Q. J. R. Meteorol. Soc. (2001), **vol**, pp. 1–12

A Probabilistic Formulation of Convective Mass Fluxes and its Relationship to Extreme Value Theory

By P. Naveau[1,2] * and M. W. Moncrieff[2]

[1] *Institut Pierre Simon Laplace, Laboratoire de Météorologie Dynamique, Ecole Polytechnique, France*
[2] *National Center for Atmospheric Research, Boulder, USA*

(Received ???; revised ???)

SUMMARY

The vertical flux of mass by atmospheric convection, and cumulus ensembles in particular, is a key quantity used in parameterization schemes for numerical weather prediction models and climate models. The concept of formulating mass fluxes in terms of statistical quantities is discussed and a probabilistic formulation of the upward mass fluxes, based on extreme value theory, is presented. This approach is illustrated by a statistical analysis of two-dimensional, cloud-resolving numerical realizations of two regimes of tropical convection, namely a squall-line and non-squall clusters. Our statistical model allows these distinct cloud systems, which play a key role in determining the shape of the mass flux profile, to be efficiently integrated. Also taken into account is the distinction between mesoscale and convective components of the mass flux.

KEYWORDS: Upward Mesoscale Convective Systems Generalized Pareto Distribution Maximum Likelihood Estimation

1. INTRODUCTION

Convective mass fluxes are widely used in the parameterization of cumulus convection in large-scale models. It is appropriate to consider cloud systems (populations or ensembles of cumulus) rather than individual cumulus. While numerically simulated cloud systems contain uncertainties arising from parameterizations of cloud microphysics and radiation, their macrophysical properties are more complete and consistent than observations provide. Atmospheric convection normally occurs on spatial and temporal scales much smaller than directly measurable. Therefore, cloud-resolving numerical models (CRM) are useful for calculating the convective mass flux explicitly. The key problem is then to reduce these synthetic data to simple analytic forms for use in convective parameterization. One way is to derive idealised dynamical models (Moncrieff 1981), while another is to use statistical methods, which is the approach we adopt in this paper.

The standard analytic formulation for the mass flux used in convective parameterizations originates from the Squires and Turner (1962) model of an entraining plume. It is given by $\mathcal{M}(z) = \mathcal{M}_0 \exp(-(\epsilon_D - \epsilon_E)z)$, where ϵ_E and ϵ_D are the fractional entrainment and detrainement rates, respectively, which have typical atmospheric values of order $10^{-4}\, m^{-1}$ and satisfy $\epsilon_D > \epsilon_E$. We refer to Siebesma (1998) for a review and assumptions implicit in this mass flux formula. While useful in special cases, the exponential decay in the analytic form of $\mathcal{M}(z)$ has limitations and does not usually correspond to mass flux profiles computed from convective realisations. Consequently, more flexibility and non-homogeneity should be incorporated into the mass flux formulation.

Since cloud system simulations provide a comprehensive source of synthetic data for tropical convection, it is natural to develop a new probabilistic upward mass flux method based on statistical principles. The parameters (coefficients) of our statistical model have a physical interpretation, being closely related to the distinct cloud populations that occur in different shear profiles. The statistical model is based on extreme value theory, which theoretically justifies this approach and is explained below. More importantly, the wind

* Corresponding author: Institut Pierre Simon Laplace, Laboratoire de Météorologie Dynamique, Ecole Polytechnique, France

extremes play a significant role in the computation of the upward mass flux, and therefore the statistical theory of extremes ought to be important in the stochastic formulation of the mass flux. In addition, a key aspect is the organising effect of environmental (i.e. mean flow) shear on convection as shown by Moncrieff (1981). Two scales are identified – convective and mesoscale, a property seen in Figures 3 and 6 from Liu and Moncrieff (1998; hereafter LM98). We show that this decomposition is closely related to the problem of separating extreme velocities from the rest of the motion spectrum.

Besides providing a new description of the upward mass flux, extreme velocities from CRM outputs is in itself of strong interest. The statistical properties of such quantities, essential to convection, have not previously received much attention, despite the increasing popularity in the statistical analysis of extremes in atmospheric science. For example, a couple studies (Zwiers et al. 1998 and Kharin et al. 2000) examined how well general circulation models simulate extremes. Our work has a similarity to such studies: the statistical analysis of numerically simulated outputs and the use of extreme value theory to quantify the tail distribution of important atmospheric variables.

Our paper is organised as follows. In the next section we present the synthetic data sets and the notation for the upward mass flux. The decomposition of this mass flux and its relationship to total condensate is undertaken in Section 3. In section 4 the distribution of vertical velocity exceedances is studied. In section 5 we describe the key statistical procedures, interpret the coefficients of the statistical model in physical terms, and describe the mass flux computation including the convective and mesoscale decomposition. The paper concludes in section 6.

2. Data sets and notation

(*a*) *Synthetic data from resolved convection*

Atmospheric convection tends to be dynamically organised or coherent on spatial scales ranging from a few kilometers to about a thousand kilometers (the mesoscale). While these mesoscale systems have been extensively analysed in field experiments and numerically simulated, the attendant effects on the large-scale circulation of the atmosphere are not well understood.

