HADLEY CENTRE TECHNICAL NOTE NO. 9

A LAGRANGIAN CLIMATOLOGY FOR THE NORTH ATLANTIC

by

John Methven, Paul Berrisford and Brian Hoskins

Department of Meteorology
University of Reading
PO Box 243, Earley Gate,
Reading, RG6 6BB UK
Email: J.Methven@reading.ac.uk

June 1999

Hadley Centre for Climate Prediction and Research
Meteorological Office
London Road
Bracknell
Berkshire RG12 2SY

NOTE: This paper has not been published. Permission to quote from it should be obtained from the Director of the Hadley Centre.

Contents

ABSTRACT

This atlas presents a climatology of the origin and history of air arriving at low levels for three clusters spaced along the North Atlantic stormtrack. The ECMWF Re-Analysis of the atmosphere (Mar 1979 to Feb 1994) was used to calculate over a million back trajectories, enabling quantification of measures relating to the spreading of each cluster following them backwards in time from the arrival grids on a daily basis. Statistics of these measures are collated to form a climatology for the seasonal characteristics of trajectories arriving over the three regions, pertaining not only to their average origin but also to the chances of bringing together two air-masses with very different origins. Seasons showing anomalous transport within the ERA period are objectively identified. In addition, the mean heating and moistening experienced along trajectories is shown as a function of air origin.

ACKNOWLEDGMENTS

The Lagrangian climatology was derived as part of a project funded by the UKMO Hadley Centre for Climate Prediction and Research. The trajectories used for the climatology were calculated for ECMWF Special Project "Routine Back Trajectories" and have been archived by the British Atmospheric Data Centre (BADC).

The authors would like to thank Kevin Hodges (Environmental Systems Statistics Centre, University of Reading) for allowing the use of his density estimation algorithms and for his helpful input throughout the project. Thanks to various people at the ECMWF for their help: Rex Gibson and the ERA team, John Greenaway from user support and Pauline Gabbitass, the tape librarian. Thanks also to Jeff Cole, Roger Brugge and Lois Steenman-Clark (Centre for Global Atmospheric Modelling, University of Reading) for their expertise with GRIB reading software, to Sam Pepler and Chunkey Lepine at the BADC for archiving the trajectories and Mike Bithell (BADC) for recommending the use of 3D Cartesian coordinates for the trajectory integrations.

DATASET

The back trajectories produced during the compilation of this Lagrangian climatology are available to researchers through the BADC. In addition to the North Atlantic trajectories there is a global dataset with 7200 trajectories arriving daily on a grid with a spacing of 10° in longitude, 0.1 in the sine of latitude and 0.1 in the ECMWF η coordinate. For further information on access to these datasets and their contents refer to the BADC's website (*http://www.badc.rl.ac.uk/*).

1 Introduction

The origin of air arriving at a location has a bearing on its temperature, moisture and trace chemical constituents. For steady flow one can infer the origin of air parcels simply by following streamlines upstream. However, in the atmosphere the wind is far from steady and over the North Atlantic is characterised by the passage of synoptic scale weather systems. In the past much attention has been given to the statistics of the tracks of storms which show that typically they track ENE from the eastern seaboard of North America and decay North of Scotland. As they grow and propagate the cyclones bring warm air from the South ahead of cold fronts and cold air from the North behind the cyclone centre, traditionally referred to as subtropical and polar air-masses respectively. Eulerian climatologies of the storm track concentrate on the time-averaged flow and variances from transient activity at fixed locations. In this framework the storms are manifest as transient eddy fluxes, for example a poleward heat flux. However, on longer timescales the trajectories of air parcels are influenced by a number of storms and the deformation flow between them and it becomes almost impossible to infer the origin of air using traditional air-mass concepts, storm tracks or Eulerian climatologies.

The aim of this atlas is to present a Lagrangian climatology for the North Atlantic based on particle back trajectories. These are calculated numerically using the 6 hourly ECMWF re-analysis of the winds to advect particles backwards in time from three "arrival grids" spaced across the storm track. Trajectories arrive at these grids every day at 12UT throughout the ERA-15 period. Further details are given in Section 2.

In Section 3 the density of origin (a normalised number density for the particle locations) 5 days before arrival is calculated on a seasonal basis. A climatology of air origin is compiled for the 15 year period, incorporating information on the vertical motion, heating and moisture change on each trajectory's approach. The interannual variability about the seasonal climatology is also examined.

Through the process of compiling the density of origin for a season much information on the spread of individual cluster releases is lost. In Section 4 the day to day variation for a wide variety of trajectory-based measures are presented. These include measures relating to the motion and spread of each cluster and the mean heating experienced along trajectories that arrive on any one day. The rate of mixing and reaction between air masses is strongly linked to the concentration gradients between them. Such gradients are enhanced by the stretching and folding of air-masses by the unsteady flow. Statistics relating to spread relative to the cluster mean position are insufficient to describe this structure within a cluster, so the finite time Lyapunov exponent and correlation dimension are used to characterize the rate of generation of "internal structure" and its nature.

The rate of decorrelation of cluster relative velocity along trajectories is also calculated in Section 5. In flows where strain is dominated by synoptic scale weather systems this rate is

faster than the separation rate between trajectories because the orientation of line elements changes faster than their length. These timescales also have a bearing on the regimes of validity for trajectory calculations.

This atlas should be of use to atmospheric chemists who are designing observational campaigns or need to assess the representivity of observations from a particular period, which may have anomalous transport characteristics, when building a chemical climatology. There is also a wealth of information on air-mass origins, as well as heating and moistening rates, in the derived climatology and interannual variability. Some of these statistics could provide sensitive tests of the climatology and variability of transport in GCMs. Investigations along these lines have been made with climate runs of the UKMO Unified Model (not presented here). It may also be possible to use the statistics of stretching rates directly in the parametrization of sub-grid scale chemical variability and reaction in GCMs.

2 The Trajectory Calculations

The back trajectories are obtained from the ERA data using the UGAMP offline trajectory model (Methven, 1997). The term **offline** indicates that the advecting wind fields are obtained from an external source which in this case is an analysis of the atmosphere derived by assimilating a global network of observations with one version of the ECMWF global circulation model (see Gibson *et al.*, 1997). The trajectories of **particles**[1] are integrated backward in time by interpolating these 4-D wind records to the current particle positions, in time and space, and then using an **integrator** scheme. Additionally, the values of meteorological fields can be interpolated to the particle positions and assigned as **attributes** for the particles.

The basic algorithm for the offline trajectory calculation is as follows:

1. The highest resolution analyses (spectral truncation T106 with 31 vertical levels) of vorticity, divergence and surface pressure are transformed to find the 3D velocity in spherical coordinates $(\dot{\lambda}, \dot{\phi}, \dot{\eta})$ where λ, ϕ and η describe the longitude, latitude and vertical coordinate of a particle. The vertical velocity is calculated on model half-levels from the equation of continuity with η as a general coordinate involving pressure and surface pressure (see Simmons and Burridge, 1981 and ECMWF Research Manual 2, 1988).

2. The arrival grids of clusters of trajectories are specified.

3. The next wind record is read.

[1] Here particles refer to material volumes that are much smaller than the smallest resolved scales in the wind field over the whole trajectory. These volumes stretch but both ends of the filament experience the same velocity gradient.

4. The position of each particle is integrated between the two wind records. This is achieved by interpolating the winds in space and time to the particle's position and using a 4th order Runge-Kutta integrator method which involves evaluation of the velocity at four points for every trajectory time-step (see Press *et al.*, 1992).

   The easiest way to implement this is to use a constant trajectory time-step, δt. Note that δt is generally far smaller than the interval between adjacent wind records, $\Delta t = t^+ - t^-$. All of the trajectories are integrated from t^- to t^+ before moving on to the next record so that only two time records are retained at any instant in the calculation. Note that the winds are interpolated to the particle's current position at each intermediate time:

$$\mathbf{u}(\mathbf{x}[t], t) = (1 - s)\ \mathbf{u}(\mathbf{x}[t], t^-) + s\ \mathbf{u}(\mathbf{x}[t], t^+) \tag{1}$$

   where the weight $s = (t - t^-)/\Delta t$. This method assumes that the Lagrangian decorrelation timescale for velocity is much greater than the interval $\Delta t/2$. The results in Section 5 show that this is indeed the case. The 4th order Runge-Kutta method has been used, with linear time interpolation, for trajectory calculations by many authors (e.g. Pierrehumbert and Yang, 1993; O'Neill *et al.*, 1994; Sutton *et al.*, 1994).

   For prolonged integrations a trajectory can keep within a specified accuracy with greater efficiency by using an adaptive step-size scheme. However, the benefits of an adaptive scheme are only realised if the required accuracy is such that the ratio $\Delta t/\langle \delta t \rangle \gg 1$, where $\langle \delta t \rangle$ is the average step-size along a trajectory. This is generally not the case for atmospheric trajectories since errors arising from wind field truncation dominate those from the integrator. Therefore the constant trajectory step-size is retained.

   In order to avoid inaccuracies with advection near the poles the horizontal velocity (i.e. velocity tangent to the sphere) is decomposed into 3 Cartesian components with their origin centred on the sphere (the z-axis is taken to pass through the poles). The back trajectory calculation then amounts to the Runge-Kutta integration of four independent coordinates:–

$$\frac{d}{d(-t)} \begin{pmatrix} x/a \\ y/a \\ z/a \\ \eta \end{pmatrix} = - \begin{pmatrix} -\cos\phi \sin\lambda\, \dot{\lambda} - \sin\phi \cos\lambda\, \dot{\phi} \\ \cos\phi \cos\lambda\, \dot{\lambda} - \sin\phi \sin\lambda\, \dot{\phi} \\ \cos\phi\, \dot{\phi} \\ \dot{\eta} \end{pmatrix} \tag{2}$$

5. Attributes are assigned to particles by the interpolation of selected fields by the same method as for the winds.

6. Repeat steps 3 to 5 until the trajectories reach the desired length.

7. Particle positions and attributes are output.

Fields at particle positions are obtained by cubic Lagrange interpolation in the vertical followed by bilinear interpolation in the horizontal and linear interpolation in time. The de-

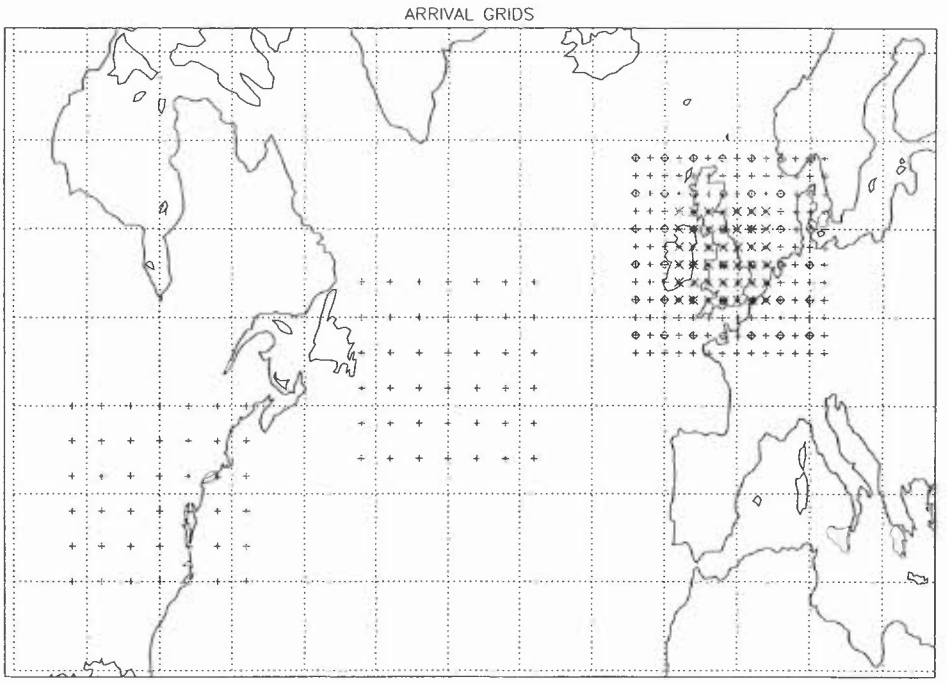


Figure 1: The arrival grids for the trajectories. The diamonds mark the trajectories used for the sparse European grid and each × marks one used for the compact grid.

tails can be found in Methven (1997) together with measurements of the errors in trajectories arising from the interpolation of discretised wind fields.

The trajectories are integrated backwards in time for 5 days from three "arrival grids" centred over Northwest Europe, the mid-Atlantic stormtrack region and the eastern seaboard of North America (see Fig. 1). The trajectories arrive daily, at 12UT, throughout the 15 year ERA period. Since the retrieval of winds from spectral analyses was the most time consuming component of the operation, efficiency was greatly enhanced by simultaneously calculating overlapping sets of back trajectories with release times which were staggered at daily intervals.

Since the intention is to build up a climatology of the origin of air affecting surface conditions in the three arrival regions, all trajectories arrive near the top of the boundary layer. A fixed value of 900hPa was chosen as the arrival level. Turbulent transport by unresolved eddies within the boundary layer is not parametrized in this trajectory model but it is assumed that such transfer is sufficiently rapid that the trajectory calculations here are relevant to surface observations of chemical constituents. The validation of the back trajectory model with trace chemical observations is discussed briefly in Section 6.3.

Note that the trajectory density on arrival at the European grid is four times that for the other two. This provides a greater density of trajectories arriving near the chemical observation sites in Europe, aiding validation studies with chemical data. It also enables tests on the sensitivity of statistics obtained in the Lagrangian climatology to the density and size

of the arrival clusters. Throughout the atlas, the "sparse" European grid refers to one with the same density as the American and mid-Atlantic clusters (the diamonds in Fig. 1) and the "compact" grid refers to the central quarter of the area of the high density grid.

3 The Origins of Trajectories

3.1 Seasonal Density of Origin

Trajectories arriving at one instant over a particular region may have disparate origins with very different chemical signatures. Moreover, trajectories in the North Atlantic storm track region vary greatly from day to day. The aim here is to quantify the probability that air arriving over each grid has originated from a given area when averaging over a season. A similar analysis was performed by Merrill (1994) who showed the probability of trajectory passage as a function of position for several years of isentropic trajectories arriving at chemical observation sites. The density of passage was found by counting all trajectory time points (separated by interval δt) falling in lat-lon grid boxes. In this atlas the density of origin is calculated using only the time points at a specified interval before trajectory arrival. The simplest method would be to count the number of trajectories originating from each lat-lon grid box. However, here we use **kernel estimation** methods which have two major advantages over grid box counting methods: there is no area bias due to box definition and the estimation is much more robust and smooth than "binning" (Silverman, 1986). The number density of particles is estimated by summing kernel functions which are placed at the position of each trajectory at a given interval (T) before arrival. Here we have used the linear kernel for density estimation on the sphere from Hodges (1996) which has the form:

$$K(\mathbf{x}.\mathbf{x}_i\,;\,C_i) = \begin{cases} \frac{2C_i}{2\pi(C_i-1)^2}\,(C_i\,\mathbf{x}.\mathbf{x}_i - 1) & : \mathbf{x}.\mathbf{x}_i \geq \frac{1}{C_i} \text{ and } C_i \geq 1 \\ 0 & : \texttt{otherwise} \end{cases} \tag{3}$$

where $\mathbf{x}_i$ is the vector from the centre of the Earth to the particle i, divided by the Earth's radius, and $\mathbf{x}$ describes any other point on the sphere. Each kernel has a finite spread which smooths the density estimate with an arc bandwidth (ABW) given by $\cos^{-1}(1/C_i)$. A density estimate is then found by summing the kernels at particle positions:

$$\rho(\mathbf{x}) = \frac{1}{n}\sum_{i=1}^{n} K(\mathbf{x}.\mathbf{x}_i\,;\,C_i). \tag{4}$$

Here adaptive smoothing has been used such that the kernel width is smaller where the density from a pilot estimate is high. The pilot kernel's width is determined by a maximum likelihood estimation which aims to keep both the bias and variance of points about the estimate to a minimum. All the estimates shown here have been made using a pilot ABW

of 5.72° and a sensitivity parameter for the adaptive smoothing of 0.5 (β in Section 2e of Hodges, 1996). The procedure is similar to finding a weighted average of a set of points, where a larger kernel spread is used where uncertainty is greater. The integral of the density over all solid angles equals one.

The first set of figures in this atlas show the density of origin, at 5 days before arrival, for each year in the ERA period. These distributions have been calculated using all trajectories which arrive in each cluster, daily, amounting to about 3800 trajectories per season for the American and mid-Atlantic clusters and four times as many for the European cluster. Note that for the European cluster a normalised density of $1.0\,\mathrm{ster}^{-1}$ corresponds to a number density of approximately three particles per $1° \times 1°$ grid box at 50°N. All of the contours in the density field enclose high values which indicates that the trajectory origins tend to clump together (refer to Section 4.1.4). In most seasons there is one main maximum, although occasionally there are two or more indicating either changes in transport regime throughout the season or that different corners of the arrival grid typically have very different origins for that season. In general it appears that there is a greater difference in the density of origin between the four seasons than there is interannual variability. In other words, individual years do not generally deviate strongly from a climatology. In the following two sections a climatology is derived and then an objective measure of the deviation from the climatology investigated.

3.2 A Climatology

The fifteen years of ERA have been used to compile a climatology of the origin of air arriving in each cluster for each season, amounting to 12 climatologies. This is simply obtained by averaging the density of origin for all the years. The particle density is shown for both 5 days and 2 days before arrival.

The maximum density of origin for all clusters shifts to the WNW from summer to winter. The two day origin is obviously much more compact than the five day origin indicating that trajectories do not tend to recirculate in a region close to arrival. The importance of the westerlies in winter is partly due to the greater vertical motion then (see below). In summer, the American cluster spreads predominantly to the North and South and the rate of spread is slow compared to other seasons and clusters, resulting in large maximum densities. The spread is also much greater in winter than summer for the European cluster. There is a notable split in the density of origin around Greenland. Trajectories commonly come from the NW (Greenland Sea) in spring, but from the W (Hudson Strait) branch in winter. Autumn shows equal probability for originating from either side. Indeed, all clusters show that SON is typically more similar to DJF than MAM is. The $T = 2$ densities of origin shown here for the European cluster in winter are similar to density of trajectory passage[2] shown

[2]The density of trajectory passage without the "geometrical correction factor" used in that paper

in Merrill (1994) for trajectories arriving at Mace Head, Ireland. The density of trajectory passage is effectively an average over $0 < T < 10$ days of the densities of origin.

In addition, the mean pressure at the origin of trajectories has been estimated using the kernel regression method:

$$\overline{p}(\mathbf{x}) = \frac{1}{n\rho(\mathbf{x})} \sum_{i=1}^{n} p_i \, K(\mathbf{x}.\mathbf{x}_i \, ; \, C_i) \tag{5}$$

where p_i denotes the pressure at each particle location. Similarly, the standard deviation in pressure at the trajectory origins is given by:

$$\sigma_p(\mathbf{x}) = \left\{ \frac{1}{n\rho(\mathbf{x})} \sum_{i=1}^{n} [p_i - \overline{p}(\mathbf{x})]^2 \; K(\mathbf{x}.\mathbf{x}_i \, ; \, C_i) \right\}^{\frac{1}{2}} . \tag{6}$$

The mean and standard deviation are calculated for each season and then combined to produce the climatology by the method given in Section 2f of Hodges (1996) . They are plotted for the 5 day origin where the density exceeds a threshold value of $0.5\,\mathsf{ster}^{-1}$. Note that the climatological mean pressure amounts to the density weighted mean of the pressure estimate for each season.

The general characteristics of trajectories arriving at 900hPa are that they originate in the mid troposphere when coming from the North but travel close to the ground when coming from the South. This arises because most trajectories are almost adiabatic so that air descends down isentropic surfaces (constant potential temperature) from the polar regions behind cyclones, but approximately follows the northward excursion of temperature contours near the ground in the warm sector of cyclones.

The strongest gradient in $\overline{p}$ is seen for the American cluster in MAM and SON and is oriented from NW to SE. Trajectories originating from the Caribbean travel almost horizontally while those originating from Alaska come from almost 600hPa on average. Remember that in the dark blue area density is below the threshold and mean and standard deviations are not estimated; $\overline{p}$ does not drop below 600hPa for any cluster.

The standard deviation in pressure is lower when $\overline{p}$ is close to the ground. For instance, for the mid-Atlantic cluster in summer $\overline{p} < 900$hPa for trajectories from the South and the standard deviation is less than 80hPa. In cases where $\overline{p}$ is lower than 650hPa, the standard deviation is about 130hPa. The standard deviation can be even higher in neighbouring regions where trajectories come from a wide range of levels (e.g. SON, American cluster and DJF, European cluster). The standard deviation also drops to 90hPa where trajectories make it over the U.S. Rockies during DJF because their mean pressure of 700-725hPa is not far above ground.

It is interesting to examine the statistics of trajectories which do not follow isentropic surfaces. Indeed, the rate of change of potential temperature following trajectories is equal to the heating rate experienced by the resolved atmospheric fields in the GCM. Typically heating occurs through radiative transfer, latent heat release from condensation of water vapour, and fluxes of sensible and latent heat transferred from the surface by the boundary layer. The analyses of temperature incorporate the effects of these processes through the assimilation of observations in regions of heating as well as through parametrization of these processes in the ECMWF model. The distribution of heating has an impact on the flow and the maintenance of the stormtrack itself (see Valdes and Hoskins, 1989).

The time integrated effect of the heating experienced following trajectories over 5 days before arrival is calculated from their potential temperature attributes as follows:

$$\theta_i' = \theta_i[-T] - \theta_i[0]. \tag{7}$$

Note that the trajectory length T is 5 days and that the sign of the θ-change for back trajectories is such that *negative* values correspond to net heating. The mean θ-change is estimated at the trajectory origins using (5) with pressure replaced by θ_i'.

A striking feature of $\overline{\theta'}$ is that it has much greater magnitudes in winter than summer. For example, for the mid-Atlantic cluster in summer $1.5 < \overline{\theta'} < 3.5$K over almost the entire origin area indicating an average heat loss of roughly $0.5\mathrm{K\,day}^{-1}$. Since the trajectories are mainly travelling at low levels during this time, exchange with the surface or radiative effects could be contributing to the heat loss. In winter the sign of the mean heating varies and has most extreme values for the European cluster (−11.7K, +8.8K). The θ-change along trajectories is very variable and origin averaged standard deviation, $\sigma_{\theta'}(\mathbf{x})$, has a maximum of 14K over the Hudson Strait decreasing to under 7K throughout the region where $\overline{\theta'}$ is positive.

For the American cluster there is a gradual increase from MAM to DJF in the strength of the positive $\overline{\theta'}$ region to the West and South. In DJF the heat loss along trajectories descending over the Rockies is as much as $1.5\,\mathrm{K\,day}^{-1}$. This cooling is probably radiative and results in trajectories descending at a steeper angle than the isentropic surfaces. This contrasts with the annual cycle for the other clusters where MAM and SON are quite similar.

Interestingly, the areas where trajectories that experience heating originate (blue and green) are approximately the same for all three arrival grids – for example, Hudson Bay and Labrador in MAM. These patterns are not simply related to the origin averaged pressure. Also, the zero line in $\overline{\theta'}$ for the European cluster (except JJA) follows the centre of the stormtrack.

The net change in specific humidity along trajectories, averaged at their origins, is calculated in an identical fashion, replacing θ attributes with specific humidity attributes,

q_i. Any change in q is associated with non-conservation of water vapour. This may arise through condensation, evaporation and fluxes from the surface, particularly over oceans. There will also be changes associated with model error arising from the finite resolution of the water vapour field in the ECMWF model and the numerical dissipation from its advection, mismatches between the model and assimilated observations and errors in the trajectory calculation. It is hoped that the origin-averaged q-changes arise through physically realistic processes.

Examination of the mean q-changes reveals that they are closely tied in with the origin-averaged pressure. For all clusters the zero line in $\overline{q'}$ coincides with $\overline{p} \approx 850$hPa in MAM and SON, $\overline{p} \approx 875$hPa in JJA and $\overline{p} \approx 825$hPa in DJF. The general picture is of drying along trajectories from the South at low levels and moistening along trajectories descending from the NW. It is also clear that the magnitudes of the q-changes are much greater in summer than winter, in contrast to the heating pattern. It appears as though q along trajectories relaxes towards a background state which has a strong temperature dependence with a rapid decrease in specific humidity with height which is stronger in summer.

There is no obvious connection between the q and θ-changes. At first sight this is surprising because large-scale ascent would be related to saturation and condensation, and consequently drying and latent heating. However, all these trajectories arrive at low levels and have descended on average. Trajectories coming from the South at low levels experience drying and heat loss, probably related to boundary layer fluxes. Moistening occurs during descent and this presumably occurs because the water vapour field in the analyses has a profile that is maintained through mixing from the surface by the boundary layer and convection, but also through assimilation of new data and numerical diffusion. In other words, q changes following the resolved flow because of sub-grid scale water vapour flux convergence associated with convection or boundary layer turbulence.

3.3 Interannual Variability

When glancing through the density of origin plots for individual years it is immediately apparent that there is interannual variability. Although less marked than the variation with season the density maxima may move thousands of kilometres. The variability seems more pronounced for the European cluster than for the American. Winter seems more variable than summer for trajectories arriving over Europe. The similarity of the density from an individual season to the climatology is assessed objectively using the **overlap** integral:

$$\mathcal{O} = 1 - \int \frac{1}{2} |\rho(\mathrm{x}) - \rho_c(\mathrm{x})| \; d\Omega \tag{8}$$

where ρ_c is the climatological density of origin and the integration is over the whole sphere. The overlap is unity if ρ is identical to the climatology and zero if the non-zero area

of ρ does not overlap with the climatological non-zero area. The sensitivities of the overlap to trajectory error and number density are discussed in Section 4.4.

The time series of the overlap measure are shown for two and five days before trajectory arrival. In almost all seasons the overlap is lower for $T = 5$ but the fluctuations are related, highlighting sensitivity of trajectory statistics to flow changes. There are a few seasons that stand out as particularly anomalous, having the lowest values of overlap, but no single year is anomalous for all seasons or for all regions. For example, the most anomalous season is DJF 1988 for the European cluster: looking back to that year's density of origin one can see that it is clearly bimodal. For the European cluster the seasons with lowest overlap often correspond to those which have a period of strong blocking and a eastern density maximum (e.g. MAM84, JJA83, SON93, DJF91). For the mid-Atlantic cluster the most anomalous summers are characterised by a stronger maximum to the South (e.g. JJA80, JJA88). Its winter densities are often bimodal, which is detected in the climatology and DJF80 is a year when the maximum to the SE is particularly strong. The most anomalous seasons for the American cluster also correspond to strong maxima in the South (e.g. MAM85, MAM91, SON85).

4 Statistics of the Daily Characteristics of Trajectory Clusters

4.1 Definitions of Measures

Every day when trajectories arrive on one of the three grids, the spread of their origins is extremely wide. Fig. 2 shows the trajectories for the European cluster for one day taken at random from the ERA period. Trajectories may cross, particles may travel together or disperse rapidly and they may come from different altitudes or experience different heating. The density of origin compiled using trajectories from a whole season and the related origin-averaged fields represent just one aspect of the rich trajectory behaviour. It is also of interest to know how widely the trajectory clusters are spread before arriving on individual days, how far the centroid of the cluster has moved and how rapidly pairs of trajectories separate on average as you trace them backwards in time. These characteristics relate to the rate of stirring and mixing of trace chemicals and thermodynamic fields by the atmospheric flow.

A variety of measures are defined below and their statistics are described in the following two sections.

4.1.1 Cluster Displacement

The cluster displacement refers to the great circle distance between the origin and arrival positions of a cluster's centroid. As with the other measures to follow, the positions of

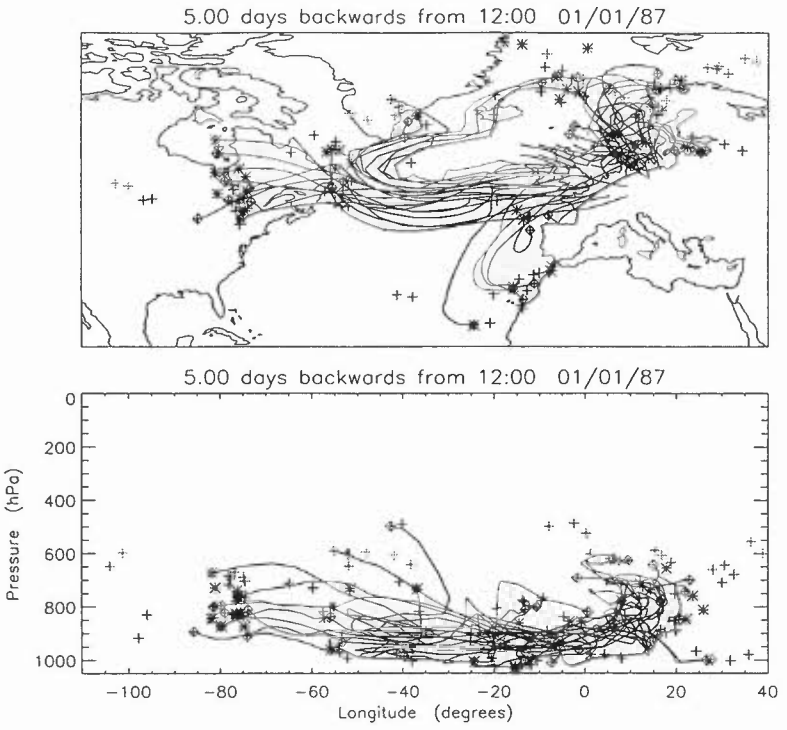


Figure 2: Trajectories arriving on the sparse European grid at 12UT on 1/1/87. The symbols marking the $T = 5$ origins of all European trajectories are the same as in Fig. 1.

particles on the Earth's surface are described in 3D Cartesian coordinates, assuming that the trajectories lie on the surface of a sphere with radius 6371km, in order to avoid problems with periodicity in longitude and the poles. A cluster centroid is found by calculating the mean value of each spatial coordinate for the particles in a cluster and then defining the unit vector, $\langle \mathbf{x} \rangle$, in the direction of the radial position vector of the centroid. The cluster displacement is given by:

$$D(t,T) = a \cos^{-1} \left(\langle \mathbf{x} \rangle [t-T] . \langle \mathbf{x} \rangle [t] \right) \tag{9}$$

where a is the Earth's radius.

4.1.2 Cluster Spread

On the sphere the RMS spread of particles from the cluster centroid can be defined as:

$$R(t,T) = a \left\langle \left[\cos^{-1} (\mathbf{x} . \langle \mathbf{x} \rangle) \right]^2 \right\rangle^{\frac{1}{2}} \tag{10}$$

where $\mathbf{x}$ denotes the particle position vectors an interval T before arrival and the angle brackets denote averaging over particles within the cluster. The arrival grids themselves have quite a wide spread which varies slightly due to the difference in their latitudes. The spreads of the American, mid-Atlantic and (sparse) European grids are 779, 714 and 645km respectively. It is also interesting to measure the spreads in the meridional and zonal directions giving information on the gross shape of each cluster. The RMS meridional spread is:

$$Y(t,T) = a \langle\, [\phi - \langle\phi\rangle\,]^2 \,\rangle^{\frac{1}{2}} \tag{11}$$

where ϕ is latitude. The meridional spread for all three arrival grids is 381km. The zonal spread is defined using a local Cartesian approximation giving $X(t,T) = (R^2 - Y^2)^{1/2}$.

4.1.3 Cluster Displacement and Spread in the Vertical

The cluster averaged change in altitude, potential temperature and specific humidity are also shown in this atlas. Firstly the change in each quantity is evaluated along each trajectory with the sign convention as in (7). The cluster "displacement" in each quantity is simply the cluster average of the changes. The "spread" in the potential temperature changes experienced along trajectories is defined as:

$$\langle [\delta\theta']^2 \rangle^{\frac{1}{2}} = \langle\, [\theta' - \langle\theta'\rangle\,]^2 \,\rangle^{\frac{1}{2}} \tag{12}$$

where $\langle\theta'\rangle$ is the cluster averaged θ-change and $\delta\theta'$ denotes a deviation from this average. Since all trajectories arrive at the same pressure (900hPa), the spread in pressure changes along trajectories is equal to the spread in trajectory pressures.

Unfortunately the displacement and spread in q cannot be calculated for the last 6 months in ERA due to an interpolation error affecting only q in the trajectory calculations. The climatology, power spectra and histograms of q-change also do not include these 6 months.

4.1.4 Correlation Dimension

If a compact cluster of particles is released in the atmosphere the cluster will tend to be stretched rapidly in the direction of the average strain following the cluster centroid. This stretching rate is exponentially fast on average and eventually the cluster becomes so elongated that its ends are caught in different eddies and it folds. The cluster continues to stretch and fold many times so that from a coarse-grained viewpoint it can resemble a cloud of particles which have mixed within the cloud (see discussion in Pierrehumbert and Yang,

1993; Maryon and Buckland, 1995). However, if there are sufficient particles, on closer inspection one would still be able to discern the filamentary structure within the cloud.

The correlation dimension characterizes the "internal structure" of a cluster of particles over a range of length scales. Following Pierrehumbert (1991) , the first step in its calculation is to find the great circle separation between all pairs of particles in the cluster ($r_{ij} = a \cos^{-1}(\mathbf{x}_i.\mathbf{x}_j)$). The correlation function (or cumulative histogram), $H(r)$, is then defined as the number of particle pairs with a separation less than r, as r is varied. If H has a power law dependence on r over a range of scales then the distribution of particles is said to possess a scale-invariant structure. If $H \propto r^d$, for $r_0 < r < \alpha r_0$, then d is the correlation dimension. Its value varies between 0 and 2; a value of zero indicates that all particles are clumped together in a few clusters with spreads less than r_0. A value of one would be obtained if the particles were strung out in a long line while d approaches two if the particles are distributed homogeneously over a region with scale greater than αr_0. Here the correlation dimension is simply calculated as:

$$d(t,T) = \frac{\ln(\, H[\alpha r_0, t-T]/H[r_0, t-T]\,)}{\ln \alpha} \tag{13}$$

where the smallest length scale r_0 is set equal to the median particle separation for the cluster's arrival grid (approximately the RMS spread at $T = 0$) and the range of scales, α, is fixed. Since half the particle separations are less than r_0 at $T = 0$, the correlation dimension is expected to be lower than one for the arrival grids and its value is expected to increase as the cluster stretches out. The value of α must be large in order to reduce d_0 towards zero, but must be small enough that the length scale αr_0 does not exceed the maximum particle separations at long times ($T = 5$ days). $\alpha = 4$ was chosen because over 5 days the RMS cluster spread increases by roughly a factor of 3 and consequently the range where $\ln H$ versus $\ln r$ follows a straight line includes the range $r_0 < r < 4r_0$. d_0 is then equal to 0.494, 0.487 and 0.510 for the American, mid-Atlantic and European clusters respectively.

4.1.5 Finite Time Lyapunov Exponent

Dense clusters of particles tend to stretch at an exponential rate forming filaments that subsequently fold and stretch further. Similarly, two particles with a small initial separation tend to separate at an exponential rate (Cocke, 1969). This continues to occur until the separation is so large that the particles are influenced by different eddies; this is equivalent to the folding stage for the compact cluster (decribed as plication by Maryon and Buckland, 1995). Folding events tend to slow the rate of separation.

The stretching rate also gives the rate at which tracer gradients at the edges of filaments increase with time (Varosi *et al.*, 1991). This tightening is arrested by mixing at the edges of filaments and reaction if the tracer is chemically active. If the stretching rate is

high, the tracer gradients are also expected to reach higher values. The sharpness of gradients between chemicals and the distribution of gradients are important to the net chemical reaction in a region if the reaction rates are nonlinearly dependent on concentrations.

A measure used to characterize the average stretching rate experienced by a line element connecting two nearby trajectories is the finite time Lyapunov exponent:

$$\lambda_{ij}(t,T) = \frac{1}{T} \ln \left(\frac{r_{ij}[t-T]}{r_{ij}[t]} \right). \tag{14}$$

The λ_{ij} values are found for the line elements connecting each particle with its four nearest neighbours on the arrival grid. The median Lyapunov exponent, $\lambda(t,T)$, is then used to represent the average stretching rate experienced by line elements within the cluster. It was found to be necessary to use four neighbours for each particle, rather than the nearest to the North for example, in order to reduce the effects of the initial orientation in the early stages when line elements rotate to align with the strain. Since the histogram of Lyapunov exponents for each cluster tends to have a long tail to high values, the median gives a closer estimate to the peak of the values than the mean. It is also less susceptible to fluctuations in the tail.

4.2 Collation of Statistics

4.2.1 Time Series

The daily values of the cluster measures described in the last section are plotted for the whole duration of the ERA period. Time series are shown for each cluster for trajectory lengths of 2 and 5 days. Some measures show much more variability than others; for instance cluster displacement has a much wider range of values than meridional spread. It is possible to pick out the dates of extreme events and sharp regime changes that maybe obscured by averaging or are lost when the distributions of values are obtained. Also trends across seasons or interannual changes are not hidden.

In addition to the rapid fluctuations, many of the measures show an extremely strong annual cycle. The strength of the annual cycle can vary with position along the stormtrack. The next section describes a power spectral analysis which confirms the dominance of the annual frequency in most measures. The distribution of values is then decomposed into histograms for each season for the three arrival regions.

4.2.2 Power Spectra of Time Series

The power spectra of the time series have been calculated using discrete Fourier transforms of the data and finding the power associated with each frequency in the range $1/W \leq f_n \leq$

$N/(2W)$ where W is the length of a time window and N is the number of discrete time points in that window (the number of days in this case). $1/W$ is the lowest frequency with a complete cycle in the window and $N/(2W)$ is the Nyquist frequency. The method is described in detail in §12.7, Press *et al.* (1992) .

If the window length was to be set to the full 15 years of ERA the power spectrum obtained would be noisy and the uncertainty associated with each frequency would be 100%. The uncertainty in the power spectral densities is reduced by obtaining the power spectrum for a number of shorter windows and averaging the densities. The most efficient use of a finite time series is made by overlapping windows by half their length, so that $M = (2X/W - 1)$ windows are used to obtain a power spectrum for a time series of length X. The variance associated with each frequency's power spectral density is then reduced by a factor of about $9M/11$. The obvious drawback of using shorter windows is that the longest period for which a power estimate can be obtained decreases and the interval between the discrete frequencies increases. There is a playoff between accuracy and resolution.

Additionally, when transforming a section of a time series using Fourier techniques it is assumed that the whole time series is periodic with period W. This introduces two problems. The first is that the transform is effectively the transform of the product of the time series with a unity top-hat function of width W. The result is that the power in a single frequency is spread into neighbouring frequencies. The integrated magnitude of this spread is reduced by multiplying the data in each time window by a window function, which changes smoothly from zero at the window edges to one at its centre, before transforming. Here a Parzen (sawtooth) window has been used (see Press *et al.*, 1992). The second problem arises if the data has a trend over each window. The amplitude of the trend can strongly influence the low frequency end of the spectrum. A better estimate of the low frequencies was obtained by detrending (subtracting a linear regression through the time series in each window) before the spectral transform.

For the power spectra shown here, a window length of 5 years was used so that the variance in the power spectral densities was reduced to about 24%. The strong peak at the annual frequency is resolved although it's power clearly spreads to the neighbouring two frequency bins (with periods of 5/4 and 5/6 years). Note that without detrending the lowest few frequencies have more power than the annual spike, but this power is an artifact that depends on window length W. The spectrum for frequencies much higher than annual is better resolved by using a shorter window length, say one year. The resulting high frequency spectral densities were found to follow a curve resembling an running average through the spectral densities shown here (with log-axes). The only two peaks to stand out above the background variance, for a range of window lengths, were the annual peak for most measures and a semi-annual peak, with a 0.5 year period, for a few of the spectra. For all spectra there is a marked drop in power for frequencies higher than $0.1\,\mathrm{day}^{-1}$. This is related to the fact that trajectory length is 5 days so that trajectory arrivals on consecutive days are experiencing different regions in the same flow. In some sense there is averaging over a 5 day

wide moving window which reduces variability in the highest frequencies.

Although the 15 year period is too short to reliably measure power at interannual frequencies, it is sufficiently long to resolve the strong annual peak. Therefore it seems that 15 years is just long enough to obtain a climatology of the seasonal cycle in the trajectory statistics. Moreover, it is also sensible to describe those years which deviate significantly from the climatology as anomalous, as was done with the overlap measure.

4.2.3 Histograms

Histograms using the daily values of each measure from the whole ERA period are shown in a matrix with arrival location in columns and seasons stratified by row. Thus by scanning across each matrix one can see the variation in the distribution of each measure with position relative to the stormtrack and by scanning down each column one can see the seasonal variations.

Three histograms are shown in each panel for the European cluster. The heavy solid line uses only trajectories from the "sparse" arrival grid to calculate the measures. The density and size of this arrival grid is the same as for the American and mid-Atlantic grids. The dotted line uses all the European trajectories and the light solid line uses only those trajectories arriving on the "compact" European arrival grid. These extra histograms are used to illustrate the dependence of the Lagrangian statistics on trajectory sampling.

4.3 Climatology and Variability

4.3.1 Dependence on Trajectory Length

The variation of each measure with the time before cluster arrival is most clearly seen in the histograms. For instance, the median RMS spread always increases with trajectory length. At 2 days before arrival RMS spread varies approximately between one and three times the arrival spread (see Section 4.1.2), with the peak biased towards lower values. In fact, there is a small probability that a cluster's spread decreases with trajectory length (discussed in Section 5). Following clusters back 5 days, the spread can be as much a five times its value on arrival and the histogram is much wider than before.

The histogram of cluster displacement widens enormously with increasing T. At $T = 2$, the median displacement approaches the median spread of the cluster although the histogram is quite different, starting from zero and with a long tail to values higher than the highest spread. The stretching of the displacement histogram towards much higher values continues with T resulting in a very flat distribution.

The pressure displacement histogram has a similar behaviour to the one for horizontal displacement; it extends from near zero to large, but negative, $\langle p' \rangle$ and this tail continues to

extend with T. Since trajectories all arrive at 900hPa it is of no surprise that most descend to this level and only a very small percentage of clusters experience average ascent. The percentage with net ascent decreases with T because trajectories sample a wider variety of flows. Over 5 days the maximum cluster descent is about 300hPa from the 600hPa level. This corresponds with the lowest values of $\overline{p}(\mathbf{x})$ indicating that on almost every occasion when the cluster average descent is so large, trajectories originate from a fairly small region.

The θ displacement histogram broadens considerably with trajectory length. For some clusters and seasons there is a significant shift in the median to be discussed below. In constrast, the histogram of q displacement does not broaden a great deal but has a consistent shift towards negative values as T increases. This ties in with the earlier conclusion that q' is tied strongly to p'. On average, the further trajectories descend the more their specific humidity increases.

