A large-scale analysis of pockets of open cells and their radiative impact

D. Watson-Parris[1], S. A. Sutherland[1*], M. W. Christensen[1†], Ryan Eastman[2]and P. Stier[1]

[1]Atmospheric, Oceanic and Planetary Physics, Department of Physics, University of Oxford, UK
[2]Department of Atmospheric Sciences, University of Washington, Seattle, Washington
[*]Now at The University of New South Wales, New South Wales, Australia
[†]Now at Pacific Northwest National Laboratory, Richland, WA, USA

Key Points:

- Convolutional Neural Networks are used to detect 8491 pockets of open cells in marine stratocumulus between 2005-2018.
- The first comprehensive analysis of their microphysical and climatological properties is presented.
- Their global radiative effect is found to be neglible. Closing all POCs would lead to an instantaneous top-of-atmosphere imbalance of only 0.01 $\mathrm{W/m^2}$

Corresponding author: Duncan Watson-Parris, duncan.watson-parris@physics.ox.ac.uk

Abstract

Pockets of open cells sometimes form within closed-cell stratocumulus cloud decks but little is known about their statistical properties or prevalence. A convolutional neural network was used to detect occurrences of pockets of open cells (POCs). Trained on a small hand-logged dataset and applied to 13 years of satellite imagery the neural network is able to classify 8,491 POCs. This extensive database allows the first robust analysis of the spatial and temporal prevalence of these phenomena, as well as a detailed analysis of their micro-physical properties. We find a large (30%) increase in cloud effective radius inside POCs as compared to their surroundings and similarly large (20%) decrease in cloud fraction. This also allows their global radiative effect to be determined. Using simple radiative approximations we find that the instantaneous global annual mean top-of-atmosphere perturbation by all POCs is only $0.01\,\mathrm{W/m^2}$.

Plain Language Summary

The amount of sunlight that reaches, and warms, the surface of the earth is heavily influenced by clouds, in particular marine stratocumulus clouds, a type of low-lying cloud that forms above cold-upwelling regions of the ocean. Marine stratocumulus clouds form in two distinct regimes; open-cells and closed-cells. Closed-cell clouds have a higher cloud cover and reflectivity than open-cell clouds. Small pockets of open cell clouds sometimes form within larger regions of closed-cell clouds, these are referred to as 'pockets of open cells'. Here we use machine learning to detect occurrences of this phenomenon and characterise them in a long-term satellite dataset. This allows their effect on the climate to be determined for the first time. Despite substantial local-scale changes in cloud properties we find that their effect on the climate is small.

1 Introduction

Stratocumulus clouds play a vital role in the global energy balance (Randall et al., 1984) and can exist in two distinct regimes: open cells and closed cells (Agee et al., 1973), which can be considered two states of a coupled oscillator (Koren & Feingold, 2011). First coined in 2004 (Bretherton et al., 2004), POCs are small regions of open cell clouds embedded in a uniform surrounding deck of closed cell clouds. Despite the importance of the marine stratocumulus decks on the global climate (Randall et al., 1984; Stevens et al., 2005; Hansen et al., 2013), studying POCs poses several difficulties due to their complex and ill defined nature. POCs also cannot be resolved by the general circulation models (GCMs) used to model the global climate due to their relatively small spatial scale and a lack of representation of mesoscale organisation in such models (Berner et al., 2013a). The global radiative impact of POCs and hence that of their absence in GCMs is currently unknown.

Closed cells have a markedly higher albedo than open cells for a given cloud fraction (McCoy et al., 2017), and this is exacerbated by their increased cloud fraction (Rosenfeld et al., 2006). Factors affecting the open/closed transition could have a dramatic effect on the total contribution of these clouds to the planetary albedo. For example, it has been proposed that anthropogenic aerosol could have a large effect on the number of POCs, and in turn lead to a large top-of-atmosphere radiative perturbation (Rosenfeld et al., 2006).

Since their discovery, many studies have investigated the properties of selected POCs. It has been observed that POCs are coherent and long-lived, lasting tens of hours (Stevens et al., 2005; Berner et al., 2013a; Wang & Feingold, 2009), and typically consisting of fewer, larger cloud droplets than the surrounding cloud (Stevens et al., 2005; Wood et al., 2011; Terai et al., 2014). They locally have a reduction in cloud optical depth, and a stronger tendency to precipitate compared to the surrounding cloud (Stevens et al.,

2005; Berner et al., 2013a). The surrounding cloud outside of a POC is moister than in uniform closed cells (Stevens et al., 2005), there is a reduced rate of entrainment drying, and more efficient coalescence scavenging in POCs than in the surrounding cloud (Stevens et al., 2005; Berner et al., 2013a; Wood et al., 2011; Terai et al., 2014). Modelling studies have shown the formation of POCs is likely driven by a sharp increase in collision-coalescence as Liquid Water Path (LWP) increases and cloud droplet number decreases (Berner et al., 2013b), and is maintained through the enhanced nucleation caused by the very efficient wet removal of aerosol just below the inversion (Kazil et al., 2011).

While painting a consistent picture, each of the observational studies described above include very few POC cases (at most five), and often used data that was not collected specifically for the analysis of POCs. Although one recent analysis presented a large hand-annotated collection of open-cellular cloud cover over the south east Atlantic (Abel et al., 2019), no global analysis of POCs has been performed. Here we use a machine learning technique to automatically detect POCs from satellite images and build up a database of almost 8500 POCs, shedding light for the first time on their spatial and temporal distributions. Using this database we are also able to make robust estimates of the average micro- and macro-physical properties of these phenomena in each of the three main stratocumulus regions of the globe.

We will outline the data used and the machine learning methods applied to the problem in Sec. 2; describe the spatial and temporal distributions of the POCs and their average physical properties in Sec. 3; before concluding in Sec. 4 with a discussion of the implications for these results and an indication of some of the many other possible uses for the database.