LM98 conducted two 12-day, two-dimensional numerical experiments of tropical convection. One was in jet-like shear, which is associated with squall-line type organisation. In the other, the wind shear was removed, but the same thermodynamic environment was used, yielding less coherent *non-squall clusters*. Specifically, we used 576 temporal snapshots from LM98 taken every 30 minutes over the spatial domain 600×26 km (the original data had a 10-second resolution). The horizontal resolution was 2 km and the vertical resolution ranged from 0.1 km in the boundary layer to 1 km in upper levels. The condensate cloud fields from these two experiments are plotted in figures 3 and 6 of LM98. These systems persist because the convective available potential energy was continuously generated by a prescribed domain-scale advection of temperature and moisture. As illustrated in Fig. 2 from LM98, these representative series of long-term data are ideal for statistical analysis. Six variables [wind velocities (U, W), potential temperature and different categories of moisture (cloud, rain and ice)] characterised the cloud systems and were available at each grid point.

(*b*) *Trajectories and exceedances of vertical velocity*

The vertical velocities play an important role in the computation of the upward mass flux. It is therefore essential to assess their intensities and locations in time and

space. As an example, we plot the locations of strong vertical velocities (i.e. greater than a given threshold, see Section 3 for the methodology used to choose the threshold). These traces correspond to the squall-line case in Fig. 1 and to the cluster pattern in Fig. 2, respectively. The intensities associated with these traces are defined as *exceedances* that measure by how much the velocity exceeds a prescribed threshold. Besides capturing

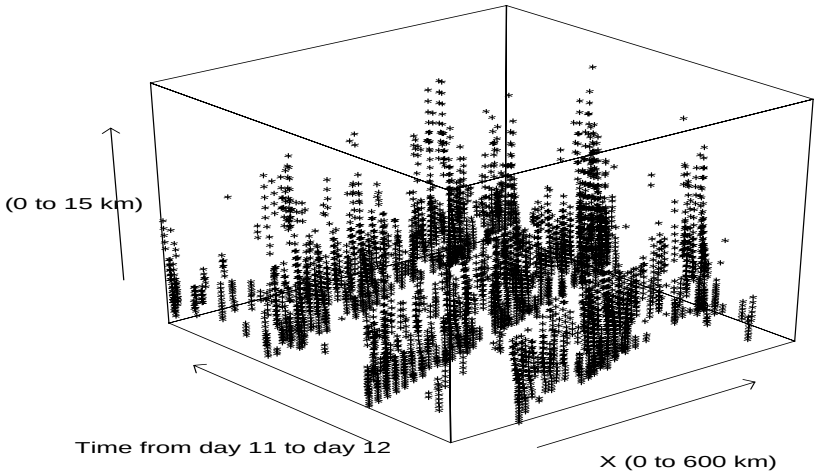

Figure 1. Squall-line case: Positions in time and space where vertical wind velocities exceed 4 ms^{-1} during the last two days.

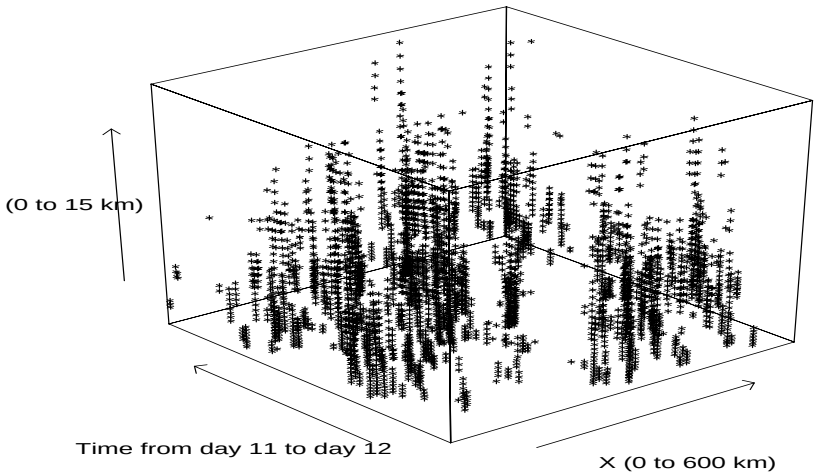

Figure 2. Non squall-line cluster case : same variable as in Fig. 1.

some of the main dynamical structures, the thresholding procedure has two other main advantages. First, it greatly reduces the data set by retaining only the key characteristics. Indeed, the majority of the vertical velocity values are approximately zero over most of the computational grid. Only large values of W contain relevant information regarding cloud trajectories for the upward mass flux computation (Fig. 1). The second advantage is that probabilistic models (Leadbetter et al 1983 and Embrechts et al

1997), as well as statistical tools (Reiss 1997, Coles 2001), can describe the distribution of exceedances. These theoretical results and statistical devices are key aspects in our statistical formulation of the upward mass flux.

