Can recurrence quantification analysis be useful in the interpretation of airborne turbulence measurements?

Stanisław Król¹, Alan M. Blyth², Steven Böing³, Leif Denby⁴, Tom A. Lachlan-Cope⁵, and Szymon P Malinowski⁶

¹Institute of Geophysics, Faculty of Physics, University of Warsaw
²NCAS, University of Leeds
³Univ of Leeds
⁴Danish Meteorological Institute
⁵British Antarctic Survey
⁶University of Warsaw

August 12, 2023

Abstract

In airborne or model data, clouds are often defined using information about Liquid Water Content (LWC). Unfortunately LWC is not enough to retrieve information about the dynamical boundary of the cloud, i.e. volume of turbulent air around the cloud. In this work, we propose an algorithmic approach to this problem based on a method used in time series analysis of variables in dynamic systems, namely Recurrence Plot (RP) and Recurrence Quantification Analysis (RQA). By studying RP’s constructed with turbulence kinetic energy, vertical velocity and temperature fluctuations and quantifying appropriate time series of laminarity (LAM) calculated using the plots, we distinguish between turbulent and non-turbulent segments along a horizontal flight leg.
Can recurrence quantification analysis be useful in the interpretation of airborne turbulence measurements?

Stanislaw Król\(^1\), Alan Blyth\(^2\), Steven Böing\(^2\), Leif Denby\(^2,4\), Tom Lachlan-Cope\(^3\), Szymon P. Malinowski\(^1\)

\(^1\)Institute of Geophysics, Faculty of Physics, University of Warsaw
\(^2\)National Centre for Atmospheric Science, University of Leeds
\(^3\)British Antarctic Survey
\(^4\)Danish Meteorological Institute

Key Points:

- A vector characterising turbulence along the aircraft track based on velocity and temperature fluctuations is constructed.
- A time series of this vector allows the construction of a Recurrence Plot, a tool used in studies of dynamical systems.
- A quantity named laminarity, derived from the Recurrence Plot, is used to detect turbulent and non-turbulent regions.

Corresponding author: Stanislaw Król, skrol@fuw.edu.pl
Abstract
In airborne or model data, clouds are often defined using information about Liquid Water Content (LWC). Unfortunately LWC is not enough to retrieve information about the dynamical boundary of the cloud, i.e. volume of turbulent air around the cloud. In this work, we propose an algorithmic approach to this problem based on a method used in time series analysis of variables in dynamic systems, namely Recurrence Plot (RP) and Recurrence Quantification Analysis (RQA). By studying RP’s constructed with turbulence kinetic energy, vertical velocity and temperature fluctuations and quantifying appropriate time series of laminarity (LAM) calculated using the plots, we distinguish between turbulent and non-turbulent segments along a horizontal flight leg.

Plain Language Summary
Cloud is defined by the presence of liquid (or solid) water in the atmosphere. By agreeing on a certain threshold of liquid water content, one can define a cloud boundary. Cloud processes result in disturbances in the airflow around the cloud (turbulence). Its presence should allow the cloud dynamical boundary to be defined in a similar manner. However, a single simple variable, which could be used in an analogous way is not well defined. In this study we propose a method to define such a variable useful in distinguishing between turbulent and non-turbulent volumes around the cloud (i.e. determining dynamic cloud interface), adopting a method used in the study of dynamical systems. On the base of recorded temperature and air velocity fluctuations characteristic for turbulence, turbulent transport and mixing, a threshold of a variable named “laminarity” is used to define the dynamic cloud boundary on the aircraft trajectory.

1 Introduction
Understanding atmospheric turbulence is crucial to properly describe transport processes in the Atmospheric Boundary Layer (ABL) and in the free atmosphere, mixing processes in clouds, and even rain production (Bodenschatz et al., 2010). Airborne measurements provide direct information on turbulent velocities and associated fluctuations of thermodynamic fields (Wendisch & Brenguier, 2013). Statistical interpretation of these measurements is, however, problematic. Time series collected in the course of airborne measurements represent conditions along a one-dimensional, complicated trajectory inside evolving, coupled fields of momentum, temperature, humidity and cloud particles. In contrast to laboratory experiments, data collected in the course of a flight do not contain information on the boundary and initial conditions of the flow, on external forcings and internal feedbacks. There is even no general rule to distinguish to what extent the observed fluctuations can be contributed to turbulence, waves, or inhomogeneities of non-turbulent nature. Thus an objective division of the collected data series into segments representing similar stages/properties of the flow is crucial to construct conditional statistics of measured variables to characterise phenomena along the flight trajectory.

Typically, additional information and assumptions are necessary to construct reasonable conditional statistics. To explain this in more detail, consider e.g. cloud/clear air turbulent mixing. While the cloud microphysical boundary can be defined in a data series by a selected threshold of liquid water content (LWC) or droplet number concentration (DNC) (Malinowski & Zawadzki, 1993), cloud dynamical boundary, i.e. the region around a cloud where the flow is affected by dynamical (Ackerman, 1958) and thermodynamical (Radke & Hobbs, 1991) processes related to the cloud, is not unique. In airborne measurements porpoising through stratocumulus cloud top (Malinowski et al., 2013) three parameters (turbulence kinetic energy, potential temperature gradient and wind shear) were used to distinguished between the four different layers in the cloud top vicinity. Neither the parameters, nor threshold values could be considered objective, yet the division was algorithmic and allowed to describe many properties of turbulence and
turbulent mixing in stratiform clouds. Such divisions are even more difficult in case of
convective clouds penetrated horizontally. Heus and Jonker (2008) have shown the ex-
istence of shells around clouds where dynamical interactions occur. While cloud mod-
ellers who obtain two- or three- dimensional thermodynamic fields may use such param-
eters as enstrophy (Mellado, 2017) or more advanced numerical approaches (Nair et al.,
2021) to obtain a border between turbulent cloud and non-turbulent environment, a sim-
ilar approach cannot be adopted to airborne data along a 1-D trajectory.

