PROCEEDINGS OF SPIE

SPIEDigitalLibrary.org/conference-proceedings-of-spie

Representative atmospheric turbulence profiles for ESO Paranal

O. J. D. Farley, J. Osborn, R. W. Wilson, T. Butterley, D. Laidlaw, et al.

O. J. D. Farley, J. Osborn, R. W. Wilson, T. Butterley, D. Laidlaw, M. Townson, T. Morris, M. Sarazin, F. Derie, M. Le Louarn, A. Chacón, X. Haubois, J. Navarrete, J. Milli, "Representative atmospheric turbulence profiles for ESO Paranal," Proc. SPIE 10703, Adaptive Optics Systems VI, 107032E (10 July 2018); doi: 10.1117/12.2312760

SPIE.

Event: SPIE Astronomical Telescopes + Instrumentation, 2018, Austin, Texas, United States

Representative atmospheric turbulence profiles for ESO Paranal

O. J. D. Farley[a], J. Osborn[a], R. W. Wilson[a], T. Butterley[a], D. Laidlaw[a], M. J. Townson[a], T. Morris[a], M. Sarazin[b], F. Derie[b], M. Le Louarn[b], A. Chacón[b], X. Haubois[b], J. Navarrete[b], and J. Milli[b]

[a]Centre for Advanced Instrumentation, University of Durham, Durham, UK
[b]European Southern Observatory, Karl-Schwarzshild-Str.2, 85748 Garching bei Muenchen, Germany

ABSTRACT

The optical turbulence profile is a key parameter in tomographic reconstruction. With interest in tomographic adaptive optics for the next generation of ELTs, turbulence profiling campaigns have produced large quantities of data for observing sites around the world. In order to be useful for Monte Carlo AO simulation, these large datasets must be reduced to a small number of profiles. There is commonly large variation in the structure of the turbulence, therefore statistics such as the median and interquartile range of each altitude bin become less representative as features in the profile are averaged out. Here we present the results of the use of a hierarchical clustering method to reduce the 2018A Stereo-SCIDAR dataset from ESO Paranal, consisting of over 10,000 turbulence profiles measured over 83 nights, to a small set of 18 that represent the most commonly observed profiles.

Keywords: Atmospheric turbulence, clustering

1. INTRODUCTION

In tomographic adaptive optics (AO) knowledge of the vertical distribution of the atmospheric turbulence, parameterized by the refractive index structure constant $C_n^2(h)$ as a function of height h, is required in order to compute the tomographic reconstruction of the turbulent phase.[1,2] The turbulence profile, being a meteorological phenomenon, is variable on timescales from several minutes to entire seasons and is specific for a particular observing site.

At ESO Paranal, an ongoing campaign with a Stereo-SCIDAR[3] instrument has produced a large database of over 10,000 measurements of the full atmosphere turbulence profile with high resolution (250 m bins).[4] However in itself databases such as this are not particularly useful for Monte Carlo AO simulation due to long computation times for a single profile. We therefore require a method of reducing this large dataset to small number of "typical" profiles.

Conventionally this reduction takes the form of binning the profiles, usually according to some integrated parameter of the turbulence such as the seeing (integrated strength) then taking an average profile for each bin, resulting in, for example, a profile associated with each seeing quartile. However there is no guarantee that the profiles in each seeing bin will share the same structure and therefore features in the profile tend to be averaged out, resulting a profile or set of profiles that may never have been measured. We opt to employ cluster analysis to group profiles according to their structure thus reducing this loss of information.