(c) *Upward mass flux definition*

In each of the two simulations described above, the cloud systems (either squall lines or non-squall clusters) maintain an overall self-similar form. It is therefore legitimate to define a composite upward mass flux as:

$$\mathcal{M}(z) = \frac{1}{\int dt dx} \int \rho(z) W(x,z,t)\ \mathcal{I}[0 < W(x,z,t)]\ \mathcal{I}[Q_c(x,z,t) > q_0] dx dt$$

where ρ is the air density which is assumed to be only a function of height z, Q_c is the total condensate (rain water, cloud droplets and ice), $q_0 = 0.1$ gm^{-3}, and $\mathcal{I}[A]$ denotes the indicator function which is equal to 1 whenever A is true and 0 if it is false. In practice, we estimate $\mathcal{M}(z)$ by first taking the expectation of all positive W such that $Q_c > q_0$ over the entire domain, and second by multiplying this expectation by $\rho(z)$. Under a probabilistic frame work (i.e. viewing W and Q_c as random variables), the upward mass flux is simply equivalent to the expectation:

$$m(z) = \rho(z)\ \mathrm{E}\left\{W(z)\ \mathcal{I}[0 < W(z)]\ \mathcal{I}[Q_c(z) > q_0]\right\}. \tag{1}$$

Note that capital letters like $Q_c(z)$ now correspond to random quantities and normal fonts to constants like q_0.

3. Mass flux decomposition

(a) *Mesoscale and convective scale*

To obtain a parametric formulation of $m(z)$, a first step is to understand the dependence between the variables W and Q_c. One could directly estimate the bivariate density of the random vector (W, Q_c) but this approach presents difficulties. In particular, the dependence cannot be simply modeled by a covariance function, because these variables are not normally distributed. We use the fact that neither the entire distribution of Q_c nor the entire dependence between W and Q_c is necessary to compute $m(z)$. We only need to understand the relationship between W and the event $\{Q_c(z) > q_0\}$, which is much easier. This fact will be used frequently herein.

Before presenting the advantages of such approach in detail, we divide the indicator function $\mathcal{I}[0 < W(z)]$ into two parts: $\mathcal{I}[0 < W(z)] = \mathcal{I}[0 < W(z) \leq \delta_z] + \mathcal{I}[\delta_z < W(z)]$, for some high positive threshold δ_z which can vary with z. Then, the upward mass flux can be decomposed into two components, $m_1(z, \delta_z)$ and its complement $m_2(z, \delta_z)$, which correspond to fluxes with large and small values of W, respectively:

$$m_1(z, \delta_z) = \rho(z)\ \mathrm{E}\left\{W(z)\ \mathcal{I}[\delta_z < W(z)]\ \mathcal{I}[Q_c(z) > q_0]\right\}$$

and $m_2(z, \delta_z) = m(z) - m_1(z, \delta_z)$. Does this mathematical decomposition have a physical meaning? To answer this question, the spatial locations of large and small W for the squall-line case during the first 5 hours of the 11th day are plotted in Fig. 3 and Fig. 4. One immediately notes that the strong vertical velocity occurs mostly within the convection towers and that weaker values occur in the anvil region. In other words, this decomposition approximately separates the mesoscale part of the cloud system from

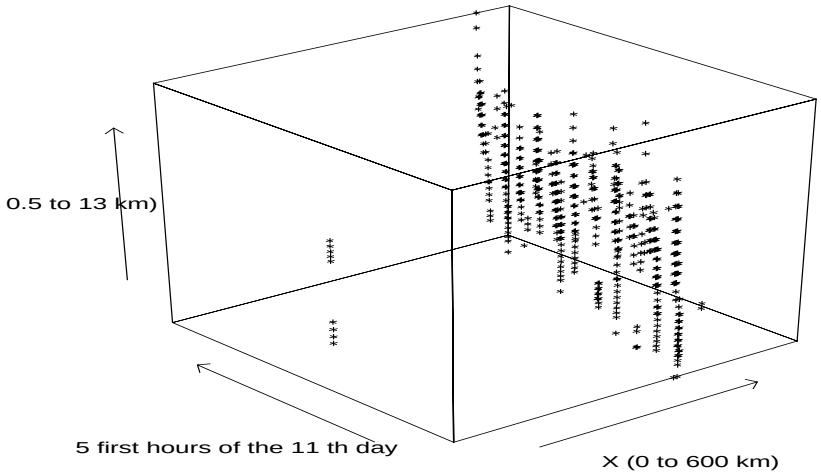

Figure 3. Decomposition of the vertical wind: locations of strong W during the first 5 hours of the 11th day for the squall-line case.

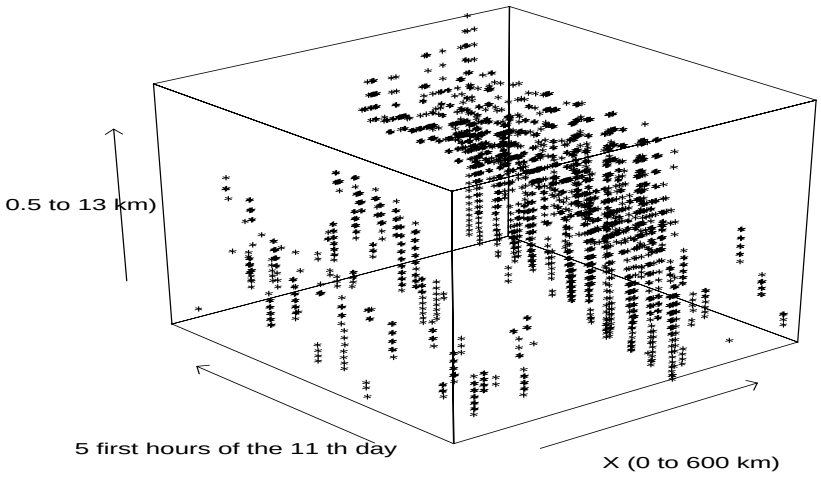

Figure 4. Decomposition of the vertical wind: locations of low W during the first 5 hours of the 11th day for the squall-line case.