This work describes an effort to construct a robust, objective method to classify
time series collected in the course of flight by “turbulent” (dynamic) properties using a
small number of a-priori assumptions. The idea is to apply a technique which is widely
adopted in nonlinear time series analysis called Recurrence Quantification Analysis (RQA)
(Marwan et al., 2007), which can be used to study transitions in the system. It revolves
around the concept of recurrence of the dynamical system into the same regions of the
phase space, which is one of the fundamental aspects in the analysis of such systems. It
has found applications in many fields such as physiology, neuroscience, economy, and earth
sciences.

In the following analysis, dynamic (turbulent) properties of clouds and clear air in
between are characterised by three variables, representing important dynamical processes
in clouds: turbulence, characterised by its kinetic energy (TKE), mixing, characterised
by temperature fluctuations ($T'$) and vertical transport (updrafts and downdrafts) char-
terised by vertical velocity ($w'$). These variables, normalised according to the RQA
approach, form a 3-D phase subspace, and recurrences of a vector representing proper-
ties of measured turbulence into the same regions of this phase subspace indicate seg-
ments of the time series representing similar dynamic properties of the flow.

To test applicability of this approach we use airborne data collected in the course
of the EUREC4A (Elucidating the role of clouds-circulation coupling in climate) exper-
iment (Stevens et al., 2021) with the British Antarctic Survey (BAS) Twin Otter (TO)
research aircraft, equipped to measure turbulence, temperature fluctuations, and cloud
properties. Cloud mask defined by a threshold of LWC, is added to the comparison, al-
lowing turbulence inside and outside of cloud to be distinguished.

The structure of the article is the following. In section 2 we describe the data, sec-
tion 3 discusses the details of the time-dependent RQA method adopted, and in section
4 the results are presented and discussed. Final conclusions can be found in section 5
of the paper.

2 Data description

EUREC4A (Stevens et al., 2021), was a measurement campaign conducted in late
January and early February 2020, with a focus on the study of coupling between atmo-
spheric boundary layer and ocean surface, and the study of marine cumulus clouds and
their spatial organisation near Barbados. One of the research aircraft present was the
British Antarctic Survey’s Twin Otter (TO). The aircraft flew at different altitudes per-
forming cloud penetrations, collecting data between the clouds as well as in the atmo-
spheric boundary layer (ABL), even near the ocean surface.

The TO was equipped with Meteorological Airborne Science Instrumentation (MASIN)
gathering data at two frequencies: 1 Hz and 50 Hz. The data set contains GPS plane
position, pressure altitude, three components of wind velocity from the Best Air Tur-
bulence (BAT) probe (Crawford & Dobosy, 1992; Garman et al., 2006), temperature and
H2O molar ratio. There were cloud spectrometers, measuring cloud microphysical prop-
erties, as well as an additional instrument to measure temperature fluctuations, the Ultra-
Fast Thermometer 2b (UFT-2b) from the UFT family (Kumala et al., 2013) with an ef-
fective resolution reaching 2000 Hz.
Not all collected time series were suitable for analysis due to instrumental problems. The UFT-2b provided useful data from 7 flights, in 5 of these cloud penetrations were recorded. In these flights three 50 Hz velocity components (turbulence data) from the BAT (Best Air Turbulence) probe were available, yet in some segments of flights with deteriorated resolution. Records from BAT and UFT-2b were used in this study. These instruments were co-located within 10 cm distance on the nose boom of the aircraft, thus after averaging temperature fluctuations to match the 50 Hz signal from the BAT probe (corresponding to ~1.3 m spatial resolution at 69 m/s TAS) data can be considered as collected in the same volume.

Additionally, LWC from cloud spectrometers, located under the wings (~5 meters apart) was used to define cloud mask. In the first two flights it was the data from the Fast-Forward Scattering Spectrometer Probe (FFSSP) (Baumgardner et al., 1993; Brenguier et al., 1998) with 50 Hz resolution, and for the remaining three, from Cloud Droplet Probe (CDP) (Lance et al., 2010) with 1 Hz resolution (~65 m spatial resolution). A total of 246 clouds defined as segments with LWC $>10^{-3}$ g/kg at 1 Hz were found. Only a fraction of these data were selected for the preliminary analysis described here: a fragment of a time series collected at constant altitude during flight TO-336, starting at around 12:42 UTC. Since the typical true air speed (TAS) of the TO was about 69 m/s, and velocity fluctuations were of order of a few m/s, the frozen flow hypothesis is applicable to data interpretation.

3 Data analysis

As mentioned in the introduction, we decided to perform time-dependent RQA using TKE (denoted in equations as $k$), $T$ and $w'$:

$$\vec{x} = [T, k, w'],$$

(1)

as indicators of airflow dynamics. $w'$ is calculated by subtracting the mean calculated over the whole horizontal flight segment.

RQA began from normalisation of the components of the recurrence vector in order to equally capture the influence of each component. The normalisation was performed according to the formula:

$$X_n = \frac{x_n - \mu_{x_n}}{\sigma_{x_n}},$$

(2)

where $\mu_{x_n}$ and $\sigma_{x_n}$ is the mean and standard deviation of $n$-th vector component, and the average is taken along the whole horizontal flight segment of the flight.