2 Method

For the POC detection process we use true-color RGB composites generated using SatPy (Raspaud et al., 2018) from the Moderate Resolution Imaging Spectrometer (MODIS) on board the NASA Terra (MODIS Science Team, 2015) satellite. The Level 1B data sets were used which provide calibrated and geolocated radiances for all 36 MODIS spectral bands at 1km resolution. Due to the relatively large size of POCs and to speed up training and detection the images were linearly resampled from 1350x2030 pixels and split in two, producing 224x224 pixel images.

In order to train the machine-learning model it is necessary to create a dataset of satellite images and hand-logged POC masks. However, determining whether a particular pattern in a cloud deck is truly a POC or not can be ambiguous as no clear definition currently exists. In order to ensure that the labelling of the POCs was consistent, a set of rules were devised. These rules were designed to balance the number of falsely identified POCs and the number of missed POCs. The finalised rules are as follows:

1. The structure of the POC and surrounding cloud must be correct: POCs must be open cell cloud, which looks 'stringy', and the surrounding cloud must be closed cell cloud, which looks 'bubbly'. It is often hard to distinguish between a thin layer of closed cell clouds, where the 'bubbles' become separated and open cell clouds, but identifying it as one of these two descriptive words helps to decide which it is.
2. At least 80% of the perimeter of the POC must be continuous closed cell cloud. This is still likely a POC, since it has formed mostly embedded, and not requiring a POC to be completely embedded significantly increases the amount of data with which to train the model.
3. POCs can be at the edge of an image. While there is no way of knowing what the cloud deck looks like beyond the bounds of the image, it is beneficial for the neu-

ral net to count these as POCs, since they share the same structure and properties as POCs elsewhere, allowing it to learn better.

4. The boundary must be 'reasonably sharp' on all edges. A sharp transition from open to closed cell clouds is characteristic of a POC, and so if this transition was too gradual the POC was ignored.

Applying these labelling rules to a selection of 1029 images resulted in a dataset of 216 images containing 715 POCs, and 813 images that are known to contain no POCs[1].

The model itself uses a modified ResNet-152 (He et al., 2015) with the dense layers replaced by three up-sampling blocks based on the second half of the ResUnet model (Zhang et al., 2017). The ResNet-152 portion of the model is pre-trained on ImageNet (Deng et al., 2009). The upsampling blocks are trained using a DICE loss function and Adam optimizer, with a learning rate that decayed by factors of 0.2 when the validation loss plateaued for 3 epochs. The final masks are refined using a second, reduced ResUnet, model that was trained in the same way as the ResNet-152 model. These models were both implemented in Keras (Chollet et al., 2015), using the TensorFlow engine (Abadi et al., 2015) and are freely available as described in the Acknowledgements.

To gauge the performance of the model in terms of true positives (TP) and negatives (TN), and false positives (FP) and negatives (FN), a balanced accuracy (bACC) score is used (L. Olson & Delen, 2008):

$$\texttt{bACC} = \frac{1}{2}\left(\frac{\texttt{TP}}{\texttt{TP}+\texttt{FN}} + \frac{\texttt{TN}}{\texttt{TN}+\texttt{FP}}\right). \tag{1}$$

These scores were initially calculated on a pixel-by-pixel basis, however the training set of POCs was constructed using masks that had straight edges, while the model can track the edge of a POC with much greater accuracy. This meant that the model was unfairly penalised since the exact shape of the POC did not match up, resulting in greater numbers of false negatives and positives. To combat this, the metrics were calculated using whether or not the image contained a POC, since if it did it is likely that the POC was placed correctly due to its distinctive nature. The balanced accuracy score of the model on the test set of 100 images was 0.863. Therefor any POC identified by the model has an 86% chance of being a correct identification, meaning the vast majority of POCs in the dataset created by the model are true positives. In other iterations of the model, it was possible to get a much higher recall (TP/(TP+FN)), however only to the detriment of the precision (TP/(TP+FP)). We choose to prefer ensuring more of the POCs found are true positives rather than detecting as many as possible and introducing false positives.

It should also be noted that, even with the rules enumerated above, labelling POCs could often be quite an ambiguous task, and visual inspection confirmed that the network performed very well. While the masks deviated from the human labels in some places, those differences were entirely reasonable and in some cases more accurate than the human labelling. An example of a POC annotated by hand and by the model can be seen in Fig. 1.

We apply the inferred POC masks to a wide selection of satellite retrievals to build statistics about the POCs and their surrounding environment. Primarily we use the retrieved MODIS Terra (MOD06_L2) cloud properties (Platnick et al., 2015) which are already available on the same coordinates and have been shown to compare favourably to in-situ observations (Painemal & Zuidema, 2011; Min et al., 2012). To complement these retrievals we have included collocated cloud retrievals from the combined cloud profiling radar (CPR; on Cloudsat) and Cloud-Aerosol Lidar with Orthogonal Polarization

[1] This resource is being made publicly available - see the Acknowledgments for details.

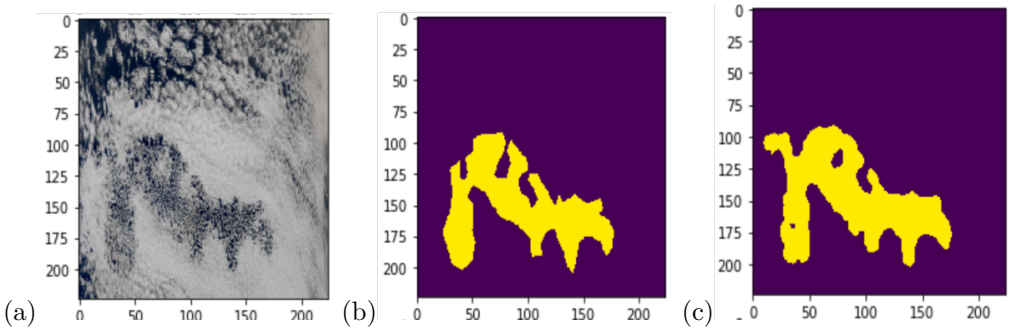

Figure 1. From left to right: (a) An example rescaled input image of a POC, (b) the hand logged test mask, and (c) the inferred mask from the machine learning model (which did not see this POC during training).