the convective part. The concept of decomposing the mass flux into a convective and mesoscale components in physical space was used by Houze (1977). Recently, Yano et al. (2001) proposed a more complex approach in which the decomposition criterion was implemented in wavelet space that represents different processes at different scales.

(*b*) *Statistical dependence between Q_c and W*

Besides distinguishing the mesoscale component from its convective counterpart, our mass flux decomposition has statistical advantages. Since large vertical velocities are usually associated with large Q_c, one anticipates that $Q_c > q_0$ whenever W is sufficiently large. To test this hypothesis, we estimate the probabilities that $Q_c > t_0$ given $W > \delta_z$ at different altitudes. These probabilities are denoted by $p(\delta_z) = \Pr[Q_c(z) > q_0 | W(z) > \delta_z]$.

Figure 5 shows the estimated values of $p(\delta_z)$ at 4 randomly chosen altitudes. As expected,

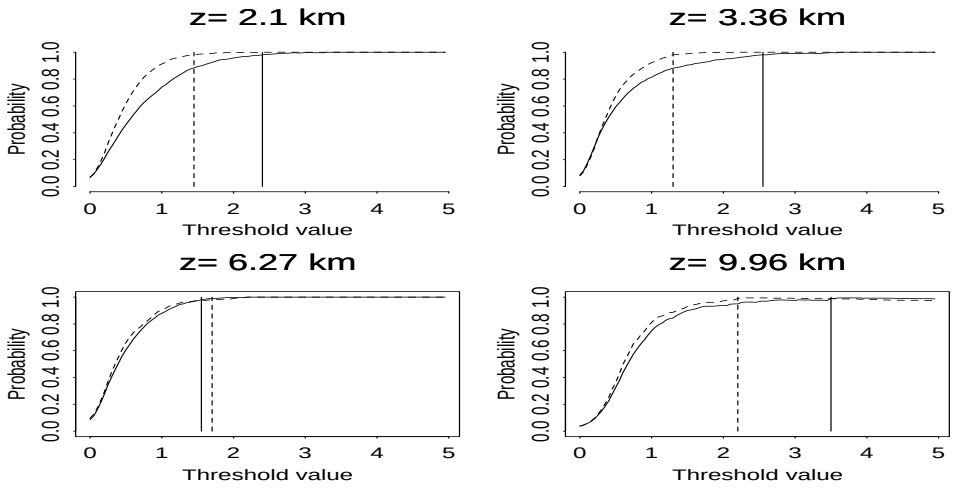

Figure 5. Probability $p(\delta_z)$ for 4 randomly chosen altitudes. The vertical lines represent the values of δ_z for which $0.998 < p(\delta_z)$. The solid line corresponds to the squall-line case and the dotted one to the non-squall cluster.

$p(\delta_z)$ increases to 1 as the threshold increases. The total condensate Q_c is almost always greater than q_0 whenever the vertical wind W is high enough. The left panel of Fig. 6 indicates that δ_z, takes roughly two values, around 1.5 ms^{-1} from 2 km to 7 km and around 3 ms^{-1} above 8 km. This can be better understood by examining the snapshot

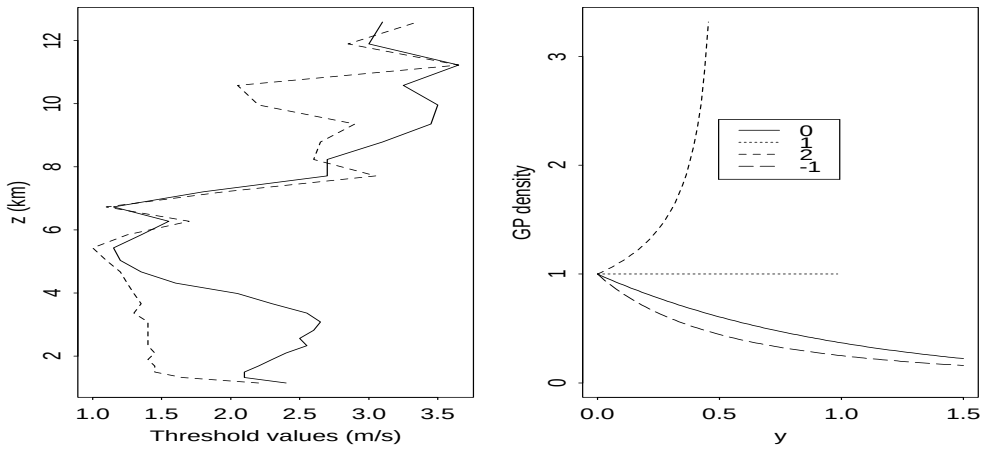

Figure 6. Left panel: Vertical profile of the estimated δ_z for which $0.998 < p(\delta_z)$. Right panel: Density of the Generalised Pareto density.

of total hydrometeor field in Fig. 6 from LM98. Precipitating convective systems occur mostly below 8 km and in less-organised but more extensive stratiform mesoscale clouds above 8 km.

