



Machine learning reveals climate forcing from aerosols is dominated by increased cloud cover

Ying Chen ^{1,11}✉, Jim Haywood ^{1,2}, Yu Wang ³, Florent Malavelle ⁴, George Jordan ², Daniel Partridge ¹, Jonathan Fieldsend ¹, Johannes De Leeuw ⁵, Anja Schmidt ^{5,6,7,12}, Nayeong Cho ⁸, Lazaros Oreopoulos ⁸, Steven Platnick ⁸, Daniel Grosvenor ⁹, Paul Field ^{4,10} and Ulrike Lohmann ³

Aerosol–cloud interactions have a potentially large impact on climate but are poorly quantified and thus contribute a substantial and long-standing uncertainty in climate projections. The impacts derived from climate models are poorly constrained by observations because retrieving robust large-scale signals of aerosol–cloud interactions is frequently hampered by the considerable noise associated with meteorological co-variability. The 2014 Holuhraun effusive eruption in Iceland resulted in a massive aerosol plume in an otherwise near-pristine environment and thus provided an ideal natural experiment to quantify cloud responses to aerosol perturbations. Here we disentangle significant signals from the noise of meteorological co-variability using a satellite-based machine-learning approach. Our analysis shows that aerosols from the eruption increased cloud cover by approximately 10%, and this appears to be the leading cause of climate forcing, rather than cloud brightening as previously thought. We find that volcanic aerosols do brighten clouds by reducing droplet size, but this has a notably smaller radiative impact than changes in cloud fraction. These results add substantial observational constraints on the cooling impact of aerosols. Such constraints are critical for improving climate models, which still inadequately represent the complex macro-physical and microphysical impacts of aerosol–cloud interactions.

Marine low-level liquid clouds have a profound impact on the energy balance of the Earth system, exerting a net cooling effect by reflecting sunlight^{1,2}. It has been previously estimated that only a 6% increase of their albedo could offset the warming from a doubling of CO₂^{3,4}. Aerosol–cloud interactions (ACIs) are postulated to enhance albedo and prolong the lifetime of liquid clouds^{5,6} and therefore counterbalance a substantial, yet poorly constrained, portion of greenhouse gas warming^{7–10}, leading to only a small net positive overall forcing. As the Earth has warmed by around 1.2 °C since pre-industrial times^{10,11}, this would imply that the Earth system is highly sensitive, and therefore vulnerable, to anthropogenic climate forcing¹². Such a high sensitivity would suggest a very limited remaining carbon budget if the +1.5 °C target of the 21st Conference of the Parties at Paris is to be met¹¹.

Despite decades of effort, ACIs still contribute substantially to uncertainties in climate projections^{1,7,9–11}. A primary reason for the large uncertainty in ACIs is the lack of suitable large-scale constraints to challenge general circulation models (GCMs)^{13–15}. ACI operates through processes whereby cloud droplets form on aerosol particles. For a fixed cloud liquid-water path (LWP), high concentrations of aerosol lead to more droplets with smaller effective radius (r_{eff} Twomey r_{eff} effect⁵), which increases cloud albedo. Smaller cloud droplets may inhibit precipitation due to weakened collision coalescence⁶, and suppressed precipitation implies clouds retain more water, leading to an increased LWP (LWP adjustment), and prolong their lifetime and areal extent, which manifests as

increased cloud fraction (CF, CF adjustment)⁶. There is clear evidence of the Twomey r_{eff} effect from numerous comprehensive satellite observations (for example, refs. ^{8,15–19}), but continuous debate surrounds the LWP adjustment with different magnitudes and signs reported^{8,9,15,20,21}, possibly due to confounding adjustments such as effects of entrainment and droplet evaporation processes^{22–26}. The CF adjustment is even more difficult to constrain owing to the large-scale impacts of meteorological co-variability²⁷, leading to long-standing and ongoing disputes in the scientific literature^{16,19,28–32}. Satellite observational constraints of ACI tend to be limited to either small-scale observations or large-scale climatological analyses³³. A typical example of a small-scale observation is ‘ship-tracks’, manifesting as brighter lines in stratocumulus cloud decks caused by ship emissions. Such small tracks are generally able to rule out confounding meteorology^{8,19}, but with a scale far below the resolution of GCMs and a short temporal signature, they are therefore not ideal constraints for these models^{33,34}. Climatological analyses examine the correlations between cloud properties and aerosol on a large spatiotemporal scale, but such correlations can be confounded by meteorological co-variability and therefore may not confirm the causal processes of ACI^{29,33,35–38}.

In this Article, we overcome these limitations by developing a meteorological reanalysis and satellite-based machine-learning approach that predicts cloud properties in a near-pristine environment, and compare the results with observations of clouds perturbed by the large-scale effusive Icelandic eruption of Holuhraun.

