = \frac{1}{N} \sum_{j=1}^{j=N} x'_{i,j}$$

and form the shifted set of time series with zero mean viz:

$$x_{i,j} = x'_{i,j} - \underline{x'}_i$$

We now wish to find a rotation of the data:

$$y_j^k = \sum_{i=1}^{i=M} x_{i,j} \phi_i^k \tag{4}$$

where k is the eigenvector index (k=1,M), under two conditions:

a. that the variance:

$$\sigma^2 = \frac{1}{N} \sum_{j=1}^{j=N} (y_j^k)^2 = \frac{1}{N} \sum_{j=1}^{j=N} \left(\sum_{i=1}^{i=M} x_{i,j} \phi_i^k \right)^2$$

be a maximum and

b. that the rotation vector normalizes to 1.0:

$$\sum_{i=1}^{i=M} (\phi_i^k)^2 = 1$$

The procedure used is Lagrange's method of undetermined multipliers:

$$\sigma^2 = \frac{1}{N}\sum_{j=1}^{j=N}\left(\sum_{i=1}^{i=M}x_{i,j}\phi_i^k\right)^2 + \lambda^k\left(1 - \sum_{i=1}^{i=M}\phi_i^k\phi_i^k\right) \tag{5}$$

where λ^k is an undetermined multiplier. Minimizing eq 5 with respect to ϕ_l leads directly to the linear set of equations:

$$\sum_{i=1}^{i=M}\left(\frac{1}{N}\sum_{j=1}^{j=N}x_{i,j}x_{l,j}\right)\phi_i^k = \sum_{i=1}^{i=M}C_{i,l}\phi_i^k = \lambda^k\vec{\phi^k} \tag{6}$$

where $C_{i,l}$ is the cross-covariance matrix.

Equation 6 is the eigenvector equation where λ^k is referred to as the eigenvalue and the $\vec{\phi^k}$ is the rotation which maximizes the variance of the output series y_j^k. In general, there will be M eigenvalues and eigenvectors denoted by $\lambda^1 > \lambda^2 > ... > \lambda^M$ and $\vec{\phi^1}, \vec{\phi^2}, ..., \vec{\phi^M}$. Since the cross-covariance matrix is symmetric (Hermitian in the case of complex quantities), it can be shown that the eigenvalues are real (and positive definite since the off diagonal elements are < diagonal elements) and the eigenvectors are orthonormal, i.e.:

$$\vec{\phi^l} \cdot \vec{\phi^k} = \delta_{l,k}$$

where $\delta_{l,k} = 0, l \neq k; = 1, l = k$. Furthermore, the time series (y_j^k) associated with each eigenvector has a variance $\sigma^{k2} = \lambda^k$ and the sum of the eigenvalues $\sum\lambda^k$ equals the total variance of the input station time series. Eigenvectors are the most efficient means of representing the variance of the data grid. From a reduced eigenvector set where noise and signal have been separated, the time series y_j^k based on transformation of eq 4 will contain approximately the same amount of information (i.e. the signal) as the original data. Normally, the eigenvector problem is solved in terms of the complete solution and written in matrix form:

$$C\Phi = \Phi\Lambda$$

where C is the cross-covariance matrix as before, Φ is the eigenvector matrix whose columns are the constituent eigenvectors $\vec{\phi^k}$

and Λ is a diagonal matrix whose elements are the constituent eigenvalues λ^k .

Step 2. Determining the number of significant eigenvectors

There are various selection rules, based on either variance or temporal criteria, that can be used to determine the number of significant (i.e. above, noise) Vigenvectors so that a reduced data set can be established. A good discussion on these can be found in Preisendorfer (1988). In our application, three simple criteria will be examined. The first involves the drawing of a "Scree" plot of the eigenvalues. in descending order, vs the i-index. There usually exists a distinct change. in the slope of the plot demarcating the break between signal and noise. The selection of the break in the slope is somewhat subjective, However, and a more objective. method is useful. The Guttman Lower Bound criteria states that for a given eigenvector to be significant, it must explain more variance than that which is contained in the average input time series from all locations. A third procedure, Selection Rule "N", involves a statistically rigorous approach utilizing a series of Monte Carlo simulations with random numbers in order to determine the inherent noise levels of the analysis. If there is zero self-correlation in the input sample data set of N values, N realizations are performed. However, in our application self-correlation is expected and one must use an equivalent number of samples, $\hat{N}_v$. $\hat{N}_v$, for large N, is approximated by

$$\hat{N}_v = \frac{N(1 - \rho^2)}{(1 + \rho^2)} \tag{7}$$

where

$$\rho = \frac{\sum z(t)\, z(t-1)}{\sum z^2(t-1)} \tag{8}$$

and z(t) is the input time series after removal of the mean. $\hat{N}_v$ is calculated for each location and generally the minimum value is used in the simulation. Using the asymptotic theory of eigenvalues of large matrices (see Preisendorfer, 1988), one can determine a value

$$\beta = \frac{\hat{N}_v}{M} \tag{9}$$

and look-up tables have been established (e.g. Table 5.1 in Preisendorfer (1988)) for β from which one obtains the equivalent eigenvalue at index i which would arise from an analysis on random

noise. If the data eigenvalue is larger than the corresponding noise eigenvalue, than the associated eigenvector is significant at the 95% confidence level. The results of these tests will be discussed in Section 3 .

Step 3. Principal Components

Another approach to the eigenvector problem is that of principal components which differs from the eigenvector analysis described in Step 1 in two major respects. First, the input time series is further normalized by the standard deviation ie.:

$$x_{i,j} = (x'_{i,j} - \underline{x'}_i)/\sigma_i \qquad (10)$$

Consequently, the eigenvalues and eigenvectors of the cross-correlation matrix are calculated. The total variance of the grid, and hence the sum of the eigenvalues, is now equal to the total number of grid points. Second, the eigenvectors are normalized to the eigenvalues:

$$\sum_{i=1}^{i=M} \phi_i^j \phi_i^k = \lambda^j \delta_{k,j} \qquad (11)$$

This procedure offers a number of advantages in the analysis of reduced data sets. Consider a set of M grid data points, reduced to P eigenvector time series based on the principal component approach. The quantity:

$$Co_i = \sum_{j=1}^{j=P} \phi_i^j \phi_i^j \qquad (12)$$

is called the communality of the rotation and is the percent of variance at each grid point accounted for by the transformation. Thus if a particular grid point is totally different from the rest of the data, it will have a small (<0.5) communality. The amount of variance explained at each location will remain constant through the factor analysis.

Step 4. Factor Analysis

Several problems can arise with eigenvector/principal component analysis. First, if there are a large number of stations closely grouped together, accounting for a large portion of the total variance, the eigenvector reduction may reflect the covariability of the proximity of this grouping rather than underlying structure.

Second, virtually all eigenvector analyses exhibit a structure based on geometry rather than physical processes. For example, mode 1 of many analyses will have elements all of the same sign suggesting mutual covariability of the data. The second mode will show two distinct groups with antivariability. Furthermore, the eigenvector elements (or loadings) will be fairly small, with any one element accounting for only a small fraction of its associated grid point variance. The approach in factor analysis is to further rotate the eigenvector matrix so that the principal component (PC) loadings reflect simple structure (initially defined by Thurstone in 1947, see Richman, 1986 for a good review). This is the situation when the eigenvector elements make only large (>25% and preferably > 50%) or negligible (< 5%) contribution to grid point variance. The required rotation is of the form:

$$F = \Phi T \tag{13}$$

where F is the rotated eigenvector matrix, now called the factor matrix with column elements $\vec{b}$ which represent the underlying structure, Φ is the eigenvector matrix of PC loadings and T is a, yet to be determined, rotation. The rotation may be either orthogonal or oblique, depending on whether or not the columns are orthonormal. A widely used orthogonal rotation is the VARIMAX rotation. The squares of the columns of the factor loadings are treated as data and we consider the rotation which minimizes the quantity:

$$V = \frac{1}{P} \sum_{i=1}^{i=P} [\frac{1}{M} \sum_{j=1}^{j=M} (\frac{b_{i,j}^2}{Co_j})^2 - (\frac{1}{M} \sum_{j=1}^{j=M} \frac{b_{i,j}^2}{Co_j})^2]^2 \tag{14}$$

The rationale is to rotate the PC loadings so that large values are emphasized and grouped into similar classes. This is essentially a minimization of the fourth moment of the distribution of the PC loadings. The factor matrix now consists of P columns each representing one factor (or region in our application) whose row elements are the loadings associated with the corresponding Al stations. The columns of the factor matrix F are examined and groupings are determined by file following criteria:

a.   The columns are searched and factor loadings > 0.5 are identified.
b.   If two columns contain loadings > 0.5, the column with the larger loading is accepted as the required factor.

Stations whose maximum loadings share the same factor form part of the Same region. The results of this analysis will be discussed in Section 3 .

2.4 Statistical Analysis

The statistical analysis was applied to annual, winter and fall data for the individual grid point data sets, regionally grouped data and the combined locations. A grouping by spectral type was also performed on the combined data set. Records whose fit residual error was greater than 20% were excluded. The analysis consists of three steps: 1. a determination of the underlying fit parameter probability distributions; 2. the development of a family of spectra with known confidence limits; and 3. the establishment of predictive relationships between the averaged, statistical fit parameters and significant waveheight.

Step 1. Probability distribution analysis.

Juszko (1991), in a similar analysis on data from a single location, showed that the 10-fit parameters are best treated as a reduced set of eight parameters given as $\tan^{-1}(\delta_1/\delta_2)$, ω_{m1}, ω_{m2} $\ln(\lambda_1)$, $\ln(\lambda_2)$, $\ln(p_1)$, $\ln(p_2)$ and $(\theta_{m1} - \theta_{m2})$. For each input data set, the probability occurrence histogram of the derived parameters plus wind speed, for each of 11 significant wave height classes (i.e. 0-1m, 1-2,2-3, ..., 9-10, >10m), was established. Outlier points and default values (set by the fit software to limit parameter values to physically acceptable ones, e.g. ω_m cannot be greater than the maximum spectral frequency, or as a result of computer overflow limits, e.g. p < 60), as well as occurrences of zero wind speed, were removed. In general, these problems were most severe for the 0-1m height class. Bounds were set on the distribution. After testing various probability functions (e.g. Rayleigh, Beta, Gamma), the two-sided bounded Gaussian distribution was found to be the most versatile. A non-linear least squares fit was applied to the probability occurrence histograms and the location of the upper and lower 95% confidence limits, modal (position of maximum occurrence) and median (50% probability level) values were, established.

Step 2. Development of the family of spectra.

The family of spectra of a given data set are obtained by sequentially selecting each parameter examined in step 1 to be a "target" parameter, determining a catchment range about the target mode, median and 95% limits, and scanning all accepted wave records for occurrences when the equivalent parameter value falls within the catchment range. Once all records have been examined, the non-target parameter values are averaged. The catchment limits were established by selecting a fixed probability range about the .025 and .975 (95% limits), and the modal probability value (this equals 0.5 for the

median value but can vary for the mode for ill-behaved distributions and is found by scanning for the maximum functional value). The range used was initially set to a.025 probability adjustment, however, as there may be zero occurrences of target values failing within the specified catchment area, the probability adjustment is sequentially increased to .05, .075, ... until at least one occurrence is encountered. For each data set, the analysis produces a family of N x 9 spectra for each height class, where N (usually N=3) is the, number of probability levels examined. The individual target parameter modal (or median) spectra are averaged together to provide one overall modal (or median) spectrum.

Step 3. Predictive relationships.

Once the families of spectra are established, the results can be generalized by determining the predictive relationship between respective fit parameters (y) and significant wave height (x). The functional form

$$y = ax^b + cx + d \qquad\qquad (15)$$

was found to be the most versatile in Handling the various x-y relationships observed.

3.0 REGIONAL ANALYSIS

The numerical techniques of empirical orthogonal function (EOF), principal component (PC) and factor analysis (FA), described in section 2.3 , were applied to the data in order to understand the underlying behavioral modes of the wave climate and to determine the physical grouping of the stations.

3.1 Considerations

The first step in the analysis is the selection of the appropriate physical parameters to be mapped. As much of the later statistical analysis is based on the separation of information into significant wave height classes, significant wave height (HSIG) is a logical choice. The significant wave height has the advantage of being a well-established wave parameter, whose behaviour is less erratic than, for example, individual spectral model fit parameters, and will exhibit regional variability due to physical environmental factors such as the average wind fetch, water depth, etc. A major justification for the use of the 10-parameter model is the requirement for modelling directional wave energy. Therefore, it would be appropriate to also choose a physical parameter which could represent the regional behaviour of directional wave energy. The directional forcing vectors are readily established from the time series of wind speed and direction. Juszko and Graham (1992) showed that a wave "directional energy" vector could be obtained in a similar fashion using HSIG (in the place of wind speed) and vector mean direction (VMD) which can also be calculated over selected frequency ranges. An EOF or PC analysis using these vector time series would not only provide regional groupings similar to those found with HSIG, but the eigenvector components would also indicate the direction associated with the wave variance. In order to obtain the wave vector time series, the 10-parameter model spectra were regenerated and HSIG and VMD values calculated over four frequency ranges: 1) all frequencies; 2) < 0.5 rps, low frequency or swell 3) 0.5-0.85 rps, mid-frequency; and 4) 0.85-2.0 rps, high-frequency.

The EOF analysis is best performed oil continuous data which posed a slight problem with the existing data set due to the numerous winter data gaps. Small data gaps were rare but these occurrences were handled through interpolation. Consequently two separate analyses were performed, the first on all 53 locations after the largest data gap period was excised from all data sets, essentially resulting in the removal of a large portion of the winter information, and the second on a reduced data set of 34 locations of annual continuous data. As mentioned in Section 2.3 , the EOF analysis may be influenced by the geographic spacing of the measurement sites. Although the later FA

results should be less sensitive to this problem, it was verified by decimating the 53 and 34 grid point data sets to 32 and 21, respectively, through removal of some of the Scotian Shelf and Grand Banks sites. The, resulting maps are shown in Fig. 6 .

3.2 EOF Analysis Results

The EOF analysis was performed on the covariance matrices obtained from the time-centered series of HSIG (real valued analysis) and five energy vectors (Complex analysis) measured at the selected grid points of the four data sets (i.e. M=34, 53, 21 and 32). The five energy vectors were the wind vector and the wave directional energy vectors calculated over the four frequency ranges described in the previous section. In our application, time-centering was performed by removing the monthly mean in order to account for a possible annual signal. The number of significant eigenvectors (i.e. signal vs noise) was determined according to the rules described ill Section 2.3 . Fig 7a and b contain the "Scree" plots for the HSIG and vector analyses based on M=34. The choice of P significant eigenvectors associated with the break in the slope is difficult. The Guttman Lower Bound criteria results are shown as the arrows in both figures. In order to apply Selection Rule "N", the equivalent number of independent samples ($\hat{N}_v$) must first be established using eq. 7 and 8. In our analysis, we had an input time series of 4382 samples (for continuous annual data). After applying eq. 7 and 8, $\hat{N}_v$ ranged from 289 to 445 points depending on measurement site. As

$$\hat{N}_v dt' * N dt = 4382 * 6.0 hrs \tag{16}$$

this implies a self-correlation time scale of dt'= 2.5 to 3.8 days (i.e. the time necessary between two samples for them to be uncorrelated) which reflects the 2 to 4 day weather time scale. A value for β (eq. 9) can now be calculated (β = 8.5 and 13.1) and the equivalent noise eigenvalues obtained from look-up tables. The equivalent noise eigenvalues are plotted on Fig. 7a as the upper and lower dashed lines. According to Rule "N", any eigenvalues lying above the dashed lines are significantly different from noise. By performing these tests, it was found that for M=53, 34 and 32, the first six eigenvectors were significant and for M=21, only the first four were significant. This implies that using only six (or four) eigenvectors, one can regenerate the input signal time series, at a given measurement site, and explain

$$100 \times \frac{\sum\limits_{i=1}^{P} \lambda^i}{\sum\limits_{i=1}^{M} \lambda^i} \qquad (17)$$

percent of the total variance at that site. This quantity will vary from location to location depending on its covariability (noise levels are assumed relatively constant at all locations as indicated by the little separation in the dashed lines on Fig. 7a for the range in $\hat{N}_v$). The amount of variance explained at each site are shown in Fig. 8 a to c (M=34) for the HSIG, wind and total wave energy vectors, respectively. An example of a regenerated time series is shown in Fig. 9a (input) and b (regenerated) for the wind and total wave energy vector (cosine and sine) time series from grid point 259. It can be seen that the major features and sign convention have been retained through the transformation (eq. 4 using P eigenvectors) acting as a check on both the analysis and programming.
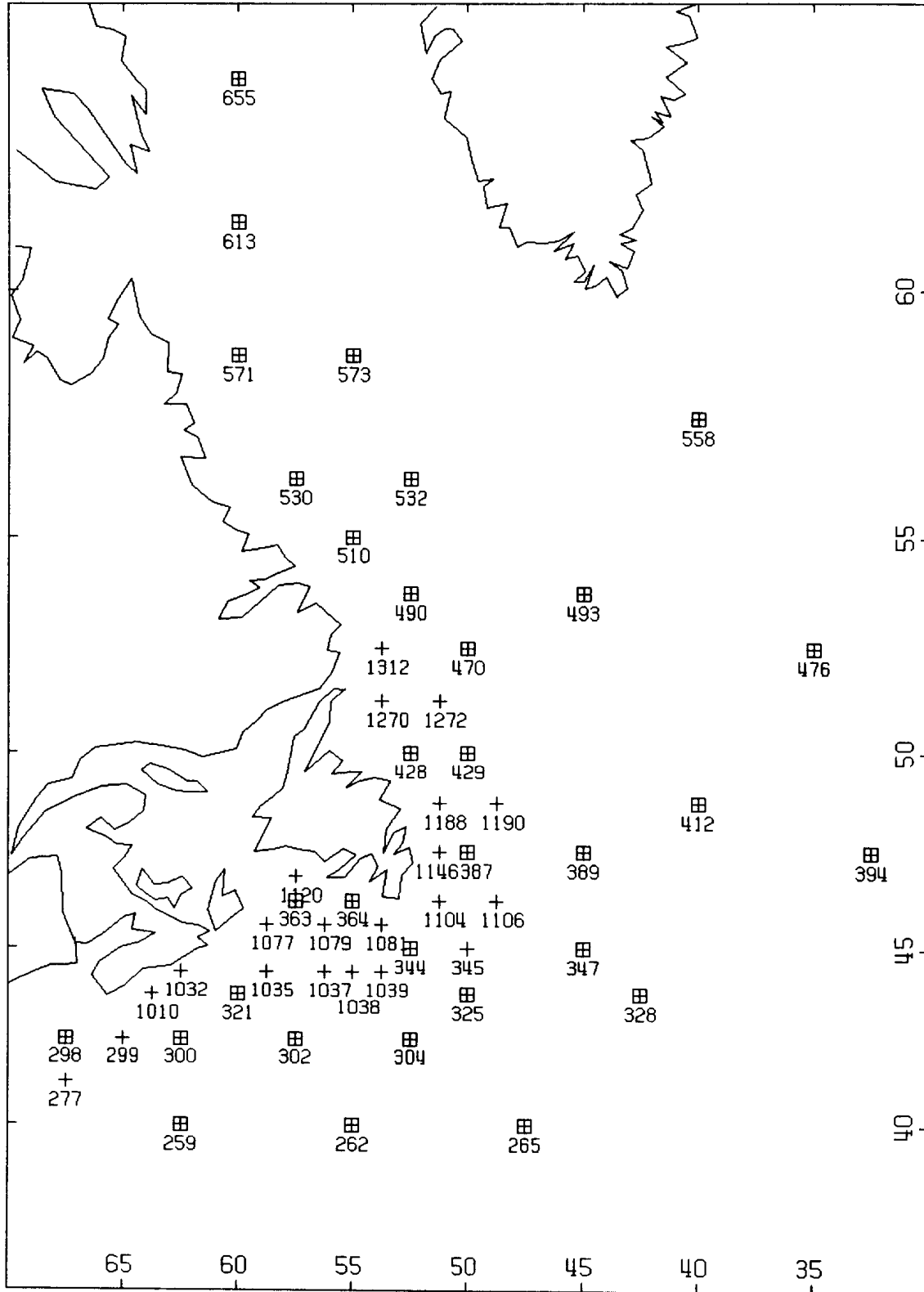
Fig. 6a Grid point locations used in the EOF analysis M=53.
        Squares indicate locations used in the reduced data set M=32.
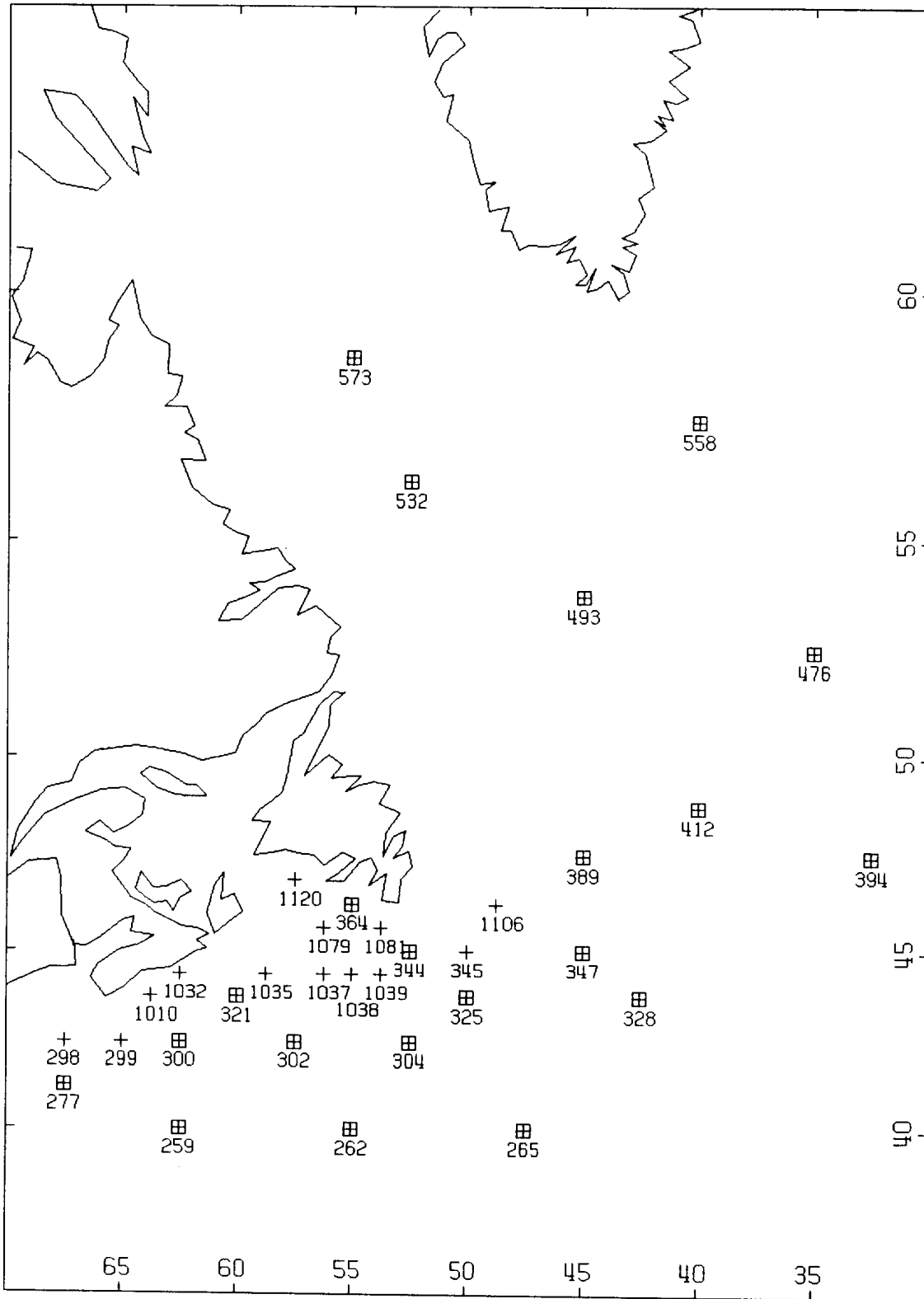
Fig. 6b Grid point locations used in the EOF analysis M=34.
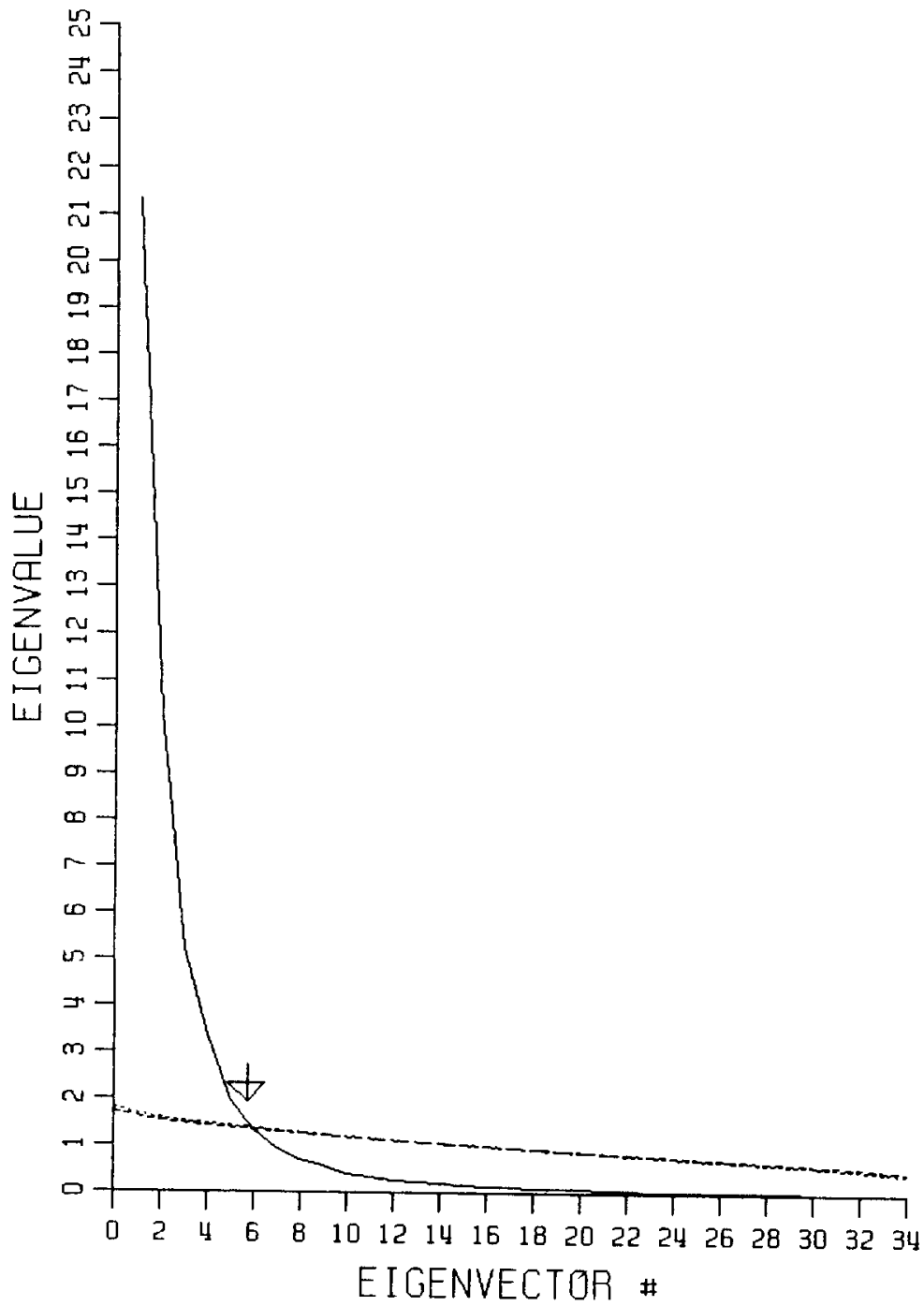Squares indicate locations used in the reduced data set M=21.

Fig. 7a Scree plot for HSIG analysis M=34. Arrow points to the Guttman Lower Bound Criteria. Dashed lines are the Selection Rule 'N' noise eigenvalue levels.
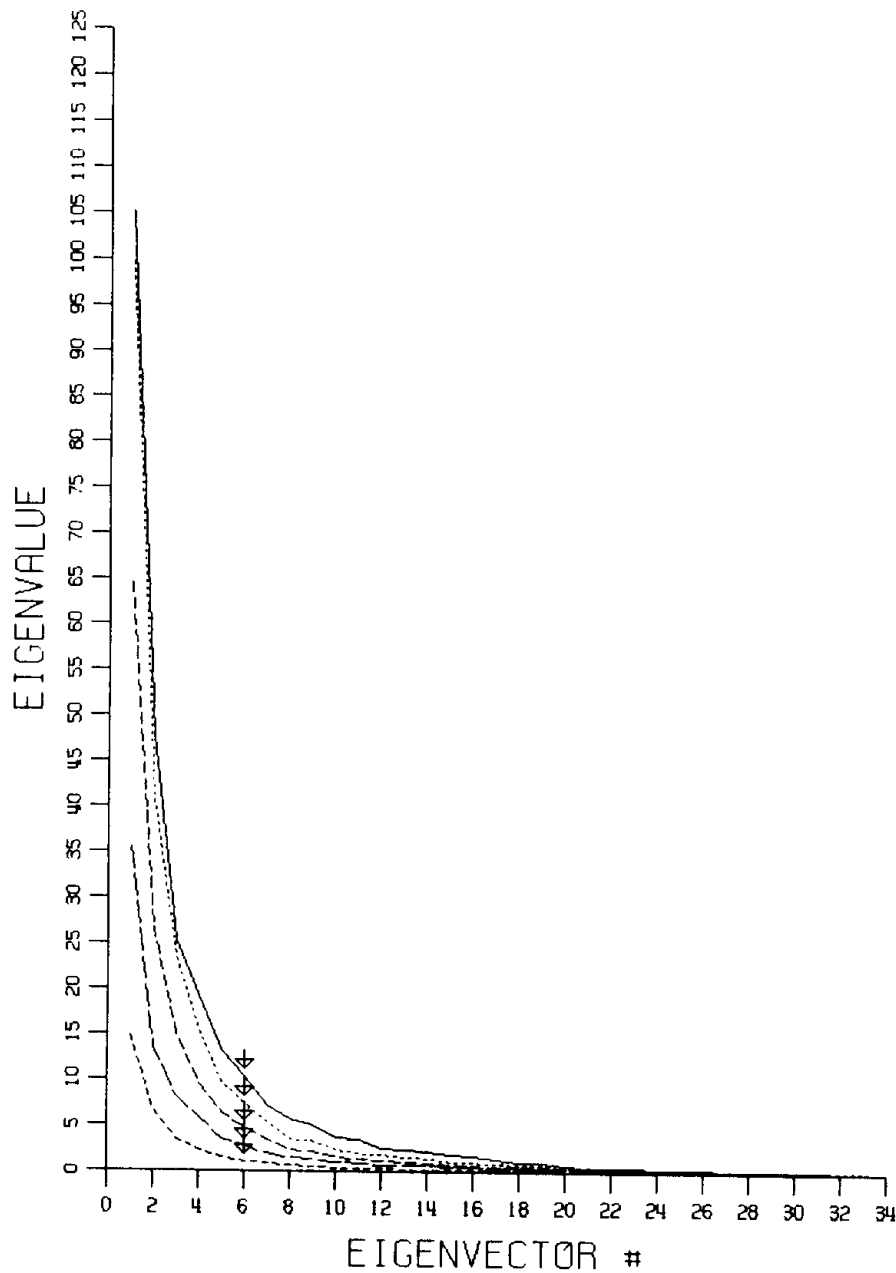
Fig. 7b Scree plot for wind (solid) and wave vector (dashed) M=34
    analyses. Arrows point to the Guttman Lower Bound Criteria.
    Increasing dash length represents the total, low, mid and
    high frequency wave component results, respectively.

Mapping of the eigenvector constituents provides information on the distribution of the variance into modes of behaviour. The first four modes for HSIG M=34 analysis are shown in Fig 10   while those for total wind and wave energy vectors are provided in Appendix 1 Figs A1   and A2 , respectively. The eigenvectors are normalized so that the sum of the constituents equals the corresponding eigenvalue. The M=34 data results are chosen for illustration as the annual, continuous time series is the best form of the data to use in an EOF analysis. In these figures, the percentage of the total variance explained by that mode is listed in the lower right corner, and in the case of Figs Al   and A2 , the eigenvector constituent consists of an amplitude and direction which are mapped as vectors, pointing toward the direction of variance propagation, and which have been normalized by the largest magnitude eigenvector constituent, also drawn in the lower right of the map. The relative magnitude of the eigenvector constituent or length of the constituent vector, provides a qualitative understanding of how well the sites group together. Opposite signs in Fig 10   indicate anti-variability and it can be seen that the modes reflect expected EOF analysis behaviour (i.e. first mode – moving together; second mode – two regions of anti-variability; etc.). The nodal lines in Fig 10   and the gyre-like features in Figs Al   and A2   appear to be geographically linked. In the case of the reduced data sets (M=21 or 32), the nodal lines and maximum coefficient values on the Scotian Shelf and Grand Banks were shifted slightly offshore, most noticeably for mode 2, reflecting the reduced weighting in this region. The vector analysis results show different first mode behaviour between wind and waves and similar behaviour at higher modes. The advantage of the EOF analysis lies in the separation of the variance into mutually independent modes which can then be examined dynamically. The eigenvector amplitude time series can be treated as any other time series and the correlation between the forcing and response signal modes, time scales, coherences, etc. can be calculated. For example, the correlation between the wind and wave vector results are shown in Table 1 . As expected, the high-frequency wave vectors are best correlated with the winds with little phase shift. The disadvantage of the EOF analysis is that the information needed to clearly differentiate covarying regions is hidden in the modal structure and can only be extracted through the use of factor analysis.
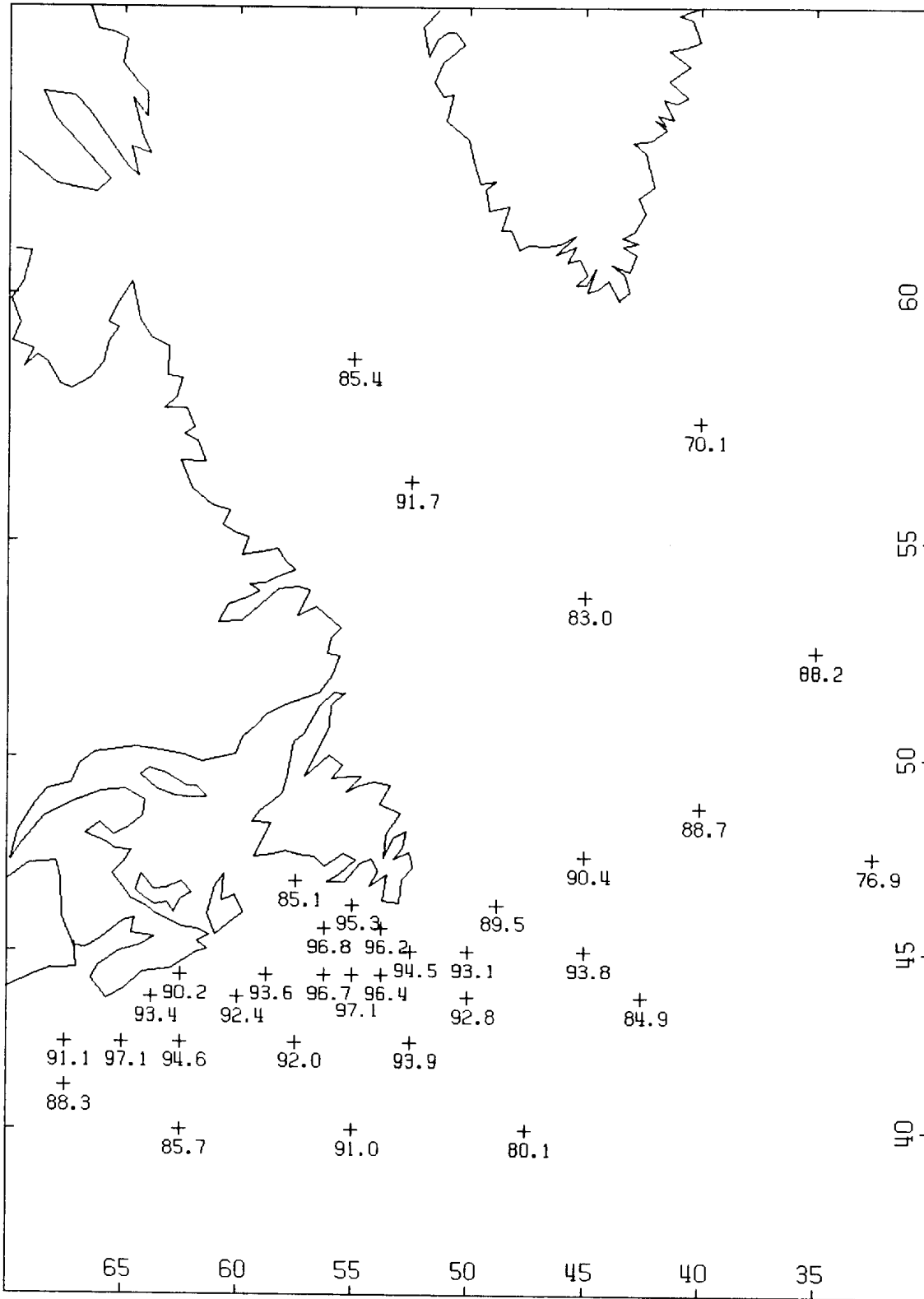
Fig. 8a The amount of variance explained by the first six
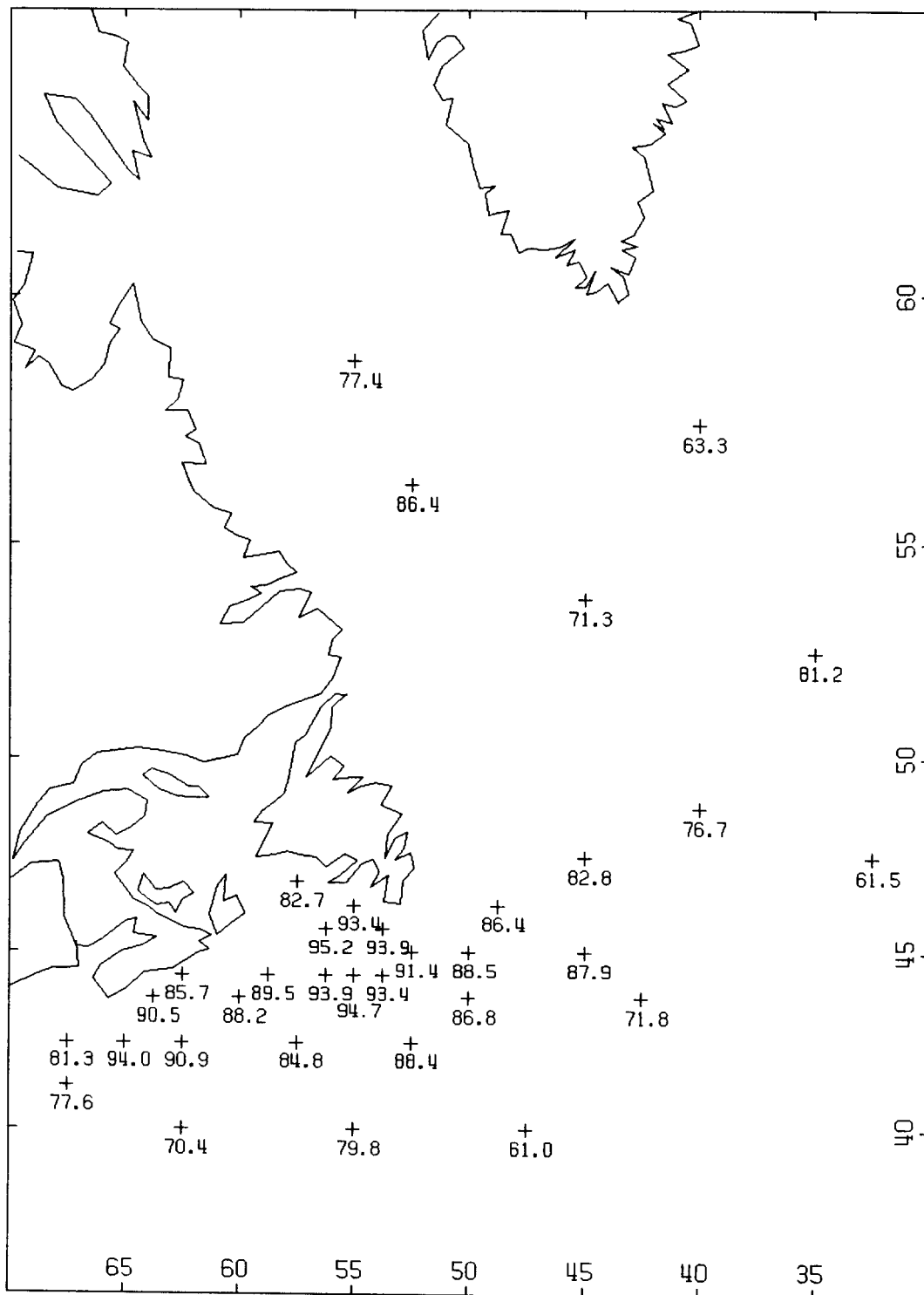   eigenvectors M=34 HSIG analysis.

Fig. 8b The amount of variance explained by the first six eigenvectors, M=34 wind vector analysis.
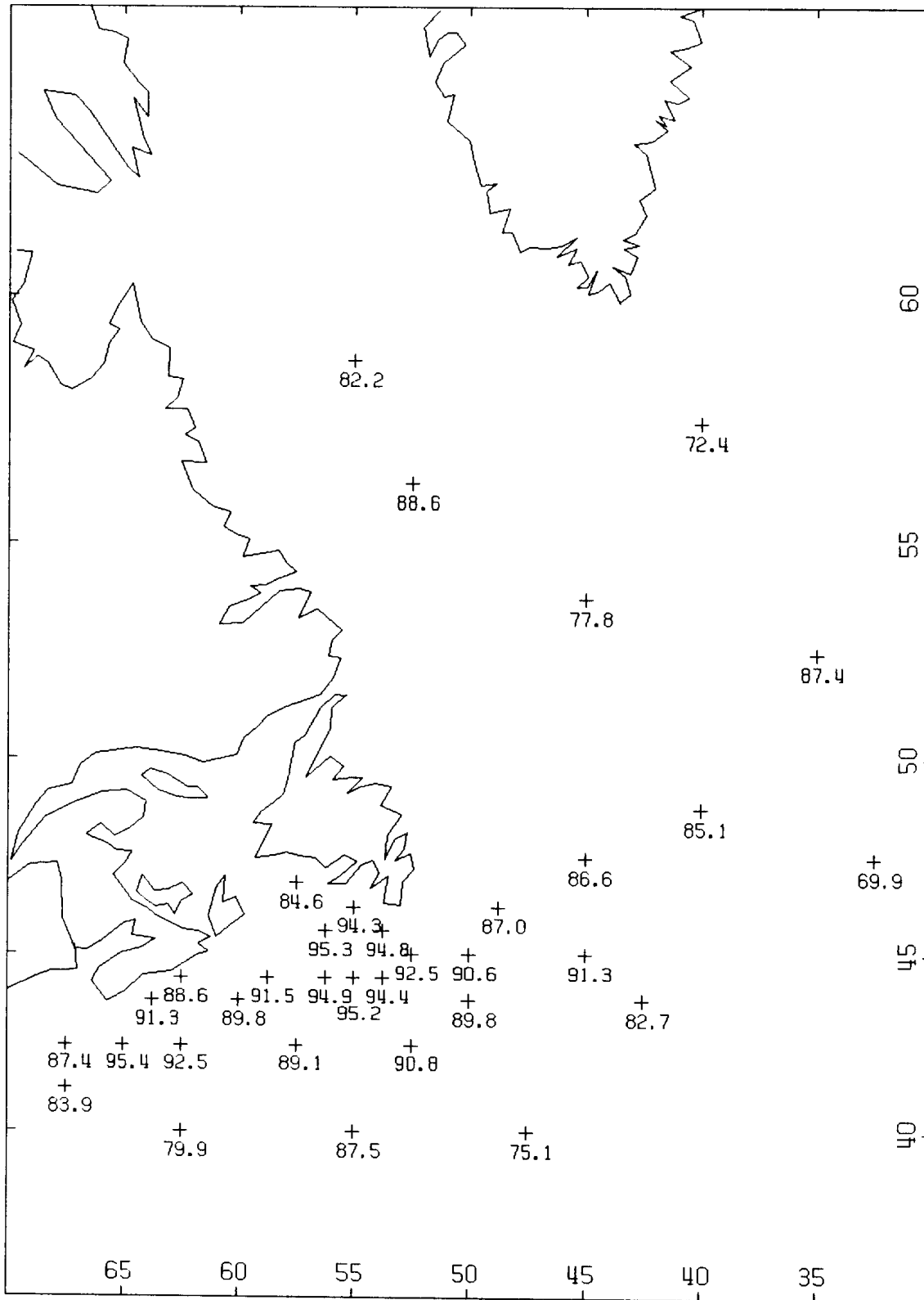
Fig. 8c The amount of variance explained by the first six
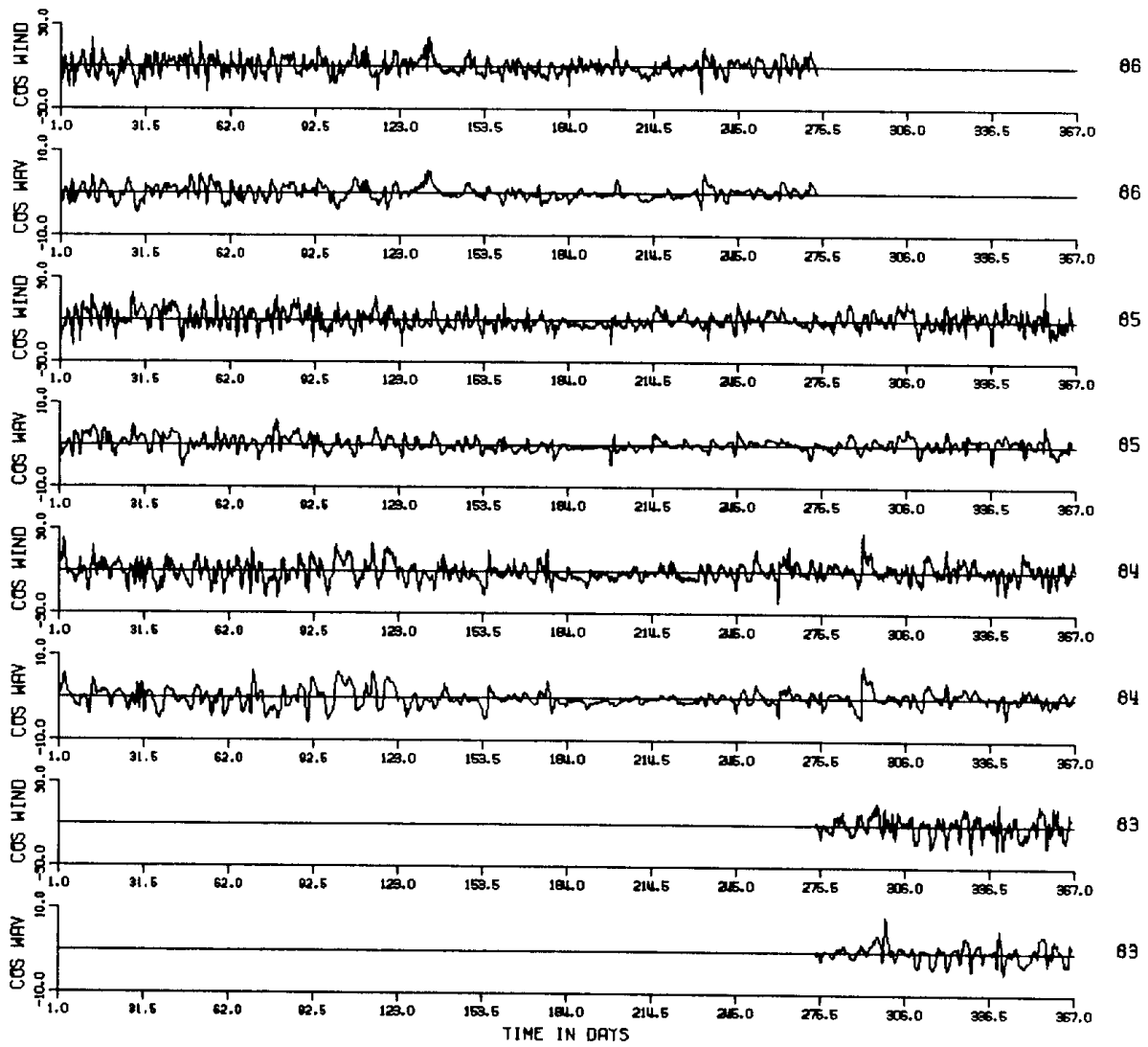eigenvectors, M=34 total wave vector analysis.