As the trajectory lengths, T, are increased line elements experience stretching rates throughout different regions of the flow and the Lagrangian average stretching rates become more similar for different line elements. Consequently, the histogram of Lyapunov exponents becomes *narrower* as T increases. The trajectories are described as chaotic if the Lyapunov histogram collapses onto a single positive value as $T \to \infty$ (Pierrehumbert, 1991). This behaviour is also seen here with the histogram of median Lyapunov exponents compiled from daily trajectory arrivals over a season and even in the climatological histogram for that season. However, the median of the climatological histogram often shifts to slightly smaller values as T increases because the line elements get so long that "folding" events slow the rate of separation or even result in a transient decrease in separation. The median of the histograms ranges from 0.2–0.4 day^{-1} which is about half the value of the histogram peaks obtained for the cases examined by Maryon and Buckland (1995) . However, these values are in close accord with the peak of the histograms in Pierrehumbert and Yang (1993) for trajectories on the 315K isentropic surface.

The correlation dimension (only shown in the time series) is seen to increase rapidly, from 0.5 to 0.7–0.8, over two days following clusters backwards and reach a value close to one by $T = 5$ days. This value is consistent with the filamentary nature of the distribution of particles in individual cluster releases. Maryon and Buckland (1995) found similar results in their case studies, even though their cluster was initially point-like and began spreading by a diffusive process superposed on the advection. The correlation dimension, defined by the maximum slope in their $\ln H$ vs $\ln r$ diagrams for the $100 < r < 10000$ km range, reached 1.0 in all cases by 5 days and about 1.5 after 10 days. The spread between their cases was large; here the spread averaged over the ERA period is quantified by the standard deviations in the time series (~ 0.2).

4.3.2 Dependence on Arrival Region and Season

The time series and power spectra show that the annual signal in RMS spread increases dramatically looking at arrival regions further along the stormtrack. In fact, there is no significant peak in the power spectral density for the American cluster but the annual peak is almost two orders of magnitude larger than any other frequency's value for the European cluster. Although this behaviour is seen in zonal spread it is not seen for meridional spread because the annual peak for the mid-Atlantic cluster is small. Interestingly, there are strong peaks at periods of exactly 0.5 years for the zonal spread of the mid-Atlantic cluster and the meridional spread of the American cluster. The latter peak arises because the meridional spread of the American cluster is much larger in JJA than in the other seasons.

The median horizontal displacement is typically greatest for the mid-Atlantic cluster and is smallest for the American cluster. Throughout the stormtrack, there is a prominent seasonal cycle with cluster displacement being smallest in summer. The annual signal is strongest for the American cluster whose displacement is three times less in summer than in winter. The other two clusters have a peak at the 0.5 year period because, in addition to the strong annual variation, displacement in SON is much greater than in MAM. The displacement spectra generally show much more power at intra-annual frequencies than the spreads' spectra up to $0.1\,\mathsf{day}^{-1}$ where the drop off associated with trajectory length is more marked.

At all arrival locations the histogram of pressure changes is far narrower in summer than for the other seasons. The histograms do not vary greatly with position in the stormtrack, although the median descent in summer is largest arriving at the American grid. Also, there is a clear semi-annual signal for the American and mid-Atlantic clusters.

The width of the histogram of θ-displacements is clearly much greater in winter than summer and for the European cluster compared with the other two. The JJA mid-Atlantic histogram lies almost entirely in the positive range as would be expected from the origin-averaged θ-change climatology shown earlier. The median of the other histograms is harder to interpret because the values contributing to the distribution are clearly a function of their origin superimposed on a fluctuation of comparable magnitude. As trajectory length increases, the American and mid-Atlantic histograms show an increase in the positive tail which outweighs the increase in the negative tail. This is probably because the longer trajectories tend to originate from higher levels where they experience radiative cooling. The American histogram also has a systematic shift towards positive values over the year from MAM to DJF. This trend is clearly seen in the raw time series which shows a ramping up from April to February and then a sudden drop back to April values. Consequently, there are strong peaks in its spectrum at annual and semi-annual frequencies (imagine a semi-annual cycle with a peak at the end of March superposed on an annual cycle with a peak in January).

The European histograms (except JJA) have long negative tails for $T = 2$ and as trajectory length increases the shape remains but the spread increases. Also, the median is

close to zero and does not change significantly with T. These observations accord with the climatological $\overline{\theta'}(\mathbf{x})$ which showed that trajectories from the NW experienced strong heating on average ($\theta' < 0$) but there was a wide variation about this mean.

As noted earlier, the q-displacement histograms are biased towards negative values and are narrowest in winter for all clusters. The American cluster's q-displacement spectrum is much steeper than its θ-displacement spectrum, showing much more variability on the intra-annual timescales and less of an annual dependence.

The finite time Lyapunov exponent has a very clean annual cycle throughout the stormtrack with a peak in winter, although of small amplitude compared with the fluctuations in displacement. Indeed, power in all other frequencies is very low and its spectrum is the flattest of all measures. The American cluster has distinctly lower stretching rates than the others for $T = 2$ but the median values increase with trajectory length. Individual cluster releases tend to stay clumped together and then stretch out at a later stage and consequently the correlation dimension is lower for the American cluster. At the other locations, where stretching is faster, the median of the Lyapunov exponent histogram decreases with T because trajectories that arrive as neighbours can have origins effectively lying either side of a fold in a tracer contour.

4.4 Sensitivity of Measures to Sampling of Trajectories

Here the sensivity of the measures presented in this atlas to the size and density of a cluster on arrival is investigated. As mentioned in Section 4.2.3, the measures were calculated for three clusters arriving over Europe: the full high density grid, a sparse grid with quarter the density and a compact grid with high density but quarter the area. The daily values of these measures were compiled into three separate histograms, the dotted, heavy solid and light solid lines respectively.

For all trajectory lengths, the histogram of RMS spread is shifted to smaller values for the compact cluster, although the high spread tail almost reaches the values attained by the larger two clusters. The sparse cluster's histogram is indistinguishable from that for the full cluster.

In horizontal, pressure, θ and q displacements the histograms widen to larger displacements upon reducing cluster size. This reflects the fact that cluster averaged fluctuations in displacement are more likely for smaller clusters which sample a smaller volume of the atmosphere. The cluster density has a negligible impact on the histograms indicating that the sparse density is high enough for a representative sampling of trajectories arriving over these regions.

The differences between the density of origin obtained from the full and sparse European clusters (for 1979) was found to be considerably smaller than the interannual variability when

measured by the overlap integral. The $T = 2$ overlap was always greater than 0.96 and the $T = 5$ overlap greater than 0.95. A similar sensitivity in the overlap measure was found to using cluster arrivals from every other day, rather than daily, although the overlap was 0.01 higher in summer and 0.02 lower in winter. This is clearly related to the much wider spread of origins in winter compared to summer and the faster rate of separation of trajectory pairs. The density of origin for the compact European cluster had a much lower overlap of about 0.90 when compared to the density for the full cluster. It is no surprise that the overlap is lower than for the sparse cluster when considering the much lower RMS spread for each cluster arrival. Nevertheless, this overlap is still higher than that obtained by comparing individual years to the climatology, which indicates that the density of origin is influenced strongly by changes in individual cluster displacements in addition to their spreads.

Both cluster size and density affect the histograms of Lyapunov exponents in different ways. The width of the histograms increases on reducing cluster size. For 2 day trajectories, the histogram of Lyapunov exponents is unaffected by cluster arrival grid density and the spread from reducing cluster size is seen to be roughly symmetrical. As trajectory length increases the histogram for the sparse cluster is shifted bodily to lower values, but the widening of the histogram for the compact cluster is centred about the full cluster's histogram. This behaviour occurs because the separation of neighbouring trajectories, as one follows a cluster backwards, is only exponentially rapid while their separation is smaller than the dominant scales in the velocity gradient (i.e. it is accurate to linearise the relative velocity between two trajectories). Once a cluster reaches this scale folding events slow its stretching rate. After 2 days the line elements connecting our trajectories are short enough that their length does not affect their stretching rate. However, by 5 days the length of line elements has become important for the sparse cluster, while the median for the full and compact clusters remains almost constant. The sparse clusters underestimate the true median of the Lyapunov exponent histogram obtained with infinitesimal line elements by approximately $0.03\,\mathrm{day}^{-1}$. Therefore, it can be concluded that a 1° arrival grid (the high density) is sufficient for a stable Lyapunov exponent histogram out to 5 day trajectory lengths. This conclusion is supported by the similarity in the peak values of the Lyapunov exponent histograms obtained by Pierrehumbert and Yang (1993) . They integrated the linearised evolution equation for line elements, which is equivalent to ensuring that trajectory pairs are arbitrarily close.

5 Temporal Characteristics of Trajectory Cluster Dispersal

5.1 Time Dependence of Cluster Spread

Cluster spread tends to increase with the length of trajectories. The spread occurs in pressure, potential temperature and specific humidity as well as in the horizontal. If each step made by a particle over interval δt was uncorrelated with the last then the trajectory would be a random walk. Moreover if the steps made at any instant by the particles in a cluster were

also uncorrelated then the cluster would spread diffusively. The model for such a diffusive increase of cluster spread in a coordinate z is:

$$\langle [\delta z']^2 \rangle = 2\sigma_z^2 \tau_z t \tag{15}$$

where the notation follows (12), $\sigma_z = \langle [\delta \dot{z}]^2 \rangle^{\frac{1}{2}}$ is the standard deviation of velocity in the z-direction and τ_z is the Lagrangian decorrelation timescale of z. The assumption for the diffusive model is that $\tau_z \ll t$ (see Sparling *et al.*, 1997). The Lagrangian decorrelation timescales will be defined and quantified in the next section.

The salient feature of the diffusive model is that the standard deviation in the *z–change along trajectories*, $\langle [\delta z']^2 \rangle^{\frac{1}{2}}$, has a $t^{1/2}$ time dependence. Note that if all trajectories start at the same z-level then the standard deviation in z values also has a $t^{1/2}$ time dependence. When there is a strong shear in $\dot{z}$ superimposed on the random velocity field, it can be shown that the standard deviation has a linear time dependence which is often termed ballistic or super-diffusive. Here the standard deviation of a cluster of *back* trajectories is assumed to have a power law time dependence on trajectory length of the form $\langle [\delta z']^2 \rangle^{\frac{1}{2}} \sim T^s$. For every cluster arrival a linear regression of $\ln(\langle [\delta z']^2 \rangle^{\frac{1}{2}})$ against $\ln T$ is performed and s is found from the slope of the fit.

Histograms of the spreading exponents, s, are shown for the increase in RMS spread as well as meridional, zonal, pressure, potential temperature and specific humidity spreads. The power law spreading model is fitted to the values of the spreads, obtained as in Section 4.1, at 6 hour intervals in trajectory length T. As for the other cluster measures, these histograms are displayed in a region versus season matrix.

The narrowest peaks are obtained for the histogram of exponent for RMS spread, s_R. Median values are lowest for the American cluster ($\sim$ 0.32) and highest for the European (0.35–0.45). The median exponent for meridional spreading, s_Y, is always higher and has a range 0.38–0.47, and the histograms are generally wider. It is the rapid slow down in the zonal spreading which drags the exponent s_R to lower values than s_Y. The zonal spreading exponents, s_X, are particularly low for the American cluster and in MAM and JJA a sizeable fraction of its histogram lies at negative values of s_X. In other words it is quite common for the American cluster's zonal spread to decrease with time during summer. This behaviour cannot be viewed as a diffusive process.

Power laws for cluster spreading do not often apply well to the dispersal of individual clusters, in the sense that the spread of the points about a plot of $\ln R$ versus $\ln T$ is typically large compared with the trend. However, it is usually the case that a localised cluster will tend to spread when advected forwards in time and may even have an exponent $s_R \approx 0.5$. A cluster of trajectories influenced only by a random velocity field will also spread diffusively with this time dependence. Although similar in this measure of spreading these two behaviours are very

different because trajectories integrated using a specified wind field are reversible whilst those influenced by the random velocity field are not. For instance, if a cluster of trajectories is advected forwards and spreads over interval T and then is integrated backwards from the final positions, the particles arrive almost exactly at their original positions (see Section 6.1). For this "final" arrangement particles the cluster spread decreases with back trajectory length. As back trajectory length increases the fraction of the set of arrangements for which it is possible to "put the genie back into the bottle" is expected to decrease because the Lagrangian decorrelation timescale is finite. The random velocity field is an extreme limit where the decorrelation timescale is zero so that the probability of putting the genie back is also zero. An estimate of this probability, P_g, for the North Atlantic clusters is given by the fraction of the histograms of spread that lies below the $T = 0$ spread. Only one day before arrival P_g is less than 2% for the European cluster, dropping to $< 0.5\%$ for $T = 2$ and undetectable for $T \geq 3$. These probabilities are not affected by arrival grid density but are larger by a factor of about 3 for the compact cluster. In contrast, the probability of a decrease in cluster spread is much higher for the American cluster: for T=1, 2, 3 and 4 the probabilities are $P_g \leq 24\%$, 9%, 2.5% and 0.5% respectively, dropping below detection for $T = 5$.

The histograms of exponents for the compact European cluster are far wider, extending to higher values of the exponents. The peak in the s_R histogram varies about 0.6 with its highest value in winter. The lower values of exponents for the large clusters imply that their rate of spread slows down more quickly than for the compact cluster or even for diffusion. There must be a degree of "sticking" of trajectories within cyclones and anticyclones. This type of behaviour was seen by Pierrehumbert and Yang (1993) to slow the rate of homogenisation of tracer, as measured by the approach of the correlation dimension of clusters towards two. Since the exponents, s, are not scale invariant it seems that the power law model is not a very good one. There is also a far greater symmetry between the zonal and meridional directions, especially in summer, for the compact cluster. The asymmetry for the large clusters is likely to be a combined result of the rectangular shape of the arrival grids and the typical eddy radius. The zonal spread on arrival is much larger relative to an eddy radius than the meridional spread and this asymmetry is reduced following the cluster dispersal. The compact cluster does not disperse widely enough for the asymmetry to be reduced. These results contrast with those obtained by Pierrehumbert and Yang (1993) for the dispersal of trajectories on the 315K isentropic surface. In the upper troposphere, the shear at the flanks of the jet stream results in a super-diffusive spreading exponent in the zonal direction. For 3D trajectories arriving at low levels the flow is far more blocked and the zonal exponent is at most diffusive.

The exponents for the spreading in pressure and θ are considerably larger, with most medians in the range 0.6–0.7. These exponents are super-diffusive which points to the importance of the systematic change in p or θ following clusters. In the case of pressure this arises through descent close to isentropic surfaces. For dispersal in θ this arises through common heating histories - for instance the radiative heat loss experienced for the higher level trajectories. Sparling *et al.* (1997) also found super-diffusive spreading in θ for clusters

in the lower stratosphere which tended to diffusive spreading on timescales greater than a month.

It appears that diffusion is a poor model for the spread in q changes. Spreading exponents are low and frequently negative. The previous Lagrangian diagnostics showed that q systematically increases following trajectories down into the lower troposphere where sub-grid scale flux convergence maintains high humidity. Since q falls so rapidly with height, its spread in values also tends to be much smaller at higher altitudes with the result that the spread in q can decrease when following a cluster backwards in time.

5.2 Lagrangian Decorrelation Timescales

Lagrangian decorrelation timescales measure the time taken for the *cluster relative velocities* of particles to become uncorrelated along their trajectories. As discussed above, for times long compared with the decorrelation timescale the cluster is expected to spread quasi-diffusively as given by (15). This timescale is obviously dependent on cluster size relative to the scales in the strain field, as well as the cluster's arrival region and season. If a cluster's arrival grid was very compact compared to the scales in the resolved strain field then this timescale would be long and related to the cluster average stretching rate. For a given flow geometry, the decorrelation rate would increase if stretching rate increased in the limit of compact clusters.

However, when the cluster is always of comparable size to the eddies the dependence on cluster size is less important and the timescale effectively measures the time taken for total velocity to decorrelate along trajectories when averaging over all trajectories. If the cluster under consideration consisted of a nearly homogeneous distribution of particles on a global isentropic surface, the decorrelation timescale for isentropic velocity would indicate the time taken for line elements with random orientation to align with contours of a long-lived tracer (Haynes and Anglade, 1997). In this limit the relationship between stretching rate and decorrelation rate is not obvious. Typically it is found that that the decorrelation timescale is short compared to stretching timescales so that there is a tendency for line elements to align rapidly with the local strain. Ngan and Shepherd (1999) point out that this property is shared by the chaotic advection scenario, where a large-scale coherent wind field varies slowly in time at a point but trajectories diverge rapidly, experiencing regions of the flow with very different velocities.

Following Sparling *et al.* (1997) , the Lagrangian autocorrelation function (LAF) is defined to be:

$$\mathcal{L}(t,T) = \frac{\langle \delta\dot{z}[t]\ \delta\dot{z}[t-T] \rangle}{\sigma_z[t]\ \sigma_z[t-T]} \tag{16}$$

where $\delta\dot{z}$ denotes the deviation in the z-component of velocity from the cluster mean and σ_z is the standard deviation in the z-component of velocity. For this diagnostic the

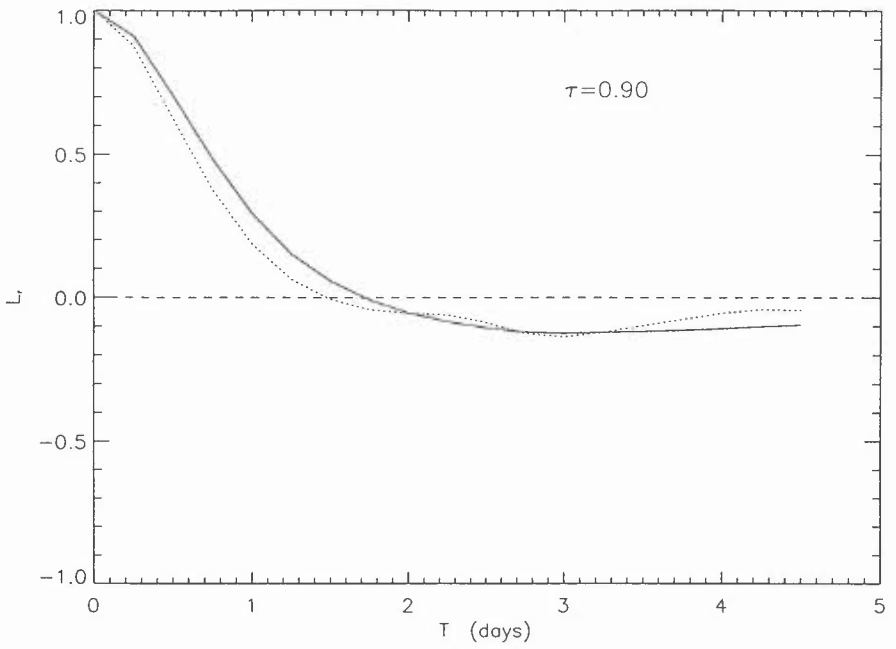


Figure 3: The seasonally averaged Lagrangian autocorrelation function for horizontal velocity versus trajectory length for MAM 1979. The solid line and timescale have been obtained using all trajectories in the European cluster and the dotted line has been obtained for the compact subset.

velocities were approximated using centred difference in time of the trajectory coordinates because they were not stored as trajectory attributes. Note that the for the horizontal decorrelation timescale the horizontal velocity was estimated using the great circle distances between trajectory points and then the LAF was defined using the dot products between the velocity deviations, $\delta\mathbf{V}$.

The Lagrangian decorrelation timescale can then be defined in various ways. One can define the decorrelation timescale as the integral of the LAF over T from 0 to ∞. However, when there are fewer trajectories the LAF can oscillate about zero for individual cluster arrivals and these oscillations affect the timescale obtained if the integral is truncated at 5 days. Therefore the LAF is averaged over cluster arrival dates, t, for a whole season which eliminates the oscillations. Even then, the seasonally averaged LAFs found here have small, but persistent negative correlations at long times (see Fig. 3). These reduce the value obtained for the timescale even though they have no bearing on the rate of fall of the LAF towards zero. They also introduce a strong dependence of the timescale on T. The negative correlations arise because the trajectories arrive at rather special locations in the flow (i.e. near the ground) where the distribution of velocities is very different to their origins. The most robust definition of a timescale which is not sensitive to the long time, low correlation behaviour is to find the time taken for the seasonally averaged LAF to fall from 1 to $1/e$.

Lagrangian decorrelation timescales, τ_z, have been calculated for the horizontal projection of trajectory velocities, the vertical velocity in pressure and isentropic coordinates

and also for $\dot{q}$. Linear interpolation between trajectory lengths at 6 hour intervals was used to find the value of T when the LAF drops below $1/e$.

The decorrelation timescale for horizontal velocity, τ_r, is much longer than for $\dot{\theta}$ or $\dot{q}$. The only variation with arrival location is that the American cluster shows a stronger annual cycle than the European cluster. τ_θ is significantly larger for the mid-Atlantic cluster, although far shorter than the value of 10 days estimated for the lower stratosphere by Sparling *et al.* (1997) . The decorrelation timescale for vertical velocity lies between τ_r and τ_θ and is also larger for the mid-Atlantic cluster. The indication is that the vertical velocity is related to both the horizontal velocity and the heating along trajectories. In the absence of heating it would be reasonable to expect a similar decorrelation timescale for horizontal and vertical velocities as air glides up and down isentropic surfaces under the influence of baroclinic waves. However, strong gravity wave activity would be expected to increase the decorrelation rate for vertical velocity. Additionally, latent heating associated with condensation in regions of large scale ascent would result in trajectories rising at steeper angles than isentropic surfaces and radiative cooling in dry, descending regions would result in trajectories descending at steeper angles than the isentropic slope. It is easy to see that rapid decorrelation in heating would shorten the decorrelation timescale for vertical velocity.

The decorrelation timescale is almost independent of arrival grid density but the compact European cluster has a shorter τ_r. This behaviour is hard to understand but must be related to the larger range of cluster mean velocities for the compact cluster.

6 Interpretation of Lagrangian Models

6.1 Wind Field Resolution

The accuracy of trajectory calculations like those presented here was investigated by Methven (1997) . Using baroclinic wave life cycle simulations it was found that:

- The calculations are sufficiently precise that they are almost reversible. For example, a cluster of trajectories was integrated forwards in time for 12 days and then backwards in time from those end points. The position errors, averaged over trajectories, upon return at the original date were 0.017° in longitude, 0.008° in latitude and 3.7×10^{-5} in η.
- The errors arising from a trajectory time-step of one hour (comparing with an adaptive step size scheme) were almost two orders of magnitude lower than those from degrading wind field resolution from T85 to T42.
- Errors incurred by reducing temporal truncation, Δt, from 3 hours to 6 hours were smaller than those from reducing horizontal truncation from T85 to T42. These errors were in turn smaller than reducing vertical truncation from L60 to L30.

When using ECMWF analyses to advect particles forwards for 2 days it was found that degrading the wind field from T106 on 31 η levels to a 2.5° × 2.5° grid on 15 pressure levels introduced RMS distance errors of 270km. RMS pressure errors in trajectory positions were 25hPa. Individual trajectories were incorrectly positioned by as much as 2000km and 240hPa. These positioning errors seem rather large at first sight. However, they mainly arise because a small phase error as a trajectory passes a deformation point in the flow can amplify into a huge error if the trajectory passes to the wrong side of the deformation point. This is a manifestation of the sensitivity of trajectories to initial conditions.

When considering clusters containing many trajectories the measures of cluster dispersal as a whole are not so dependent upon the phase errors of trajectories relative to the wind field because each trajectory is displaced slightly from its neighbours on the arrival grid which makes some allowance for the sensitivity to initial conditions. Once a number of cluster arrivals have been combined to produce the statistics for a season, such as density of origin, the impact of wind field resolution is far less severe (e.g. Fig. 4 in Methven, 1997).

Moreover, experiments using the contour advection technique have shown (Methven and Hoskins, 1999) that most of trajectory error lies in the along-filament direction. The locations of the centres of tracer filaments are remarkably insensitive to wind field resolution. On average filament displacement errors are *an order of magnitude less* than the smallest resolved scale in the wind field (in adiabatic life cycle experiments or the lower stratosphere).

Filament displacement errors and contour stretching rates are however more sensitive to the temporal truncation of the wind than the simple trajectory distance errors suggest. In fact, we expect large differences in the formation and positioning of filaments unless $\Delta t \ll \tau_r$, where τ_r is the Lagrangian decorrelation timescale for the horizontal wind on the average isentropic surface followed by the trajectories. In this atlas τ_r was found to be about 1.1 days which is almost 10 times longer than $\Delta t/2$ indicating that 6 hours is a suitable interval between wind records (see discussion below (1)). However, the decorrelation timescales for vertical velocity, $\dot{\theta}$ and $\dot{q}$ are shorter than this. It would be desirable to reduce Δt to 3 hours to obtain greater accuracy in the vertical motion and even shorter intervals maybe required if the heating and moistening experienced by individual trajectories are to be followed. Indeed these short decorrelation timescales may have been overestimated by calculations using a 6 hour interval. Nevertheless it is hoped that the statistics of θ and q changes are representative of real changes following "air-masses" whose motions are represented by the resolved flow in the analyses.

6.2 Sub-Grid Scale Transport

The trajectories shown in this atlas follow the path of a particle advected by the resolved velocity field in the ERA dataset. However, there will always be sub-grid scale motions which advect material in the atmosphere.

Turbulent motions with decorrelation timescales much shorter than the interval between wind records and spatial scales much smaller than the smallest scales resolved in the model can be considered to act like diffusion, mixing concentrations of tracers between neighbouring grid boxes. A real particle (with density equal to air density) would tend to follow a path resembling the resolved flow plus a random walk. Following filaments of tracer there is a competition between spreading at a diffusion rate dependent on the sub-grid scale velocity variance and thinning due to the resolved straining motions. The values of a passive tracer's mixing ratio following the resolved flow change due to the volume-averaged convergence of sub-grid scale tracer fluxes. The flux convergence is often modelled by relating the fluxes to the resolved gradient in tracer mixing ratio.

There is another class of "sub-grid scale" motions where their timescales maybe short but their lengthscales are comparable or even longer than the grid scale in the vertical. The resolved velocity in the ECMWF model and the re-analyses respects mass continuity. Since the model is also hydrostatic, the velocity is three dimensionally non-divergent in pressure coordinates. To the extent that the atmosphere is hydrostatic on scales close to the grid scale, the sub-grid scale velocities are also non-divergent. Therefore net transport into a volume (in pressure coordinates) is again only possible through velocity-tracer correlations. For example, consider that buoyant convection is active over some fraction, a_u, of each grid box's area and that the area-averaged updraft velocity, ω_u, is known. If the atmosphere is then filled with particles we can expect a fraction a_u to experience a vertical velocity $(\omega_r + \omega_u)$ and the remaining fraction to move with velocity $(\omega_r - a_u\omega_u/(1-a_u))$. The rate of dispersal of a localised cluster of particles will be more rapid if convective transport is accounted for but the tracer concentrations will only change within a grid-box if there is a tracer gradient.

If particles are allowed to follow the resolved flow plus the parametrized convective velocities in a GCM, then the rate of change of q and θ following particles will be dominated by condensation in updrafts, plus evaporation and radiative transfer in the environment. However, following the resolved flow the net changes in potential temperature are given by first integrating the heating within updrafts and their environment over grid box volumes and the interval between analyses, and then monitoring those values following the trajectories. Expressions are given for the grid box averaged tendencies arising from convection in Gregory and Miller (1989) . The values obtained for these θ and q changes are not trivially interpreted but it is precisely the heating following the trajectories of the resolved flow that influences the dynamics in the GCM and analyses. Furthermore, Valdes and Hoskins (1989) have shown that the heating distribution has a profound influence on the existence and nature of the stormtrack itself.

6.3 Validation of Trajectories with Chemical Observations

Absolute measures of trajectory accuracy are hard to define and typically displacement errors in trajectories are defined relative to a control simulation which is the best attainable.

However, if the trajectories of air parcels in the atmosphere are sufficiently similar to those following the resolved flow in the analyses, then it is possible to validate trajectory model performance using observations of chemical concentrations, combined with Lagrangian modelling of chemistry. There are many problems inherent in this approach including:

- Even if a trace chemical is truely passive such that $D\chi/Dt = 0$, the rate of change of mixing ratio following the resolved flow, $\dot{\chi}$, will not be zero if sub-grid scale flux convergence is important. These effects include mixing by small-scale turbulence and transport by convection. However, no account of these fluxes is made in the results presented here.

- Surface emissions of chemicals are typically mixed very rapidly within the boundary layer but only slowly from there by "entrainment" at the top of the boundary layer. There are many uncertainties in Lagrangian modelling associated with the parametrization of boundary layer height and the estimation of source rates associated with sub-grid scale flux convergence within the boundary layer and the emissions inventories.

Here results are presented using the CiTTyCAT Lagrangian chemical model from the Cambridge University Department of Chemistry to integrate reactions along trajectories calculated using the UGAMP trajectory model (as used for this atlas). The trajectories all arrive at Mace Head, Ireland during the EASE96 chemical observation campaign. Three time series of methane concentration at Mace Head are shown here by way of example. For a description of the Lagrangian chemistry model, the field used for initialisation and the comparison of other chemical observations with the model refer to Evans *et al.* (1999) .

The heavy solid line in Fig. 4 shows the observed methane trace (Derwent *et al.*, 1994). The dotted line is obtained by integrating trajectories backwards from each observation time at Mace Head and using the value of methane at the $T = 5$ origin obtained from a climatological model of the methane concentration field (Law and Pyle, 1993; Law and Nisbet, 1996). It is immediately apparent that most of the sudden changes in methane observed at Mace Head are attributable to significant changes in trajectory origin alone. The thin solid line has been obtained by integrating the Lagrangian chemistry model along each trajectory, using the climatological value for initialisation at $T = 5$, and plotting the modelled value of methane upon arrival. Most of the time the chemical model pulls the trace towards the observed values. A paper in preparation shows that the agreement between chemical observation and Lagrangian modelling can be improved still further using the origin-averaging techniques described in this atlas. Origin-averaging reduces the occurrence of narrow spikes in time traces associated with displacement errors of trajectories relative to emissions and the effects of sub-grid scale chemical fluxes and reactions. The focus is then drawn to significant changes in air origin (several kernel widths), although it is hoped that the statistics of the rapid concentration fluctuations in the model are in accord with atmospheric values so that the effects of nonlinear dependence of reaction on concentrations are not misrepresented.

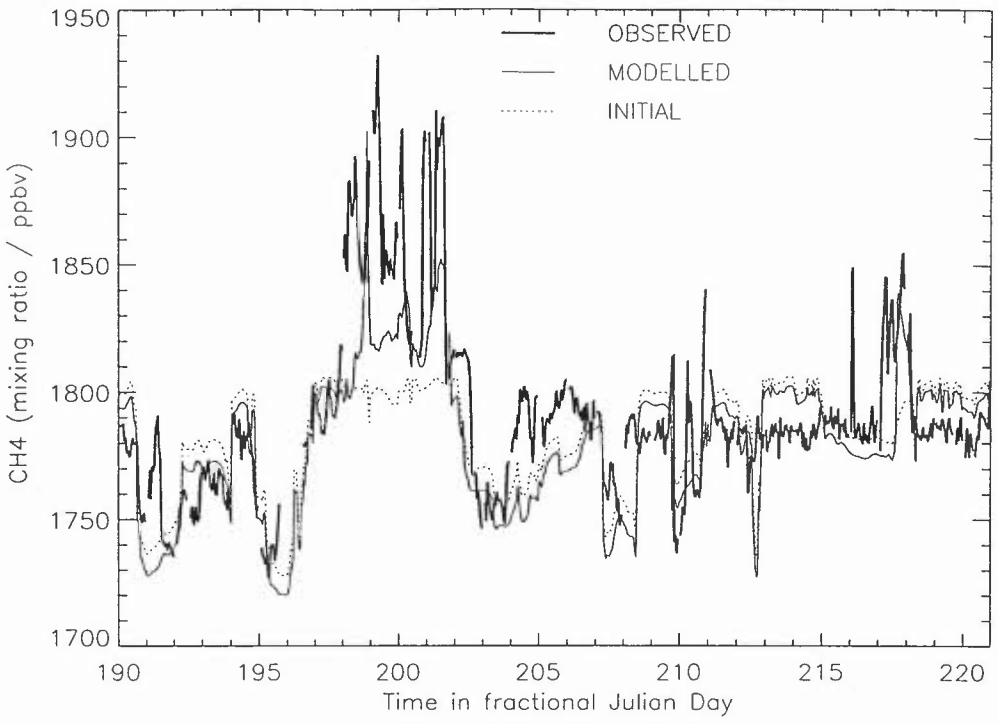


Figure 4: Time series of methane at Mace Head, Ireland during the EASE96 campaign. Thick line: observations. Dotted line: values obtained by following trajectories back 5 days from each time and taking the climatological value at the trajectory origin. Thin solid line: values obtained by integrating chemistry along each trajectory.

It is clear that the quality of the trajectories is sufficiently high that phase errors in the passing of air masses over Mace Head are of the order of an hour. These results are quite surprising bearing in mind the limitations on associating trajectories of the analysed flow with air parcel trajectories.

7 Synopsis

A Lagrangian climatology has been derived based on the analysis of trajectories arriving daily at low levels at three grids across the North Atlantic stormtrack. The annual cycle was found to be dominant in trajectory behaviour and therefore the climatology was stratified by arrival location and season. The 15 year ERA period was sufficiently long to resolve the annual peak clearly in the power spectra of Lagrangian diagnostics, and is therefore long enough for a representative climatology of the seasons. The overlap integral was devised to objectively identify seasons with trajectory origins that differ significantly from the climatology.

Lagrangian decorrelation timescales suggest that while 6 hours between analyses is good for determining the horizontal projection of 3D trajectories, it maybe too long to accurately locate the altitude, and especially potential temperature and specific humidity, of

individual trajectories. However, the measures presented here were found to be insensitive to trajectory errors and the sub-sampling of trajectory clusters when compared to the natural variability.

The decorrelation rate for horizontal velocity ($1.1\,\mathsf{day}^{-1}$) is faster than the stretching rate of line elements connecting trajectories ($0.3\,\mathsf{day}^{-1}$) suggesting that contours of a locally released tracer will tend to align rapidly with long-lived background tracers (Haynes and Anglade, 1997). In this regime, for integration lengths $T \gg \tau_r$ small errors in the initial shape of tracer contours cascade to such small scales that they become negligable compared to displacement errors from wind field truncation (Methven and Hoskins, 1999). This partly explains the success of the modelled time traces in Fig. 4 for example, even though initial conditions were only known on very large scales.

As trajectory lengths increase they become comparable to the timescale for changes in fields following trajectories. In the example of methane shown above, although it is long-lived the modelled changes over 5 days do have a significant effect on the time trace. Since the lifetime of this chemical is known to be long is it is reasonable to suppose that any change has been monotonic over the trajectory, although it may have been more rapid over a sector passing over emissions. It is then possible to associate a trajectory *history* with its *origin*. In the interpretation of the potential temperature changes in this atlas it is similarly assumed that, over a 5 day timescale, a negative origin-averaged θ-change is associated with heating along a trajectory although it may occur more rapidly during a segment of a trajectory. In this way, the useful length of trajectories is limited by interest in particular attributes. For instance, it would be fruitless to examine θ-changes over 30 day trajectories because this length exceeds even the slowest timescales associated with radiative heating. Moreover, since θ and q changes are also associated with passing through convective regions or the boundary layer, and these events are connected with vertical motions in baroclinic waves, over 10 days trajectories may experience several episodes of heating/cooling and water vapour gain/loss and the association of origin-averaged changes with trajectory history is lost.

References

Cocke, W. (1969). Turbulent hydrodynamic line stretching: Consequences of isotropy. *Phys.Fluids*, **12**, 2488–2492.

Derwent, R., Simmonds, P., and Collins, W. (1994). Ozone and carbon monoxide measurements at a remote maritime location, Mace Head, Ireland, from 1990–1992. *Atmos.Environment*, **28**, 2623–2637.

ECMWF Research Manual 2 (1988). *ECMWF Forecast Model: Adiabatic Part.*

Evans, M., Shallcross, D., Law, K., Wild, J., Simmonds, P., Spain, T., Berrisford, P., Methven, J., Lewis, A., McQuaid, J., Pilling, M., Bandy, B., Penkett, S., and Pyle, J. (1999). Validation of a Lagrangian box model using field measurements from EASE 1996. *Atmos.Environment*, **submitted**.

Gibson, J., Kallberg, P., Uppala, S., Hernandez, A., Nomura, A., and Serrano, E. (1997). ERA description. Technical Report 1, ECMWF re-analysis.

Gregory, D. and Miller, M. (1989). A numerical study of the parametrization of deep tropical convection. *Quart.J.Roy.Met.Soc.*, **115**, 1209–1241.

Haynes, P. and Anglade, J. (1997). The vertical-scale cascade in atmospheric tracers due to large-scale differential advection. *J.Atmos.Sci.*, **54**, 1121–1136.

Hodges, K. (1996). Spherical nonparametric estimators applied to the UGAMP model integration for AMIP. *Mon.Wea.Rev.*, **124**, 2914–2932.

Law, K. and Nisbet, E. (1996). Sensitivity of the methane growth rate to changes in methane emissions from natural gas and coal. *J.Geophys.Res.*, **101**, 14387–14397.

Law, K. and Pyle, J. (1993). Modelling trace gas budgets in the troposphere. 1: Ozone and odd nitrogen. *J.Geophys.Res.*, **98**, 18377–18400.

Maryon, R. and Buckland, A. (1995). Tropospheric dispersion: the first ten days after a puff release. *Quart.J.Roy.Met.Soc.*, **121**, 1799–1833.

Merrill, J. (1994). Isentropic airflow probability analysis. *J.Geophys.Res.*, **99(D12)**, 25881–25889.

Methven, J. (1997). Offline trajectories: calculation and accuracy. Technical Report 44, UGAMP (UK Universities Global Atmospheric Modelling Programme).

Methven, J. and Hoskins, B. (1999). On the advection of high resolution tracers by low resolution winds. *J.Atmos.Sci.*, **in press**.

Ngan, K. and Shepherd, T. (1999). A closer look at chaotic advection in the stratosphere. Part I: Geometric structure. *J.Atmos.Sci.*, **submitted**.

O'Neill, A., Grose, W., Pope, V., MacLean, H., and Swinbank, R. (1994). Evolution of the stratosphere during Northern winter 1991/92 as diagnosed from U.K.M.O. analyses. *J.Atmos.Sci.*, **51**, 2800–2817.

Pierrehumbert, R. (1991). Large-scale horizontal mixing in planetary atmospheres. *Phys.Fluids*, **3A**, 1250–1260.

Pierrehumbert, R. and Yang, H. (1993). Global chaotic mixing on isentropic surfaces. *J.Atmos.Sci.*, **50**, 2462–2480.

Press, W., Flannery, B., Teukolsky, S., and Vetterling, W. (1992). *Numerical Recipes: The Art of Scientific Computing.* Cambridge. 702 pp.

Silverman, B. (1986). *Density estimation for statistics and data analysis.* Chapman and Hall.

Simmons, A. and Burridge, D. (1981). An energy and angular-momentum conserving vertical finite-difference scheme in hybrid vertical coordinates. *Mon.Wea.Rev.*, **109**, 758–766.

Sparling, L., Kettleborough, J., Haynes, P., McIntyre, M., Rosenfield, J., Schoeberl, M., and Newman, P. (1997). Diabatic cross-isentropic dispersion in the lower stratosphere. *J.Geophys.Res.*, **102(D22)**, 25817–25829.

Sutton, R., MacLean, H., Swinbank, R., and O'Neill, A. (1994). High-resolution stratospheric tracer fields estimated from satellite observations using Lagrangian trajectory calculations. *J.Atmos.Sci.*, **51**, 2995–3005.

Valdes, P. and Hoskins, B. (1989). Linear stationary wave simulations of the time-mean climatological flow. *J.Atmos.Sci.*, **46**, 2509–2527.

Varosi, F., Antonsen, T., and Ott, E. (1991). The spectrum of fractal dimensions of passively convected scalar gradients in chaotic fluid flows. *Phys.Fluids*, **A3**, 1017–1028.

A Glossary of Terms

t	time, especially trajectory arrival time
T	trajectory length (magnitude)
δt	trajectory integrator time-step
Δt	interval between wind records
z	any variable
z_i	value of z along trajectory i
$z[T]$	z evaluated at T
z_0	z at arrival time ($T = 0$)
$\dot{z}$	Lagrangian derivative of z (following resolved winds)
z'	change in z along trajectory ($z[-T] - z[0]$)
$\langle z \rangle$	mean of z over cluster
δz	deviation from cluster mean
$\overline{z}$	mean of z over arrival dates
$\overline{z}(\mathbf{x})$	origin averaged value of z
$\sigma_z(\mathbf{x})$	standard deviation in z about origin averaged value
σ_z	standard deviation of z-component of velocity about cluster mean
τ_z	Lagrangian decorrelation timescale for z
a	Earth's radius
λ	longitude OR finite time Lyapunov exponent
ϕ	latitude
η	hybrid pressure-sigma vertical coordinate
p	pressure
θ	potential temperature
q	specific humidity
$\mathbf{x}$	unit radial position vector for a point on the sphere
r_{ij}	interparticle great circle separation
K	kernel function for density estimates on the sphere
C	smoothing parameter
n	number of particles used for density estimate
ρ	density of origin (normalised)
$\mathcal{O}$	overlap integral
D	horizontal cluster displacement
R	RMS great circle spread of a cluster

Y	RMS meridional spread
X	RMS zonal spread
H	correlation function
d	correlation dimension
α	range of scales used to estimate d
W	time window width
N	number of discrete points within W
M	number of windows used for whole ERA series
s	time exponent for cluster spread
$\mathcal{L}$	Lagrangian autocorrelation function

MAM	March, April, May
JJA	June, July, August
SON	September, October, November
DJF	December, January, February

BADC	British Atmospheric Data Centre
EASE96	Eastern Atlantic Summer Experiment 1996
ECMWF	European Centre for Medium-Range Weather Forecasts
ERA	ECMWF Re-Analysis 15 year dataset (Jan 79 to Feb 94)
GCM	Global Circulation Model
UKMO	United Kingdom Meteorological Office
UGAMP	U.K. Universities Global Atmospheric Modelling Programme

B Description of Figures

B.1 Seasonal Density of Origin

The density of origin, 5 days before arrival, is shown in a region versus season matrix for each year of the ERA-15 period. The arrival regions are given in Fig. 1 (the full European grid is used for these calculations). The outermost contour value is $0.5\,\mathrm{ster}^{-1}$ and the other contour values increase at an interval of $0.5\,\mathrm{ster}^{-1}$. A polar stereographic projection has been used and the graticules have a spacing of 15° in longitude and latitude.

B.2 Overlap Integrals

The overlap of each season's density of origin with the climatological estimate is shown as a function of the year since 1979, in a region versus season matrix. The solid lines are for the $T = 5$ days density of origin and the dashed lines for $T = 2$ days.