By construction of the threshold (see Fig. 5), we can assume for the remainder of this paper that $W > \delta_z$ implies $Q_c > q_0$. This assumption greatly facilitates the computation

of $m_1(z, \delta_z)$ which simply becomes

$$m_1(z, \delta_z) = \rho(z)\ \mathrm{E}\left\{W(z)\ \mathcal{I}[\delta_z < W(z)]\right\} \tag{2}$$

and does not depend anymore on Q_c This implies that we only need to determine the distribution of large vertical velocity to estimate (2). In the following section we will show how to complete this statistical task by using extreme values theory.

4. Distribution of vertical velocity exceedances

Most statistical methods used in climate studies are concerned primarily with what goes on in the center of a probability distribution (the mean) and on the deviation from it (the variance). One reason for the popularity of such methods is that the assumed distribution is Gaussian, which is uniquely characterised by its mean and its variance. Such assumption is often justified by one of the most famous theorem in probability, the Central Limit Theorem, which states the distribution of the sample mean is approximately Gaussian for large samples. A less known, but nevertheless similar, result also holds for the extremes. Indeed, the distribution of the maximum in large random samples is not unknown but instead follows a specific distribution called the *Generalized Extreme Value* distribution (GEV). In recent years, the statistical methodology in extremes has shifted from working with maxima towards modeling *exceedances*, i.e. amounts above a given threshold. The advantages of using exceedances are its flexibility and the fact that more data are retained. From a probabilistic point of view, the two methods are equivalent, the GEV distribution simply takes an analog but different form, called the *Generalized Pareto Distribution* described in the next section. The extreme values theory has long been applied to a variety of problems in finance, hydrology, forestry; however its use for analyzing climate models has been more limited (Kharin et al. 2000, Zwiers et al. From a probabilistic standpoint, the book by Leadbetter et al. (1983) is recommended. For more recent works, the reader is referred to Embrechts et al. (1996) for theoretical aspects and to Coles (2001), Reiss (1997) and Katz (1999) for a more applied approach.

(*a*) *Exceedances and Generalized Pareto Distribution*

It is known from theory that the conditional distribution of exceedances $Y = W - \delta$ given that $W > \delta$ should approximately follow the so-called *Generalized Pareto Distribution* (GPD) defined by the probability distribution function $F_{k,\alpha}(y) = 1 - (1 - ky/\alpha)^{1/k}$ if $k \neq 0$ and $F_{k,\alpha}(y) = 1 - \exp(-y/\alpha)$ otherwise. The parameters α and k are the scale and shape coefficients respectively and the range of y is $0 \leq y < \infty$ for $k \leq 0$ and $0 \leq y \leq \alpha/k$ for $k > 0$. To illustrate the importance of the coefficient k on the distribution shape, we plot in the right panel of Fig. 6 the generalized Pareto density for 4 different values of k and $\alpha = 1$. From this graph, one can see that the precise estimation of k is essential for assessing the right density to the exceedances (see Appendix for the estimation procedure of the GPD parameters).

(*b*) *Diagnostic tools*

From the extreme value theory, one expects that the distribution of the exceedances Y should follow the GPD if the correct threshold sequence is chosen. In Section 3b, we derived the threshold values δ_z according to the dependence between W and Q_c. To check if the exceedances above such chosen thresholds really follow a GPD, we use two classical

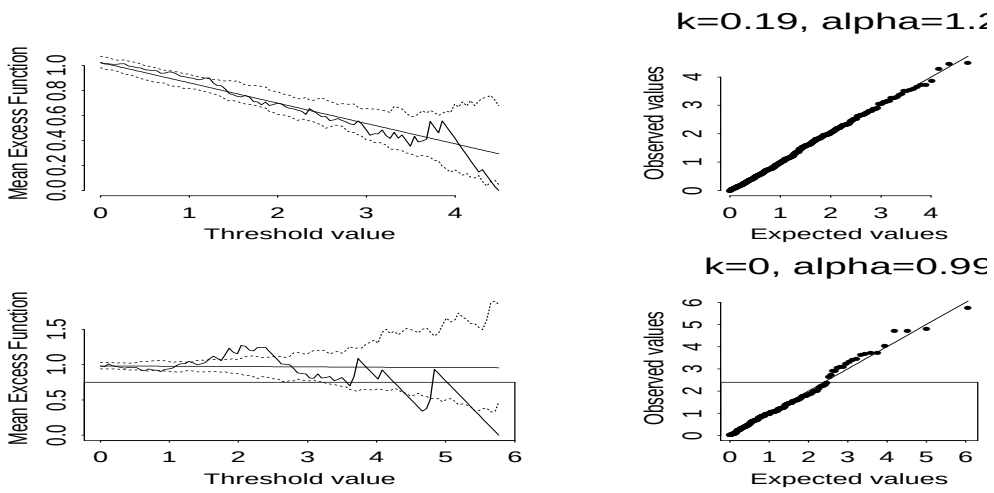

Figure 7. Mean excess function and qq-plot for 2 altitudes (2.10, 8.23 km, i.e. δ_z=(2.4, 2;7)) for the squall-line case during the last 10 days. The dotted line corresponds to the $5\% - 95\%$ confidence intervals and the straight solid line to the expected mean excess function.