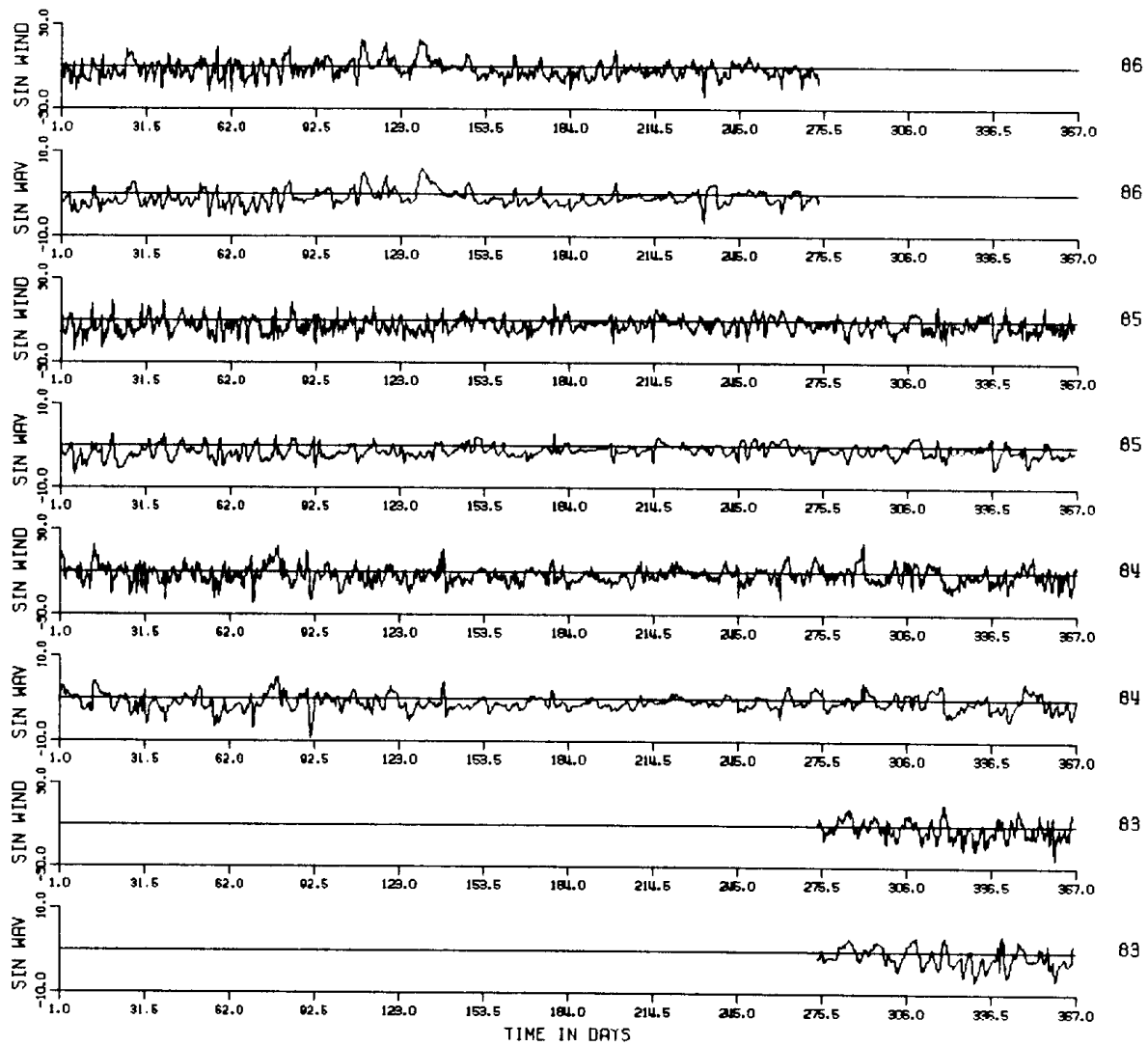Fig. 9a Input cosine time series of wind and total wave energy at grid point 259.

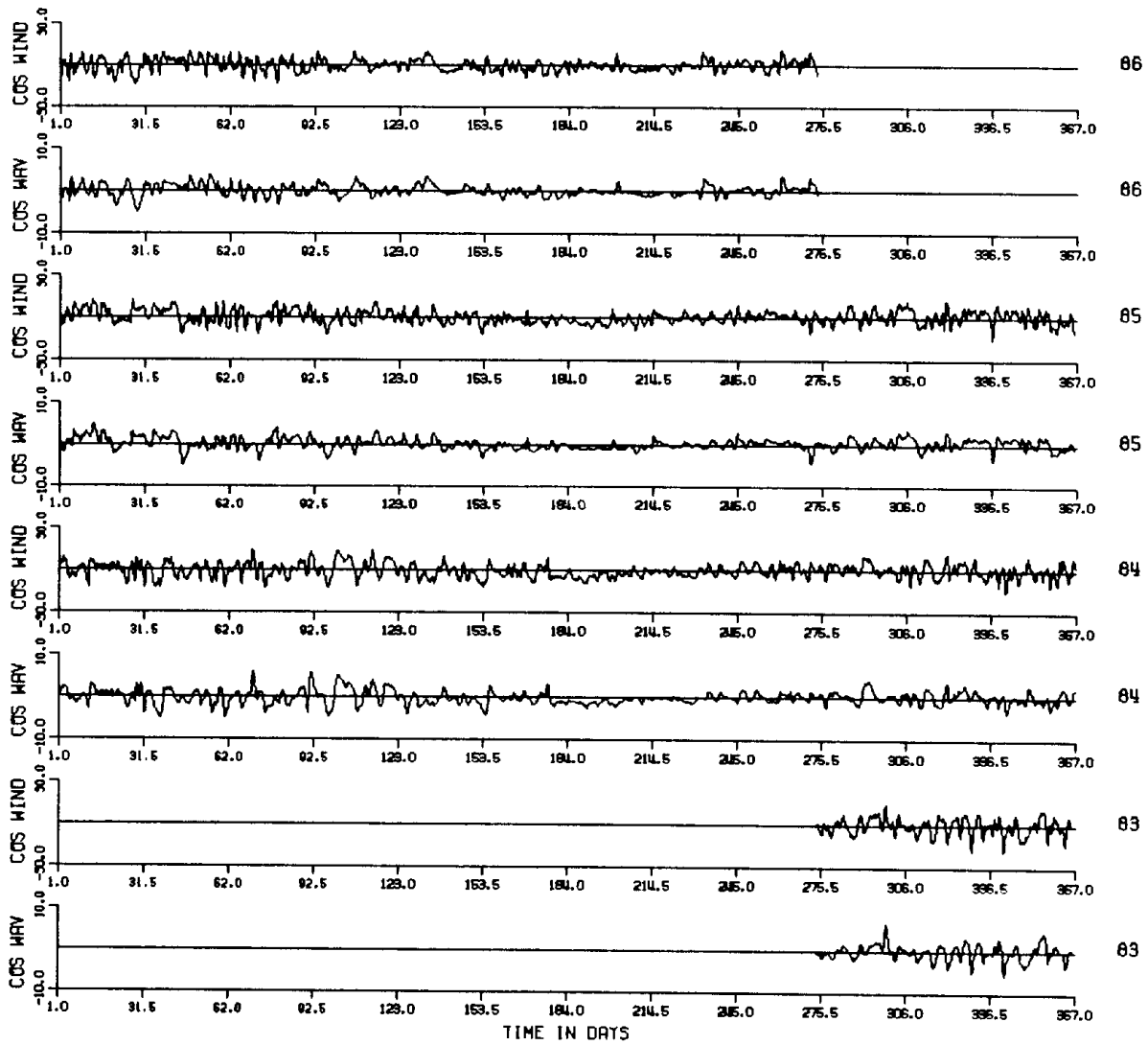Fig. 9a Input sine time series of wind and total wave energy
at grid point 259.

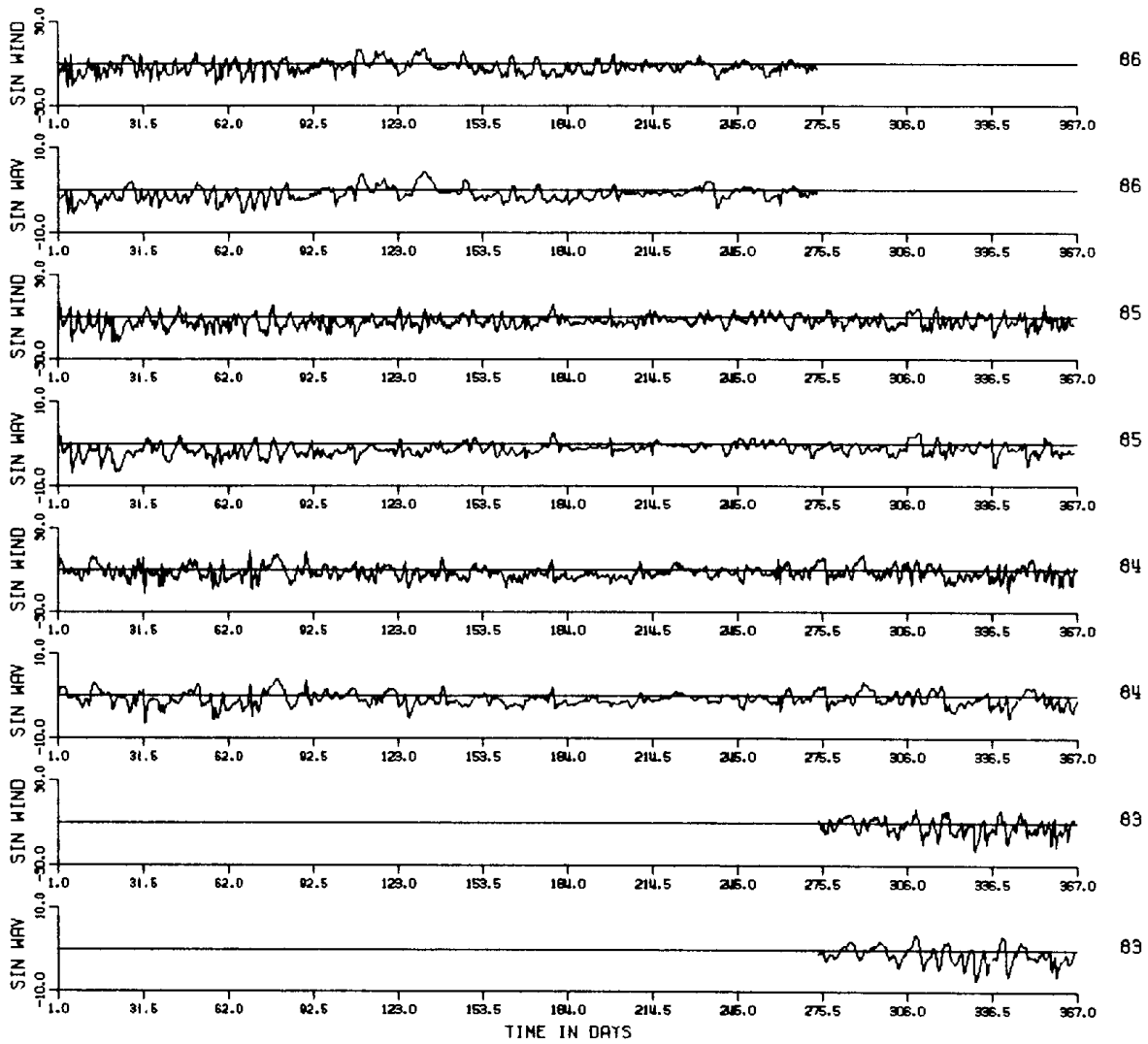Fig. 9b Regenerated cosine time series of wind and total wave energy at grid point 259.

Fig. 9b Regenerated sine time series of wind and total wave
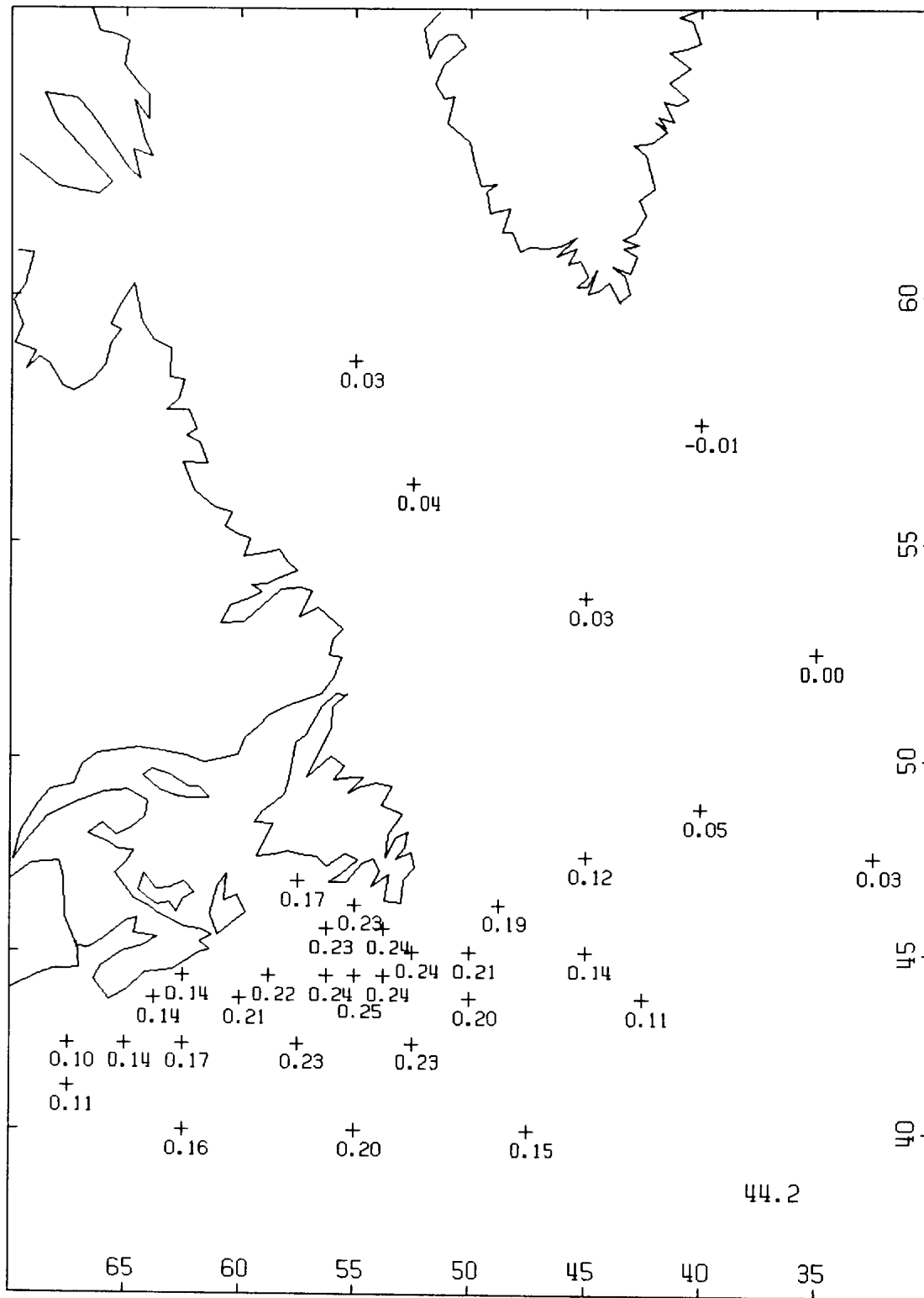energy at grid point 259.

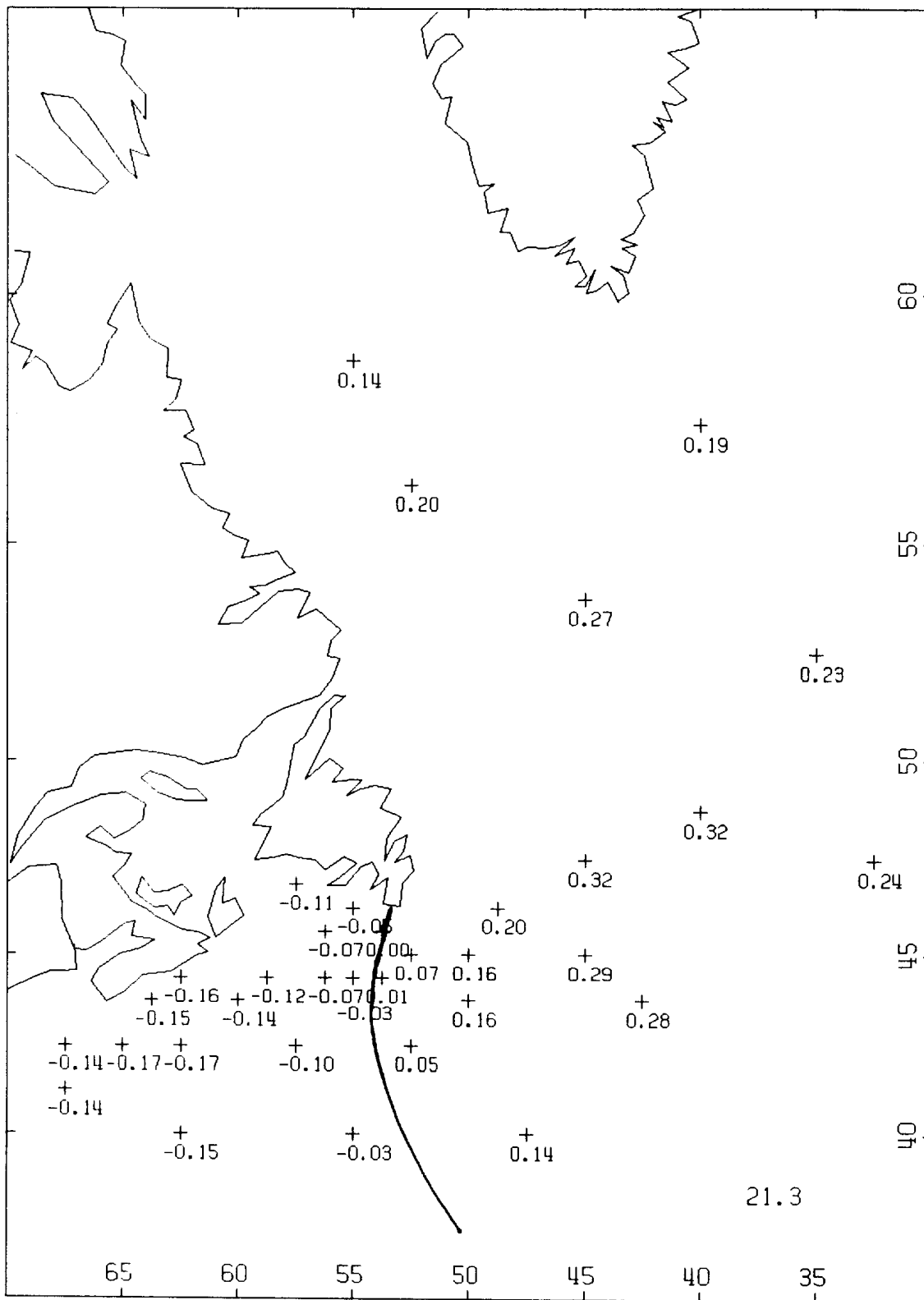Fig. 10a Mode 1 map of eigenvector constituent values,
        M=34 HSIG analysis.

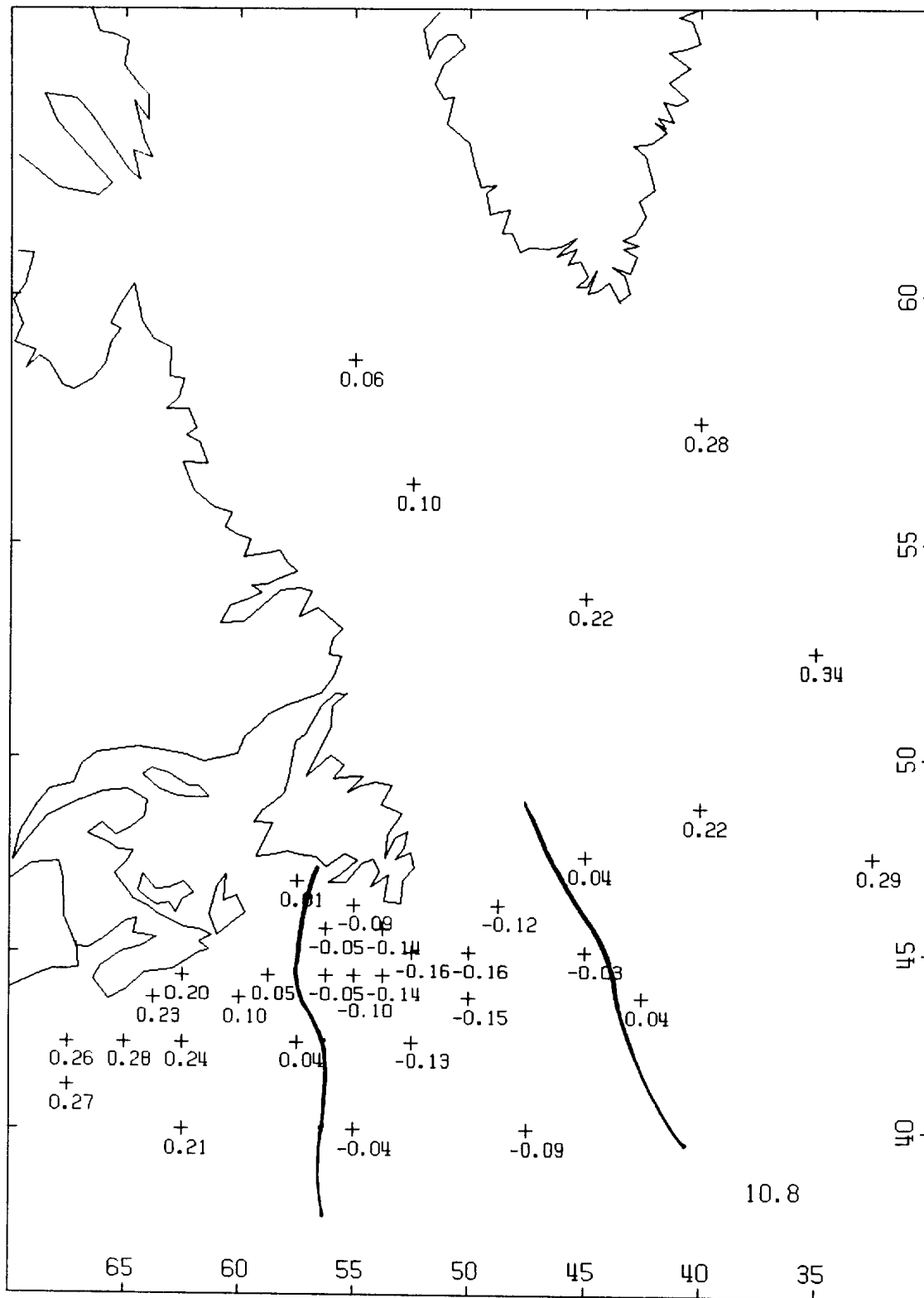Fig. 10b Mode 2 map of eigenvector constituent values,
M=34 HSIG analysis.

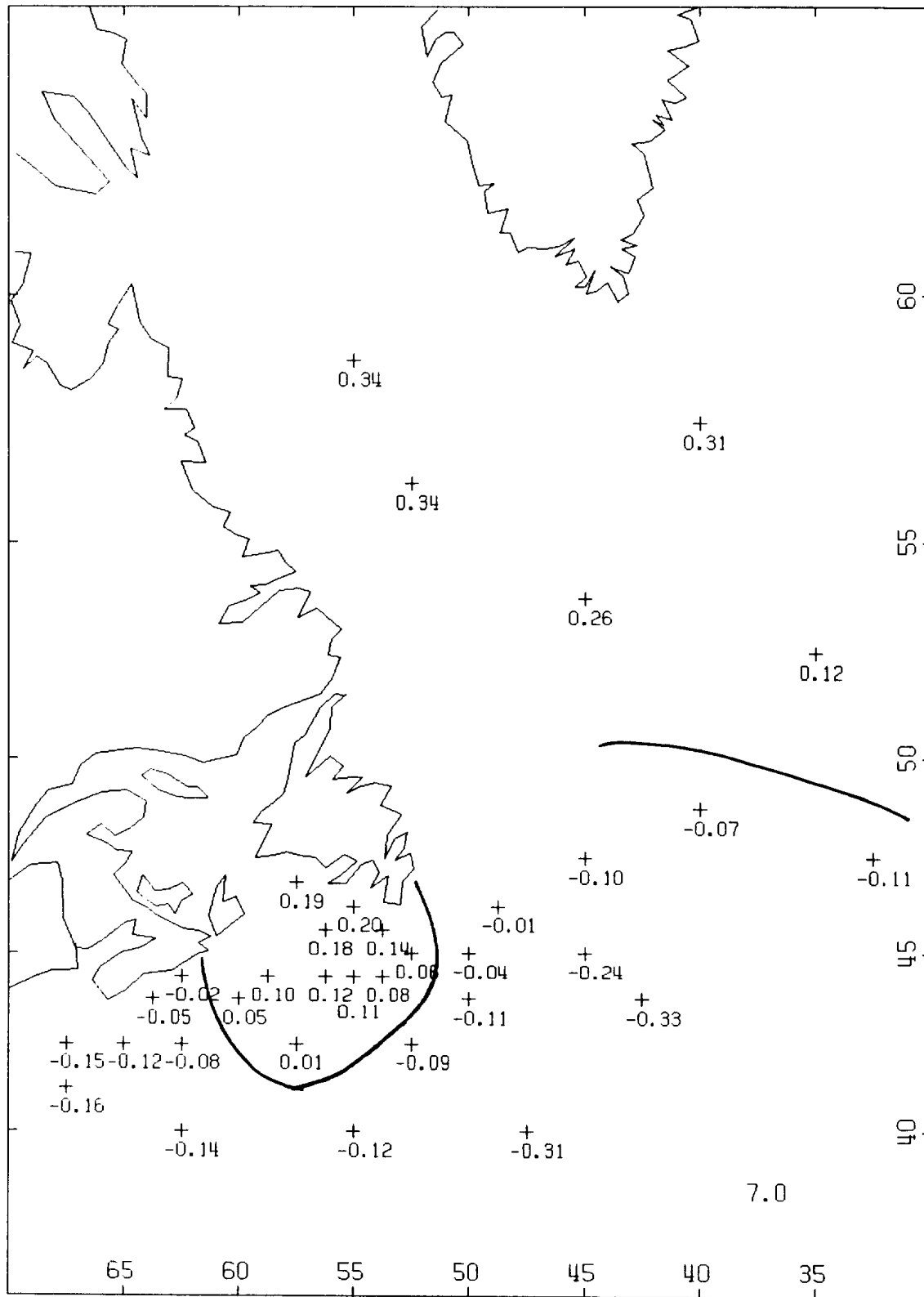Fig. 10c Mode 3 map of eigenvector constituent values,
M=34 HSIG analysis.

Fig. 10d Mode 4 map of eigenvector constituent values.
M=34 HSIG analysis.

Table 1. Correlation coefficients and phases (clockwise) between wind and wave eigenvector modes

MODE #	1				2		3		4
	CC	DEG.	CC	DEG.	CC	DEG.	CC		DEG
TOTAL	.78	278	.53	347	.52	10.8	.62		349
LOW	.38	210	.11	359	.23	352	.13		334
MID	.71	250	.41	364	.41	12.7	.51		346
HIGH	.86	325	.78	356	.75	351	.78		354

3.3 Factor Analysis Results

As described in Section 2.3 , FA uses the results from a principal component analysis applied to the P significant eigenvectors found in the previous section. The output consists of a table of factor loadings whose columns are searched for absolute values greater than 0.5 (i.e. indicating a communality of at least 25%). Sites, whose maximum loadings share the same factor, are grouped together as regions. For example, Table 2a   and b   lists the factor loadings for the HSIG analysis when M=34 and 53, respectively. These are mapped in Fig. 11 a   and b . Similar results are shown for the wind and total wave energy vectors in Tables 3 a ,b   and 4 a ,b   and Figs. 12 and 13 , respectively. In this case, the loadings are forced positive (i.e. one examines the magnitude of the complex vector). The M=21 and M=32 analyses, as well as the three wave vector component analyses, did not provide any significant additional information. The boundaries of the regions may shift slightly for the different data sets as marginal sites often share a high level of communality with more than one region and dominance is determined by the selected averaging. Grid points 262 and 328 were the two stations most likely treated as individual factors (i.e. low communality with the remaining 52 stations). There appears to be six regional groupings of the stations:

1.   Labrador Shelf – GP 655, 613, 571, 573, 530, 532, and 510
2.   Eastern Offshore Open Ocean – GP 558, 493, 476, 412 and 394
3.   Northern Coast of Nfld. – GP 490, 1312, 470, 1270, 1272, 428, 429, 1188, 1190, 1146, and 387
4.   Eastern Grand Banks and Flemish Cap – GP 389, 1104, 1106, 345, 347, 325, 328, 304, 262, and 265
5.   Western Grand Banks – GP 1120, 3631 364, 1077, 1079, 1081, 344, 10351, 1037, 1038, 1039, 321, and 302
6.   Scotian Shelf – GP 1032, 1010, 298, 299, 300, 277, and 259.

It is rewarding to note that the regions obtained with factor analysis
correspond to geographic areas which might be chosen intuitively.
However, we now have a strong statistical justification for their
selection and know how much variance is being accounted for if the
station is included in a regional averaging.

Table 2a. Factor loadings for HSIG M=34 factor analysis

GP#	F1	F2	F3	F4	F5	F6
259	0.283	-0.038	0.812*	-0.026	0.033	-0.339
262	0.512	-0.284	0.328	-0.016	0.069	-0.674*
265	0.175	-0.703*	0.084	-0.026	0.016	-0.518
277	0.055	-0.061	0.934*	-0.016	0.047	-0.028
298	0.071	-0.042	0.949*	-0.038	0.036	0.036
299	0.207	-0.005	0.962*	-0.044	0.004	-0.026
300	0.390	0.021	0.878*	-0.045	-0.014	-0.144
302	0.705*	-0.109	0.506	-0.009	0.017	-0.393
304	0.649*	-0.494	0.158	0.013	0.069	-0.494
1010	0.384	0.050	0.882*	-0.068	-0.044	0.023
321	0.724*	0.001	0.616	-0.024	0.000	-0.141
325	0.538	-0.750*	0.005	0.090	0.037	-0.256
328	0.003	-0.864*	0.036	0.004	-0.305	-0.090
1032	0.444	0.090	0.827*	-0.090	-0.032	0.059
1035	0.808*	-0.014	0.522	-0.015	0.005	-0.098
1037	0.902*	-0.150	0.319	0.015	0.031	-0.170
1038	0.905*	-0.244	0.226	0.028	0.052	-0.193
1039	0.873*	-0.360	0.148	0.039	0.068	-0.208
344	0.822*	-0.490	0.062	0.071	0.061	-0.132
345	0.618	-0.723*	-0.008	0.124	0.023	-0.103
347	0.154	-0.924*	-0.018	0.099	-0.221	-0.037
1079	0.937*	-0.084	0.287	0.018	0.027	-0.005
1081	0.920*	-0.303	0.120	0.052	0.064	-0.040
364	0.944*	-0.137	0.191	0.042	0.044	0.046
1106	0.531	-0.754*	-0.041	0.191	-0.052	0.066
1120	0.835*	0.084	0.359	0.003	-0.002	0.133
389	0.152	-0.823*	-0.057	0.236	-0.362	0.116
394	-0.066	-0.311	0.032	-0.061	-0.812*	0.060
412	-0.036	-0.565	-0.037	0.164	-0.725*	0.111
476	-0.060	-0.083	-0.005	0.198	-0.912*	0.002
493	-0.006	-0.255	-0.055	0.677*	-0.551	0.030
532	0.072	-0.187	-0.072	0.919*	-0.164	0.021
558	-0.069	0.075	-0.041	0.526	-0.633*	-0.107
573	0.083	-0.083	-0.079	0.913*	-0.036	0.013

```
Table 2b. Factor loadings for HSIG M=53 factor analysis

GP#      F1          F2          F3          F4          F5          F6

259     0.197     -0.814*     -0.013      -0.064       0.307      -0.086
262     0.579*    -0.370       0.084      -0.062       0.449      -0.131
265     0.832*    -0.150       0.004       0.060       0.101       0.032
277     0.094     -0.913*     -0.092      -0.010       0.101      -0.024
298     0.052     -0.915*     -0.111      -0.023       0.143       0.003
299     0.049     -0.936*     -0.069      -0.034       0.267       0.006
300     0.081     -0.861*     -0.025      -0.051       0.439      -0.015
302     0.337     -0.515       0.080      -0.014       0.678*     -0.083
304     0.729*    -0.191       0.057       0.102       0.525      -0.090
1010   -0.019     -0.842*     -0.058      -0.029       0.450       0.045
321     0.142     -0.577       0.058       0.017       0.752*     -0.045
325     0.841*    -0.030       0.053       0.290       0.362       0.004
328     0.765*    -0.104      -0.046       0.248      -0.075       0.363
1032   -0.065     -0.779*     -0.070      -0.052       0.519       0.028
1035    0.147     -0.473       0.068       0.049       0.830*     -0.048
1037    0.303     -0.302       0.088       0.095       0.866*     -0.072
1038    0.404     -0.220       0.087       0.133       0.841*     -0.086
1039    0.521     -0.148       0.077       0.178       0.779*     -0.089
344     0.618     -0.052       0.052       0.309       0.675*     -0.048
345     0.771*     0.000       0.052       0.405       0.419       0.030
347     0.808*    -0.019      -0.004       0.380       0.016       0.288
1077    0.032     -0.404       0.058       0.097       0.867*     -0.019
1079    0.159     -0.244       0.076       0.156       0.926*     -0.049
1081    0.392     -0.098       0.066       0.251       0.846*     -0.068
363     0.003     -0.300       0.059       0.144       0.909*     -0.021
364     0.174     -0.151       0.066       0.222       0.916*     -0.054
1104    0.609*     0.024       0.042       0.503       0.548       0.015
1106    0.690*     0.024       0.042       0.561       0.308       0.127
1120   -0.090     -0.290       0.038       0.153       0.876*     -0.001
1146    0.516      0.029       0.022       0.665*      0.451       0.048
387     0.555      0.051       0.036       0.684*      0.352       0.102
389     0.606*     0.042       0.007       0.553      -0.002       0.434
394     0.200     -0.048      -0.047       0.114      -0.079       0.767*
1188    0.388      0.046       0.055       0.817*      0.317       0.089
1190    0.443      0.059       0.063       0.792*      0.176       0.215
412     0.366      0.014      -0.001       0.325      -0.113       0.775*
428     0.241      0.025       0.123       0.886*      0.237       0.086
429     0.280      0.054       0.128       0.885*      0.144       0.190
1270    0.159     -0.004       0.225       0.870*      0.161       0.097
1272    0.149      0.032       0.229       0.912*      0.101       0.179
1312    0.101     -0.028       0.337       0.854*      0.066       0.157
470     0.074      0.018       0.311       0.822*     -0.006       0.329
476     0.000      0.016       0.042       0.142      -0.050       0.902*
490     0.053      0.013       0.502       0.786*      0.025       0.191
493     0.048      0.055       0.308       0.530      -0.053       0.626*
510     0.058      0.002       0.661*      0.654       0.047       0.130
530     0.060      0.016       0.785*      0.485       0.077       0.083
532     0.012      0.032       0.684*      0.576       0.000       0.244
558    -0.084      0.071       0.222       0.187      -0.030       0.681*
571     0.020      0.026       0.879*      0.260       0.115       0.012
573    -0.005      0.060       0.857*      0.360       0.016       0.147
613     0.007      0.063       0.915*      0.113       0.086       0.023
655     0.031      0.095       0.826*      0.013       0.097       0.033
```
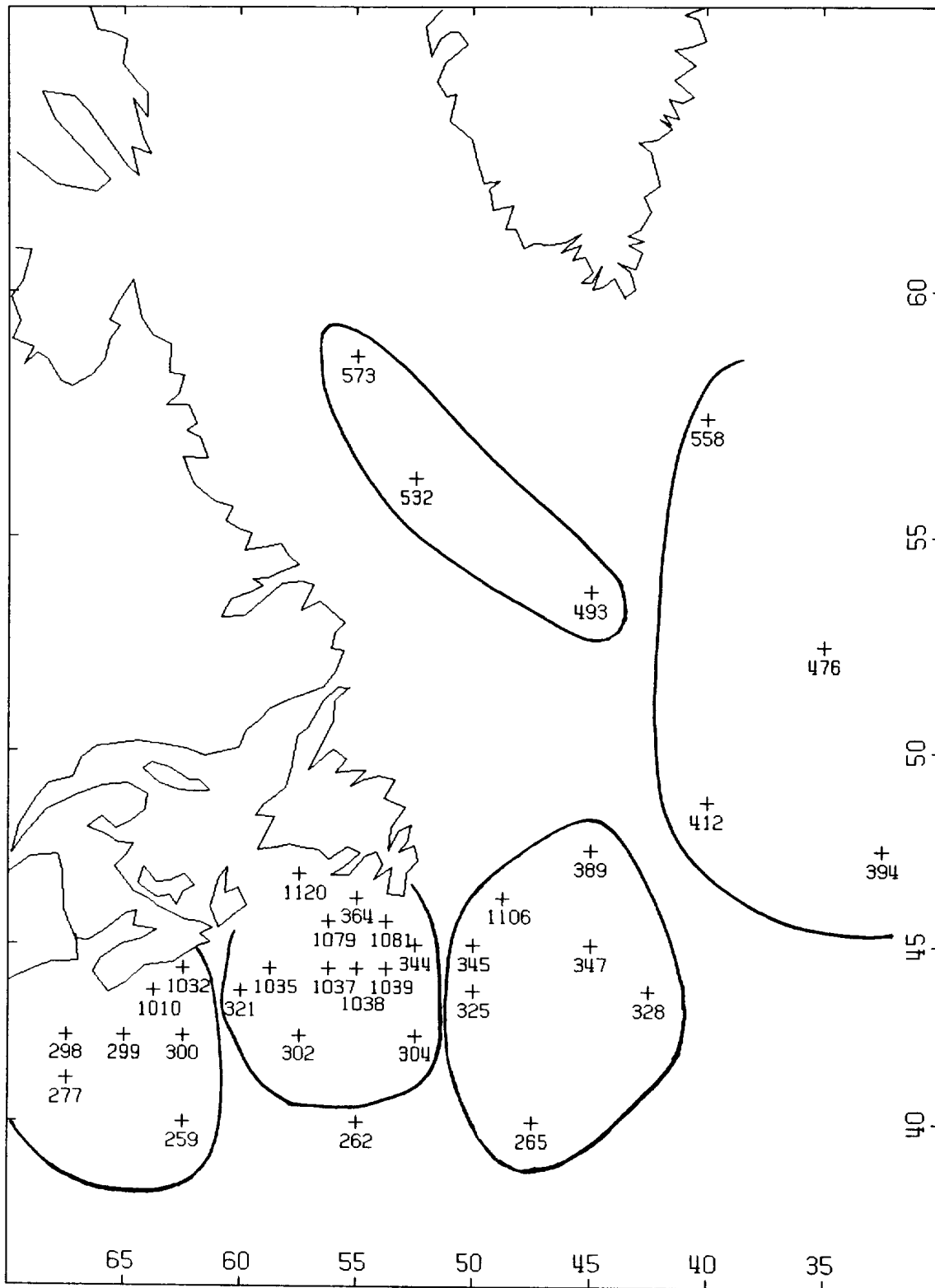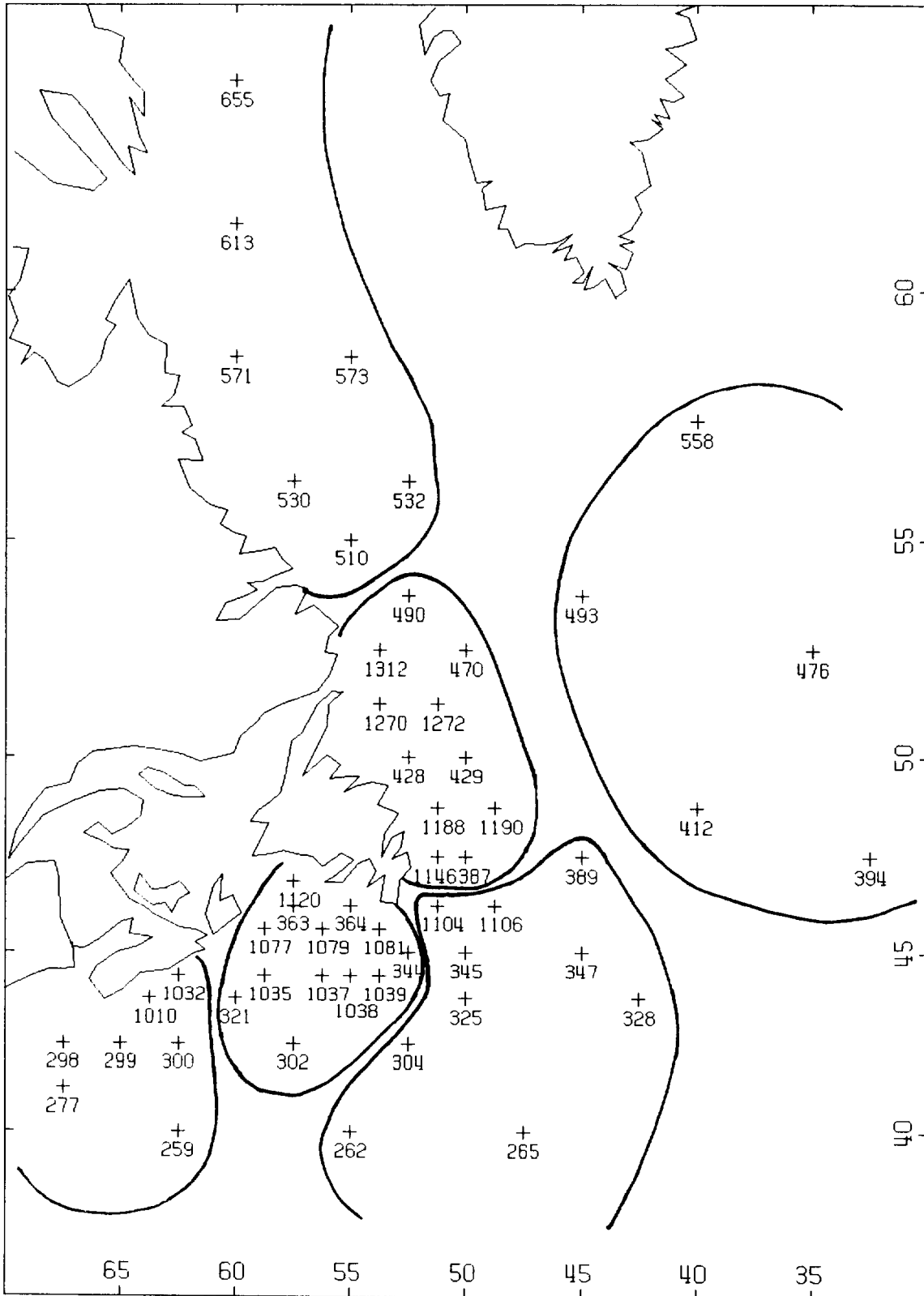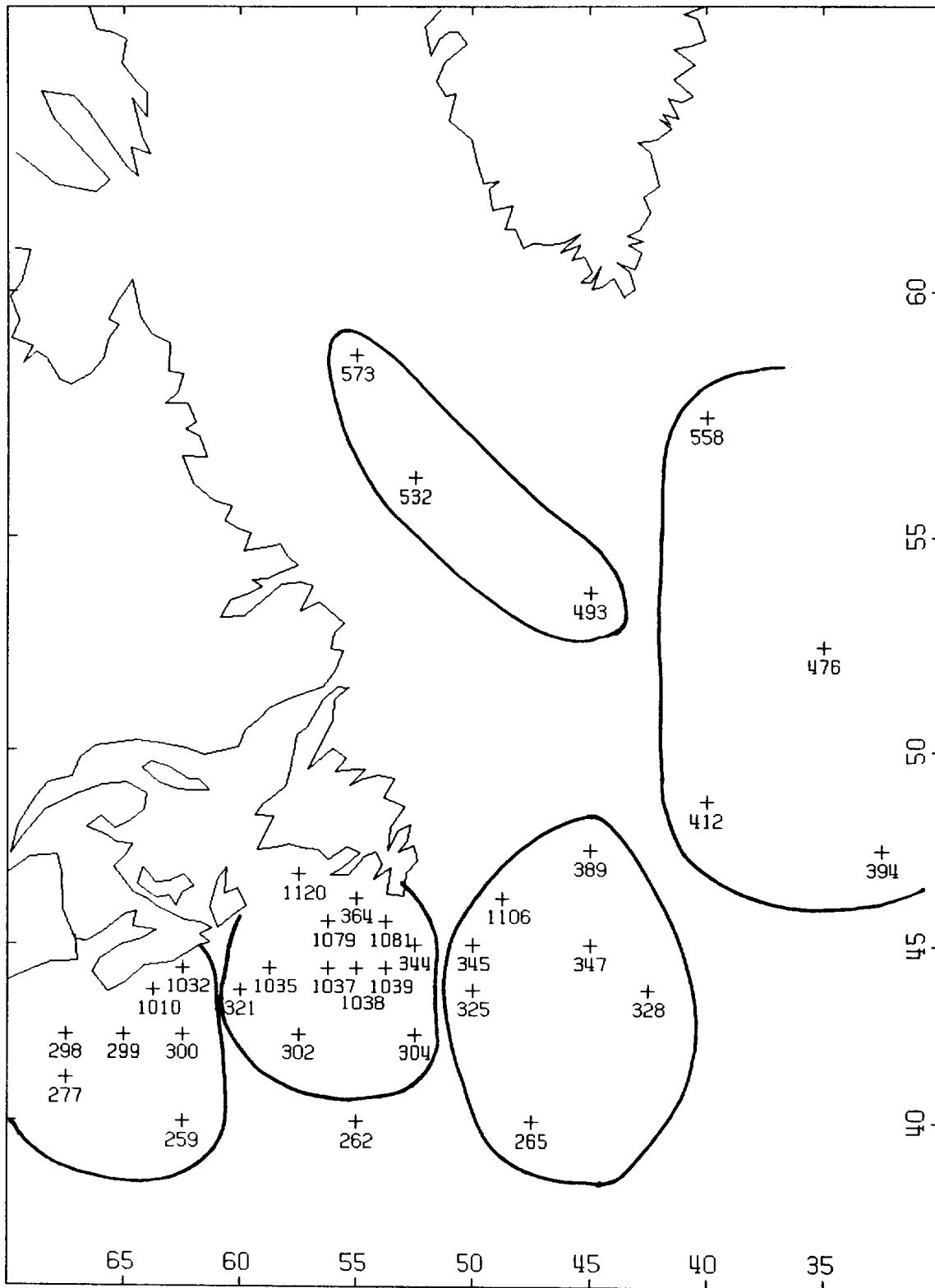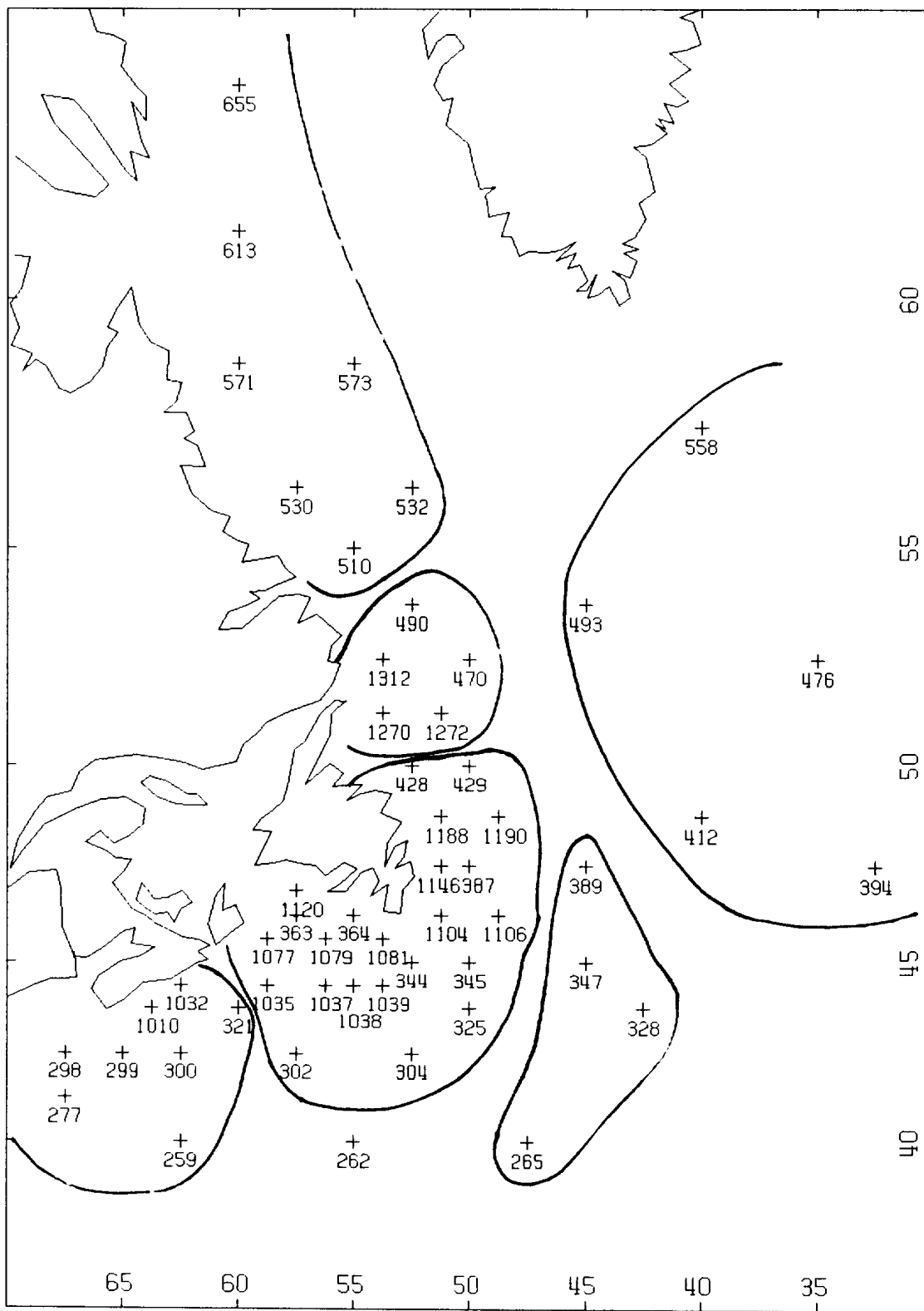
Fig. 11a Regions obtained with the M=34 HSIG factor analysis.