B.3 Time Series

The time series of Lagrangian measures relating to the horizontal dispersal of the daily cluster arrivals are shown for each arrival region at a trajectory length $T = 2$ days and then for each region at $T = 5$ days. The diagnostics for the sparse European cluster are shown so that the arrival density is the same as for the American and mid-Atlantic clusters.

Time series are then shown for the Lagrangian measures relating to vertical motion, heating and moisture change. The ordering for cluster region and trajectory length is the same as before and the mean and standard deviation for each 15 year time series is shown above the plot. Note that the time series for the specific humidity changes stop 6 months short due to an error in the attribution of q for this final period.

B.4 Power Spectra

Power spectra for the ERA time series of each Lagrangian measure were calculated by averaging power spectral densities for overlapping 5 year windows (see Section 4.2.2). They are shown for the American, mid-Atlantic and sparse European clusters for $T = 5$ days. The lowest frequency bin corresponds to a period of 5 years and the highest to a period of 2 days and the frequency interval is $5.482 \times 10^{-4}\,\mathsf{days}^{-1}$. Both axes are shown with log scales and the units of power are km^2 for the spreads and horizontal displacement, days^{-2} for the median Lyapunov exponent, hPa^2 for pressure displacement, K^2 for θ displacement and $\mathsf{g}^2\mathsf{kg}^{-2}$ for q displacement. Since the trajectory length is 5 days but the clusters arrive daily there is a degree of similarity in the conditions experienced by consecutive cluster arrivals which results in the steep drop in power for frequencies $> 0.1\,\mathsf{days}^{-1}$.

B.5 Histograms for the Lagrangian Measures

The values for the Lagrangian measures are collated for all four seasons, using data from the whole ERA period, and shown in histograms of the number of arrivals per bin versus bin value. The histograms are shown in region versus season matrices, for trajectory lengths of 2 and 5 days, for RMS spread, horizontal displacement, median Lyapunov exponent plus pressure, potential temperature and specific humidity displacements. The median value for each distribution is plotted in the top right corner of each panel. Note that the Lyapunov exponents are already the median values taken from the histogram of all nearest neighbour trajectory pairs for each cluster arrival and thus the climatological histogram shown here represents the distribution of median values from individual realisations. In the European panels the heavy solid line is obtained for the sparse cluster, the light solid line for the compact cluster and the dotted line for the full European cluster.

B.6 Histograms for the Time Dependence of Cluster Spread

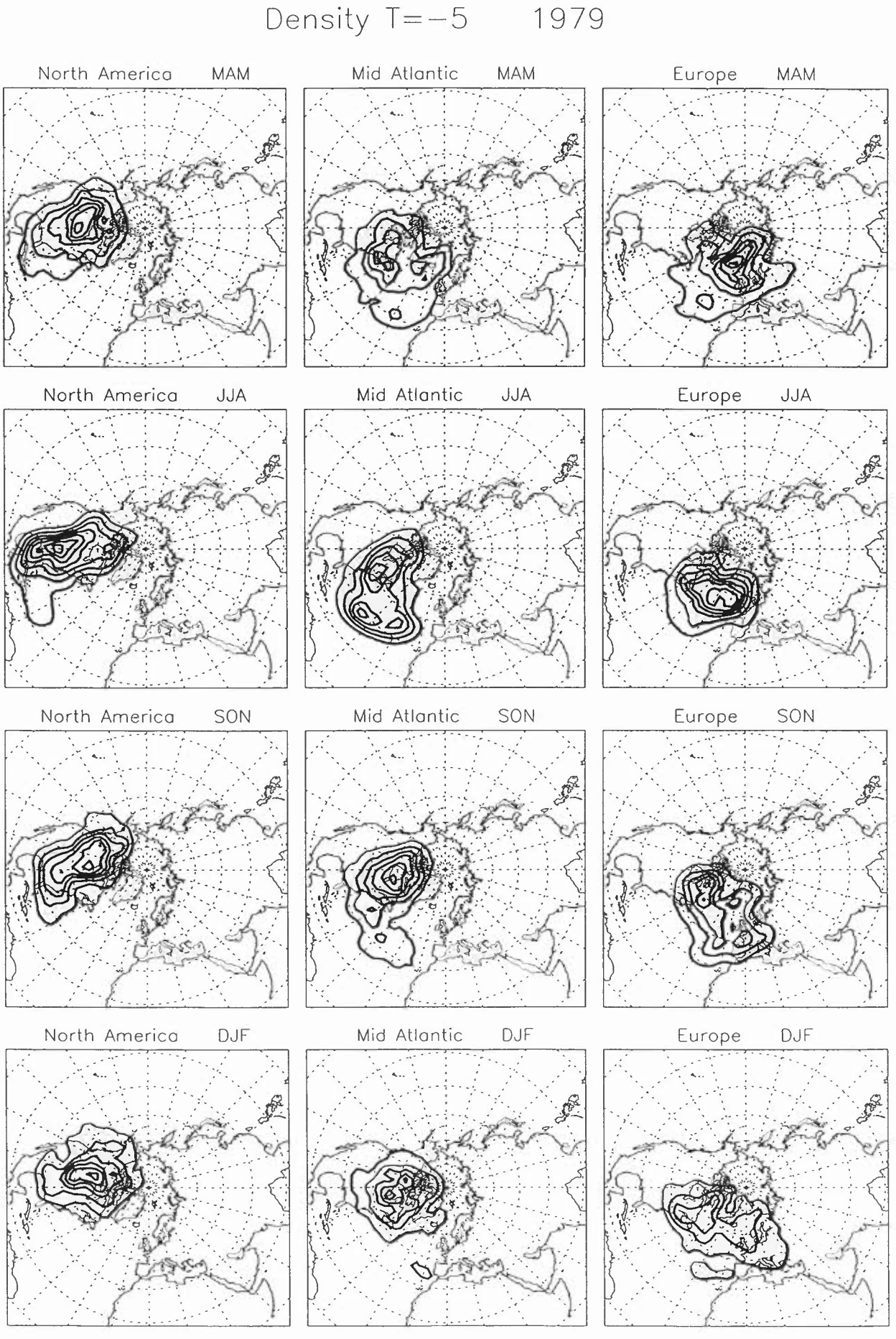
The dependence of cluster spread on trajectory length was fitted to the power law $\langle[\delta z']^2\rangle^{\frac{1}{2}} \sim T^s$ for every cluster arrival, for the values of spreads at 6 hr intervals in T. The resulting exponents, s, have been collated for each of the four seasons to form histograms, shown in a region versus season matrix as above. The exponents have been fitted for the time dependence of R, Y, X, $\langle[\delta p']^2\rangle^{\frac{1}{2}}$, $\langle[\delta\theta']^2\rangle^{\frac{1}{2}}$ and $\langle[\delta q']^2\rangle^{\frac{1}{2}}$ as defined in Section 4.1.

B.7 Lagrangian Decorrelation Timescales

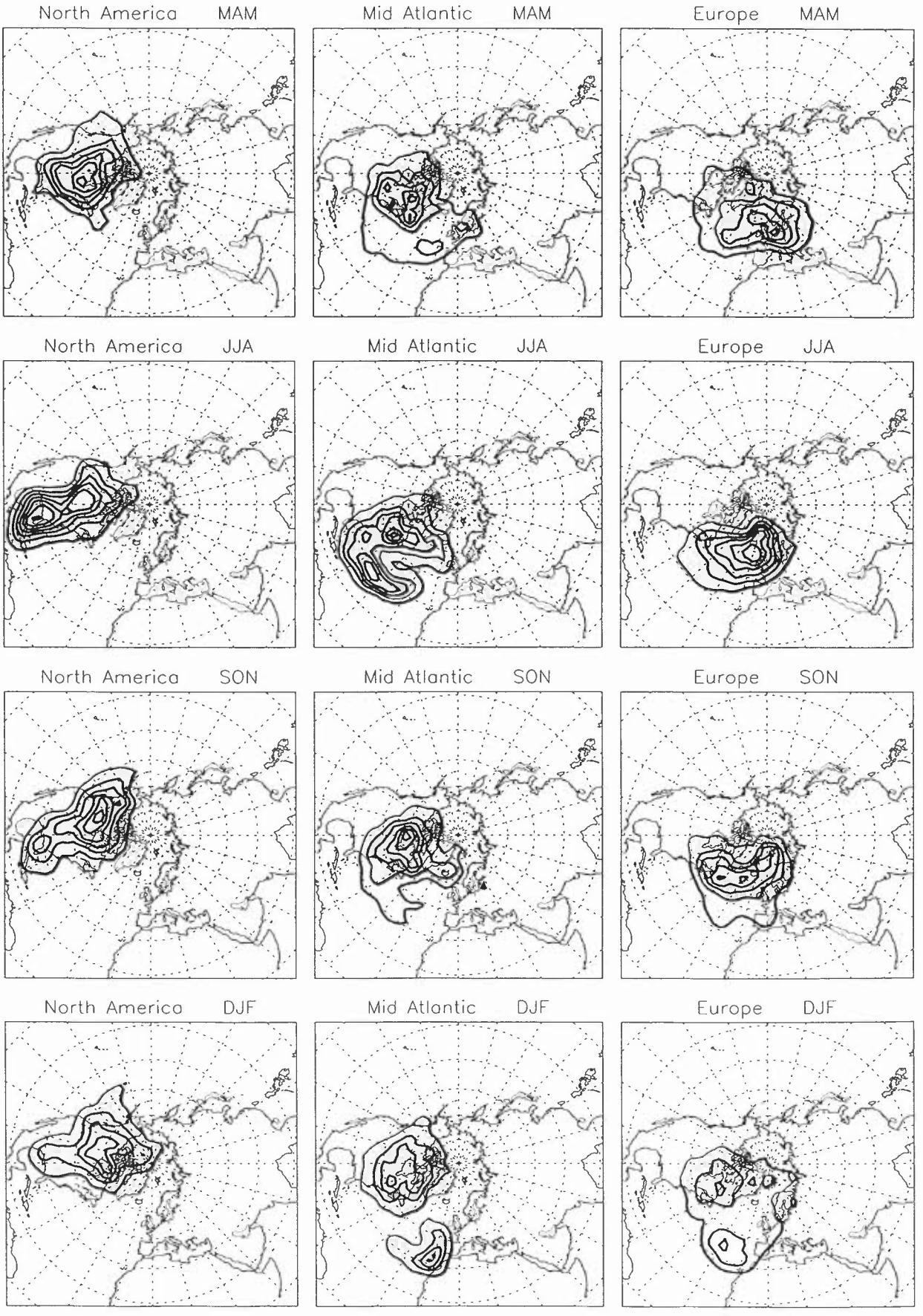
The decorrelation timescales for horizontal velocity, vertical velocity, $\dot{\theta}$ and $\dot{q}$ have been calculated from the seasonally averaged Lagrangian autocorrelation functions and are shown as a function of year since MAM79. The timescales are shown for each arrival region in turn and the mean and standard deviation are given at the top of each panel. For the European plots, the solid line was obtained for the full European cluster (although the sparse cluster gave almost identical values) and the dotted line was obtained for the compact cluster.

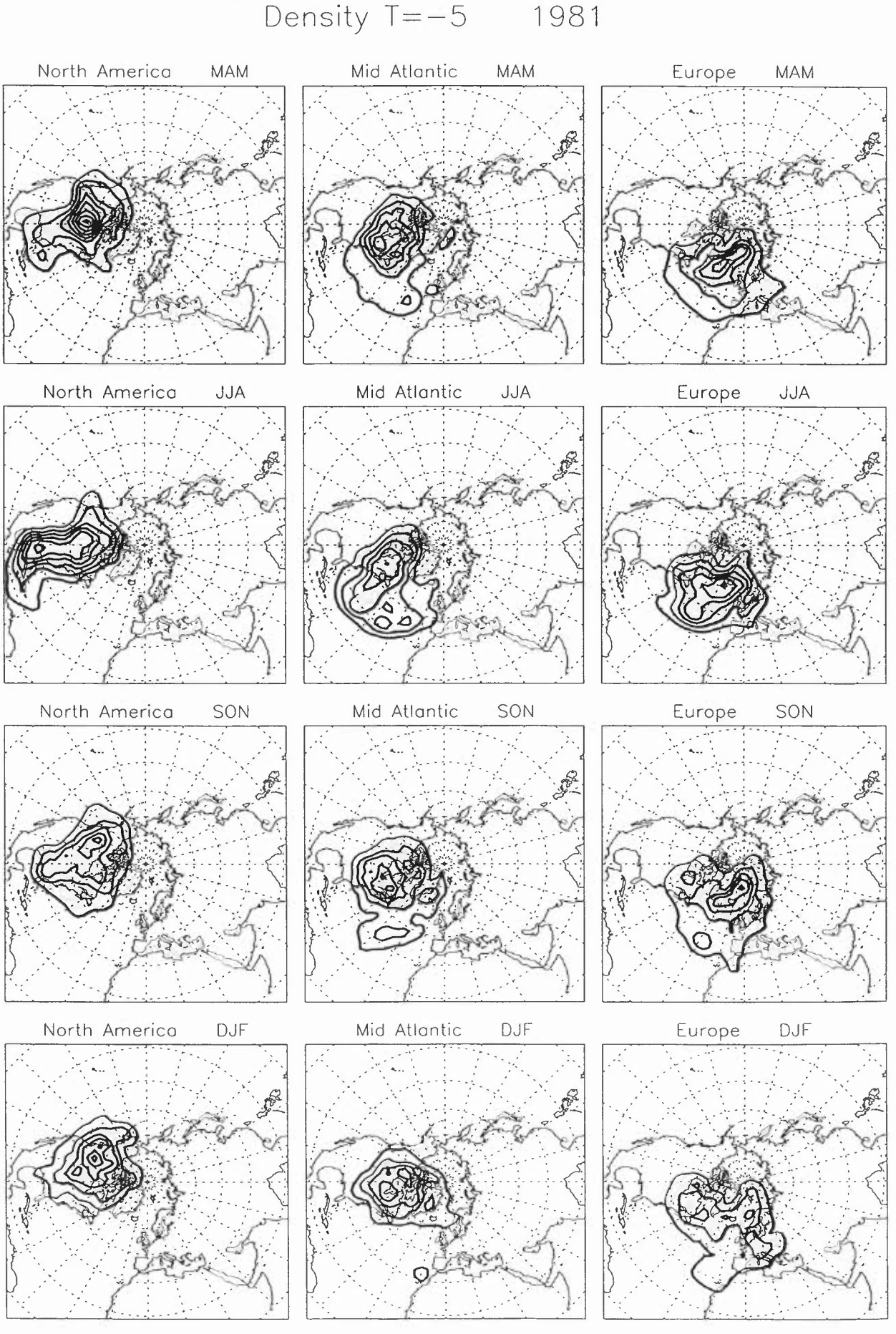
B.8 The Lagrangian Climatology

The colour plots at the end of the atlas show the climatologies of trajectory origin for each season using all trajectories arriving at the American, mid-Atlantic and (full) European grids (see Fig. 1). The map projection shown is the same as for the seasonal densities of origin. On each page there are 6 fields: the density of origin 5 days before arrival ($T = 5$) and the density of origin at $T = 2$, followed by the mean pressure, standard deviation in pressure, mean θ-change and mean q-change obtained from kernel regression techniques. A linear kernel with an arc band width of 5.72° was used in all estimates (Hodges, 1996) and the mean and standard deviation fields are only shown where the $T = 5$ density exceeds $0.5\,\mathsf{ster}^{-1}$. The units for density of origin, $\overline{p}$, σ_p, $\overline{\theta'}$ and $\overline{q'}$ are ster^{-1}, hPa, hPa, K and $\mathsf{g\,kg}^{-1}$ respectively. Note that all trajectories arrive at 900 hPa and that the sign of the θ and q changes is such that *negative* values correspond to net heating and moistening following trajectories over the last 5 days before arrival.

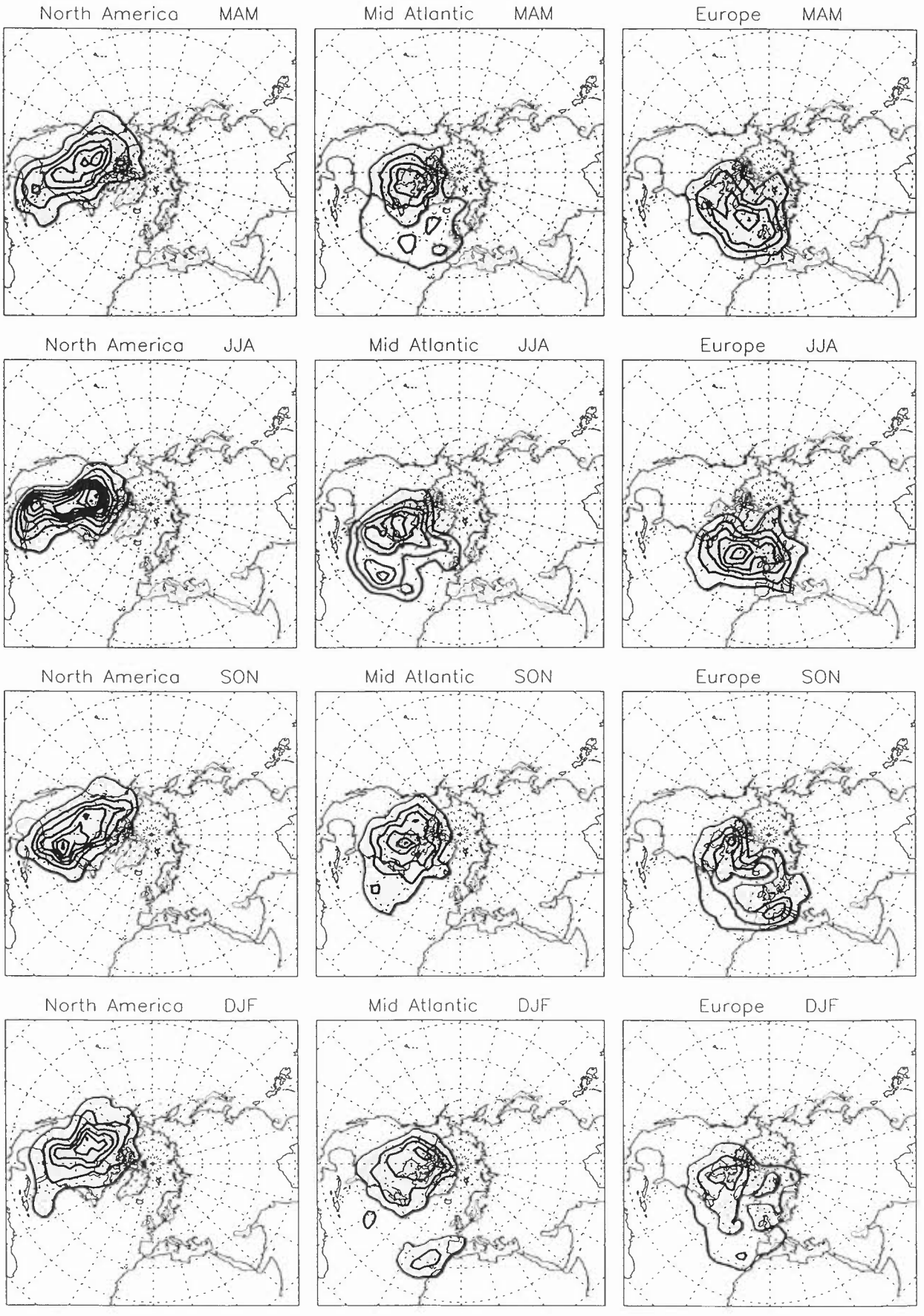

Density T=−5 1979
North America MAM
Mid Atlantic MAM
Europe MAM
North America JJA
Mid Atlantic JJA
Europe JJA
North America SON
Mid Atlantic SON
Europe SON
North America DJF
Mid Atlantic DJF
Europe DJF

Density T=−5 1980

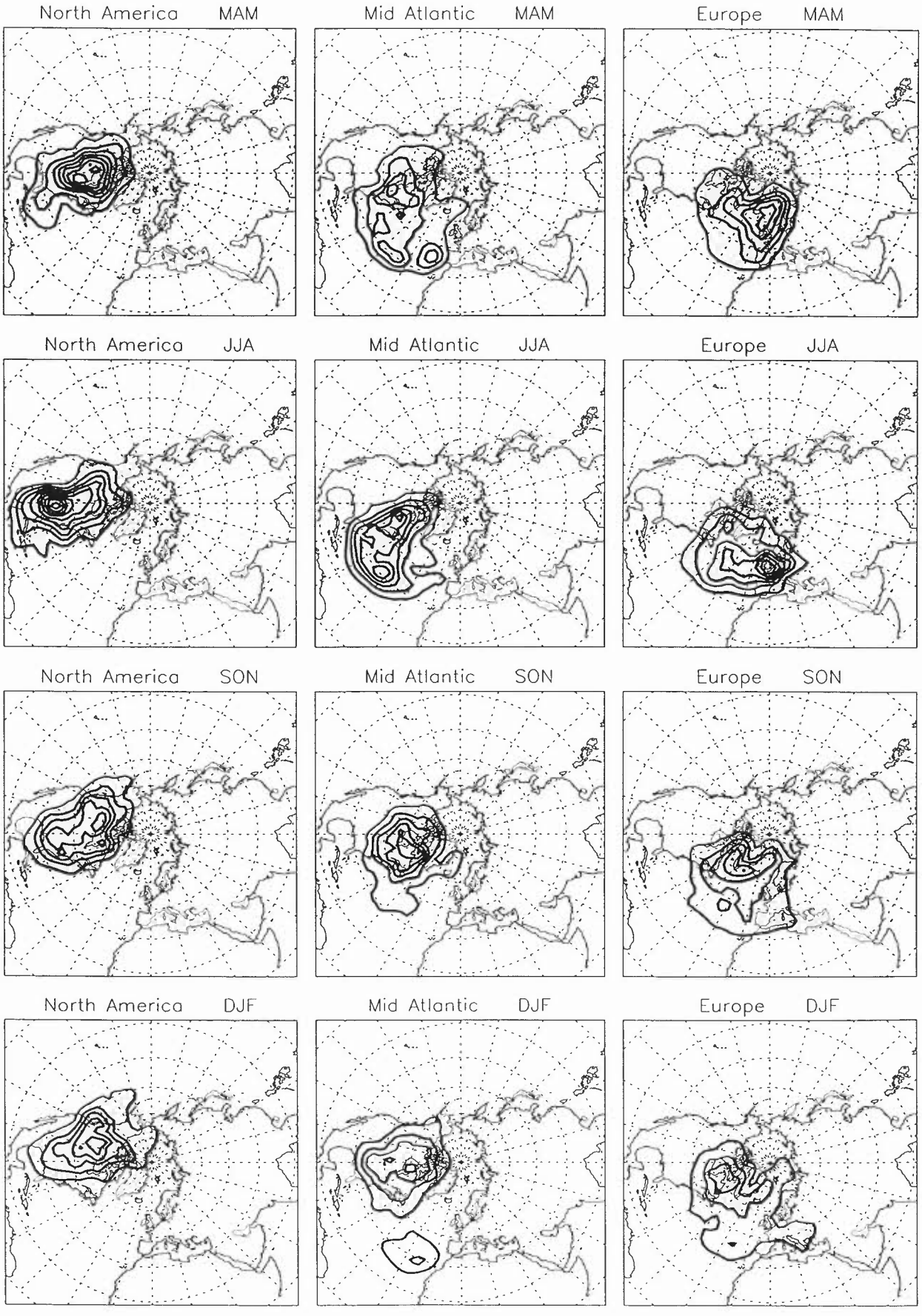



Density T=−5 1981
North America MAM
Mid Atlantic MAM
Europe MAM
North America JJA
Mid Atlantic JJA
Europe JJA
North America SON
Mid Atlantic SON
Europe SON
North America DJF
Mid Atlantic DJF
Europe DJF

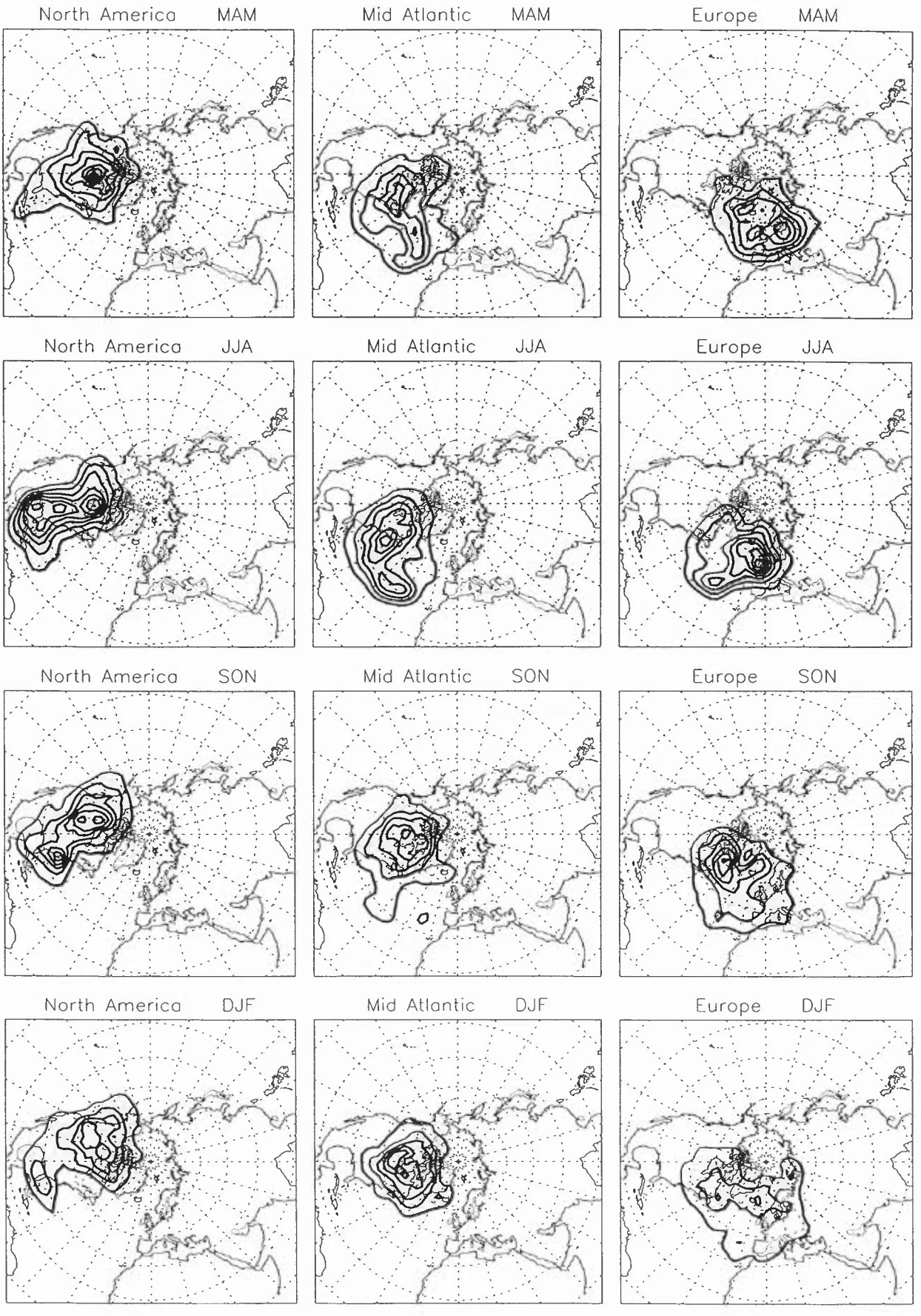
Density T=−5 1982

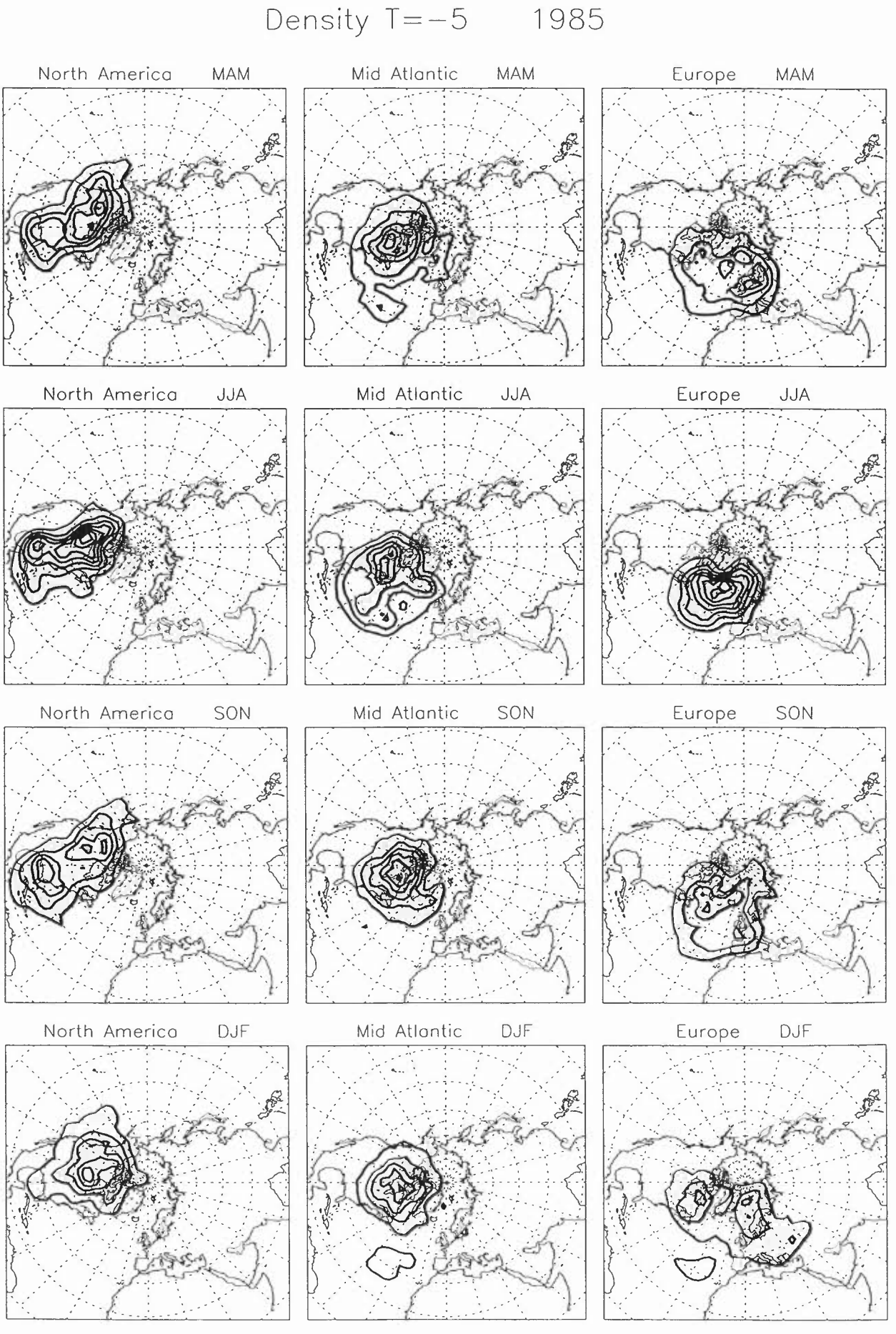


Density T=−5 1983

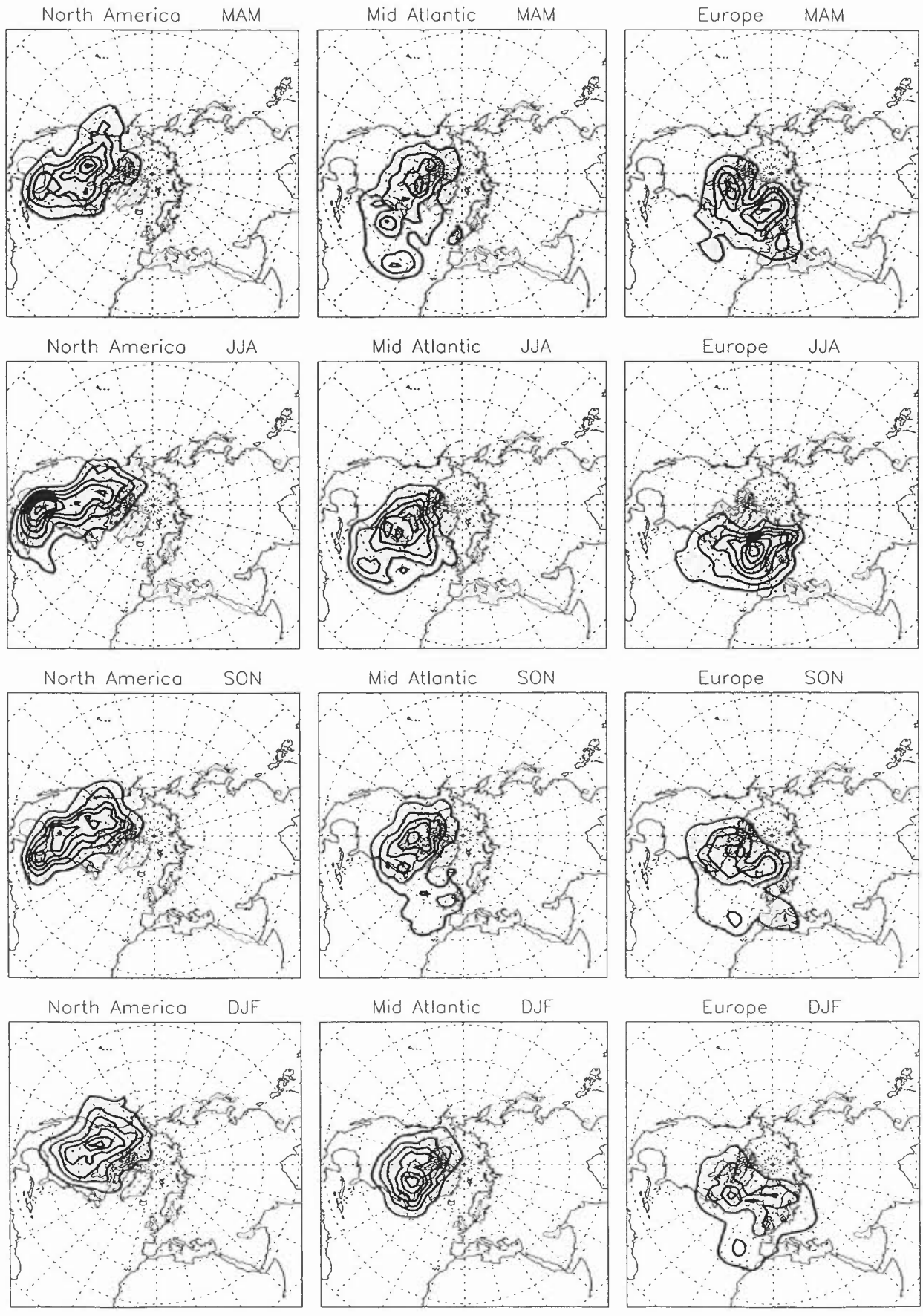


Density T=−5 1984

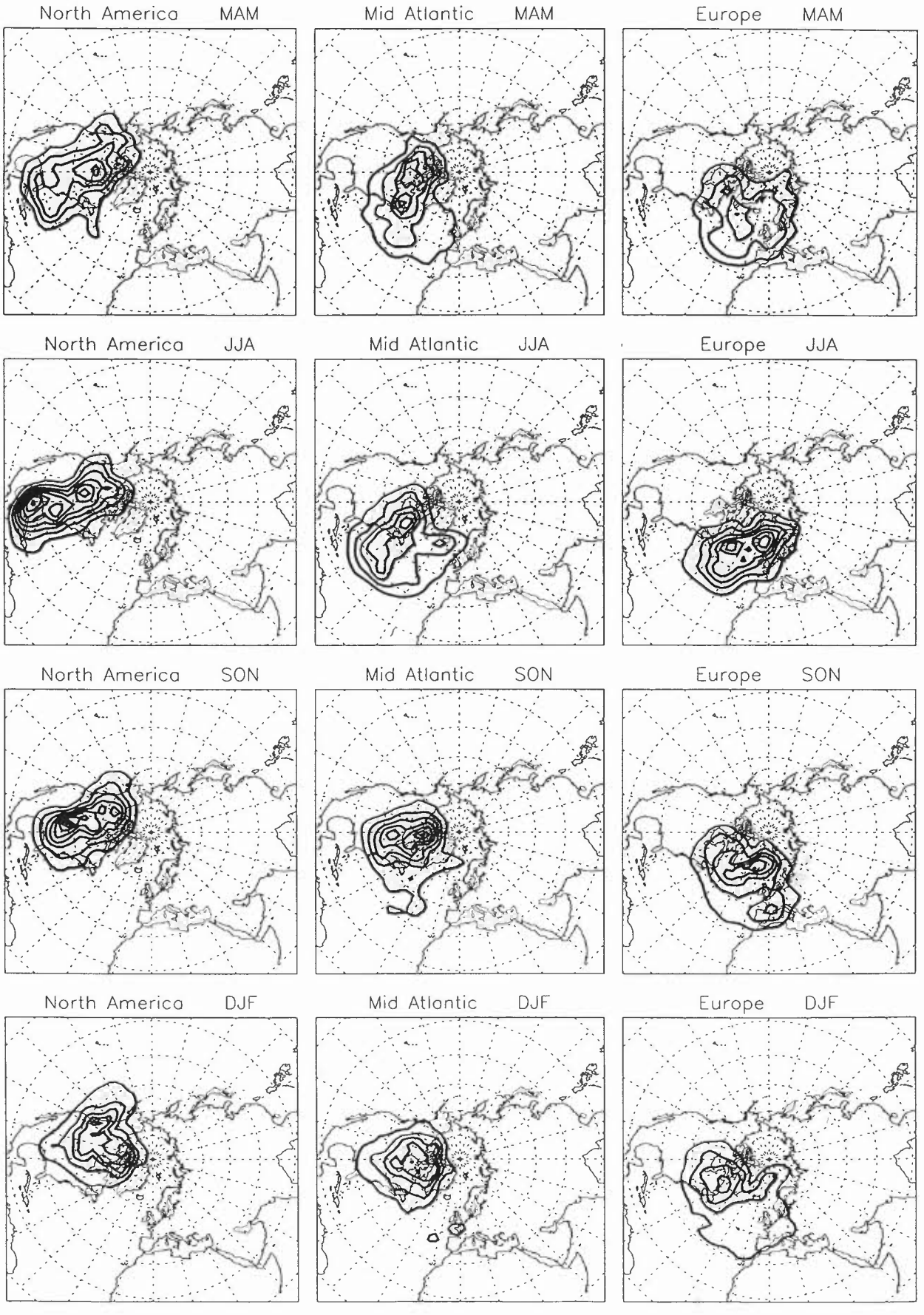



Density T=-5 1985
North America MAM
Mid Atlantic MAM
Europe MAM
North America JJA
Mid Atlantic JJA
Europe JJA
North America SON
Mid Atlantic SON
Europe SON
North America DJF
Mid Atlantic DJF
Europe DJF