(CALIOP; on Calipso) to determine the cloud base and top height (2B-GEOPROF-LIDAR; Mace & Zhang, 2014), as well as cloud and rain water paths (2C-RAIN-PROFILE; Lebsock & L'Ecuyer, 2011). We have also included liquid water path estimates from the Advanced Microwave Scanning Radiometer for EOS (AMSR-E) on board Aqua (Wentz & Meissner, 2004); the Advanced Microwave Scanning Radiometer 2 (AMSR2) onboard the GCOM-W1 (Chang et al., 2015). An additional subset of 1572 POCs were detected in MODIS Aqua imagery in order to facilitate collocation between these other products retrieved from the A-train constellation, 398 of which intersected the narrow CloudSat-CALIOP swath.

For visualisation of the background stratocumulus amount in each region, as well as calculation of the radiative effect of POCs we use a combination of nearly 2 million individual MODIS stratocumulus classifications (Wood & Hartmann, 01 May. 2006) and MODIS monthly cloud fraction (Hubanks et al., 2008) between 2003 and 2011.

Due to their very irregular shapes and sizes it is not possible to create an average, or composite POC. Instead, using OpenCV (Bradski, 2000) to extract regions of fixed distance from the edge of each POC we can plot the average MODIS properties as a function of distance from the edges of all of the detected POCs. For CloudSat-CALIOP and AMSR, the sparser sampling of these instruments necessitates a coarser aggregation and so their results are reported only as average quantities inside and outside of the POC boundary.

3 Results and discussion

The model was run on all MODIS images which intersected the three main marine stratocumulus regions in the north-east Pacific (off the coast of California), south-east Pacific (off the coast of Peru) and south-east Atlantic (off the coast of Namibia), as defined by Klein and Hartmann (1993), between 2005 to 2018. From the 51,164 images inspected 8,491 POCs were detected, in 4,729 of the images.

Figure 2 shows the temporal distributions of POC occurrence. These have been normalised using the number of images used and the average stratocumulus cloud amount taken from ISCCP data (Young et al., 2018) in order to remove the strong seasonal cycle in the quantity of stratocumulus cloud. One of the most striking features is the magnitude of Peruvian distribution which shows more than three times the number of POCs compared to the other regions. All regions show a well defined peak around local wintertime, coinciding with the maximum stratocumulus amount. The Californian stratocu-

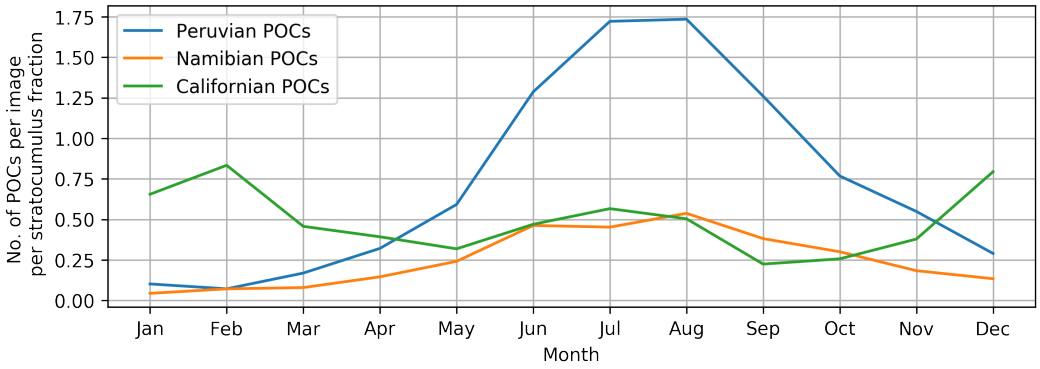

Figure 2. Temporal distribution of POC occurrences detected in MODIS true-color imagery by our algorithm for the three regions studied, normalised by the number of images used and the average stratocumulus cloud amount in that region.

mulus deck, however, also shows a second peak during June to August, the local summertime.

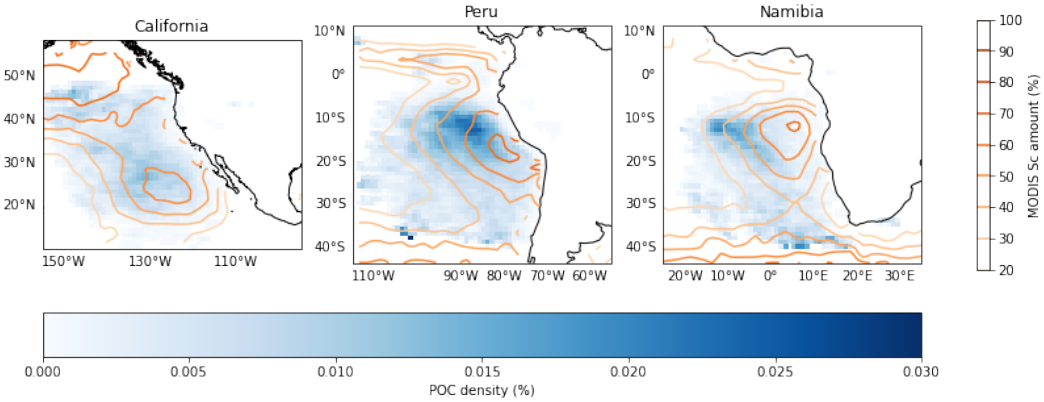

Figure 3. The percentage density of detected POCs in the three regions of interest with the MODIS climatological stratocumulus amount overlaid. The POC dataset represents an average between 2005 to 2018, while the MODIS stratocumulus amount is averaged over 2003 to 2011.