For each data point on the leg a recurrence matrix Marwan et al. (2007), characterising the dynamical state of the volume of air along a certain segment of the flight trajectory, centred on this point was calculated according to the formula:

$$R_{ij} = \Theta(\epsilon - |\vec{X}_i - \vec{X}_j|),$$

(3)

where $\Theta$ is the Heaviside step function, $\epsilon$ is the so-called recurrence threshold and $\vec{X}_i$ is the vector $\vec{X}$ at time $i$, $i \in (1, ..., N)$, where $N$ is the length of $\vec{X}$. As pointed out in Marwan et al. (2007), there is no single universal method for choosing the best $\epsilon$, but there are several “rules of thumb”. According to these rules, a recurrence threshold $\epsilon = 1$ was chosen, which corresponds to about 10% of the diameter of the phase space. The example recurrence matrix and corresponding elements of equation (1) are plotted in figure 1.
Figure 1. Top panel: recurrence plot for 256 experimental points long segment centred on a selected point (a cloud edge) of the time series of the measurement data, bottom panels: plots of individual components of the vector.

According to Marwan et al. (2007) visual inspection of patterns appearing in the recurrence matrix makes it possible to distinguish different types of trajectories of the
investigated vector in phase space. In particular, white areas or bands correspond to rapid
changes, chess-board like structures to periodic events, and diagonal lines suggest par-
allel segments of trajectories in phase space. In figure 1 a change of the recurrence be-
haviour of the time series after 3 seconds from a slowly changing state to rapid fluctu-
ations is visible. Additionally, two vertical lines around 4 s and 4.5 s correspond to the
values of $[T, k, w']$ similar to that in the first 3 s of the plot, visible as a black area on
the matrix.

Some of the above-mentioned properties of each matrix can be characterised by one
or more measures proposed by Marwan et al. (2007) in the overview of recurrence quan-
tification analysis (RQA). Time series of such measures can be interpreted in terms of
varying properties of turbulence along the trajectory of the aircraft.

Since there is no general rule allowing for selection of the recurrence matrix size,
we assumed that the matrix built on a segment corresponding to a few integral scales
of turbulence should be the right choice. We performed a series of tests, starting from
the size corresponding to 128 data points (which corresponds to $\sim 176$ m at 69 m/s TAS)
to 1024 data points, and we decided on a 256 point matrix, corresponding to 352 m, i.e.
1-3 typical integral length scales (Craig & Dörnbrack, 2008; Wadawczyk et al., 2022).
In parallel we had to decide which RQA parameter serves our purpose best. We have
tested such parameters as recurrence rate, determinism and laminarity (Marwan et al.,
2007). Since our time series does not characterise the particular dynamical system, but
the actual state of turbulence along the flight track, and the goal is to distinguish be-
tween non-turbulent and turbulent regions, we decided to focus on a metric called lam-
inarity $LAM$, which, according to Marwan et al. (2007) “represents the probability of
occurrence of laminar (slowly changing) states of the system”:

$$LAM = \frac{\sum_{v=v_{\text{min}}}^{V_P(v)} P(v)}{\sum_{v=1}^{V_P(v)} P(v)},$$

where $v$ is the length of a black line in the recurrence plot, $P(v)$ is the number of
lines of length $v$ and $v_{\text{min}}$ is the minimal line length. In other words, $LAM$ is defined
as a number of black points forming vertical lines of length equal to at least $v_{\text{min}}$, nor-
malised by the number of all black points in a Recurrence Plot. For laminar states $LAM$
is close to 1 (almost all points form vertical lines), whereas for states with high fluctu-
ations it is closer to 0 (few vertical lines, mostly solitary black points). The choice of pa-
rameter $v_{\text{min}}$ is important, since it excludes shorter lines from calculations. In this study
$v_{\text{min}} = 10$, corresponding to a $\sim 14$ m distance, was chosen. The argument is that we
look for fluctuations of sizes corresponding to large scales of the turbulence cascade, small
scale variability should be studied after selecting segments for more detailed analysis.

Finally, in order to understand the relative importance of turbulence kinetic en-
ergy, mixing and updrafts/downdrafts analogous analysis, construction of recurrence ma-
trices and following RQA, was performed for each vector component separately. The ex-
ample application of the procedure described in the previous section is presented in fig-
ure 2.

The top panel of figure 2 shows three time series of laminarity calculated for var-
ious sizes (512, 256 and 128 points corresponding to $\sim 704$, $\sim 352$ and $\sim 176$ m distance
in physical space) of the RP. It can be seen that the black curve representing an RP of
size $512 \times 512$ hardly reflects dynamic disturbances related to small clouds, such as the
one at $t \approx 7350$ s. The blue curve representing $LAM$ calculated for a 128 point long
segment is most sensitive to changes in the environment, but shows strong oscillations,
which are hard to interpret in terms of turbulence properties. In general, increasing the
size of the matrix smooths out the time-dependent RQA, and broadens the peaks of the
quantities. This is an effect of capturing a bigger picture around a selected point and
related smoothing, as seen for example at $t \approx 7250\text{s}$, where the black curve does not give insight into the variability of dynamics captured by the red and blue curves corresponding to smaller matrices. Additionally, 256 points / $\sim 352\text{ m}$ corresponds roughly to 3–4 integral scales of turbulence, and distinguishing between turbulent/non turbulent regions on a distance corresponding to a few integral scales seems reasonable.

The bottom panel of figure 2 shows a plot of $LAM$ time series for the whole vector (light-blue curve) and for its components ($T$, $k$, $w'$). Grey shading indicates the separate use of a cloud mask indicating the presence of cloud water (at 1 Hz, i.e. 69 m spatial resolution). It can be seen that considering the whole vector gives the complete picture of turbulent regions within and around the clouds, while laminarity calculated for each component separately shows the relative importance of $k$, $T$ and $w'$, i.e. mechanical turbulence, mixing and vertical transport which change along the time series.

