Can statistical entropy measures be used to quantify mixing in the Antarctic atmosphere?

Nikolai Krützmann, Adrian McDonald, Steve George
Department of Physics and Astronomy, University of Canterbury, Christchurch, New Zealand

Introduction

"Climate is what we expect, weather is what we get." – Mark Twain

Although there have been significant improvements in weather forecasting since this statement in the 19th century, it is essentially still valid today. The source of the underlying problem now has a name: complexity. The large number of factors that have an influence on the dynamics of the atmosphere, and their interdependence, can lead to almost random changes in these dynamics, making predictions very difficult, or even impossible. One of the main sources of complexity within the atmosphere is mixing. Large fluctuations in chemical concentrations of trace gases (gases that can only be found in relatively low quantities) can be observed at the boundary between distinct regions in the atmosphere, and dynamical variations in the boundary-regions can cause sudden mixing of previously separated air masses.

The Antarctic polar-vortex is a strong barrier, separating the cold polar winter air from warmer mid-latitude air. Its intensification in austral springtime severely restricts meridional (north-south) mixing and ultimately contributes to the *Antarctic Ozone Hole*. This is not the case in the northern hemisphere, as the polar-vortex is weaker and allows more mixing with warmer, ozone-laden air across its borders. Entropy measures are examined to determine whether they can be used to quantify the presence and intensity of complicated mixing processes occurring at the polar-vortex boundaries. The present study considers sample entropy and Rényi entropy, which are both measures of the complexity of the underlying distribution of data.

We aim to use these measures to understand mixing of chemical constituents in the atmosphere and particularly the Antarctic atmosphere. Using artificial data, we determine how to use these measures optimally. Application of entropy measures to tracers (chemically inert species), simulated using the chemistry-climate-model SOCOL (which runs on the University of Canterbury Super Computer) show clear patterns that have been identified by previous studies which have examined mixing. Our method has the significant benefit that it is data driven and requires considerably less computational effort than previous studies. Application of the same entropy measures to observations made by the MLS instrument onboard the Aura satellite show qualitatively similar structures to those determined from SOCOL data.

Sample entropy

Sample entropy (SampEn) is a statistical measure, proposed by Richman and Moorman (2000), which quantifies the variability of time-series by comparing sequences of consecutive data points. It provides a measure of the regularity or predictability of a time-series (high sample entropy is related to low predictability / high complexity). Sample entropy is derived from the conditional probability that a sequence of data points is within a certain tolerance range r for m steps. This tolerance r is usually measured in units of the standard deviation (STD) of the time-series. Hence, sample entropy depends on the length of the data series N, the length m of sequences to be compared and the tolerance range r specified.

Climatological datasets generally display regularly varying patterns on some time-scale (e.g. annual variations are common) and will be affected by noise. Using artificial time-series, the sensitivity (meaning its ability to distinguish between different datasets) of sample entropy on such signals was tested, in order to determine the optimal set of parameters N, m and r.

Figure 1 and *2* show the dependence of sample entropy on data length N for different types of artificial signals (e.g. white noise, red noise, white noise with superimposed sine waves of period 6, 63 and 314 data points, red noise with superimposed sine waves).

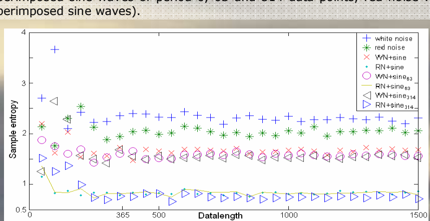

Figure 1: Dependence of sample entropy on data length N (with r = 0.18*STD)

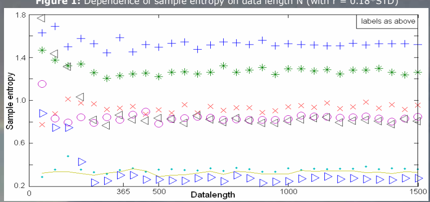

Figure 2: Dependence of sample entropy on data length N (with r = 0.4*STD)

Figure 1 and 2 show that sample entropy can be used to distinguish between noise and noisy data which also contains a regular signal. Irregularities in these patterns can therefore lead to changes in sample entropy. While $N \geq 1000$ data points is found to be optimal for clearly separating the different types of signals, reasonable differentiation can achieved even with 365 data points (e.g. daily measurements for one year).