Fig. 11b Regions obtained with the M=53 HSIG factor analysis.

Table 3a. Factor loadings for wind vector M=34 factor
analysis.

GP#	F1	F2	F3	F4	F5	F6
259	0.331	0.660*	0.153	0.038	0.184	0.316
262	0.487	0.115	0.299	0.068	0.307	0.599*
265	0.313	0.114	0.184	0.124	0.577*	0.342
277	0.145	0.809*	0.102	0.094	0.280	0.043
298	0.144	0.815*	0.102	0.098	0.311	0.108
299	0.265	0.880*	0.059	0.075	0.293	0.029
300	0.432	0.809*	0.050	0.046	0.236	0.085
302	0.745*	0.380	0.157	0.047	0.115	0.330
304	0.657*	0.155	0.249	0.063	0.456	0.392
1010	0.389	0.814*	0.031	0.064	0.279	0.092
321	0.711*	0.597	0.061	0.038	0.100	0.059
325	0.564	0.225	0.212	0.026	0.654*	0.160
328	0.324	0.164	0.110	0.140	0.741*	0.074
1032	0.513	0.725*	0.007	0.050	0.228	0.121
1035	0.824*	0.452	0.075	0.039	0.013	0.073
1037	0.912*	0.249	0.102	0.041	0.172	0.058
1038	0.905*	0.191	0.115	0.040	0.271	0.058
1039	0.860*	0.176	0.139	0.036	0.371	0.068
344	0.796*	0.195	0.139	0.009	0.470	0.048
345	0.593	0.237	0.166	0.027	0.667*	0.067
347	0.302	0.221	0.073	0.080	0.846*	0.106
1079	0.911*	0.220	0.077	0.028	0.178	0.188
1081	0.861*	0.176	0.093	0.019	0.369	0.146
364	0.875*	0.174	0.067	0.034	0.269	0.246
1106	0.468	0.230	0.101	0.099	0.725*	0.215
1120	0.804*	0.239	0.055	0.051	0.106	0.327
389	0.234	0.197	0.206	0.149	0.778*	0.256
394	0.171	0.113	0.682*	0.249	0.179	0.118
412	0.205	0.085	0.603*	0.128	0.568	0.124
476	0.086	0.057	0.842*	0.124	0.116	0.252
493	0.100	0.065	0.450	0.650*	0.254	0.100
532	0.028	0.069	0.049	0.922*	0.067	0.035
558	0.062	0.098	0.545*	0.489	0.058	0.283
573	0.028	0.073	0.098	0.871*	0.011	0.007

Table 3b. Factor loadings for wind vector M=53 analysis.

| GP#	F1	F2	F3	F4	F5	F6
259	0.160	0.718*	0.027	0.148	0.193	0.138
262	0.379	0.377	0.031	0.335	0.229	0.293
265	0.225	0.313	0.113	0.169	0.525*	0.156
277	0.041	0.834*	0.099	0.068	0.208	0.116
298	0.038	0.853*	0.106	0.087	0.168	0.117
299	0.118	0.926*	0.076	0.039	0.188	0.109
300	0.253	0.880*	0.038	0.114	0.184	0.088
302	0.570*	0.505	0.037	0.361	0.080	0.167
304	0.629*	0.285	0.051	0.392	0.327	0.218
1010	0.230	0.886*	0.070	0.058	0.192	0.087
321	0.534	0.686*	0.048	0.228	0.248	0.085
325	0.654*	0.271	0.087	0.300	0.488	0.104
328	0.219	0.235	0.187	0.098	0.615*	0.313
1032	0.354	0.803*	0.063	0.097	0.247	0.078
1035	0.673*	0.533	0.059	0.265	0.284	0.085
1037	0.816*	0.308	0.071	0.344	0.170	0.105
1038	0.846*	0.234	0.078	0.361	0.066	0.109
1039	0.847*	0.200	0.083	0.360	0.060	0.113
344	0.860*	0.175	0.097	0.311	0.144	0.081
345	0.753*	0.238	0.110	0.267	0.413	0.067
347	0.404	0.266	0.175	0.153	0.653*	0.281
1077	0.721*	0.430	0.076	0.174	0.366	0.075
1079	0.868*	0.223	0.089	0.232	0.272	0.069
1081	0.908*	0.127	0.105	0.259	0.051	0.065
363	0.805*	0.275	0.084	0.141	0.369	0.077
364	0.904*	0.116	0.100	0.182	0.213	0.063
1104	0.856*	0.170	0.125	0.257	0.188	0.063
1106	0.723*	0.244	0.139	0.263	0.424	0.142
1120	0.788*	0.235	0.085	0.079	0.385	0.085
1146	0.836*	0.156	0.117	0.342	0.116	0.086
387	0.788*	0.194	0.127	0.367	0.225	0.121
389	0.464	0.218	0.174	0.301	0.518*	0.401
394	0.163	0.059	0.198	0.099	0.050	0.638*
1188	0.775*	0.151	0.082	0.497	0.096	0.088
1190	0.685*	0.195	0.106	0.498	0.250	0.188
412	0.175	0.106	0.179	0.168	0.361	0.711*
428	0.699*	0.117	0.090	0.606	0.160	0.083
429	0.668*	0.169	0.072	0.627	0.122	0.120
1270	0.594	0.088	0.220	0.640*	0.219	0.111
1272	0.603	0.142	0.173	0.701*	0.122	0.055
1312	0.503	0.095	0.362	0.657*	0.203	0.098
470	0.493	0.142	0.273	0.703*	0.075	0.129
476	0.055	0.090	0.083	0.153	0.175	0.806*
490	0.427	0.111	0.486	0.646*	0.129	0.009
493	0.256	0.119	0.214	0.463	0.144	0.519*
510	0.325	0.093	0.675*	0.479	0.155	0.081
530	0.212	0.084	0.778*	0.280	0.165	0.120
532	0.273	0.103	0.714*	0.438	0.047	0.103
558	0.043	0.092	0.234	0.187	0.212	0.553*
571	0.089	0.073	0.853*	0.049	0.123	0.086
573	0.153	0.075	0.877*	0.186	0.044	0.096
613	0.048	0.044	0.867*	0.028	0.056	0.043
655	0.029	0.020	0.751*	0.075	0.019	0.046

48

Fig. 12a Regions obtained with the M=34 wind vector factor analysis.

Fig. 12b Regions obtained with the M=53 wind vector factor analysis.

Table 4a. Factor loadings for total wave vector M=34 factor
          analysis.

GP#	F1	F2	F3	F4	F5	F6
259	0.245	0.788*	0.090	0.050	0.048	0.324
262	0.557	0.381	0.152	0.083	0.059	0.621*
265	0.544*	0.219	0.171	0.128	0.489	0.349
277	0.129	0.872*	0.036	0.088	0.217	0.073
298	0.159	0.870*	0.035	0.094	0.238	0.159
299	0.231	0.924*	0.071	0.072	0.173	0.084
300	0.330	0.888*	0.093	0.051	0.113	0.059
302	0.622*	0.590	0.096	0.059	0.195	0.323
304	0.791*	0.289	0.182	0.055	0.045	0.400
1010	0.376	0.840*	0.095	0.067	0.153	0.174
321	0.558	0.721*	0.127	0.021	0.220	0.043
325	0.876*	0.198	0.144	0.031	0.188	0.188
328	0.601	0.170	0.230	0.118	0.603*	0.083
1032	0.436	0.780*	0.113	0.054	0.184	0.197
1035	0.643*	0.616	0.151	0.029	0.298	0.099
1037	0.784*	0.463	0.136	0.053	0.308	0.062
1038	0.841*	0.387	0.130	0.056	0.266	0.067
1039	0.882*	0.321	0.129	0.050	0.195	0.078
344	0.908*	0.258	0.133	0.023	0.124	0.023
345	0.915*	0.192	0.122	0.029	0.126	0.033
347	0.757*	0.162	0.170	0.051	0.530	0.048
1079	0.779*	0.402	0.161	0.036	0.346	0.193
1081	0.880*	0.290	0.129	0.030	0.243	0.116
364	0.822*	0.317	0.149	0.028	0.308	0.220
1106	0.881*	0.162	0.076	0.090	0.205	0.110
1120	0.679*	0.374	0.194	0.018	0.327	0.317
389	0.736*	0.145	0.244	0.138	0.434	0.189
394	0.250	0.066	0.675*	0.251	0.313	0.121
412	0.502	0.128	0.589*	0.138	0.433	0.167
476	0.213	0.088	0.856*	0.195	0.202	0.099
493	0.350	0.112	0.498	0.608*	0.114	0.116
532	0.117	0.080	0.080	0.926*	0.018	0.042
558	0.091	0.085	0.638*	0.526	0.129	0.093
573	0.005	0.071	0.028	0.904*	0.001	0.014

Table 4b. Factor loadings for total wave vector M=53 analysis.

GP#	F1	F2	F3	F4	F5	F6
259	0.154	0.048	0.154	0.176	0.042	0.806*
262	0.496	0.239	0.391	0.291	0.022	0.379
265	0.461	0.190	0.598*	0.129	0.136	0.182
277	0.035	0.107	0.092	0.098	0.103	0.876*
298	0.081	0.099	0.059	0.082	0.112	0.885*
299	0.144	0.055	0.043	0.074	0.081	0.945*
300	0.242	0.036	0.053	0.099	0.038	0.917*
302	0.550	0.223	0.059	0.219	0.054	0.625*
304	0.732*	0.272	0.318	0.223	0.048	0.279
1010	0.299	0.024	0.091	0.027	0.076	0.881*
321	0.516	0.120	0.169	0.139	0.043	0.753*
325	0.816*	0.243	0.333	0.105	0.082	0.198
328	0.472	0.195	0.534*	0.373	0.203	0.159
1032	0.382	0.010	0.186	0.061	0.052	0.825*
1035	0.613	0.152	0.255	0.185	0.071	0.648*
1037	0.746*	0.230	0.196	0.188	0.074	0.491
1038	0.798*	0.262	0.129	0.171	0.066	0.410
1039	0.840*	0.275	0.040	0.141	0.063	0.332
344	0.881*	0.255	0.019	0.098	0.054	0.245
345	0.878*	0.227	0.209	0.086	0.075	0.174
347	0.660*	0.172	0.474	0.315	0.161	0.167
1077	0.624*	0.114	0.369	0.224	0.098	0.555
1079	0.767*	0.177	0.330	0.204	0.083	0.427
1081	0.863*	0.234	0.175	0.116	0.070	0.293
363	0.699*	0.122	0.392	0.234	0.096	0.450
364	0.821*	0.178	0.304	0.168	0.085	0.332
1104	0.901*	0.245	0.021	0.050	0.059	0.165
1106	0.865*	0.227	0.197	0.131	0.087	0.154
1120	0.692*	0.096	0.411	0.231	0.091	0.404
1146	0.868*	0.334	0.030	0.075	0.037	0.145
387	0.863*	0.325	0.065	0.112	0.060	0.139
389	0.670*	0.259	0.311	0.415	0.147	0.166
394	0.149	0.260	0.356	0.573*	0.188	0.077
1188	0.804*	0.472	0.040	0.085	0.045	0.133
1190	0.795*	0.410	0.090	0.201	0.065	0.154
412	0.376	0.270	0.331	0.661*	0.181	0.161
428	0.679*	0.629	0.098	0.088	0.126	0.099
429	0.729*	0.565	0.015	0.162	0.063	0.148
1270	0.537	0.708*	0.125	0.099	0.255	0.085
1272	0.617	0.701*	0.049	0.122	0.167	0.115
1312	0.481	0.731*	0.077	0.081	0.340	0.093
470	0.538	0.690*	0.027	0.244	0.205	0.133
476	0.088	0.206	0.304	0.785*	0.142	0.132
490	0.431	0.715*	0.050	0.115	0.415	0.113
493	0.381	0.452	0.097	0.568*	0.171	0.153
510	0.313	0.601	0.048	0.063	0.609*	0.100
530	0.172	0.416	0.046	0.061	0.759*	0.086
532	0.294	0.538	0.044	0.190	0.631*	0.126
558	0.057	0.143	0.144	0.637*	0.208	0.138
571	0.052	0.154	0.041	0.045	0.898*	0.076
573	0.168	0.264	0.037	0.175	0.851*	0.109
613	0.024	0.031	0.029	0.058	0.908*	0.074
655	0.015	0.027	0.022	0.067	0.811*	0.057
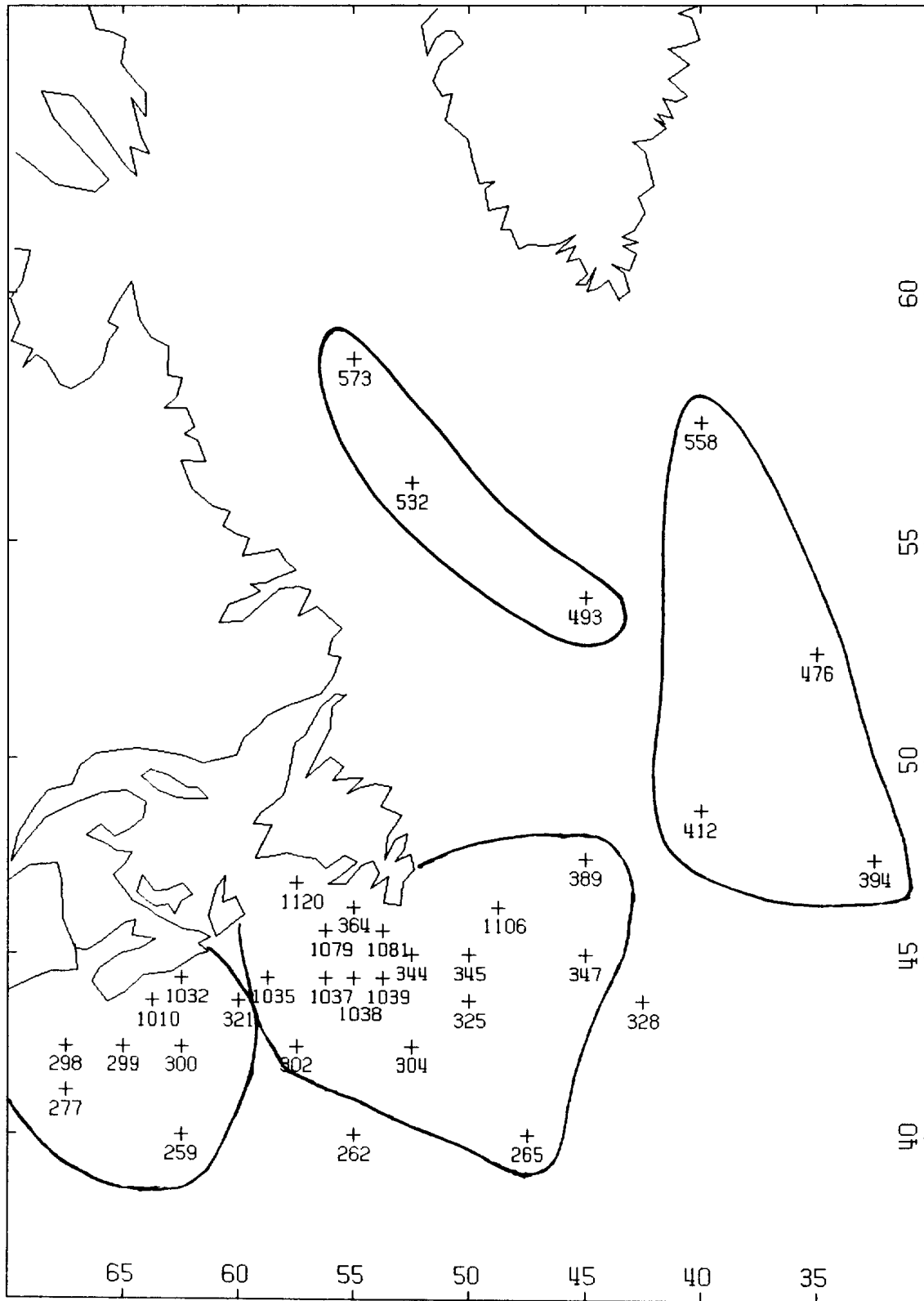
Fig. 13a Regions obtained with the M=34 total wave vector factor analysis.
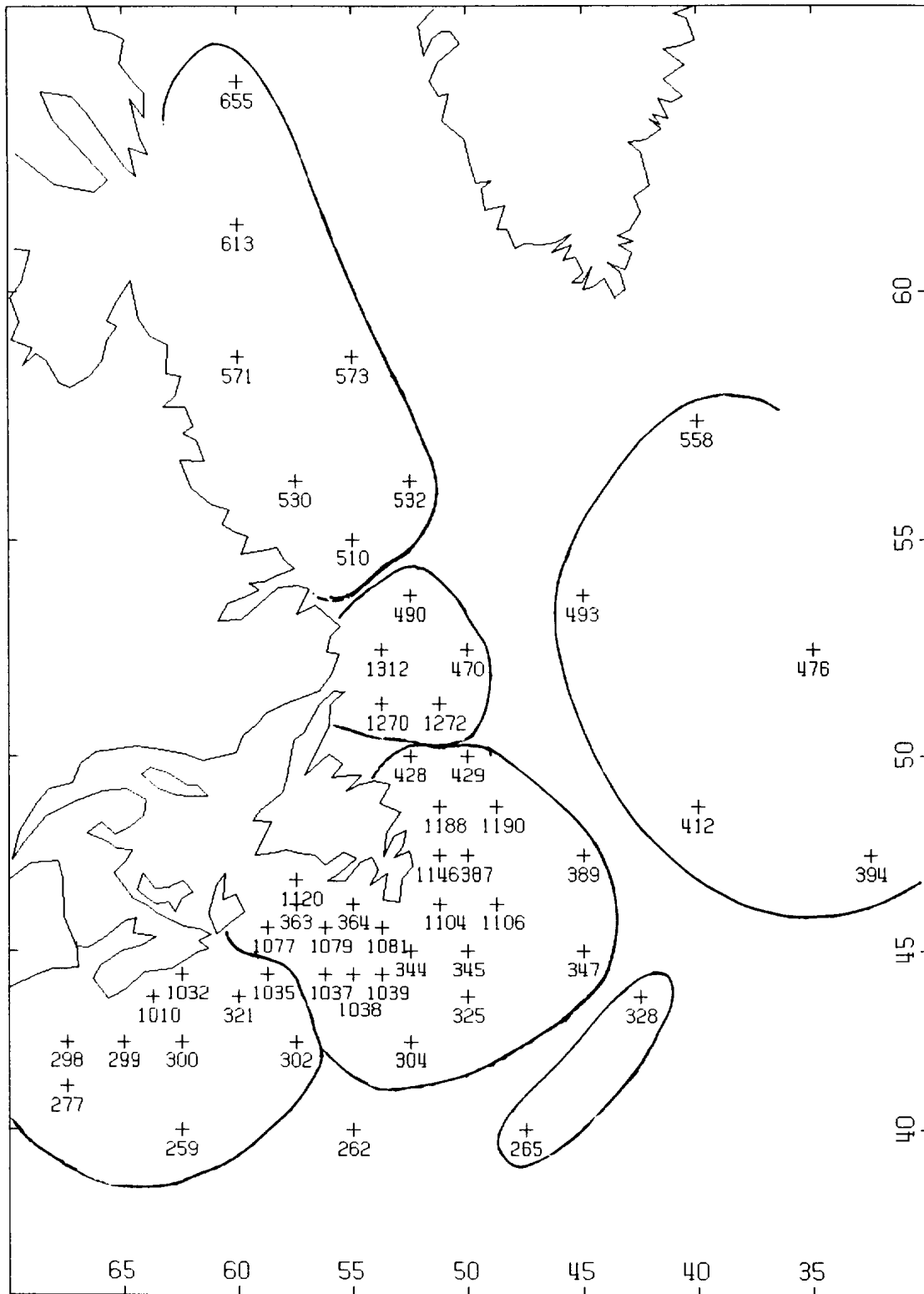
Fig. 13b Regions obtained with the M=53 total wave vector factor analysis.

4.0 MODEL STATISTICAL ANALYSIS

The statistical analysis, as described in Section 2.4 , consists of: 1) the performance of a probability analysis on the occurrence histograms of a set of eight derived model fit parameters plus wind speed for each of 11 significant waveheight classes; 2) the development of a family of probability spectra; and 3) establishing the regression relations between the fit parameters of the probability spectra and HSIG. The analysis was performed on data subsets based on temporal periods (annual, winter and fall) and spatial groupings (individual grid points, the six regions found in Section 3 and all locations combined). A grouping according to spectral type was also examined for the combined data set. Because of the volume of data and results, only selected examples will be discussed. All the results can be made available on computer storage medium.

4.1 Probability Analysis

A two-sided bounded Gaussian distribution was fit, by means of a non-linear iterative procedure, to the probability occurrence histograms of the eight derived fit parameters $\tan^{-1}(\delta_1/\delta_2)$, ω_{m1}, ω_{m2} $\ln(\lambda_1)$, $\ln(\lambda_2)$, $\ln(p_1)$, $\ln(p_2)$ and $(\theta_{m1} - \theta_{m2})$, plus wind speed. This resulted in a set of nine probability functions per wave height class per temporal-spatial data subset. A minimum of 20 data values was required to perform the fit. An example of the probability fits for three wave height classes of the combined annual analysis, is shown in Fig. 14 . The fits for the other height classes are provided in Fig A3 of the Appendix. The occurrence histograms are generally well represented by the bounded Gaussian. All data set probability fits were examined visually and the behaviour was consistent. The 0-1m and 1-2m wave height classes showed the most variability with some occurrences of bimodal distributions. At larger waveheights, the smaller sample sizes could result in an non-standard, "gap-tooth," and, at times, flat distribution. The bounded Gaussian could still be fit mathematically to these cases however the interpretation of the probability limits, choice of mode, etc. becomes questionable.

The probability functions allow for the determination of any confidence limit parameter value, the mode of the distribution (i.e. peak location of the probability curve) and the median (i.e. 50% probability value). For well-behaved distributions, the mode and median would have similar values. The bounded Gaussian allows for truncated distributions, in which case this is no longer true. In our application, the 95% confidence limits (i.e. probability values of .025 and .975) of the parameter distribution was used in the generation of the family of spectra.
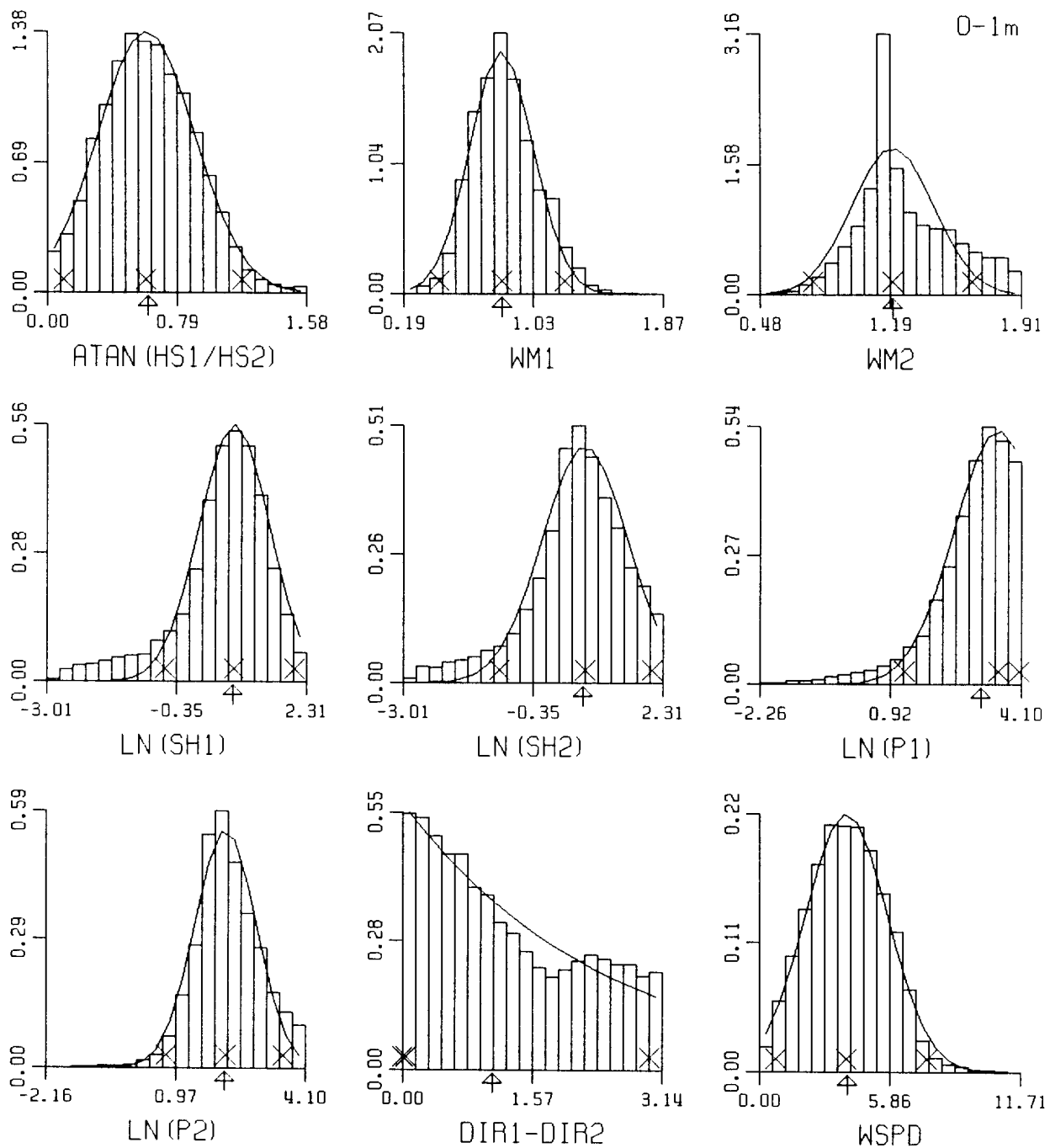
Fig. 14a Probability occurrence histograms and bounded Gaussian
fit for the eight derived model parameters and wind speed.
Combined annual data, HSIG: 0-1m.
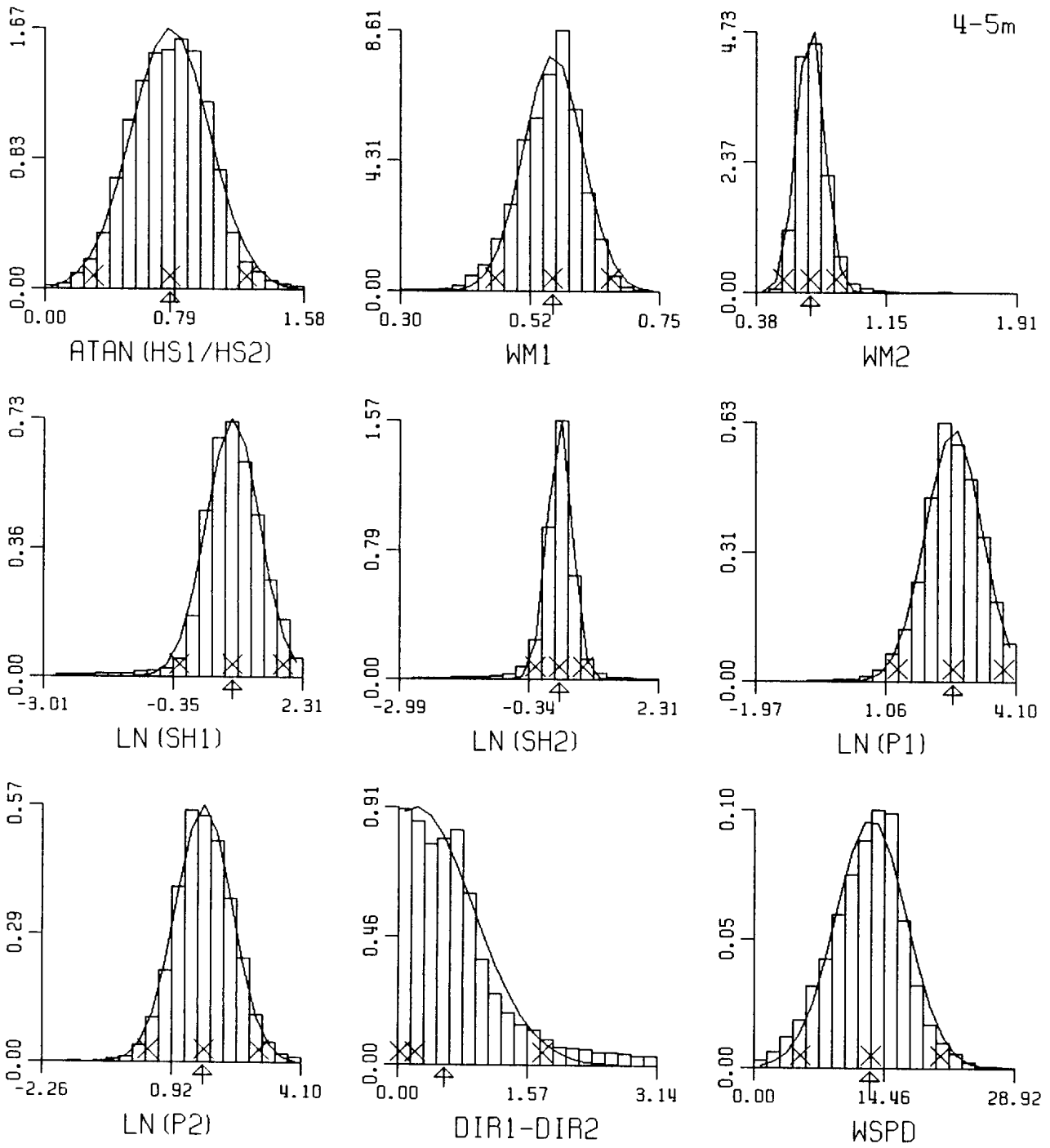
Fig. 14b HSIG: 4-5m.
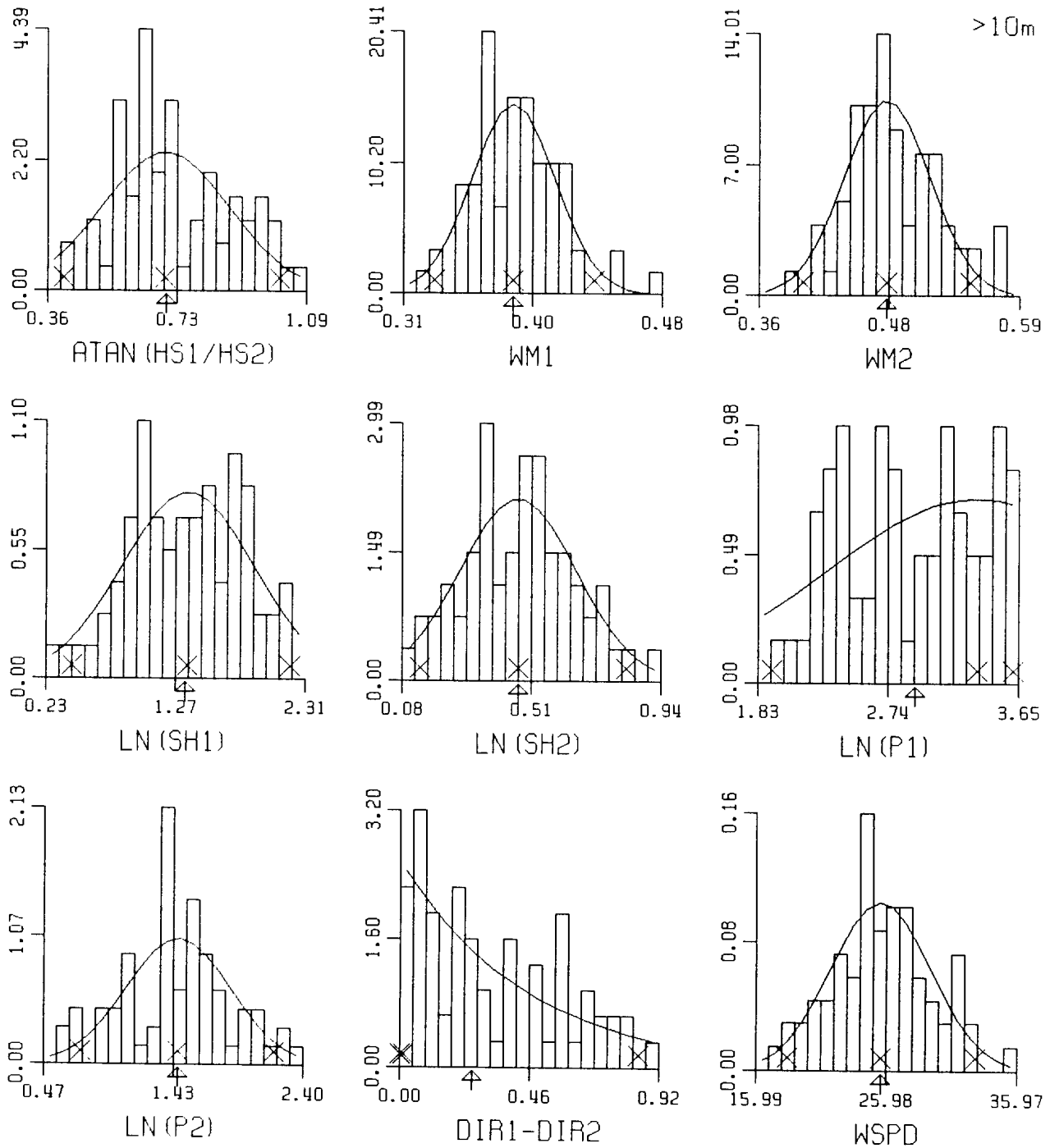
Fig. 14c HSIG: >10m.

4.2 Family of Spectra

The family of spectra, for each temporal-spatial data subset, was developed according to the method described in Section 2.4 . Fig. 15 shows the averaged modal spectra, per height class, per region for the annual data sets. The spectrum associated with the spatially combined data set is provided in the lower right corner. Contour intervals for all directional spectra were set to .01, .025, .05, 0.1, 0.25, 0.5, 1., 2., 4., 6., 8., 10., 15., 20. and 30. m^2/(rps – rad). The spectra are positioned on the map to correspond to the six established regions. There appears to be very little regional variability in the average modal spectra except at the lowest significant wave heights where the presence or absence of swell may most influence the fit. Fig A4   in the Appendix contains the lower and upper 95% confidence limit spectra when ω_{m1} is the target parameter. The target parameter, wave height bin and confident(, level are noted in the upper right of the maps. The regional variation is again slight. At higher significant wave heights, differences may reflect a reduced number of averagings in the establishment of the family of spectra (i.e. a small sample size for high energy records) which, due to the nonlinear nature of the 10-parameter model, could lead to unexpected spectral shapes. The absence of large regional differences, when examining data grouped into HSIG classes, is not surprising. Sea spectra are self-similar so that a spectrum associated with a significant wave height of, for example, 6m will look the same wherever it is measured (due to the strong relationships between height, peak period and self-similar spectral shape). As we, are working with model wave spectra, there may be some concern that the spectral shape is a product of the ODGP numerics, however self-similarity in field spectra has been well established. The regional variation will be reflected in the temporal information, lost in the statistical averaging, and in the number of occurrences of, for example, 6m spectra at one location compared to another. The latter can be. seen in Fig 16 (a   to c) which maps the percent occurrence of HSIG by region and overall.

4.3 Predictive Relations

In order to generalize the family of spectra to any significant waveheight, the predictive relationships between the statistically determined fit parameter (y) and mean height per wave height class (x) was established using eq. 15 and a non-linear fit procedure. The results for the annual combined data set averaged modal spectra are shown in Fig. 17 . Similar results for the 95% confidence limits are shown in Fig. 18 . It can be seen that the functional fit has various degrees of success, best when the chosen parameter is the target