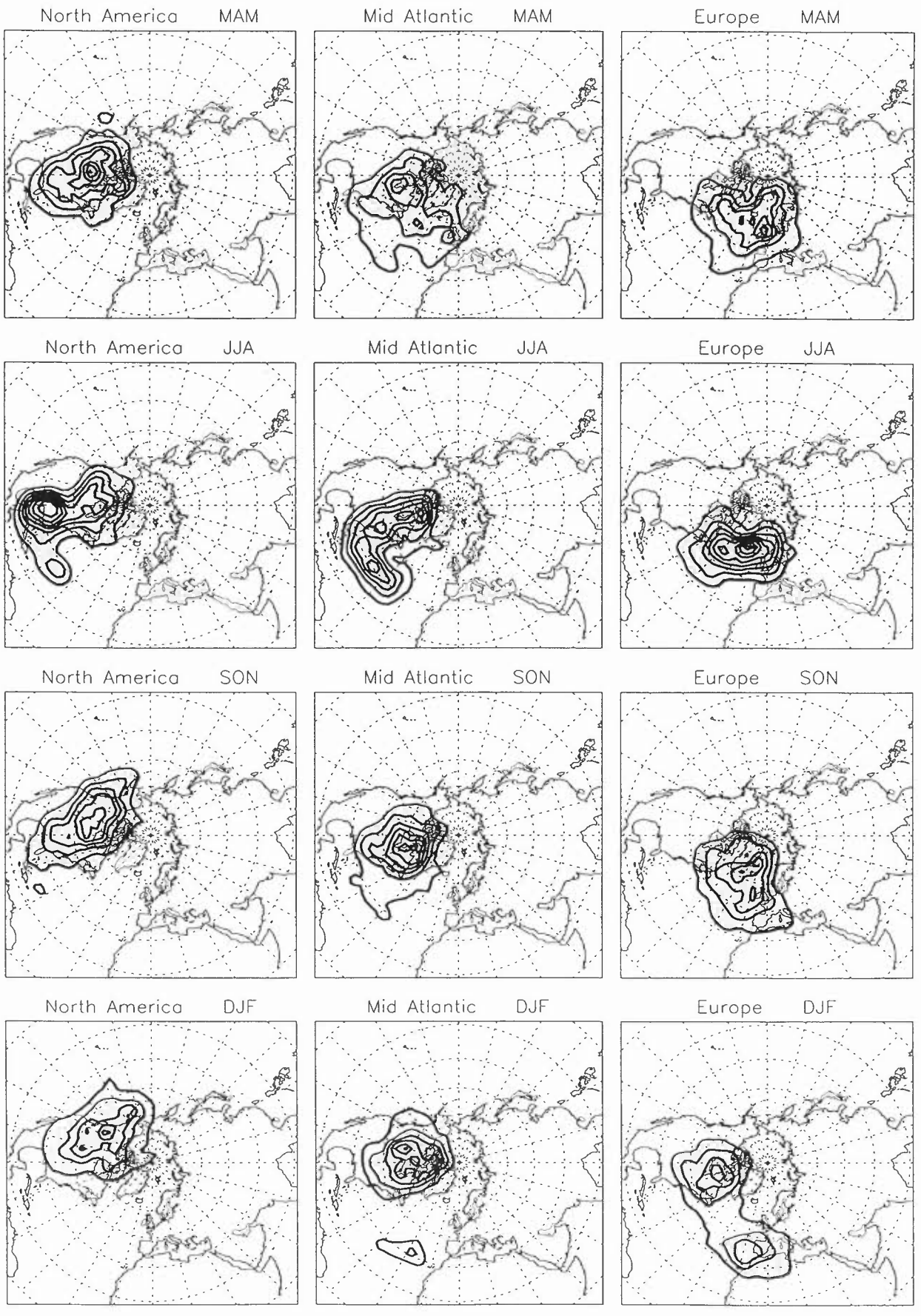
Density T=−5 1986



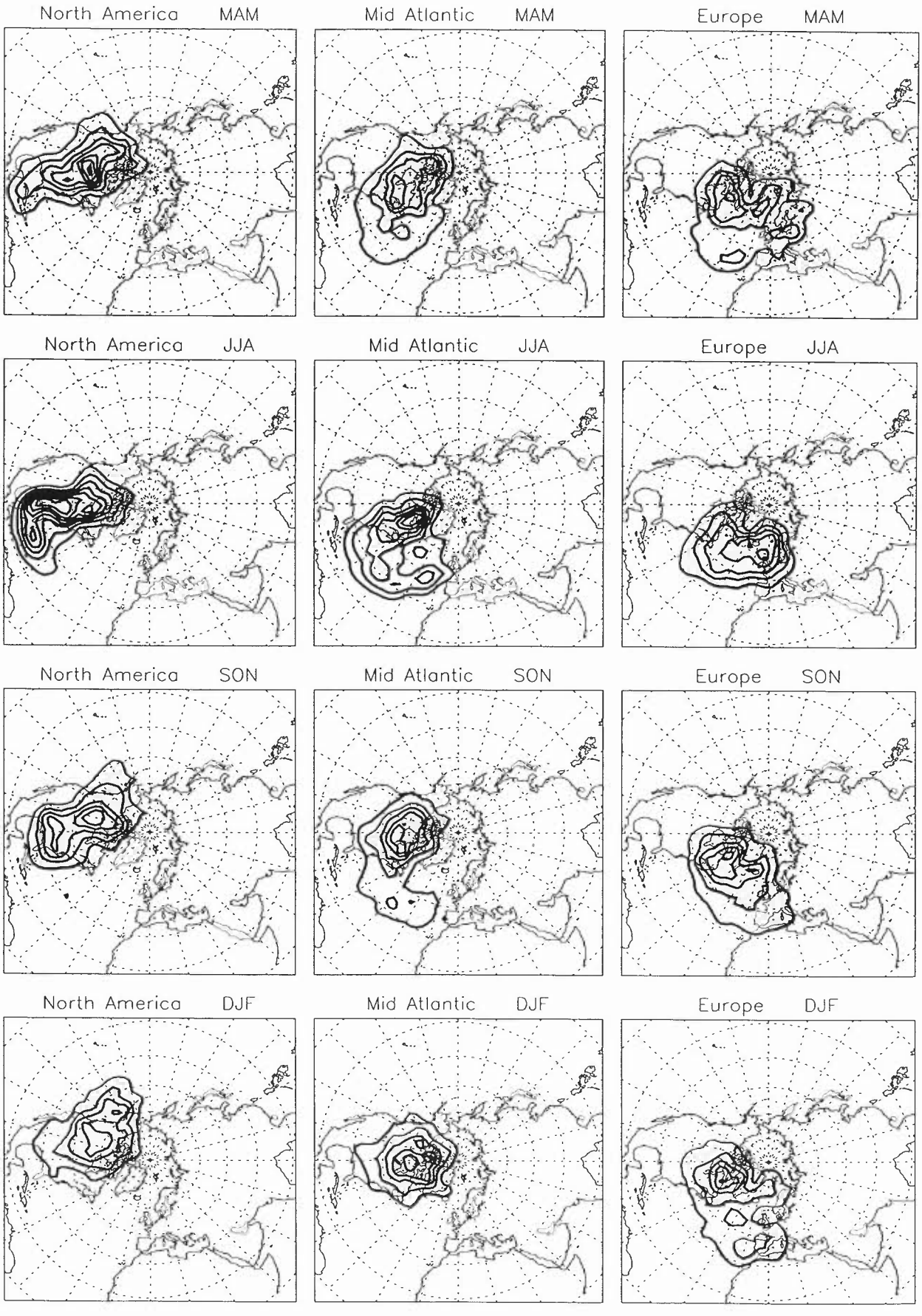
Density T=−5 1987



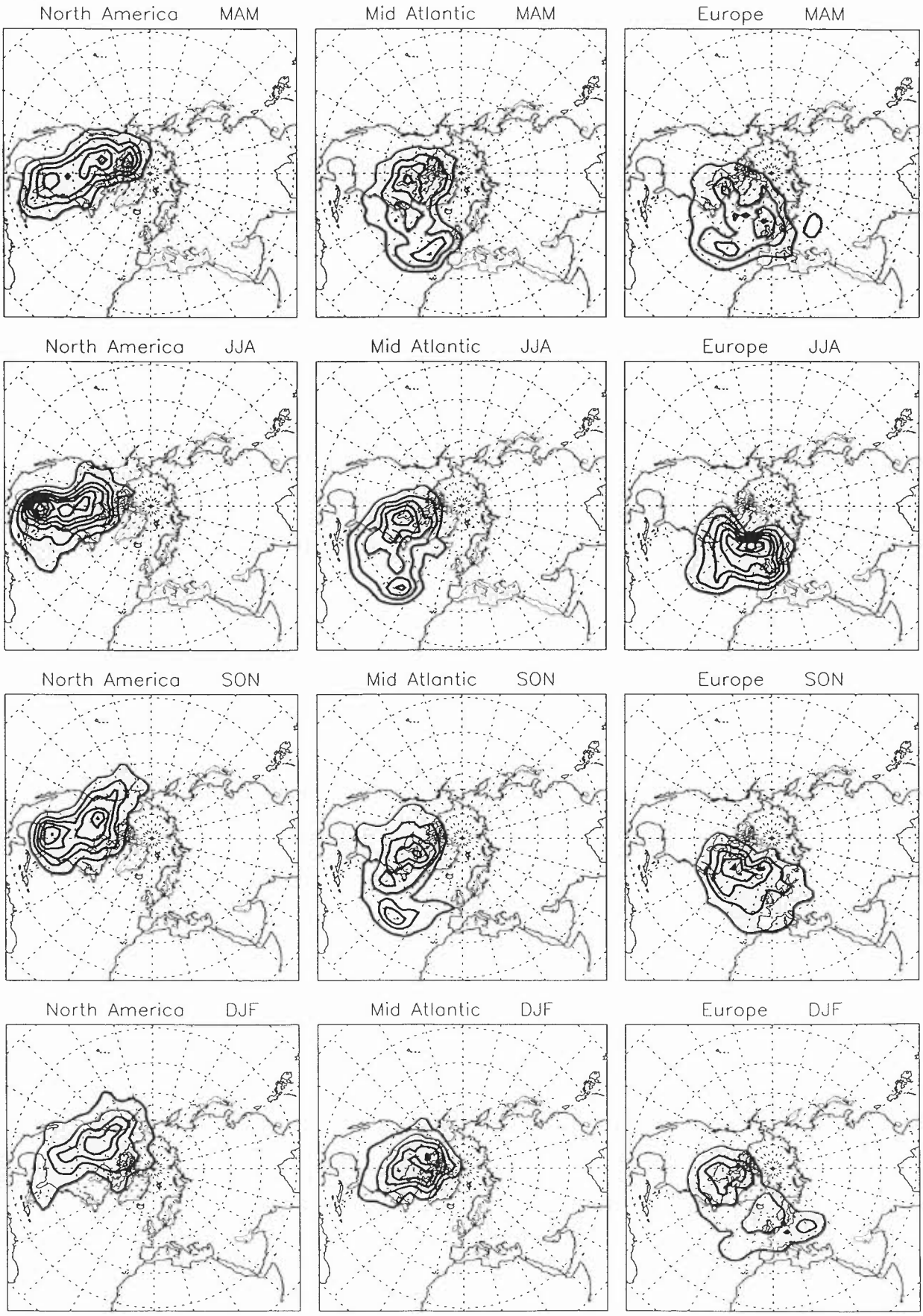
Density T=−5 1988



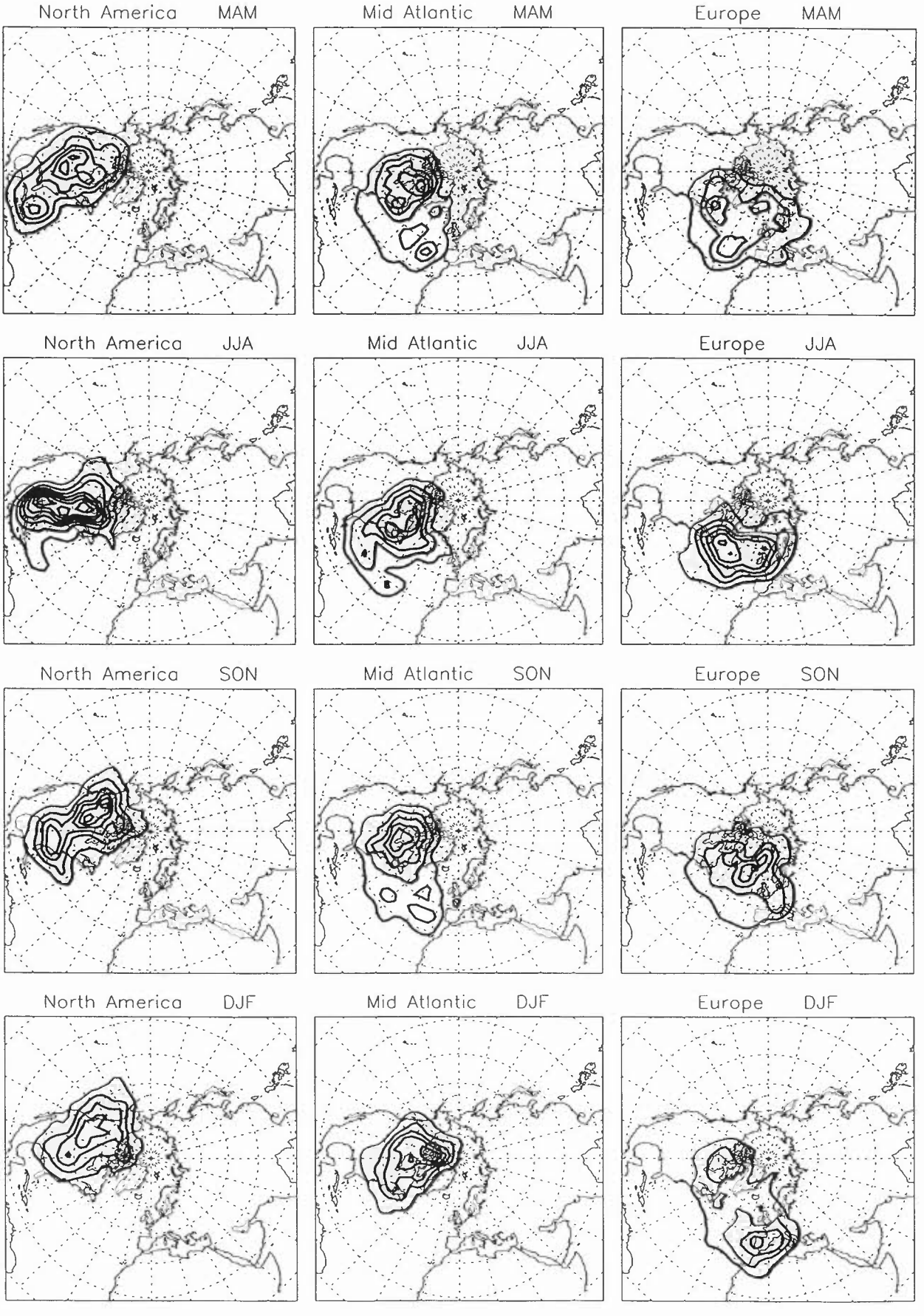
Density T=−5 1989



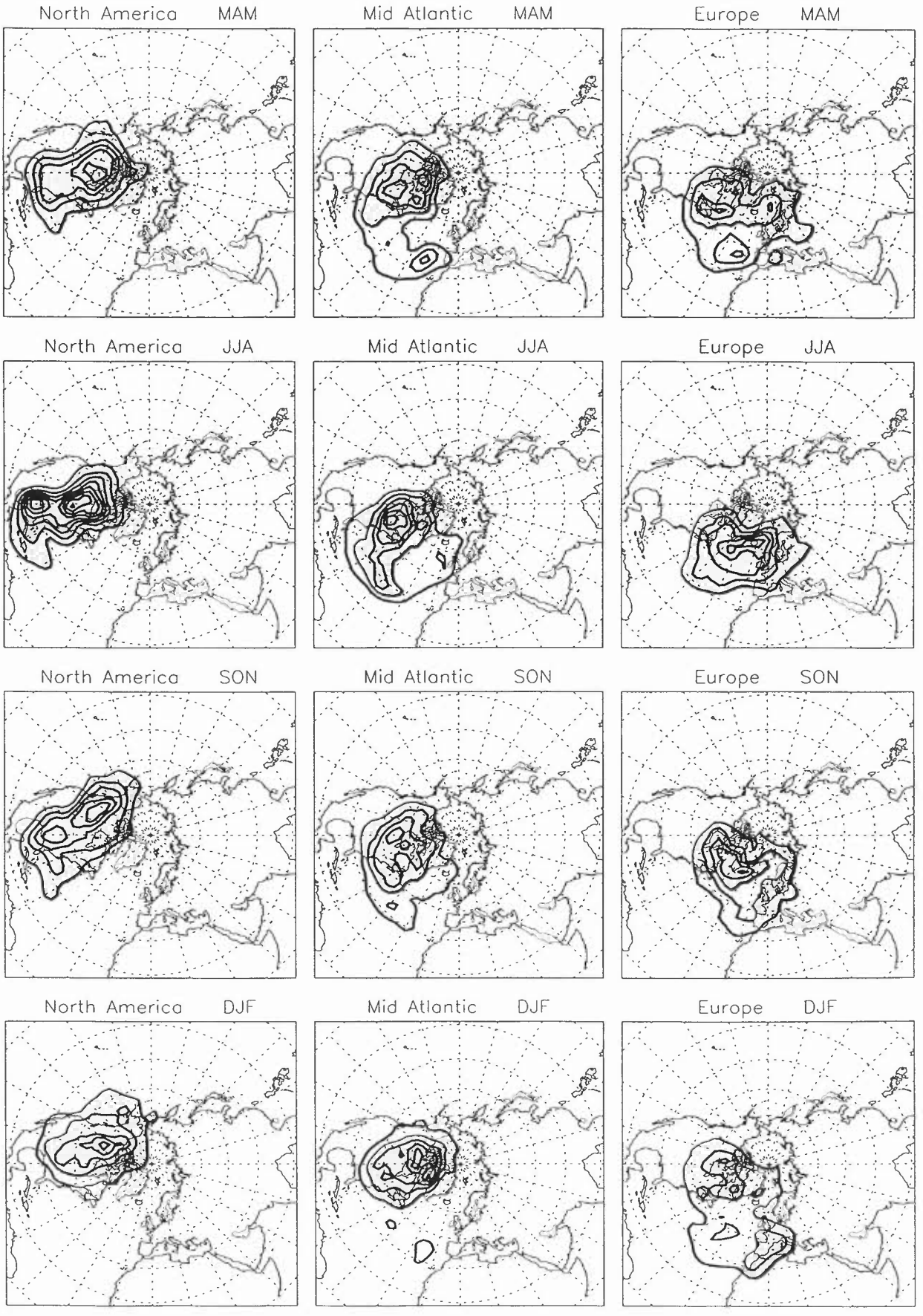
Density T=-5 1990



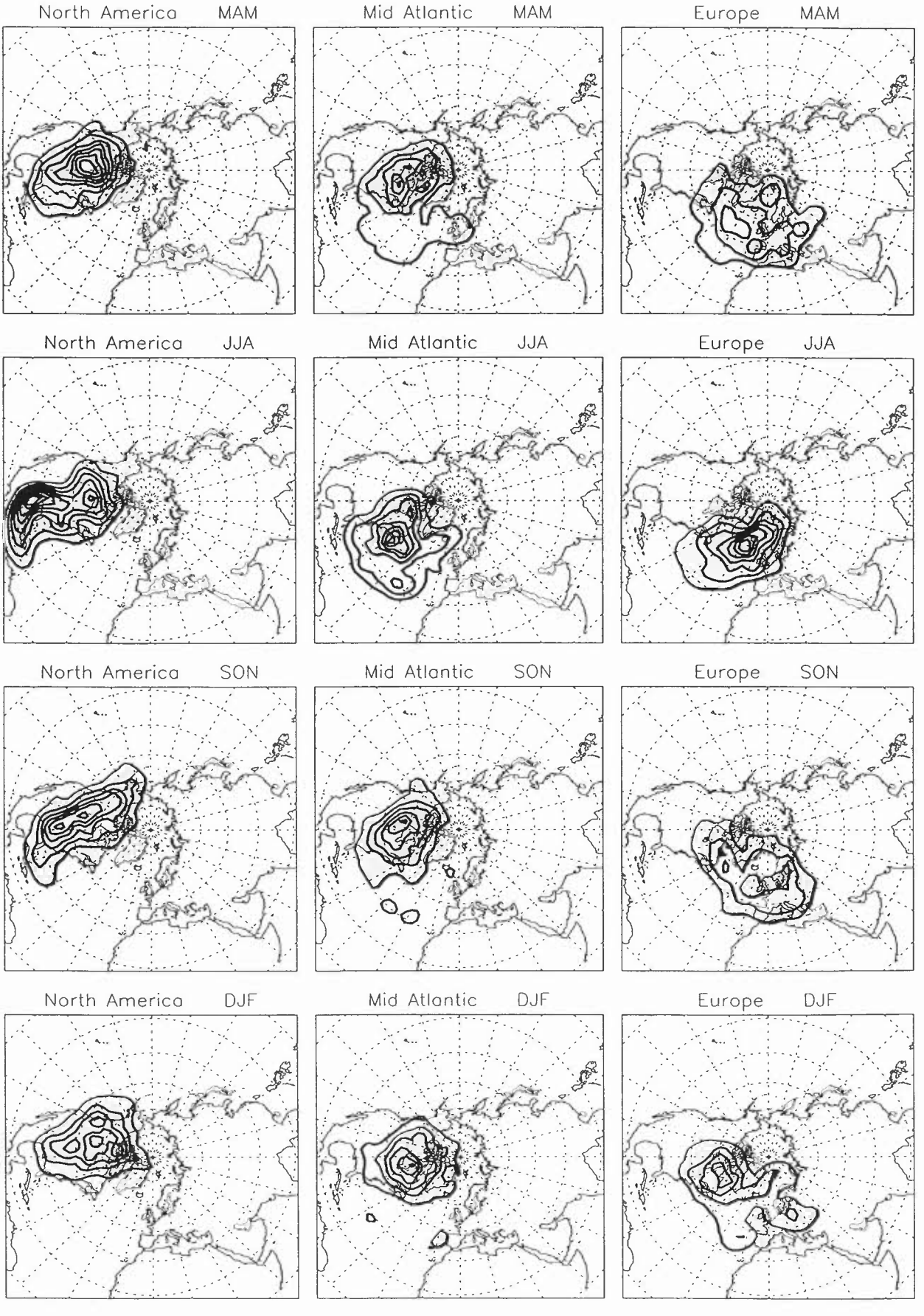
Density T=−5 1991



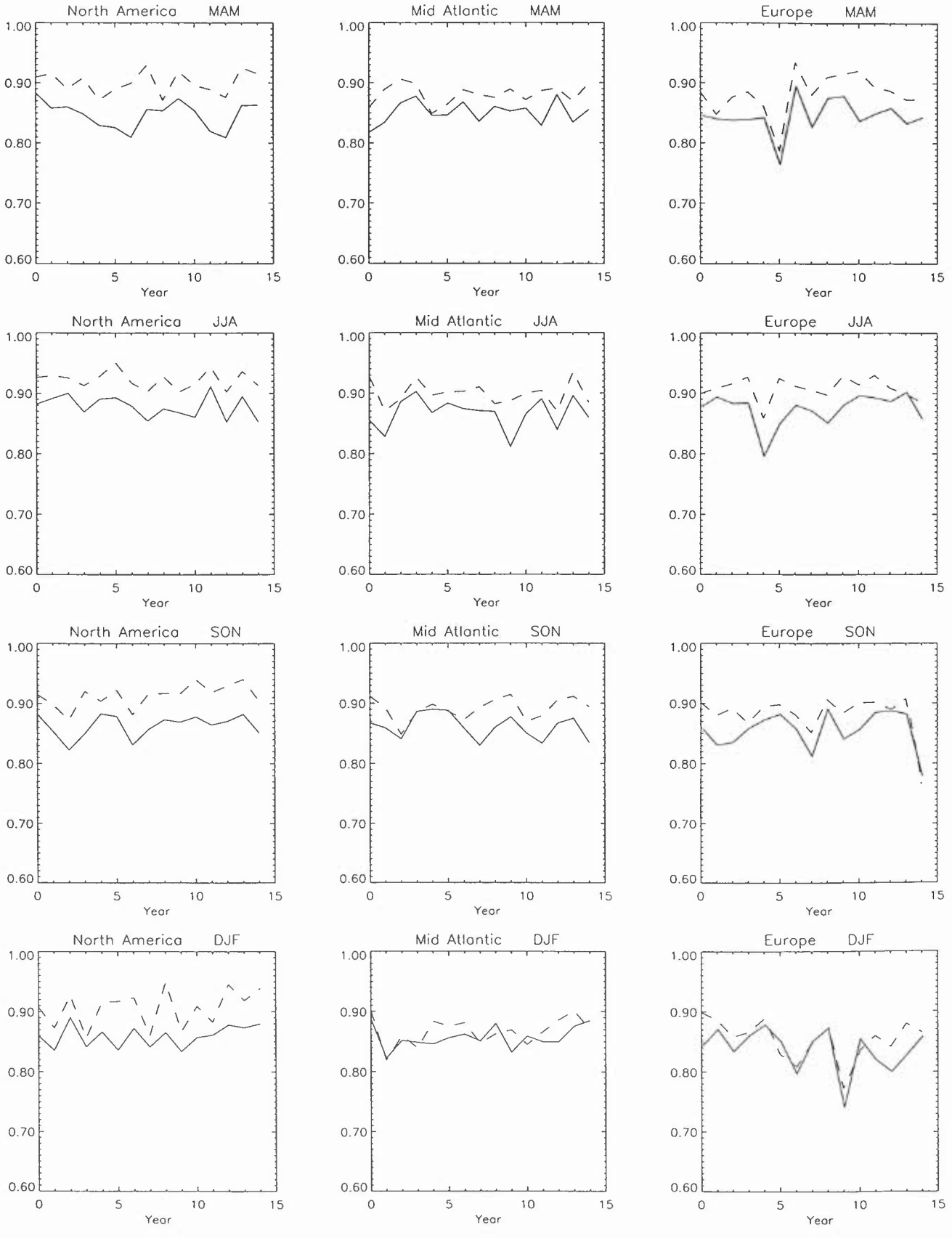
Density T=−5 1992



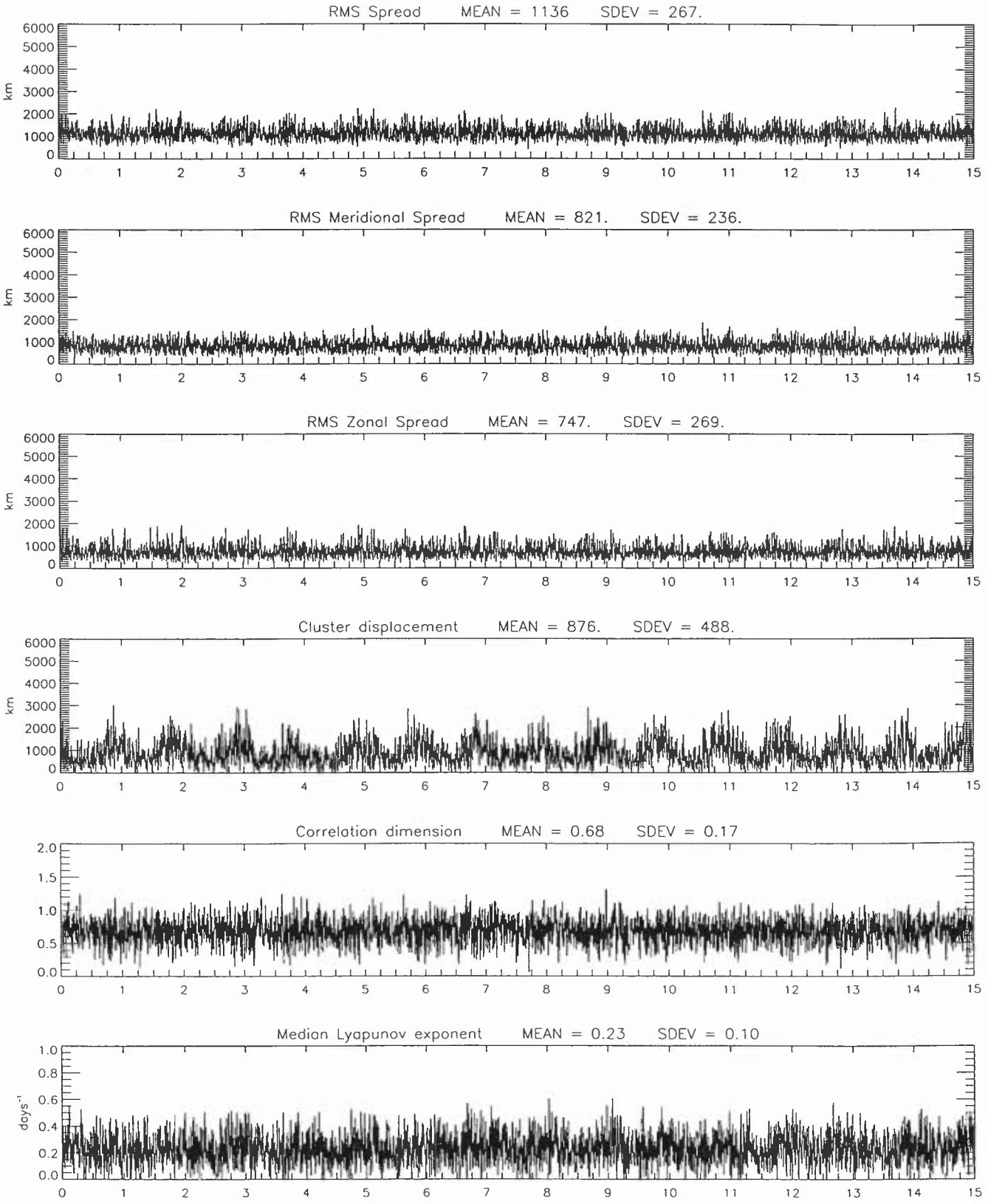
Density T=−5 1993



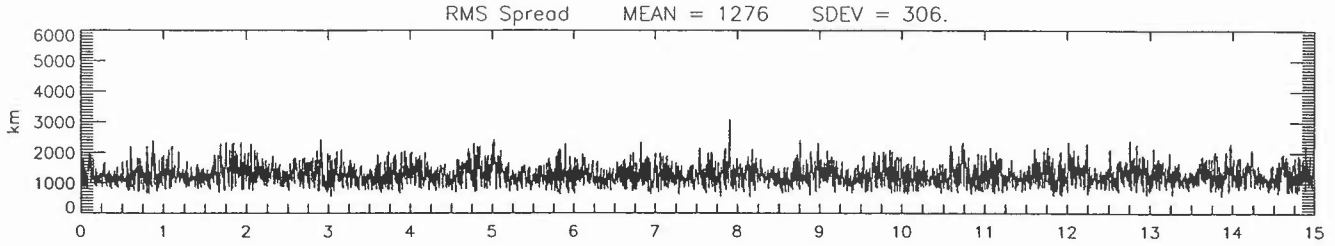
Overlap

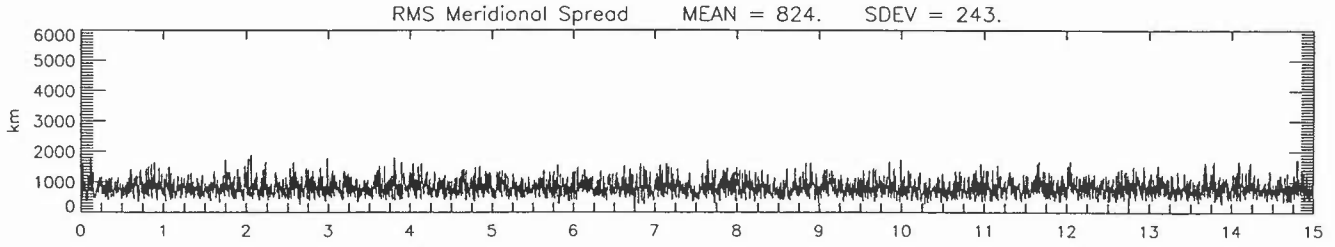

North America MAM
Mid Atlantic MAM
Europe MAM
North America JJA
Mid Atlantic JJA
Europe JJA
North America SON
Mid Atlantic SON
Europe SON
North America DJF
Mid Atlantic DJF
Europe DJF
1.00
0.90
0.80
0.70
0.60
0
5
10
15
Year