It is interesting to note that sinusoids of different frequencies converge to almost equal values of sample entropy. It follows, that sample entropy can not be used to distinguish between signals of similar form but different frequency. This is not unexpected, a signal which contains noise and has a certain period is no more complex than the same quantity of data but with a different periodicity – assuming we adequately sample both variations.

Figure 2 is the same diagram as *Figure 1*, but calculated using a different value of r, showing that the absolute value of sample entropy and the relative separation of the signals, and therefore its sensitivity to changes in the 'dynamics' of the time series, strongly depend on r. Testing a wide range of values, we conclude that $r = 0.4$ is optimal for our purposes since this value produces the highest relative separation between the signals, compared to the variability of the value of sample entropy.

Although sample entropy is defined for any m, we use the commonly chosen value $m = 2$ (e.g. Lake et al., 2002).

Sample entropy - application

While sample entropy has been used by Shuangcheng et al. (2006) to analyze daily temperature measurements in Southwest China and identify regions associated with climates of different complexity, we aim to apply this metric to data on a global scale. Using the parameters determined previously, we apply the sample entropy calculation to methane (CH_4) data from a model run of the chemistry-climate-model (CCM) SOCOL (Egorova et al., 2005), in order to examine whether large-scale mixing-patterns can be identified.

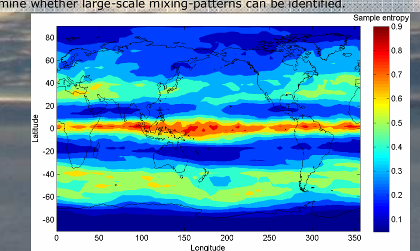

Figure 3: Global sample entropy ($r = 0.4$, $m = 2$) of 1980 methane data at approximately 28 km altitude.

Figure 3 shows sample entropy of the global ozone field from 1980 SOCOL data at a pressure of 15.85 mBar (≈ 28 km altitude). The tropical band of higher values is in accordance with previous studies which have highlighted the existence of a region separated from the rest of the atmosphere, sometimes referred to as the "tropical pipe". The ongoing chemical processes and the upwelling branch of the Brewer-Dobson circulation, are likely to cause fluctuations in tracer concentrations in this area, and therefore more complex dynamics. This is successfully identified by sample entropy.

On the other hand, the values of sample entropy, e.g. over the Antarctic, are surprisingly low, as more mixing would be expected in this region. While this could be due to model deficiencies, the fact that we have to use a full year of data could mean that strong mixing processes occurring over shorter periods of time are 'averaged out'.

Rényi entropy

Therefore, we turn our attention to Rényi entropy (RE), which is more suitable for use on non-temporal datasets, as it does not take into account the order of the data points. Instead, it is based on the occurrence-rate (or probability) p of certain values in a dataset and describes the complexity of the probability distribution function (PDF) of data. It is defined as:

$$RE(\alpha, b, N) = \left(\frac{1}{1-\alpha}\right) \cdot \ln\left(\sum_{i=1}^{b} p_i^{\alpha}(b, N)\right)$$

While Rényi entropy is defined for any value of α, the name often refers to the case of $\alpha = 2$, which is also used in this study. The probabilities $p_i(b,N)$ are calculated from a histogram of the dataset (N data points) with b bins of equal width. We choose to normalize by $\ln(N)$, the maximum value of Rényi entropy independent of α.

Depending on the distribution of the data points, different values of b are necessary to represent the underlying structure (see also *Figure 4*). In order to determine the value of b that best represents the shape of the underlying distribution, we use the optBINS algorithm by Knuth et al. (2005). It requires a minimum of 100–150 data points N, making Rényi entropy applicable to even smaller datasets than sample entropy.

Similar to sample entropy, we tested Rényi entropy with various artificial datasets, but here we show an example with SOCOL ozone data, as it demonstrates clearly how Rényi entropy can be used to characterize tracer dynamics. *Figure 4* shows the distribution, the PDF and the Rényi entropy of ozone in southern hemisphere winter/early spring (September) and summer (February).

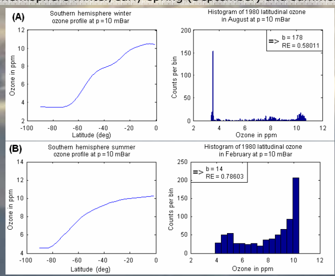

Figure 4: Ozone profile of the southern hemisphere (left), corresponding PDF (right) and resulting Rényi entropy for a southern hemisphere winter **(A)** and summer month **(B)**.