4 Results

Figure 3 shows the time series for all three components of the vector, LWC series as the grey cloud mask and $LAM$ calculated for the whole vector, with yellow shaded regions corresponding to $LAM$ below the selected threshold, i.e. to the regions where a signature of turbulence is observed. Since the goal is to distinguish between the turbulent and laminar regions, a threshold just below 1 should be the cutoff. After visual inspection the specific value of $LAM = 0.95$ was chosen. Selection of $LAM = 0.99$ resulted in selection of the majority of the record as turbulent, while $LAM = 0.90$ resulted in omission of segments where some clear signatures of turbulence were present. The exact value of the threshold can be discussed, however the goal of the selection is not the detailed turbulence analysis, but selection of the segments on which such analysis should be performed so the detailed number is of secondary importance.

Figure 3 shows that the investigated region is composed of clouds with vast volumes around the clouds affected by turbulence and laminar regions with occasional sig-
Figure 3. Time series of (from top to bottom): temperature, TKE, vertical velocity fluctuations (red - positive values, blue - negative values), LWC, and LAM along the analysed flight segment. Grey rectangles mark the cloud segments derived from LWC signal. Yellow rectangles mark the segments with LAM below 0.95.

natures of turbulence, not related to visible amounts of LWC. All clouds in the analysed data are turbulent, with clearly visible local minima of LAM values below ~0.75. The local minima of laminarity reflect coincident input of all the components of the vector, but there is no clear dependence to the actual value of LWC.

Additionally, for each cloud, regions with LAM <0.95 around the cloud can be considered “turbulent cloud shells”, and borders of these shells can be determined by the LAM threshold: such a division will help in more detailed investigation of cloud dynamics.

To close this analysis we performed initial statistical analysis of turbulence in the time series. Table 1 summarises the average turbulence kinetic energy, vertical velocity variance and temperature variance in 3 different types of segments along the trajectory: clouds (LWC above threshold), turbulent regions with no clouds (LAM above threshold) and non-turbulent regions in between. Not surprisingly, these values are very different, reflecting different dynamics along the flight trajectory.

Table 1. Values of the mean TKE, and standard deviations of the temperature and the vertical wind velocity variable component for three different masks: LAM$^>$ - LAM above 0.95 threshold, LWC - LWC above threshold, LAM$^<$ - LAM below 0.95 threshold, but with LWC below threshold.

<table>
<thead>
<tr>
<th>LAM$^&gt;$</th>
<th>LWC</th>
<th>LAM$^&lt;$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu_k$</td>
<td>0.79</td>
<td>2.46</td>
</tr>
<tr>
<td>$\sigma_T$</td>
<td>0.32</td>
<td>0.32</td>
</tr>
<tr>
<td>$\sigma_{w'}$</td>
<td>0.58</td>
<td>1.31</td>
</tr>
</tbody>
</table>

-8-
As we can see the values of the mean TKE are the highest for cloud regions, but they are also significantly higher for regions with LAM below threshold than for LAM above threshold. Standard deviation of vertical velocity follow the same trend as the mean TKE, while standard deviation of temperature is not significantly different between the regions. However, more detailed inspection of the data indicates, that the standard deviation of the temperature in laminar region comes from large scale smooth temperature variations, while in turbulent regions it comes from small scale variability. Consequently, the mask based on the changes of LAM allows for better division of data in terms of turbulence characteristics.

5 Summary and outlook

In this study we proposed a method to algorithmically divide time series of airborne measurements according to dynamic (turbulent) properties of the sampled air, combining the simple signature properties characterising turbulence, turbulent transport and turbulent mixing. The method is based on recurrence quantification analysis and the assumptions adopted are based on well-documented physical characteristics of atmospheric turbulence. We believe that this approach will facilitate future investigations of atmospheric turbulence in, within and around clouds, which requires a more systematic approach. The main advantage of the RQA method is that the constructed vector can be expanded with more measured quantities, such as LWC. However, the method is useful even when the microphysical signal is of poor quality, or in situations where it is independently used to determine the cloud-clear air interface.

We used only one metric from the RQA, laminarity, but RQA offers a variety of metrics, therefore it is possible to expand the studies and/or use metrics that are more suitable for other types of analyses. The possibilities in the future with regards to this type of study include imposing an additional condition based on the LWC signal (of a better resolution) added to the vector, as well as applying the method to data gathered during other campaigns, where the requirements to perform the similar analysis (straight, horizontal flight leg, presence of clouds or abrupt changes) hold.

We have applied this method to an environment with clouds, which have a strong contrast between regions with stronger turbulence and the surrounding environment. The method is not universally applicable, since typical flight patterns in airborne measurements consist of slant profiles and horizontal legs, and the latter can include cloud penetrations or not. A cloud penetration is characterised by abrupt changes in temperature and vertical wind velocity, therefore such events affect normalisation, which in turn affects the distribution of points in the phase space, which has a direct connection to how the RP looks, and the values obtained in the RQA. Therefore, the data studied with this method should in principle contain such abrupt events, although their magnitude can vary.

The selected threshold is subjective and its value should not be interpreted universally. On the other hand, all the distances and scales used in the analysis as well as the values and fluctuations present in the analysed time series are typical and in agreement with published data. Thus, the method proposed here should be easily adaptable to other similar data sets as well as to the interpretation of numerical simulations, in which e.g. independent estimates of cloud boundary based on LWC and on dynamic properties of the flow should help not only to characterise properties of turbulence inside and around the cloud, but to enhance our understanding of entrainment, mixing and vertical transport in clouds.
6 Open Research

The archiving of the data used in this study is underway. The data is temporarily available in the supporting information. The MASIN data will be available at CEDA (https://www.ceda.ac.uk). The raw UFT data is available at the repository of open data RepOD (https://doi.org/10.18150/1WMJ3Z). The calibrated UFT data will be available at the repository of open data RepOD. This work used JASMIN, the UK’s collaborative data analysis environment (https://jasmin.ac.uk).