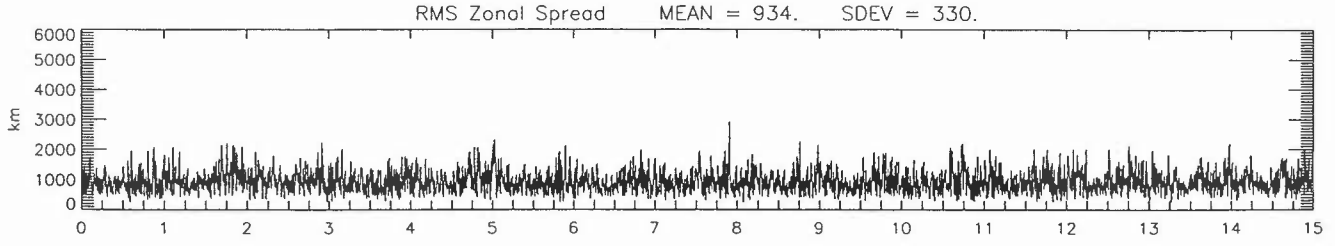
North America 2 days backwards

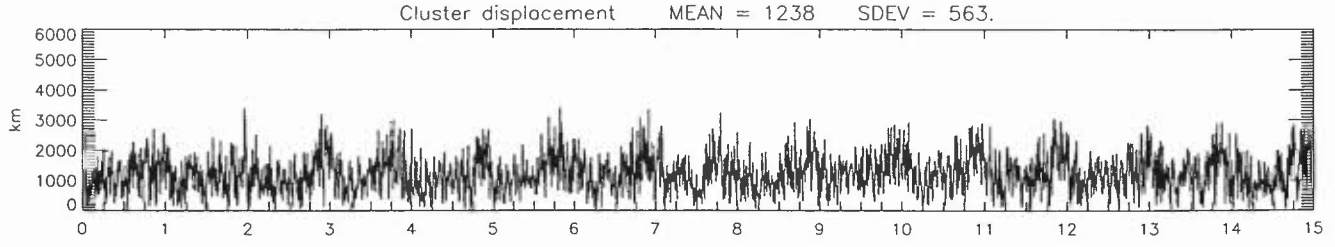


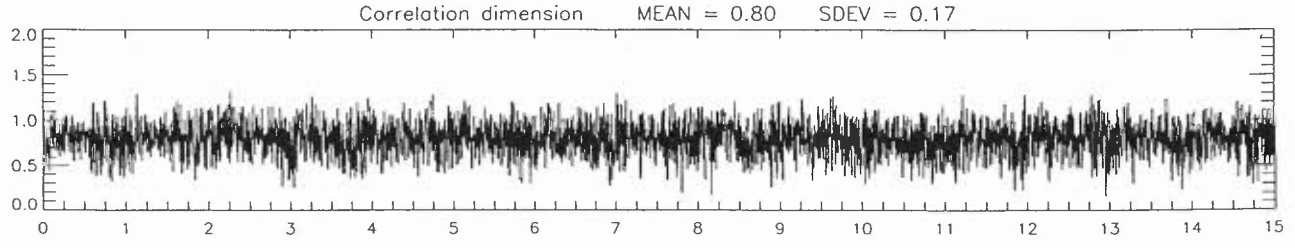
Mid Atlantic 2 days backwards

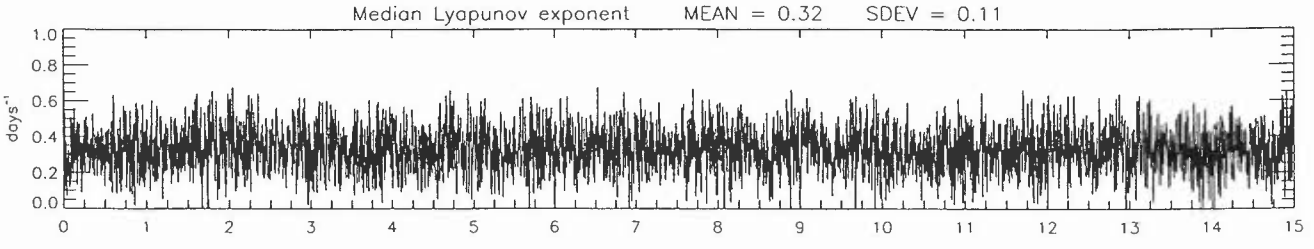






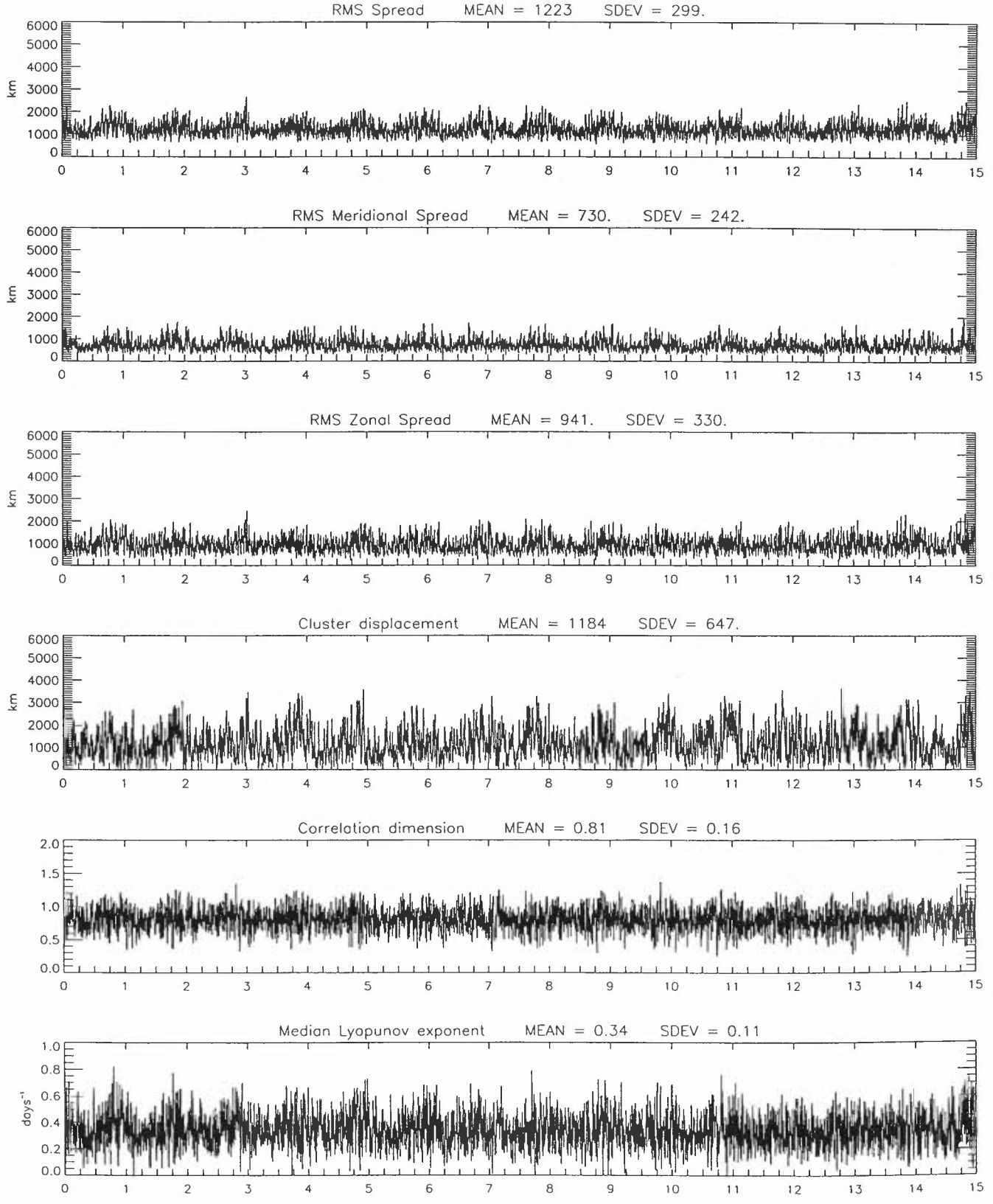




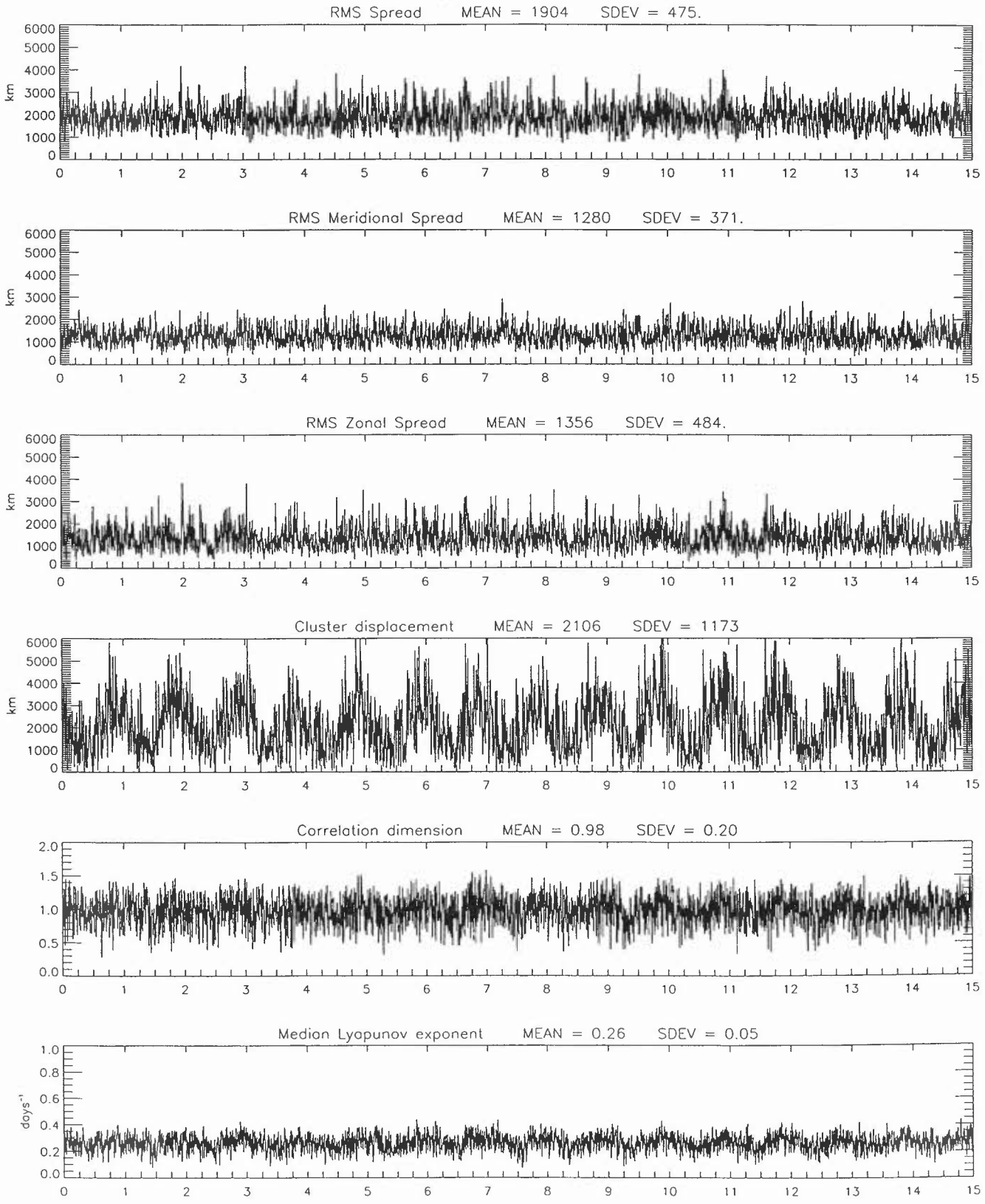


Years since 790301

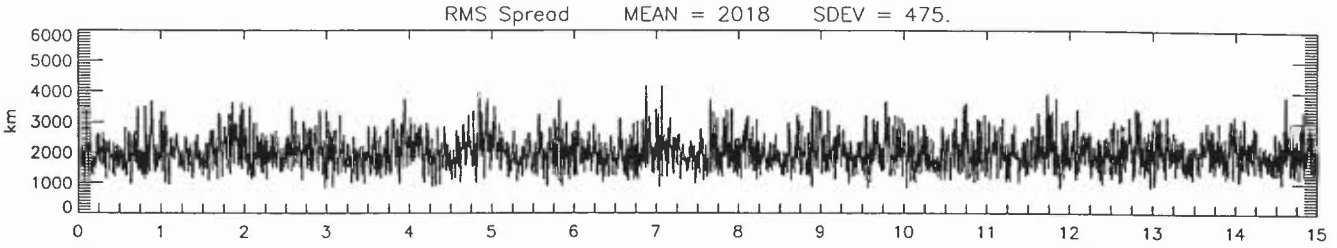
Europe, sparse 2 days backwards

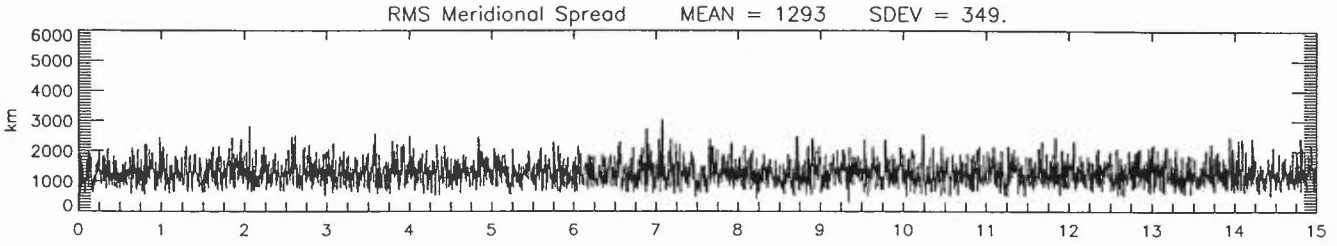


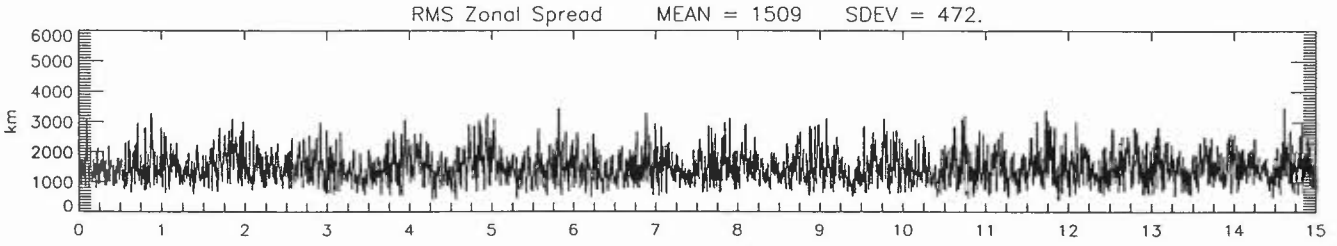
North America 5 days backwards

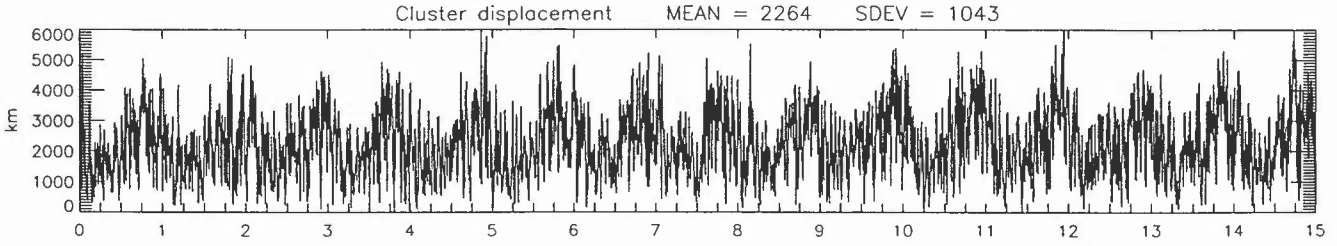


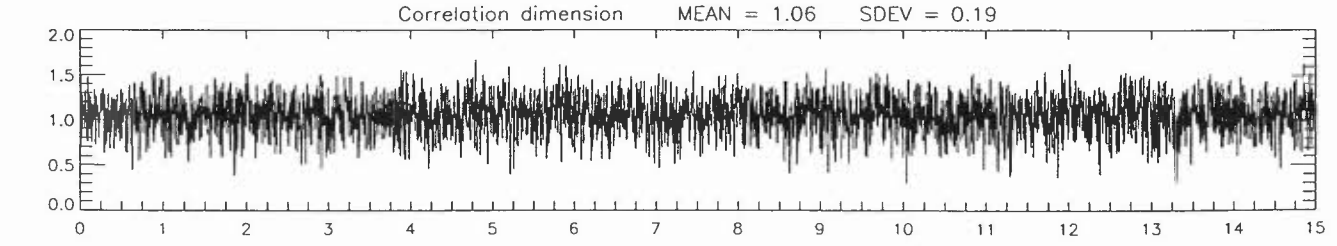
Mid Atlantic 5 days backwards

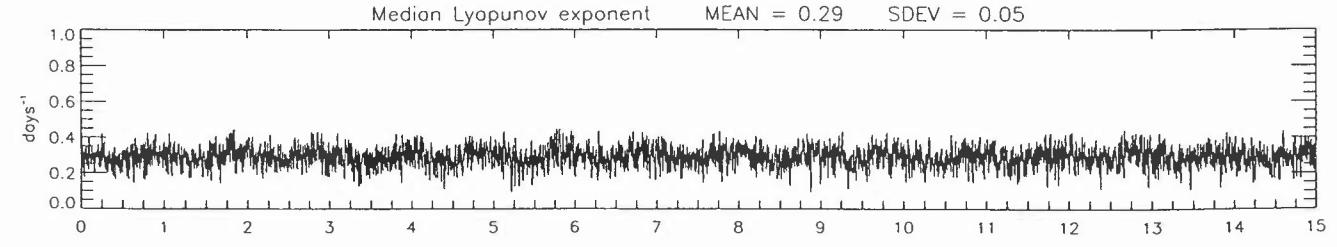






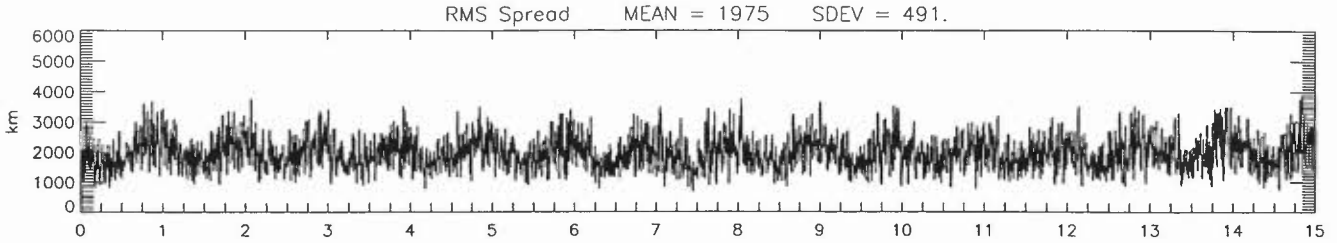


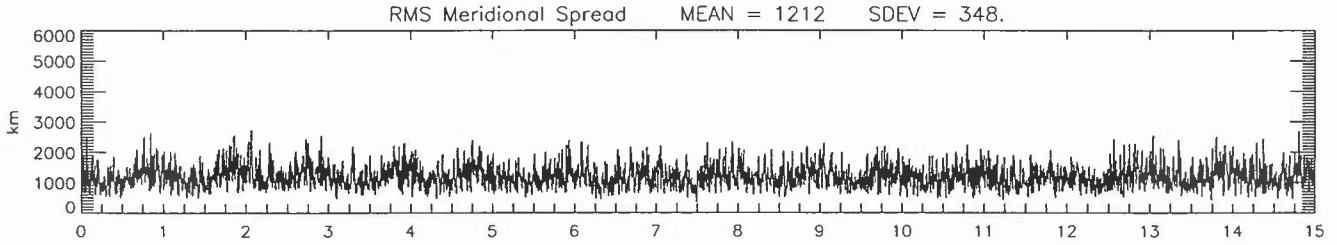


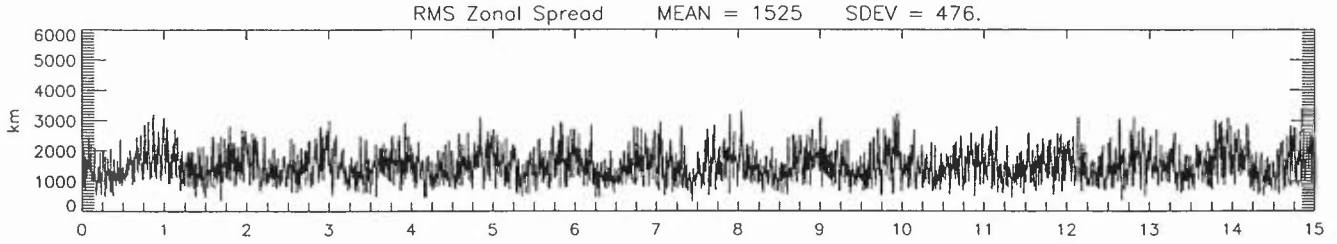


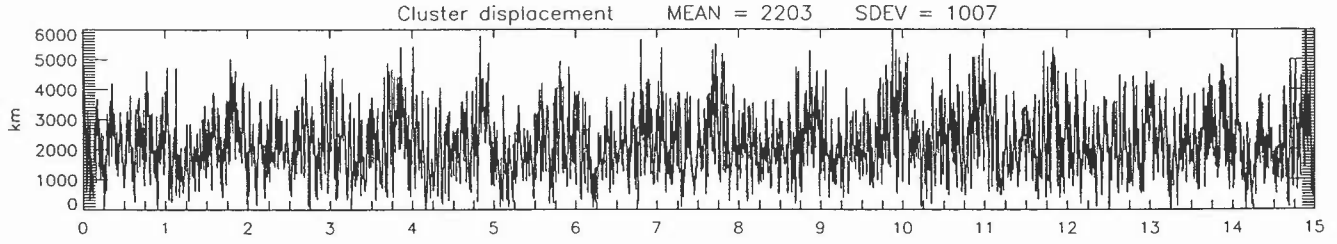
Years since 790301

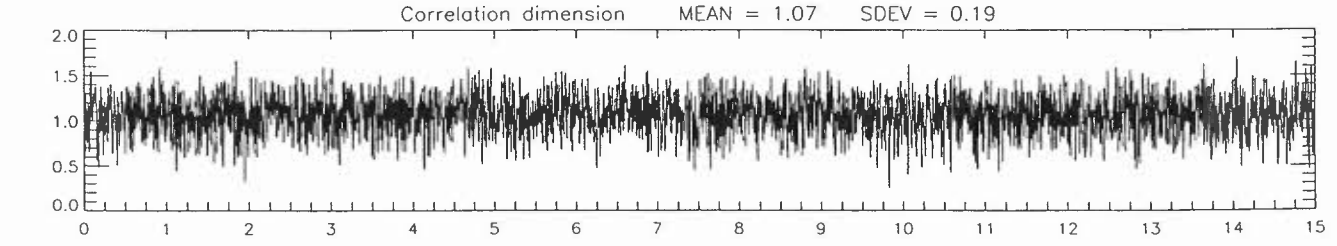
Europe, sparse 5 days backwards

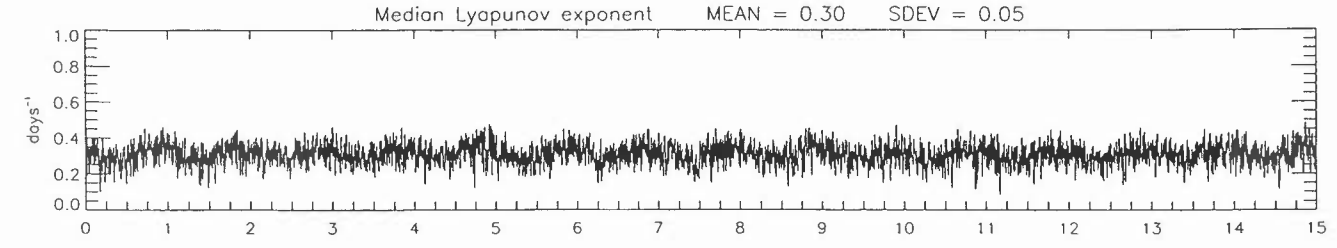






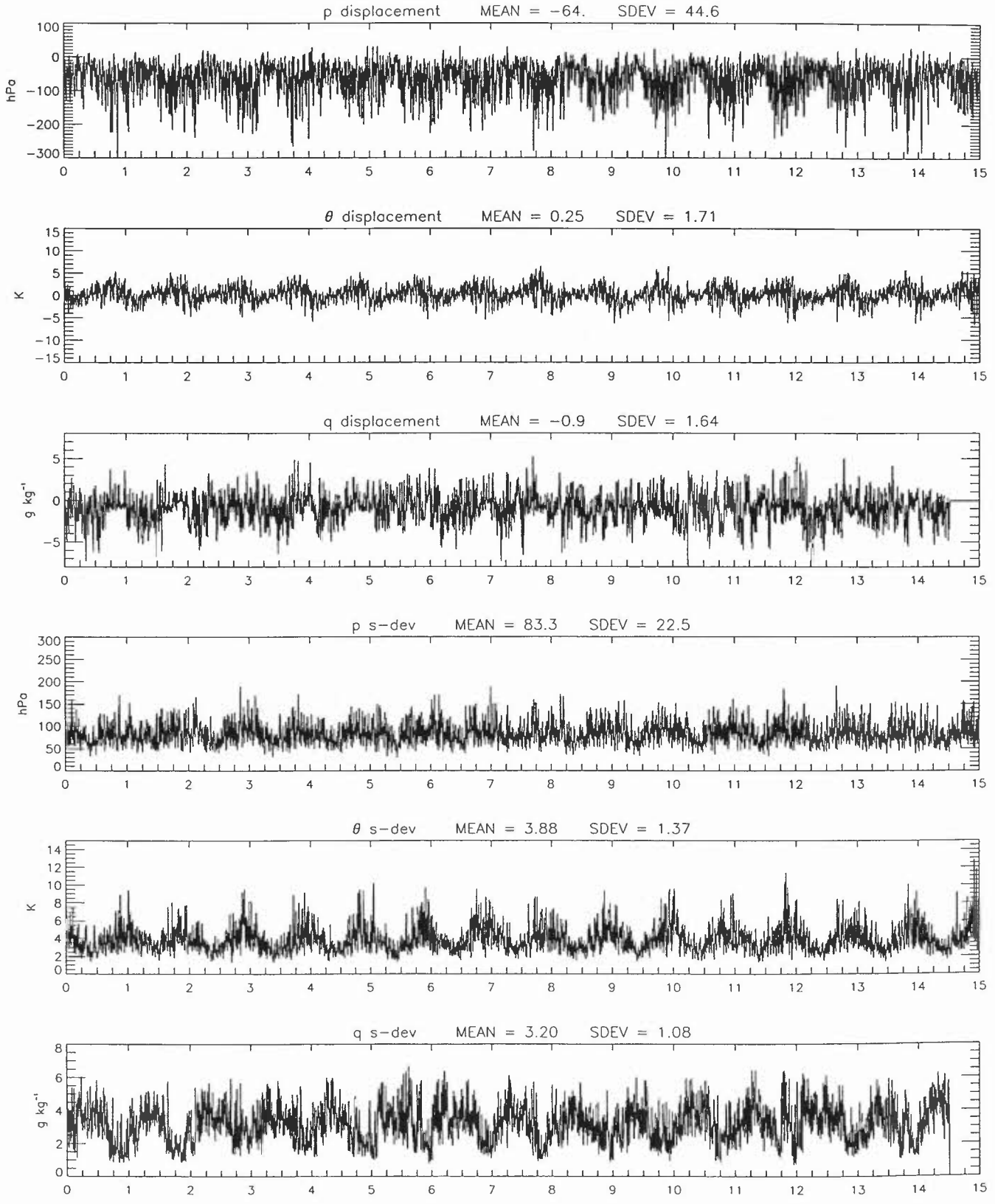






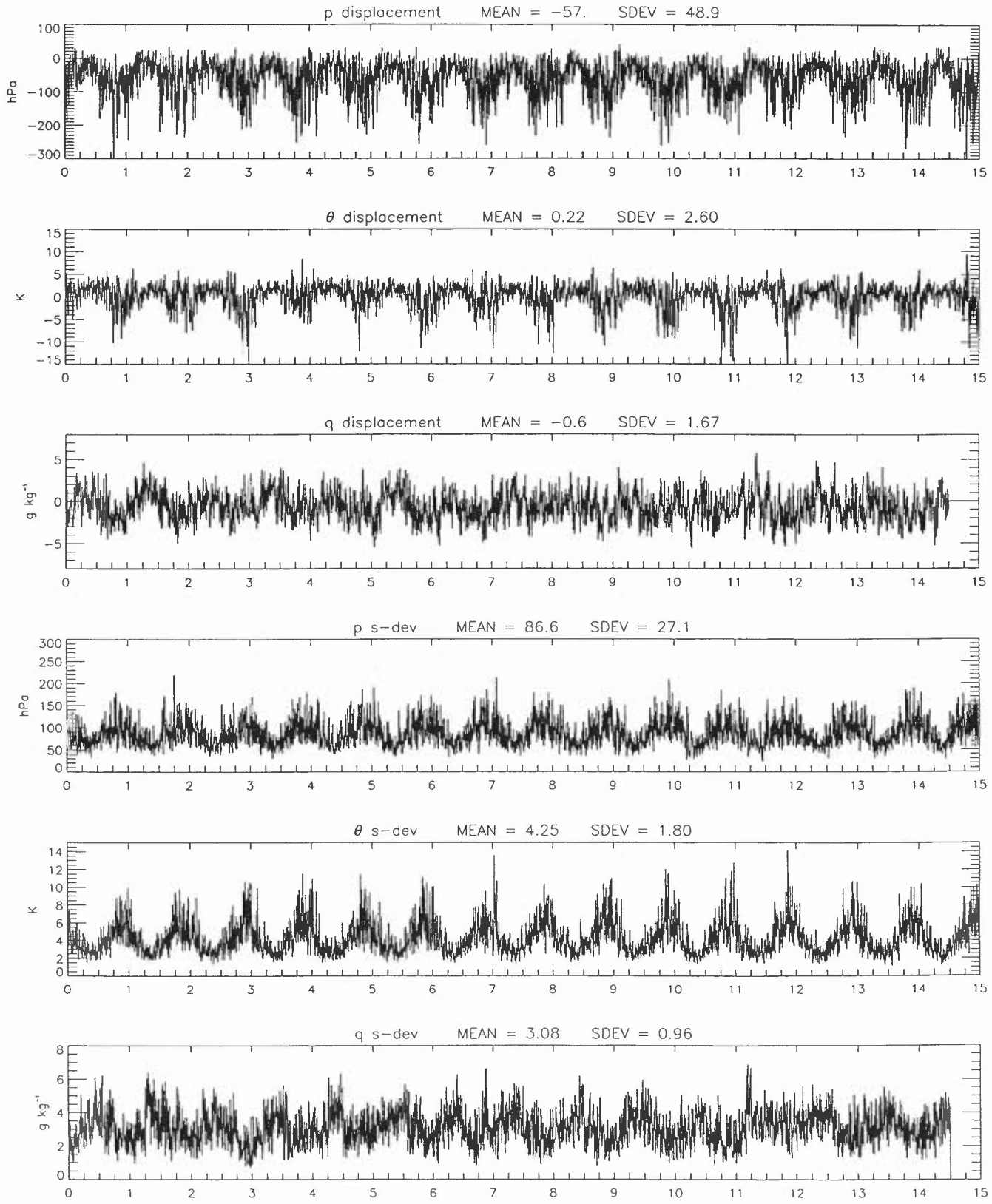
Years since 790301

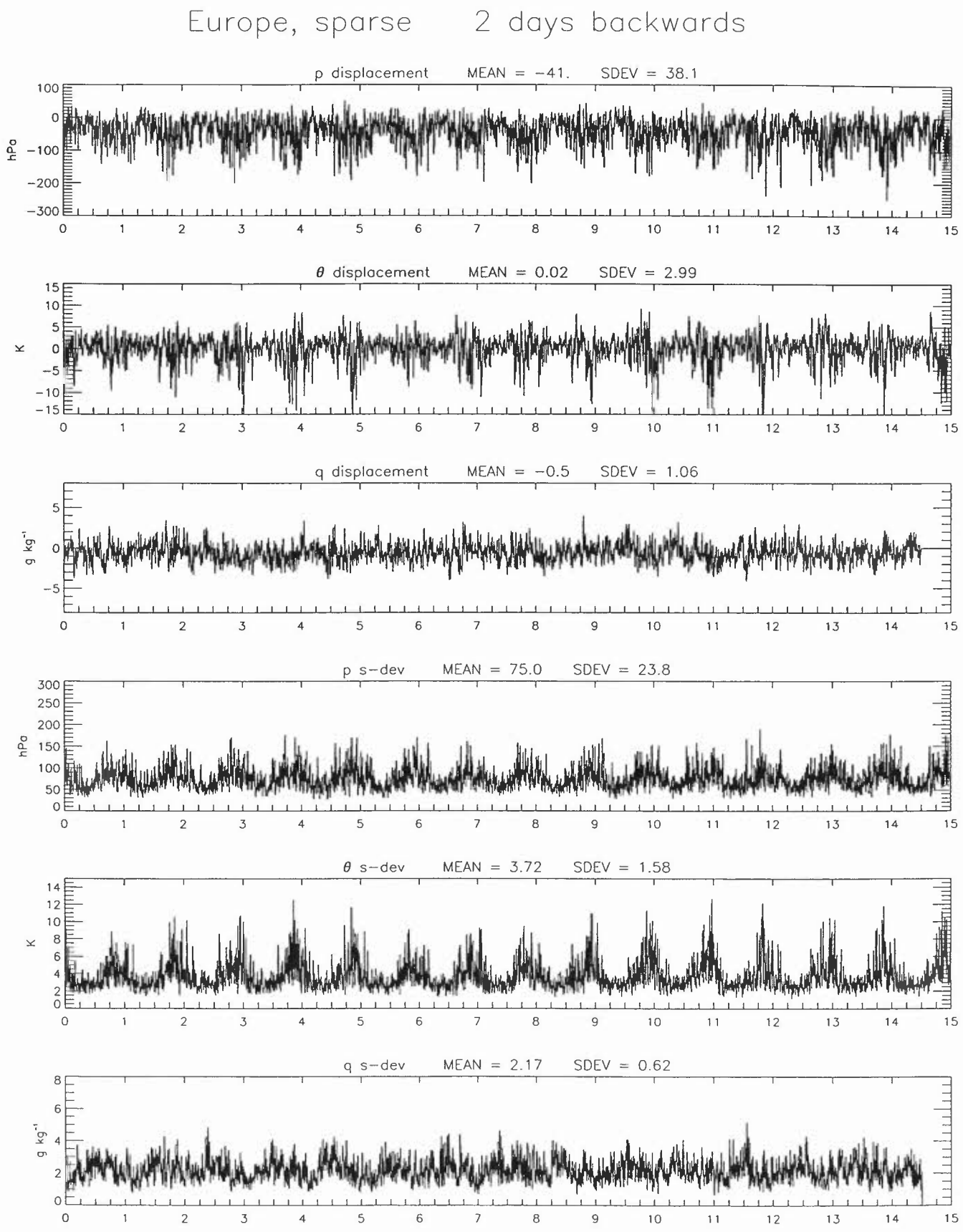
North America 2 days backwards

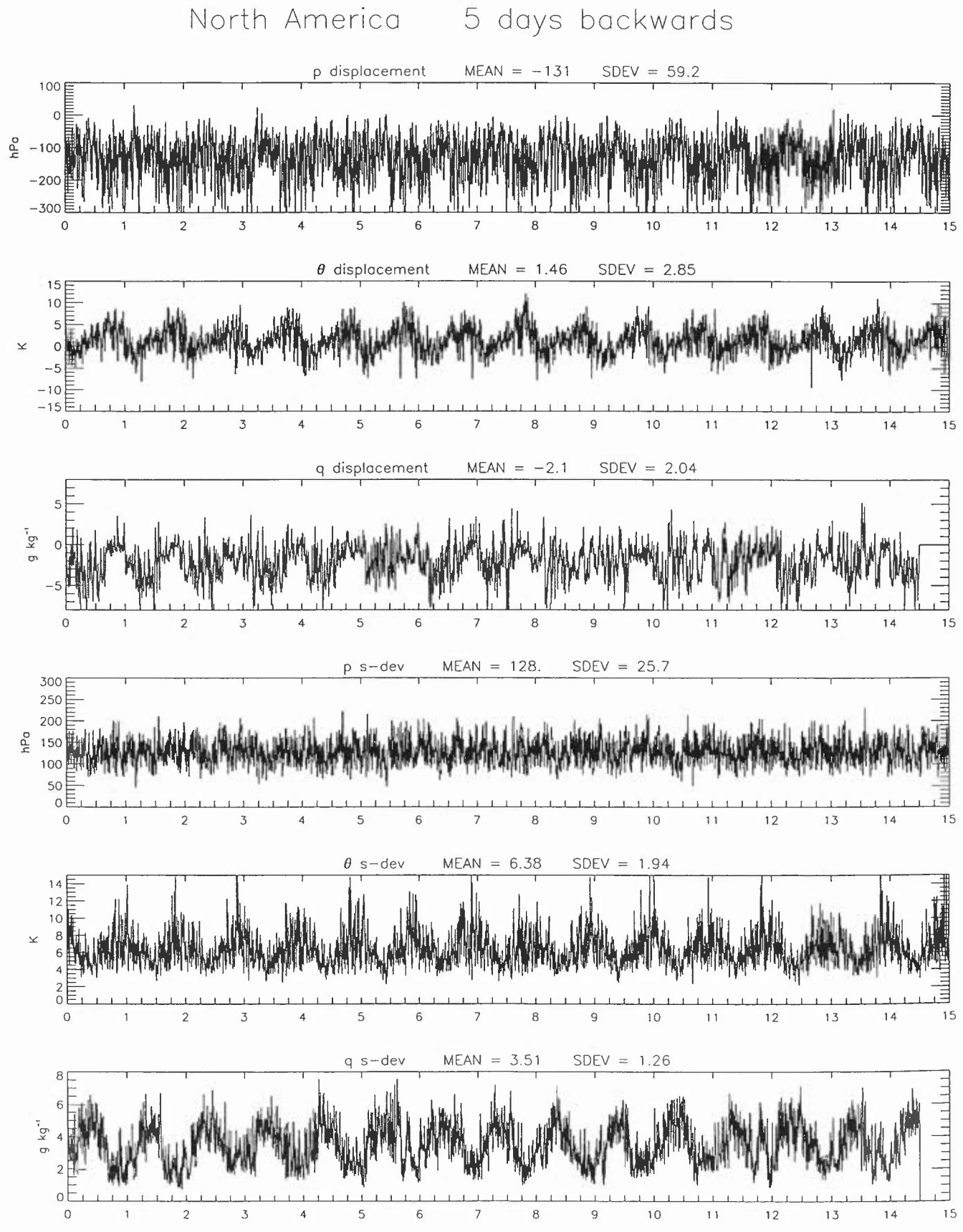


Years since 790301

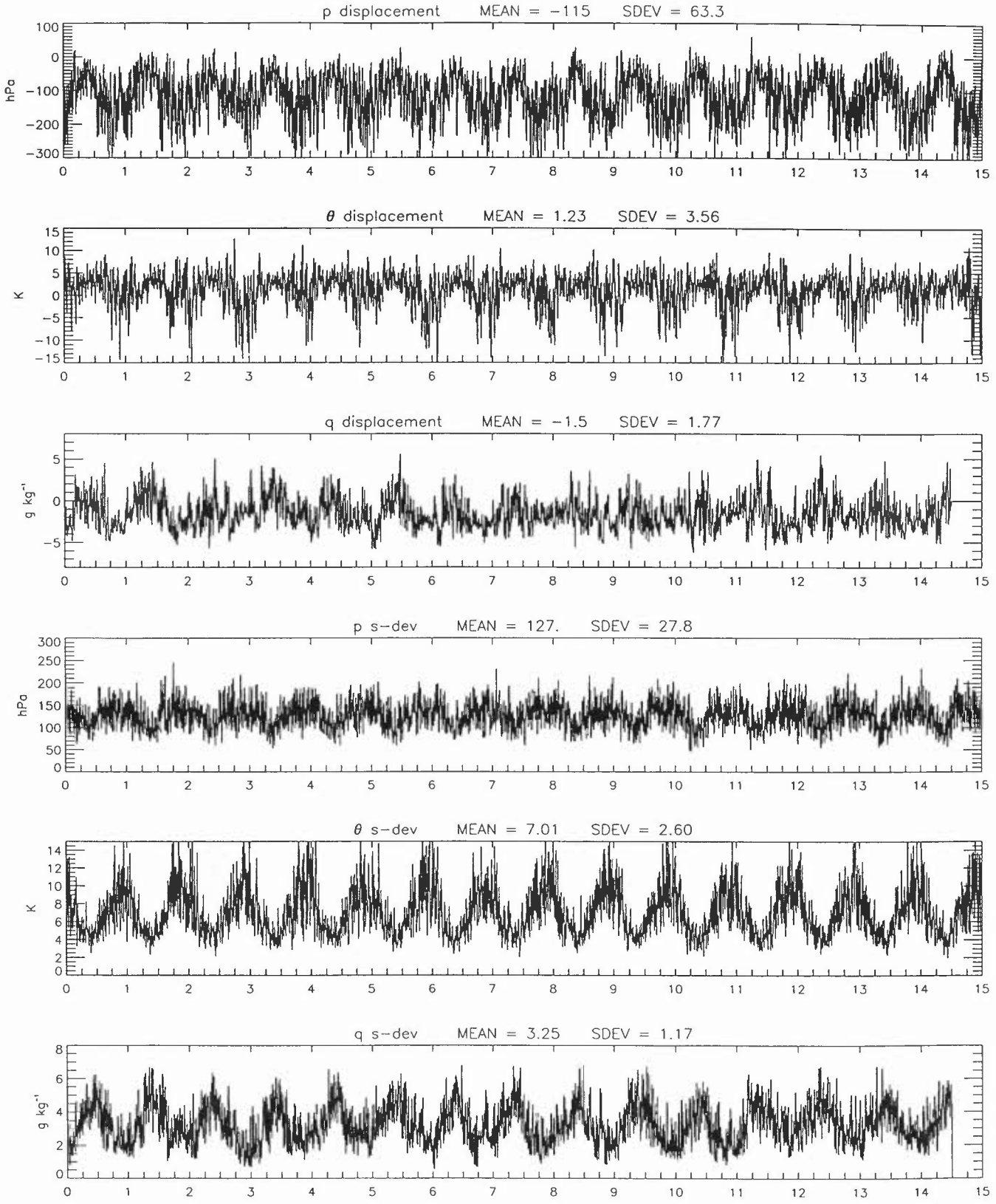
Mid Atlantic 2 days backwards



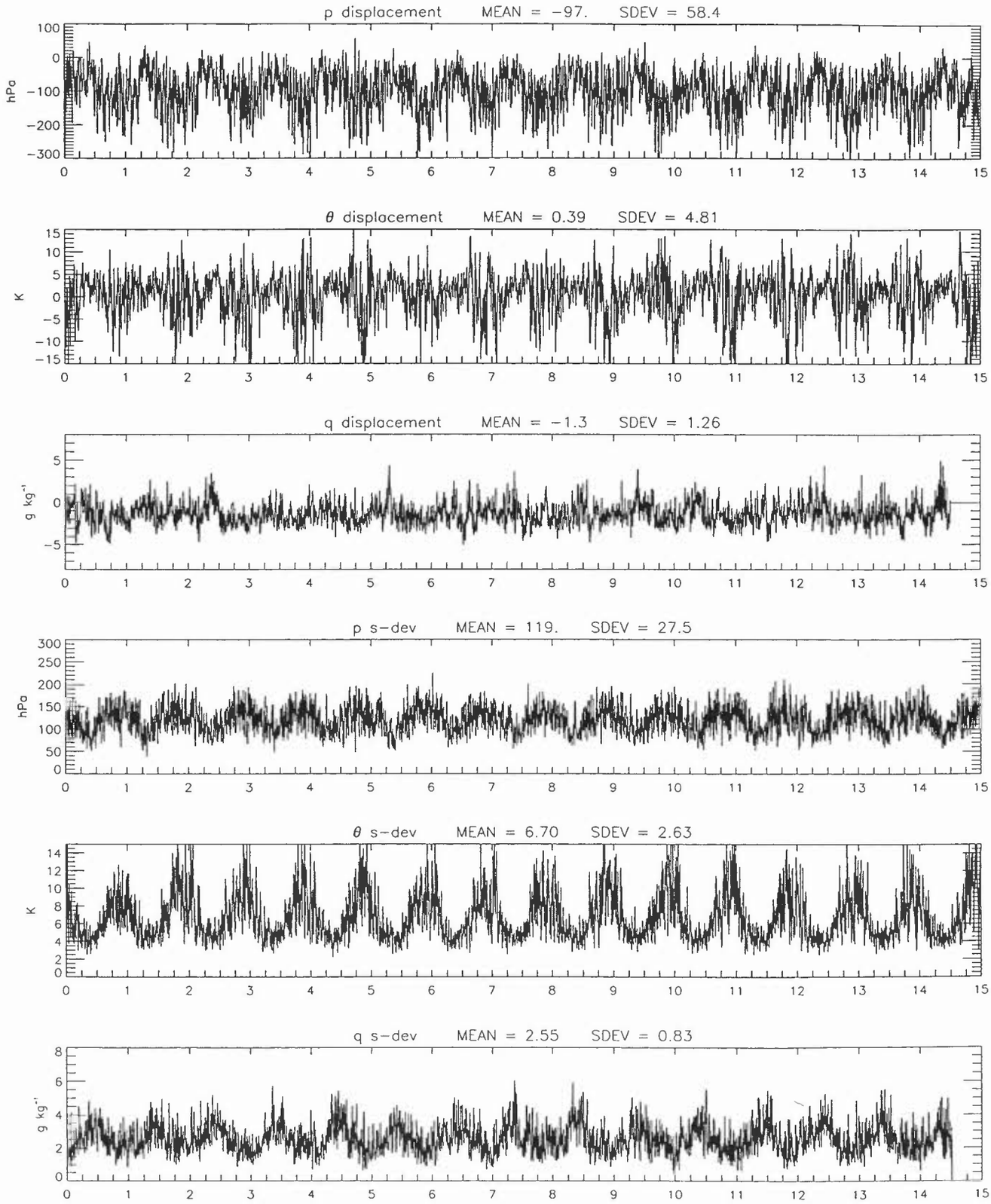

Europe, sparse 2 days backwards
p displacement MEAN = −41. SDEV = 38.1
hPa
θ displacement MEAN = 0.02 SDEV = 2.99
K
q displacement MEAN = −0.5 SDEV = 1.06
g kg⁻¹
p s−dev MEAN = 75.0 SDEV = 23.8
hPa
θ s−dev MEAN = 3.72 SDEV = 1.58
K
q s−dev MEAN = 2.17 SDEV = 0.62
g kg⁻¹
Years since 790301


North America 5 days backwards
p displacement MEAN = −131 SDEV = 59.2
hPa
θ displacement MEAN = 1.46 SDEV = 2.85
K
q displacement MEAN = −2.1 SDEV = 2.04
g kg⁻¹
p s−dev MEAN = 128. SDEV = 25.7
hPa
θ s−dev MEAN = 6.38 SDEV = 1.94
K
q s−dev MEAN = 3.51 SDEV = 1.26
g kg⁻¹
Years since 790301