Three distinct regions of ozone concentration can be observed in winter (top panels in *Figure 4*). The polar area with low ozone values, the "surf zone" with higher concentrations and the tropical region with even higher levels of ozone. Mixing between these air masses is relatively low, and the value of Rényi entropy is *0.580*. In contrast, the summer PDF shows less distinct regions, as high- and mid-latitude air is relatively well mixed, due to the absence of the polar vortex as a barrier. This change in the dynamics is reflected by a shift of Rényi entropy towards higher values (RE = *0.786*).

Rényi entropy - application

We use SOCOL methane data for calculating the temporal evolution of Rényi entropy in the southern hemisphere. We then compare our calculations with previous studies which have attempted to quantify mixing in the Antarctic stratosphere, in order to analyze the utility of Rényi entropy for this task. Rényi entropy was calculated for a moving average of 10 days of data around a circle of latitude. *Figure 5* shows the results at a pressure of 10 mBar.

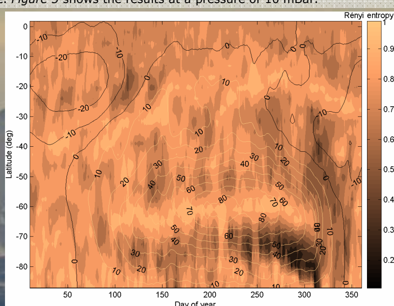

Figure 5: Latitude over time plot of Rényi entropy of 1980 methane data (SOCOL) in the southern hemisphere with zonal (east-west) wind contours at 10 mBar (≈ 31km altitude).

During Antarctic autumn, winter and early spring, a band of high values of Rényi entropy separates intermediate values at mid-latitudes from low values at high latitudes *(Figure 5)*. This shows the strong influence of the polar-vortex (which can also be seen in the zonal wind contours), restricting mixing between high- and mid-latitudes. The bright region (high Rényi entropy) is at the centre of the strongest zonal winds. Therefore, we identify it with the "vortex-edge" region, which was found to be particularly unmixed by Lee et al. (2001). This structure can be identified even more clearly in *Figure 6 (A)*, which shows an altitude-latitude-profile of the monthly zonal Rényi entropy in the southern hemisphere, averaged over September, October and November. It also shows the same diagram for the months March, April and May *(B)*. During this time there is no polar-vortex acting as a mixing barrier, the patterns are, on the average, more complex and no structures similar to *(A)* can be found. Furthermore, a band of high values at lower latitudes seems to separate tropical air from mid-latitudes. This is in accordance with the "tropical pipe" region of isolated air at the equator, mentioned previously.

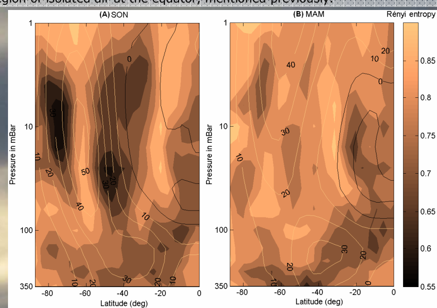

Figure 6: Altitude (350 mBar ≈ 7.5 km; 10 mBar ≈ 31 km) over latitude plot of quarterly averaged monthly Rényi entropy with zonal wind contours (SOCOL, 1980). **(A)** September, October, November **(B)** March, April, May

We compare our findings to those of Haynes and Shuckburgh (2000), who use effective diffusivity for quantifying mixing (*Figure 7*). The red L-shaped region between July and December is very similar to the structure of medium to low Rényi entropy we find in *Figure 5* during this period (note the different timescales). The separation of air masses between high- and mid-latitudes can be clearly seen.

These patterns also agree well with the findings of Allen and Nakamura (2001) and Garny et al. (2007). All these groups use very different and much more computationally demanding approaches than Rényi entropy for their calculations.

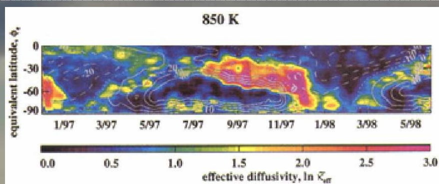

Figure 7: Plot of effective diffusivity showing similar structures as *Figure 5*. Note the different timescales. 850K potential temperature is approximately equal to 29.5 km altitude. From Haynes and Shuckburgh, 2000.