parameter and poorest for the shape, and spread fit parameters due to their intrinsically noisy nature. In the case of (θ_{m1} – θ_{m2}), the mode is generally close to zero and the regression could lead to small negative values at low heights (<5 degrees) as it attempts to fit the, small variability of the parameter with HSIG (i.e. the range in (θ_{m1} – θ_{m2}) is less than the direction resolution of the ODGP spectrum of 15 degrees). Caution would be required when applying the prediction to HSIG values lying beyond the range in the observations.
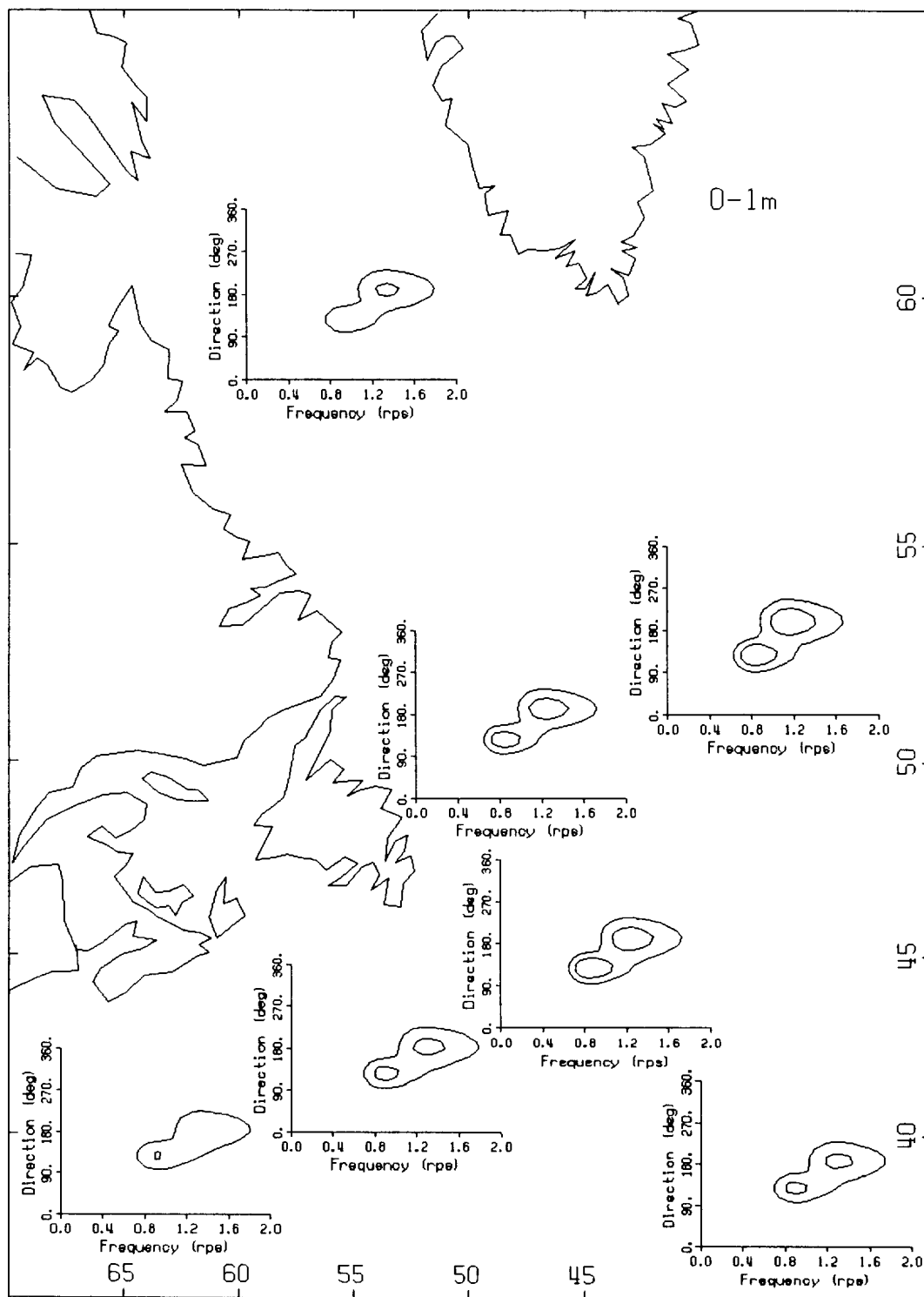
Fig. 15 Contoured average regional and overall (lower right) modal probability spectra - annual data. Contours set to: .01, .025, .05, 0.1, 0.25, 0.5, 1., 2., 4., 6., 8., 10., 15., 20., and 30. m*m/rps-rad. a) 0-1m HSIG bin.
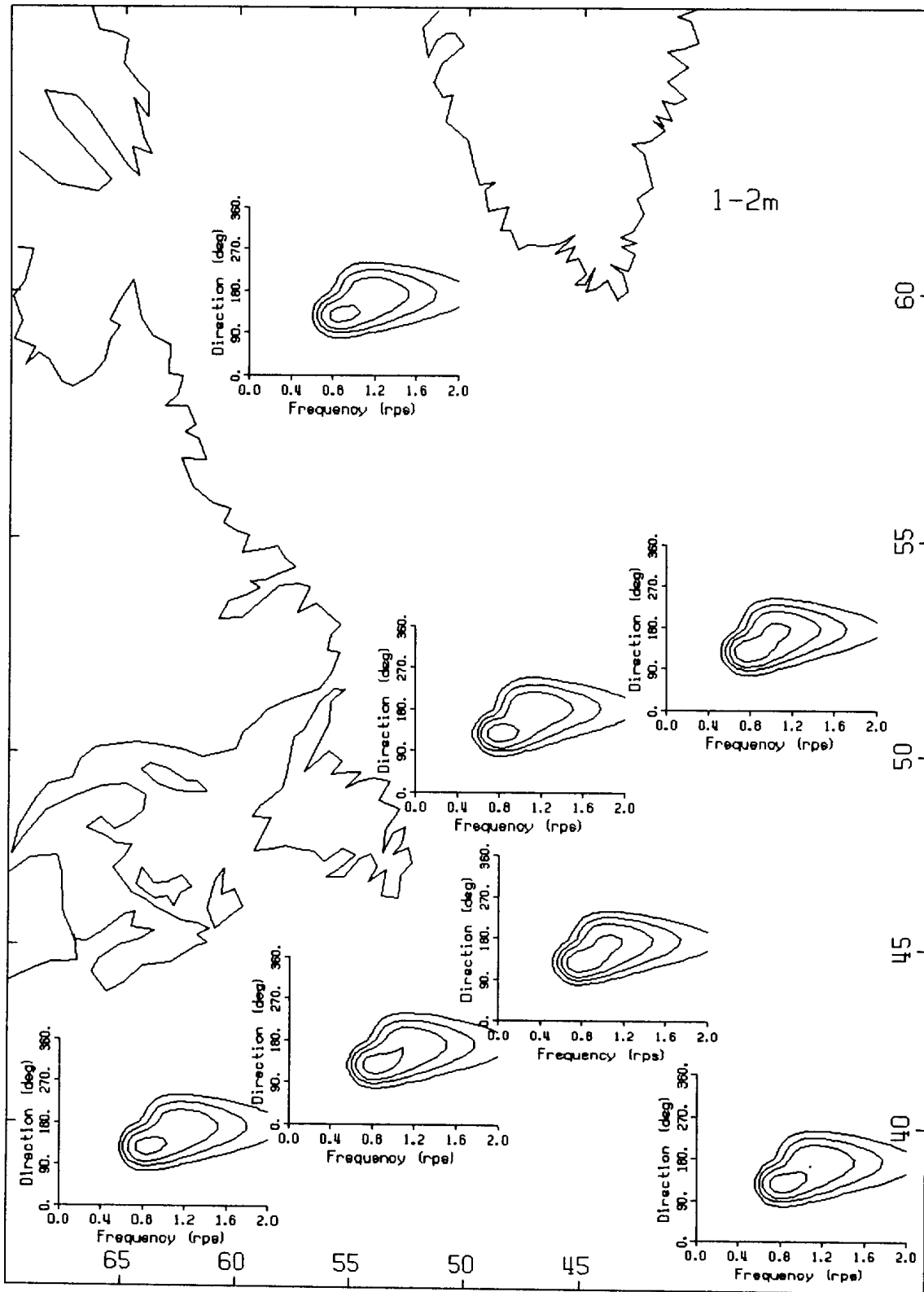
Fig. 15 (continued) b) 1-2m HSIG bin.
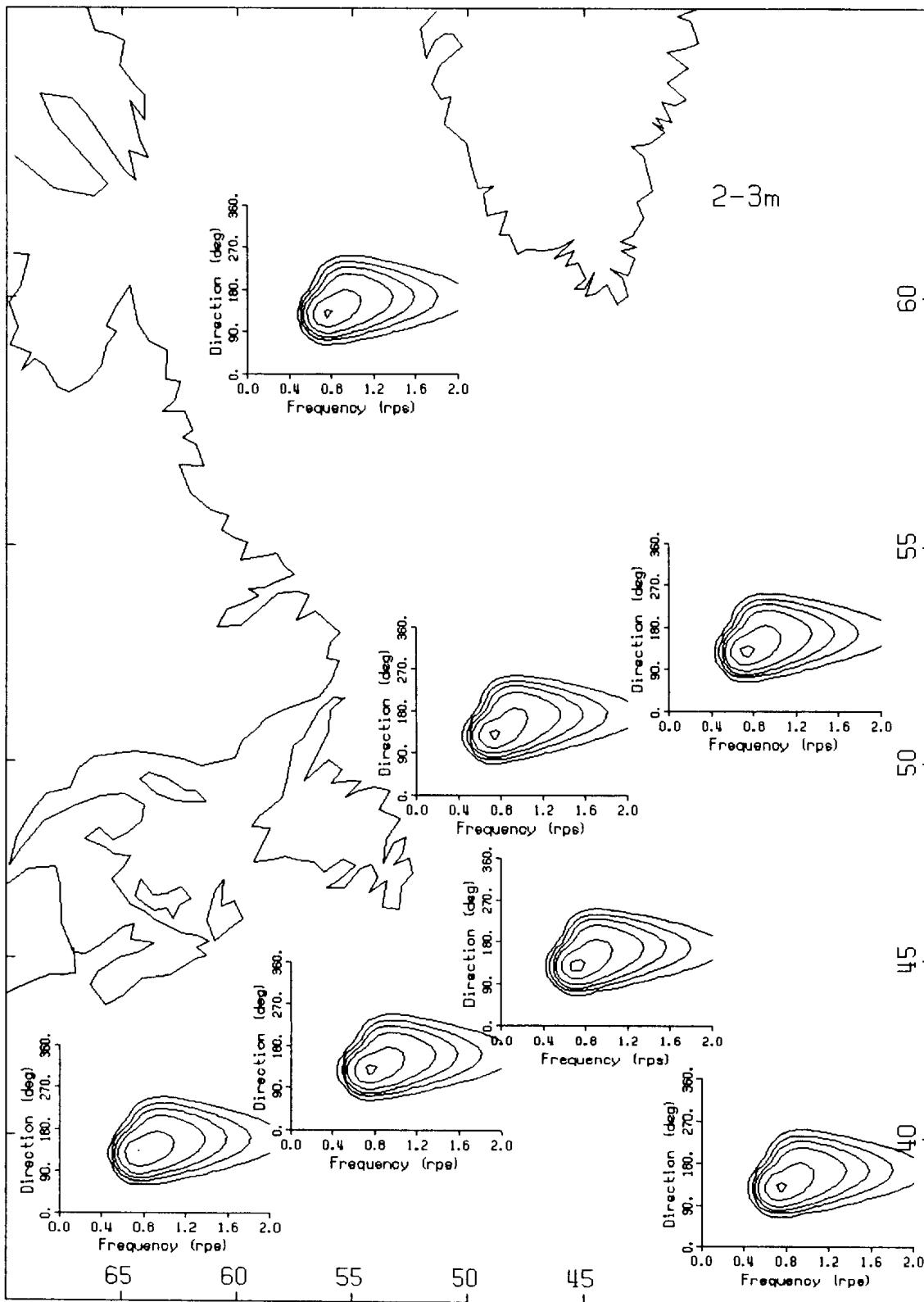
Fig. 15 (continued) c) 2-3m HSIG bin.
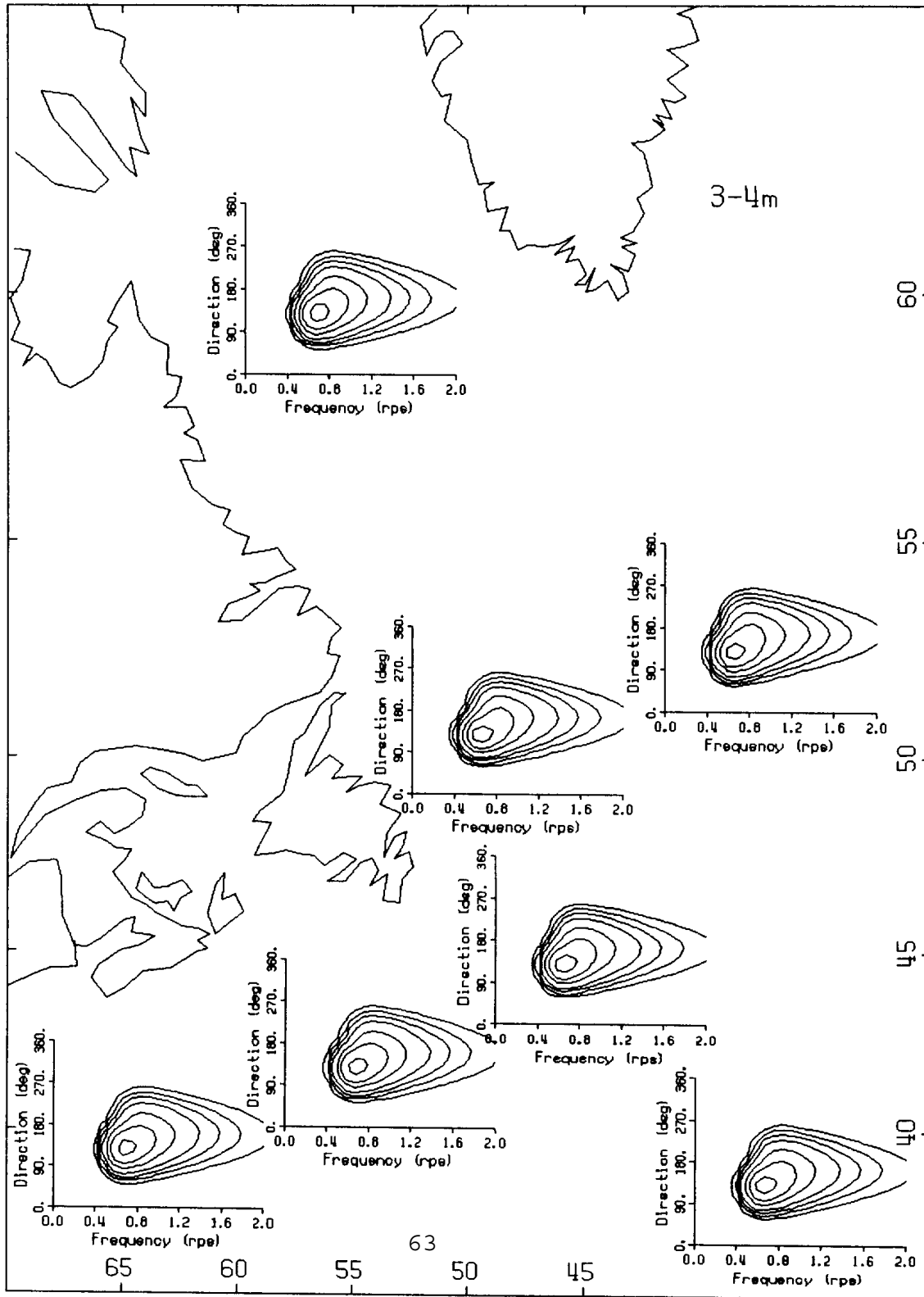
Fig. 15 (continued) d) 3-4m HSIG bin.
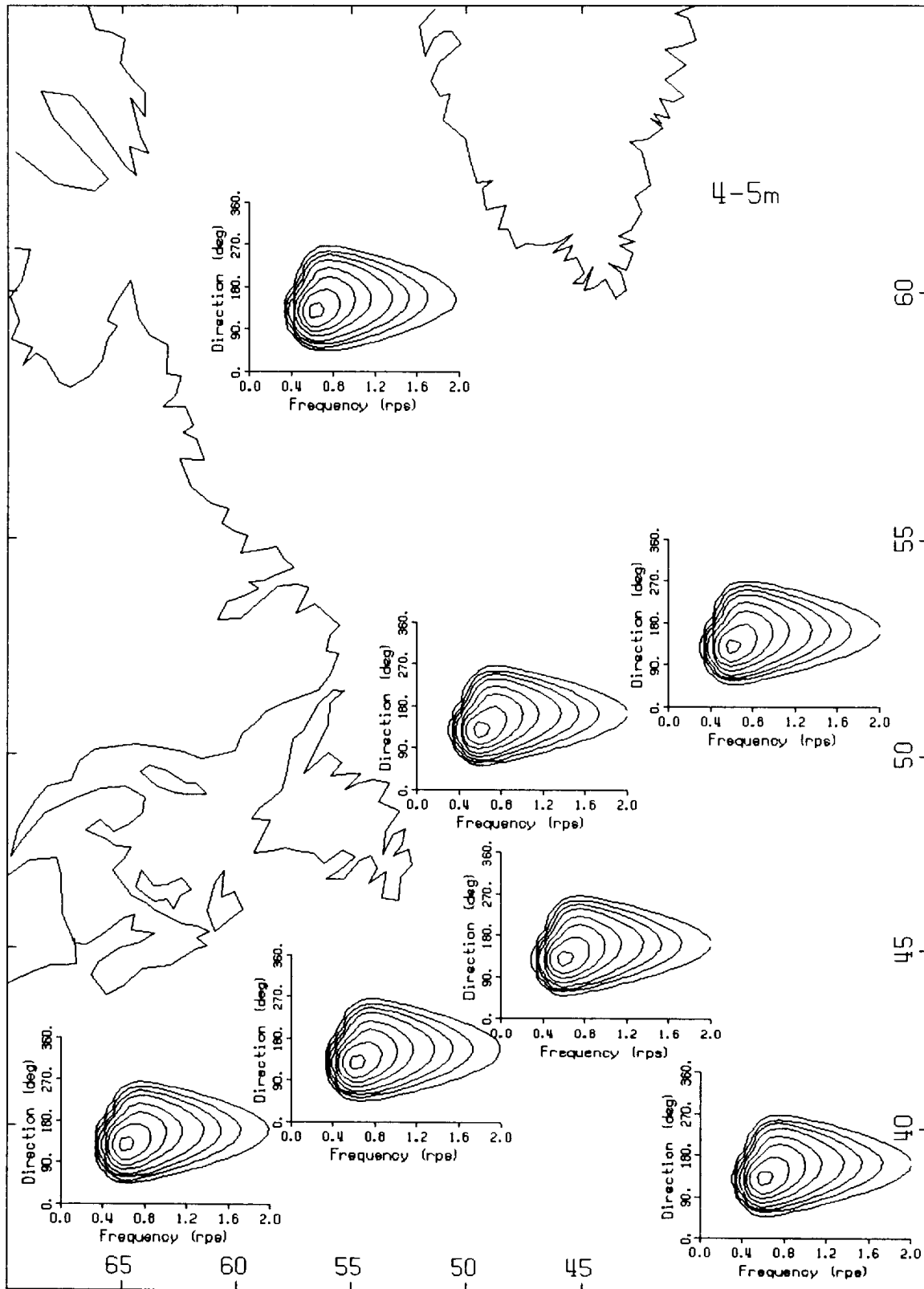
Fig. 15 (continued) e) 4-5m HSIG bin.
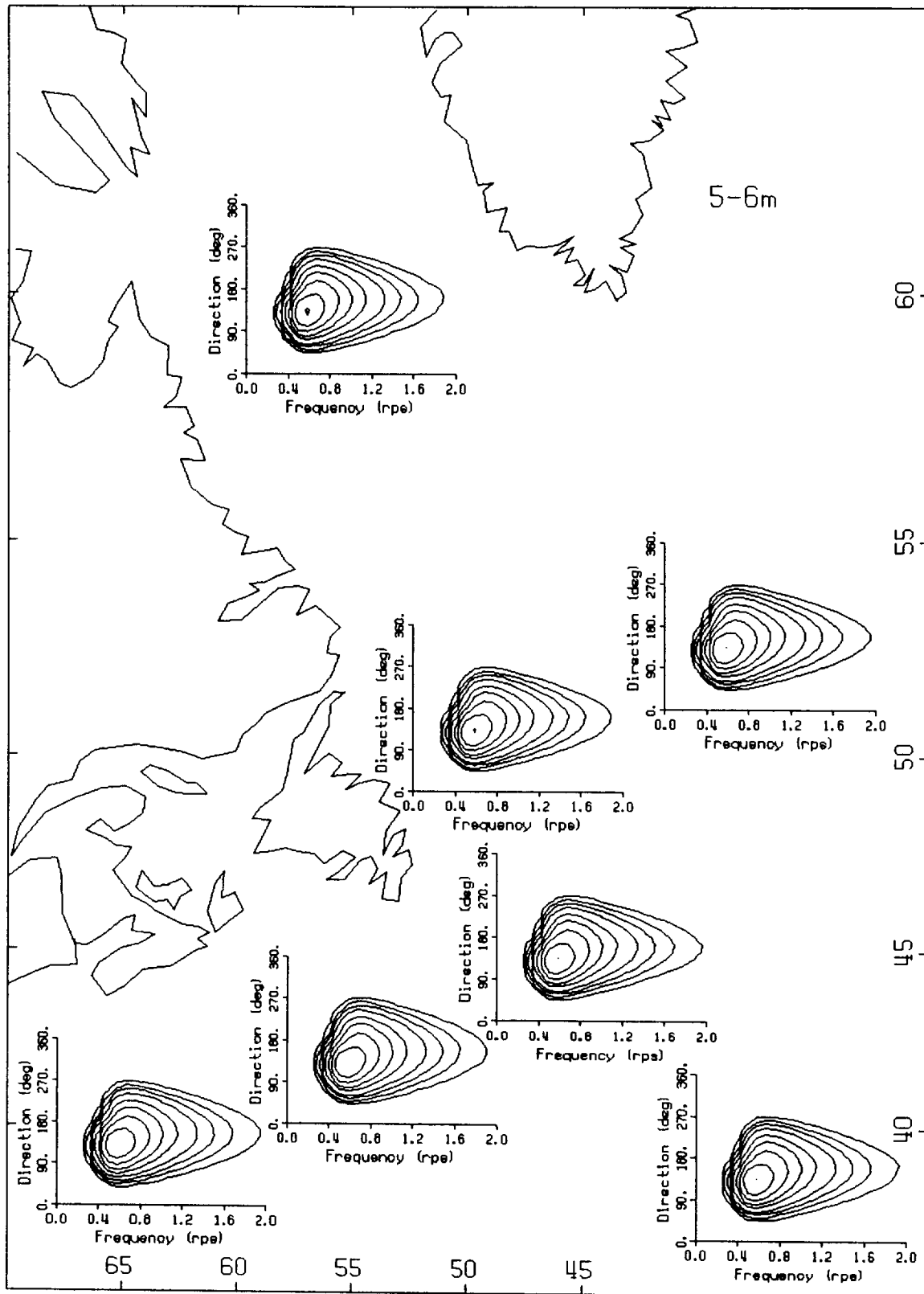
Fig. 15 (continued) f) 5-6m HSIG bin.
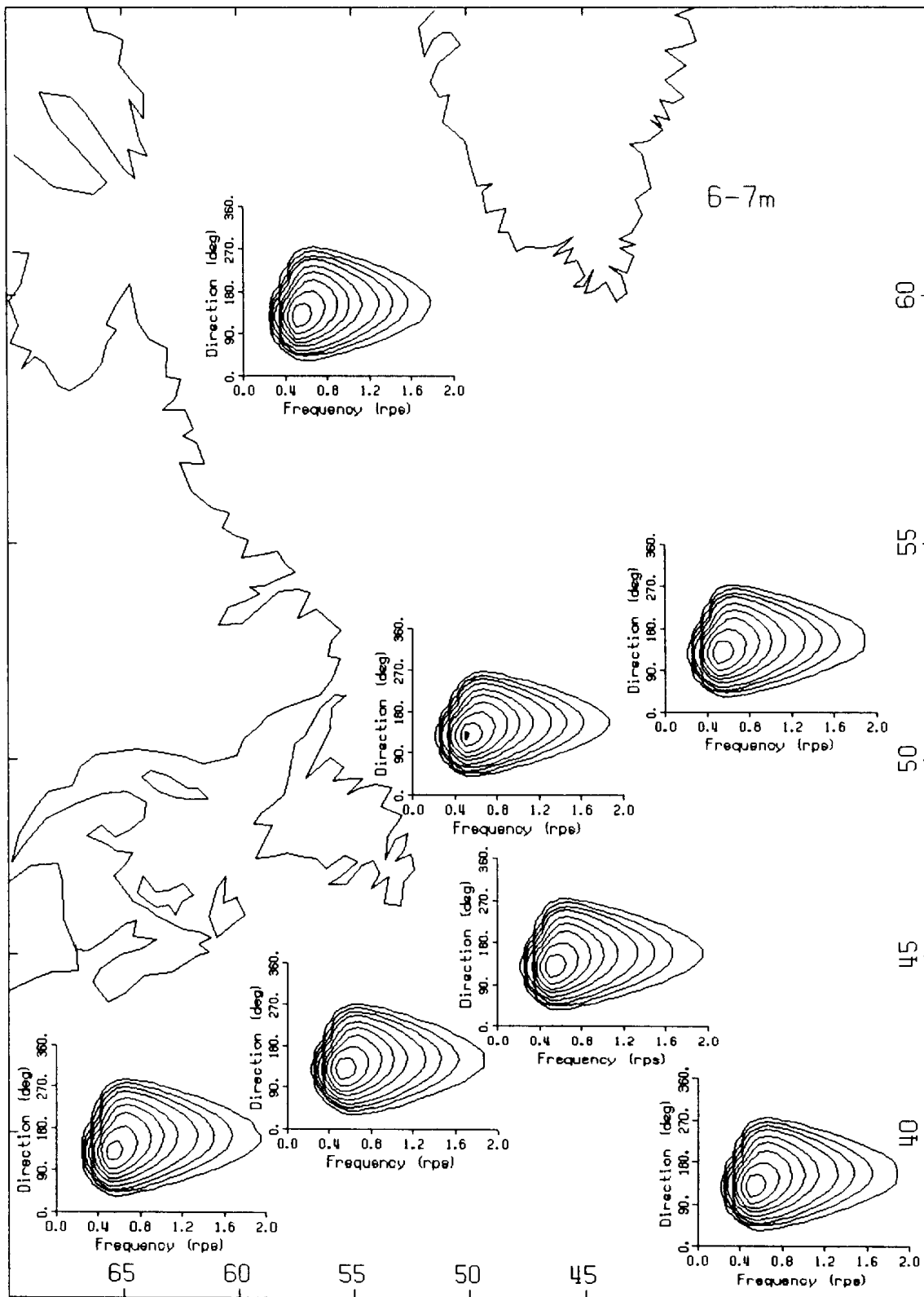
Fig. 15 (continued) g) 6-7m HSIG bin.
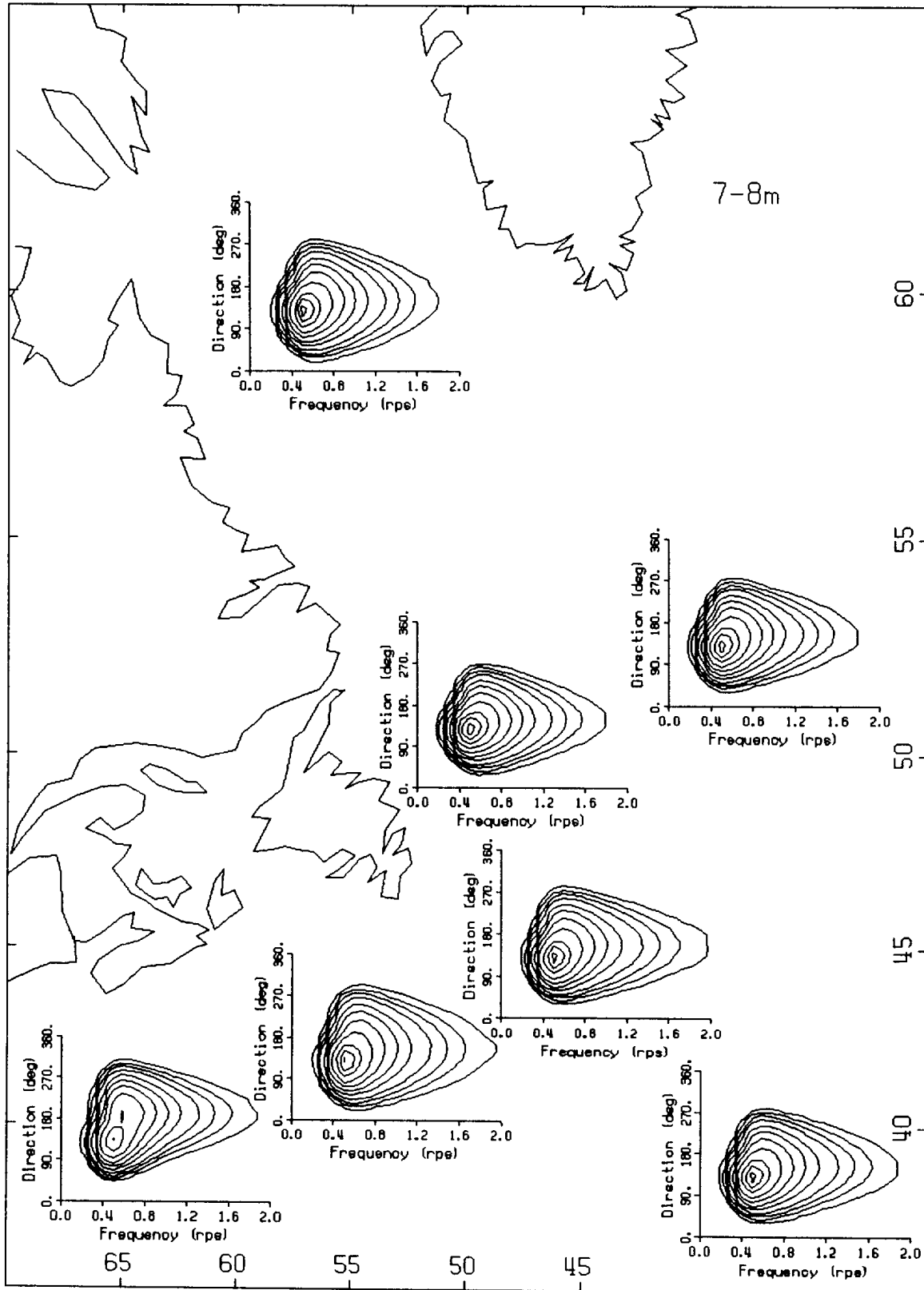
Fig. 15 (continued) h) 7-8m HSIG bin.
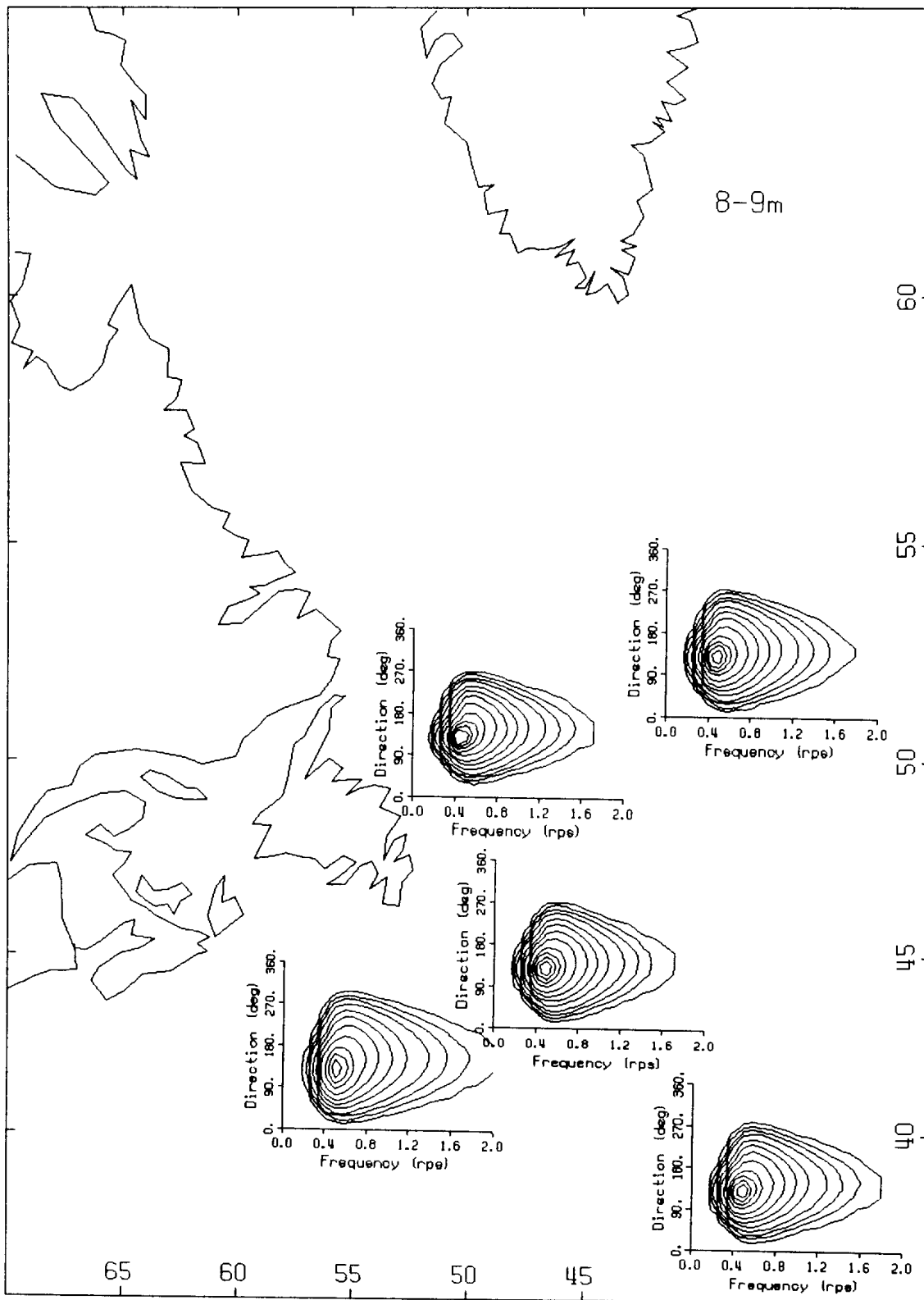
Fig. 15 (continued) i) 8-9m HSIG bin.
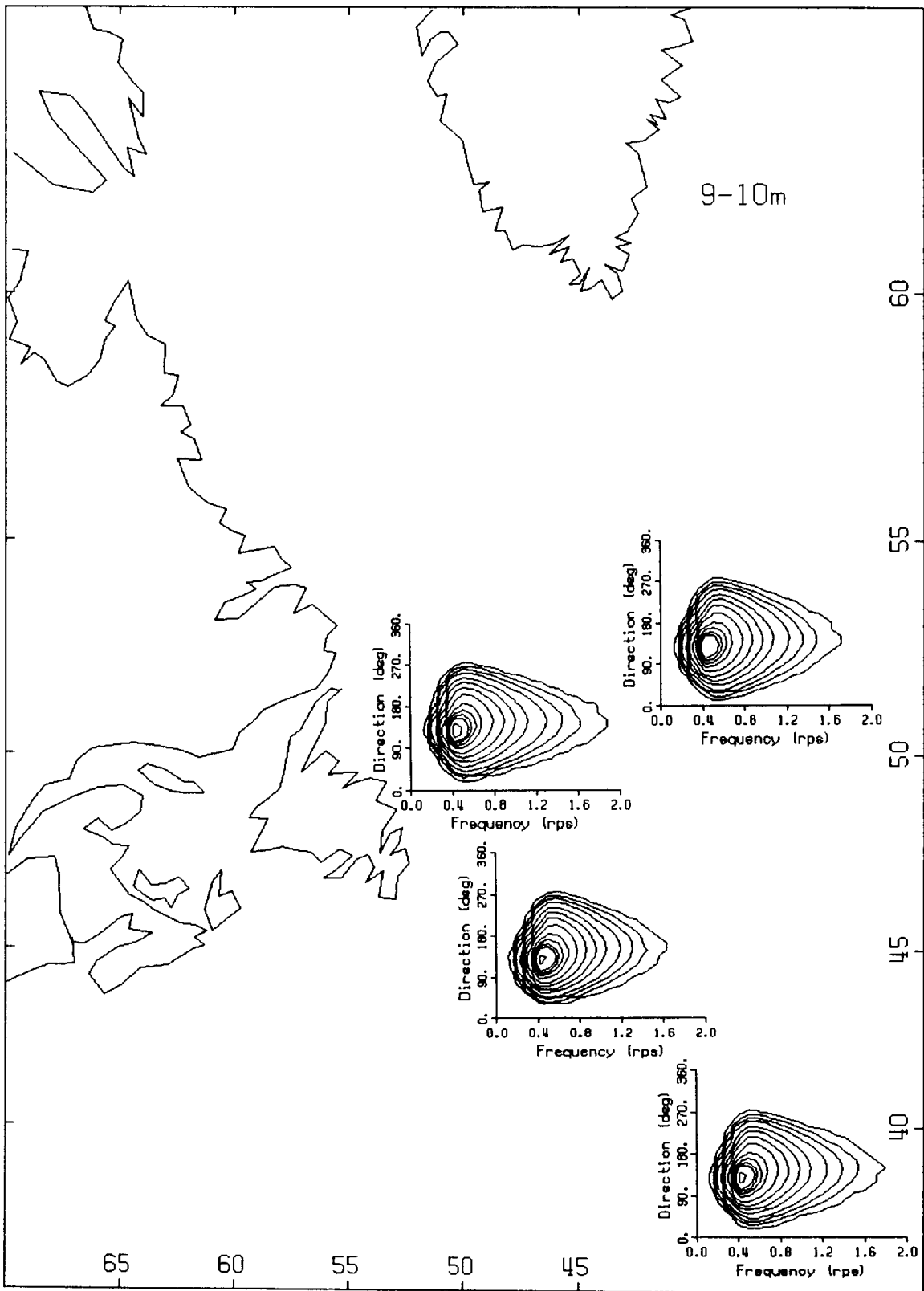
Fig. 15 (continued) j) 9-10m HSIG bin.

>10m
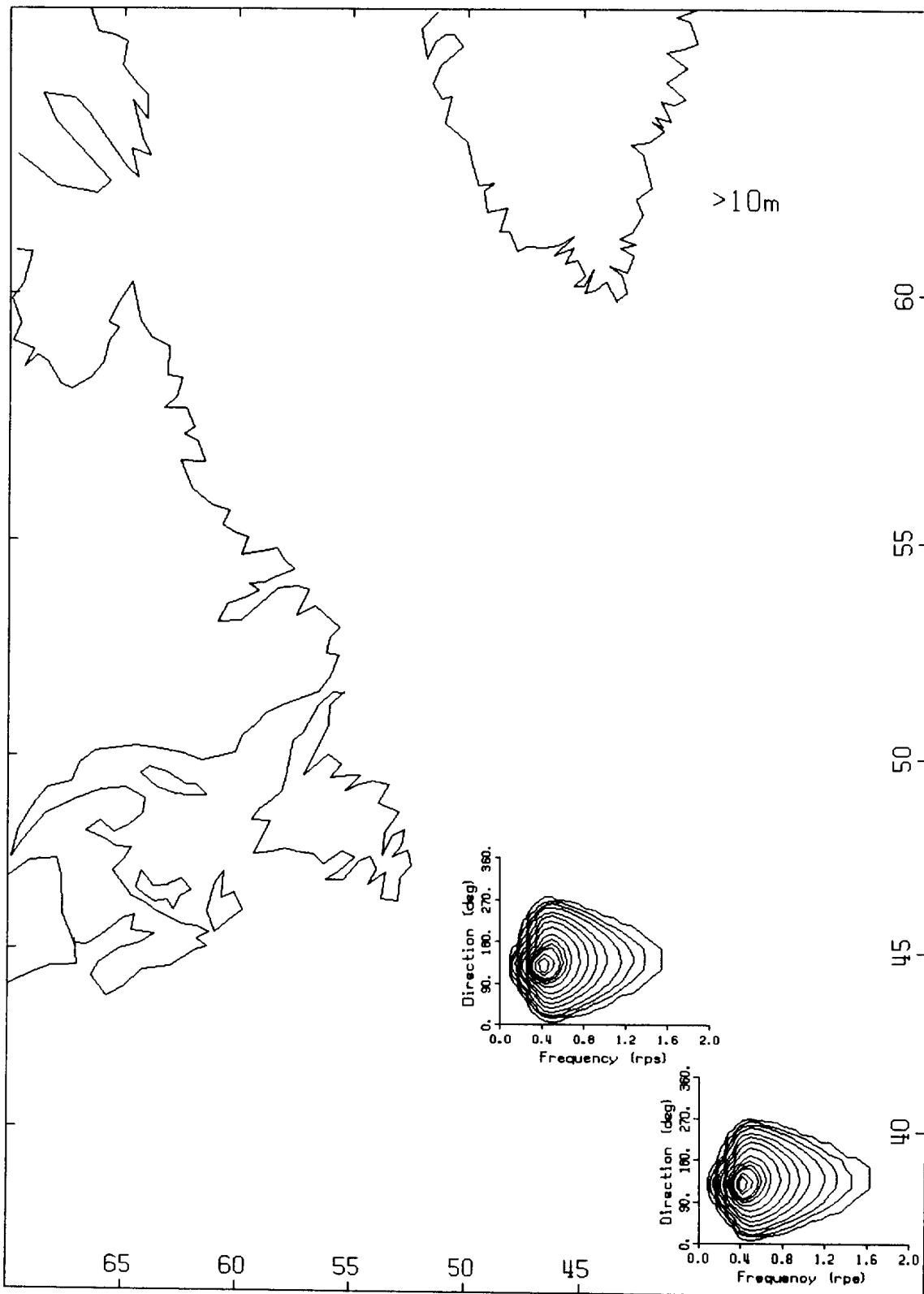
Fig. 15 (continued) k) >10m HSIG bin.
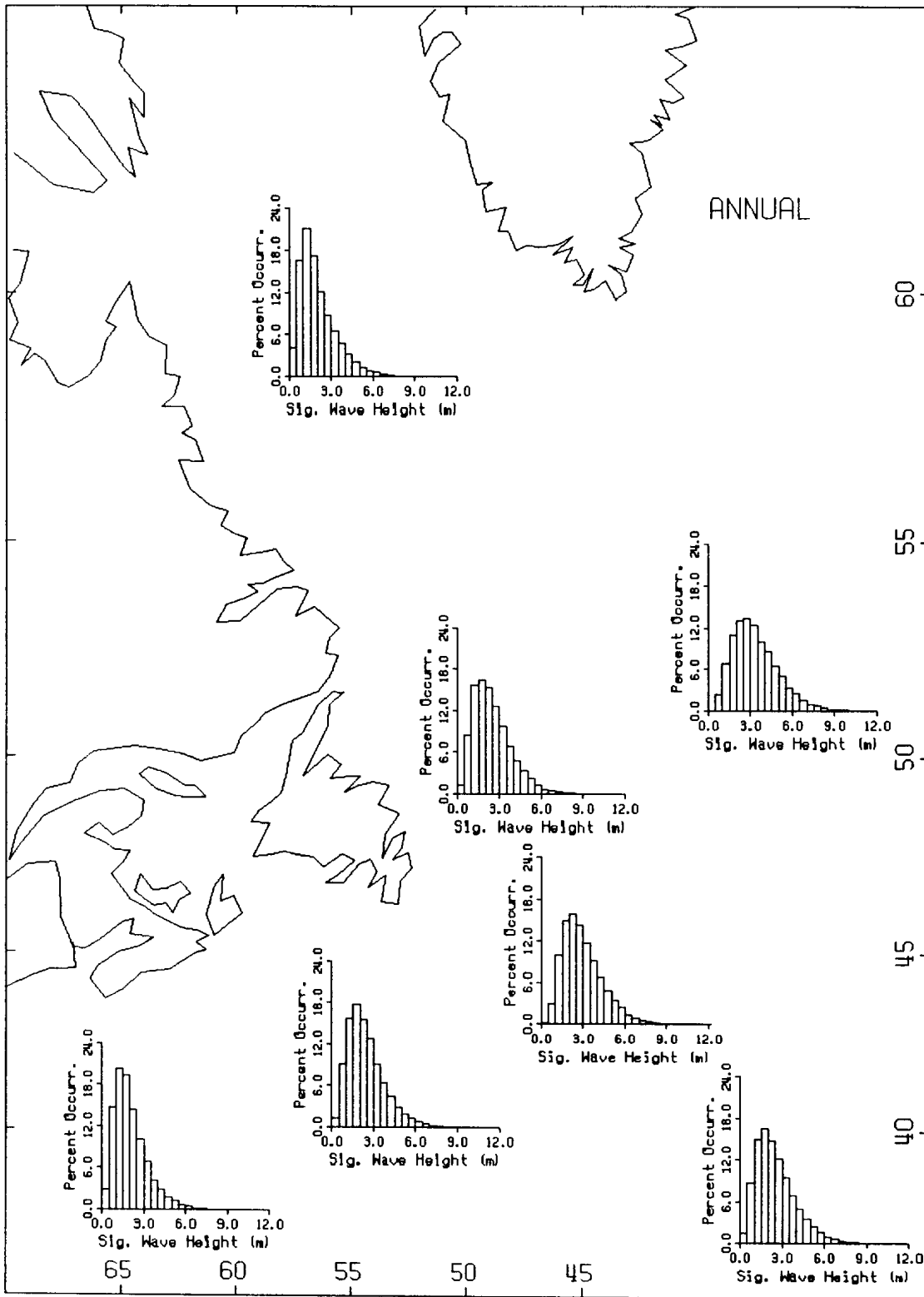
ANNUAL

Fig. 16a Annual percent occurrence histograms of HSIG for
each region and overall.
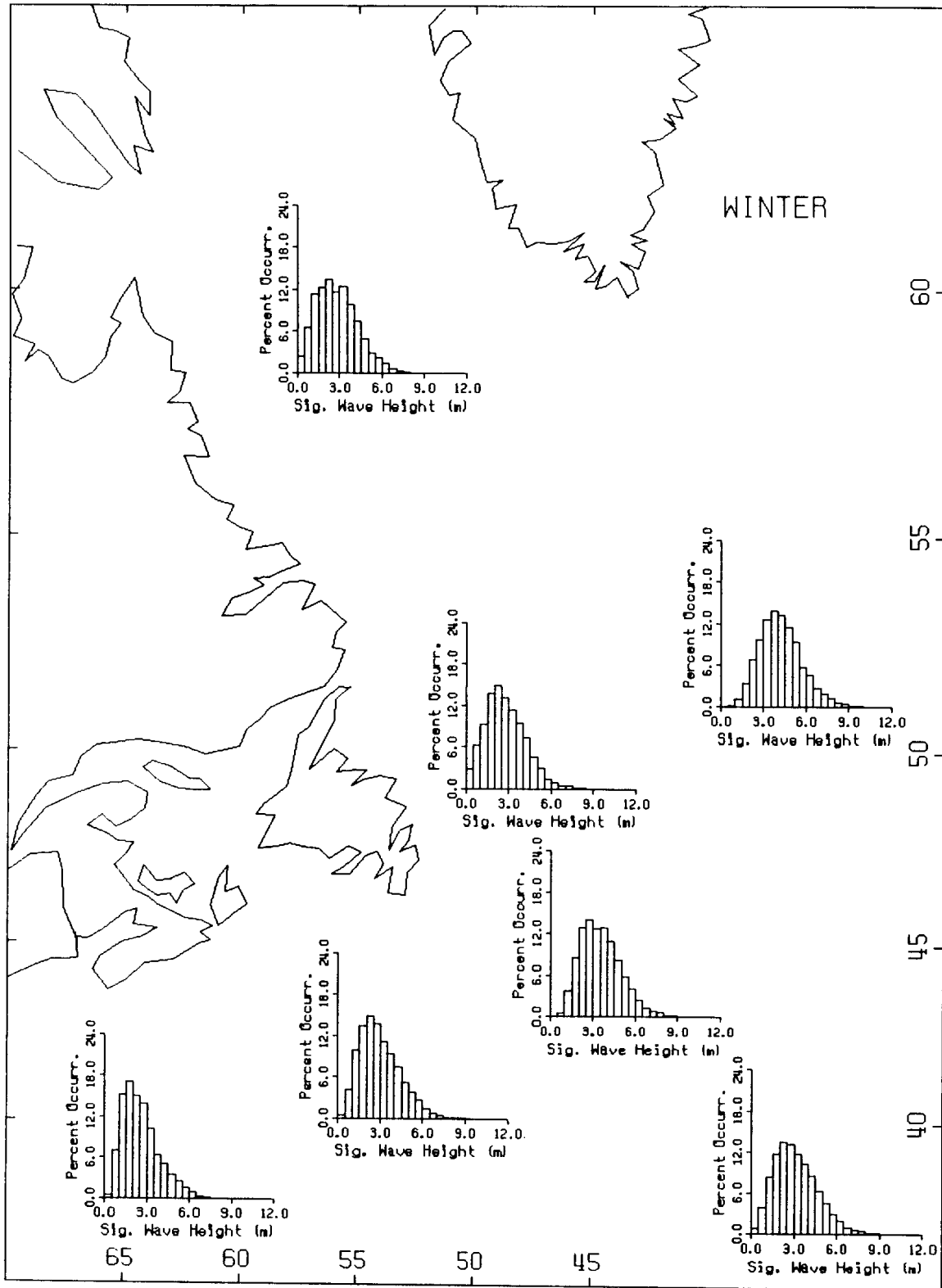
WINTER