Email: o.j.d.farley@durham.ac.uk (OJDF)

Adaptive Optics Systems VI, edited by Laird M. Close, Laura Schreiber, Dirk Schmidt, Proc. of SPIE Vol. 10703, 107032E · © 2018 SPIE
CCC code: 0277-786X/18/$18 · doi: 10.1117/12.2312760

2. CLUSTERING METHOD

We opt for hierarchical clustering[5] since we have found this produces good results for turbulence profiles. In addition we are limited by our choice of metric to measure the distance between profiles. The commonly used euclidean distance is too sensitive for C_n^2 measurements that span many orders of magnitude, resulting in many small outlier clusters that are not particularly useful or representative. We find the cosine or angular distance produces good results, defined as

$$\delta_{ij}^{\cos} = 1 - \frac{\mathbf{x}_i \cdot \mathbf{x}_j}{\|\mathbf{x}_i\|_2 \|\mathbf{x}_j\|_2}, \tag{1}$$

where $\mathbf{x}_i$ and $\mathbf{x}_j$ are two turbulence profile measurement vectors and $\|\mathbf{x}\|_2$ denotes the euclidean or L2 norm of the vector $\mathbf{x}$. Note that the cosine distance is sensitive only to changes in the shape of the profile and not changes in integrated strength. This metric as defined above assumes that all the vector elements (altitude bins) are independent. This is clearly not the case for turbulence profiles measured with an instrument with finite altitude resolution. Thus we modify the cosine distance by including a matrix s that describes the similarity between altitude bins.[6] We define this matrix according to the Stereo-SCIDAR response functions for the average star separation of 12.5 arcseconds.

In addition we choose to normalise each altitude bin by its mean value before clustering. This transformation ensures that turbulence at all altitudes is considered equally, i.e. the strong ground layer is reduced and the weak high altitude turbulence is increased slightly.

In order to decide on the number of clusters we employ two metrics. The first is the within-cluster variance, defined as the sum of cosine distances between members of each cluster $\mathbf{x}_{im}$ and the cluster centre $\bar{\mathbf{x}}_m$, summed over all clusters m:

$$W_N = \sum_{m=1}^{N} \sum_{i=1}^{n_m} \delta^{\cos}(\mathbf{x}_{im}, \bar{\mathbf{x}}_m). \tag{2}$$

This metric will decrease as the number of clusters increases and the profiles in each cluster become more tightly grouped. The rate of decrease will fall off as the clustering becomes less effective and it is at this point that we define the number of clusters, a technique known as the "elbow method".

The second metric we employ is the silhouette score.[7] For a single profile i this is defined as

$$s_i = \frac{b_i - a_i}{\max\{a_i, b_i\}}, \tag{3}$$

where a_i is the mean distance of i from the other profiles in its cluster and b_i is the mean distance of i from the profiles in all other clusters. Thus a positive silhouette score indicates this profile is closer on average to profiles in its own assigned cluster than to profiles in other clusters. We therefore seek a peak in the silhouette score to indicate the optimum number of clusters.

3. RESULTS

The metrics for choosing the number of clusters are shown in Fig. 1. We can see a peak in the silhouette score at 17 to 19 clusters which corresponds with an elbow in the within-cluster variance. We therefore suggest that using this method we extract 18 clusters from the 2018A dataset.

We can assess the clustering visually using a dendrogram as seen in the upper panel of Fig. 2. This also allows us to order the profiles according to the leaves of the dendrogram as can be seen on the lower panel of Fig. 2. The particular combination of clustering parameters chosen gives a reasonable set of interesting profiles, as can be seen in Fig. 3. Most obvious is the bias towards single strong turbulent layers. This is as a result of our cosine distance metric, where the dot product of two strong turbulent layers, sometimes up to an order of magnitude larger than adjacent C_n^2 measurements, dominates the resulting distance between the two profiles. To capture multi-layer variation we would require many more clusters, a departure from our initial goal to produce a small number of typical profiles that can be feasibly used in Monte-Carlo simulations.

Also apparent is the propensity for high altitude turbulence at Paranal. 10 of the 18 clusters exhibit turbulent layers at or above 10 km, representing around 60% of the dataset. In addition 3 clusters (1, 3 and 6) exhibit