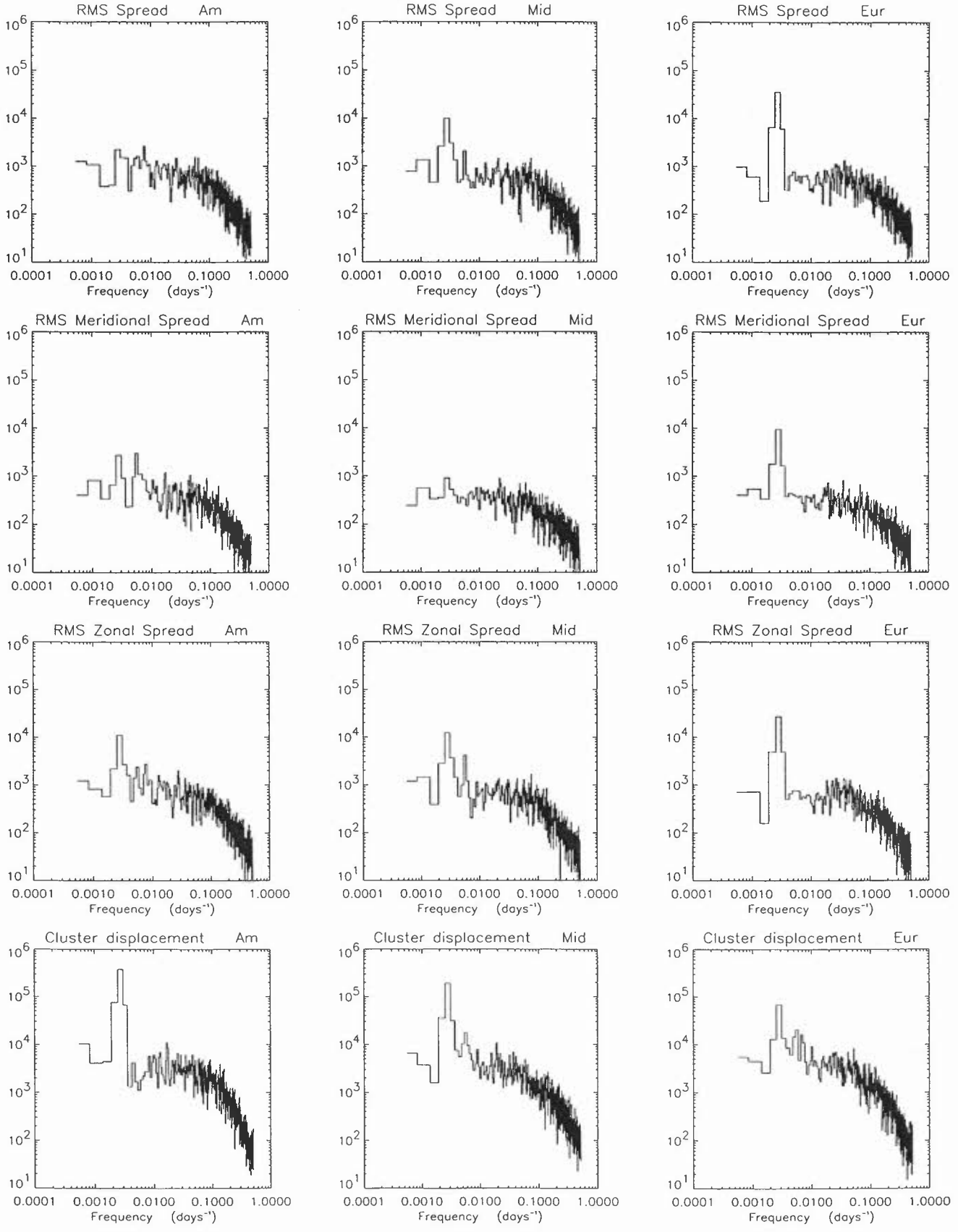
Mid Atlantic 5 days backwards



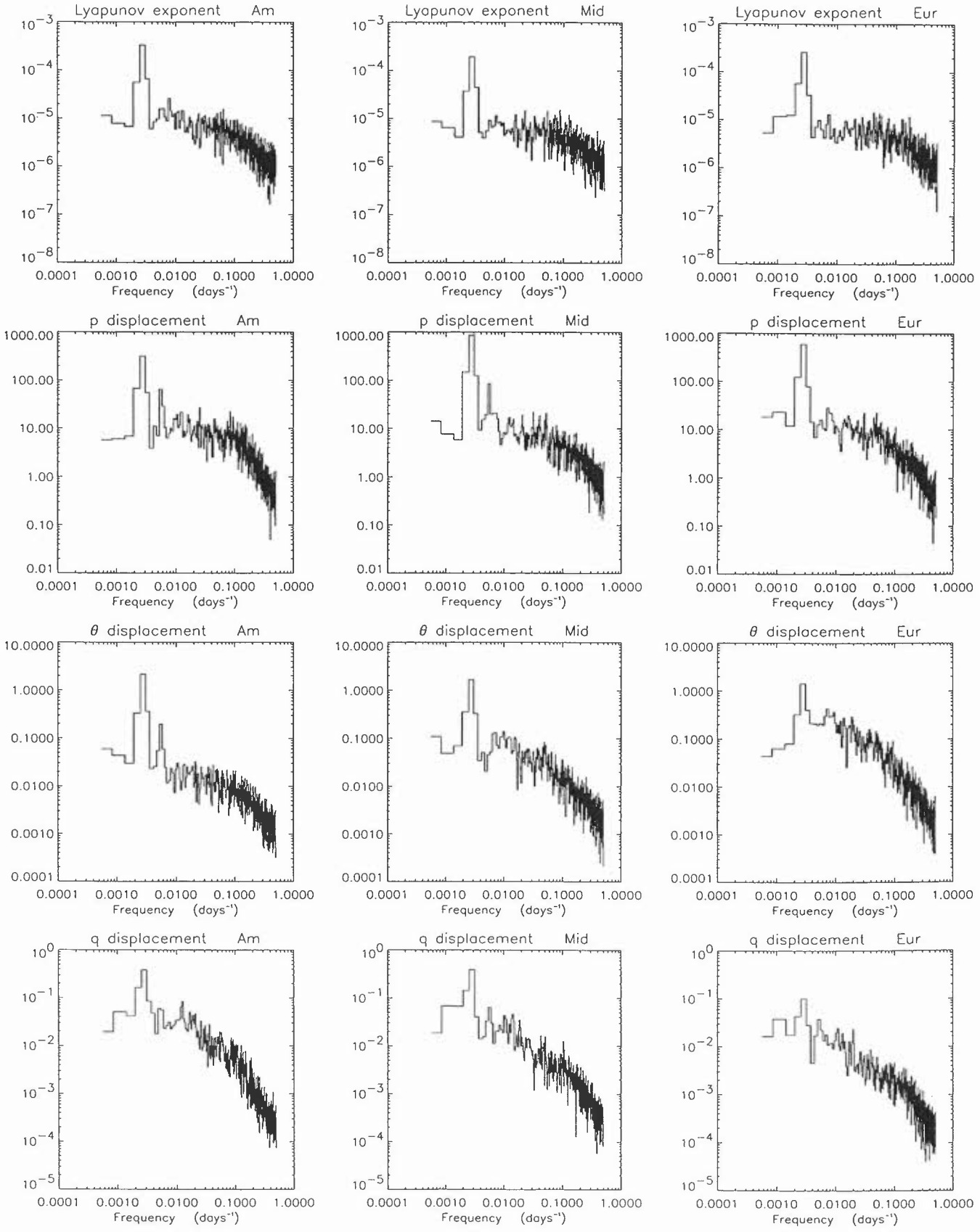
Europe, sparse 5 days backwards



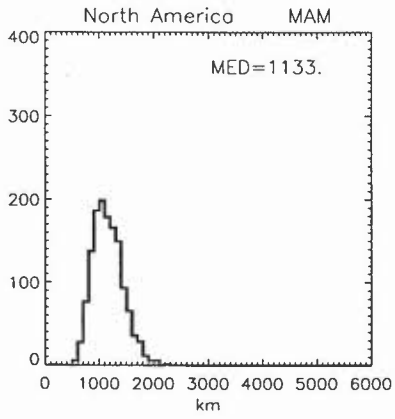
Power Spectra 5 days backwards

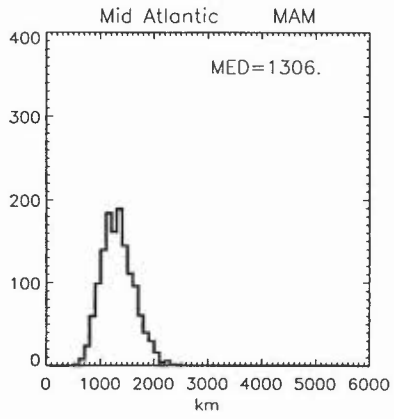


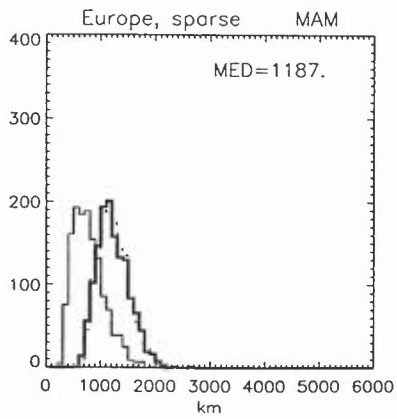
Power Spectra 5 days backwards

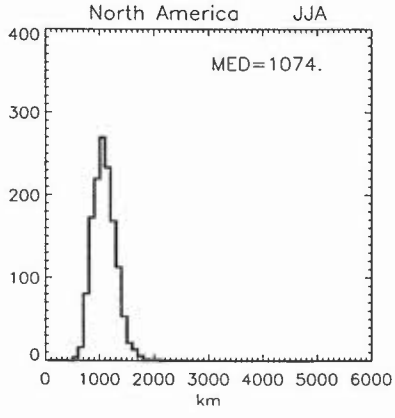


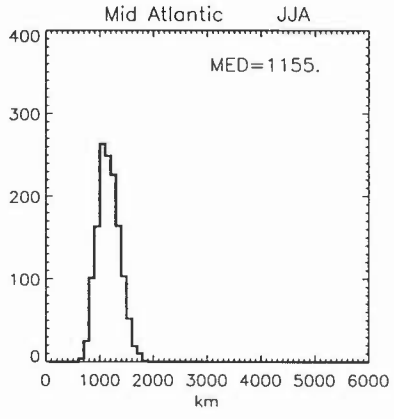
RMS Spread 2 days backwards

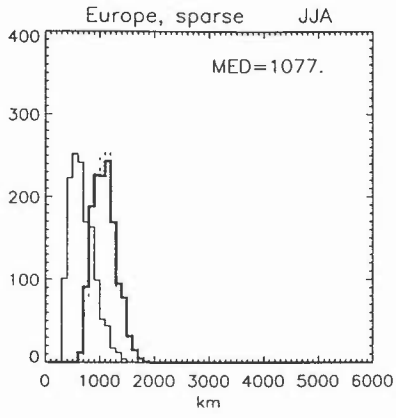


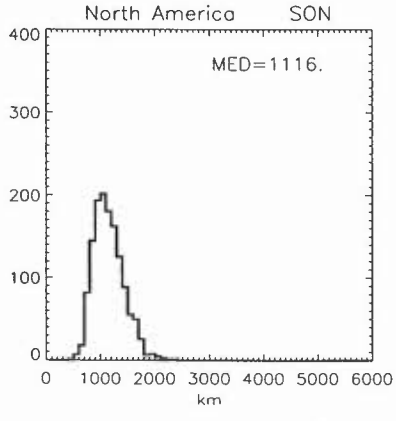


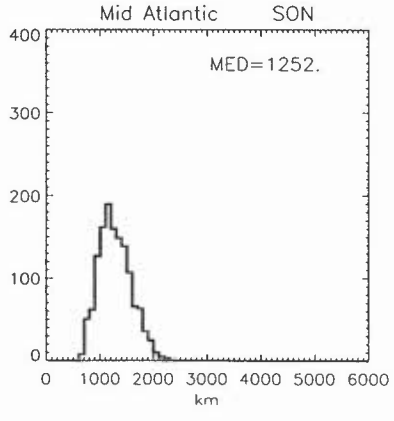


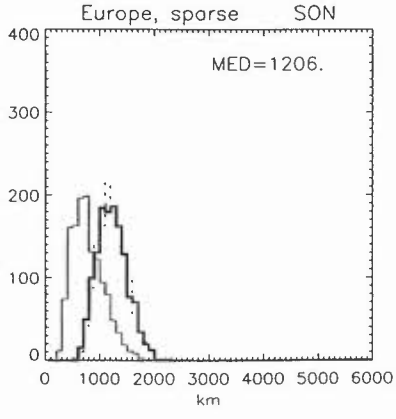


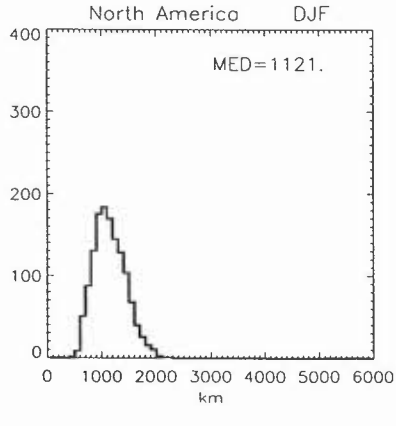


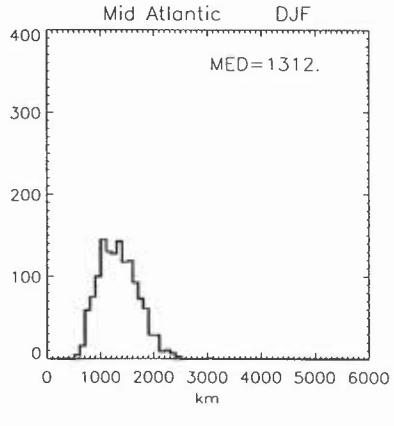


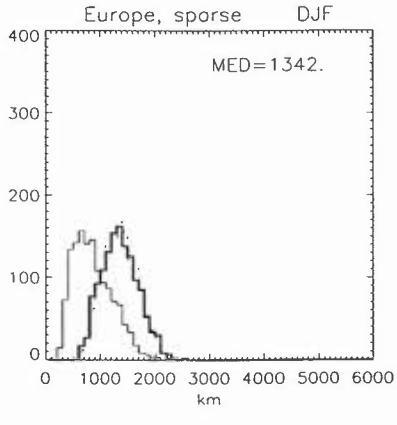




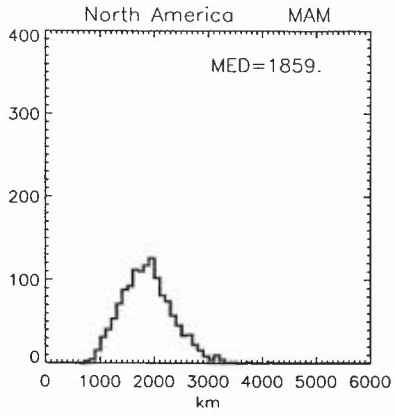


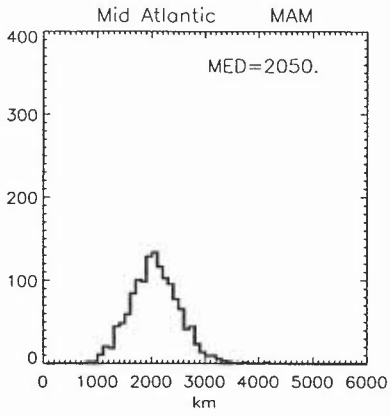


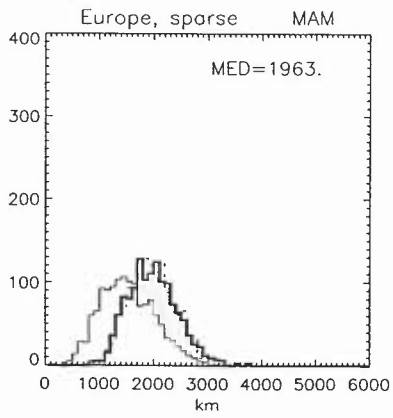


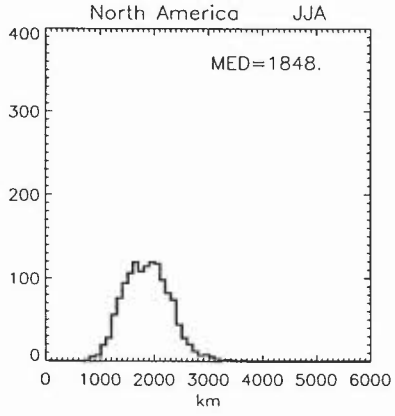


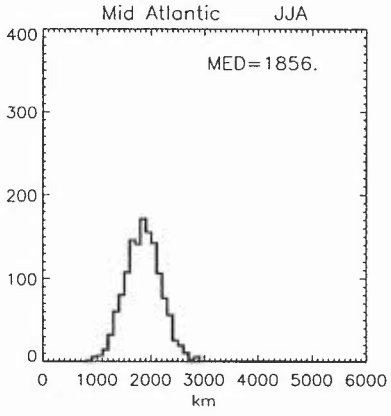
RMS Spread 5 days backwards

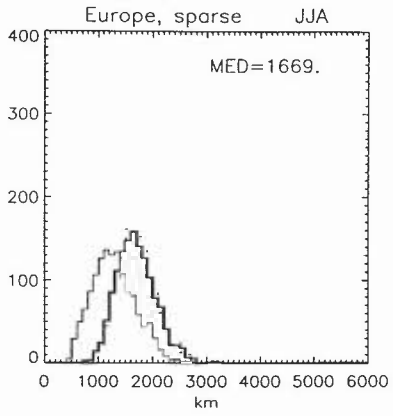


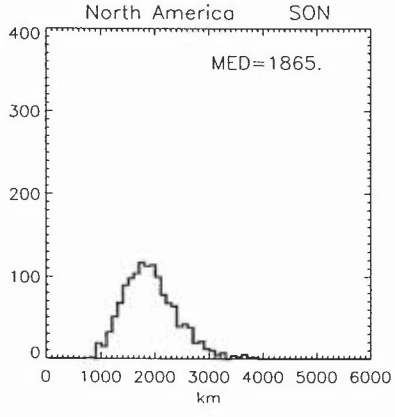


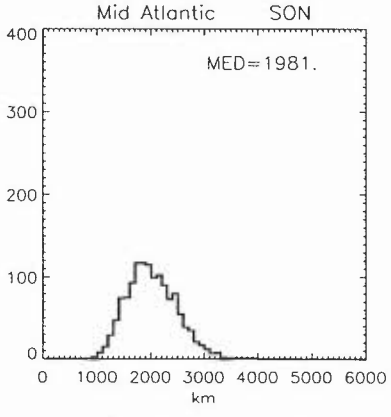


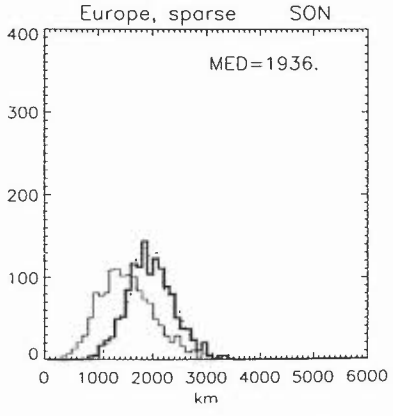


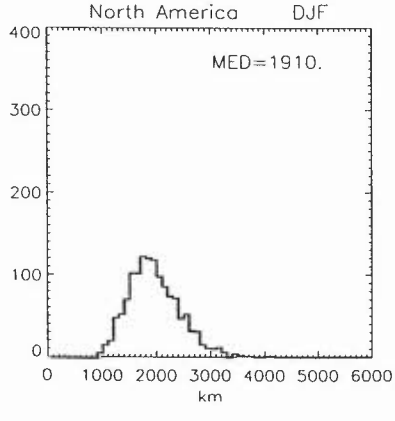


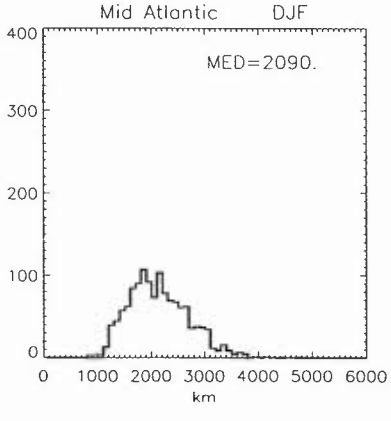


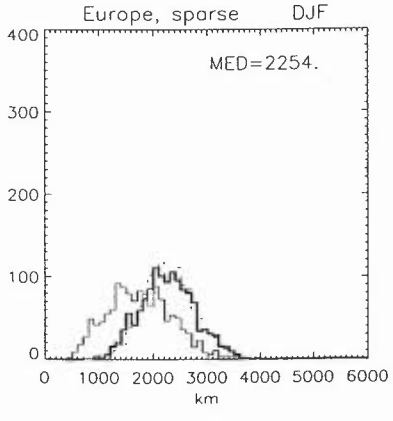




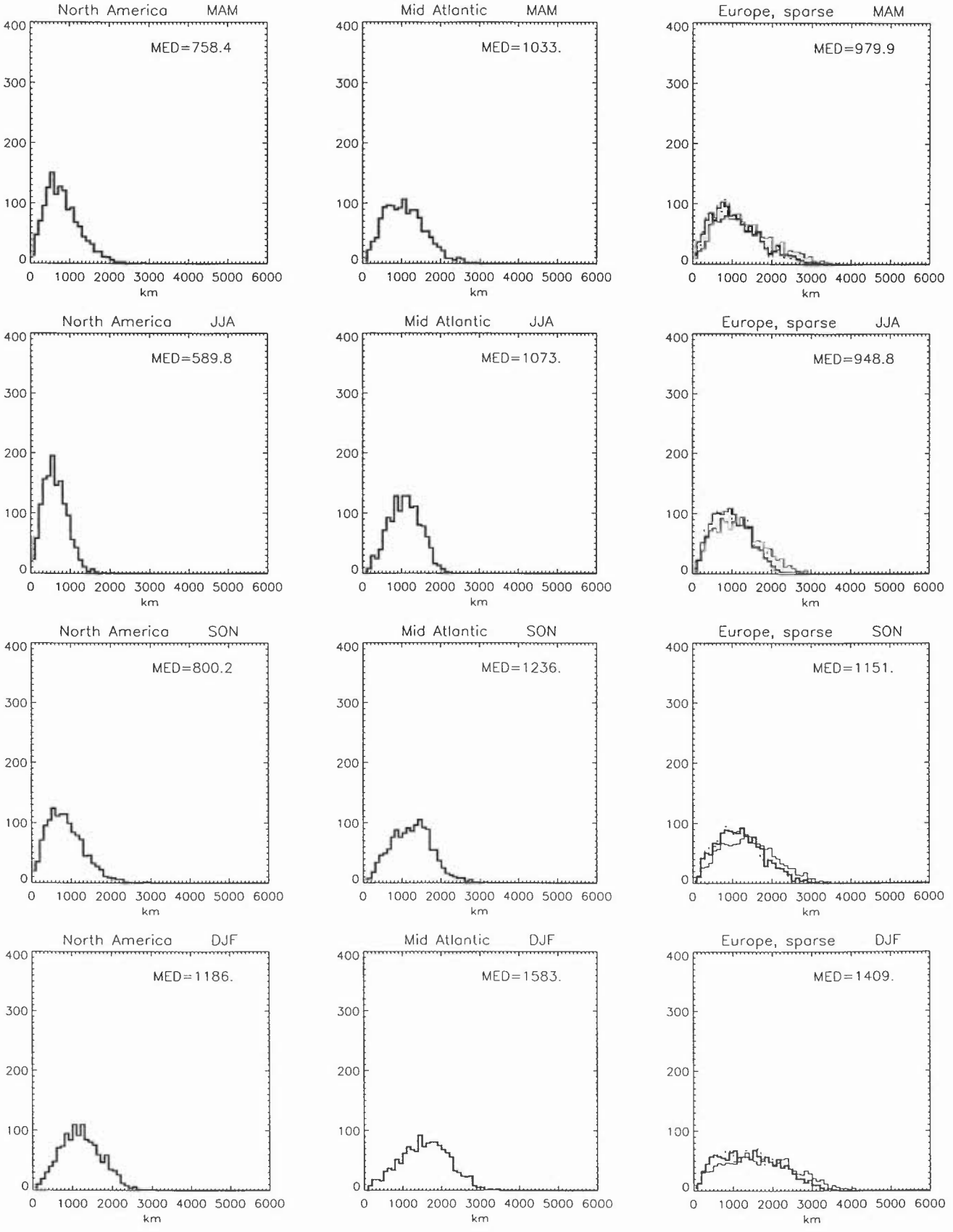




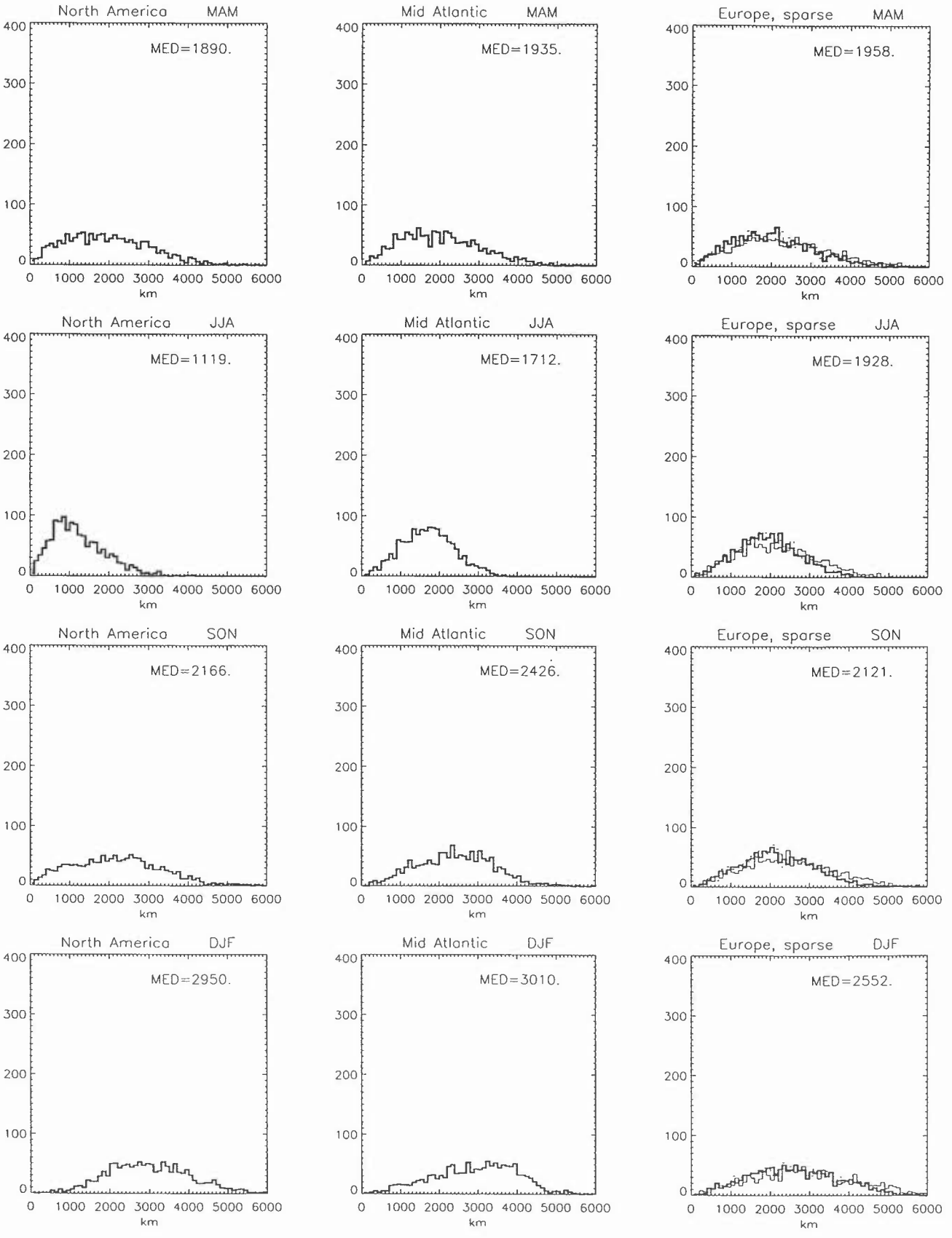




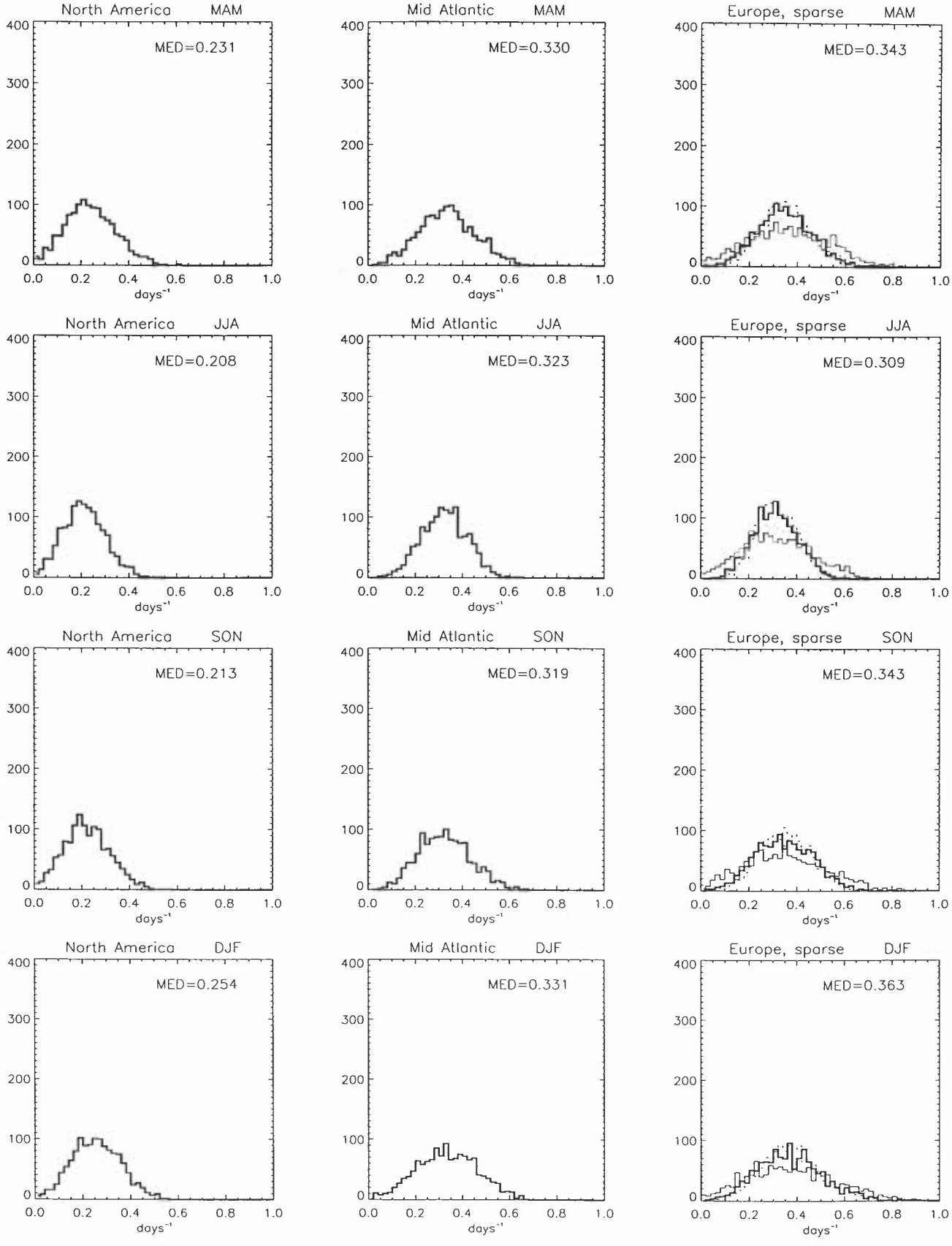
Cluster displacement 2 days backwards



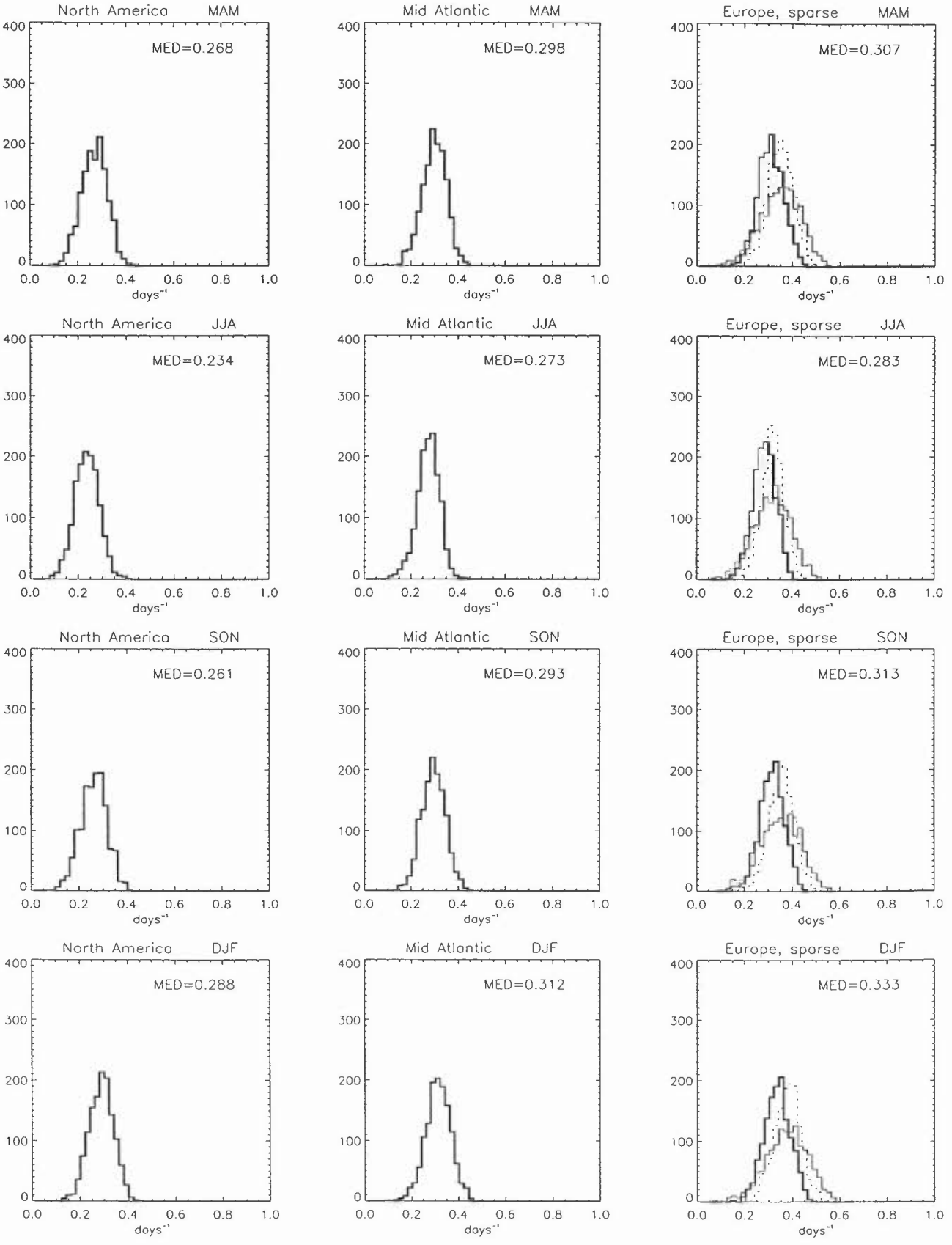
Cluster displacement 5 days backwards



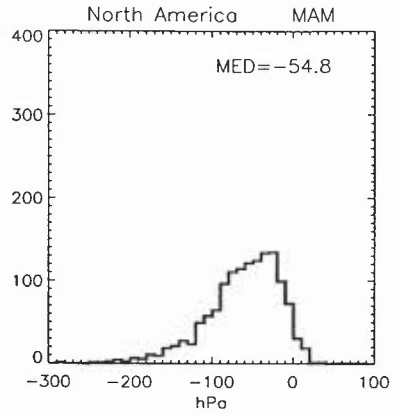
Median Lyapunov exponent 2 days backwards

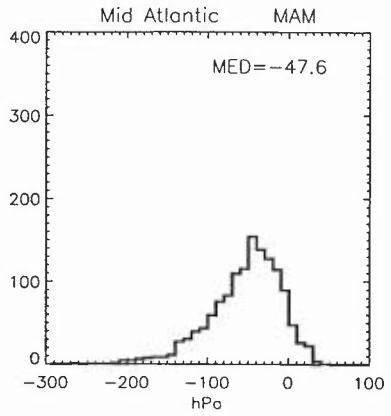


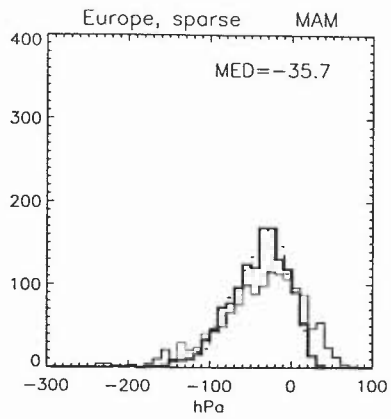
Median Lyapunov exponent 5 days backwards