Fig. 16b Winter percent occurrence histograms of HSIG for
each region and overall.
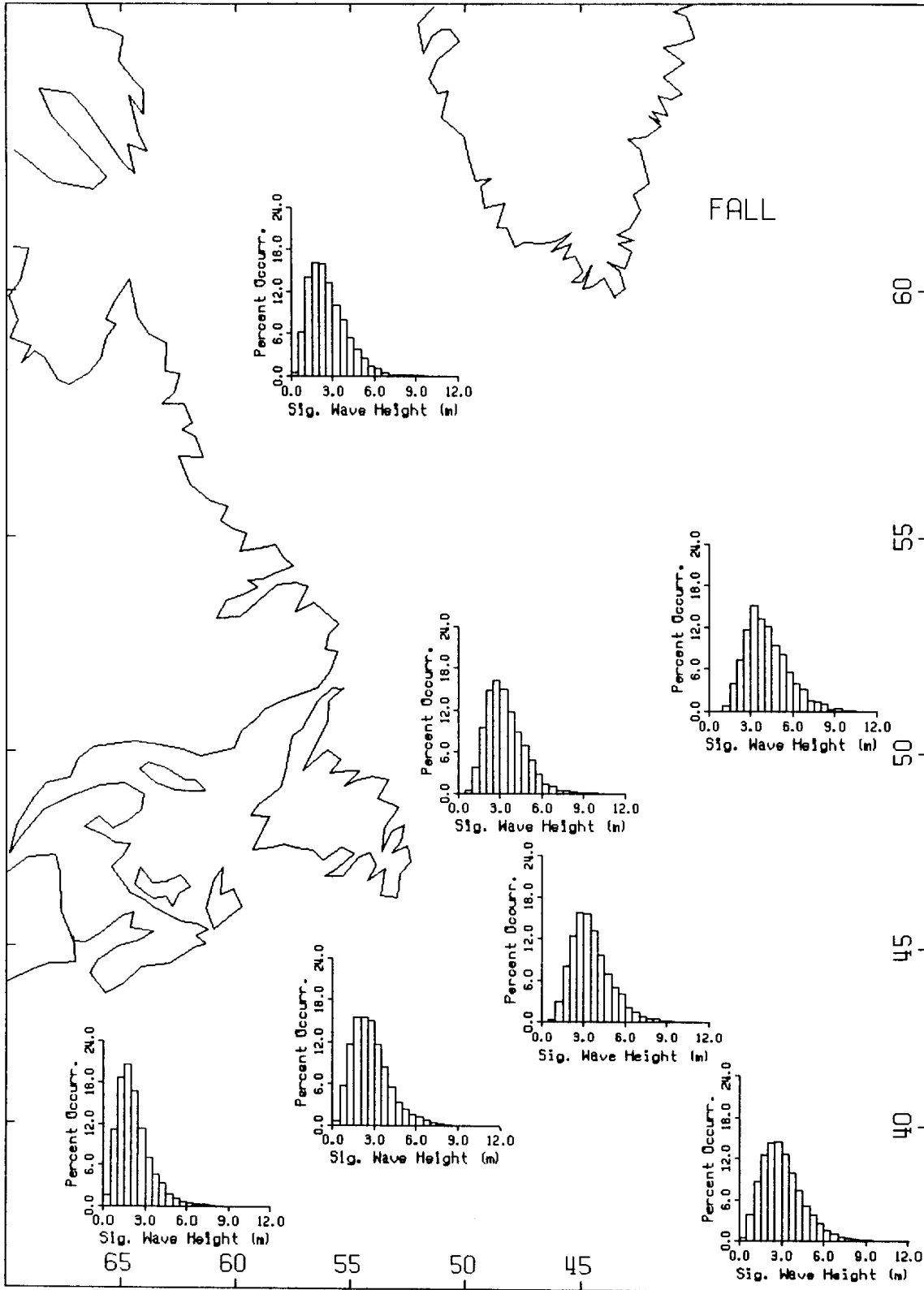
Fig. 16c Fall percent occurrence histograms of HSIG for each region and overall.
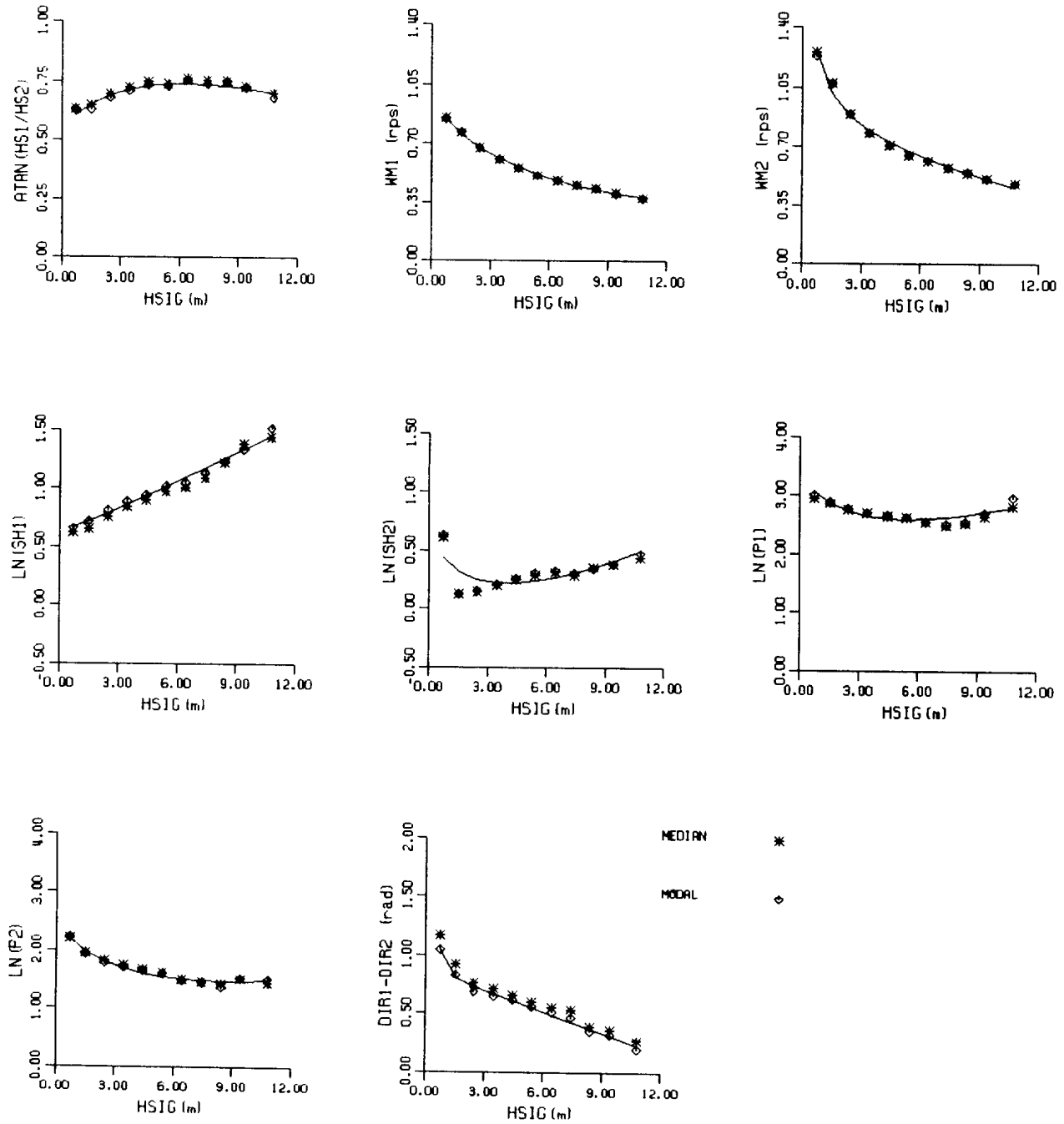
Fig. 17 Regression fits for the average modal spectra –
combined annual data.
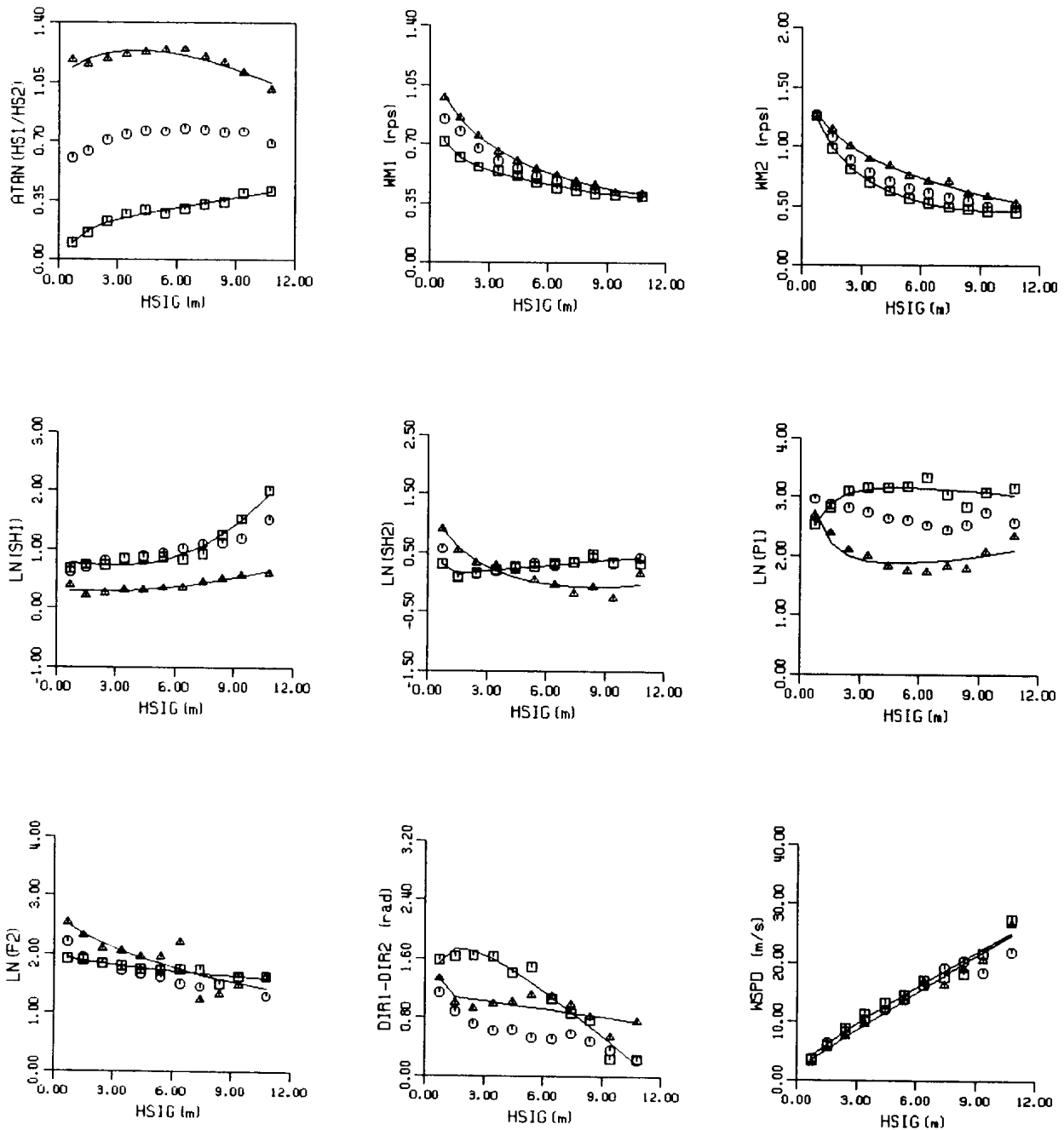
Fig. 18 Regression fits for the 95% confidence spectra - combined annual data. The target parameter plot is noted by the enclosing square. Squares: lower limit; circles: mode; triangles: upper limit.
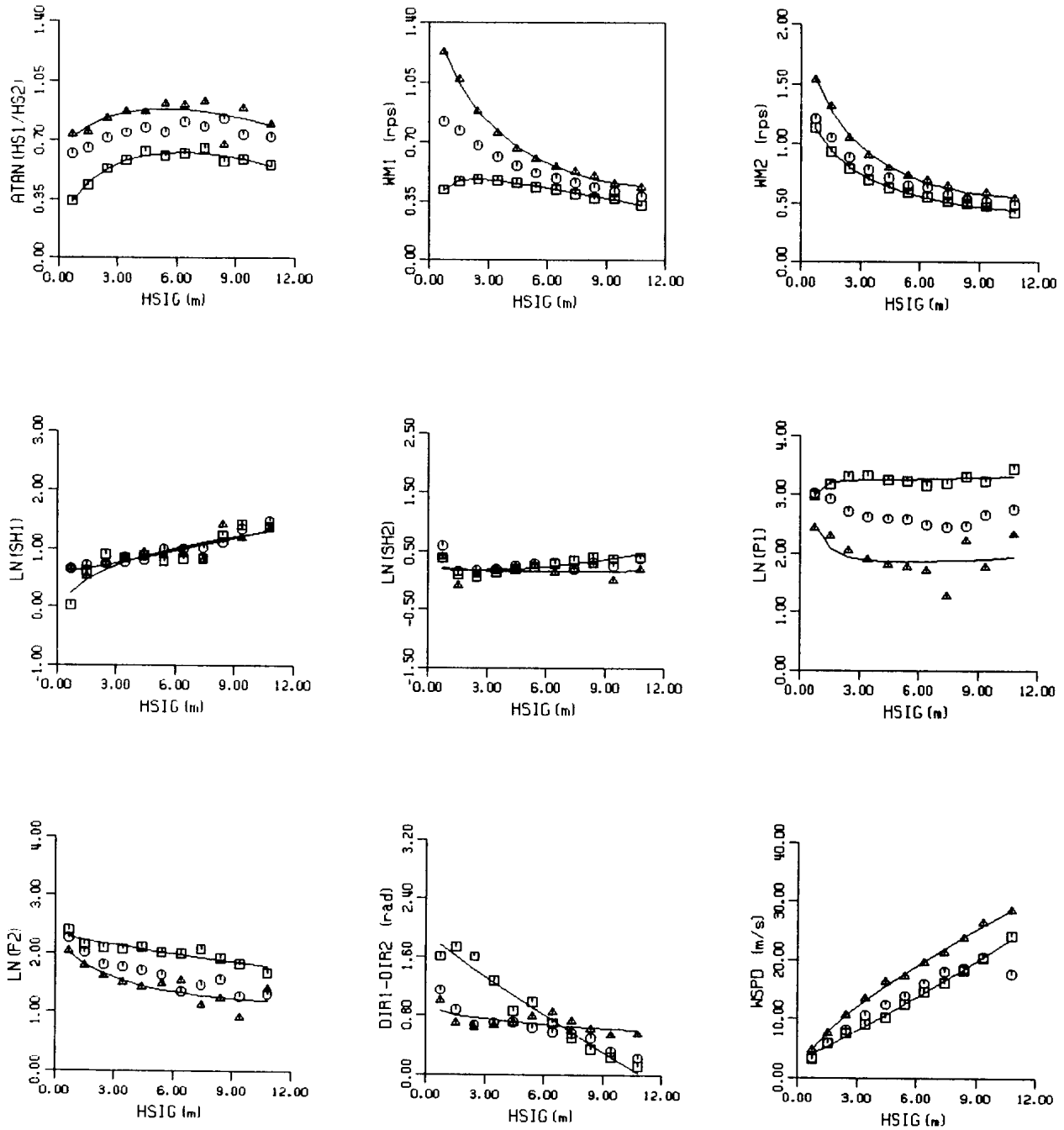
Fig. 18 (continued).
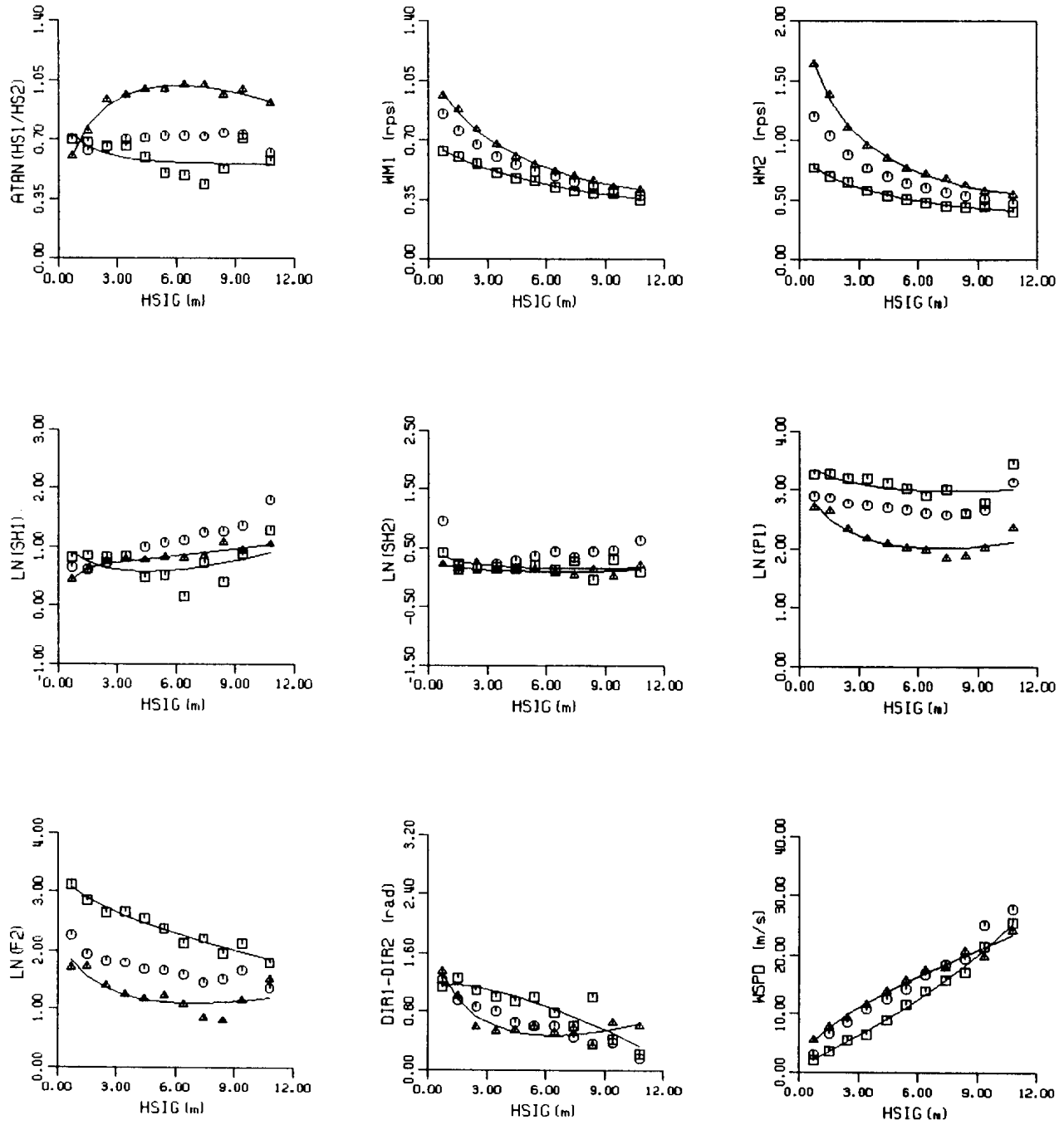
Fig. 18 (continued).   77
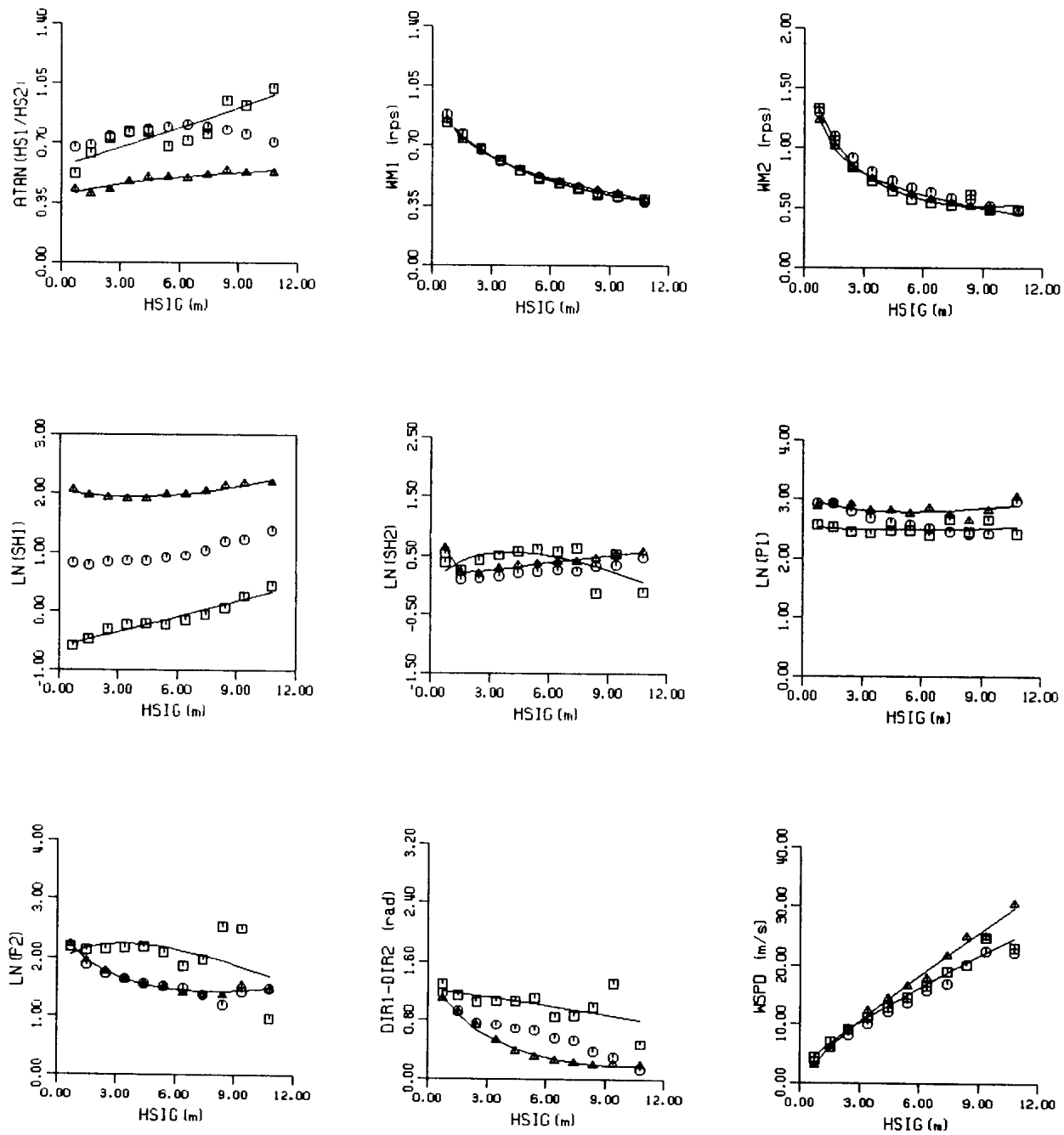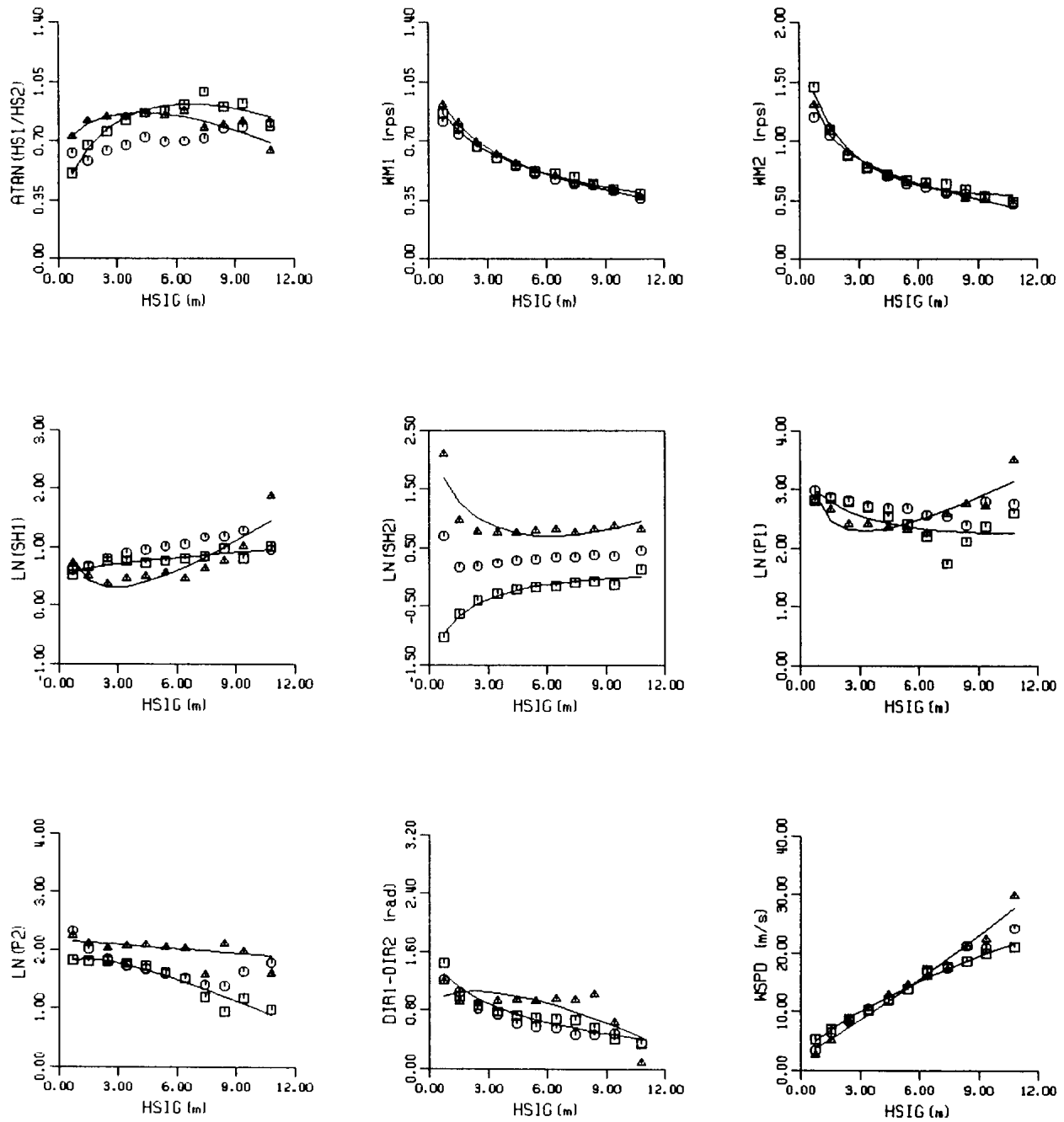
Fig. 18 (continued).
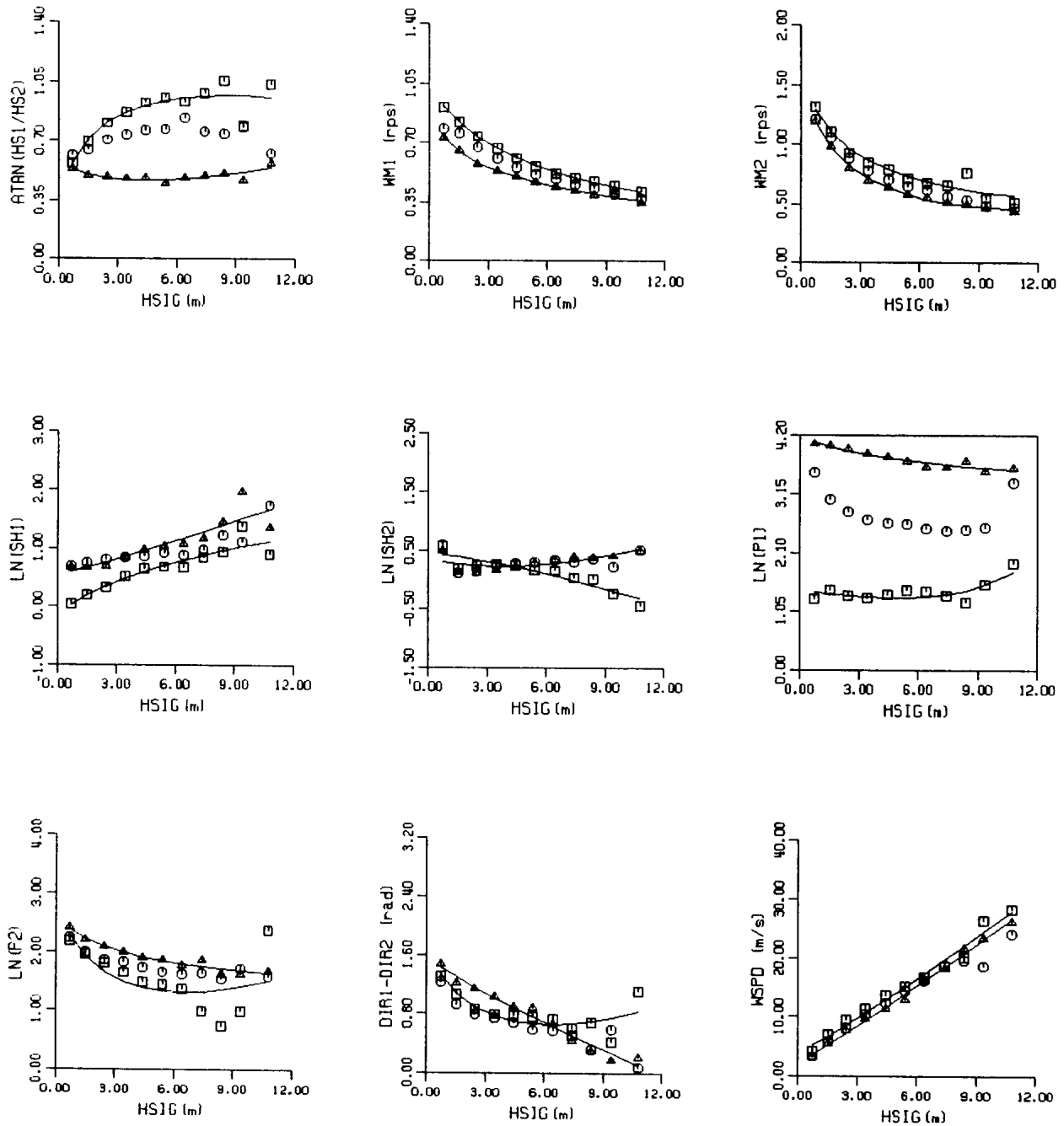
Fig. 18 (continued).
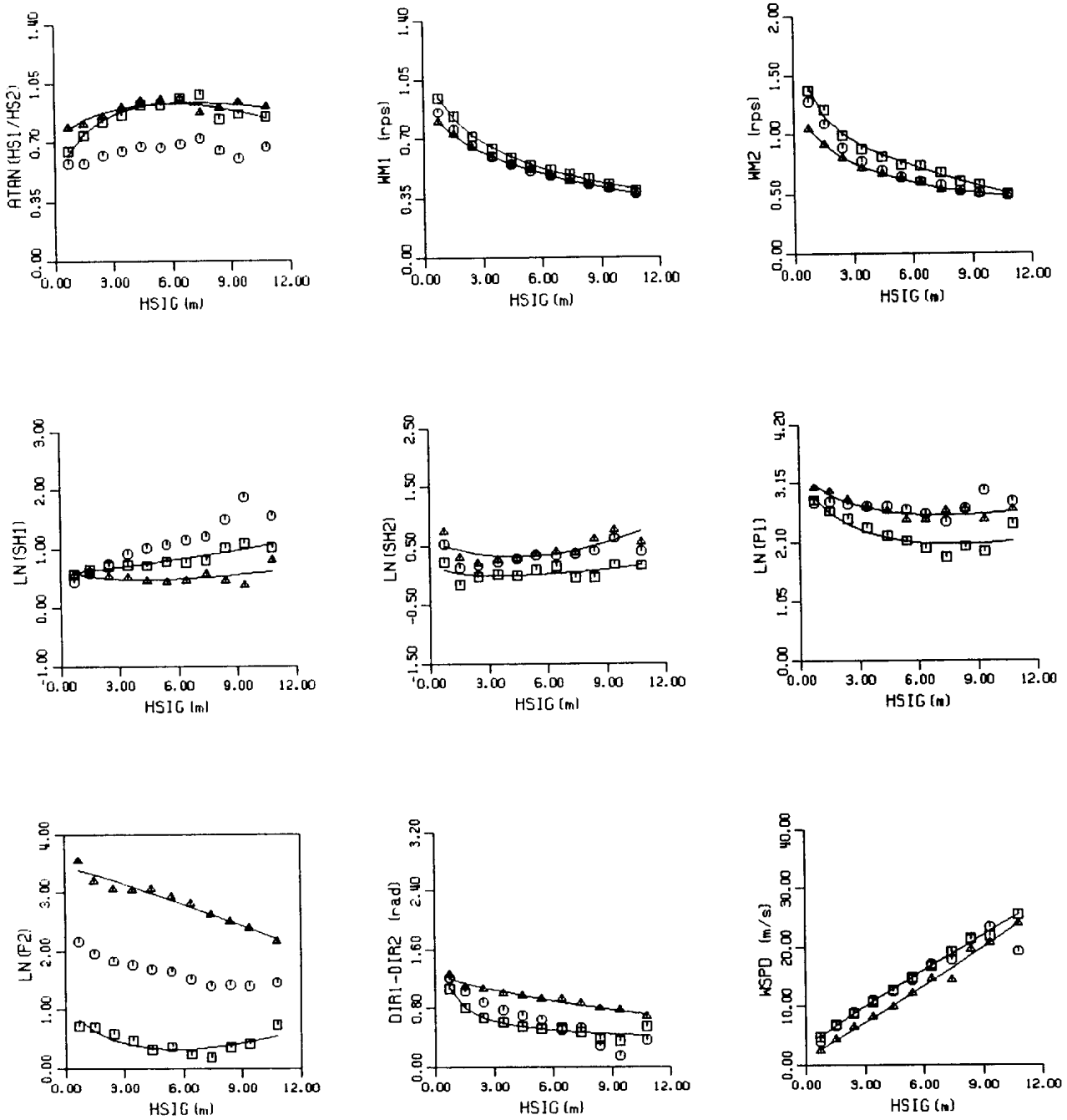
Fig. 18 (continued).
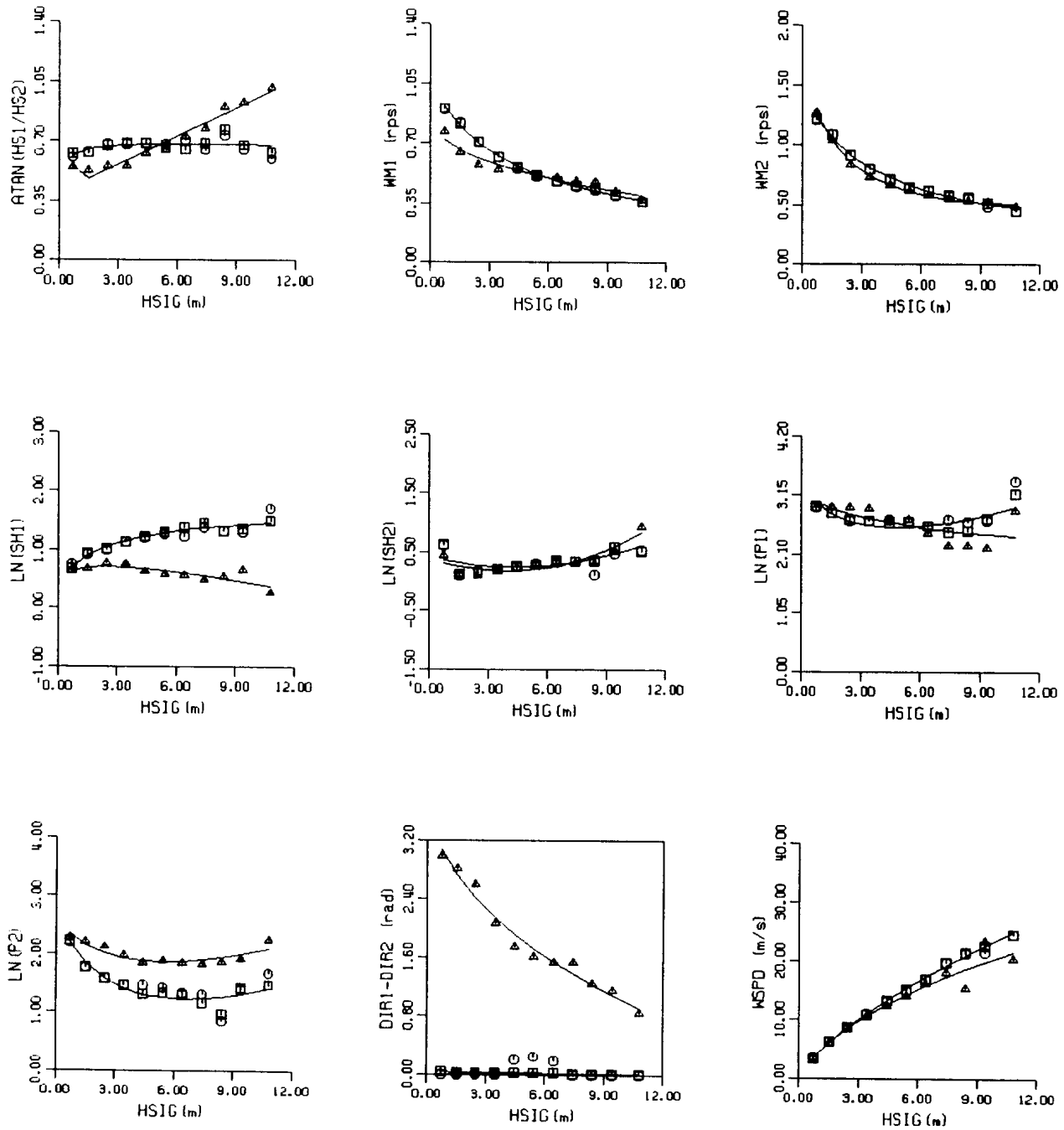
Fig. 18 (continued).
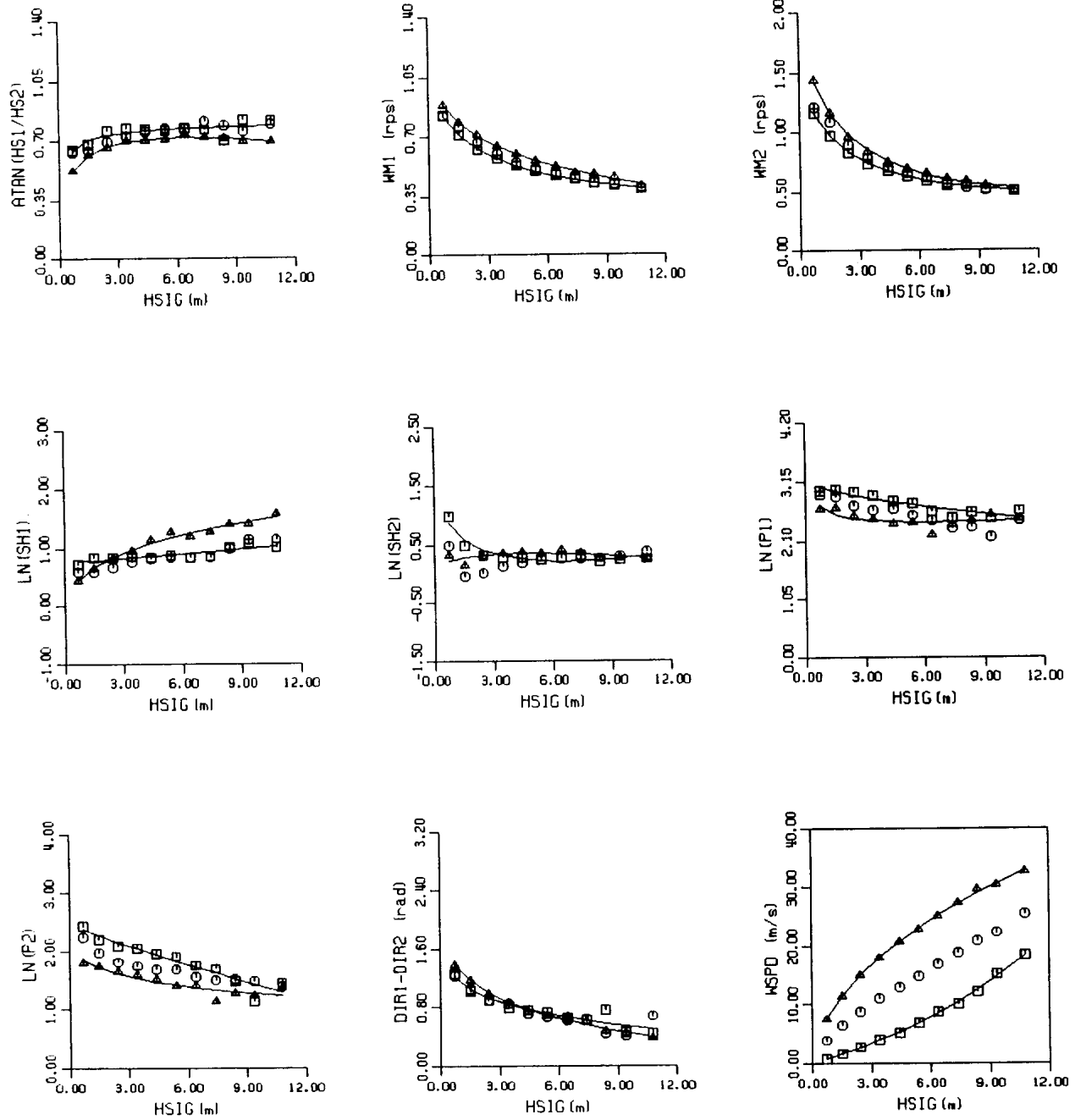
Fig. 18 (continued).

Fig. 18 (continued).

In order to examine the regional variability, the target parameter regressions, for the six regions, and at the three probability levels are shown in Fig 19 . In general, the regional regressions show similar behaviour with the greatest variability seen in the shape and spread parameters. An examination of the catchment range about the target parameters indicated that the separation between regressions may be significant for the majority of the modal fits as well as for selected confidence. fits (e.g. the catchment range for the lower 95% confidence fit of wind speed do not overlap for regions 1 and 2). The relative importance of the slight differences will depend on the sensitivity of the given application.

4.4 Role of the Shape Parameter

The predictability of the shape parameter from other fit parameters, with particular focus on the modal frequency, was examined in order to determine the possibility of reducing the number of adjustable parameters in ship motion modelling. Scatterplots of $1n(\lambda)$ vs ω_m were found to show high levels of noise with λ being approximately constant for frequencies above 0.65 rps while the scatter below this frequency, possibly reflecting fit behaviour as opposed to any physical relationship, was too large to draw any conclusions. This is supported by examination of Fig. 18   when the two shape parameters are the targets. It can be seen that, despite the wide separation in $1n(\lambda)$ for the 95% confidence limits at a particular HSIG level, there is almost no variation in the corresponding averaged ω_m value. The ω_m averages, however, do vary with other target parameters. λ values are sensitive to the energy "split" between the two components of the model (observe, the response of the $\tan^{-1}(\delta_1/\delta_2)$ parameter). There is a slight increase of λ_1 with HSIG possibly reflecting sea peak steepening coinciding with growth while λ_2 remains relatively constant for HSIG 3.0m. A proper assessment of the role of the shape parameters would require re-fitting of the 10-parameter model to the input ODGP spectra, using software which incorporates the different functional relationships for λ, and examining the associated residual errors.
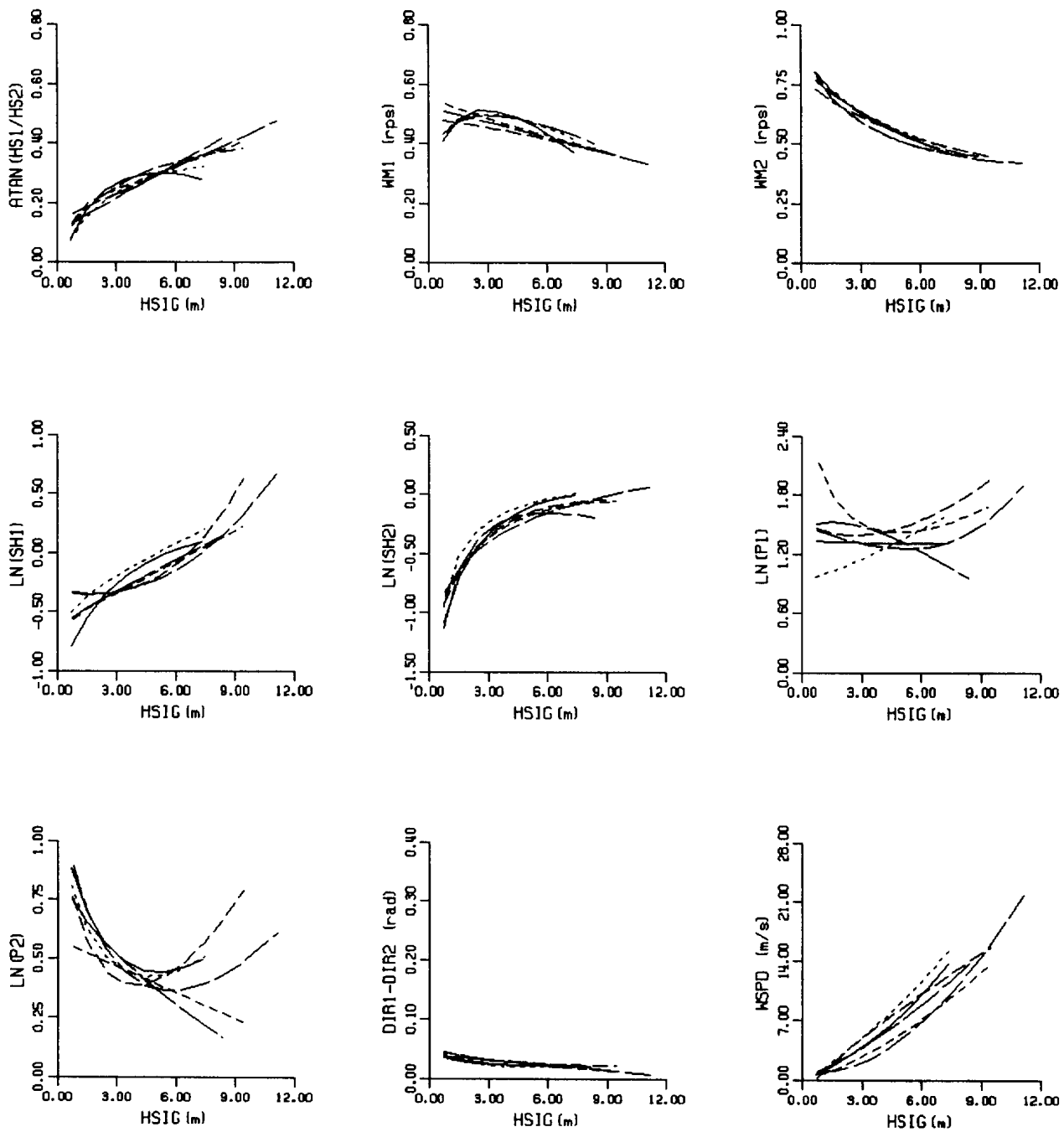
Fig. 19a Regression fits for the lower 95% confidence spectra
for the six regions when the given parameter is the
target value. Regions are noted by the increasing dash
length from areas 1 to 6 (solid).
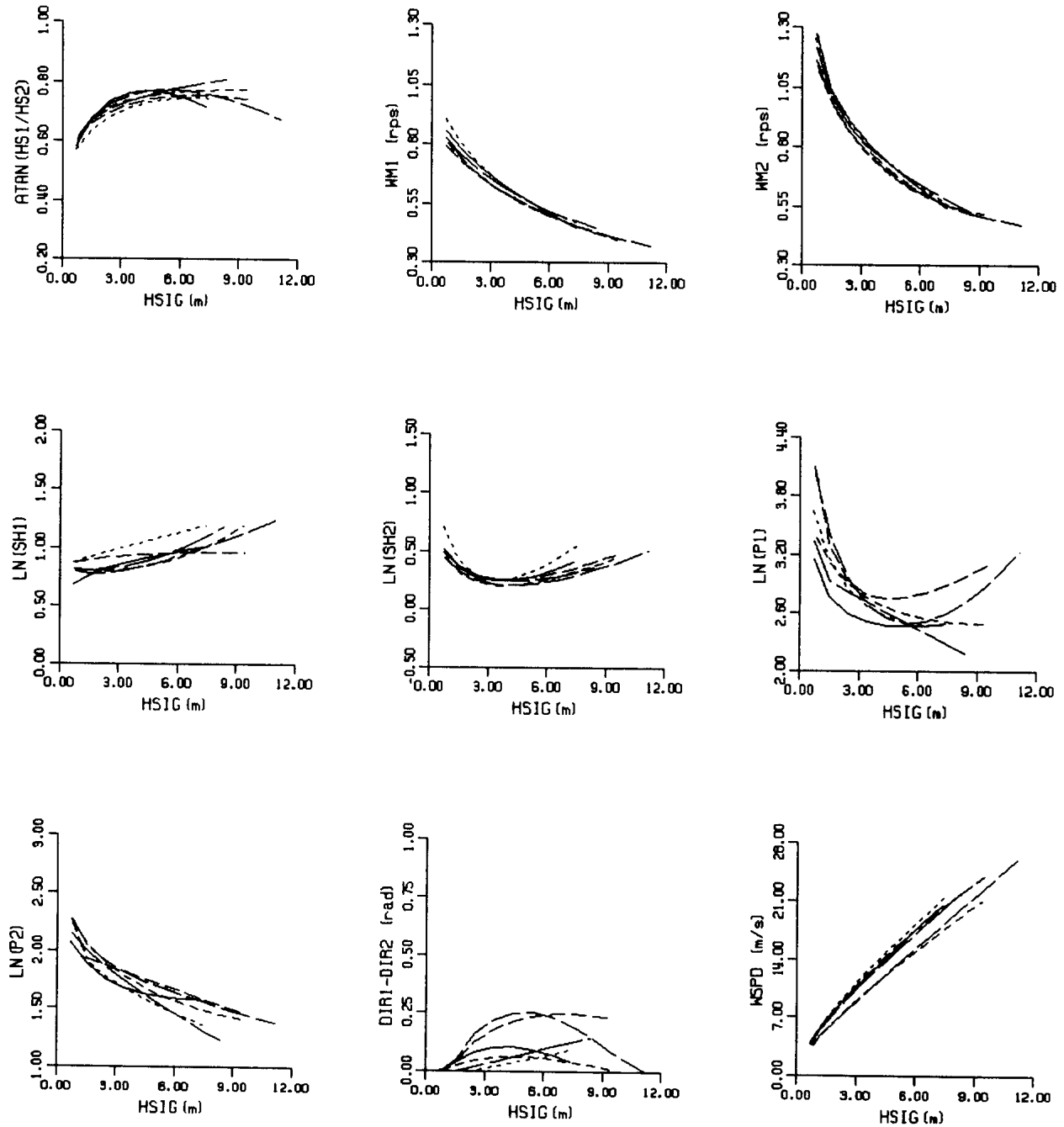
Fig. 19b Regression fits for the modal spectra.
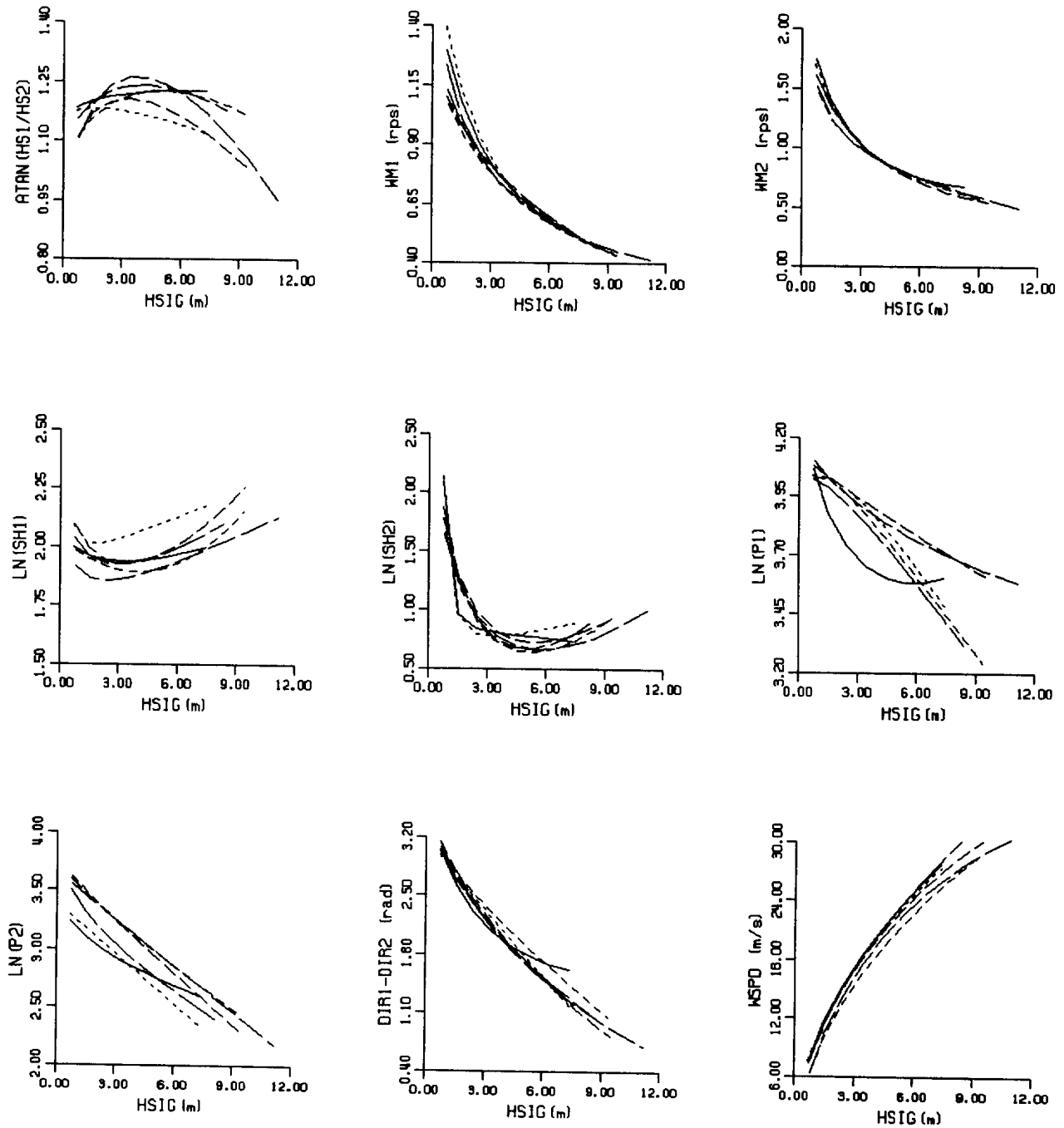
Fig. 19c Regression fits for the upper 95% confidence spectra.

5.0 SUMMARY

In this study, 220151 ODGP hindcast directional spectra, from 53 locations, were fit by a 10-parameter directional wave model. It was found that the fit for over 93% of the records (204875 spectra) was acceptable based on residual error estimates. It was also noted that, for the combined data set, 55.2% of all spectra (52.4% of acceptable fits) were judged to have more than one directional spectral peak. Hence the use of a model which can adequately represent at least two directional peaks was proven to be necessary for proper statistics and the chosen 10-parameter model was validated.

The numerical techniques of empirical orthogonal function, principal component and factor analysis were shown to be useful in the determination of a regional separation or clustering of the measurement sites, based on wave and wind features. The analyses distinguished six regions which corresponded well with known geographic divisions. These were: 1) Labrador Shelf; 2) Eastern Offshore Open ocean; 3) Northern Coast of Newfoundland; 4) Eastern Grand Banks and Flemish Cap; 5) Western Grand Banks; and 6) Scotian Shelf.

The probability and statistical analyses were performed in order to obtain a family of spectra, having known confidence limits, which could represent the range of directional spectral energy distributions observed. The averaging was performed both temporally (annual, winter and fall) and spatially (individual sites, six regions and overall). The predictive relations between the statistical fit parameters and significant wave height were established. A set of probability spectra, chosen to represent a specific temporal and spatial scale, can now be generated for any input significant waveheight as required by the given application.