Figure 3 shows the spatial distribution of POCs for the three regions with the average MODIS stratocumulus amount for 2003 to 2011 overlaid. It is interesting to note that in the Californian and Namibian cases there appears to be higher densities of POCs further from the coast, slightly offset from the peak Sc densities. This could be a consequence of the deepening of the boundary layer with increasing sea surface temperature away from coast, which favours precipitation and stratocumulus breakup. Off the coast of Namibia, the pattern of POC formation is quite distinct, with a clear hotspot that lies away from the centre of the climatological stratocumulus distribution. This matches the distribution recently demonstrated by Abel et al. (2019), which they attribute to the influence of biomass burning aerosol on the open- to closed-cell transition. In the Peruvian case the POC density broadly coincides with the main stratocumulus deck.

Considering the properties of the POCs themselves, we find that the MODIS cloud fraction is lower and has greater variability within the POC, which is to be expected, since cloud fraction is the defining characteristic of a POC. We find mean cloud fractions inside the POC of ∼70-80%, which are in excellent agreement with detailed in-situ studies Terai et al. (2014) which report values of 56-83% .

Figure 4 shows the retrieved mean cloud properties as a function of distance from the POC boundary. There are some key features present in these results. Firstly, the increase in cloud droplet effective radius (4a) and reduction in cloud optical thickness (4b) inside the POCs, which is also in good agreement with previous findings and has been attributed to a propensity for POCs to form precipitation (Stevens et al., 2005). This is supported by a clear increase in Rain Water Path (RWP) as observed by CloudSat within the POCs (4f). Interestingly, the effective radius transitions more slowly than the cloud optical thickness. Since the MODIS retrieved LWP is directly proportional to the effective radius multiplied by the optical thickness this leads to a dip in the LWP (4b) at the edge of the POCs. While the CloudSat retrievals are unable to resolve this dip, they do show a broadly constant LWP inside and outside the POC, corroborating the relationship of effective radius and optical thickness seen in MODIS (despite a constant bias in the respective retrievals). The AMSR LWP is only shown outside the POC, since the large footprint is unable to resolve the open-cellular structure, and suggests a slightly higher LWP than CloudSat.

While the MODIS cloud top height (CTH) retrieval suffers from artifacts due to the broken clouds within the POCs (not shown), Figure 4e shows the average cloud base height (CBH) and CTH estimated from the CALIOP-CloudSat retrievals. These show a robust decrease in CBH of around 100m inside the POCs as compared to the surrounding closed cell stratocumulus, again in good agreement with in-situ observations which show a decrease of around 50m Terai et al. (2014). The average CTH within the POC also shows a decrease of a similar magnitude, although this disappears when the MODIS cloud mask is applied to the CALIOP-CloudSat retrieval suggesting that the decrease in CTH is primarily caused by optically thinner veil clouds which have recently been associated with the breakup of stratocumulus (O et al., 2018).

One property we can discern directly from the POC masks, and for the first time robustly quantify, is the spatial area of the POCs. The probability distributions shown in Fig. 4d show a clear bi-modal, log-normal distribution. The peak in the smaller mode occurs at $\sim 350\text{km}^2$, while the larger distribution (containing most of the POCs) peaks at $\sim 6750\text{km}^2$. It is not clear why POCs should form in these two distinct size regimes, however it is possible that this is a consequence of growth dynamics of POCs. Figure A1 shows the average size of each POC detected in an image - ordered by its size within that image and demonstrates that when many POCs are present in an image they tend to smaller, while larger POCs tend to appear alone. We hypothesize that the smaller POCs (the first mode in Fig. 4d) represent an initial population of individual POCs while the larger (second mode of) POCs is primarily the result of merging of these smaller POCs. This is corroborated by Fig. A2, which shows the growth of the maximum and total POC size in a given region where POCs were detected in consecutive days. While the total number of POCs present over more than a few consecutive days decreases very rapidly, on average, the POCs in a given region appear to grow over a period of several days.

Finally, and perhaps most interestingly, these properties all seem to have very little dependence on the region in which they are found, implying that a POC is a universal phenomenon whose formation and properties do not depend on its location or the differences in their large scale environment.

Combining these derived properties with the density distributions, the effect of POC formation on the radiative properties of the stratocumulus decks can be estimated. The albedo of the cloud was calculated using the Eddington approximation (Christensen &

	Inside POC	Outside POC
Cloud Fraction (f_c)	0.8086 ± 0.1285	0.9573 ± 0.05398
Cloud Optical Thickness (τ_c)	7.996 ± 3.25	13.58 ± 3.55
Cloud albedo (α_c)	0.375	0.51
Scene albedo (α)	0.33	0.49

Table 1. Cloud properties and albedo inside and outside POCs.

Stephens, 2011; Stephens, 1994), and then combined with the cloud fraction to give the overall scene albedo of both the regions inside and outside the POCs:

$$\alpha_c = \frac{(1-g)\,\tau_c}{2+(1-g)\,\tau_c}, \tag{2}$$

$$\alpha = f_c\alpha_c + (1-f_c)\,\alpha_o, \tag{3}$$

where α_c is the albedo of the cloud, τ_c is the cloud optical depth, g is the asymmetry parameter, which for cloud droplets is taken to be ∼0.85, α is the average albedo for the scene, f_c is the cloud fraction, and α_o is the albedo of the ocean, which was taken to be 0.15 in this case. The interior quantities were averaged over the region from the centre of the POC to 35 km from the boundary, and the exterior quantities were averaged over the region from 35 km from the boundary to 150 km from the boundary in order to avoid the ∼70 km transition zone, as summarised in Table 1. This leads to an albedo of the interior region of POCs (α_{POC}) of 0.33, whereas the exterior region was calculated to be 0.49 ($\alpha_{\text{closed cell}}$), giving a difference of 0.16.