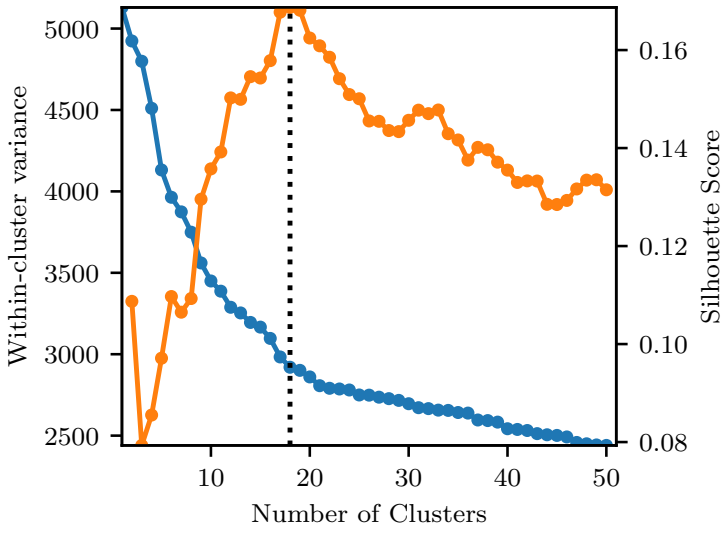

Figure 1. Silhouette score (orange) and within-cluster variance (blue) metrics for the 2018A dataset. Soft cosine distance and average linkage hierarchical clustering are used here. Dotted vertical line at 18 clusters.

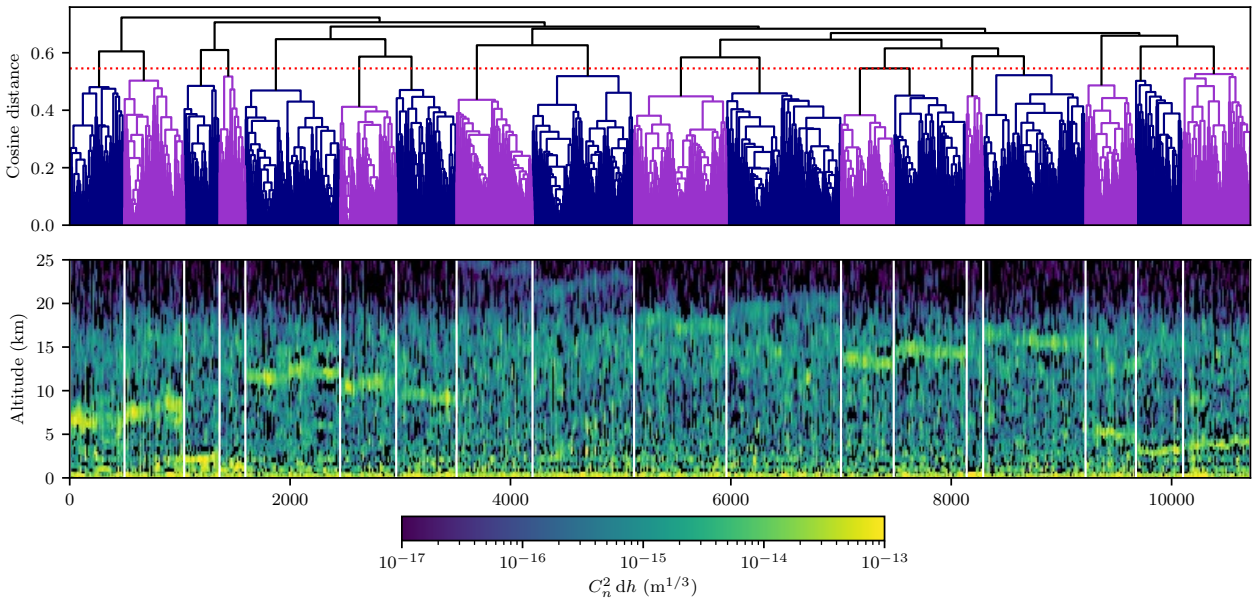

Figure 2. *Upper*: Dendrogram illustrating the average linkage hierarchical clustering of the 2018A dataset. *Lower*: The turbulence profiles in the dataset ordered according to the leaves of the dendrogram, with white vertical lines indicating the partitioning of the dataset into clusters.