APPENDIX 1. ADDITIONAL FIGURES
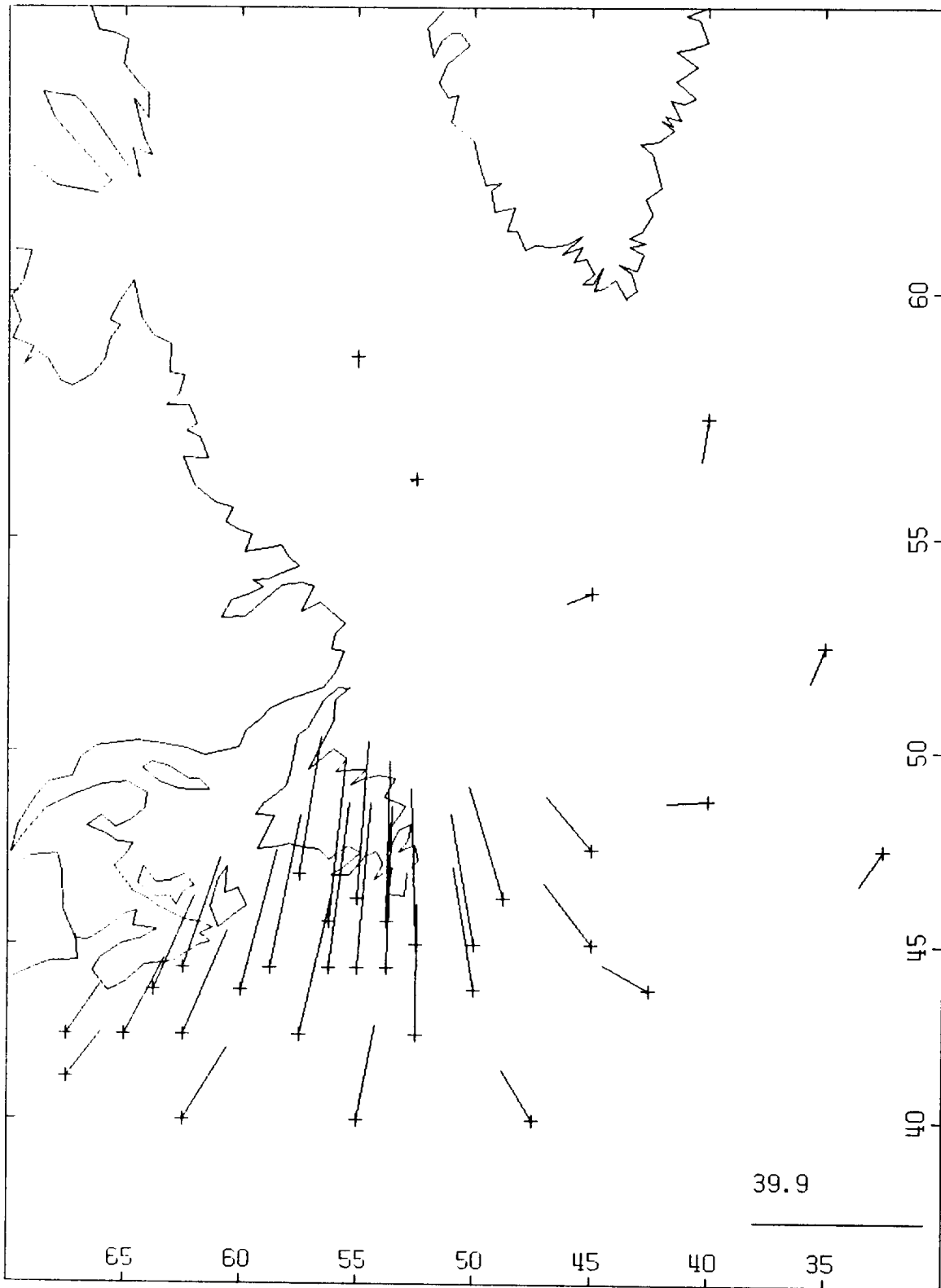
Fig. A1 Mode 1 map of eigenvector constituent values for the
M=34 wind vector analysis.
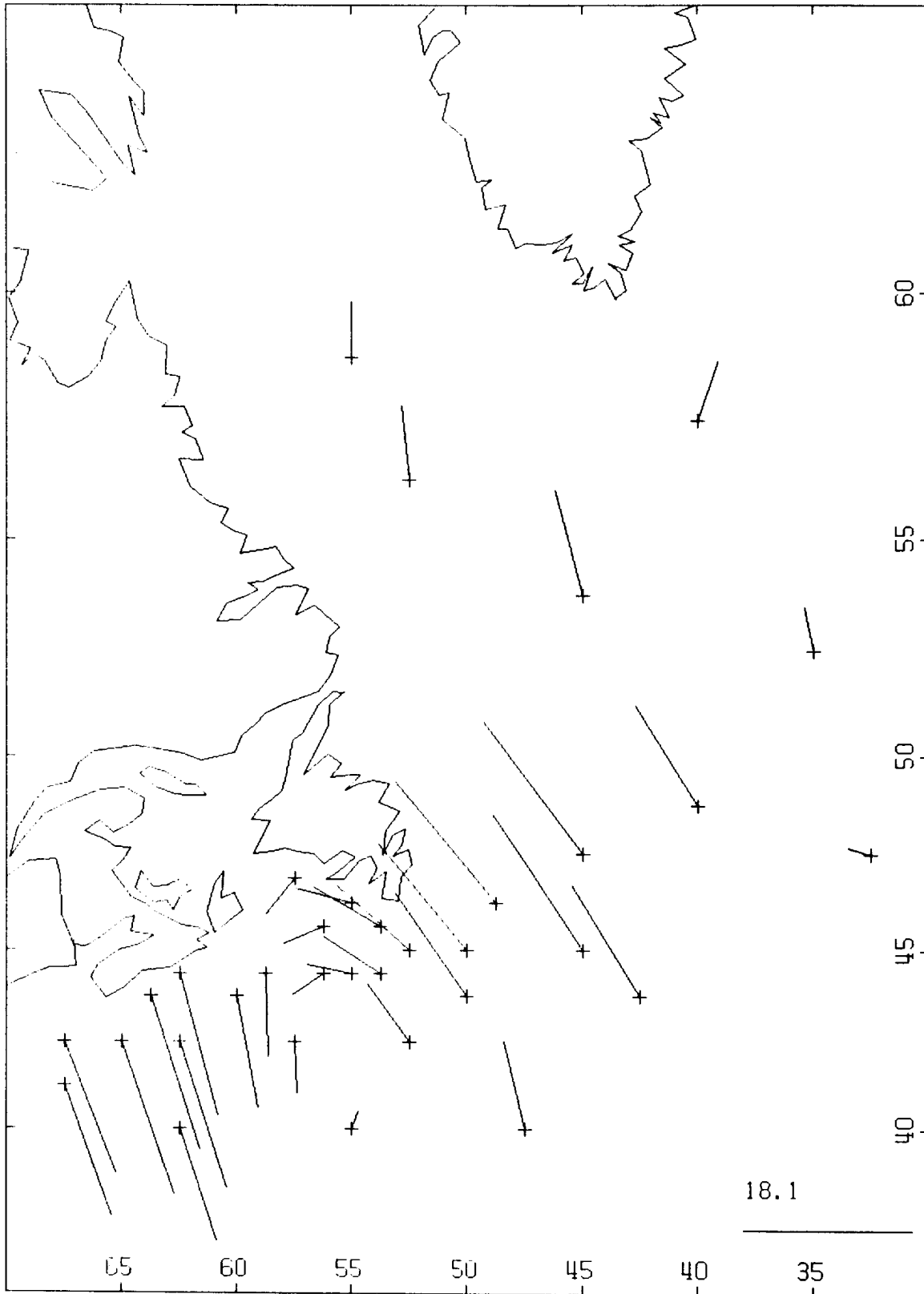
Fig. A1 (continued) Mode 2.
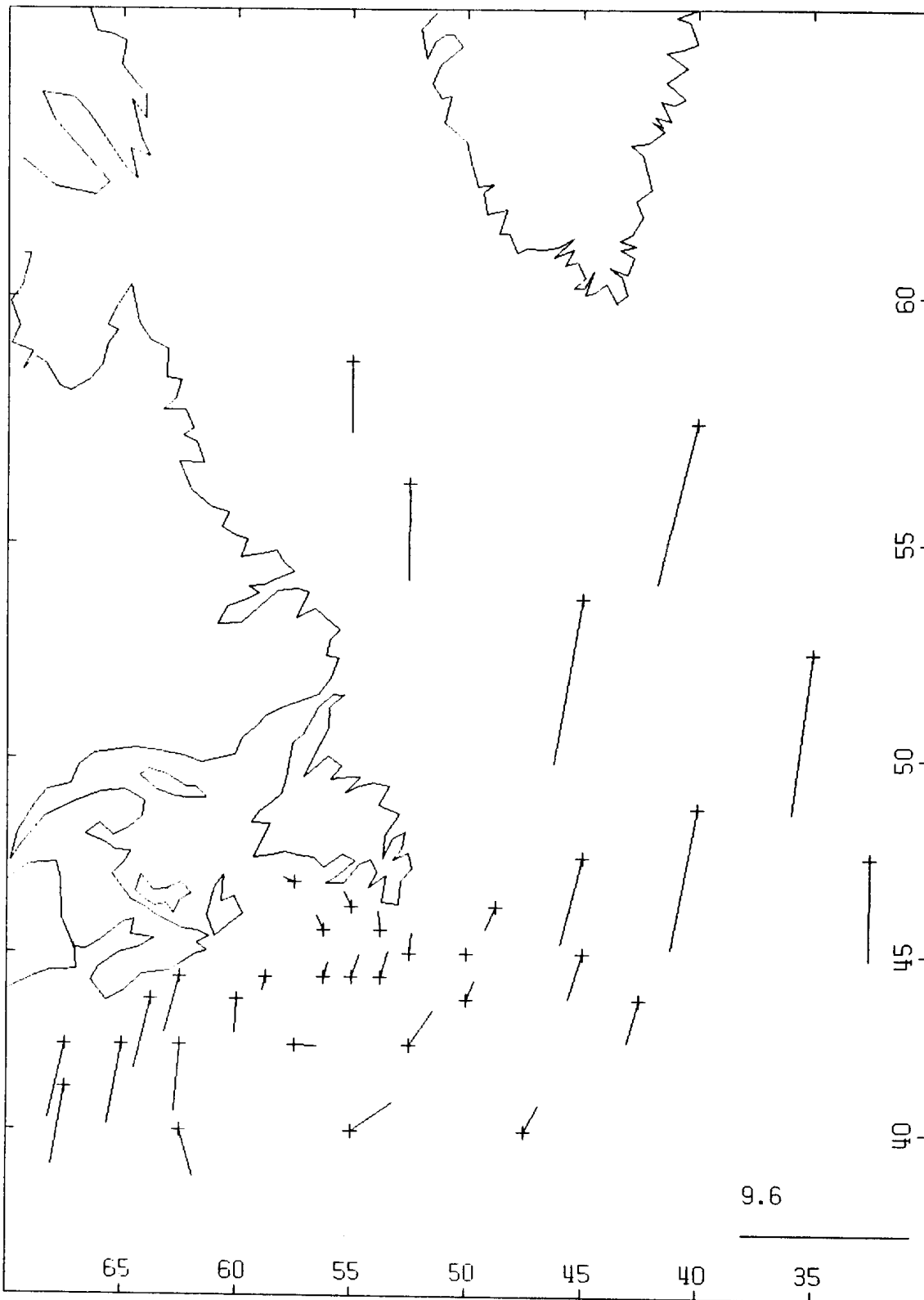
Fig. A1 (continued) Mode 3.
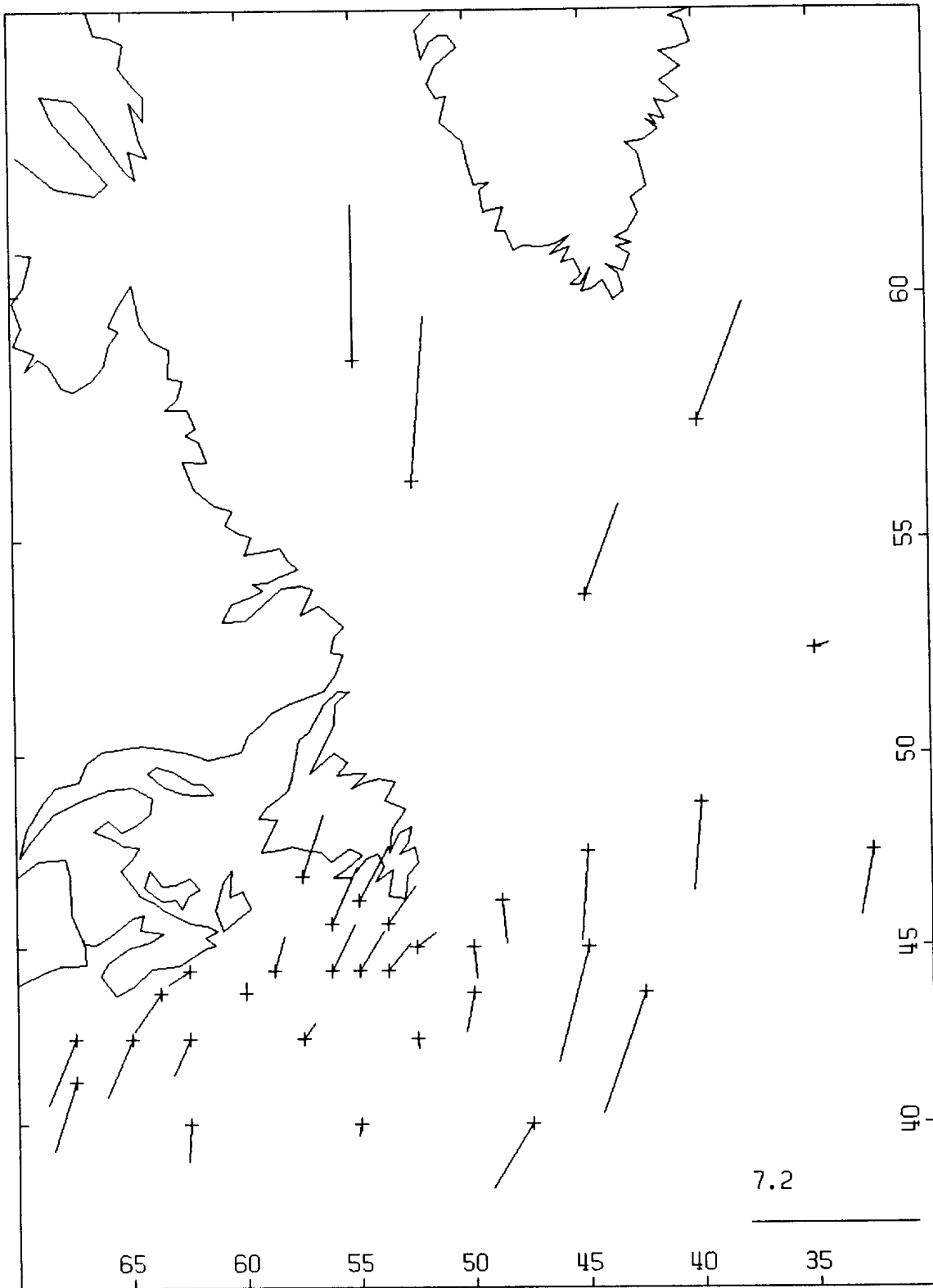
Fig. A1 (continued) Mode 4.
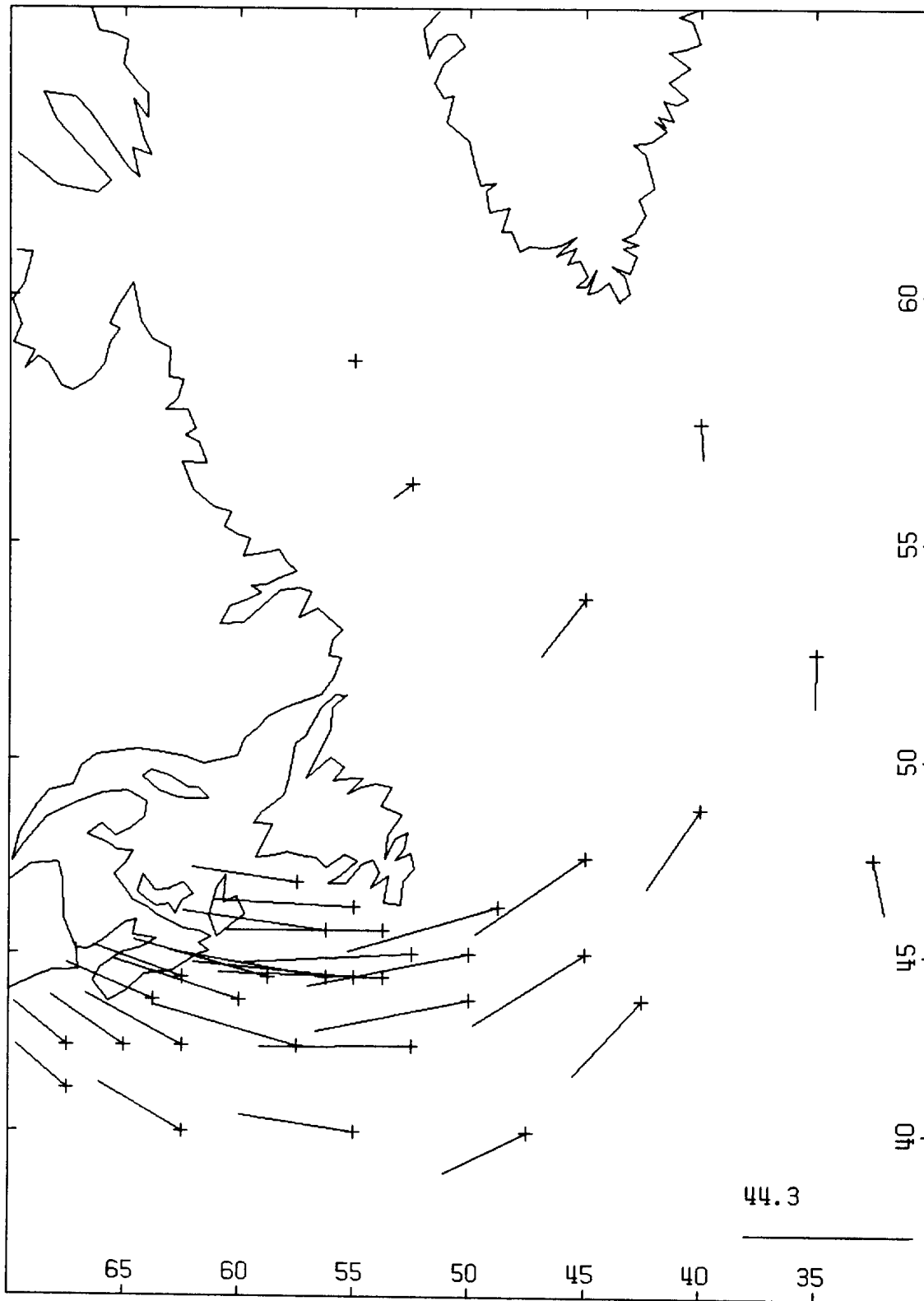
Fig. A2 Mode 1 map of eigenvector constituent values for the
M=34 total wave vector analysis.
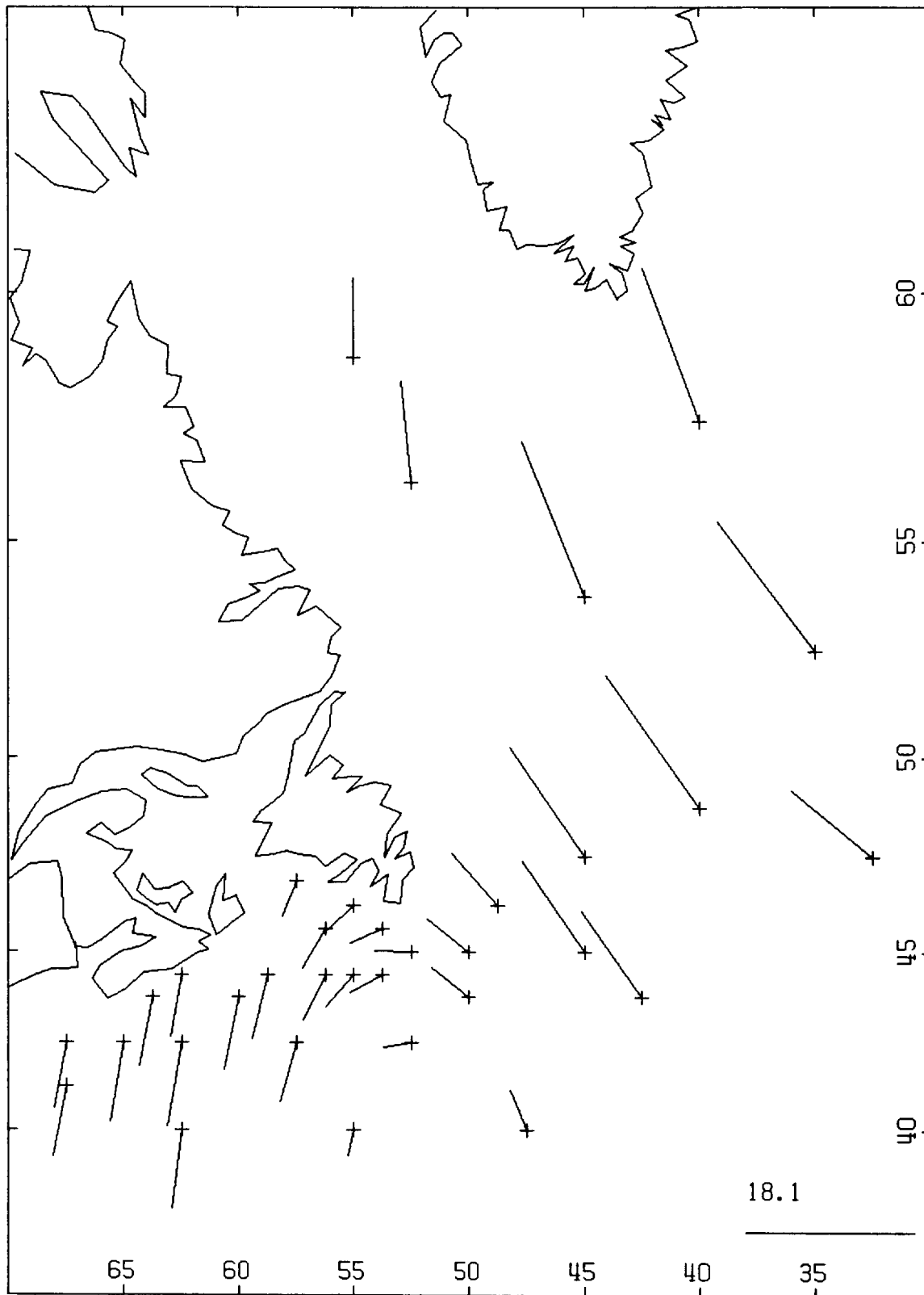
Fig. A2 (continued) Mode 2.
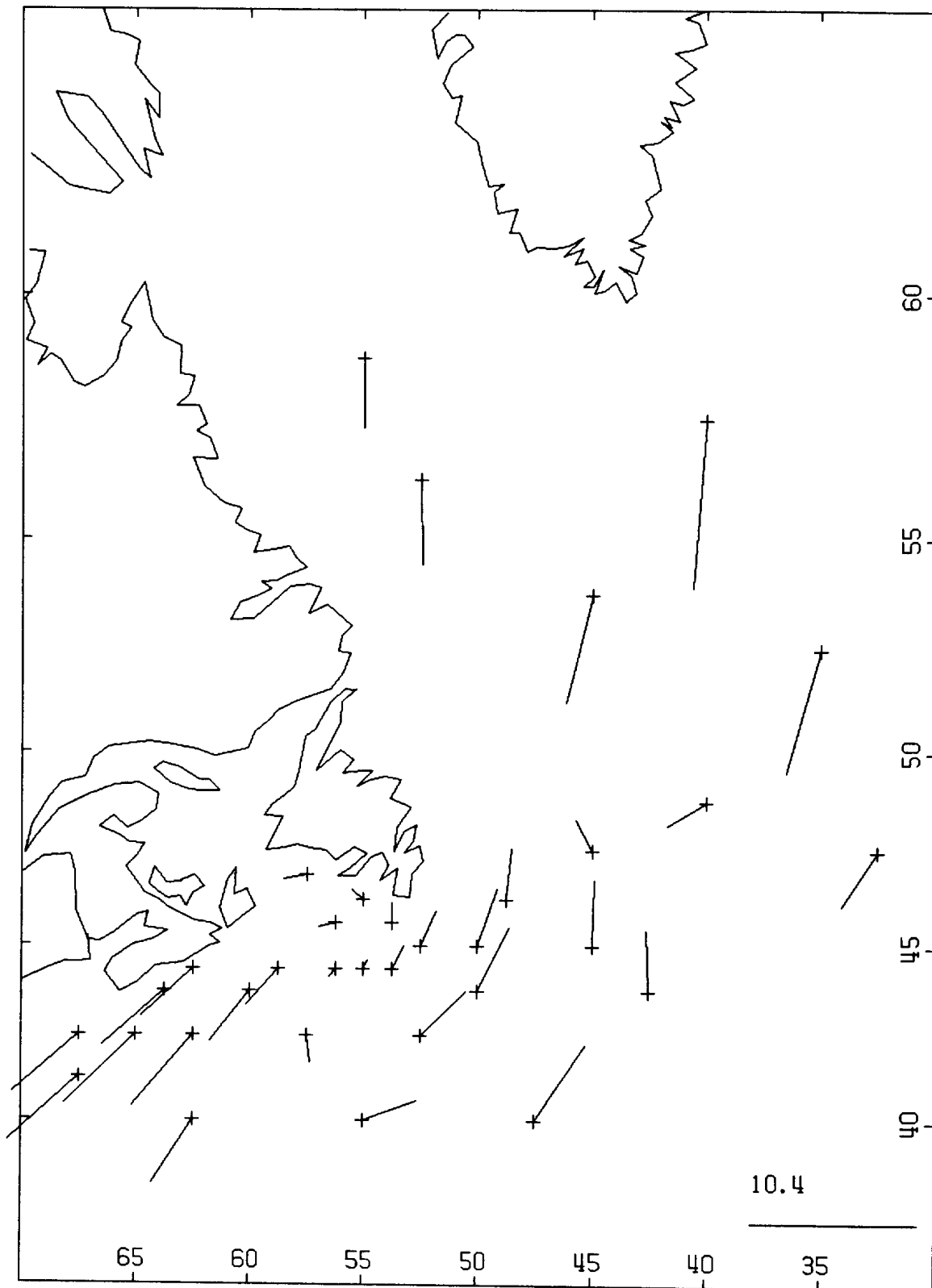
Fig. A2 (continued) Mode 3.
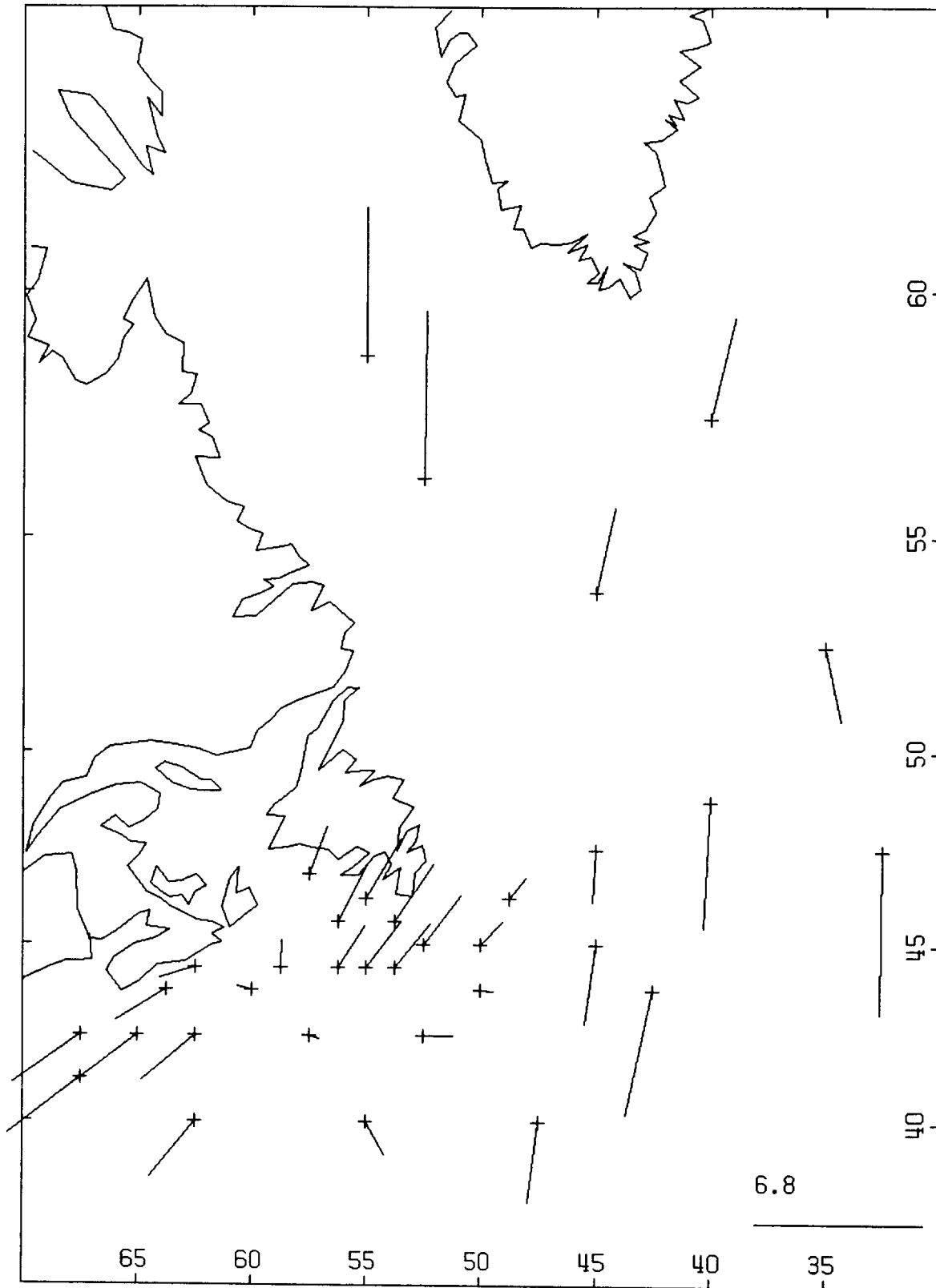
Fig. A2 (continued) Mode 4.
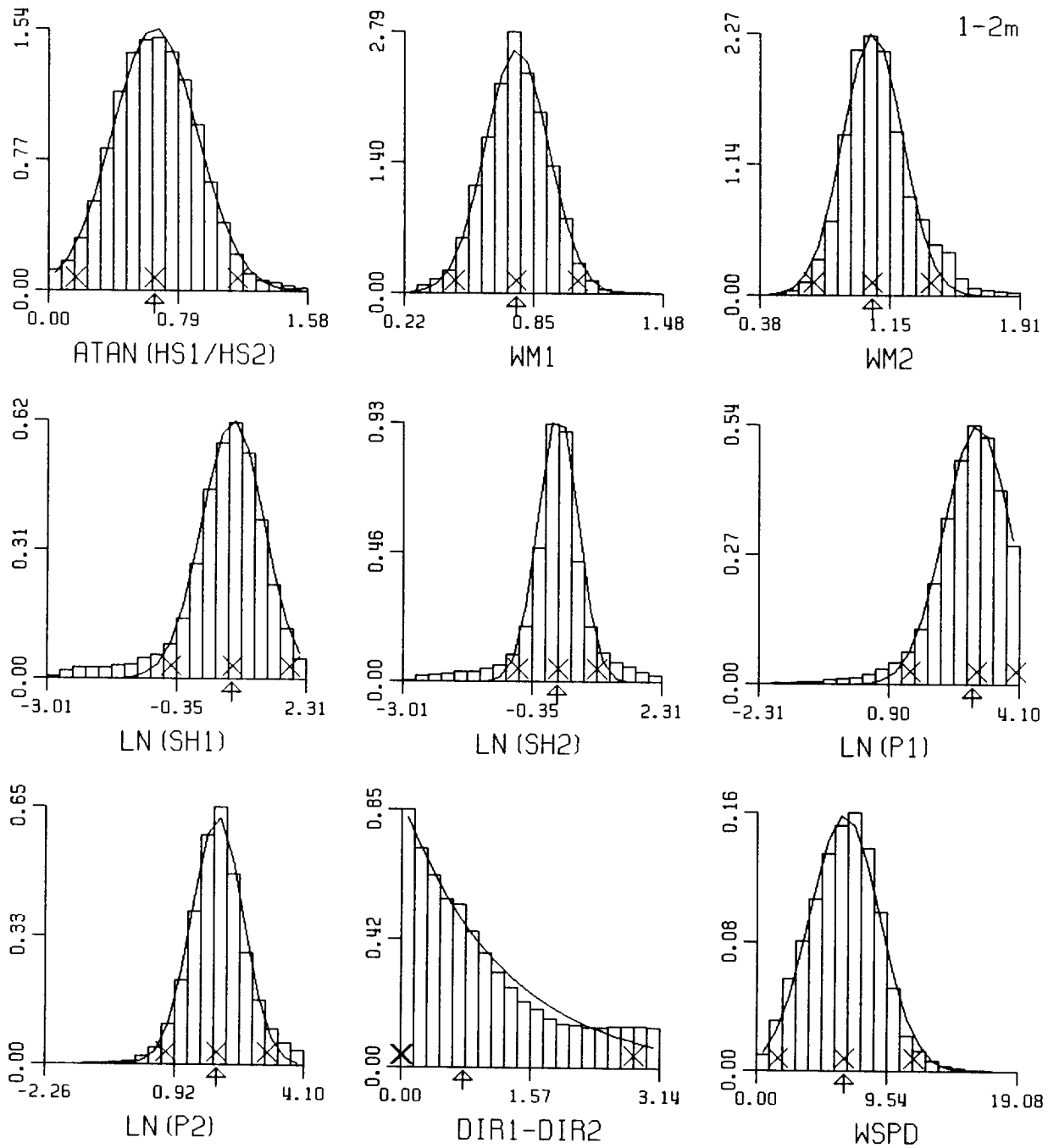
Fig. A3 Probability occurrence histograms and bounded Gaussian
fit for the eight derived model parameters and wind speed.
Combined annual data, HSIG: 1-2m.
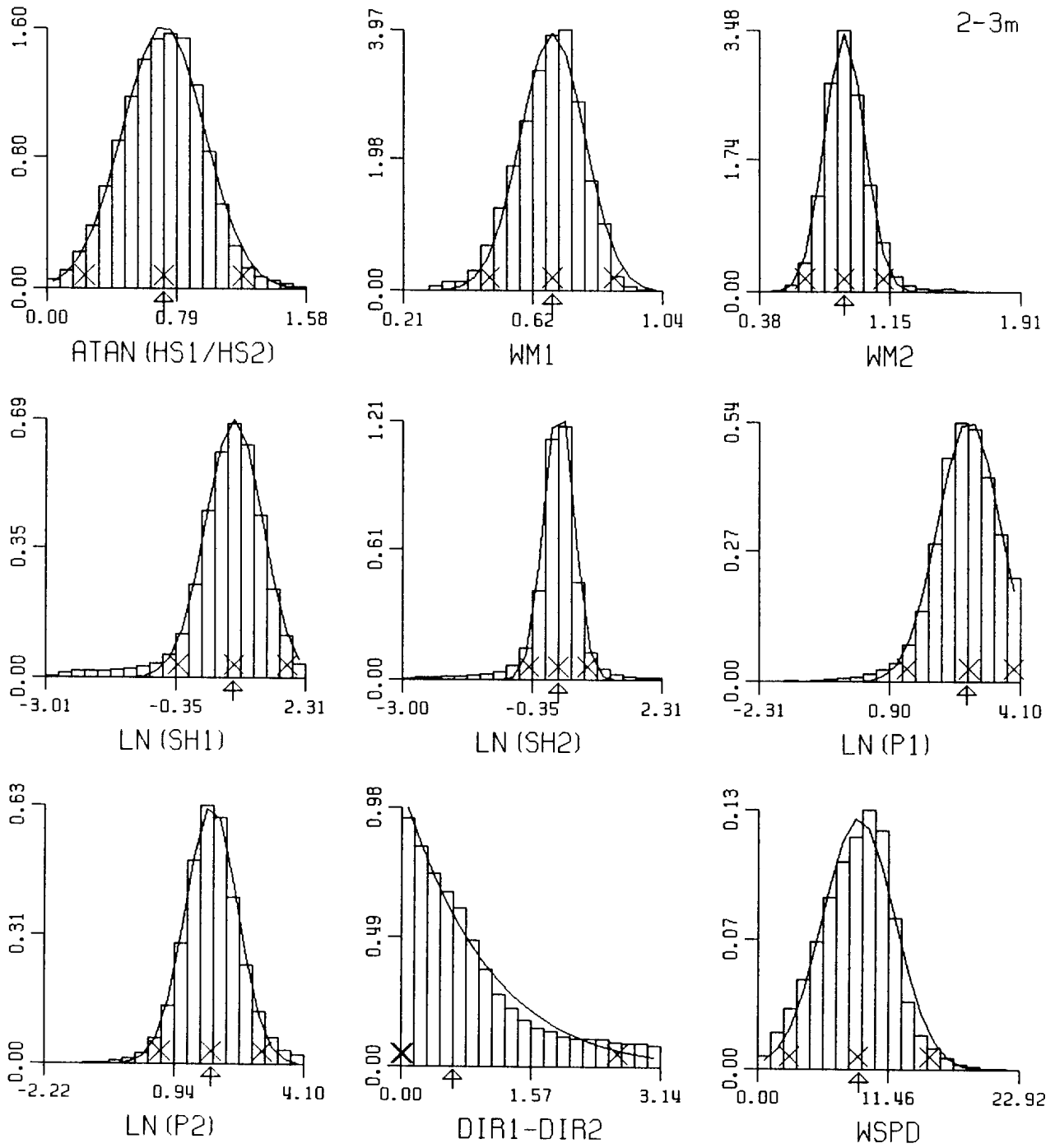
Fig. A3 (continued) HSIG: 2-3m.
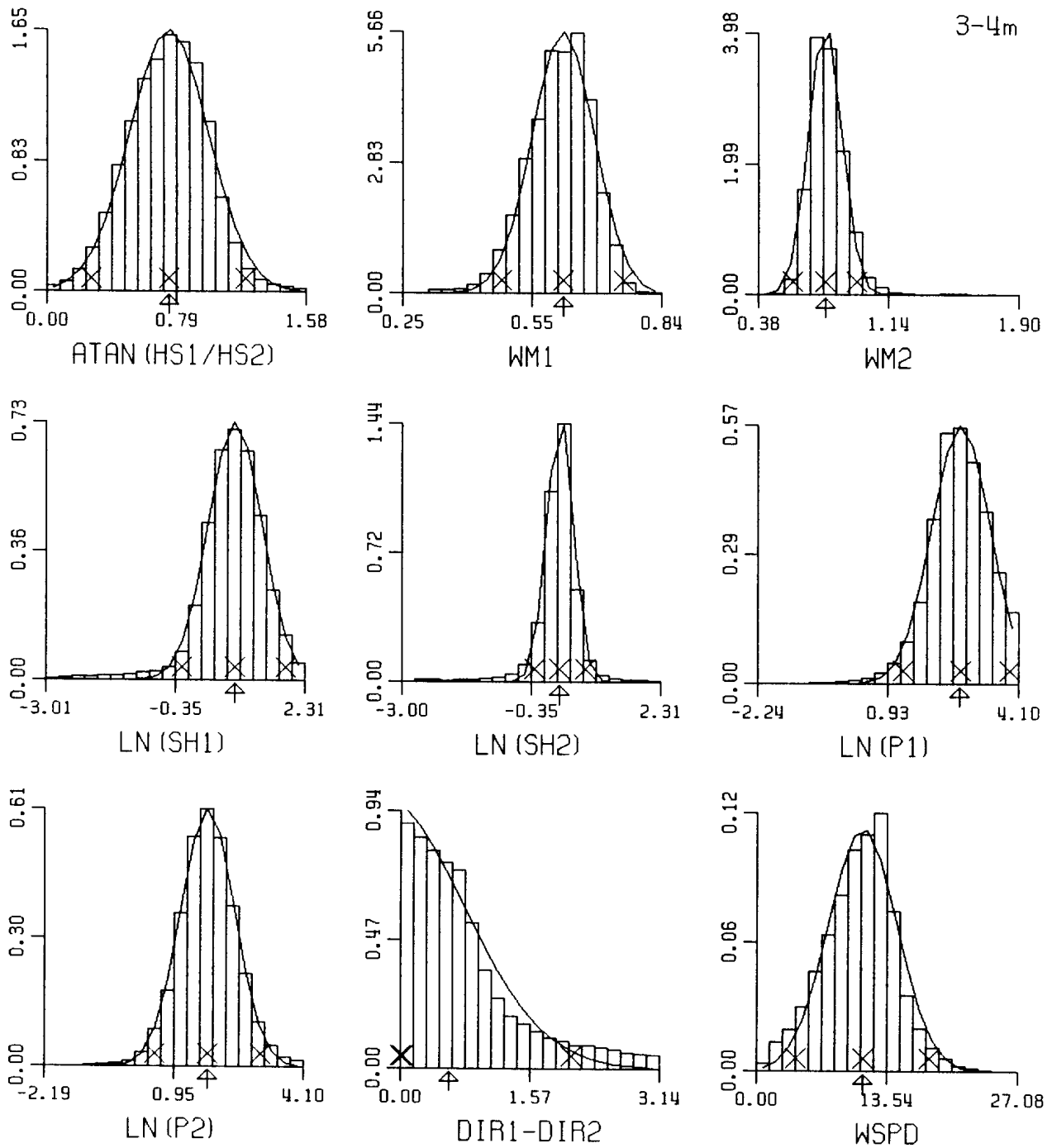
Fig. A3 (continued) HSIG: 3-4m.
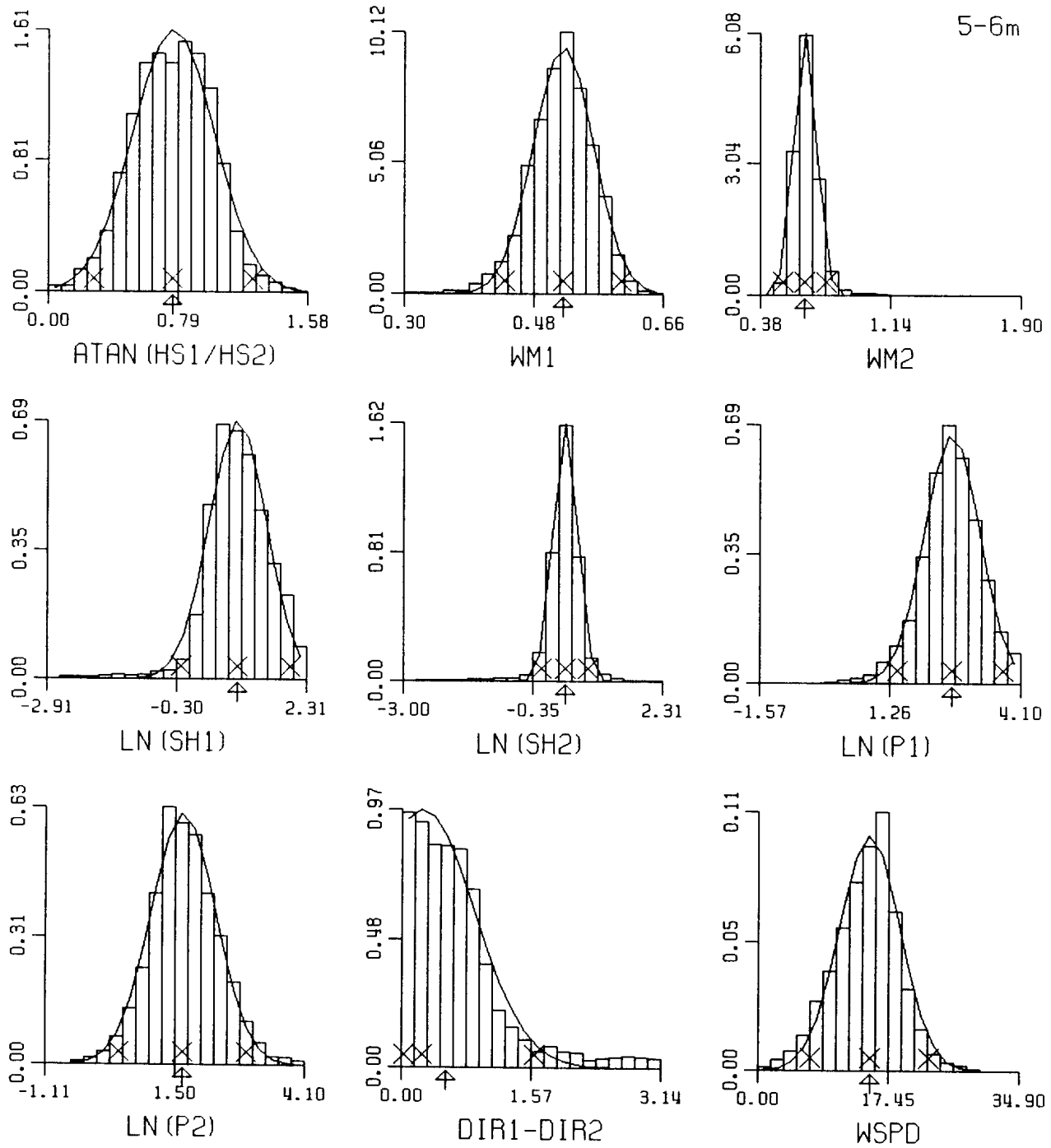
Fig. A3 (continued) HSIG: 5-6m.
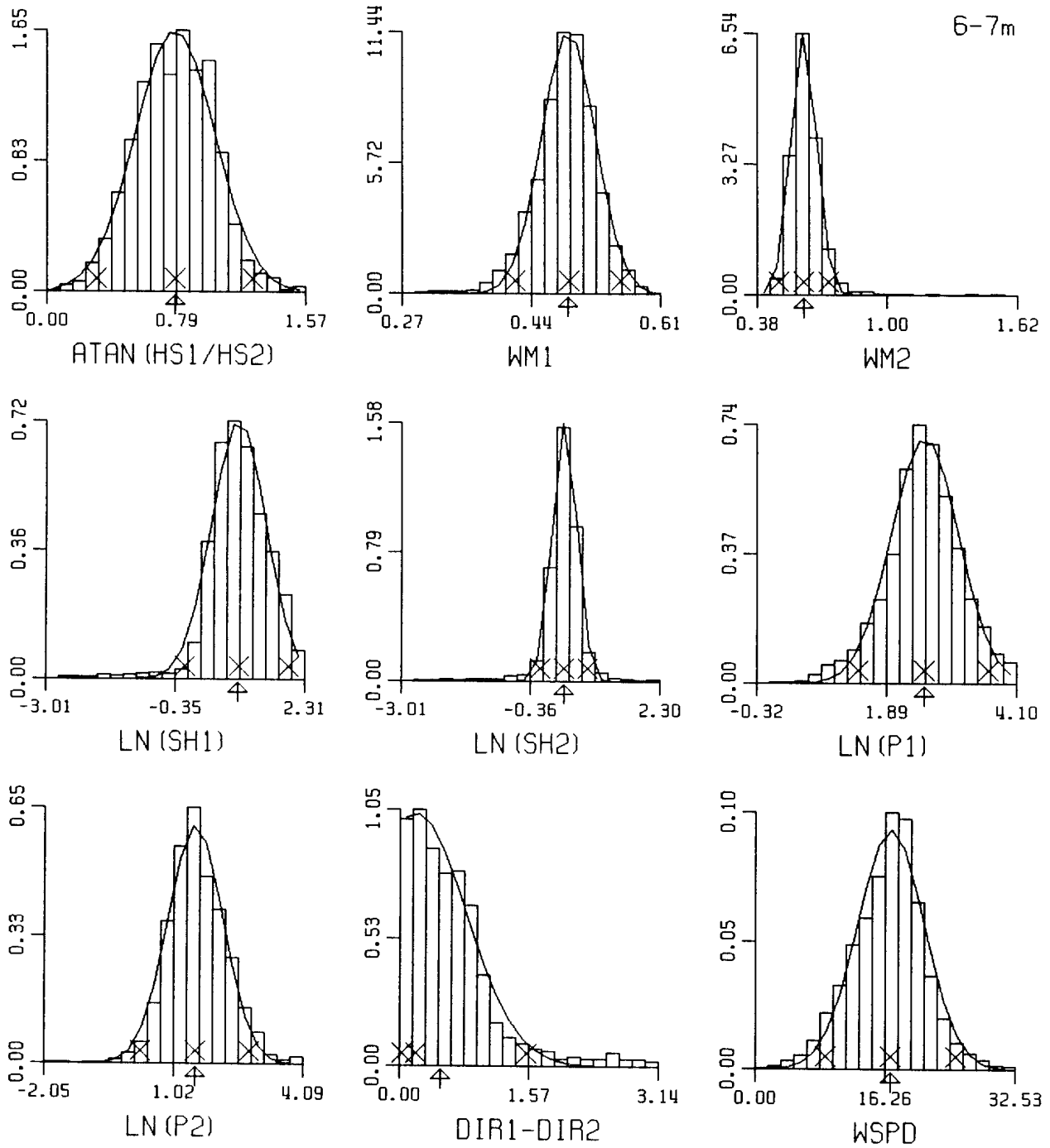
Fig. A3 (continued) HSIG: 6-7m.
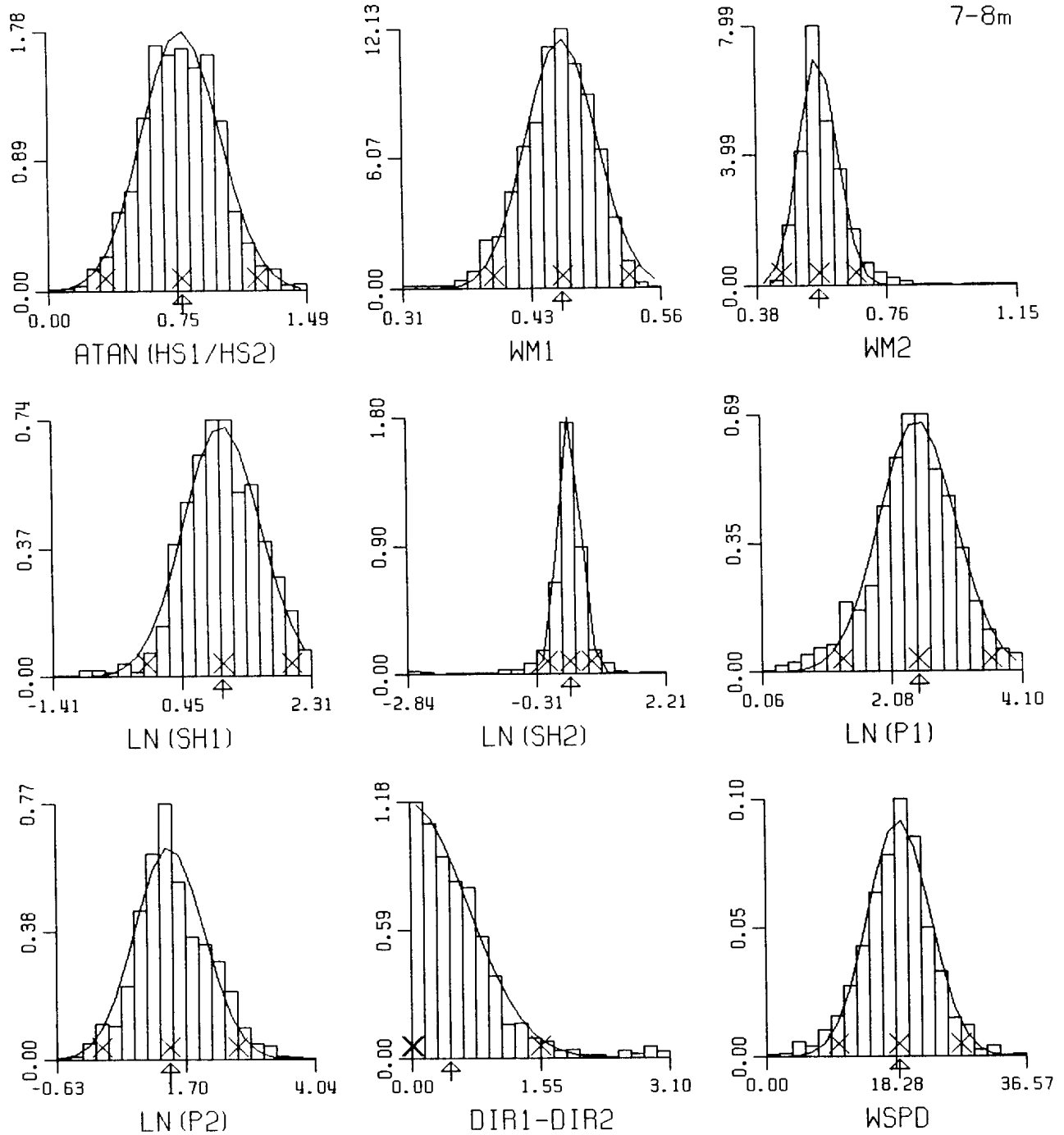
Fig. A3 (continued) HSIG: 7-8m.
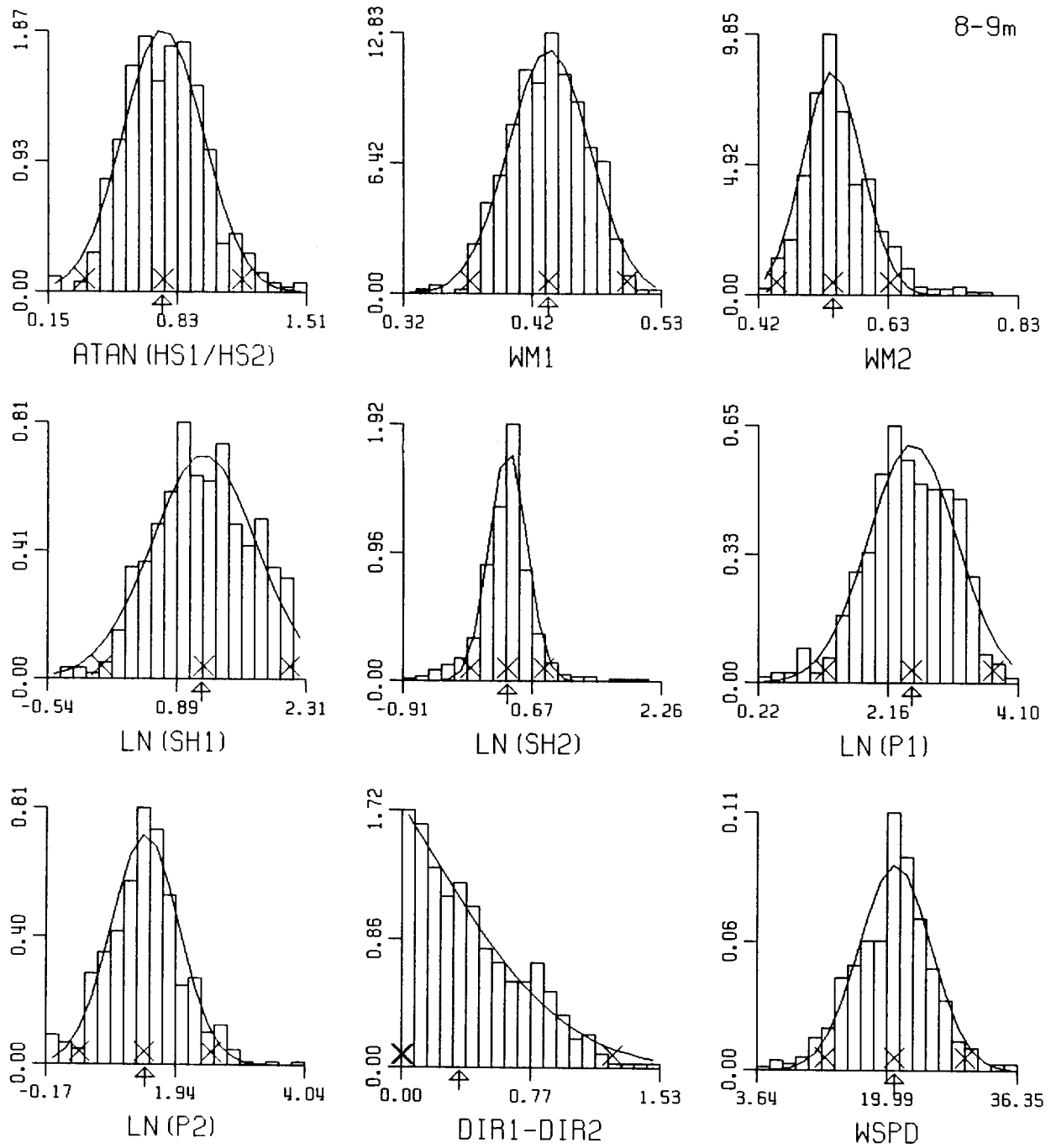
Fig. A3 (continued) HSIG: 8-9m.
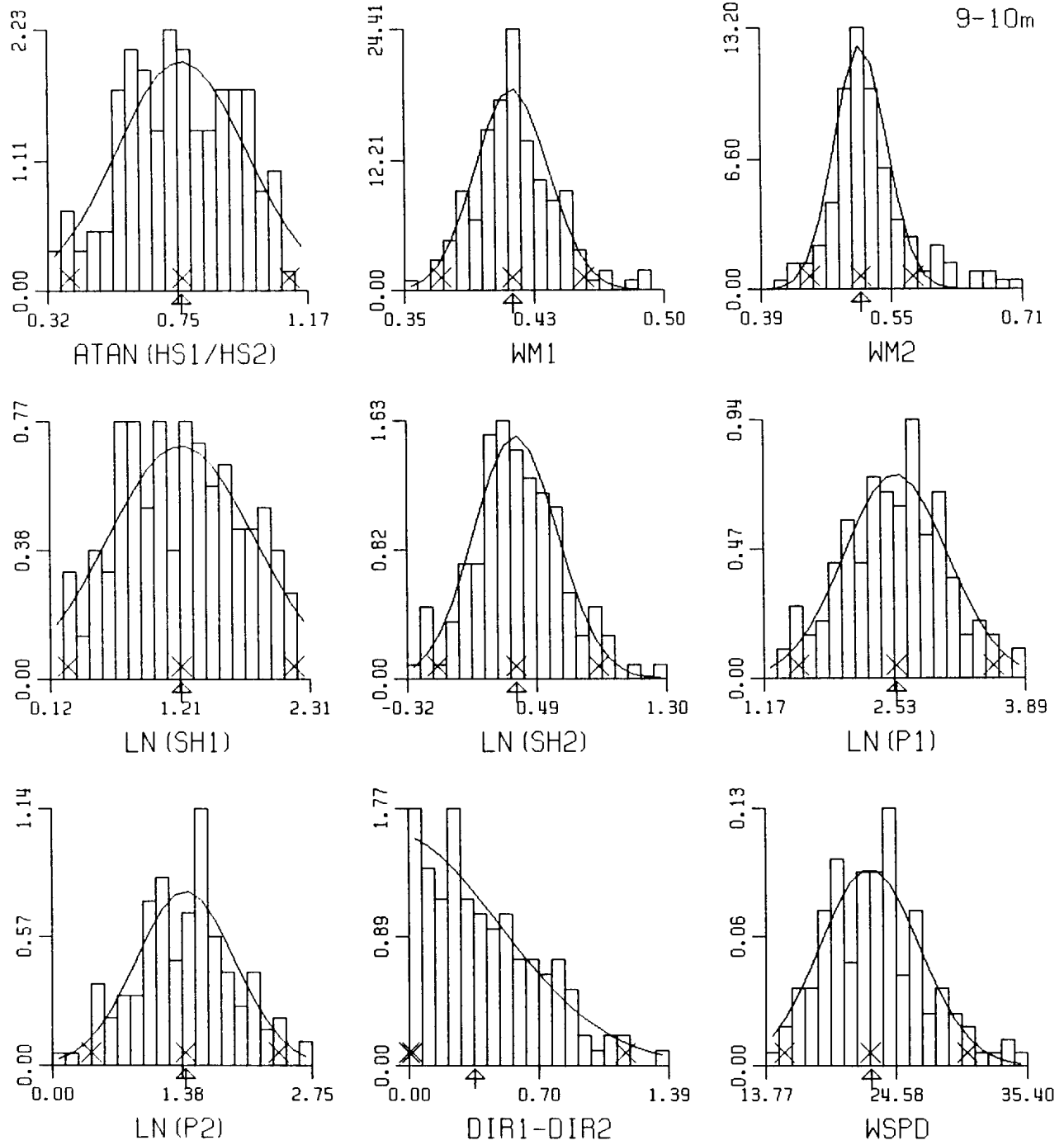
Fig. A3 (continued) HSIG: 9-10m.
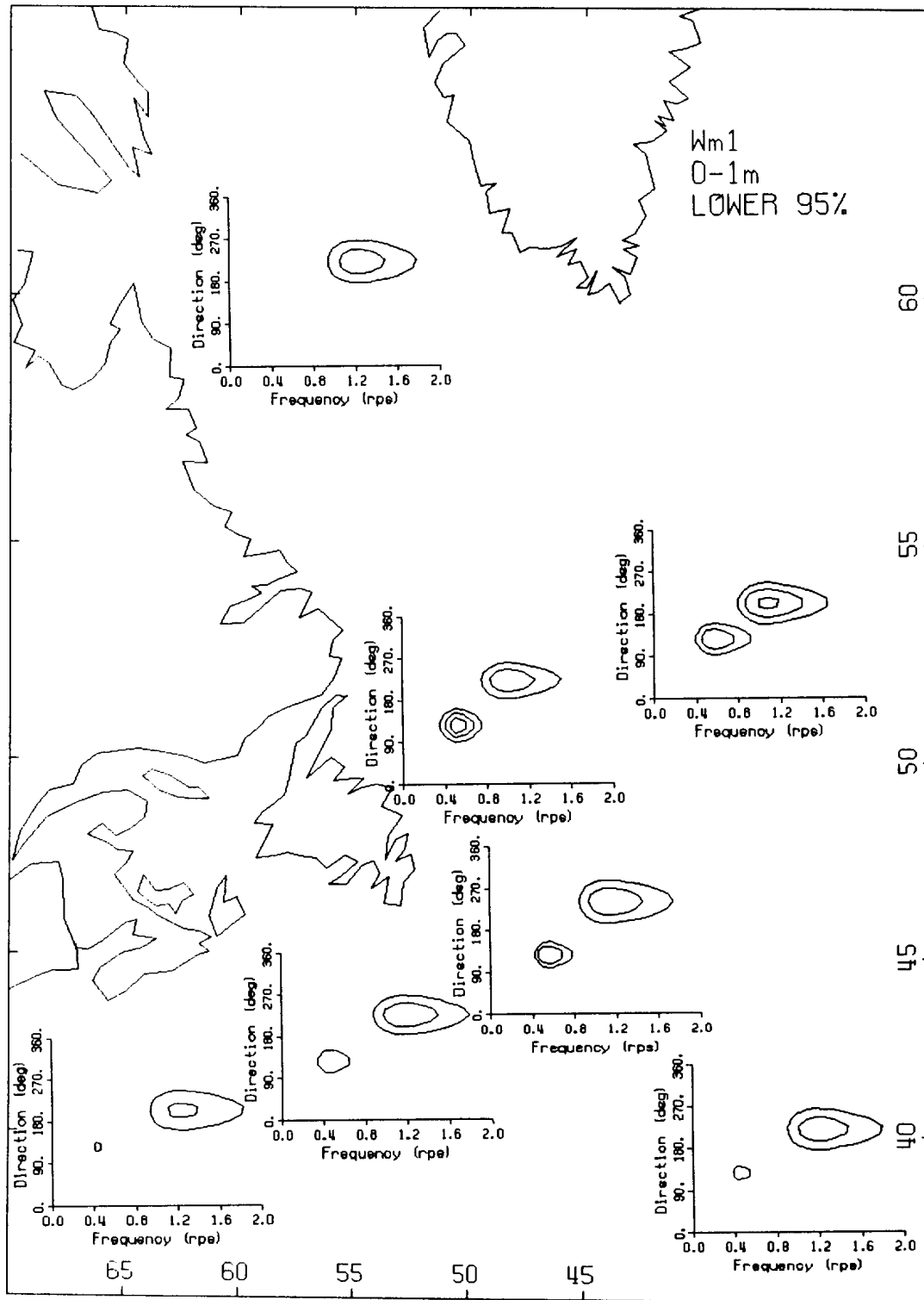
Fig. A4 Contoured 95% confidence spectra per region and overall (lower right). Contour levels set as in Fig. 15. Target parameter, confidence level and HSIG bin are noted in the upper right corner.
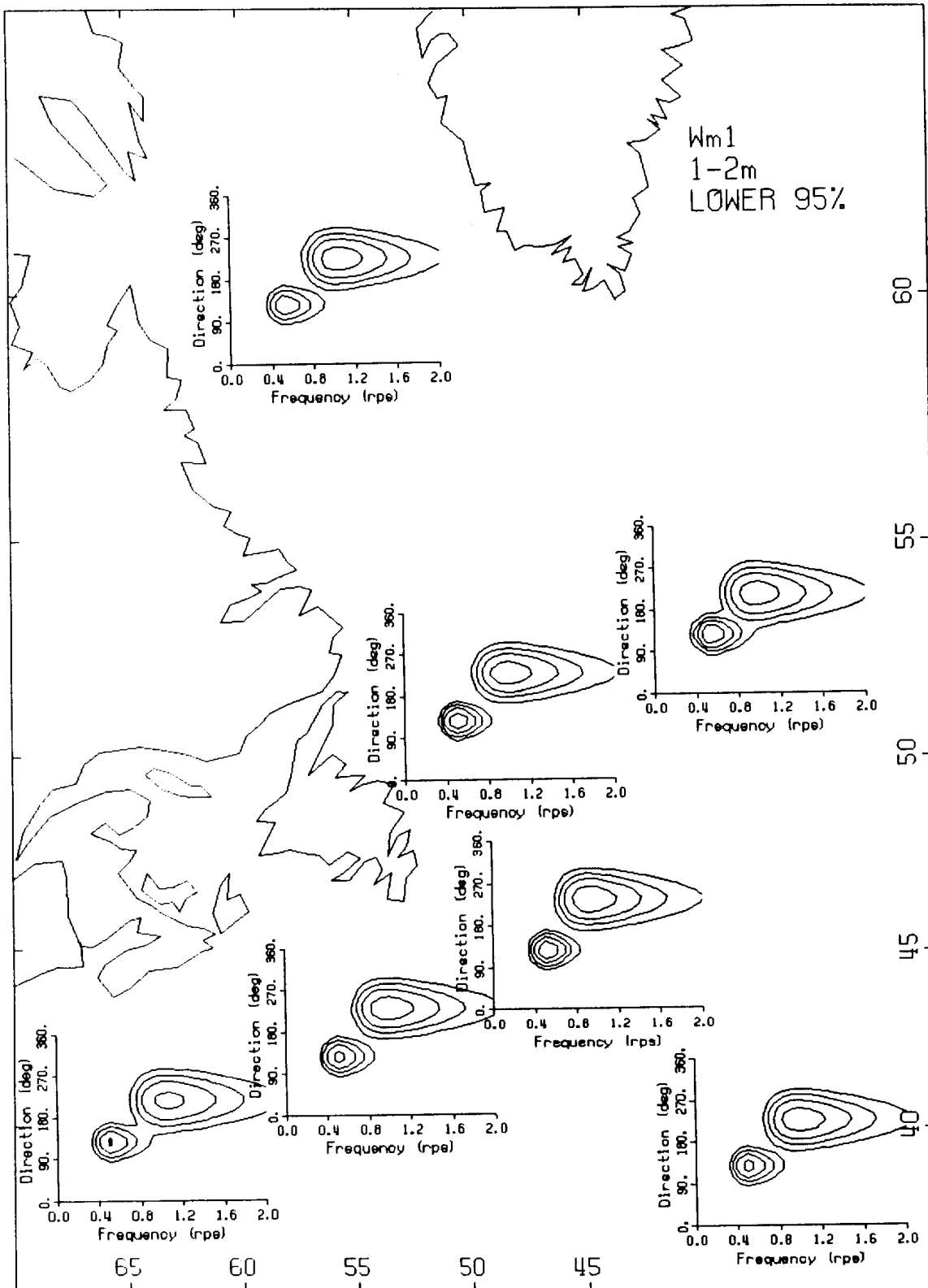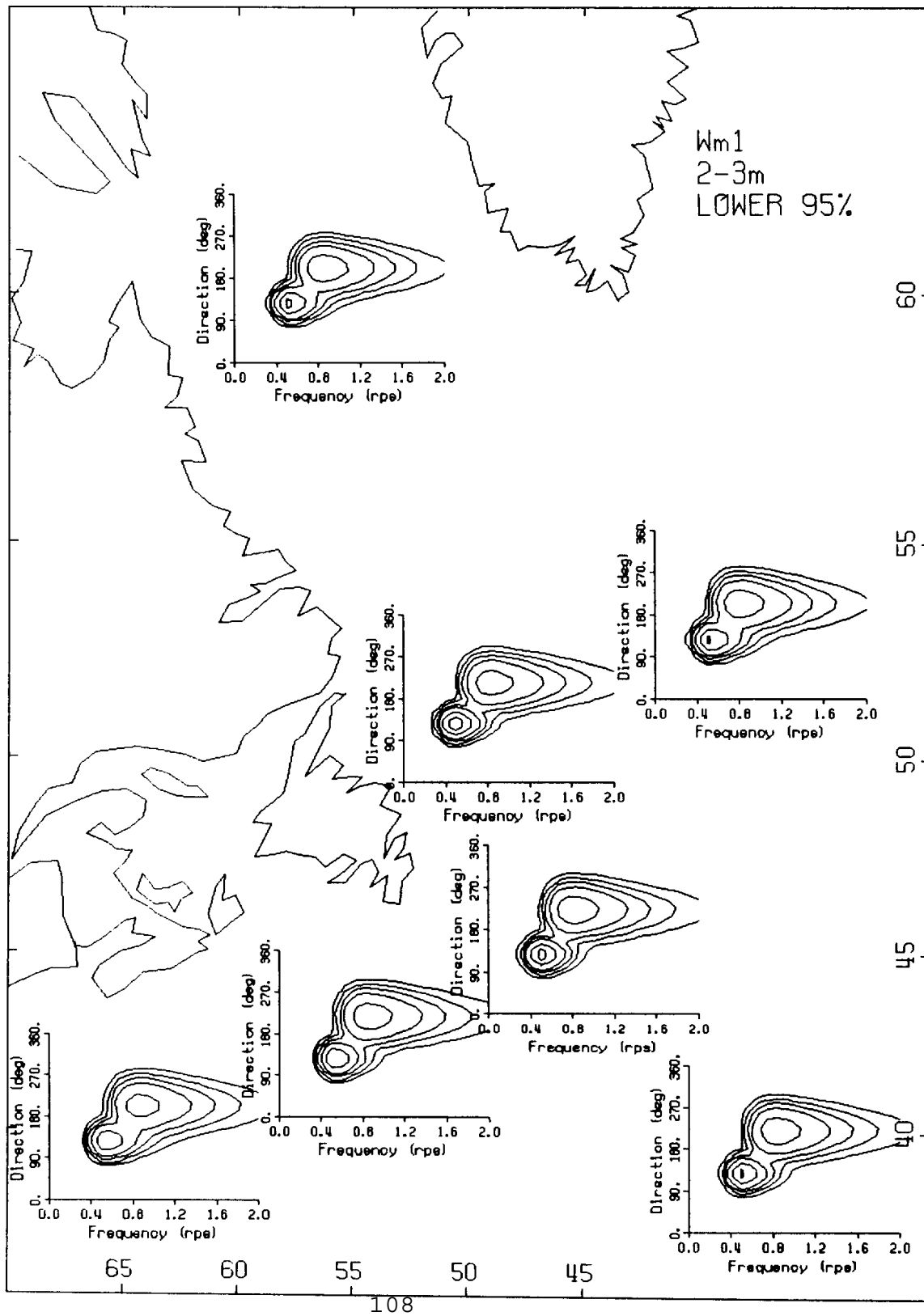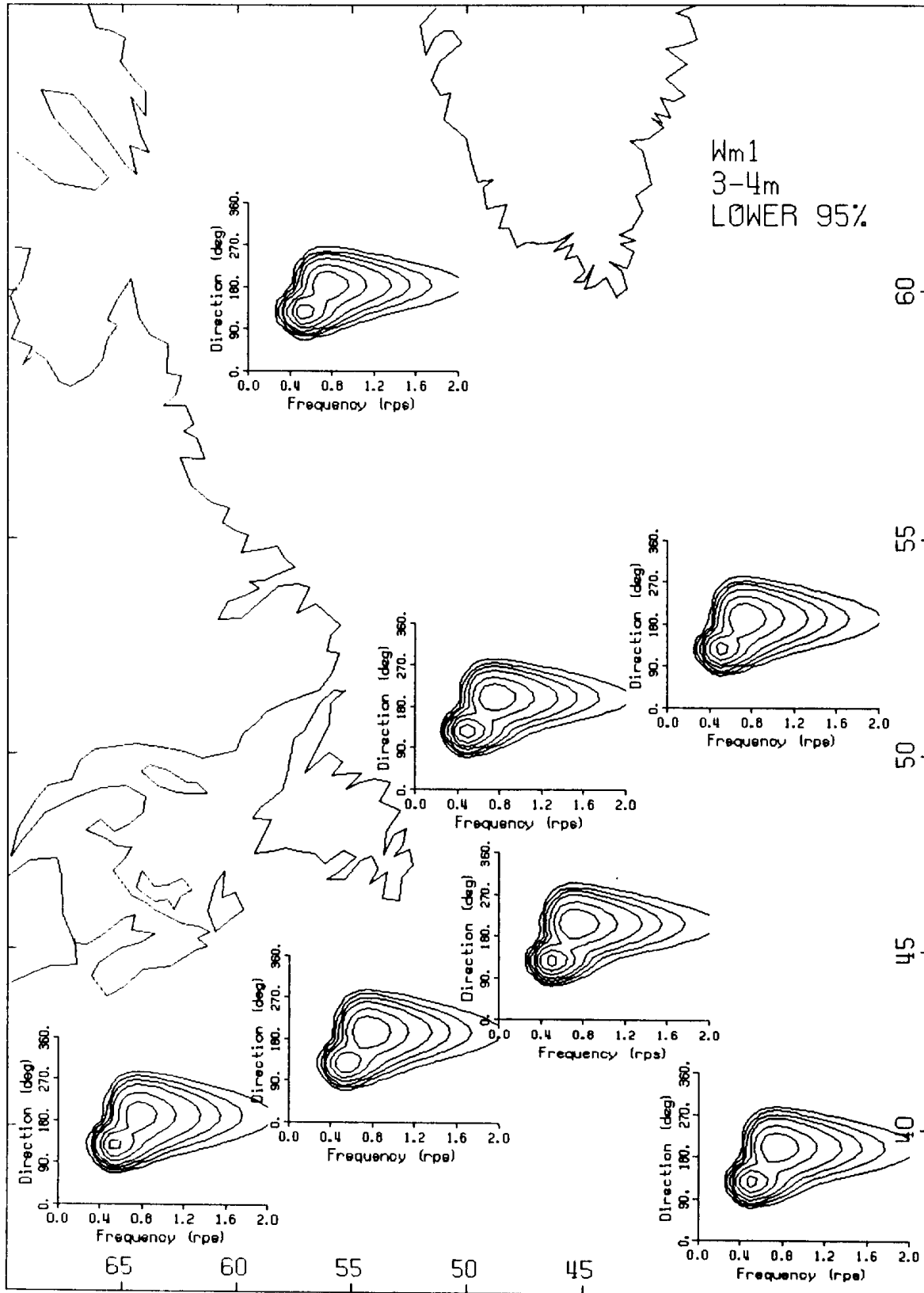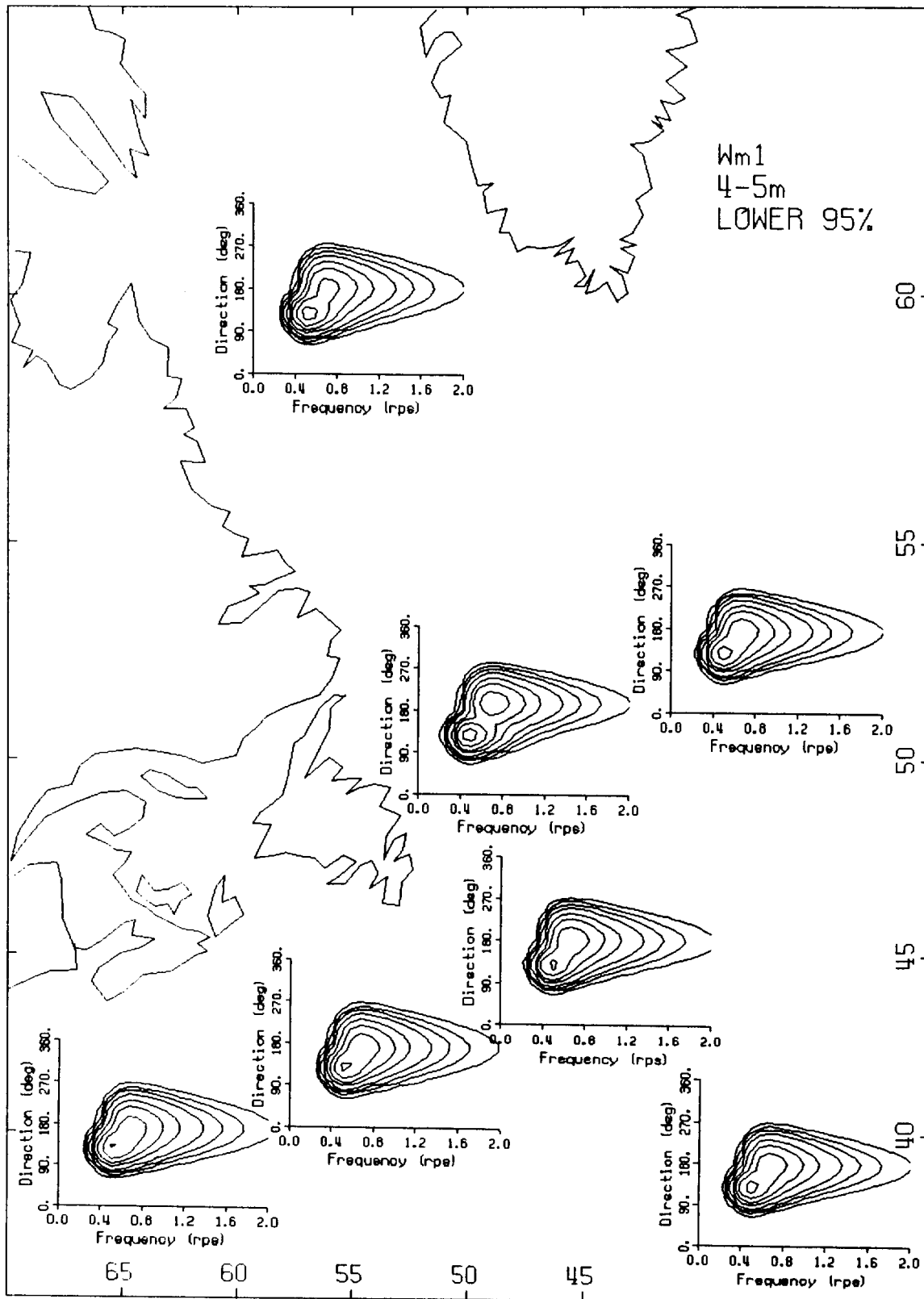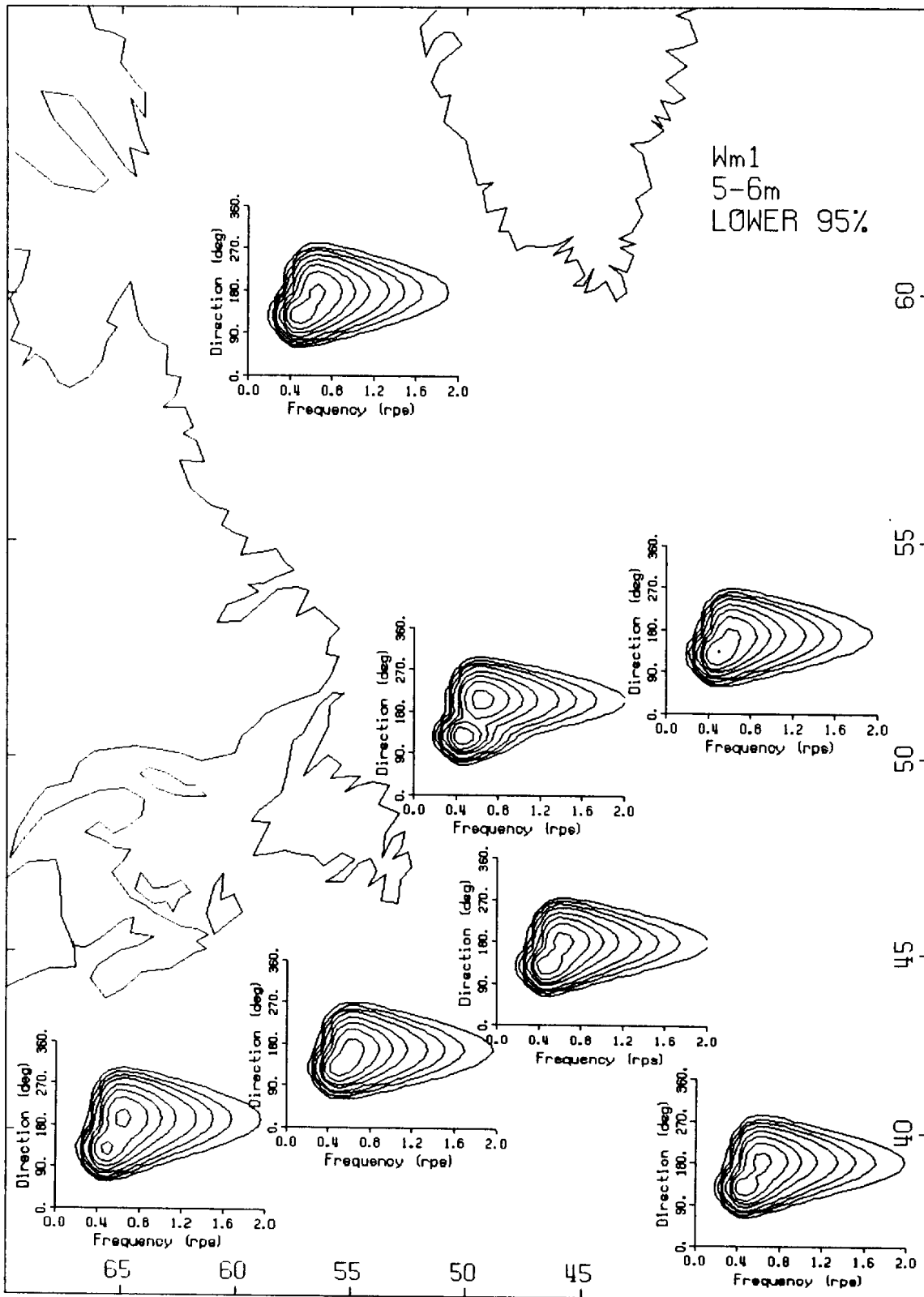
Wm1
1-2m
LOWER 95%