This difference in albedo can be combined with the average incident solar radiation ($R \approx 350\mathrm{Wm}^{-2}$), and the appropriate cloud amounts to estimate the difference in top-of-atmosphere radiation caused by the presence of POCs using:

$$R_{\mathrm{diff}} = R_{\mathrm{inc}} \times \left(\alpha_{\text{no POCs}} - \alpha_{\text{with POCs}}\right), \tag{4}$$

$$\alpha_{\text{no POCs}} = \left|f_{\mathrm{strat}}\alpha_{\text{closed cell}}\right|, \tag{5}$$

$$\alpha_{\text{with POCs}} = \left|f_{\mathrm{strat}}\left(f_{\mathrm{POC}}\alpha_{\mathrm{POC}} + (1-f_{\mathrm{POC}})\,\alpha_{\text{closed cell}}\right)\right|, \tag{6}$$

where f_{strat} is the MODIS stratocumulus amount, f_{POC} is the density of POCs (both shown in Fig. 3) and the vertical bars represent averages taken over the entire spatial region of available data. This leads to estimates of $\alpha_{\text{with POCs}}$ and $\alpha_{\text{no POCs}}$ of 0.019803 and 0.019843 respectively (a 0.2% change), and an overall radiative effect of $R_{\mathrm{diff}} \approx 0.01\mathrm{Wm}^{-2}$. The locally averaged albedo changes are similar for Californian and Namibian POCs at 0.3%, while it is around twice as large in the Peruvian POCs at 0.6% driven by the higher rate of occurrence there. Using ISCCP climatological stratocumulus amount (Young et al., 2018), or surface derived values (Hahn & Warren, 2008) results in identical values to within the accuracy reported here. This small value reflects the relatively low spatial density of POCs (c.f. Fig. 3) and suggests that any change of POC amount via anthropogenic activity (c.f. Rosenfeld et al. (2006)) would not have a large effect on the Earth's radiation balance. It is possible that the observed POC occurrence is already affected by anthropogenic aerosol, however the similarity in micro-physical properties and rates of occurrence across the different regions would seem to suggest this is unlikely.

4 Conclusion

We have created a large-scale database of all POCs present in three of the main stratocumulus decks over the last 13 years (nearly 8,500) and have analysed their spa-

tial and temporal distributions. The properties of the POCs themselves were also studied providing significant evidence in support of previous observations of their properties, including a 33% increase in effective radius and a 20% reduction cloud fraction compared to the surrounding cloud. The LWP shows a marked reduction at the boundary of the POCs, which along with a corresponding increase in rain water path within the POCs, reinforces the role of precipitation in POC formation. Finally, the properties and climatology of POCs were combined to obtain an estimated radiative effect of $0.01\,\mathrm{W/m^2}$, indicating that closing all the POCs in the atmosphere may not have as big an impact as previously postulated (Rosenfeld et al., 2006). These are nevertheless interesting phenomena due to their relevance for stratocumulus to cumulus transition (with potentially much larger effects) and future work tracking POC development and dissipation in geostationary satellite imagery should shed light on these mechanisms. This will also enable an in-depth analysis of their growth dynamics, including testing our hypothesis of the possible role of merging in forming large POCs. The hand-logged training database and automatically detected POCs are made freely available to the community for further analysis.

Appendix A Supplementary Figures

Acknowledgments
We gratefully acknowledge the support of Amazon Web Services through an AWS Machine Learning Research Award. We also acknowledge the support of NVIDIA Corporation with the donation of a Titan Xp GPU used for this research. DWP and PS acknowledge funding from the Natural Environment Research Council projects NE/L01355X/1 (CLARIFY) and NE/S005390/1 (ACRUISE). PS and MC acknowledge funding from the European Research Council project RECAP under the European Union's Horizon 2020 research and innovation programme with grant agreement 724602.

The model, including the hand-labeled masks and imagery used for training is available here: `https://github.com/climate-processes/poc-detection`. Our pre-generated POC database is freely available through the JASMIN data portal.

References

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., ... Zheng, X. (2015). *TensorFlow: Large-scale machine learning on heterogeneous systems.* `https://www.tensorflow.org/`. (Software available from tensorflow.org)

Abel, S. J., Barrett, P. A., Zuidema, P., Zhang, J., Christensen, M., Peers, F., ... Flynn, M. (2019). Open cells can decrease the mixing of free-tropospheric biomass burning aerosol into the south-east atlantic boundary layer. *Atmospheric Chemistry and Physics Discussions*, *2019*, 1–44. Retrieved from `https://www.atmos-chem-phys-discuss.net/acp-2019-738/` doi: 10.5194/acp-2019-738

Agee, E. M., Chen, T. S., & Dowell, K. E. (1973). a review of mesoscale cellular convection. *Bulletin of the American Meteorological Society*, *54*(10), 1004-1012. Retrieved from `https://doi.org/10.1175/1520-0477(1973)054<1004:AROMCC>2.0.CO;2` doi: 10.1175/1520-0477(1973)054⟨1004:AROMCC⟩2.0.CO;2

Berner, A. H., Bretherton, C. S., Wood, R., & Muhlbauer, A. (2013a). Marine boundary layer cloud regimes and poc formation in a crm coupled to a bulk aerosol scheme. *Atmospheric Chemistry and Physics*, *13*, 12549-12572. Retrieved from `https://doi.org/10.5194/acp-13-12549-2013` doi: 10.5194/acp-13-12549-2013

Berner, A. H., Bretherton, C. S., Wood, R., & Muhlbauer, A. (2013b). Marine

boundary layer cloud regimes and poc formation in a crm coupled to a bulk aerosol scheme. *Atmospheric Chemistry and Physics*, *13*(24), 12549–12572. Retrieved from https://www.atmos-chem-phys.net/13/12549/2013/ doi: 10.5194/acp-13-12549-2013

Bradski, G. (2000). The OpenCV Library. *Dr. Dobb's Journal of Software Tools*.

Bretherton, C. S., Uttal, T., Fairall, C. W., Yuter, S. E., Weller, R. A., Baumgardner, D., ... Raga, G. B. (2004). The epic 2001 stratocumulus study. *Bulletin of the American Meteorological Society*, *85*(7), 967-978. Retrieved from https://doi.org/10.1175/BAMS-85-7-967 doi: 10.1175/BAMS-85-7-967

Chang, P., Jelenak, Z., & Alsweiss, S. (2015). *Algorithm theoretical basis document: Gcom-w1/amsr2 day-1 edr version 1.0.* Retrieved from https://suzaku.eorc.jaxa.jp/GCOM_W/data/data_w_product-1.html (accessed June, 2020)

Chollet, F., et al. (2015). *Keras.* https://keras.io.