Acknowledgments

We acknowledge funding by Poland’s National Science Centre grant no. UMO-2018/30/M/ST10/00674 and the Natural Environmental Research Council grant numbers NE/S015868/1 and NE/S015779/1.

References


Lance, S., Brock, C., Rogers, D., & Gordon, J. (2010). Water droplet calibration of the cloud droplet probe (CDP) and in-flight performance in liquid, ice and mixed-phase clouds during ARCPAC. *Atmospheric Measurement Techniques*, 3(6), 1683 – 1706. doi: 10.5194/amt-3-1683-2010


Can recurrence quantification analysis be useful in the interpretation of airborne turbulence measurements?

Stanislaw Król¹, Alan Blyth², Steven Böing², Leif Denby²,⁴, Tom Lachlan-Cope³, Szymon P. Malinowski¹

¹Institute of Geophysics, Faculty of Physics, University of Warsaw
²National Centre for Atmospheric Science, University of Leeds
³British Antarctic Survey
⁴Danish Meteorological Institute

Key Points:

• A vector characterising turbulence along the aircraft track based on velocity and temperature fluctuations is constructed.
• A time series of this vector allows the construction of a Recurrence Plot, a tool used in studies of dynamical systems.
• A quantity named laminarity, derived from the Recurrence Plot, is used to detect turbulent and non-turbulent regions.

Corresponding author: Stanislaw Król, skrol@fuw.edu.pl
Abstract

In airborne or model data, clouds are often defined using information about Liquid Water Content (LWC). Unfortunately LWC is not enough to retrieve information about the dynamical boundary of the cloud, i.e. volume of turbulent air around the cloud. In this work, we propose an algorithmic approach to this problem based on a method used in time series analysis of variables in dynamic systems, namely Recurrence Plot (RP) and Recurrence Quantification Analysis (RQA). By studying RP’s constructed with turbulence kinetic energy, vertical velocity and temperature fluctuations and quantifying appropriate time series of laminarity (LAM) calculated using the plots, we distinguish between turbulent and non-turbulent segments along a horizontal flight leg.

Plain Language Summary

Cloud is defined by the presence of liquid (or solid) water in the atmosphere. By agreeing on a certain threshold of liquid water content, one can define a cloud boundary. Cloud processes result in disturbances in the airflow around the cloud (turbulence). Its presence should allow the cloud dynamical boundary to be defined in a similar manner. However, a single simple variable, which could be used in an analogous way is not well defined. In this study we propose a method to define such a variable useful in distinguishing between turbulent and non-turbulent volumes around the cloud (i.e. determining dynamic cloud interface), adopting a method used in the study of dynamical systems. On the base of recorded temperature and air velocity fluctuations characteristic for turbulence, turbulent transport and mixing, a threshold of a variable named “laminarity” is used to define the dynamic cloud boundary on the aircraft trajectory.

1 Introduction

Understanding atmospheric turbulence is crucial to properly describe transport processes in the Atmospheric Boundary Layer (ABL) and in the free atmosphere, mixing processes in clouds, and even rain production (Bodenschatz et al., 2010). Airborne measurements provide direct information on turbulent velocities and associated fluctuations of thermodynamic fields (Wendisch & Brenguier, 2013). Statistical interpretation of these measurements is, however, problematic. Time series collected in the course of airborne measurements represent conditions along a one-dimensional, complicated trajectory inside evolving, coupled fields of momentum, temperature, humidity and cloud particles. In contrast to laboratory experiments, data collected in the course of a flight do not contain information on the boundary and initial conditions of the flow, on external forcings and internal feedbacks. There is even no general rule to distinguish to what extent the observed fluctuations can be contributed to turbulence, waves, or inhomogeneities of non-turbulent nature. Thus an objective division of the collected data series into segments representing similar stages/properties of the flow is crucial to construct conditional statistics of measured variables to characterise phenomena along the flight trajectory.

Typically, additional information and assumptions are necessary to construct reasonable conditional statistics. To explain this in more detail, consider e.g. cloud/clear air turbulent mixing. While the cloud microphysical boundary can be defined in a data series by a selected threshold of liquid water content (LWC) or droplet number concentration (DNC) (Malinowski & Zawadzki, 1993), cloud dynamical boundary, i.e. the region around a cloud where the flow is affected by dynamical (Ackerman, 1958) and thermodynamical (Radke & Hobbs, 1991) processes related to the cloud, is not unique. In airborne measurements porpoising through stratocumulus cloud top (Malinowski et al., 2013) three parameters (turbulence kinetic energy, potential temperature gradient and wind shear) were used to distinguished between the four different layers in the cloud top vicinity. Neither the parameters, nor threshold values could be considered objective, yet the division was algorithmic and allowed to describe many properties of turbulence and
turbulent mixing in stratiform clouds. Such divisions are even more difficult in case of
convective clouds penetrated horizontally. Heus and Jonker (2008) have shown the ex-
istence of shells around clouds where dynamical interactions occur. While cloud mod-
ellers who obtain two- or three- dimensional thermodynamic fields may use such param-
eters as enstrophy (Mellado, 2017) or more advanced numerical approaches (Nair et al.,
2021) to obtain a border between turbulent cloud and non-turbulent environment, a sim-
ilar approach cannot be adopted to airborne data along a 1-D trajectory.