diagnostic tools. The first one is directly based on the definition of the GPD and it is called *mean excess function* and defined as the expectation of $Y - \delta'$ given that $Y > \delta'$

$$e(\delta') = \mathrm{E}(Y - \delta' | Y > \delta') = \frac{\alpha - k\delta'}{1+k}, \text{ for any } \delta' > 0 \text{ and } k < 1. \tag{3}$$

To estimate the mean exceedance function, we first calculate $Y - \delta'$ for each observation Y, then take the mean and finally plot the result against δ'. Smith with co-authors wrote a series of papers (1991,1989,1990) on the mean excess function and its applications. The choice of appropriate thresholds is based on Eq. (3), if the excesses over a threshold δ indeed follow a GPD with coefficients (δ, k), then the mean excess over any level $\delta' > \delta$ should vary with δ' according to a straight line of slope $-k/(1+k)$. The second diagnostic tool to test the quality of the GPD fit is the *qq-plot* (quantile-quantile plot) which is a probability drawn after estimating α and k. If the GPD distribution fits the data to sufficient accuracy, the qq-plot should be roughly a straight line of unit slope through the origin. In the next section, we will use these two diagnostic tools.

(c) *Vertical variability and physical interpretation of the coefficients*

To enable our exceedance-based strategy to be successful, we must take the vertical variability into account. We follow the approach of Davison and Smith (1990) where a GPD model is fitted for different subgroups of the data set. Herein, a subgroup is defined by the values of W at a specific altitude z. This involves choosing threshold values δ_z for each altitude z, varying from 0.5 km to 12.5 km (beyond this range, the variable W is negligible). Figure 7 displays the mean excess function and the qq-plot for 2 randomly chosen altitudes $z = 2.10, 8.23$ km (the dotted lines represent the boundaries of a Monte Carlo 90% confidence interval). The two diagnostics tools indicates reasonable fit, specially for the qq-plot. Some abrupt changes occur in the mean excess plot, but only at high thresholds and the curve remains within the confidence intervals except for some short sections.

Figure 8 portrays the values of α and k as a function of z. Striking similarities are noticeable between the vertical behavior of α and the fractional vertical cloudiness of

Fig. 9 from LM98. In particular, three "peaks" are exhibited in the left panel in Fig. 8

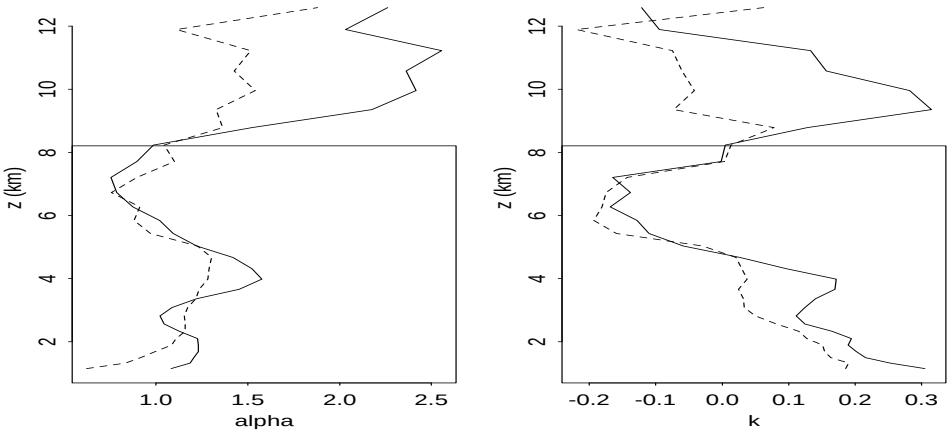

Figure 8. Vertical estimates of α and k over the domain and during the last 10 days. The solid line corresponds to the squall-line case and the dotted one to the non-squall cluster.

for the squall-line case, indicating three main cloud populations. The highest value of α corresponds to deep cumulonimbi and stratiform clouds, the second bump is located around the melting level and the smallest peak in the lowest 2 km corresponds to shallow cumulus. For the non-squall cluster case (dotted line), only two cloud populations are present and the low-level peak is absent. This may be due to the lower hydrometeor content in the non-squall cluster. For the coefficient k, the melting level and deep cumulonimbi and stratiform clouds are still visible and there is also a overall trend, i.e. k roughly decreases with altitude up to the melting level. This indicates that the tails of the distribution of W becomes heavier and heavier from 1 km to 8 km, i.e. the vertical velocity is progressively larger at higher altitudes.

5. Probabilistic formulation

To define a new statistical expression for the mass flux, we use the results in Section 4 to develop a probabilistic formulation of the convective term $m_1(z, \delta_z)$ defined by Eq. 2.