Christensen, M. W., & Stephens, G. L. (2011). Microphysical and macrophysical responses of marine stratocumulus polluted by underlying ships: Evidence of cloud deepening. *Journal of Geophysical Research: Atmospheres*, *116*(D3). Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2010JD014638 doi: 10.1029/2010JD014638

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., & Fei-Fei, L. (2009). ImageNet: A Large-Scale Hierarchical Image Database. In *Cvpr09.*

Hahn, C., & Warren, S. (2008). *A gridded climatology of clouds over land (1971-96) and ocean (1954-97) from surface observations worldwide.* Numeric Data Package NDP-026E, CDIAC, Department of Energy, Oak Ridge, Tennessee. Retrieved from https://atmos.uw.edu/CloudMap/ (Accessed December 2020)

Hansen, J., Lacis, A., Rind, D., Russell, G., Stone, P., Fung, I., ... Lerner, J. (2013). Climate sensitivity: Analysis of feedback mechanisms. In *Climate processes and climate sensitivity* (p. 130-163). American Geophysical Union (AGU). Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/GM029p0130 doi: 10.1029/GM029p0130

He, K., Zhang, X., Ren, S., & Sun, J. (2015). Deep residual learning for image recognition. *CoRR*, *abs/1512.03385*. Retrieved from http://arxiv.org/abs/1512.03385

Hubanks, P. A., King, M. D., Platnick, S., & Pincus, R. (2008). *Modis atmosphere l3 gridded product. nasa algorithm theoretical basis doc. atbd-mod-30.*

Kazil, J., Wang, H., Feingold, G., Clarke, A. D., Snider, J. R., & Bandy, A. R. (2011). Modeling chemical and aerosol processes in the transition from closed to open cells during vocals-rex. *Atmospheric Chemistry and Physics*, *11*(15), 7491–7514. Retrieved from https://www.atmos-chem-phys.net/11/7491/2011/ doi: 10.5194/acp-11-7491-2011

Klein, S. A., & Hartmann, D. L. (1993). The seasonal cycle of low stratiform clouds. *Journal of Climate*, *6*(8), 1587-1606. Retrieved from https://doi.org/10.1175/1520-0442(1993)006<1587:TSCOLS>2.0.CO;2 doi: 10.1175/1520-0442(1993)006⟨1587:TSCOLS⟩2.0.CO;2

Koren, I., & Feingold, G. (2011, 7). Aerosolcloudprecipitation system as a predator-prey problem. *Proceedings of the National Academy of Sciences*, *108*(30), 12227–12232. doi: 10.1073/pnas.1101777108

Lebsock, M. D., & L'Ecuyer, T. S. (2011). The retrieval of warm rain from cloudsat. *Journal of Geophysical Research: Atmospheres*, *116*(D20). Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2011JD016076 doi: https://doi.org/10.1029/2011JD016076

L. Olson, D., & Delen, D. (2008). *Advanced data mining techniques* (1st ed.). Springer-Verlag Berlin Heidelberg. (pp. 138) doi: 10.1007/978-3-540-76917-0

Mace, G. G., & Zhang, Q. (2014). The cloudsat radar-lidar geometrical profile product (rl-geoprof): Updates, improvements, and selected results. *Journal of Geo-*

physical Research: Atmospheres, *119*(15), 9441-9462. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2013JD021374 doi: https://doi.org/10.1002/2013JD021374

McCoy, I. L., Wood, R., & Fletcher, J. K. (2017). Identifying meteorological controls on open and closed mesoscale cellular convection associated with marine cold air outbreaks. *Journal of Geophysical Research: Atmospheres*, *122*(21), 11,678-11,702. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2017JD027031 doi: https://doi.org/10.1002/2017JD027031

Min, Q., Joseph, E., Lin, Y., Min, L., Yin, B., Daum, P. H., ... Lee, Y.-N. (2012). Comparison of modis cloud microphysical properties with in-situ measurements over the southeast pacific. *Atmospheric Chemistry and Physics*, *12*(23), 11261–11273. Retrieved from https://acp.copernicus.org/articles/12/11261/2012/ doi: 10.5194/acp-12-11261-2012

MODIS Science Team. (2015). *Mod021km modis/terra calibrated radiances 5-min l1b swath 1km.* Level 1 and Atmosphere Archive and Distribution System (LAADS). (http://modaps.nascom.nasa.gov/services/about/products/c6/MOD021KM.html) doi: 10.5067/modis/mod021km.006

O, K.-T., Wood, R., & Tseng, H.-H. (2018). Deeper, precipitating pbls associated with optically thin veil clouds in the sc-cu transition. *Geophysical Research Letters*, *45*(10), 5177-5184. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2018GL077084 doi: https://doi.org/10.1029/2018GL077084

Painemal, D., & Zuidema, P. (2011). Assessment of modis cloud effective radius and optical thickness retrievals over the southeast pacific with vocals-rex in situ measurements. *Journal of Geophysical Research: Atmospheres*, *116*(D24). Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2011JD016155 doi: https://doi.org/10.1029/2011JD016155

Platnick, S., Ackerman, S. A., King, M. D., Meyer, K., Menzel, W. P., Holz, R. E., ... Yang, P. (2015). *Modis atmosphere l2 cloud product (06_l2).* NASA MODIS Adaptive Processing System, Goddard Space Flight Center. Retrieved from http://modaps.nascom.nasa.gov/services/about/products/c6/MOD06_L2.html doi: 10.5067/modis/mod06_l2.006

Randall, D. A., Coakley, J. A., Fairall, C. W., Kropfli, R. A., & Lenschow, D. H. (1984). Outlook for research on subtropical marine stratiform clouds. *Bulletin of the American Meteorological Society*, *65*(12), 1290-1301. Retrieved from https://doi.org/10.1175/1520-0477(1984)065<1290:OFROSM>2.0.CO;2 doi: 10.1175/1520-0477(1984)065⟨1290:OFROSM⟩2.0.CO;2