This work describes an effort to construct a robust, objective method to classify
time series collected in the course of flight by "turbulent" (dynamic) properties using a
small number of a-priori assumptions. The idea is to apply a technique which is widely
adopted in nonlinear time series analysis called Recurrence Quantification Analysis (RQA)
(Marwan et al., 2007), which can be used to study transitions in the system. It revolves
around the concept of recurrence of the dynamical system into the same regions of the
phase space, which is one of the fundamental aspects in the analysis of such systems. It
has found applications in many fields such as physiology, neuroscience, economy, and earth
sciences.

In the following analysis, dynamic (turbulent) properties of clouds and clear air in
between are characterised by three variables, representing important dynamical processes
in clouds: turbulence, characterised by its kinetic energy (TKE), mixing, characterised
by temperature fluctuations ($T'$) and vertical transport (updrafts and downdrafts) char-
acterised by vertical velocity ($w'$). These variables, normalised according to the RQA
approach, form a 3-D phase subspace, and recurrences of a vector representing proper-
ties of measured turbulence into the same regions of this phase subspace indicate seg-
ments of the time series representing similar dynamic properties of the flow.

To test applicability of this approach we use airborne data collected in the course
of the EUREC4A (Elucidating the role of clouds-circulation coupling in climate) exper-
iment (Stevens et al., 2021) with the British Antarctic Survey (BAS) Twin Otter (TO)
research aircraft, equipped to measure turbulence, temperature fluctuations, and cloud
properties. Cloud mask defined by a threshold of LWC, is added to the comparison, al-
lowing turbulence inside and outside of cloud to be distinguished.

The structure of the article is the following. In section 2 we describe the data, sec-
tion 3 discusses the details of the time-dependent RQA method adopted, and in section
4 the results are presented and discussed. Final conclusions can be found in section 5
of the paper.

2 Data description

EUREC4A (Stevens et al., 2021), was a measurement campaign conducted in late
January and early February 2020, with a focus on the study of coupling between atmo-
spheric boundary layer and ocean surface, and the study of marine cumulus clouds and
their spatial organisation near Barbados. One of the research aircraft present was the
British Antarctic Survey’s Twin-Otter (TO). The aircraft flew at different altitudes per-
forming cloud penetrations, collecting data between the clouds as well as in the atmo-
spheric boundary layer (ABL), even near the ocean surface.

The TO was equipped with Meteorological Airborne Science INstrumentation (MASIN)
gathering data at two frequencies: 1 Hz and 50 Hz. The data set contains GPS plane
position, pressure altitude, three components of wind velocity from the Best Air Tur-
bulence (BAT) probe (Crawford & Dobosy, 1992; Garman et al., 2006), temperature and
H2O molar ratio. There were cloud spectrometers, measuring cloud microphysical prop-
erties, as well as an additional instrument to measure temperature fluctuations, the Ultra-
Fast Thermometer 2b (UFT-2b) from the UFT family (Kumala et al., 2013) with an ef-
fective resolution reaching 2000 Hz.
Not all collected time series were suitable for analysis due to instrumental problems. The UFT-2b provided useful data from 7 flights, in 5 of these cloud penetrations were recorded. In these flights three 50 Hz velocity components (turbulence data) from the BAT (Best Air Turbulence) probe were available, yet in some segments of flights with deteriorated resolution. Records from BAT and UFT-2b were used in this study. These instruments were co-located within 10 cm distance on the nose boom of the aircraft, thus after averaging temperature fluctuations to match the 50 Hz signal from the BAT probe (corresponding to ~1.3 m spatial resolution at 69 m/s TAS) data can be considered as collected in the same volume.

Additionally, LWC from cloud spectrometers, located under the wings (~ 5 meters apart) was used to define cloud mask. In the first two flights it was the data from the Fast-Forward Scattering Spectrometer Probe (FFSSP) (Baumgardner et al., 1993; Brenguier et al., 1998) with 50 Hz resolution, and for the remaining three, from Cloud Droplet Probe (CDP) (Lance et al., 2010) with 1 Hz resolution (~ 65 m spatial resolution). A total of 246 clouds defined as segments with LWC > 10^{-3} g/kg at 1 Hz were found. Only a fraction of these data were selected for the preliminary analysis described here: a fragment of a time series collected at constant altitude during flight TO-336, starting at around 12:42 UTC. Since the typical true air speed (TAS) of the TO was about 69 m/s, and velocity fluctuations were of order of a few m/s, the frozen flow hypothesis is applicable to data interpretation.

3 Data analysis

As mentioned in the introduction, we decided to perform time-dependent RQA using TKE (denoted in equations as $k$), $T$ and $w'$:

$$\vec{x} = [T, k, w'],$$

as indicators of airflow dynamics. $w'$ is calculated by subtracting the mean calculated over the whole horizontal flight segment.

RQA began from normalisation of the components of the recurrence vector in order to equally capture the influence of each component. The normalisation was performed according to the formula:

$$X_n = \frac{x_n - \mu_{x_n}}{\sigma_{x_n}},$$

where $\mu_{x_n}$ and $\sigma_{x_n}$ is the mean and standard deviation of $n$-th vector component, and the average is taken along the whole horizontal flight segment of the flight.

For each data point on the leg a recurrence matrix Marwan et al. (2007), characterising the dynamical state of the volume of air along a certain segment of the flight trajectory, centred on this point was calculated according to the formula:

$$R_{ij} = \Theta(\epsilon - |\vec{X}_i - \vec{X}_j|),$$

where $\Theta$ is the Heaviside step function, $\epsilon$ is the so-called recurrence threshold and $\vec{X}_i$ is the vector $\vec{X}$ at time $i$, $i \in (1, ..., N)$, where $N$ is the length of $\vec{X}$. As pointed out in Marwan et al. (2007), there is no single universal method for choosing the best $\epsilon$, but there are several “rules of thumb”. According to these rules, a recurrence threshold $\epsilon = 1$ was chosen, which corresponds to about 10% of the diameter of the phase space. The example recurrence matrix and corresponding elements of equation (1) are plotted in figure 1.
Figure 1. Top panel: recurrence plot for 256 experimental points long segment centred on a selected point (a cloud edge) of the time series of the measurement data, bottom panels: plots of individual components of the vector.