Fig. A4 (continued).

Wm1
2-3m
LOWER 95%

Fig. A4 (continued).

Fig. A4 (continued).

Wm1
4-5m
LOWER 95%

Fig. A4 (continued).

Wm1
5-6m
LOWER 95%

Fig. A4 (continued).

Wm1
6-7m
LOWER 95%

Fig. A4 (continued).

Wm1
7-8m
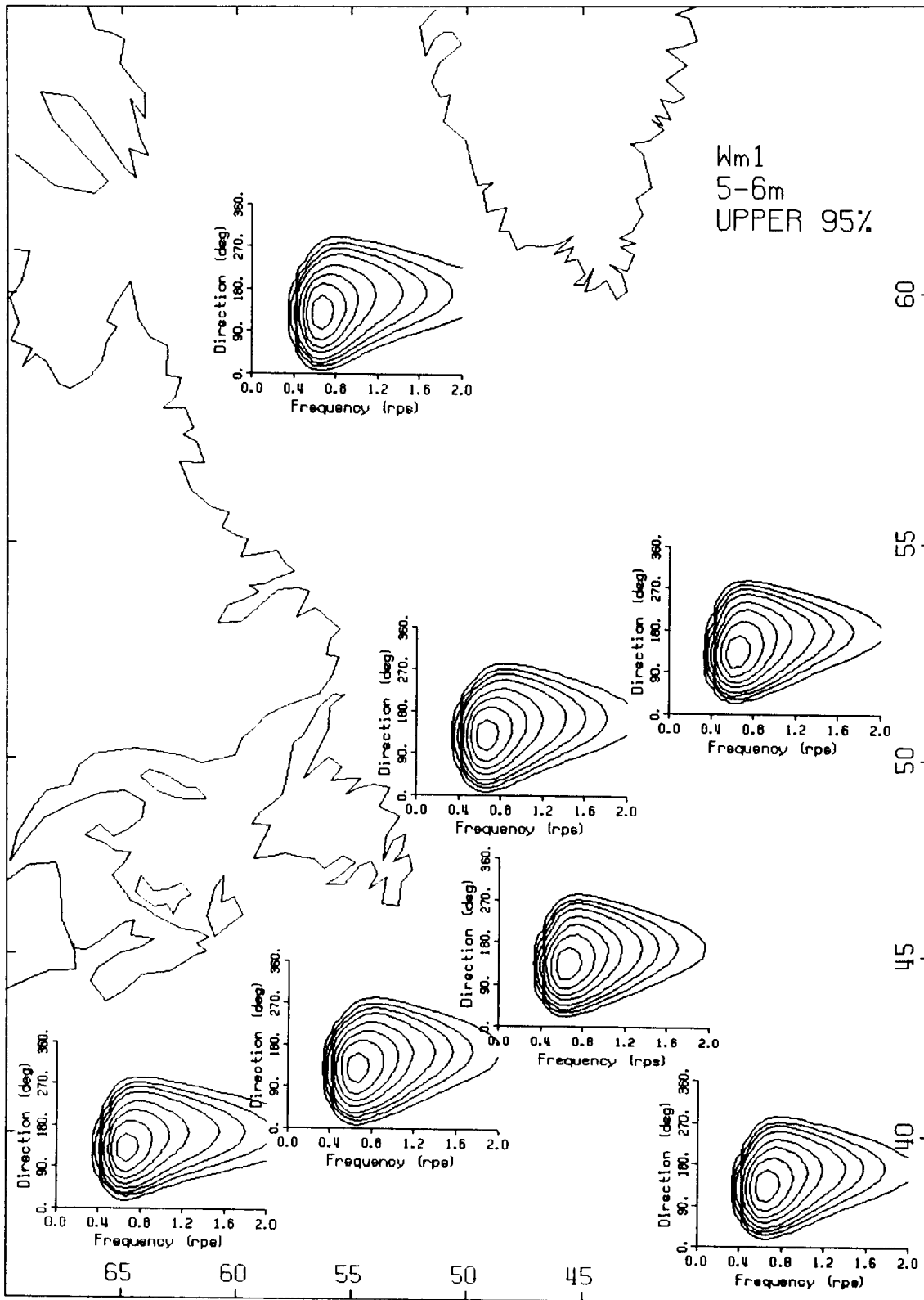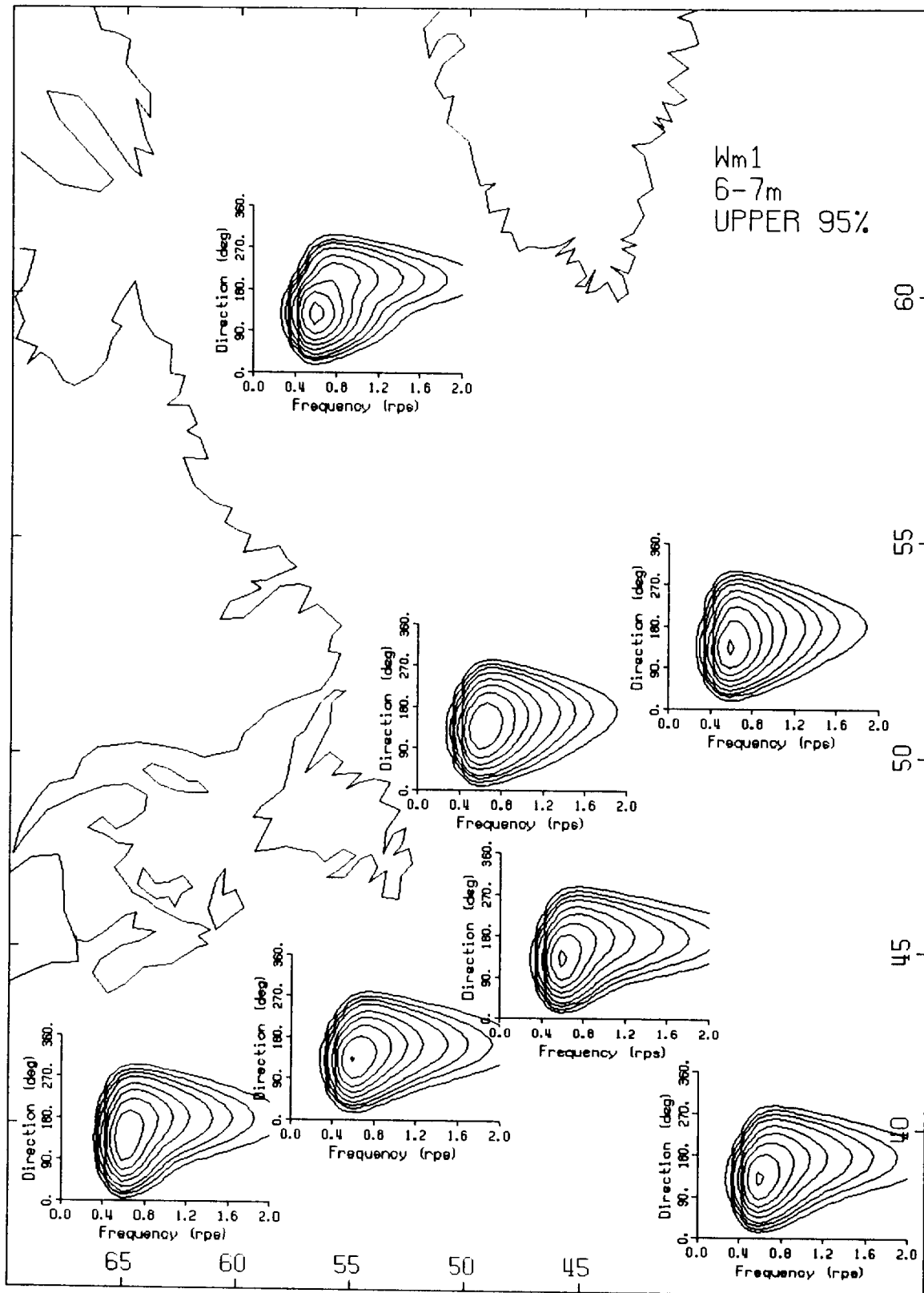LOWER 95%
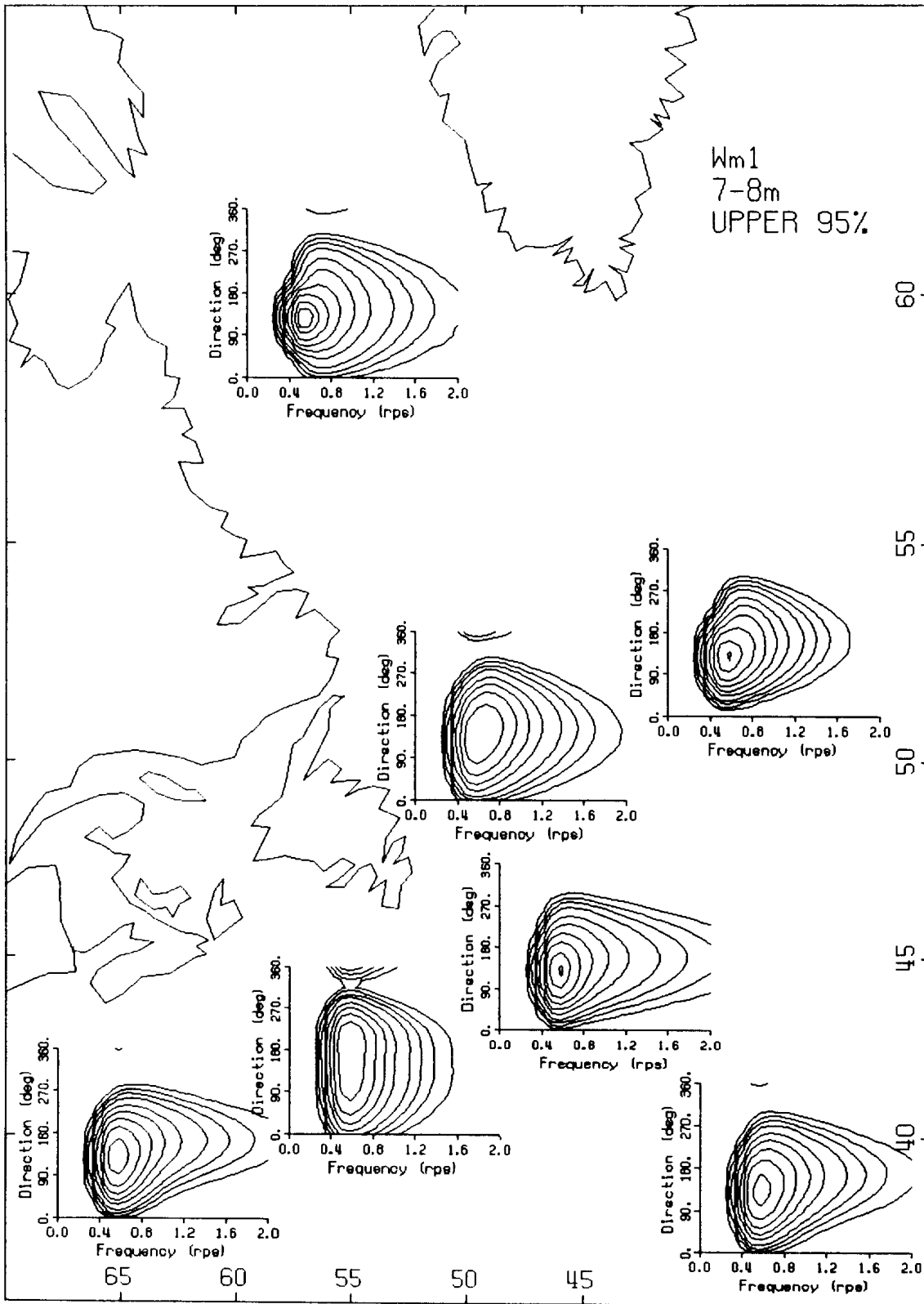
Fig. A4 (continued).

Fig. A4 (continued).

Wm1
9-10m
LOWER 95%

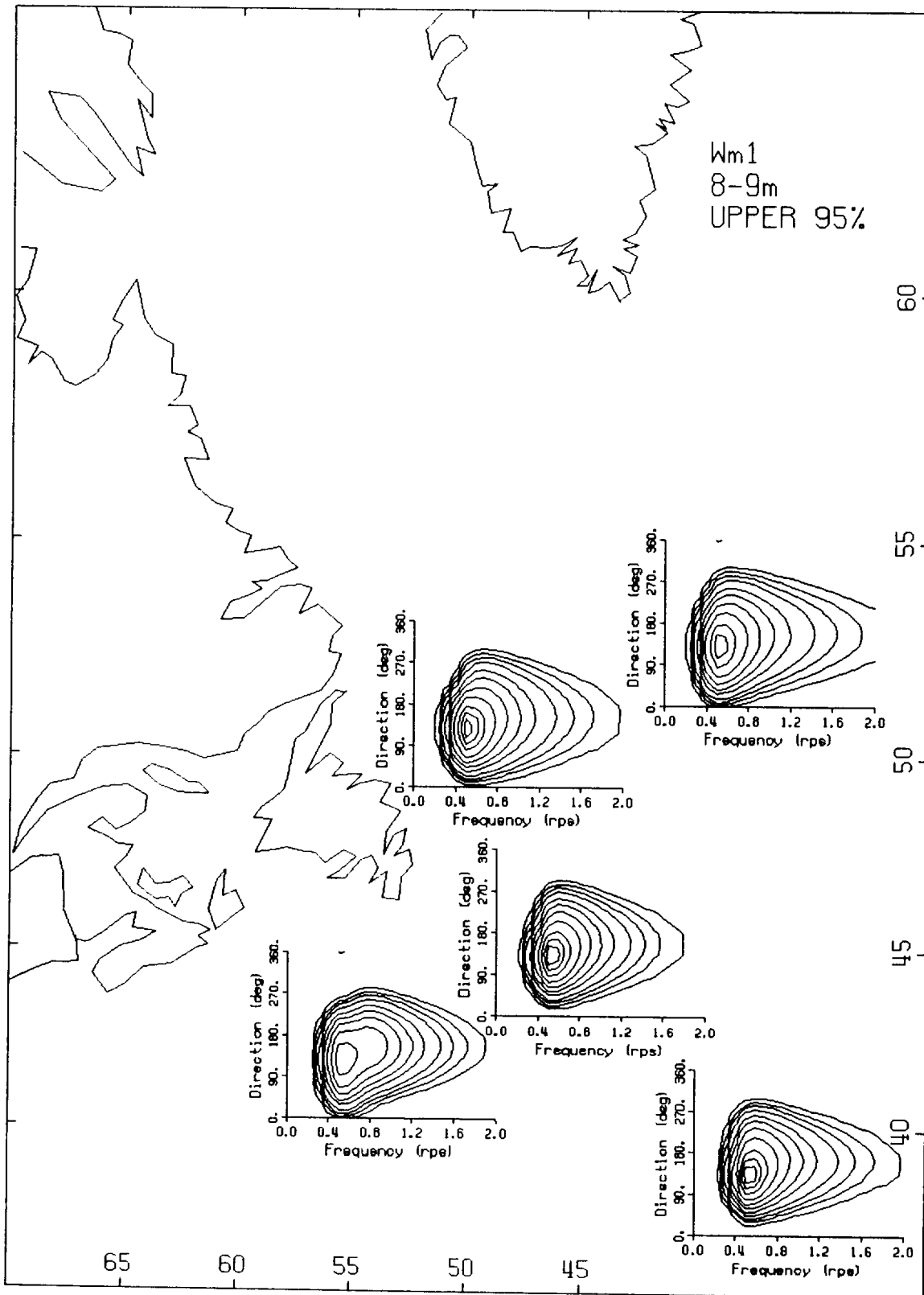Fig. A4 (continued).

Fig. A4 (continued).

Fig. A4 (continued).

Wm1
1-2m
UPPER 95%

Fig. A4 (continued).

Fig. A4 (continued).

Fig. A4 (continued).

Fig. A4 (continued).

Fig. A4 (continued).