Raspaud, M., Hoese, D., Dybbroe, A., Lahtinen, P., Devasthale, A., Itkin, M., ... Thorsteinsson, H. (2018, July). PyTroll: An open-source, community-driven python framework to process earth observation satellite data. *Bulletin of the American Meteorological Society*, *99*(7), 1329–1336. Retrieved from https://doi.org/10.1175/bams-d-17-0277.1 doi: 10.1175/bams-d-17-0277.1

Rosenfeld, D., Kaufman, Y. J., & Koren, I. (2006). Switching cloud cover and dynamical regimes from open to closed benard cells in response to the suppression of precipitation by aerosols. *Atmospheric Chemistry and Physics*, *6*, 2503-2511. Retrieved from https://doi.org/10.5194/acp-6-2503-2006 doi: 10.5194/acp-6-2503-2006

Stephens, G. (1994). *Remote sensing of the lower atmosphere: An introduction.* Oxford University Press. Retrieved from https://books.google.co.uk/books?id=2FcRAQAAIAAJ

Stevens, B., Vali, G., Comstock, K., Wood, R., van Zanten, M. C., Austin, P. H., ... Lenschow, D. H. (2005). Pockets of open cells and drizzle in marine stratocumulus. *Bulletin of the American Meteorological Society*, *86*(1), 51-58. Retrieved from https://doi.org/10.1175/BAMS-86-1-51 doi: 10.1175/BAMS-86-1-51

Terai, C. R., Bretherton, C. S., Wood, R., & Painter, G. (2014). Aircraft observations of aerosol, cloud, precipitation, and boundary layer properties in pockets of open cells over the southeast pacific. *Atmospheric Chemistry and Physics*, *14*, 8071-8088. Retrieved from https://doi.org/10.5194/acp-14-8071-2014 doi: 10.5194/acp-14-8071-2014,2014

Wang, H., & Feingold, G. (2009). Modeling mesoscale cellular structures and drizzle in marine stratocumulus. part i: Impact of drizzle on the formation and evolution of open cells. *Journal of the Atmospheric Sciences*, *66*(11), 3237-3256. Retrieved from https://doi.org/10.1175/2009JAS3022.1 doi: 10.1175/2009JAS3022.1

Wentz, F. J., & Meissner, T. (2004). *Amsr-e/aqua l2b global swath ocean products derived from wentz algorithm, version 2. [AE-Ocean LWP].* (accessed June, 2020) doi: https://doi.org/10.1002/2017JD027031

Wood, R., Bretherton, C. S., Leon, D., Clarke, A. D., Zuidema, P., Allen, G., & Coe, H. (2011). An aircraft case study of the spatial transition from closed to open mesoscale cellular convection over the southeast pacific. *Atmospheric Chemistry and Physics*, *11*, 2341-2370. Retrieved from https://doi.org/10.5194/acp-11-2341-2011 doi: 10.5194/acp-11-2341-2011

Wood, R., & Hartmann, D. L. (01 May. 2006). Spatial variability of liquid water path in marine low cloud: The importance of mesoscale cellular convection. *Journal of Climate*, *19*(9), 1748 - 1764. Retrieved from https://journals.ametsoc.org/view/journals/clim/19/9/jcli3702.1.xml doi: 10.1175/JCLI3702.1

Young, A. H., Knapp, K. R., Inamdar, A., Hankins, W., & Rossow, W. B. (2018). The international satellite cloud climatology project h-series climate data record product. *Earth System Science Data*, *10*(1), 583–593. Retrieved from https://www.earth-syst-sci-data.net/10/583/2018/ doi: 10.5194/essd-10-583-2018

Zhang, Z., Liu, Q., & Wang, Y. (2017). Road extraction by deep residual u-net. *CoRR*, *abs/1711.10684*. Retrieved from http://arxiv.org/abs/1711.10684