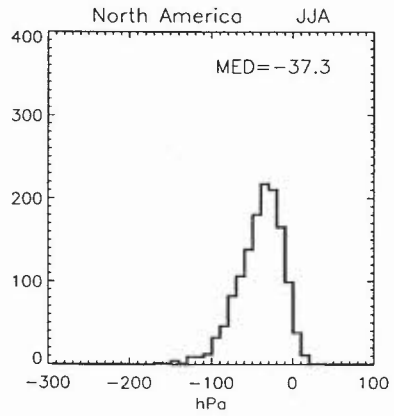


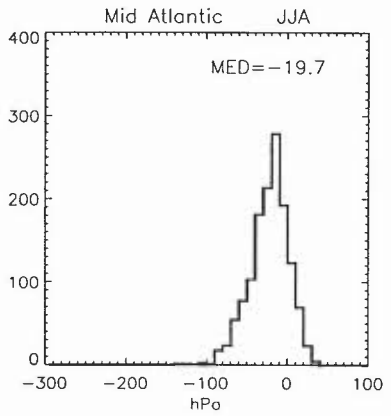
p displacement 2 days backwards

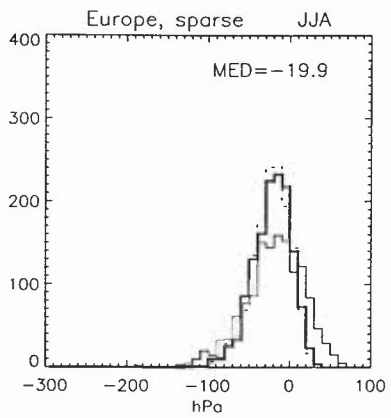


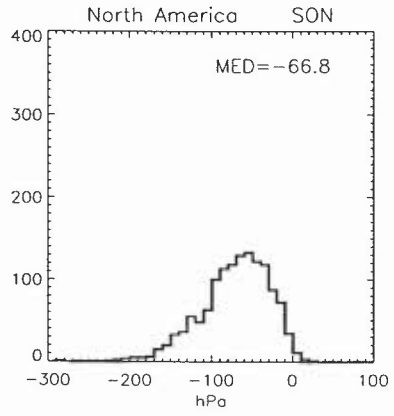


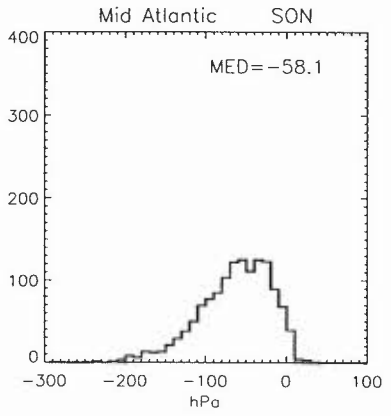


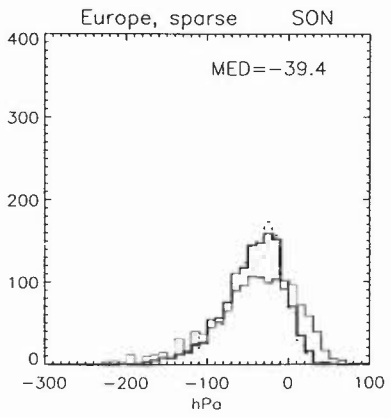


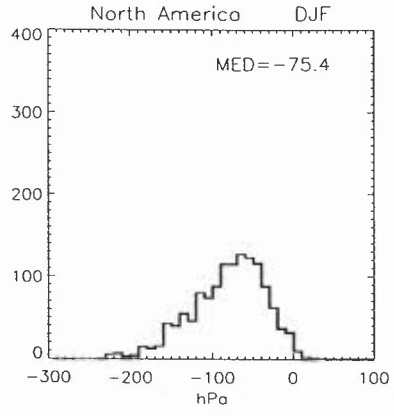


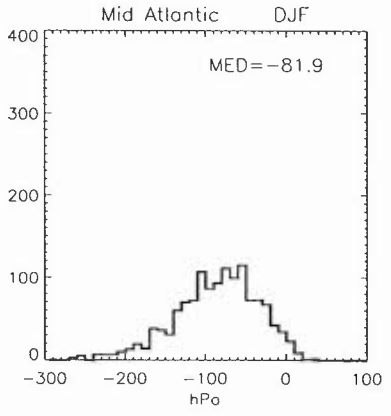


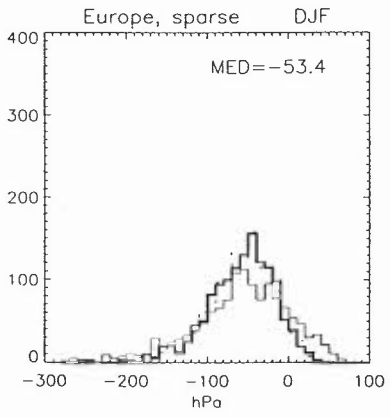




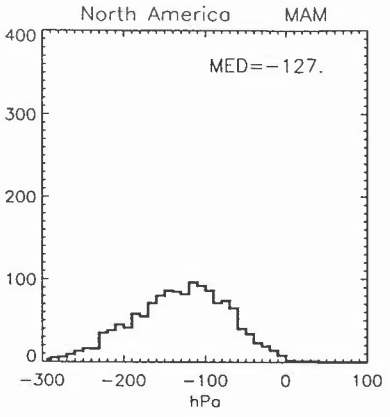


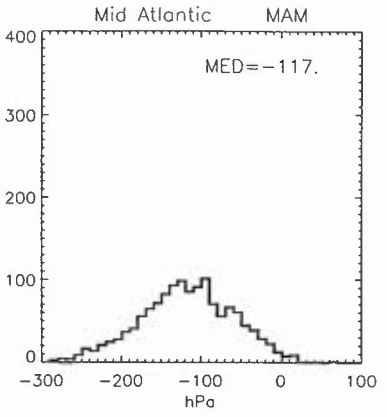


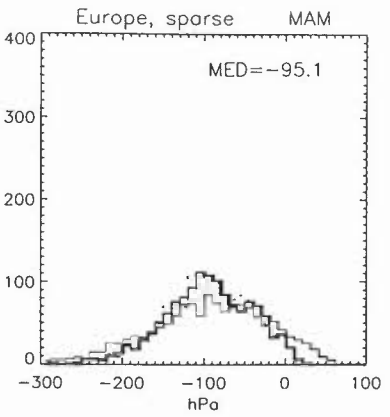


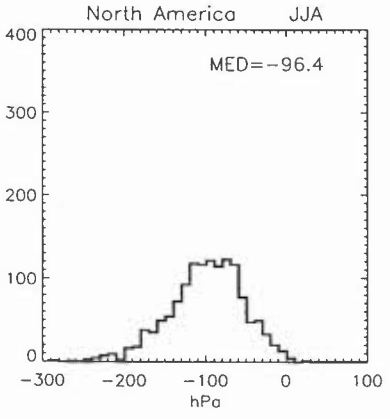


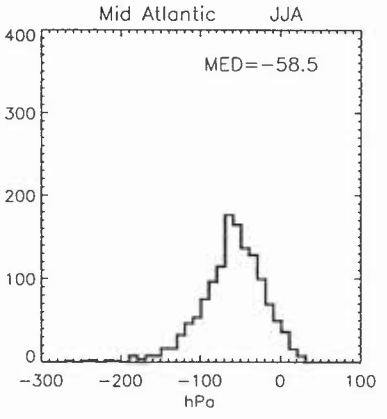
p displacement 5 days backwards

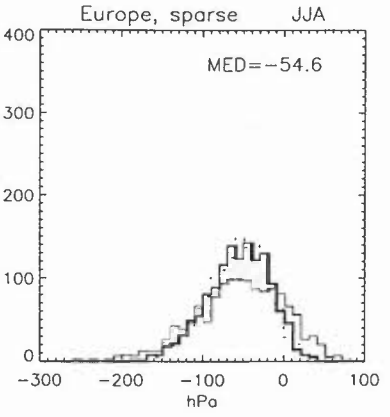


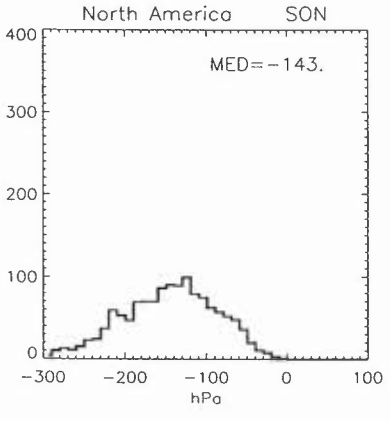


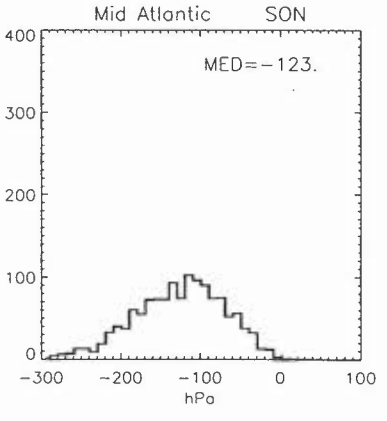


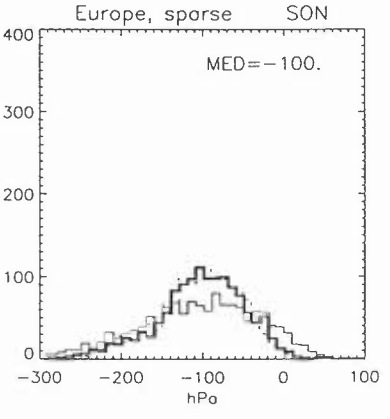


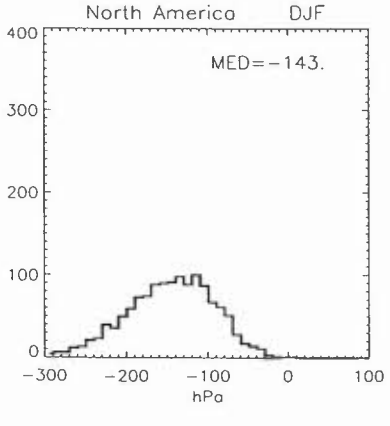


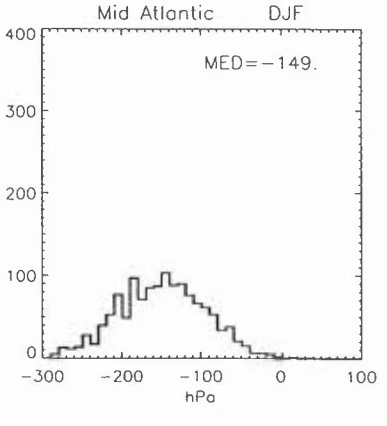


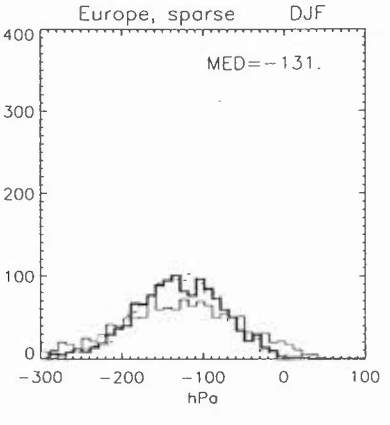




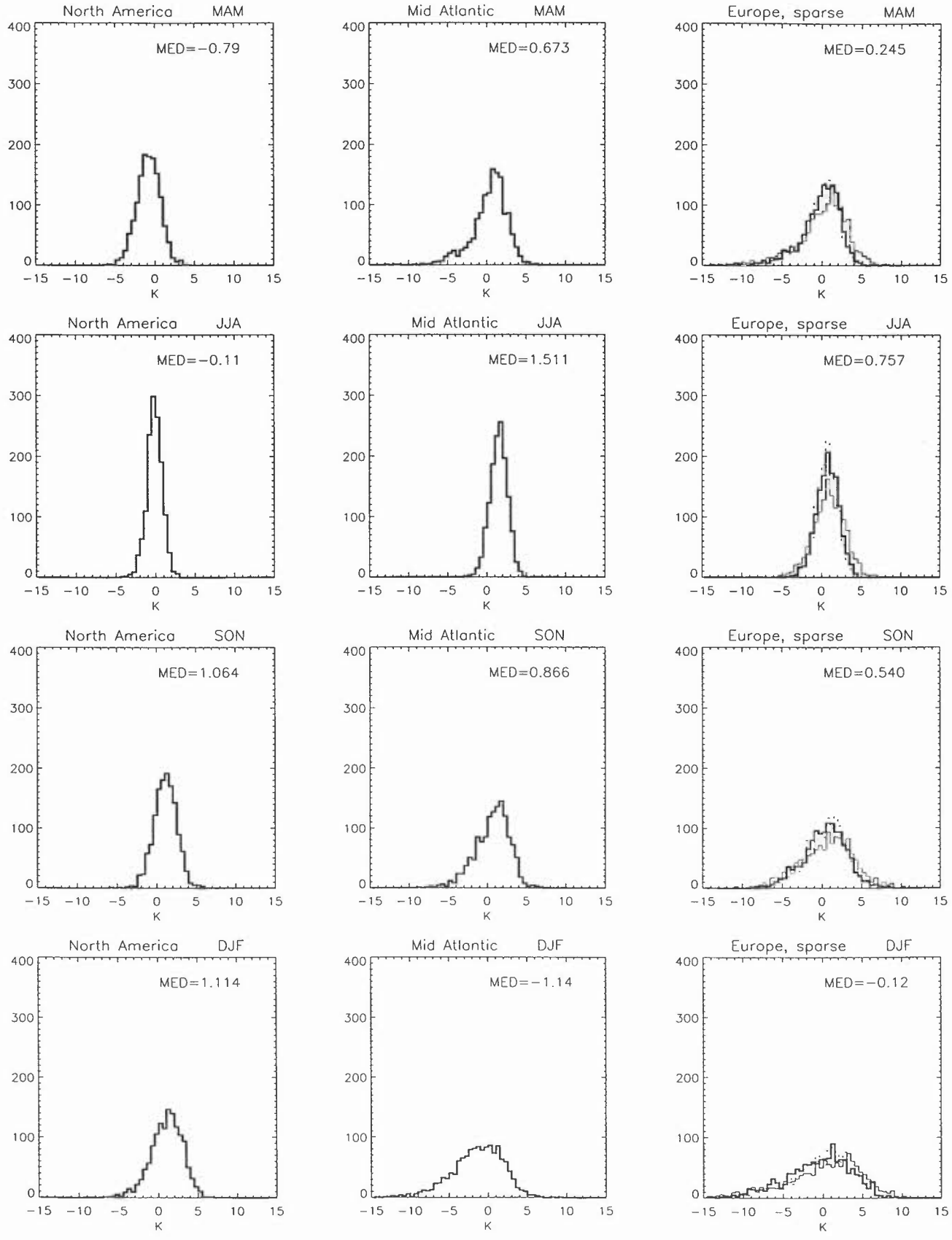




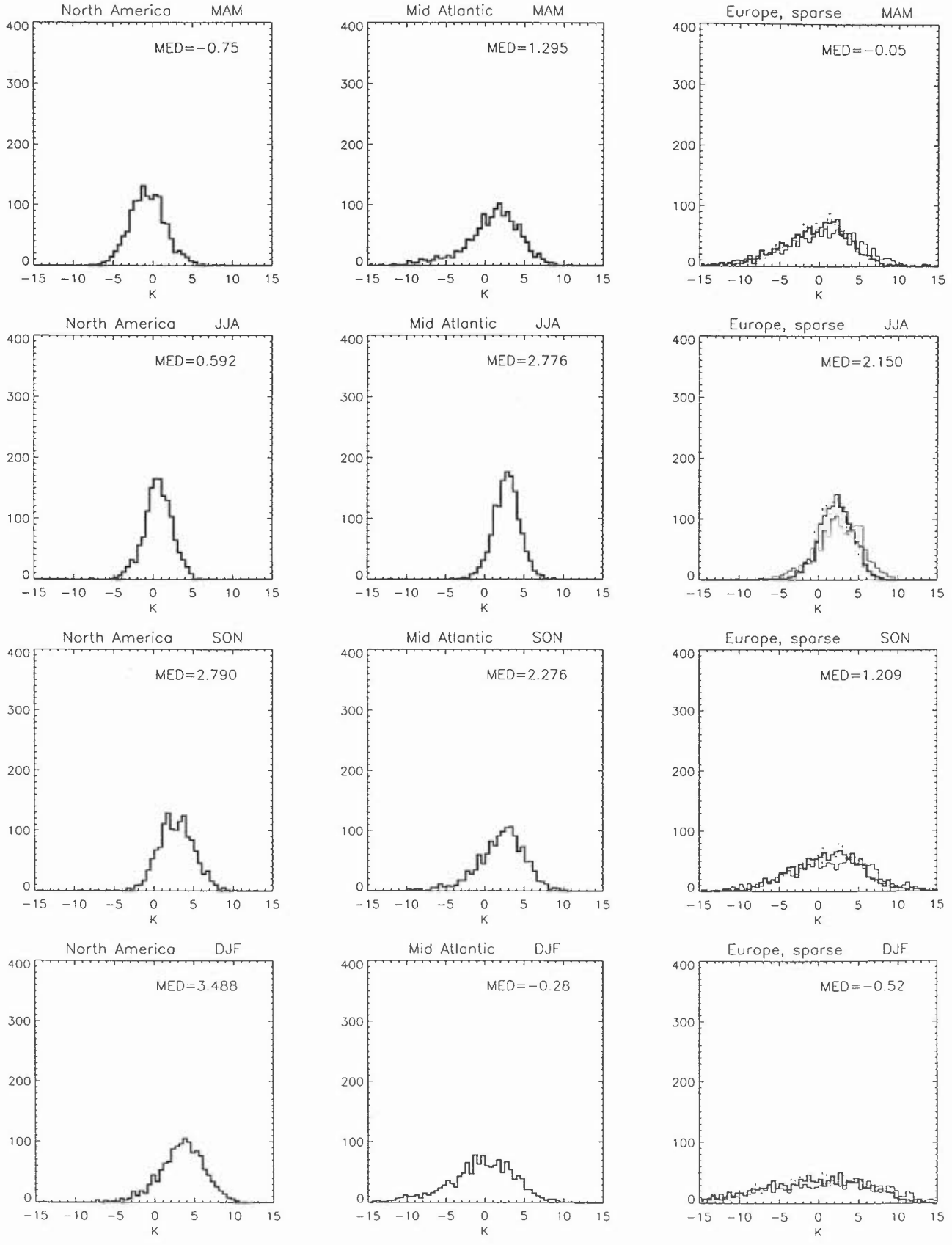




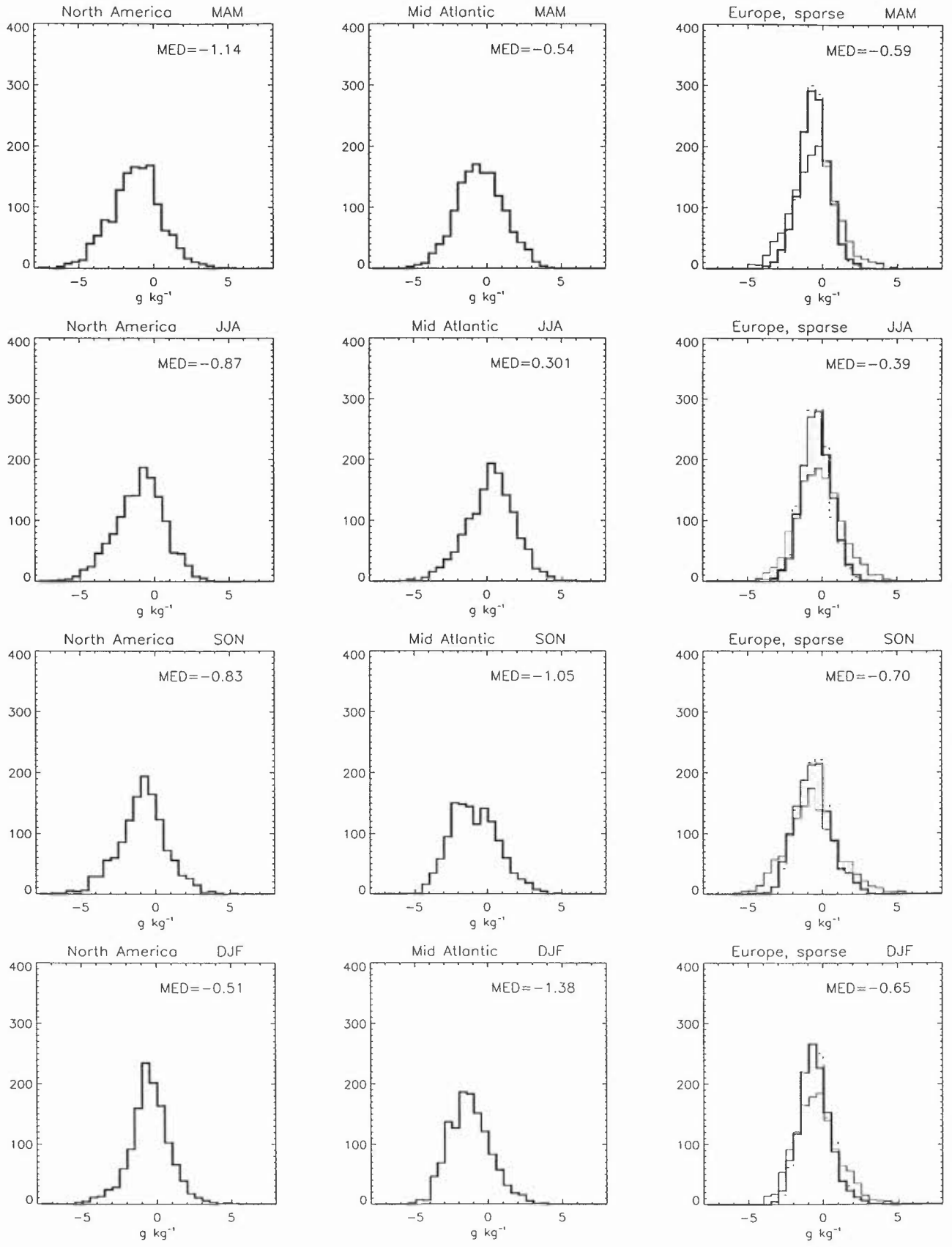
θ displacement 2 days backwards



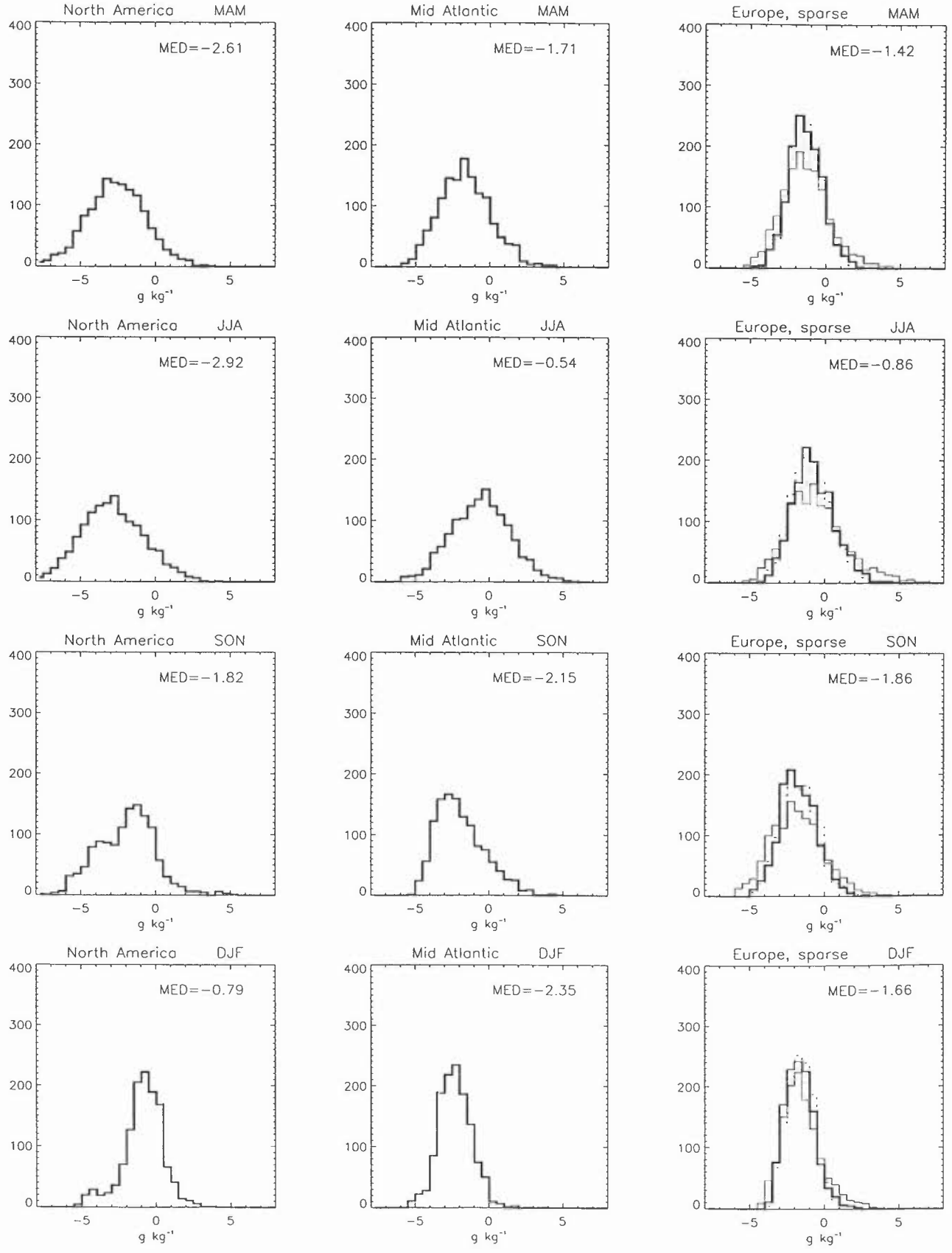
θ displacement 5 days backwards



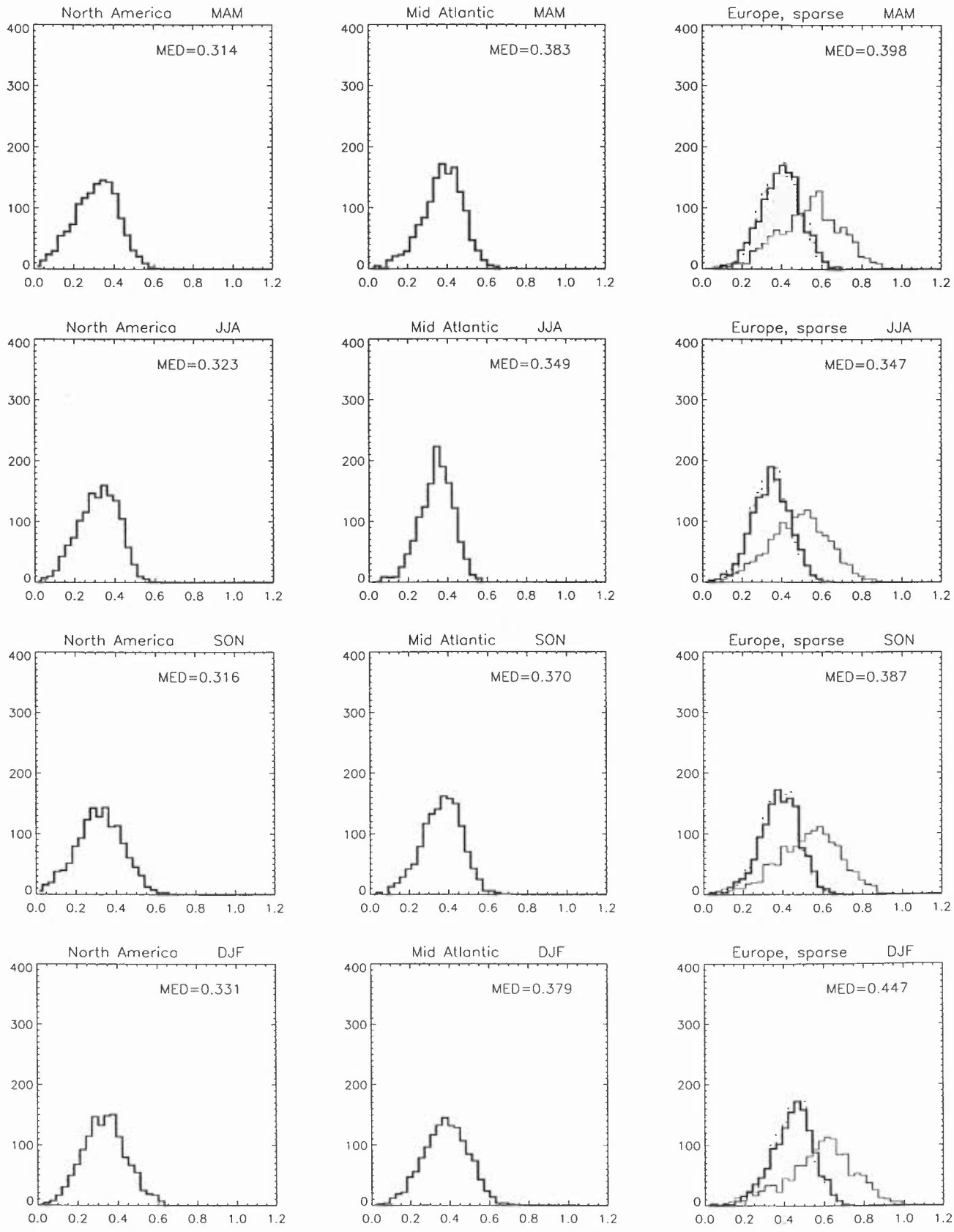
q displacement 2 days backwards



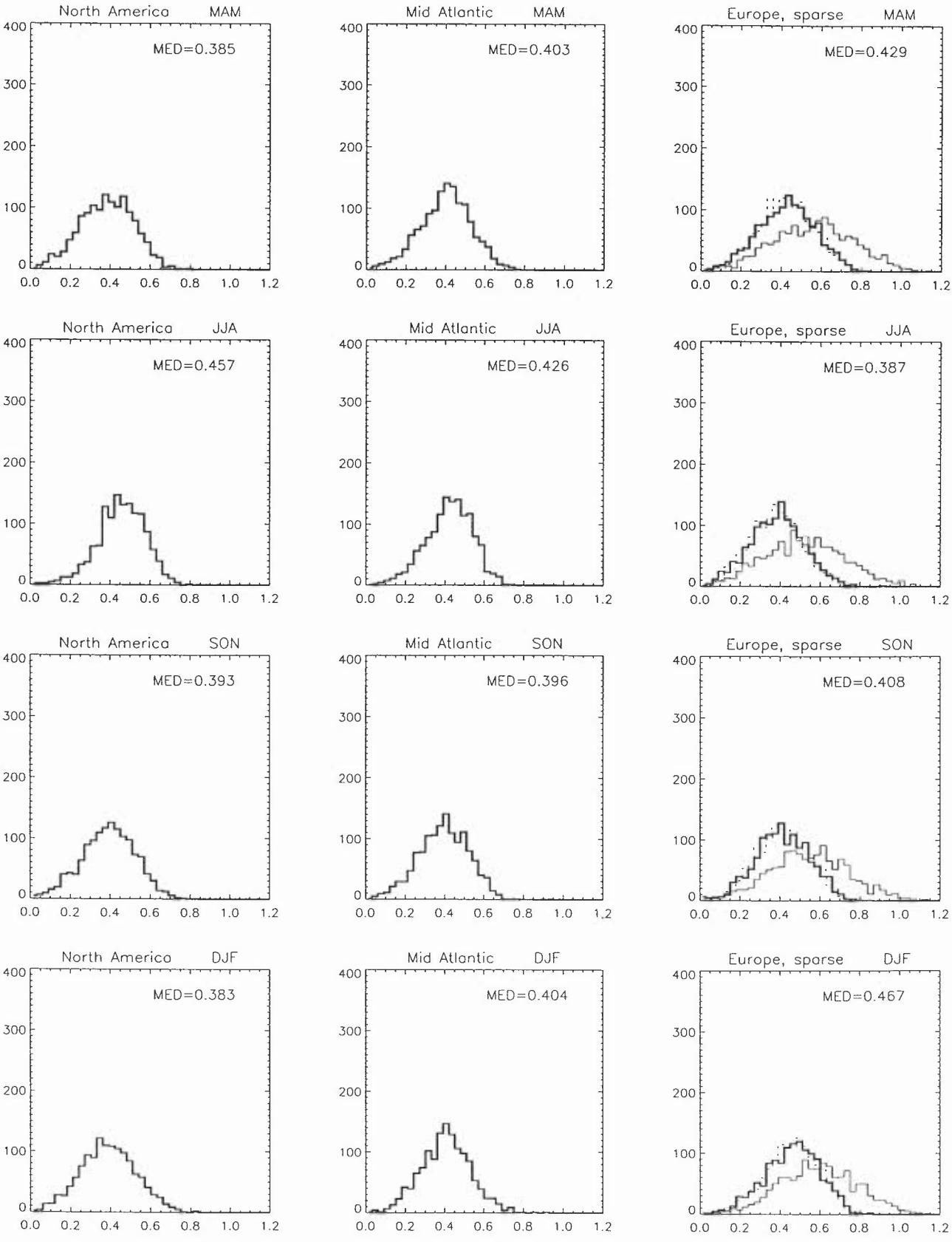