According to Marwan et al. (2007) visual inspection of patterns appearing in the recurrence matrix makes it possible to distinguish different types of trajectories of the
investigated vector in phase space. In particular, white areas or bands correspond to rapid changes, chess-board like structures to periodic events, and diagonal lines suggest parallel segments of trajectories in phase space. In figure 1 a change of the recurrence behaviour of the time series after 3 seconds from a slowly changing state to rapid fluctuations is visible. Additionally, two vertical lines around 4 s and 4.5 s correspond to the values of \([T, k, w']\) similar to that in the first 3 s of the plot, visible as a black area on the matrix.

Some of the above-mentioned properties of each matrix can be characterised by one or more measures proposed by Marwan et al. (2007) in the overview of recurrence quantification analysis (RQA). Time series of such measures can be interpreted in terms of varying properties of turbulence along the trajectory of the aircraft.

Since there is no general rule allowing for selection of the recurrence matrix size, we assumed that the matrix built on a segment corresponding to a few integral scales of turbulence should be the right choice. We performed a series of tests, starting from the size corresponding to 128 data points (which corresponds to \(\sim 176\) m at 69 m/s TAS) to 1024 data points, and we decided on a 256 point matrix, corresponding to 352 m, i.e. 1-3 typical integral length scales (Craig & Dörnbrack, 2008; Wacławczyk et al., 2022). In parallel we had to decide which RQA parameter serves our purpose best. We have tested such parameters as recurrence rate, determinism and laminarity (Marwan et al., 2007). Since our time series does not characterise the particular dynamical system, but the actual state of turbulence along the flight track, and the goal is to distinguish between non-turbulent and turbulent regions, we decided to focus on a metric called laminarity \(\text{LAM}\), which, according to Marwan et al. (2007) “represents the probability of occurrence of laminar (slowly changing) states of the system”:

\[
\text{LAM} = \sum_{v=\nu_{\min}}^{v_{\max}} \frac{v P(v)}{\sum_{v=1}^{\nu_{\max}} v P(v)}, \tag{4}
\]

where \(v\) is the length of a black line in the recurrence plot, \(P(v)\) is the number of lines of length \(v\) and \(v_{\min}\) is the minimal line length. In other words, \(\text{LAM}\) is defined as a number of black points forming vertical lines of length equal to at least \(v_{\min}\), normalised by the number of all black points in a Recurrence Plot. For laminar states \(\text{LAM}\) is close to 1 (almost all points form vertical lines), whereas for states with high fluctuations it is closer to 0 (few vertical lines, mostly solitary black points). The choice of parameter \(v_{\min}\) is important, since it excludes shorter lines from calculations. In this study \(v_{\min} = 10\), corresponding to a \(\sim 14\) m distance, was chosen. The argument is that we look for fluctuations of sizes corresponding to large scales of the turbulence cascade, small scale variability should be studied after selecting segments for more detailed analysis.

Finally, in order to understand the relative importance of turbulence kinetic energy, mixing and updrafts/downdrafts analogous analysis, construction of recurrence matrices and following RQA, was performed for each vector component separately. The example application of the procedure described in the previous section is presented in figure 2.

The top panel of figure 2 shows three time series of laminarity calculated for various sizes (512, 256 and 128 points corresponding to \(\sim 704\), \(\sim 352\) and \(\sim 176\) m distance in physical space) of the RP. It can be seen that the black curve representing an RP of size 512 \(\times\) 512 hardly reflects dynamic disturbances related to small clouds, such as the one at \(t \approx 7350\) s. The blue curve representing \(\text{LAM}\) calculated for a 128 point long segment is most sensitive to changes in the environment, but shows strong oscillations, which are hard to interpret in terms of turbulence properties. In general, increasing the size of the matrix smooths out the time-dependent RQA, and broadens the peaks of the quantities. This is an effect of capturing a bigger picture around a selected point and
related smoothing, as seen for example at $t \approx 7250s$, where the black curve does not
give insight into the variability of dynamics captured by the red and blue curves corres-
ponding to smaller matrices. Additionally, 256 points / $\sim 352$ m corresponds roughly
to 3-4 integral scales of turbulence, and distinguishing between turbulent/non turbulent
regions on a distance corresponding to a few integral scales seems reasonable.

The bottom panel of figure 2 shows a plot of $LAM$ time series for the whole vec-
tor (light-blue curve) and for its components ($T$, $k$, $w'$). Grey shading indicates the sep-
arate use of a cloud mask indicating the presence of cloud water (at 1 Hz, i.e. 69 m spa-
tial resolution). It can be seen that considering the whole vector gives the complete pic-
ture of turbulent regions within and around the clouds, while laminarity calculated for
each component separately shows the relative importance of $k$, $T$ and $w'$, i.e. mechan-
ical turbulence, mixing and vertical transport which change along the time series.