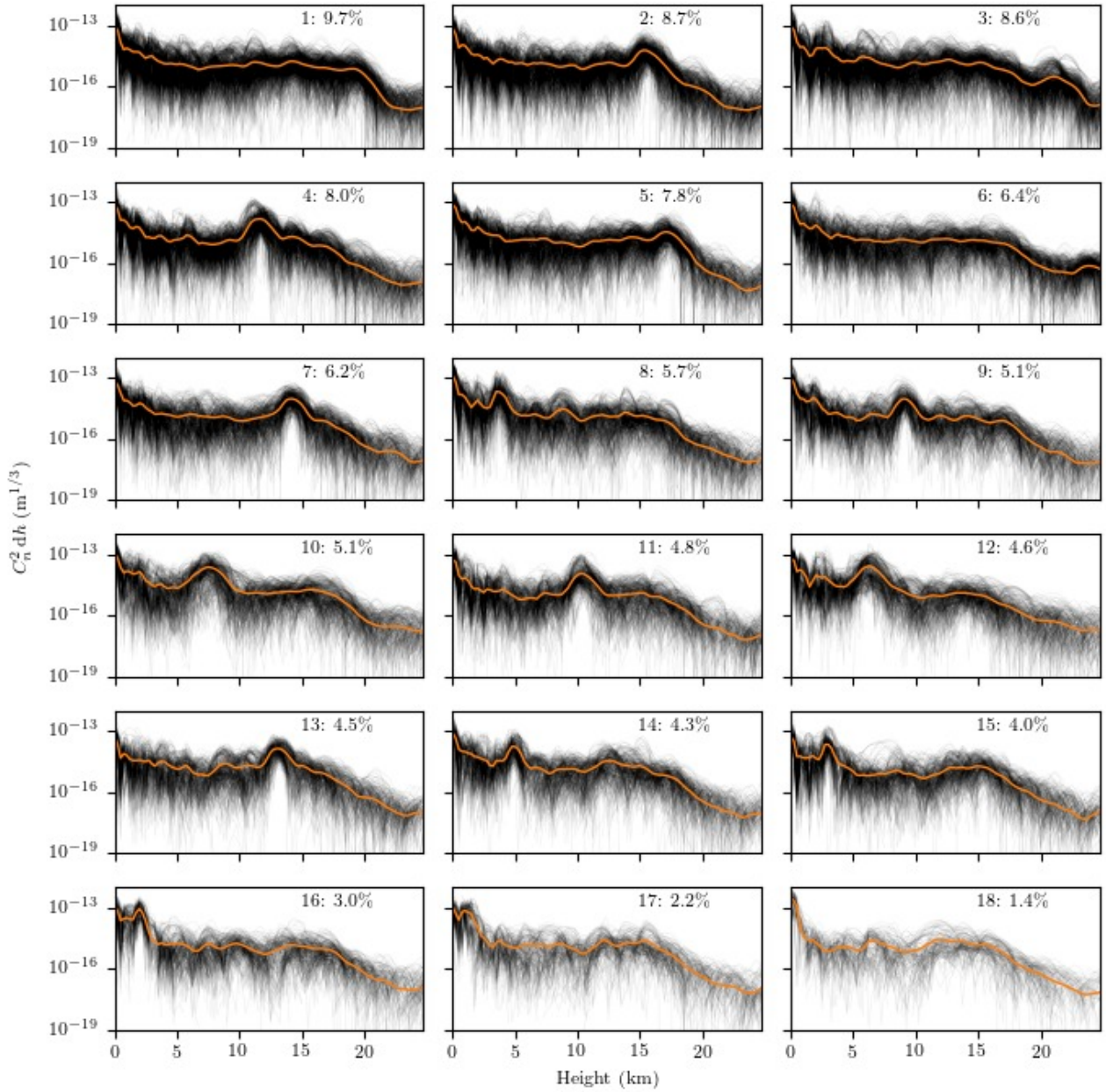

Figure 3. The 18 clusters, with each profile in a cluster displayed in black. The orange line represents the centre of each cluster. Each cluster is numbered 1-18 and has an associated size, represented as a percentage of the size of the total dataset.

Table 1. Atmospheric parameters for the 18 clusters. In all cases a wavelength of 500 nm is assumed. Ground layer here is defined as the first three altitude bins, i.e. all turbulence below 625 m.