q displacement 5 days backwards



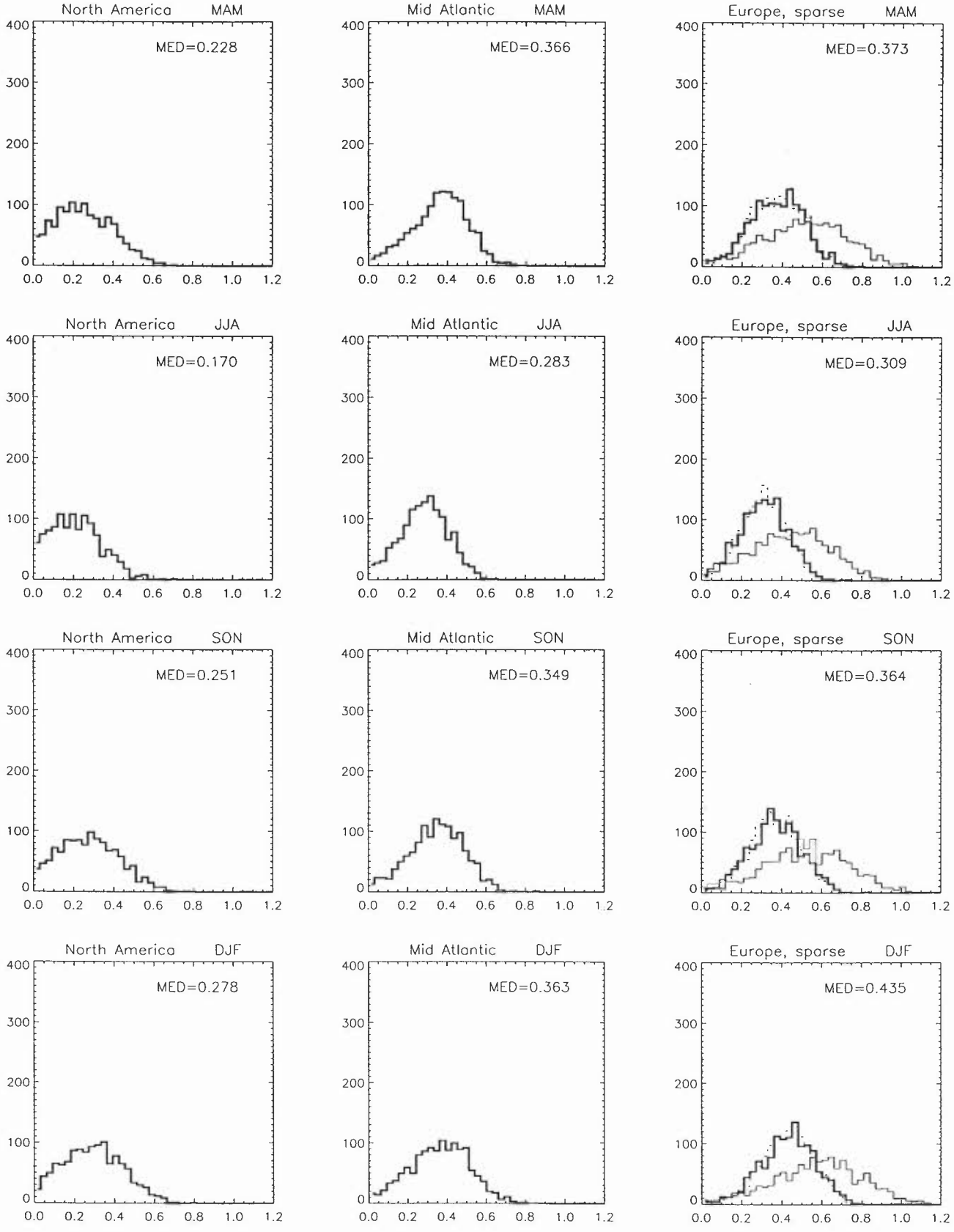
Time Exponent for RMS Spread



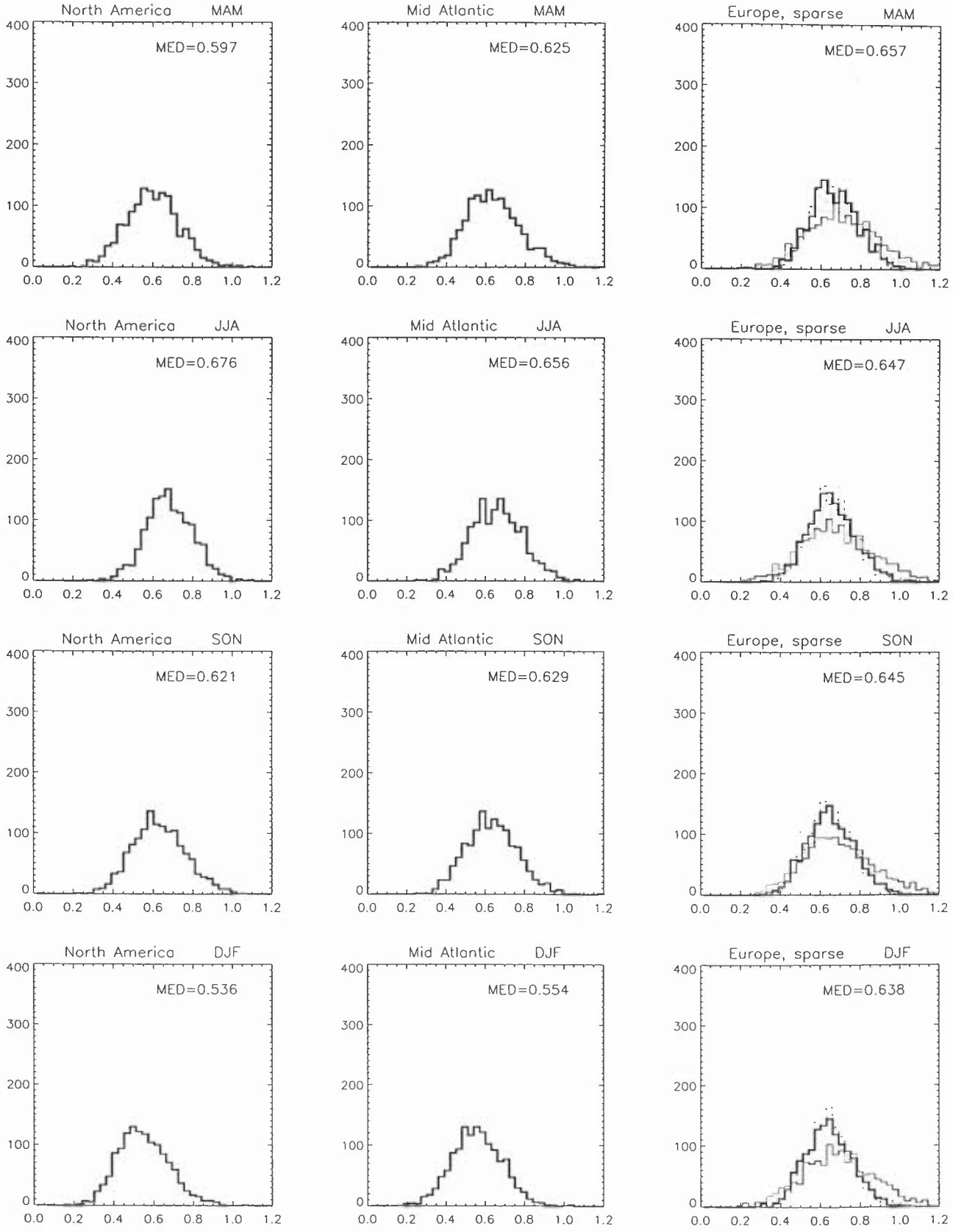
Time Exponent for Meridional Spread



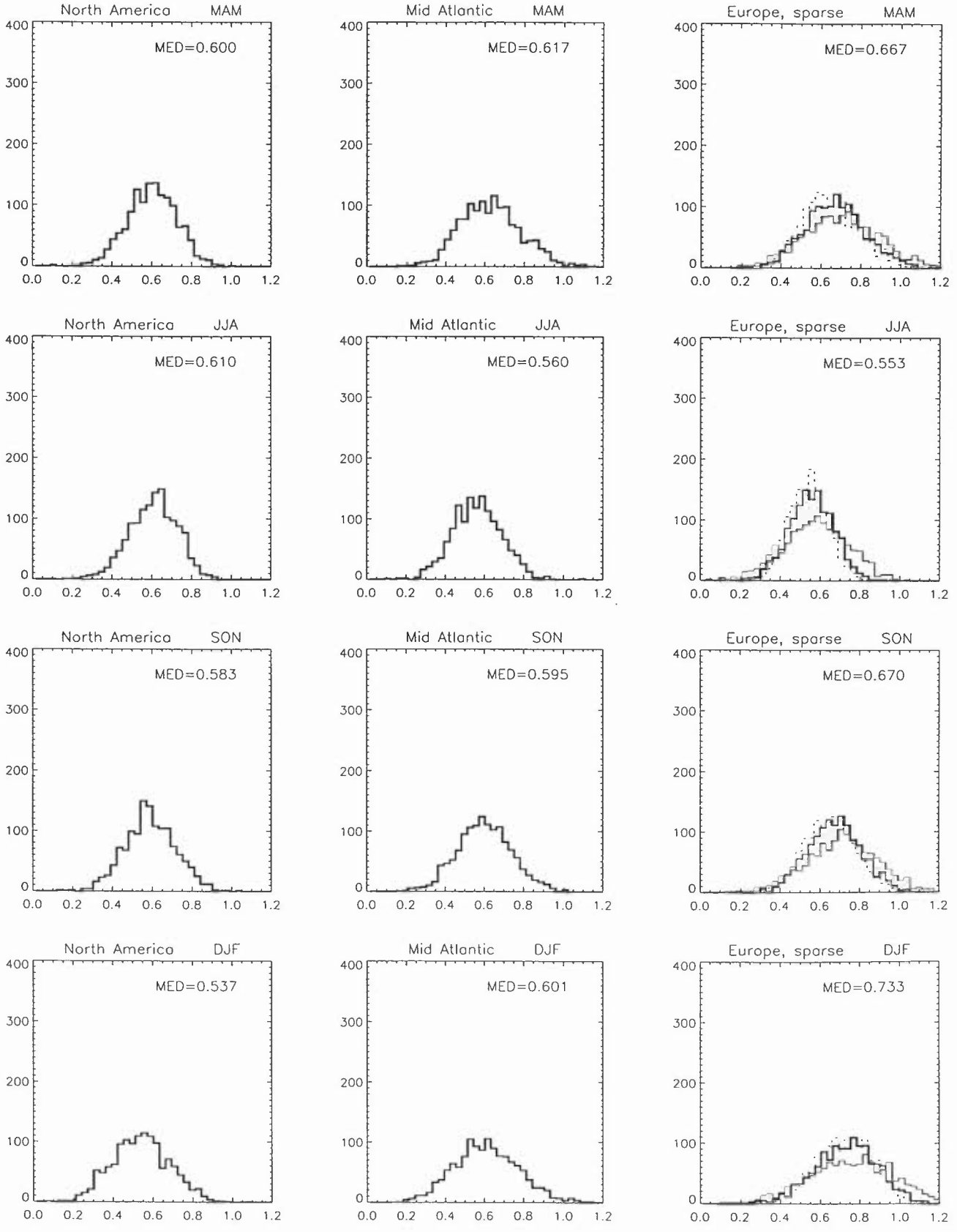
Time Exponent for Zonal Spread



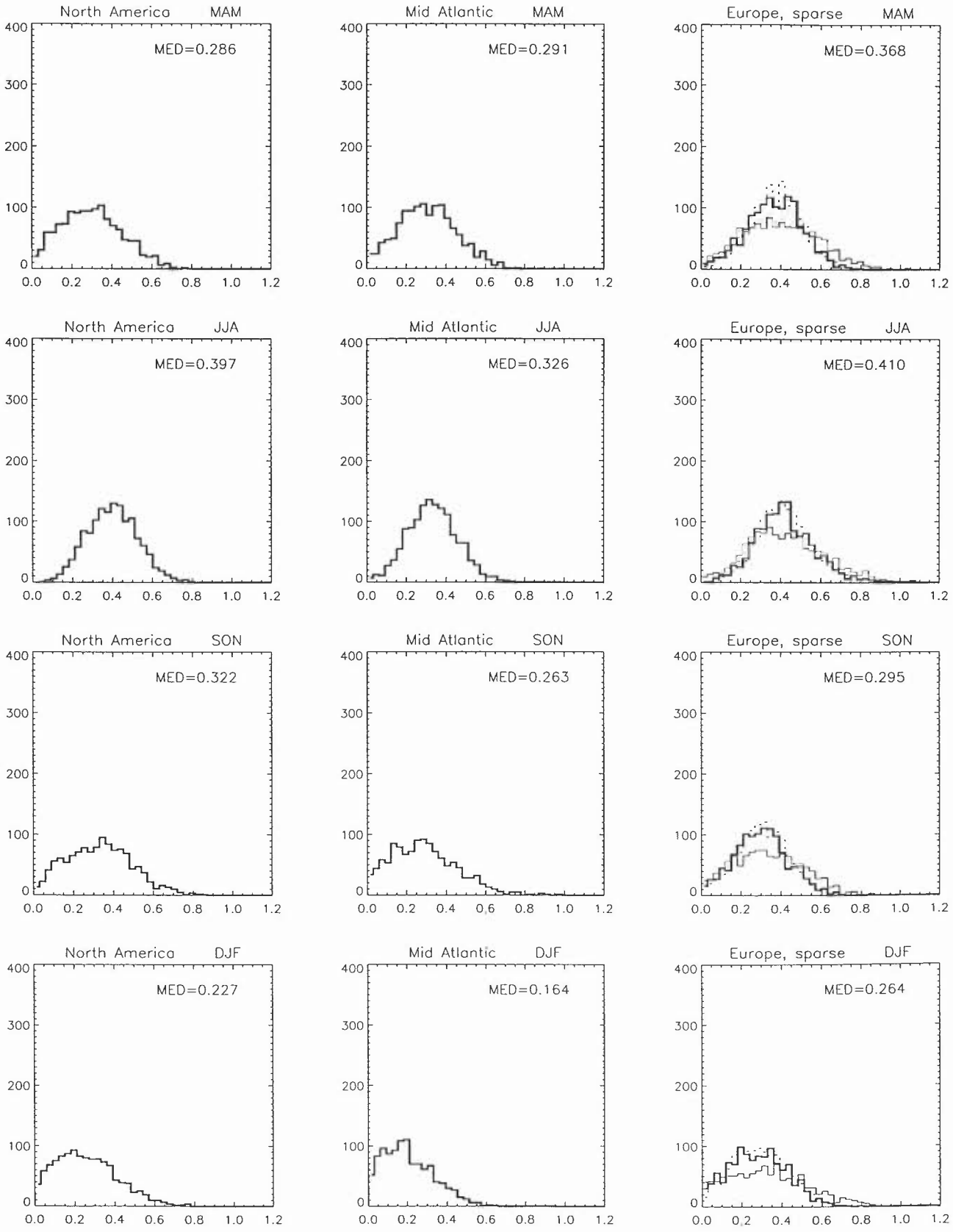
Time Exponent for Pressure Spread



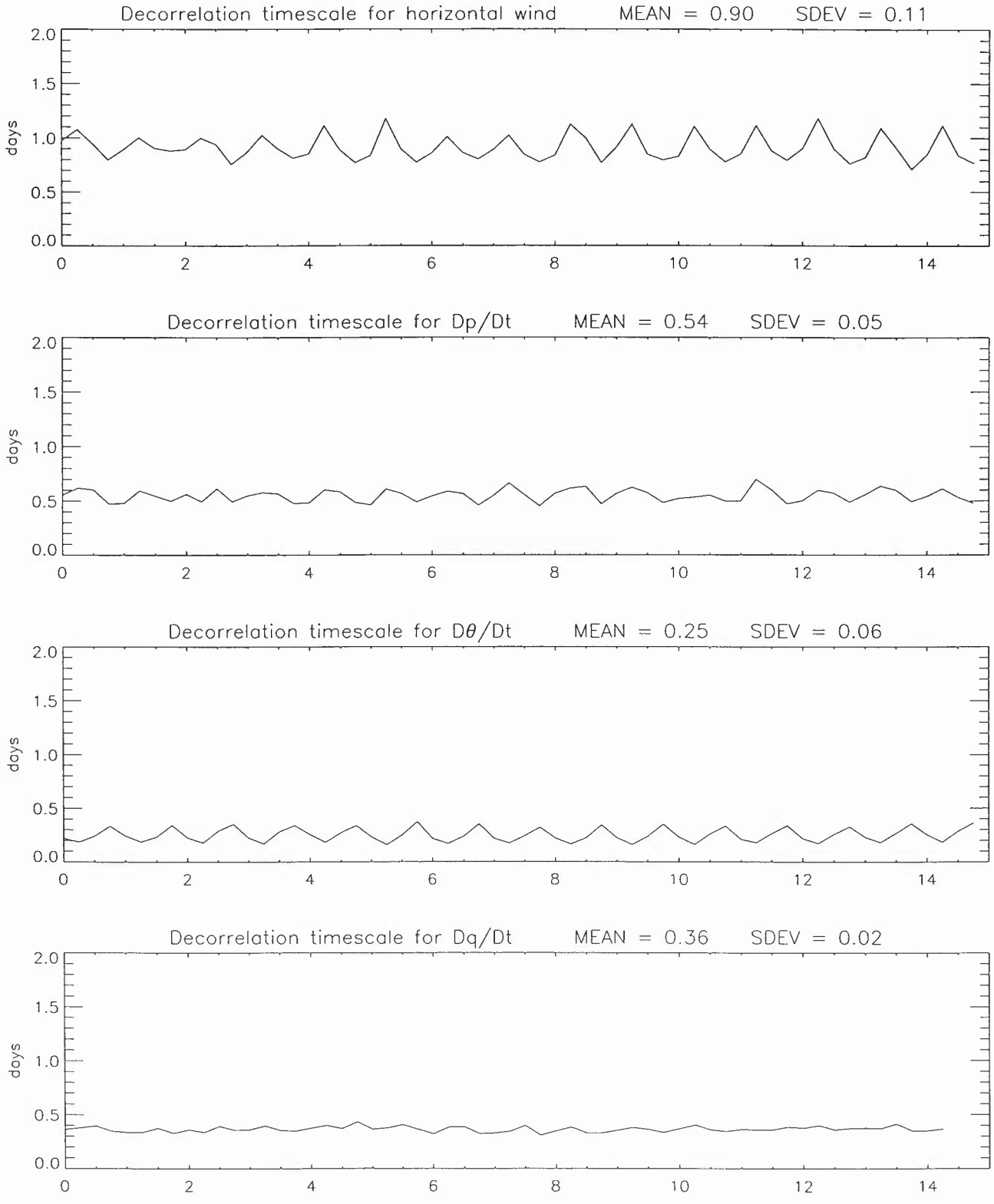
Time Exponent for θ Spread



Time Exponent for q Spread

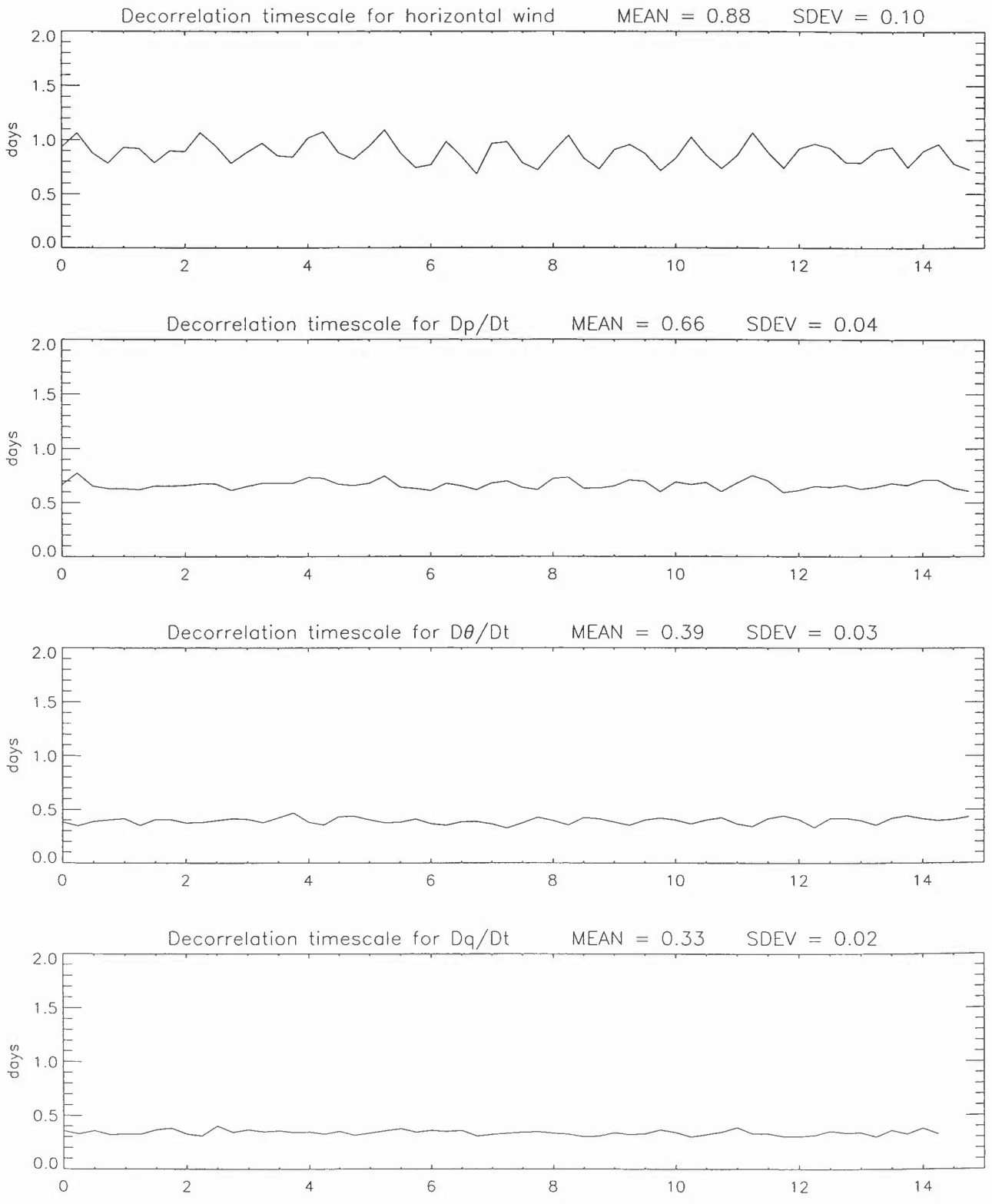


North America



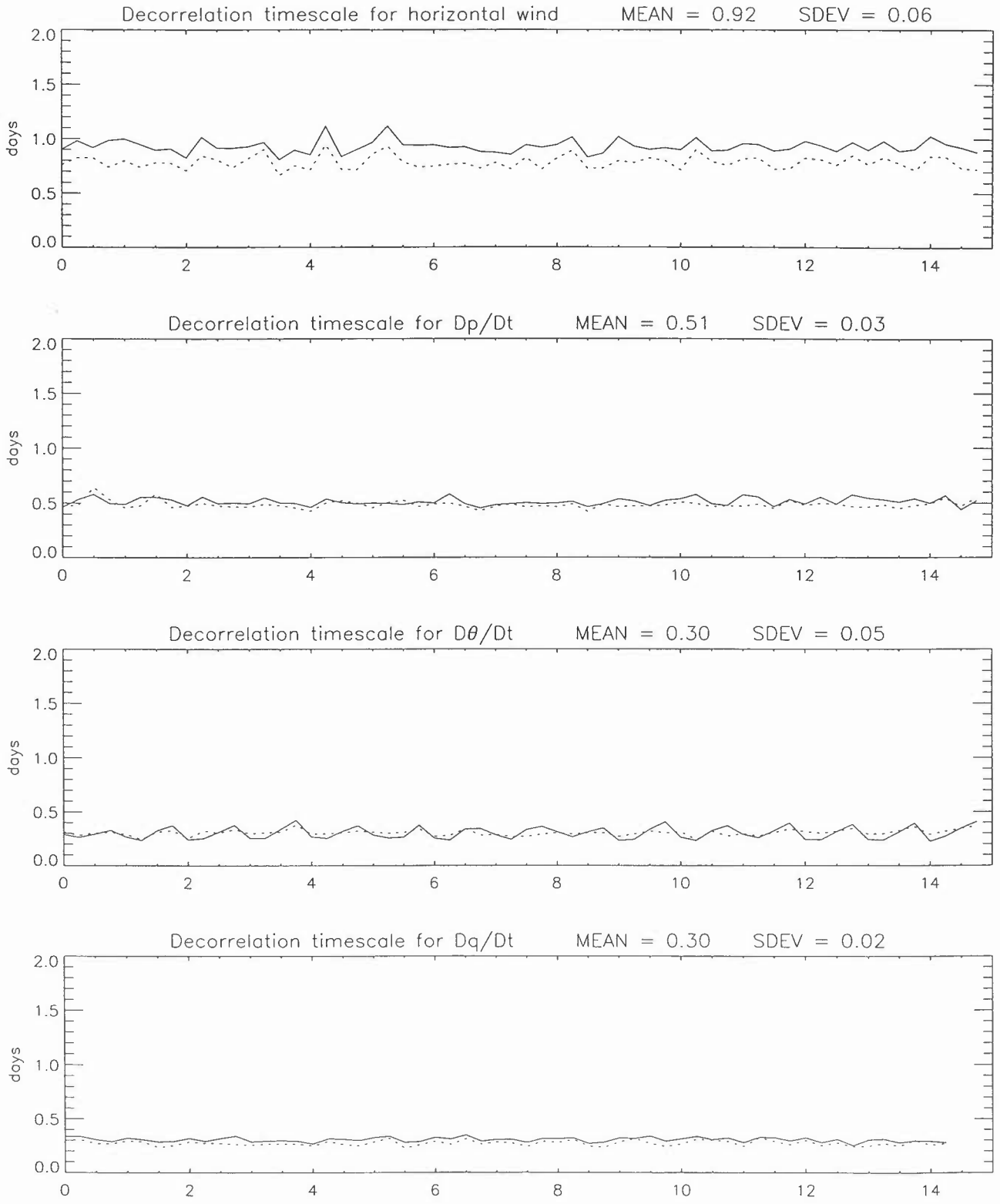
Years since 790301

Mid Atlantic



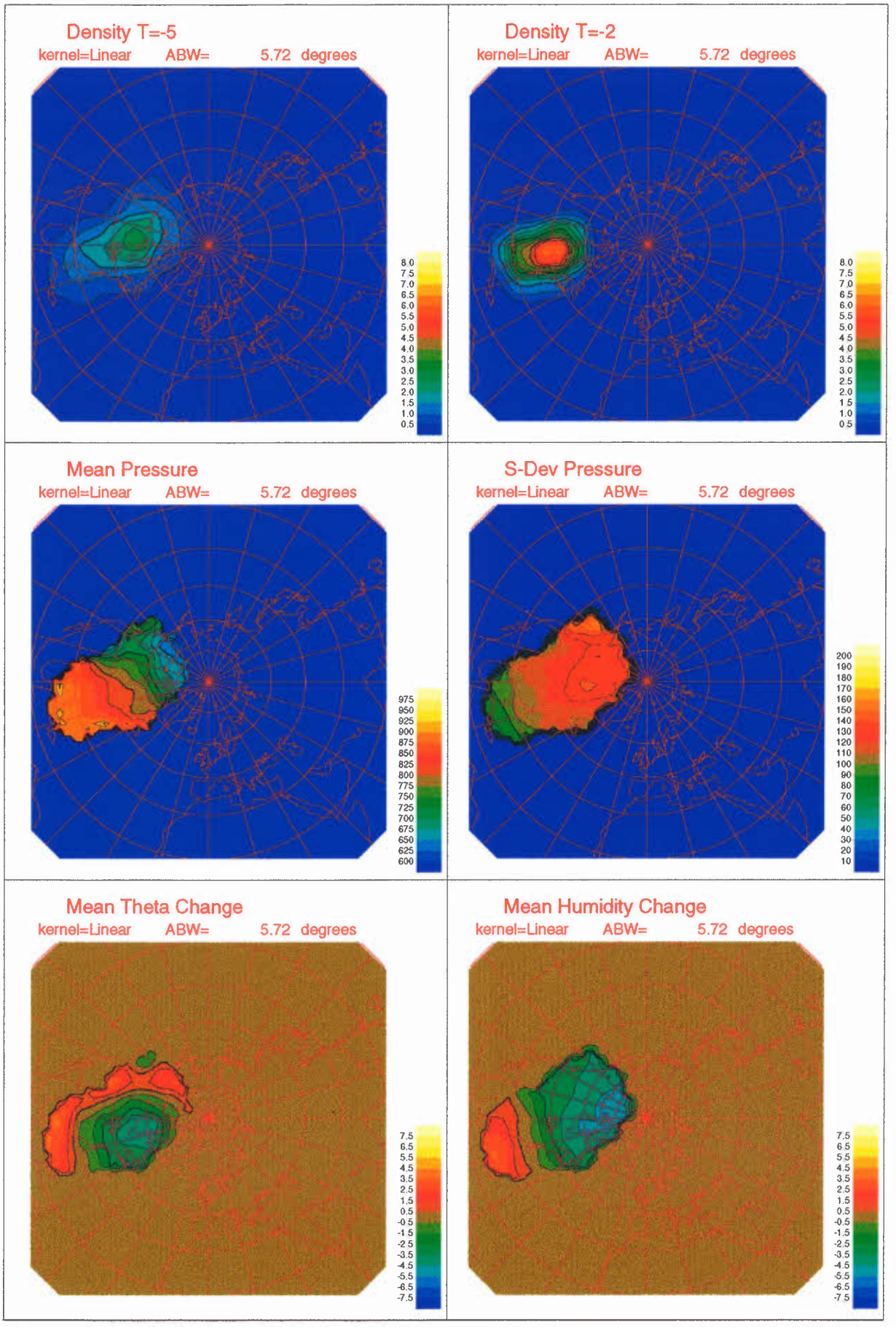
Years since 790301

Europe

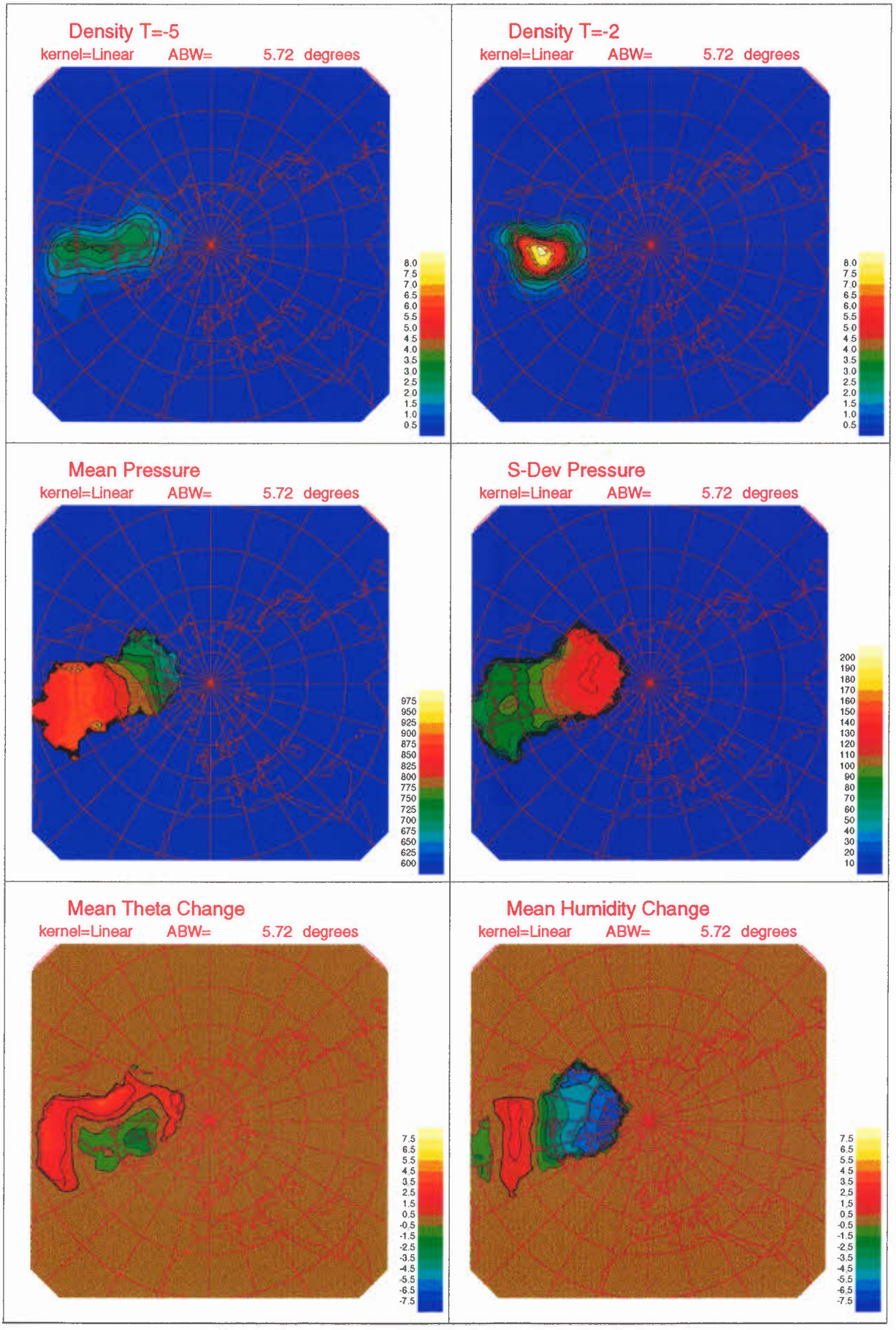


Years since 790301

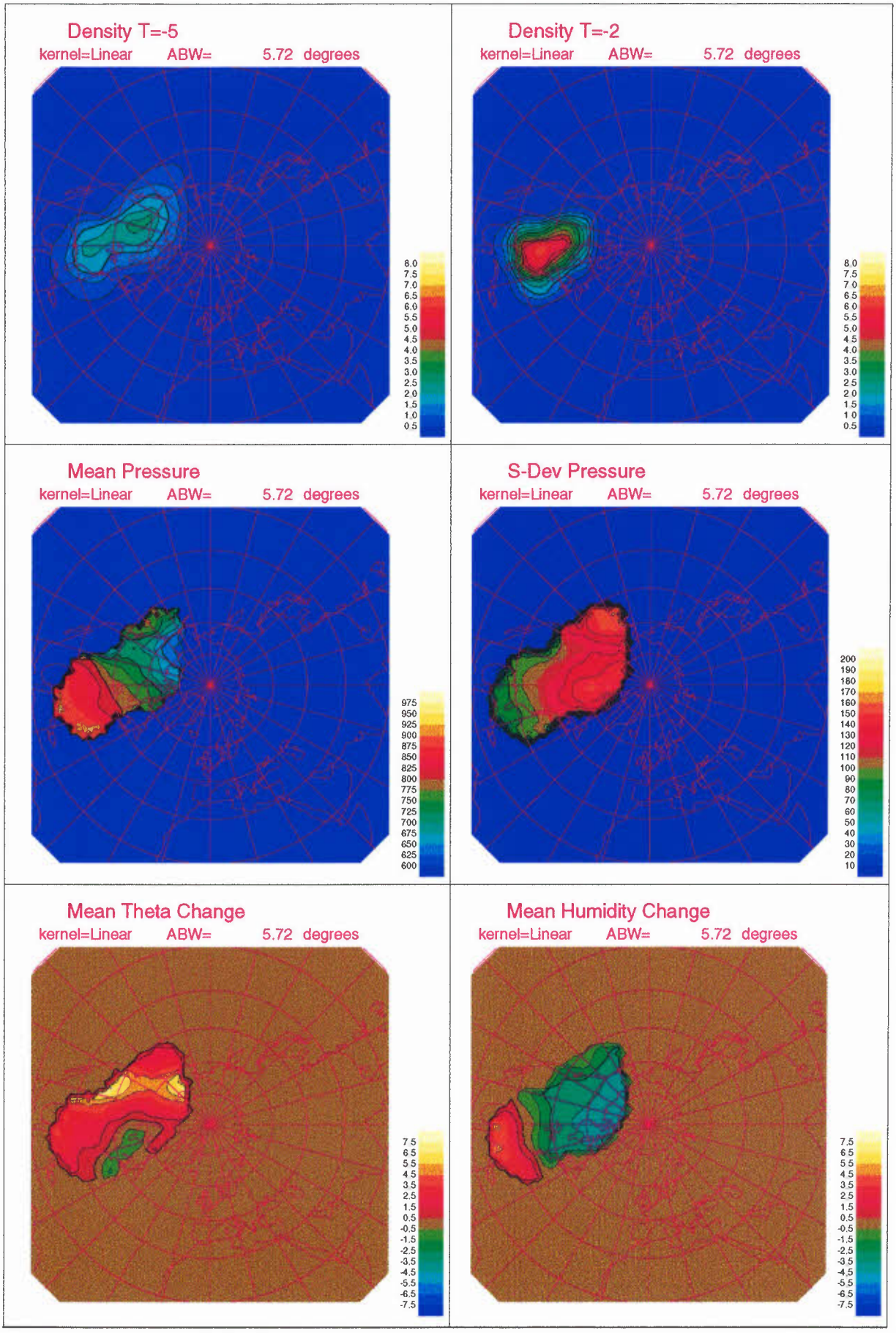
Northeast America MAM



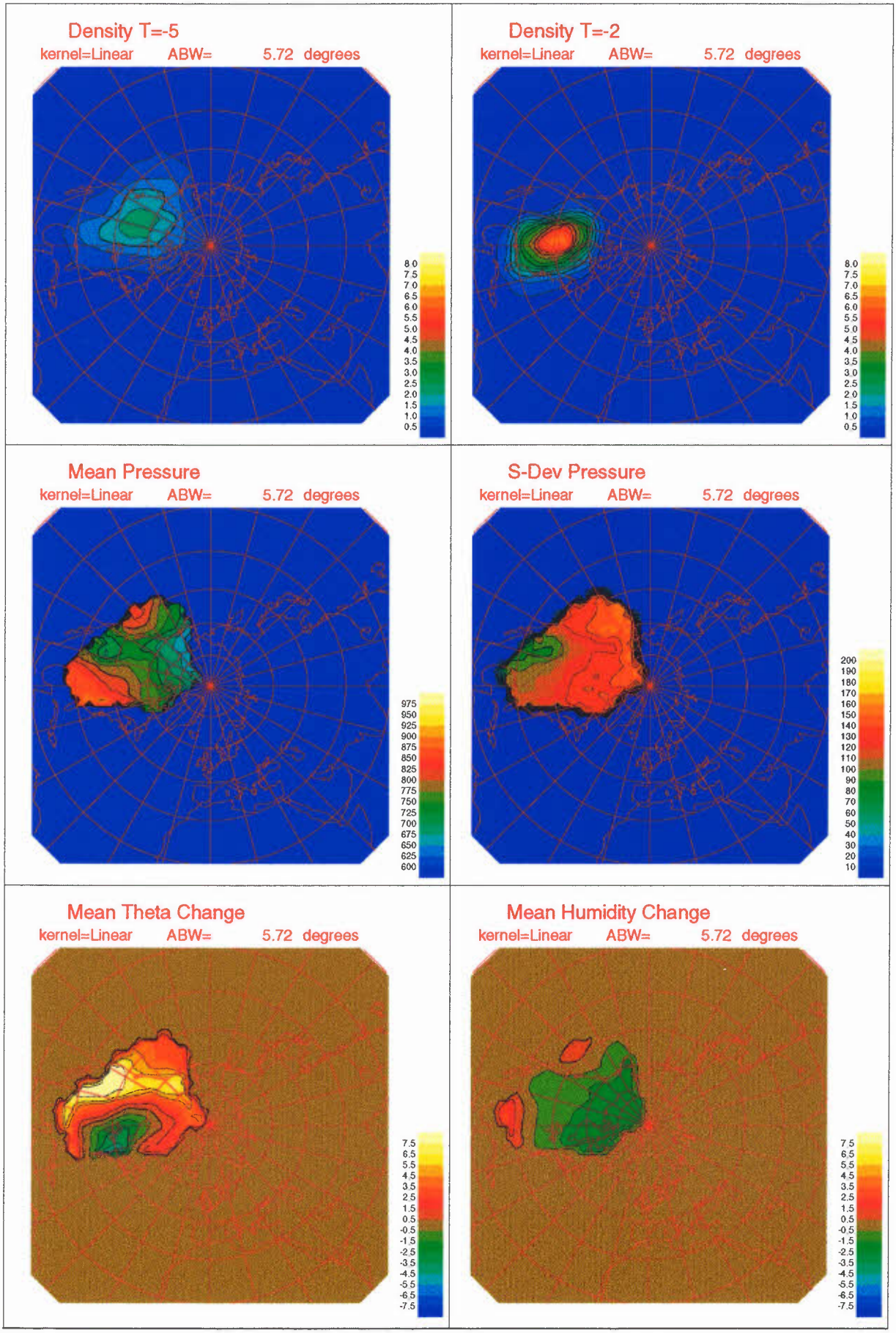
Northeast America JJA



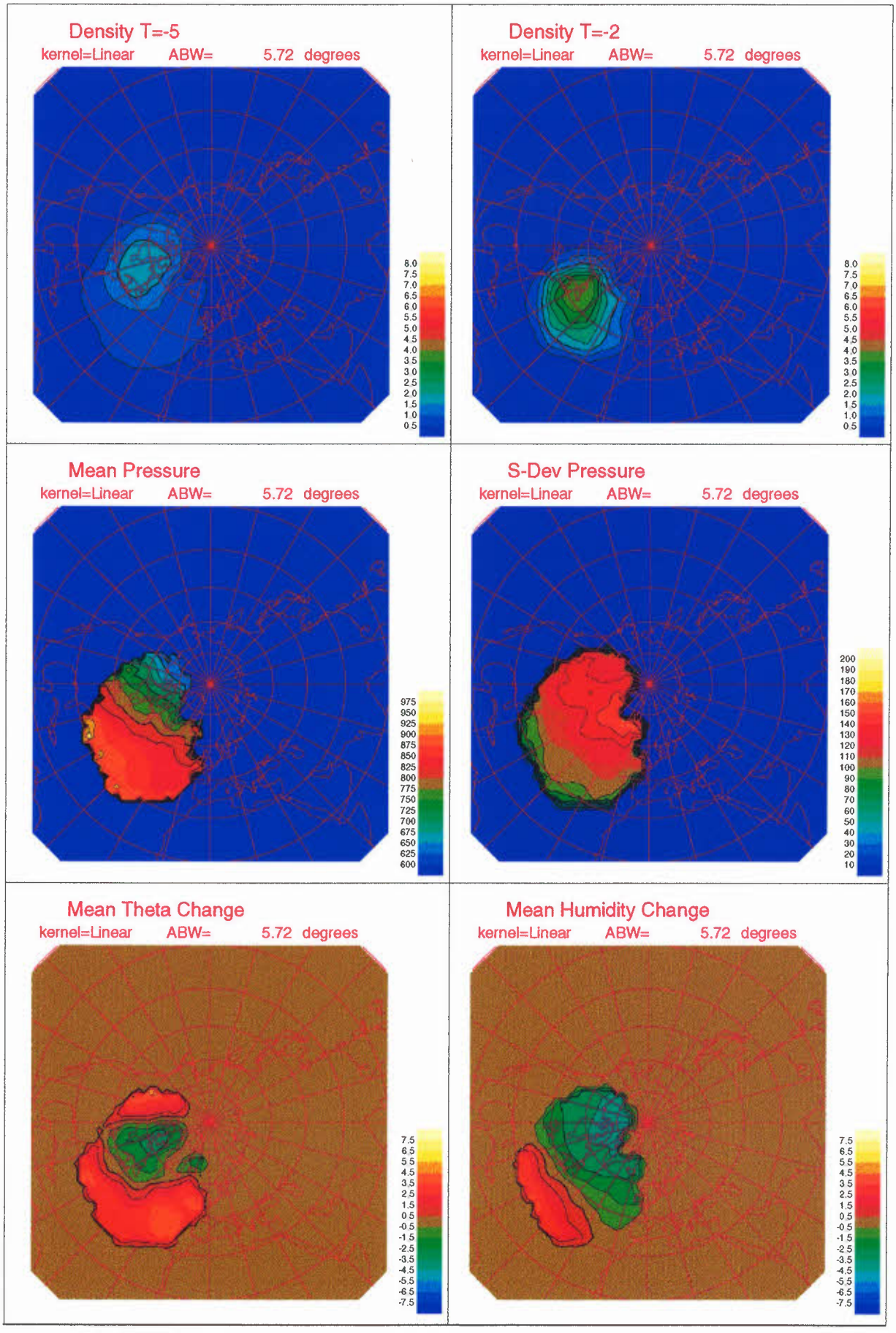
Northeast America SON



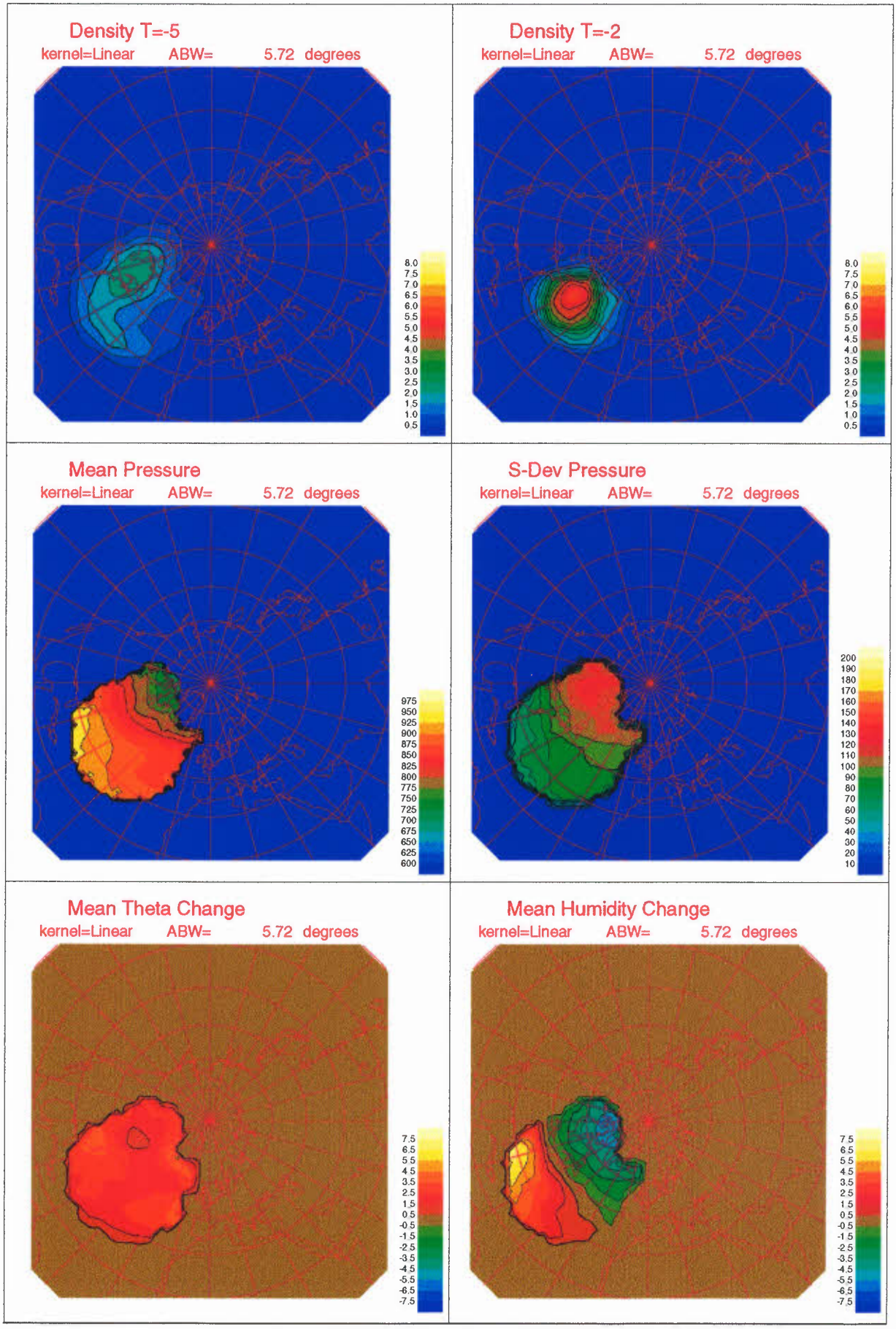
Northeast America DJF



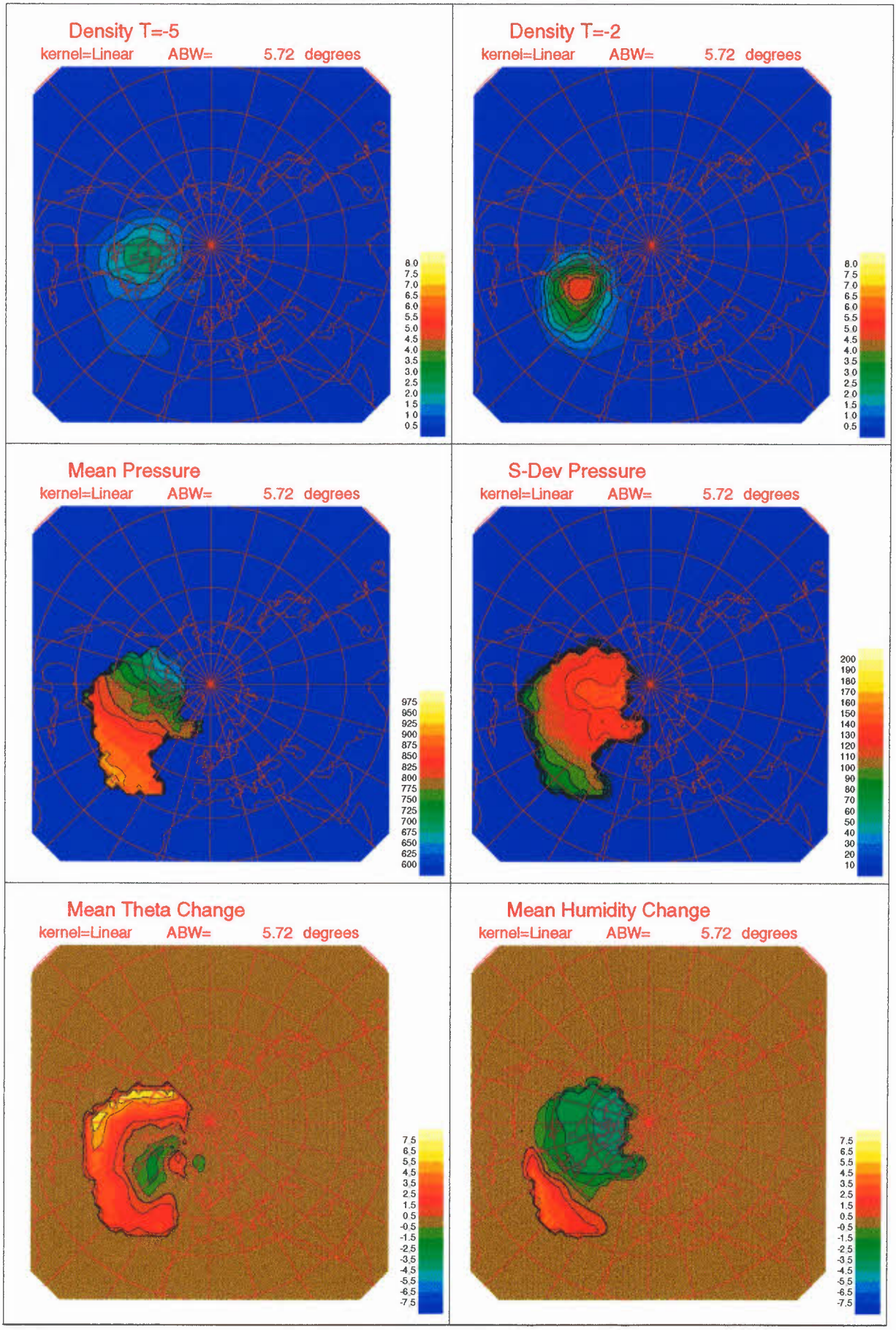
Mid Atlantic MAM



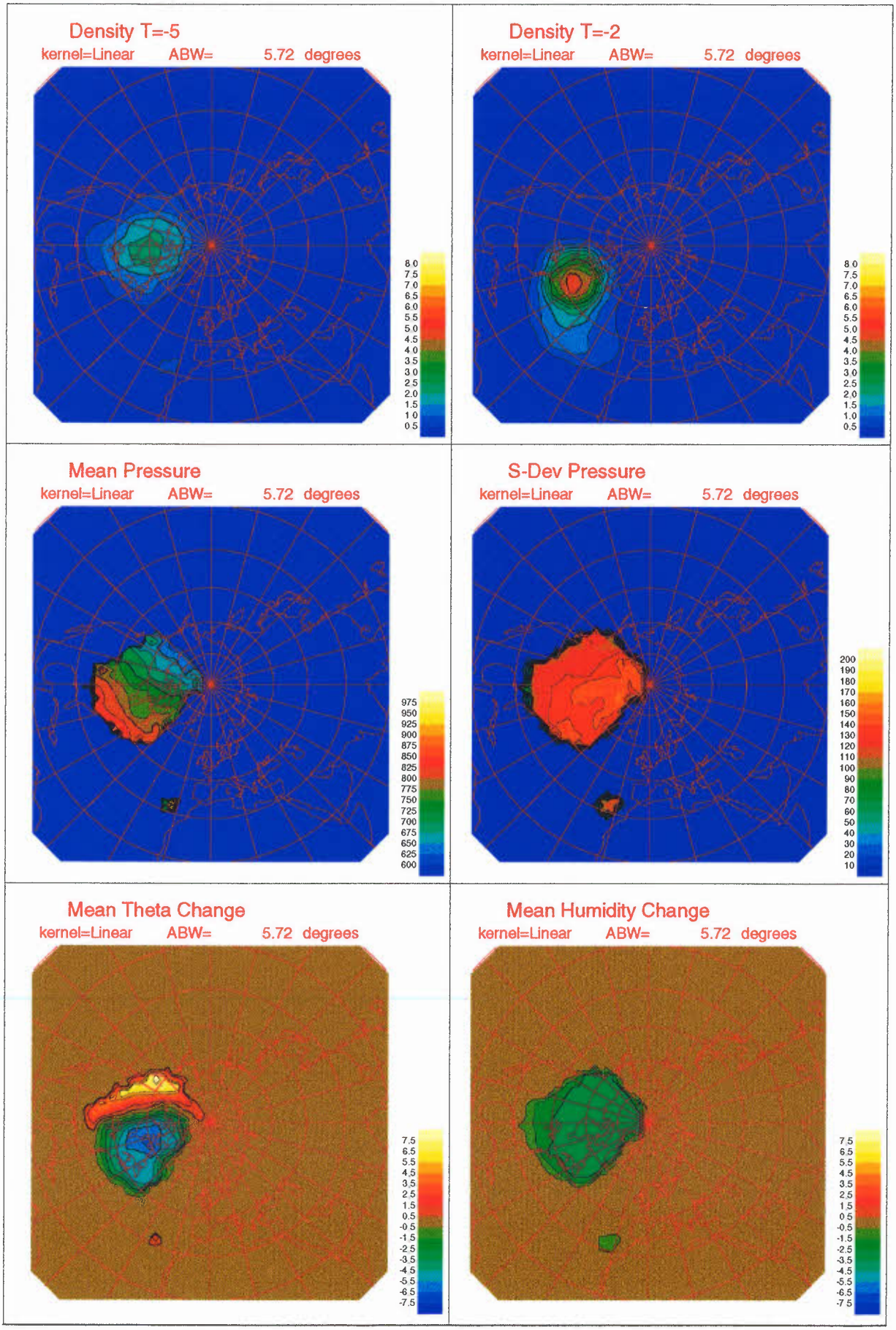
Mid Atlantic JJA



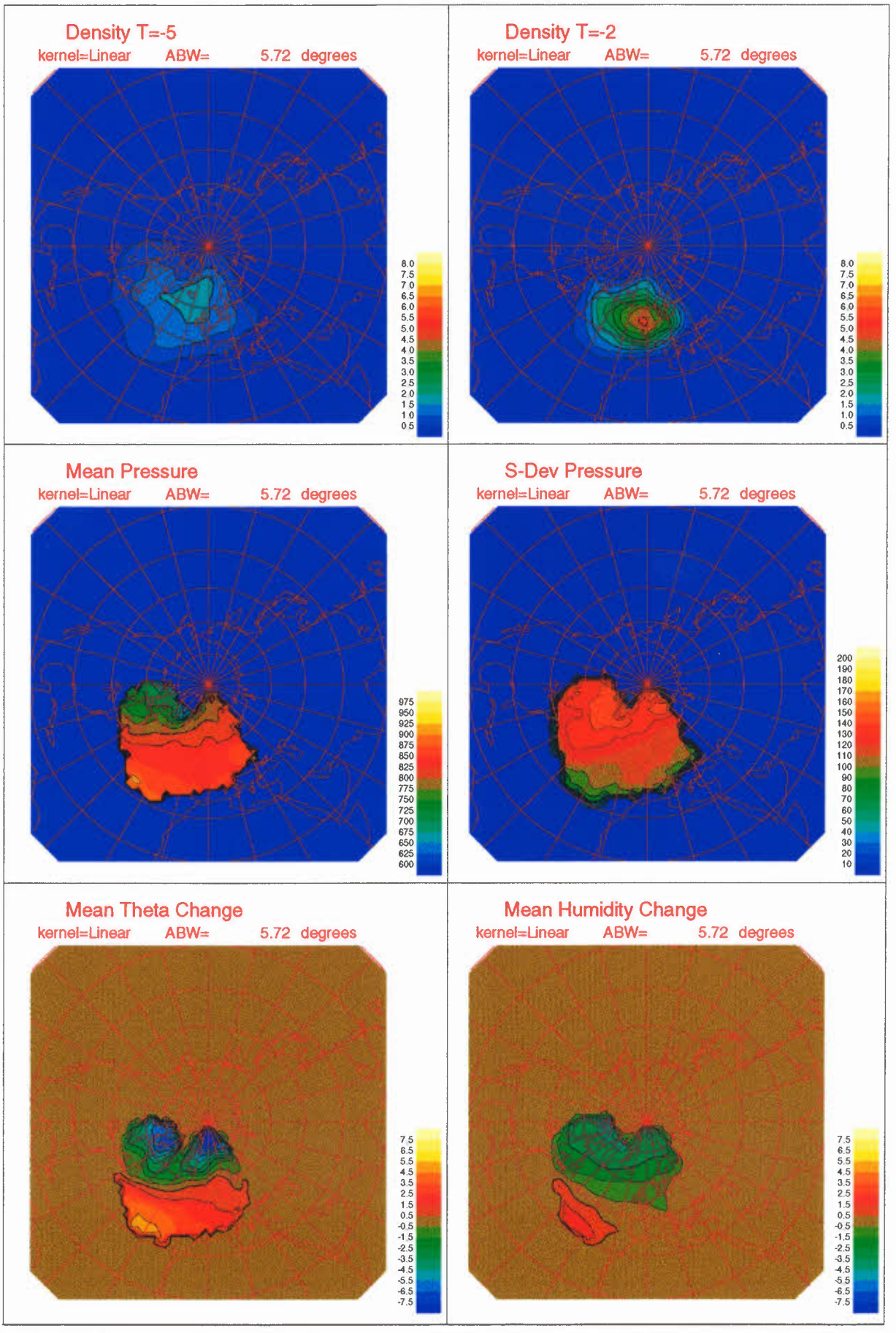
Mid Atlantic SON



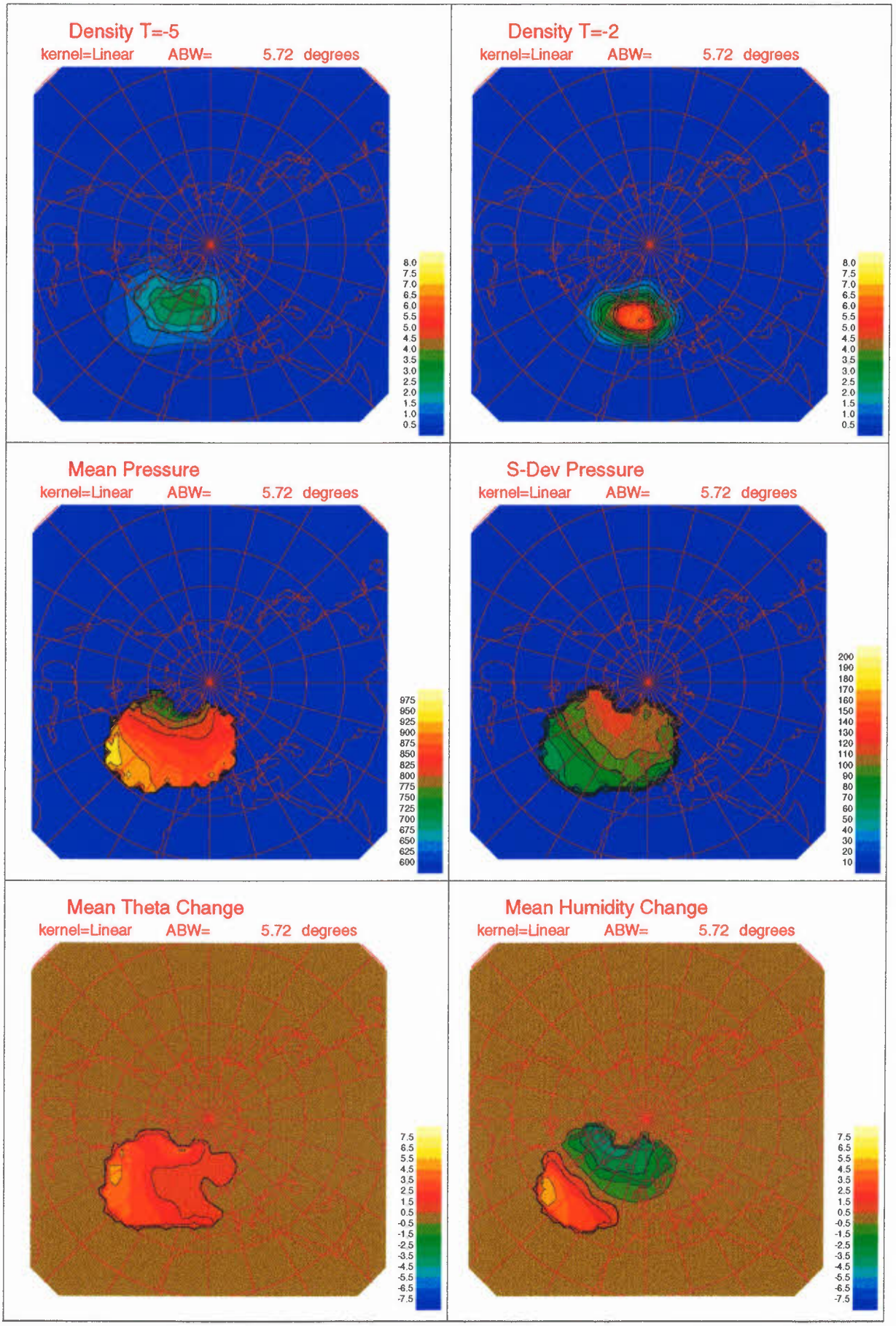
Mid Atlantic DJF



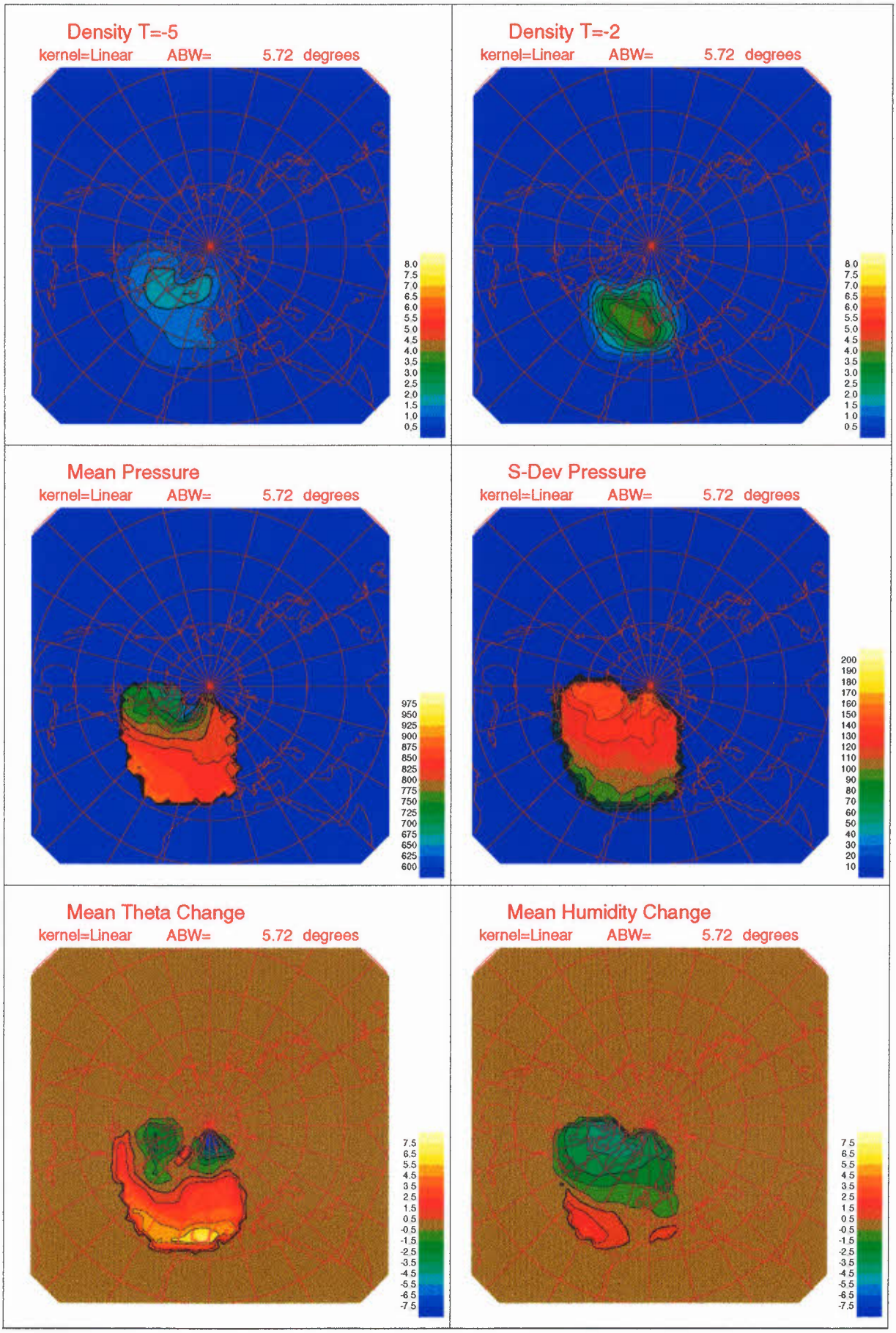
Northwest Europe MAM



Northwest Europe JJA



Northwest Europe SON



Northwest Europe DJF