4 Results

Figure 3 shows the time series for all three components of the vector, LWC series
as the grey cloud mask and $LAM$ calculated for the whole vector, with yellow shaded
regions corresponding to $LAM$ below the selected threshold, i.e. to the regions where
a signature of turbulence is observed. Since the goal is to distinguish between the tur-
bulent and laminar regions, a threshold just below 1 should be the cutoff. After visual
inspection the specific value of $LAM = 0.95$ was chosen. Selection of $LAM = 0.99$ re-
sulted in selection of the majority of the record as turbulent, while $LAM = 0.90$ resulted
in omission of segments where some clear signatures of turbulence were present. The ex-
act value of the threshold can be discussed, however the goal of the selection is not the
detailed turbulence analysis, but selection of the segments on which such analysis should
be performed so the detailed number is of secondary importance.

Figure 3 shows that the investigated region is composed of clouds with vast vol-
umes around the clouds affected by turbulence and laminar regions with occasional sig-
natures of turbulence, not related to visible amounts of LWC. All clouds in the analysed data are turbulent, with clearly visible local minima of $LAM$ values below $\sim0.75$. The local minima of laminarity reflect coincident input of all the components of the vector, but there is no clear dependence to the actual value of LWC.

Additionally, for each cloud, regions with $LAM < 0.95$ around the cloud can be considered “turbulent cloud shells”, and borders of these shells can be determined by the $LAM$ threshold: such a division will help in more detailed investigation of cloud dynamics.

To close this analysis we performed initial statistical analysis of turbulence in the time series. Table 1 summarises the average turbulence kinetic energy, vertical velocity variance and temperature variance in 3 different types of segments along the trajectory: clouds (LWC above threshold), turbulent regions with no clouds ($LAM$ above threshold) and non-turbulent regions in between. Not surprisingly, these values are very different, reflecting different dynamics along the flight trajectory.

<table>
<thead>
<tr>
<th>$LAM^&gt;$</th>
<th>LWC</th>
<th>$LAM^&lt;$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu_k$</td>
<td>0.79</td>
<td>2.46</td>
</tr>
<tr>
<td>$\sigma_T$</td>
<td>0.32</td>
<td>0.32</td>
</tr>
<tr>
<td>$\sigma_{w'}$</td>
<td>0.58</td>
<td>1.31</td>
</tr>
</tbody>
</table>
As we can see the values of the mean TKE are the highest for cloud regions, but they are also significantly higher for regions with $LAM$ below threshold than for $LAM$ above threshold. Standard deviation of vertical velocity follow the same trend as the mean TKE, while standard deviation of temperature is not significantly different between the regions. However, more detailed inspection of the data indicates, that the standard deviation of the temperature in laminar region comes from large scale smooth temperature variations, while in turbulent regions it comes from small scale variability. Consequently, the mask based on the changes of $LAM$ allows for better division of data in terms of turbulence characteristics.

5 Summary and outlook

In this study we proposed a method to algorithmically divide time series of airborne measurements according to dynamic (turbulent) properties of the sampled air, combining the simple signature properties characterising turbulence, turbulent transport and turbulent mixing. The method is based on recurrence quantification analysis and the assumptions adopted are based on well-documented physical characteristics of atmospheric turbulence. We believe that this approach will facilitate future investigations of atmospheric turbulence in, within and around clouds, which requires a more systematic approach. The main advantage of the RQA method is that the constructed vector can be expanded with more measured quantities, such as LWC. However, the method is useful even when the microphysical signal is of poor quality, or in situations where it is independently used to determine the cloud-clear air interface.

We used only one metric from the RQA, laminarity, but RQA offers a variety of metrics, therefore it is possible to expand the studies and/or use metrics that are more suitable for other types of analyses. The possibilities in the future with regards to this type of study include imposing an additional condition based on the LWC signal (of a better resolution) added to the vector, as well as applying the method to data gathered during other campaigns, where the requirements to perform the similar analysis (straight, horizontal flight leg, presence of clouds or abrupt changes) hold.

We have applied this method to an environment with clouds, which have a strong contrast between regions with stronger turbulence and the surrounding environment. The method is not universally applicable, since typical flight patterns in airborne measurements consist of slant profiles and horizontal legs, and the latter can include cloud penetrations or not. A cloud penetration is characterised by abrupt changes in temperature and vertical wind velocity, therefore such events affect normalisation, which in turn affects the distribution of points in the phase space, which has a direct connection to how the RP looks, and the values obtained in the RQA. Therefore, the data studied with this method should in principle contain such abrupt events, although their magnitude can vary.

The selected threshold is subjective and its value should not be interpreted universally. On the other hand, all the distances and scales used in the analysis as well as the values and fluctuations present in the analysed time series are typical and in agreement with published data. Thus, the method proposed here should be easily adaptable to other similar data sets as well as to the interpretation of numerical simulations, in which e.g. independent estimates of cloud boundary based on LWC and on dynamic properties of the flow should help not only to characterise properties of turbulence inside and around the cloud, but to enhance our understanding of entrainment, mixing and vertical transport in clouds.
6 Open Research

The archiving of the data used in this study is underway. The data is temporarily available in the supporting information. The MASIN data will be available at CEDA (https://www.ceda.ac.uk). The raw UFT data is available at the repository of open data RepOD (https://doi.org/10.18150/1WMJ3Z). The calibrated UFT data will be available at the repository of open data RepOD. This work used JASMIN, the UK’s collaborative data analysis environment (https://jasmin.ac.uk).

Acknowledgments

We acknowledge funding by Poland’s National Science Centre grant no. UMO-2018/30/M/ST10/00674 and the Natural Environmental Research Council grant numbers NE/S015868/1 and NE/S015779/1.

References


Lance, S., Brock, C., Rogers, D., & Gordon, J. (2010). Water droplet calibration of the cloud droplet probe (CDP) and in-flight performance in liquid, ice and mixed-phase clouds during ARCPAC. *Atmospheric Measurement Techniques, 3*(6), 1683 – 1706. doi: 10.5194/amt-3-1683-2010