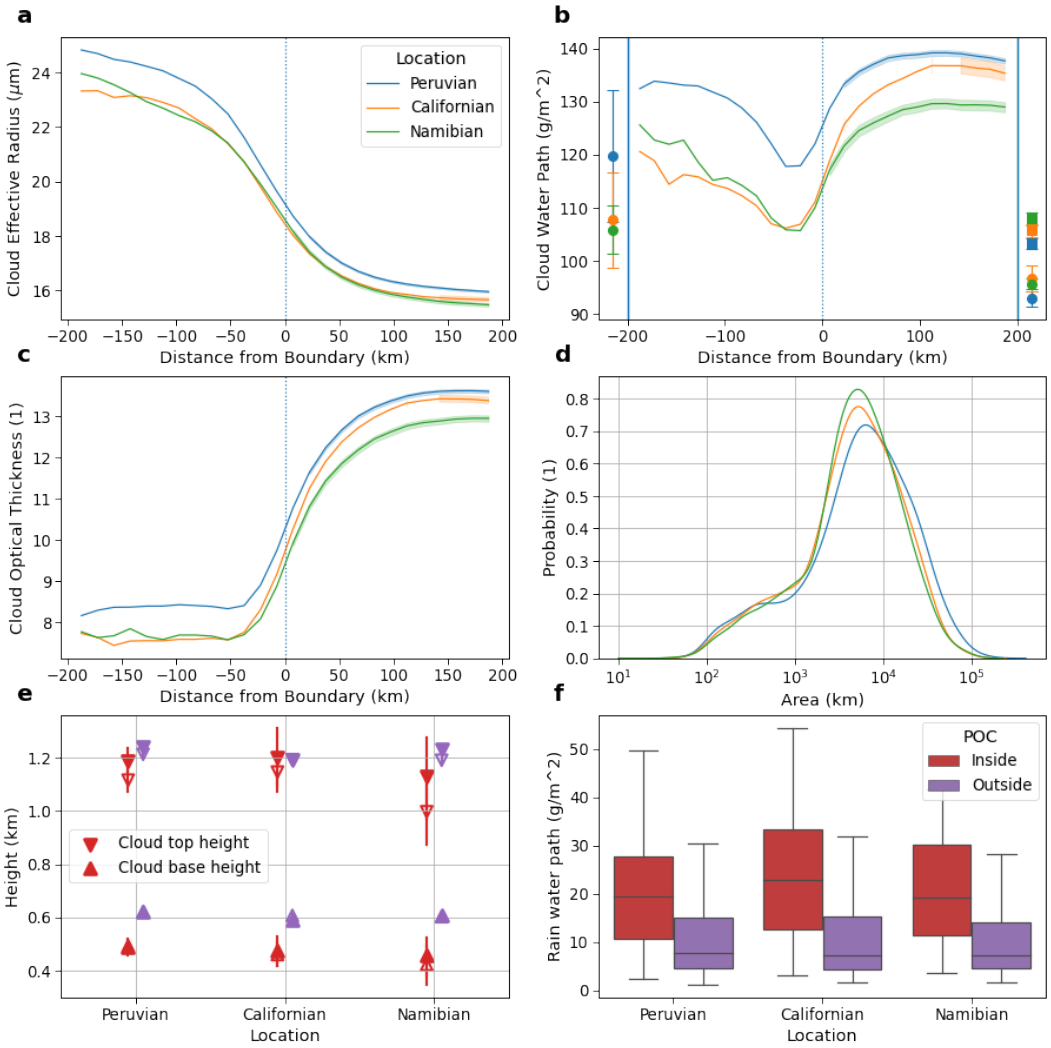

Figure 4. Retrieved cloud properties as a function of distance from the boundary of POCs: (a) Cloud effective radius as retrieved from MODIS; (b) Cloud liquid water path as retrieved from MODIS (solid lines) and CloudSat (circles) and AMSR-E (squares) shown as averages inside and outside the POCs; (c) Cloud optical thickness as retrieved from MODIS. The vertical line at 0 is the boundary between the POC and the surrounding cloud, with negative values on the x axis being inside the POC and positive values being outside. d) shows the probability distribution of the areas of the POCs in each region. The different colours represent different geographical regions, and the shaded regions represent the standard error of the data. e) CALIOP-CloudSat retrieved cloud base and top height for each region with CALIOP cloud mask (hollow) and MODIS cloud mask (solid). f) Rain water path retrieved from CloudSat for each region both inside and outside the POCs.

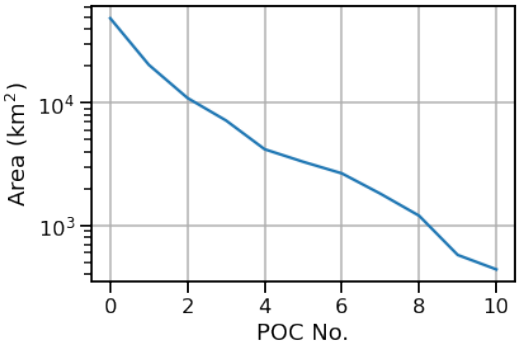

Figure A1. The mean POC area ordered by the largest to smallest POC in each image.

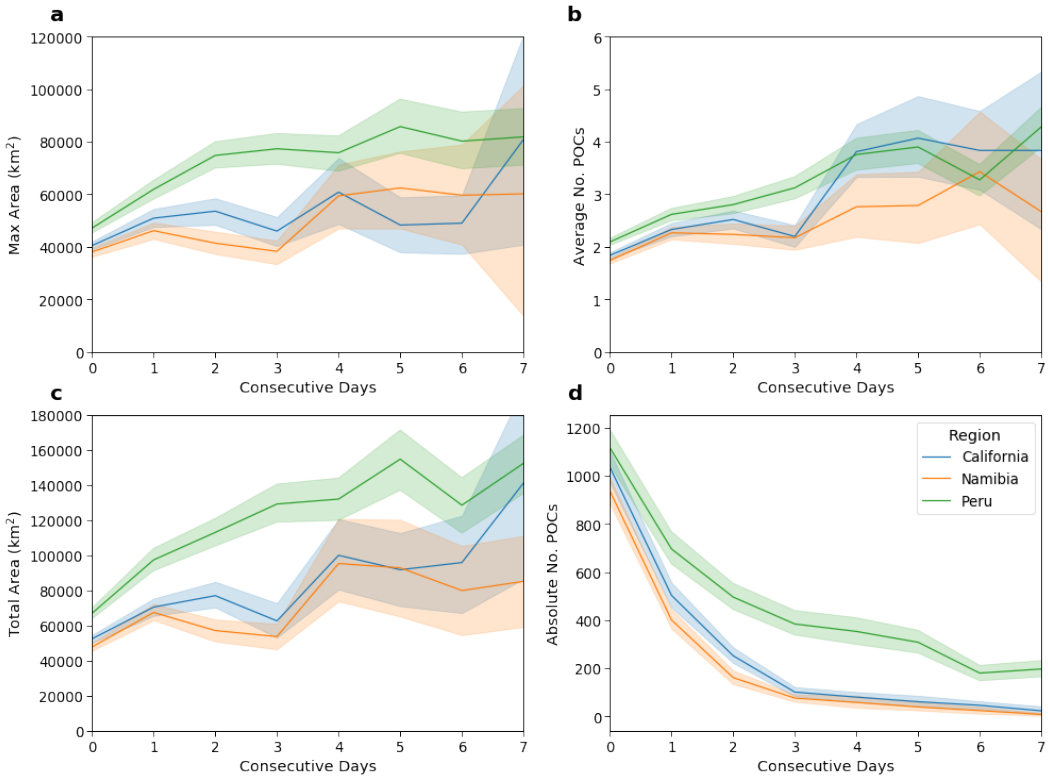

Figure A2. POC properties as a function of consecutive days that a POC was detected in that region: (a) the average maximum area of the POCs in a given scene; (b) average number of POCs in each scene; (c) average total area of POCs in each scene; (d) absolute number of POCs across all scenes for each consecutive day bin.