Finally, we apply Rényi entropy to a set of actual measurements of nitrous-oxide (N_2O). It is an atmospheric tracer with properties similar to methane and should therefore display the same mixing patterns. The measurements, made in 2005 by the MLS instrument onboard the Aura satellite, lead to qualitatively similar structures to those determined from SOCOL data.

Figure 8 also shows the band of high Rényi entropy separating intermediated values at mid-latitudes from low values in Antarctic regions seen in *Figure 5*. The fact that even slightly similar patters can be identified in calculations for simulated data of 1980 and observational data from 2005, shows that potentially valid deductions can be made from SOCOL simulations and that Rényi entropy is a promising measure of complexity in the Antarctic atmosphere.

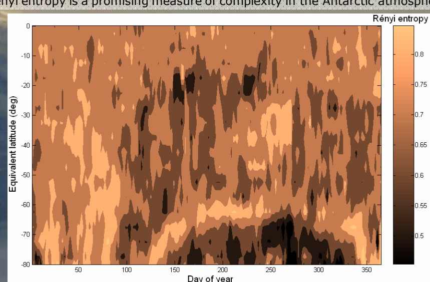

Figure 8: Latitude over time plot of Rényi entropy of 2005 MLS measurements of N_2O in the southern hemisphere at 850 K potential temperature (≈ 29.5 km altitude).

Conclusion

We have studied the statistical entropy measures Sample entropy and Rényi entropy and determined the most suitable parameters for our aim of using them as measures of complexity and mixing in the Antarctic atmosphere. While Sample entropy is most useful for quantifying the regularity and noisiness of recurring patterns in time-series, Rényi entropy is well suited for analyzing changes in the PDFs of all kinds of data, e.g. zonally and monthly averaged tracer fields.

We applied both measures to tracers from runs of the chemistry-climate-model SOCOL. We were able to identify patterns similar to those reported by several other groups, who use very different measures for quantifying mixing. Rényi entropy clearly shows the separation of high- and mid-latitude air masses in the Antarctic atmosphere associated with the strong polar-vortex in winter and early spring. Our method has the advantage that it is data driven and requires significantly less computational effort than the ones used in other studies. The validity of our approach was underlined by showing that qualitatively similar patterns can be seen when using observational satellite data of a different tracer.

We aim to improve our understanding of mixing in the Antarctic atmosphere by applying Rényi entropy to further observational data of several tracers and comparing them with more SOCOL model simulations.

References

Allen, D.R., Nakamura, N., (2001), *A seasonal climatology of effective diffusivity in the stratosphere*, J. Geophys. Res., **106**:7917-7935.

Egorova, T., Rozanov, E., et al., (2005), *Chemistry-climate model SOCOL: a validation of the present-day climatology*, Atmospheric Chemistry and Physics, **5**:1557-1576.

Garny, H., Bodeker, G.E., Dameris, M., (2007), *Trends and variability in stratospheric mixing: 1979–2005*, Atmos. Chem. Phys. Discuss., **7**:6189–6228.

Haynes, P., Shuckburgh, E., (2000), *Effective diffusivity as a diagnostic of atmospheric transport 1. Stratosphere*, J. Geophys. Res., **105**:22,777-22,794.

Lake, D.E., Richman, J.S., Griffin, M.P., Moorman, J.R., (2002), *Sample entropy analysis of neonatal heart rate variability*, Am. J. Physiol. Regulatory Integrative Comp. Physiol., **283**:789-797.

Knuth, K.H., Gotera, A., Curry, C.T., Huyser, K.A., Wheeler, K.R., Rossow, W.B., (2005), *Revealing Relationships among Relevant Climate Variables with Information Theory*, Presentation given at the *Earth-Sun Systems Technology Conference 2005*.

Lee, A.M., Roscoe, H.K., Jones, A.E, Haynes, P.H., Shuckburgh, E.F., Morrey, M.W., Pumphrey, H.C., (2001), *The impact of the mixing properties within the Antarctic stratospheric vortex on ozone loss in spring*, J. Geophys. Res., **106**:3203-3211.

Richman, J.S., Moorman, J.R., (2000), *Physiological time-series analysis using approximate entropy and sample entropy*, Am. J. Physiol. Heart Circ. Physiol., **278**:H2039-H2049.

Shuangcheng, L., Qiaofu, Z., Shaohong, W., Erfu, D., (2006), *Measurement of climate complexity using sample entropy*, Int. J. Climatol., **26**:2131-2139.