(*a*) *Convective component*

The convective component $m_1(z, \delta_z)$ is closely related to the GPD. Since we have chosen the threshold δ_z such that the exceedances above the threshold δ_z follow a GPD and that $W > \delta_z$ implies $\mathcal{Q}_c > q_0$, we can use this information when computing the expectation $m_1(z, \delta_z)$:

$$\begin{aligned} m_1(z,\delta_z) &= \rho(z)\ \Pr(W>\delta_z)\ \mathrm{E}(W|W>\delta_z), \text{ from Eq. (2)},\\ &= \rho(z)\ \Pr(W>\delta_z)\ [e(\delta_z,z)+\delta_z],\\ &= \rho(z)\ \Pr(W>\delta_z)\ \frac{\alpha(z)+\delta_z}{1+k(z)}, \text{ from Eq. (3)}. \end{aligned}$$

The right-hand side of the last equality is our new formulation of the upward mass fluxes for large W based on statistical principles. To test the validity of such approximation, we plot in Fig. 9 the values of $m_1(z, \delta_z)$ directly computed from the simulations (represented by triangles in Fig. 9) and the new one (dotted lines). According to this figure, the two computations are in excellent agreement. In particular, the differences observed between the squall-lines and non-squall clusters are well represented. This indicates that the proposed probabilistic formulation is good way of describing the upward mass flux for large W.

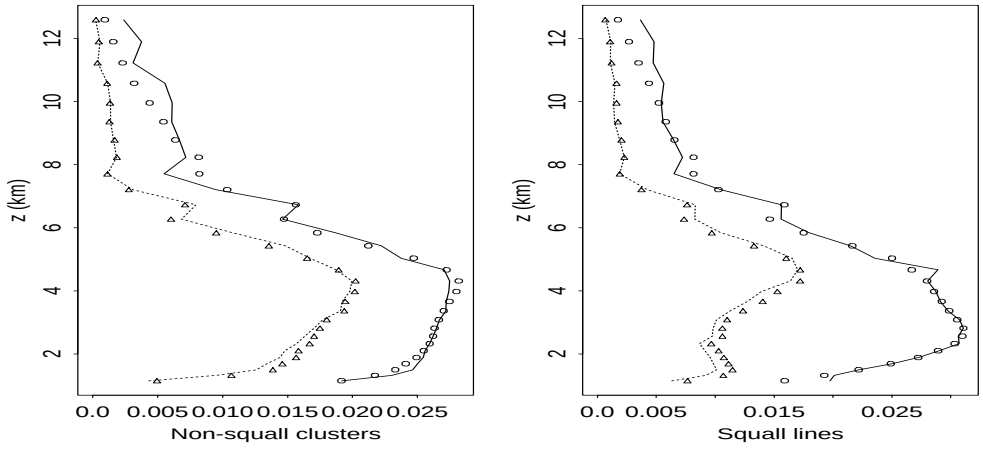

Figure 9. Comparison between the probabilistic upward fluxes ($m_1(z, \delta_z)$ in dotted line and $m(z)$ in solid line) and the ones directly computed from the simulations (triangles for convective part and circles for total flux). The squall lines case, respectively the non-squall cluster case, is represented in the right panel, respectively the left panel.

(*b*) *The probabilistic formulation of* $m(z)$

For the non-squall clusters case, there exists a striking similarity in shape between the convective upward mass flux (triangles in the left panel of Fig. 9) and the over-all mass flux (circles). This indicates that the probabilistic convective flux $m_1(z, \delta_z)$ could be a good predictor for the probabilistic flux $m(z)$. To test this hypothesis, we introduce the following model

$$\log m(z) = \begin{cases} a_0 + a_1 \log m_1(z, \delta_z) + \epsilon(z) & \text{if } z > 5 \text{ km}, \\ b_0 + b_1 \log m_1(z, \delta_z) + \epsilon(z) & \text{if } z \leq 5 \text{ km}, \end{cases} \tag{4}$$

where $\epsilon(z)$ represents some white noise. To estimate the coefficients of this statistical model, we use a least trimmed squares regression (Rousseeuw et Leroy 1987). After computation, the estimated profile of $m(z)$ is represented by the solid line in the left panel of Fig. 9 and it provides a good approximation of the upward mass flux (circles). This result shows that the quantity $m_1(z, \delta_z)$ which was built exclusively from strong velocities is sufficient to determine the whole profile of the upward mass flux.

The squall-lines case presents an additional difficulty, the similarity between $m_1(z, \delta_z)$ and $m(z)$ vanishes under 5 km (see right panel of Fig. 9). To solve this problem, we replace the part corresponding to $z \leq 5$ km in (4) by

$$b_0 + b_1 \log m_1(z, \delta_z) + b_2 \log \Pr[0 < W(z) \leq \delta_z, Q_c(z) > q_0] + \epsilon(z).$$

Adding this new predictor to the regression provides an excellent fit of the upward mass flux by the probabilistic model (see the circles and the solid line in the right panel of Fig. 9).