Wm1
6-7m
UPPER 95%
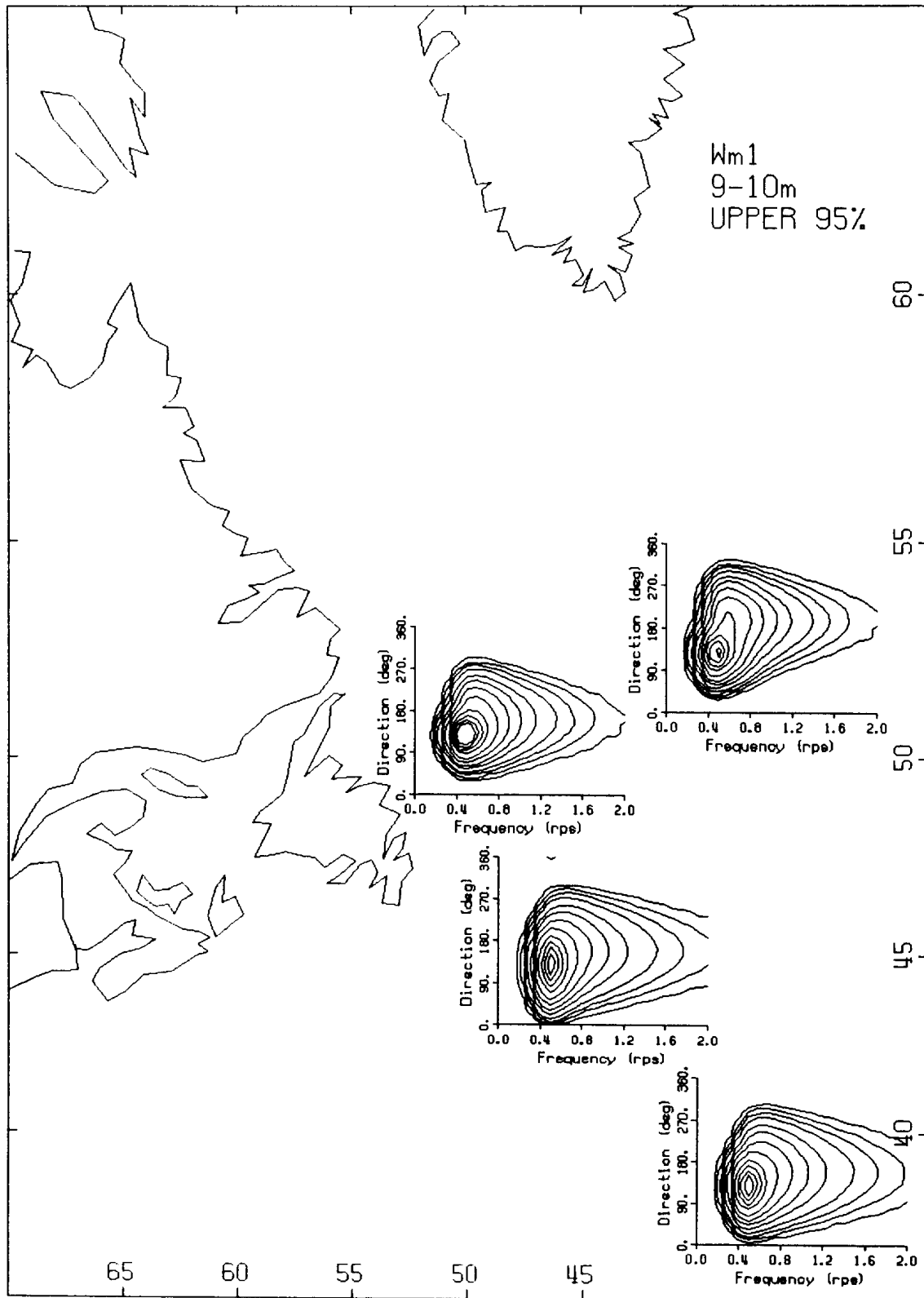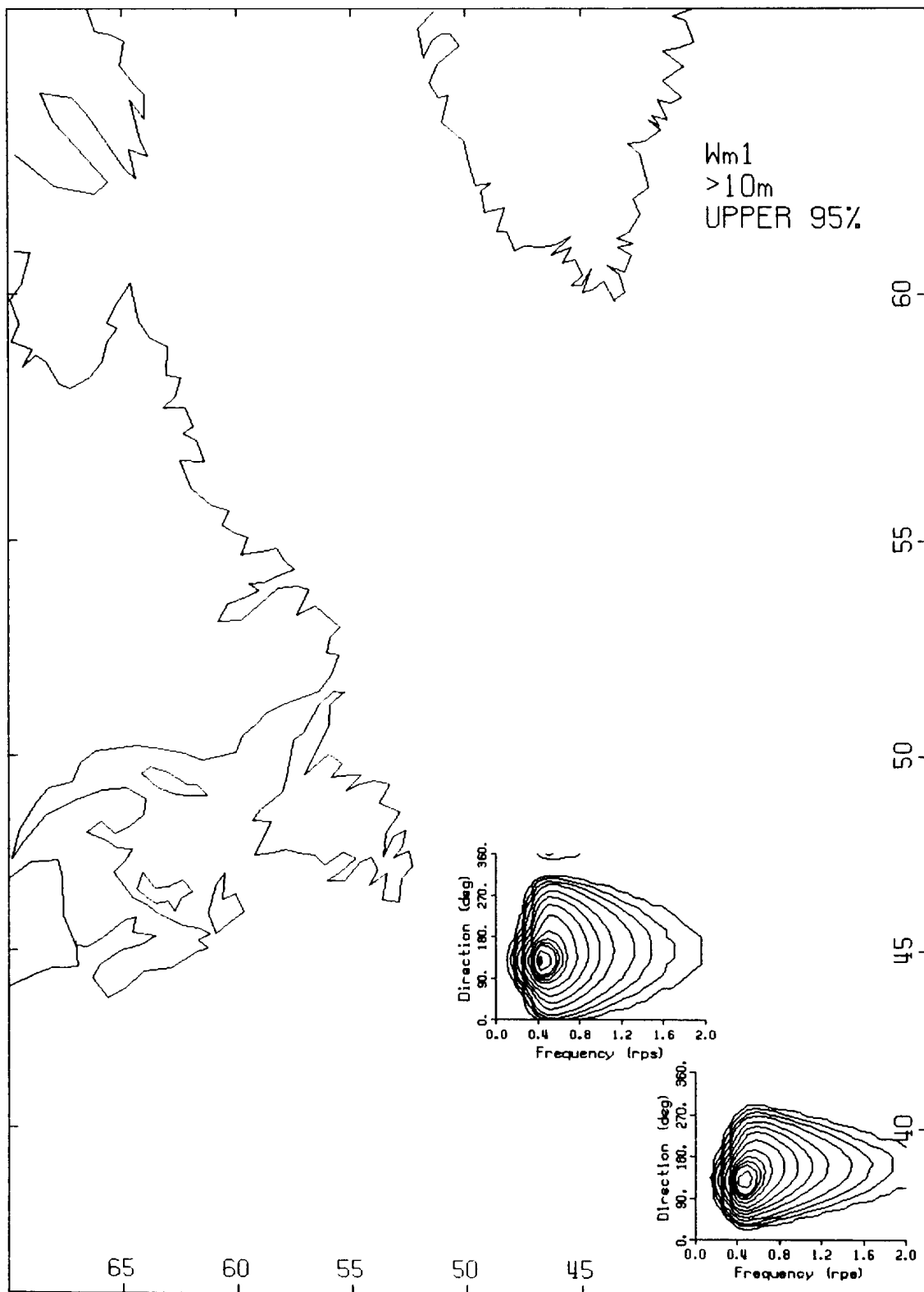
Fig. A4 (continued).

Fig. A4 (continued).

Fig. A4 (continued).

Wm1
9–10m
UPPER 95%

Fig. A4 (continued).

Wm1
>10m
UPPER 95%
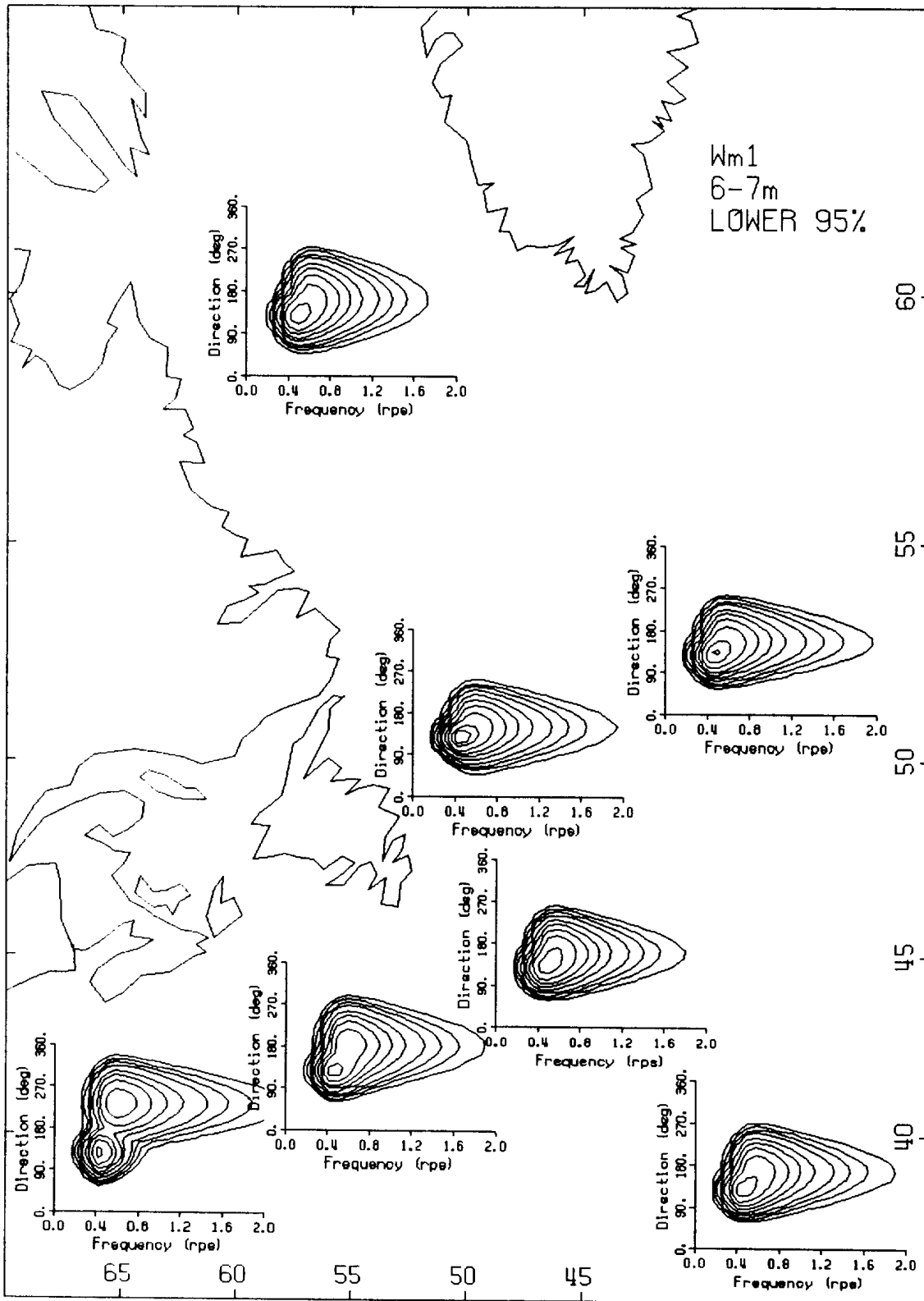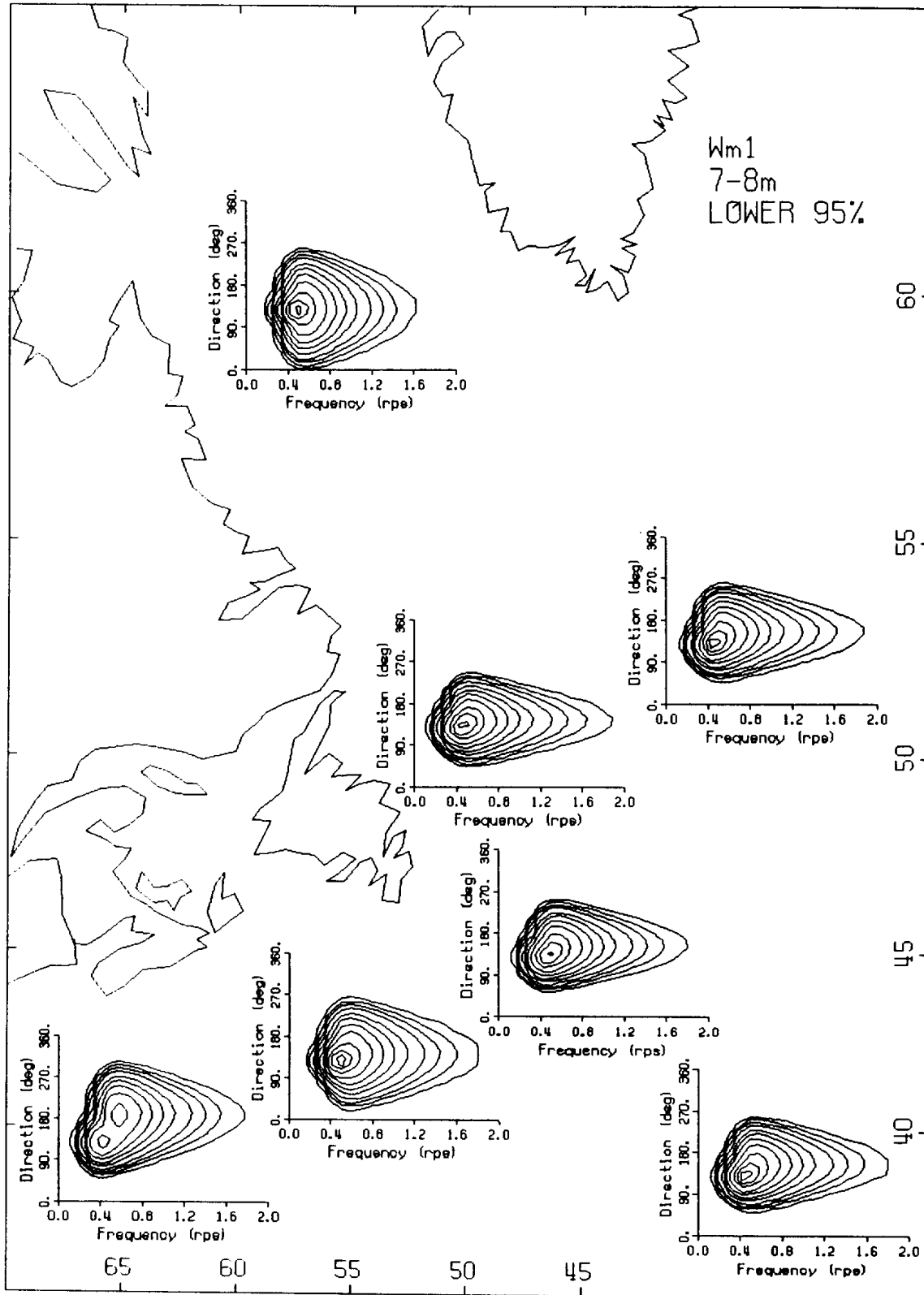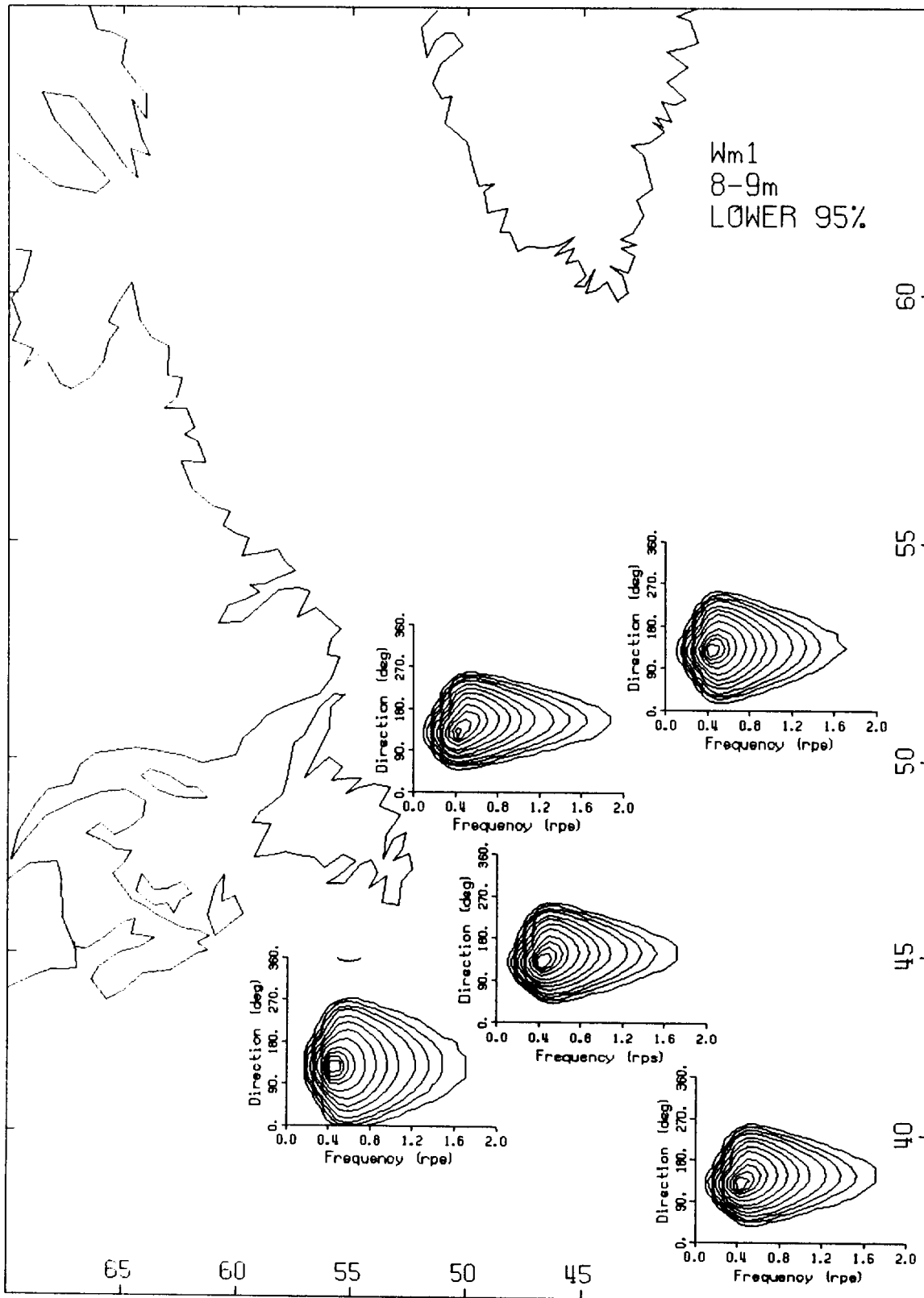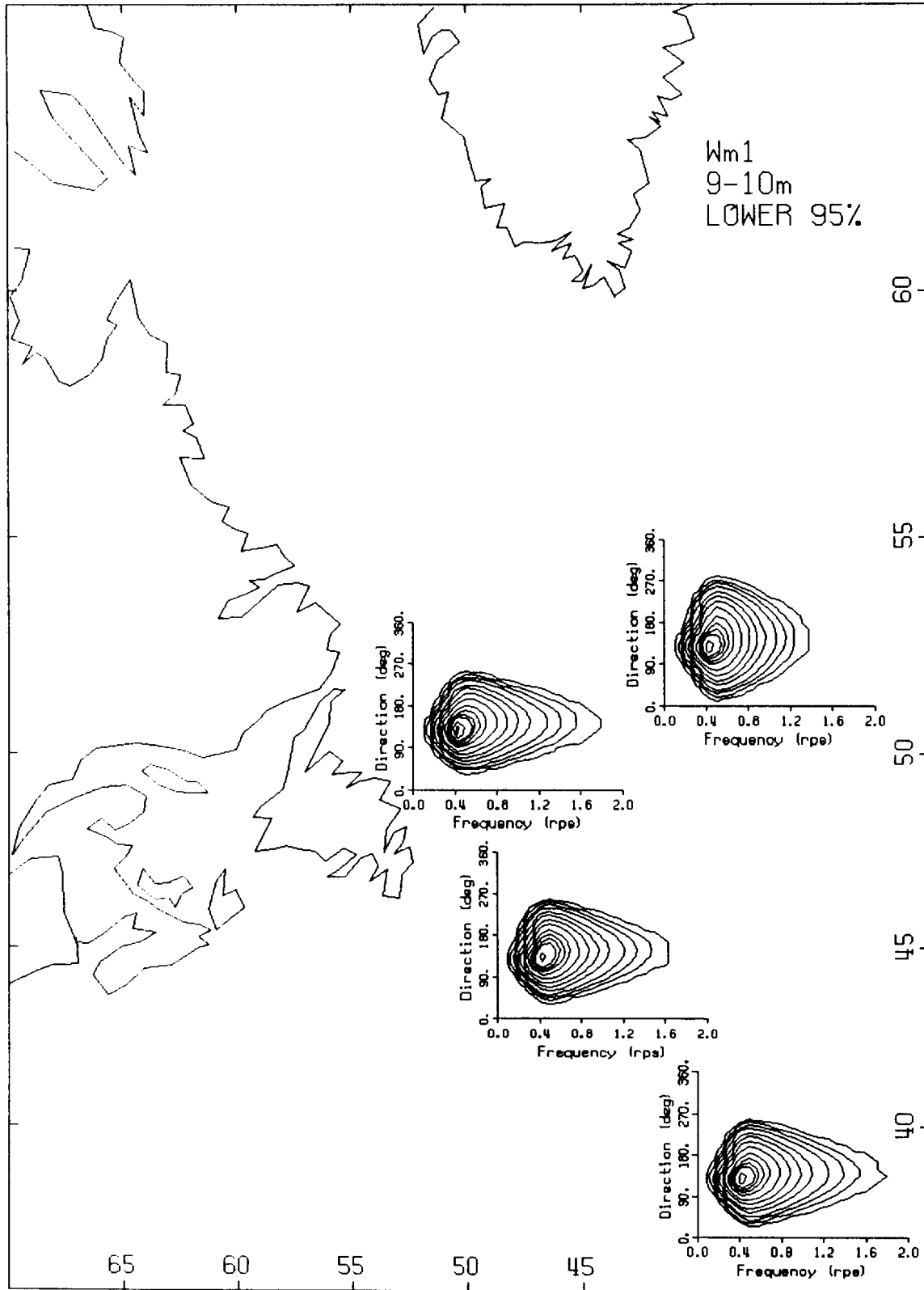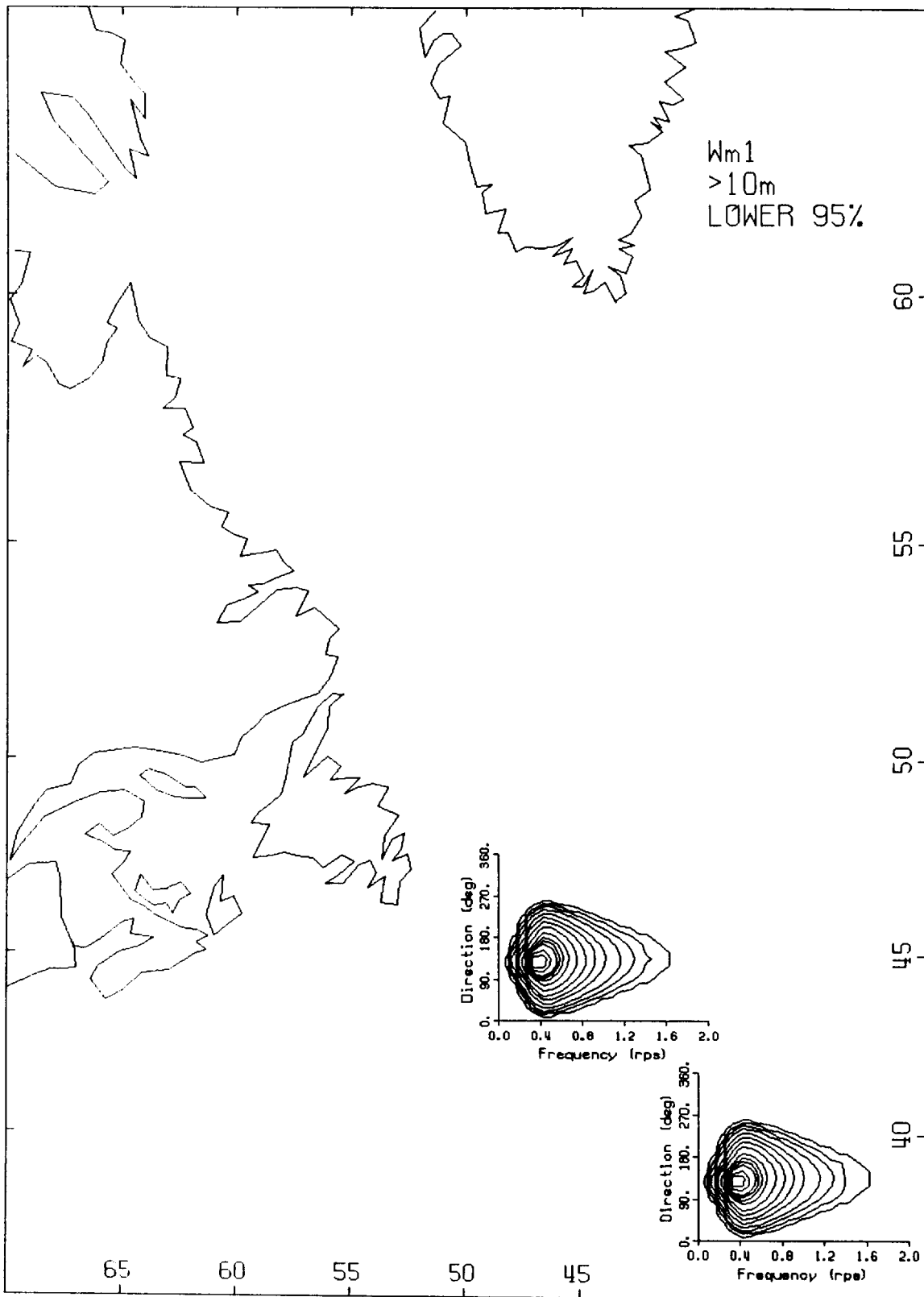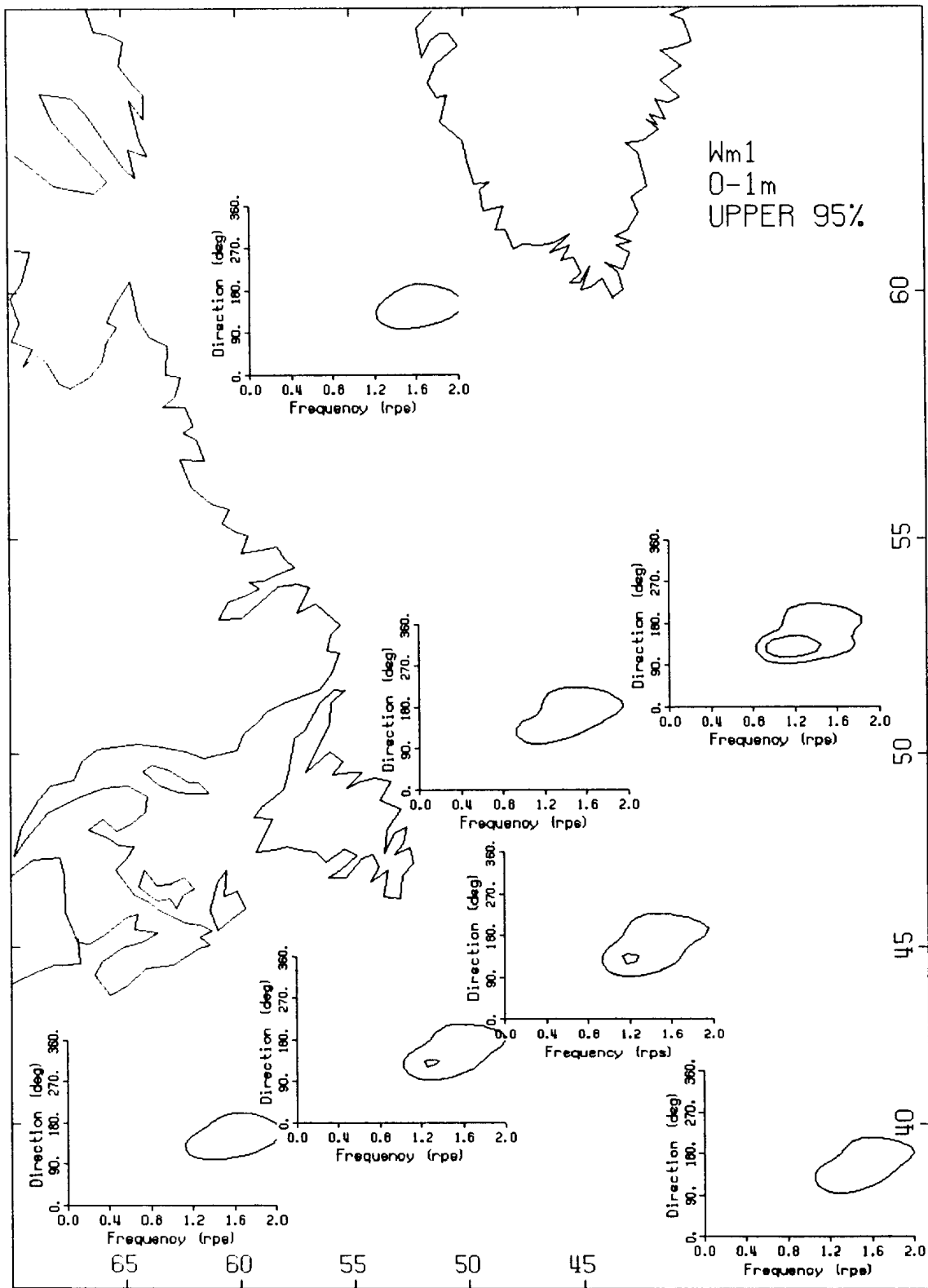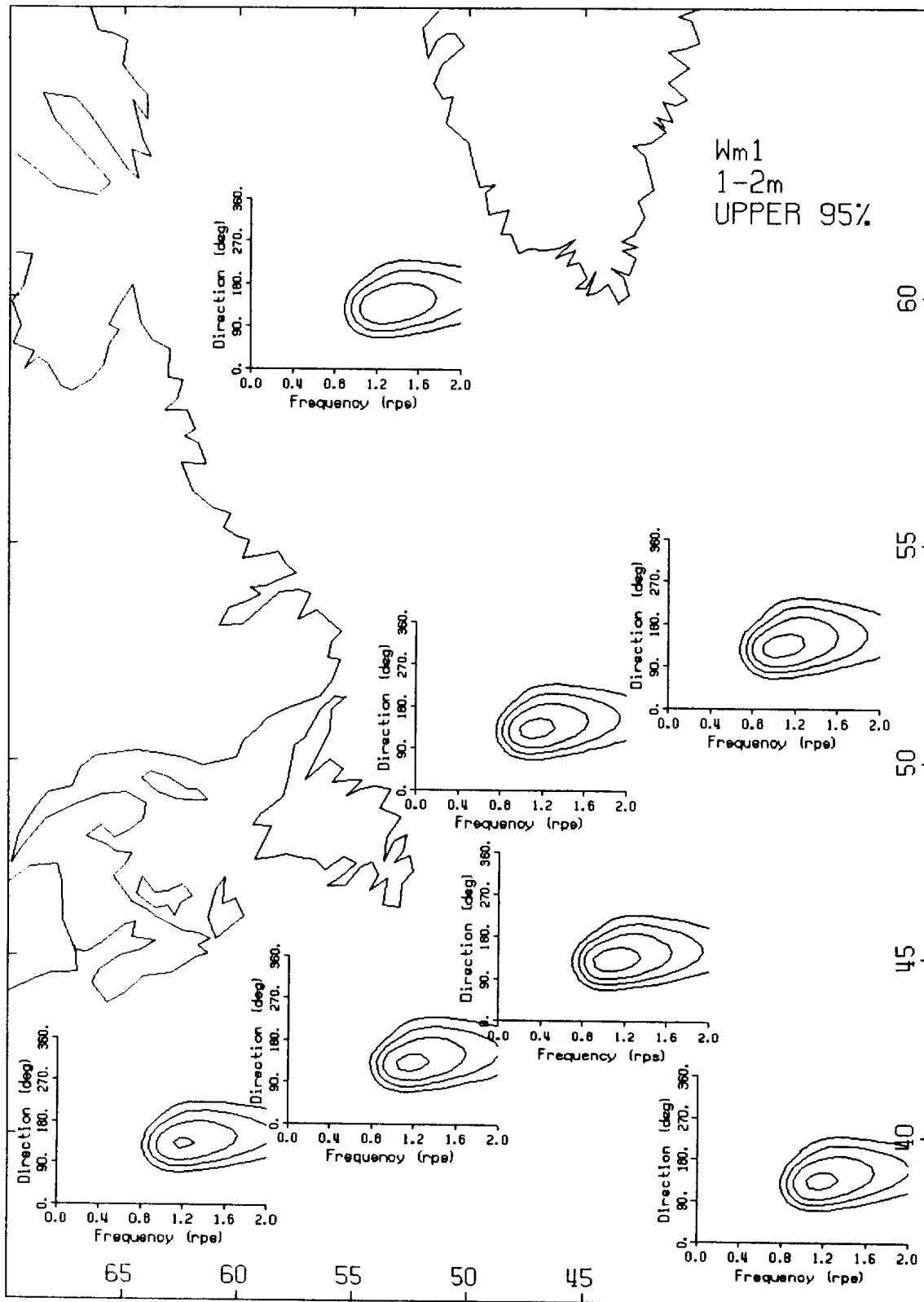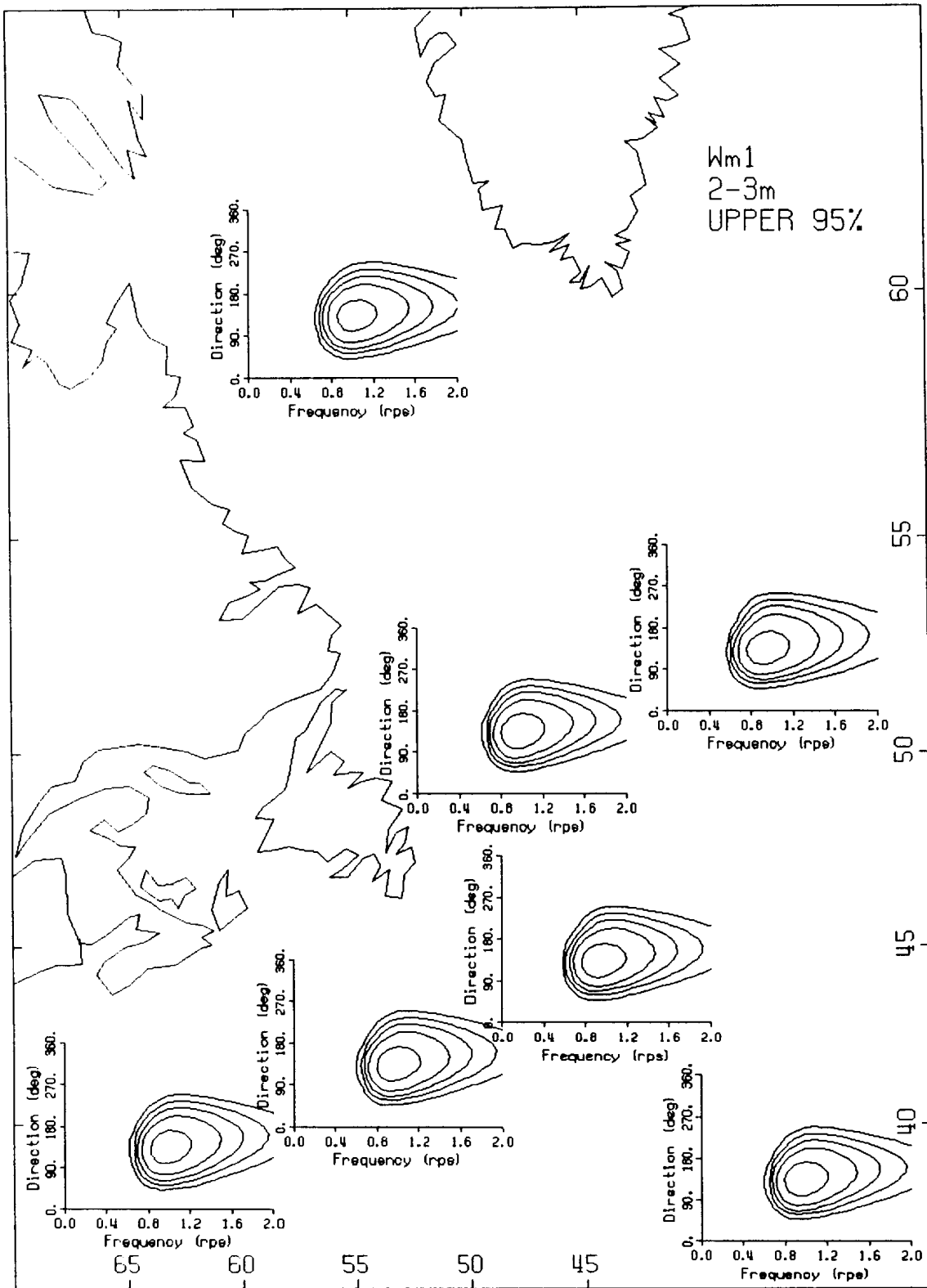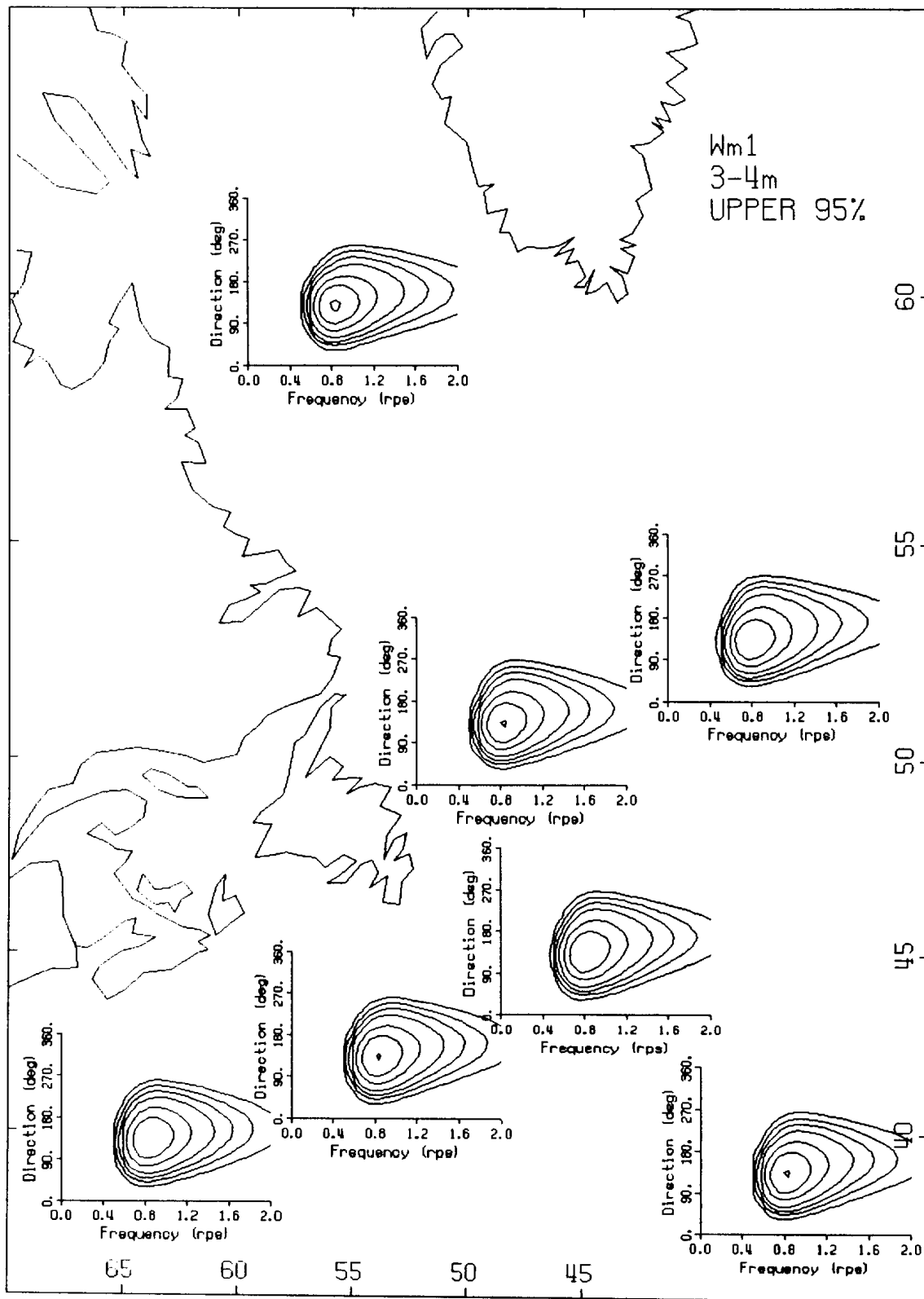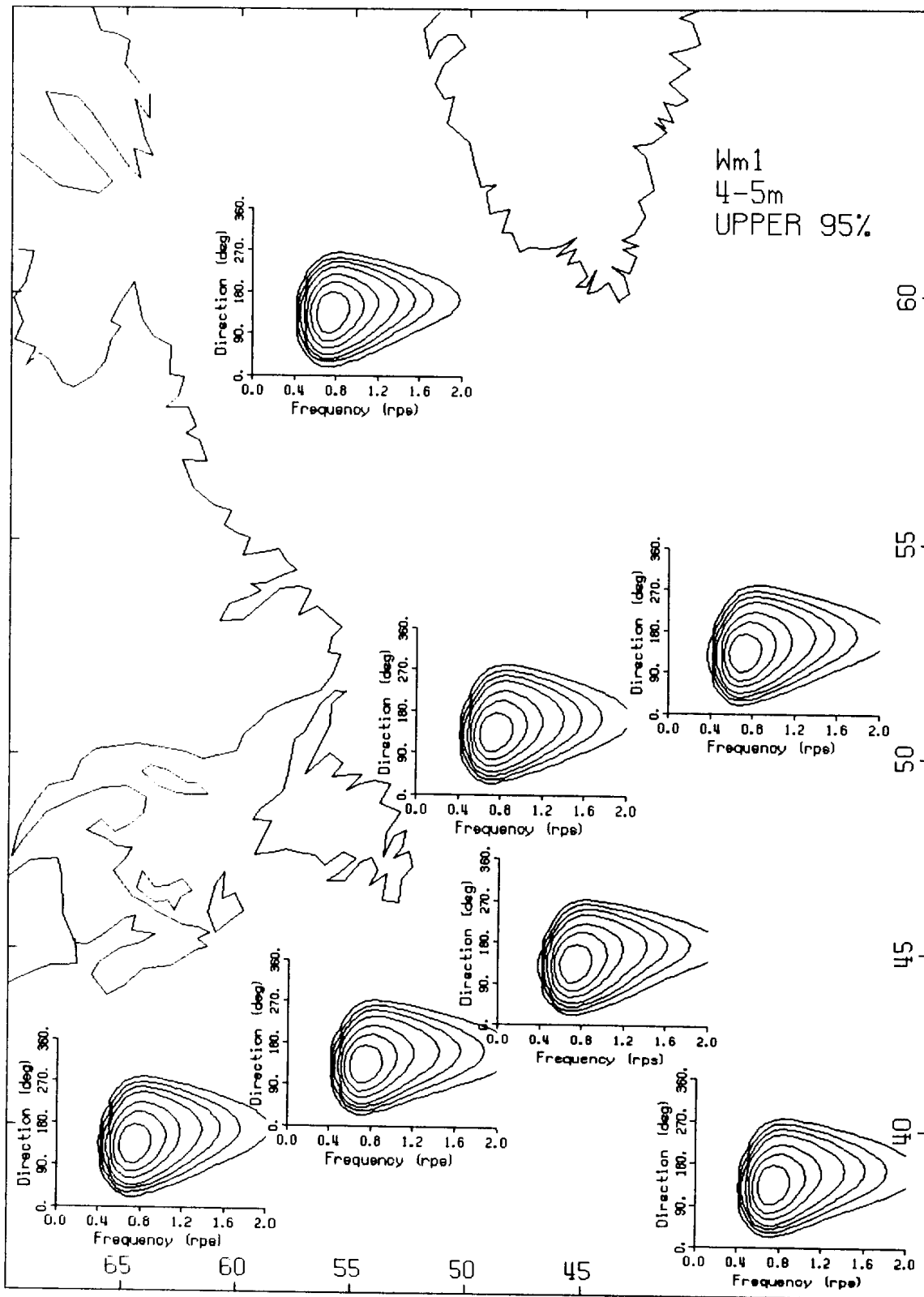
Fig. A4 (continued).

REFERENCES

Juszko B.A. 1989a. Parameterization of directional spectra – Part 1. Defence Research Establishment Atlantic Contractor Report CR/89/414, March 1989. 141pp.

Juszko B.A. 1989b. Parameterization of directional spectra – Part 2. Defence Research Establishment Atlantic Contractor Report CR/89/445, December 1989. 115pp.

Juszko B.A. 1991. Development of a statistical wave climate based on 10-parameter spectra. Vol. 1. Defence Research Establishment Atlantic Contractor Report CR/91/458, September 1991. 217pp.

Juszko, B.A. and R. Graham, 1992. Frequency and directional evaluation of the Ocean Data Gathering Program (ODGP) wave spectra at Hibernia. AtmosphereOcean 30: 441–456.

Ochi M.K. and E.N. Hubble, 1976. Six-parameter wave spectra. In the Proceedings of the 15th Coastal Engineering Conference, Honolulu, Hawaii. p 301–328.

Preisendorfer R.W.'Principal Component Analysis in Meteorology and Oceanography'. Edited by C.D. Mobley. Elsevier Science Pub. New York. 1988.

Richman M.B., 1986. Rotation of Principal Components. Journal of Climatology, Vol. 6, 293–335.

Thurstone L.L., 1947. Multiple Factor Analysis. University of Chicago Press, p360–361.

DOCUMENT CONTROL DATA

(Security classification of title, body of abstract and indexing annotation must be entered when the overall document is classified)

1. ORIGINATOR (the name and address of the organization preparing the document Organizations for whom the document was prepared, e.g. Establishment sponsoring a contractor's report, or tasking agency, are entered in section 8.) Juszko Scientific Services	2. SECURITY CLASSIFICATION (overall security classification of the document including special warning terms if applicable). UNCLASSIFIED

3. TITLE (the complete document title as indicated on the title page. Its classification should be indicated by the appropriate abbreviation (S,C,R or U) in parentheses after the title).

Statistical Description of the East Coast Directional Wave Climate using 10-Parameter Spectra

4. AUTHORS (Last name, first name, middle initial. If military, show rank, e.g. Doe, Maj. John E.)

Juszko, Barbara-Ann

5. DATE OF PUBLICATION (month and year of publication of document) March 1994	6a. NO OF PAGES (total containing information Include Annexes, Appendices, etc). 134	6b. NO. OF REFS (total cited in document) 8

7. DESCRIPTIVE NOTES (the category of the document, e.g. technical report, technical note or memorandum. If appropriate, enter the type of report, e.g. interim, progress, summary, annual or final. Give the inclusive dates when a specific reporting period is covered).

Contractor Report

8. SPONSORING ACTIVITY (the name of the department project office or laboratory sponsoring the research and development. Include the address).

Defence Research Establishment Atlantic, P.O. Box 1012, Dartmouth, Nova Scotia, Canada, B2Y 3Z7

9a. PROJECT OR GRANT NO. (if appropriate, the applicable research and development project or grant number under which the document was written. Please specify whether project or grant). 1AP	9b. CONTRACT NO. (if appropriate, the applicable number under which the document was written). W7707-3-2618
10a. ORIGINATOR'S DOCUMENT NUMBER (the official document number by which the document is identified by the originating activity. This number must be unique to this document). DREA CR 94/410	10b. OTHER DOCUMENT NOS. (Any other numbers which may be assigned this document either by the originator or by the sponsor).

11. DOCUMENT AVAILABILITY (any limitations on further dissemination of the document, other than those imposed by security classification)

(X) Unlimited distribution
( ) Distribution limited to defence departments and defence contractors; further distribution only as approved
( ) Distribution limited to defence departments and Canadian defence contractors; further distribution only as approved
( ) Distribution limited to government departments and agencies; further distribution only as approved
( ) Distribution limited to defence departments; further distribution only as approved
( ) Other (please specify):

12. DOCUMENT ANNOUNCEMENT (any limitation to the bibliographic announcement of this document. This will normally correspond to the Document Availability (11). However, where further distribution (beyond the audience specified in 11) is possible, a wider announcement audience may be selected).

13. ABSTRACT (a brief and factual summary of the document. It may also appear elsewhere in the body of the document itself. It is highly desirable that the abstract of classified documents be unclassified. Each paragraph of the abstract shall begin with an indication of the security classification of the information in the paragraph (unless the document itself is unclassified) represented as (S), (C), (R), or (U). It is not necessary to include here abstracts in both official languages unless the text is bilingual).

Ten parameter model spectra were fit by MEDS to the three years of archived ODGP directional wave spectra from 53 locations in the Western North Atlantic. There were acceptable fits for over 93% of the records. Empirical orthogonal function and factor analyses were used to identify six regions, encompassing locations having a high degree of co-variability, to allow for an assessment of regional behavior. A statistical analysis was performed on annual, winter and fall data for the individual, regionally grouped and all combined locations. This consisted of a probability analysis on the model fit parameters, the development of a family of directional spectra, having known confidence levels, and the establishment of predictive relationships between the fit parameters and significant wave height. A set of probability directional wave spectra, chosen to represent a desired spatial scale, can now be generated for any input significant wave height as required by the given application.

14. KEYWORDS, DESCRIPTORS or IDENTIFIERS (technically meaningful terms or short phrases that characterize a document and could be helpful in cataloguing the document. They should be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military project code name, geographic location may also be included. If possible keywords should be selected from a published thesaurus. e.g. Thesaurus of Engineering and Scientific Terms (TEST) and that thesaurus-identified. If it not possible to select indexing terms which are Unclassified, the classification of each should be indicated as with the title).

Waves
Ocean Waves
Wave Spectra
Hindcast
Directional Wave Spectra