Cluster	r_0 (cm)	θ_0 (arcsec)	Ground layer frac.
1	17.1	2.1	0.58
2	17.9	1.7	0.44
3	15.9	2.0	0.50
4	15.4	1.4	0.38
5	17.1	1.8	0.49
6	17.1	2.2	0.53
7	16.1	1.5	0.44
8	12.9	1.9	0.49
9	15.0	1.9	0.49
10	10.8	1.2	0.37
11	17.7	1.7	0.41
12	12.8	1.6	0.27
13	16.5	1.3	0.32
14	14.8	1.7	0.39
15	18.0	2.2	0.35
16	10.8	2.1	0.32
17	11.5	2.0	0.34
18	9.8	2.1	0.85

non negligible very high altitude turbulence above 20 km. The amount of high altitude turbulence present in the clusters is not a surprise since it reflects the small ground layer fractions that tend to be found in the profiles.[4] We do however obtain one strongly ground layer dominated cluster (18), representing only around 1% of the dataset.

In Table 1 we show integrated atmospheric parameters for the clusters. We can see that there is not a large spread in these parameters. Since we have performed clustering only on the shape of C_n^2 profiles, the dataset has been partitioned into clusters that reflect the changes in structure of the profile but not changes in atmospheric parameters, for example r_0. Therefore the profiles in each cluster follow similar distributions of these parameters. We emphasize that these are typical profiles, not extreme cases in the parameter space. Such clusters could be produced by first selecting profiles with extreme atmospheric parameters then performing the clustering to obtain the different profiles.

4. CONCLUSIONS

We have presented the results of hierarchical clustering of the 2018A Stereo-SCIDAR turbulence profile dataset from Paranal, producing 18 turbulence profiles. Choice of the number of clusters is informed by two metrics derived from the data. In addition, the partitioning of the dataset by clustering associates an approximate probability of observation with each cluster.

These profiles represent the most common atmospheric conditions observed in the dataset, therefore they will be useful for Monte-Carlo simulation of AO systems for the VLT at Paranal or the ELT, only 20 km away. The method developed for these clusters would be applicable to any other large database of turbulence profiles.

Further work will investigate the temporal properties of clusters, both on short (minute) and long (seasonal) timescales.

ACKNOWLEDGMENTS

OJDF acknowledges the support of STFC (ST/N50404X/1).

REFERENCES

[1] Fusco, T., Conan, J.-m., Rousset, G., Mugnier, L. M., and Michau, V., "Optimal wave-front reconstruction strategies for multiconjugate adaptive optics," *Journal of the Optical Society of America A* **18**, 2527 (10 2001).

[2] Vidal, F., Gendron, E., and Rousset, G., "Tomography approach for multi-object adaptive optics," *Journal of the Optical Society of America A* **27**, A253 (11 2010).

[3] Shepherd, H. W., Osborn, J., Wilson, R. W., Butterley, T., Avila, R., Dhillon, V. S., and Morris, T. J., "Stereo-SCIDAR: optical turbulence profiling with high sensitivity using a modified SCIDAR instrument," *Monthly Notices of the Royal Astronomical Society* **437**, 3568–3577 (2 2014).

[4] Osborn, J., Wilson, R. W., Sarazin, M., Butterley, T., Chac, A., Derie, F., Farley, O. J. D., Laidlaw, X. H. D., Lelouarn, M., Masciadri, E., Milli, J., and Townson, M. J., "Optical turbulence profiling with Stereo-SCIDAR for VLT and ELT," *Monthly Notices of the Royal Astronomical Society* (2018).

[5] Everitt, B. S., Landau, S., Leese, M., and Stahl, D., [*Cluster Analysis*], John Wiley & Sons, Ltd., 5th ed. (2011).

[6] Sidorov, G., Gelbukh, A., Gómez-Adorno, H., and Pinto, D., "Soft Similarity and Soft Cosine Measure: Similarity of Features in Vector Space Model," *Computación y Sistemas* **18**, 491–504 (9 2014).

[7] Kaufman, L. and Rousseeuw, P. J., [*Finding Groups in Data: An Introduction to Cluster Analysis*], John Wiley & Sons, Inc. (2005).