6. Conclusions

Besides proposing a new way of determining mass fluxes, our probabilistic approach underlines the importance of the extremes, specially in the vertical velocity. This variable is the corner stone of our stochastic formulation of the upward mass flux. The extreme values of W have a more important influence than previously thought. The wind exceedances represent a miniscule part of the data set (less than 0.2%) but their impact on the mass flux is crucial. Hence, strong vertical velocities cannot be appreciated from their numbers but, rather, from their impact on the upward mass flux. In addition, our probabilistic approach allows us to separate the convective part from the mesoscale component. The criterion used for this dichotomic decomposition is the dependence between W and Q_c which seems to play an important physical role (more research should be done in this direction). The coefficients of the GPD in our probabilistic formulation correspond to the vertical variability present the tropical convective system. Finally, while two-dimensional simulations were used as an illustration, the method could be extended to three spatial dimensions.

Appendix

We recall the basic assumptions used to estimate the parameters of our statistical model.

1. The number, $N(z)$, of exceedances at altitude z and above the level δ_z follows a Poisson distribution with mean $\lambda(z)$ and
2. conditional on $N(z) > 0$, the exceedances are identically and independently distributed with a common GPD whose coefficients are $(\alpha(z), k(z))$.

To determine the values of the coefficient vector $\theta(z) = (\alpha(z), k(z), \lambda(z))$, we estimate $\theta(z)$ separately for each altitude with the maximum likelihood method. This approach consists of two steps. First, we write down the likelihood function $l(\theta(z))$ which is the probability of observing our data as a function of the coefficients. Second, we choose the parameters $\hat{\theta}(z)$ which maximizes the likelihood function. Because of assumptions 1 and 2, we write

$$l(\theta(z)) = \frac{(\lambda(z)D)^{N(Z)} \exp(-\lambda(z)D)}{N(z)!} \prod_{i=1}^{N(z)} \left(\frac{1}{\alpha(z)} (1 - \frac{k(z) y_i}{\alpha(z)})^{1/k(z)-1} \right),$$

where D is the dimension of the grid at altitude z.

Note that we estimated the parameters independently for each altitude. A possible extension is to undertake a global maximum likelihood estimation. This would be possible whenever a vertical dependence structure is assumed.

Acknowledgement

The authors would like to thank Doug Nychka for insightful discussions and Changhai Liu for supplying the CRM data sets. The major part of this research was done in Boulder USA and supported by the NCAR Clouds in Climate program and by the NCAR Geophysical Statistics Project.

References

Coles, S. 2001 *An Introduction to Statistical Modeling of Extreme Values*, Springer, New York

Davison, A.C., Smith R.L. 1990 Models for Exceedances over High Thresholds. *J. R. Statist. Soc.*, **52**, page–page

Embrechts P., Kluppelberg C., Mikosch T. 1997 *Modelling Extremal Events for Insurance and Finance*, Springer Verlag, Berlin

Houze, R. A. Jr. 1977 Structure and Dynamics of a Tropical Squall-line system Observed during GATE. *Mon. Wea. Rev.*, **105**, 1540–1567

Katz, R.W. 1999 Extreme value theory for precipitation: Sensitivity analysis for climate change. *Advances in Water Resources*, **23**, 133–139

Kharin, V.V, Zwiers, F.W. 2000 Changes in the extremes in an ensemble of transient climate simulations with a coupled atmosphere-ocean GCM. *Journal of Climate*, **13**, 3760–3788

Leadbetter, M.R., Lindgren, G., Rootén, H. 1983 *Extremes and Related Properties of Random Sequences and Processes*, Springer, New York

Liu, C., Moncrieff, M.W. 1998 A Numerical Study of the Diurnal Cycle of Tropical Oceanic Convection. *Journal of Atmospheric Sciences*, **55**, 2329–2344

Moncrieff, M.W. 1981 A Theory of Organized Steady Convection and its Transport Properties. *Quart. J. Roy. Meteor. Soc.*, **107**, 29–50

Reiss, R.D., Thomas 1997 *Statistical Analysis of Extreme Values*, Birkhauser, Basel

Rousseeuw, P.J., Leroy, A.M. 1987 *Robust Regression and Outlier Detection*, John Wiley, New York

Siebesma, A.P. 1998 *Buoyant Convection in Geophysical Flows*, E.J. Plate, E. E. Fedorovich, D. X. Viegas and J. Wyngaard, Eds, Kluwer Academic Publishers, 441–486

Squires, P., Turner, J.S. 1962 An Entraining Jet Model for Cumulonimbus Updraughts. *Tellus*, **14**, 422–434

Smith, R.L. 1991 Extreme value theory. *Handbook of Applicable Mathematics*, **7**, 437–472

Smith, R.L. 1989 Extreme value analysis of environmental time series: An application to trend detection in ground-level ozone. *Statistical Science*, **4**, 367–393

Yano, J.I., Moncrieff, M.W., Wu, X., Yamada, M. 2001 Wavelet Analysis of Simulated Tropical Convective Cloud Systems Part II: Decomposition of Convective and Meso– Scales. *J. Atmos. Sci.*, **58**, 868–876

Zwiers, F.W.,Kharin, V.V. 1998 Changes in the extremes of the climate simulated by CCC GCM2 under CO2 doubling. *Journal of Climate*, **11**, 2200–2222