¹College of Engineering, Mathematics, and Physical Sciences, University of Exeter, Exeter, UK. ²Met Office Hadley Centre, Exeter, UK. ³Institute for Atmospheric and Climate Science, ETH Zurich, Zurich, Switzerland. ⁴Met Office, Exeter, UK. ⁵Centre for Atmospheric Science, Yusuf Hamied Department of Chemistry, University of Cambridge, Cambridge, UK. ⁶Department of Geography, University of Cambridge, Cambridge, UK. ⁷Meteorological Institute, Ludwig Maximilian University of Munich, Munich, Germany. ⁸Earth Sciences Division, NASA GSFC, Greenbelt, MD, USA. ⁹National Centre for Atmospheric Sciences, University of Leeds, Leeds, UK. ¹⁰School of Earth and Environment, University of Leeds, Leeds, UK. ¹¹Present address: Laboratory of Atmospheric Chemistry, Paul Scherrer Institut, Villigen, Switzerland. ¹²Present address: Institute of Atmospheric Physics (IPA), German Aerospace Center (DLR), Oberpfaffenhofen, Germany. ✉e-mail: y.chen6@exeter.ac.uk

The machine-learning approach is enabled by an almost threefold expansion of satellite data from Moderate Resolution Imaging Spectroradiometer (MODIS) compared with the earlier work¹⁵, thus offering a robust training dataset. The machine-learning approach allows us to quantify ACI-induced cloud responses (aerosol's fingerprint in cloud) and show an unmistakable increase in cloud cover. It also allows us to infer the relative contributions to ACI radiative effect from the Twomey effect, and the LWP and CF adjustments. Our results improve current understanding of cloud-induced climate change and provide robust large-scale constraints for climate models.

Volcanic aerosol perturbation

The effusive volcanic eruption at Holuhraun in Iceland emitted about 40,000 tonnes of SO₂ per day on average during its eruptive phase in September–October 2014 and 120,000 tonnes per day at the peak of eruption^{15,39}. The sulfate aerosol formed from volcanic SO₂ interacting with liquid-water clouds creating an invaluable natural experiment for testing ACI hypotheses at a large scale¹⁵. Detecting CF changes above meteorological noise requires a larger data volume and was left unexplored in the previous study¹⁵, which uses the MODIS Aqua 2002–2014 dataset. Here, by extending the satellite data to both MODIS Aqua and Terra and the length of the analysis period to 2001–2020, we have sufficient training data to develop a robust machine-learning approach for quantitatively disentangling Holuhraun eruption ACI signals from the noise of meteorological co-variability (Methods). We focus primarily on October 2014 because, in this second month of the eruption, the volcanic plume dispersed sufficiently across the entire region of about 3,000 km × 6,000 km (~45–75°N; ~60°W–30°E; see supplementary fig. S6.2 in ref. ¹⁵). This region is an otherwise near-pristine environment and encompasses the whole spectrum of liquid-dominated cloud regimes, with their frequencies of occurrence being comparable to those observed globally (Extended Data Fig. 1)^{15,40}.

To disentangle the ACI signal from the noise of meteorological co-variability, we train a machine-learning surrogate MODIS (ML-MODIS) using historical meteorology and MODIS observations during 2001–2020 but excluding the year of the volcanic perturbation (2014; Methods). ML-MODIS is designed to predict cloud properties for given meteorological conditions when unperturbed by volcanic aerosol. Our ‘leave-one-year-out’ cross validation (Methods) shows that the surrogate ML-MODIS can reproduce the MODIS observations of cloud droplet number concentration (N_d), r_{eff} , LWP and CF when no volcanic aerosol perturbation exists, as shown in the left column of Fig. 1. However, differences between the ML-MODIS predictions and MODIS observations are evident in the presence of the volcanic perturbation in October 2014 (right column of Fig. 1). Similar results are found for September 2014 (Supplementary Discussion section 1).

We examine the ACI corresponding to the increase in N_d instead of aerosol optical depth (AOD) because MODIS aerosol products are hampered by the overcast nature of the geographical region, and using N_d has several advantages as a mediating variable²⁹. We first quantify the increase in N_d and then estimate the susceptibility of other cloud properties: $\text{dln}r_{\text{eff}}/\text{dln}N_d$, $\text{dlnLWP}/\text{dln}N_d$ and $\text{dlnCF}/\text{dln}N_d$. The volcano-induced increase in N_d is observed across nearly the entire region, with a positive signal across the zonal means (Fig. 2a). We also observe a clear shift of the N_d probability distribution towards larger values due to the volcanic perturbation, with an average increase of 20 cm⁻³.

We perform Monte Carlo analyses (Methods) to estimate the uncertainty of ML-MODIS and to quantify the impact of ACI on relevant cloud properties. In assessing the statistical uncertainties, we follow the Intergovernmental Panel on Climate Change uncertainty guideline⁴¹ and use the 90% probabilities (which are assigned ‘very likely’). A validation of ML-MODIS by MODIS for conditions

unperturbed by Holuhraun is further achieved by these results, with median and average values close to the 1:1 line (Fig. 3) and with a 90% probability of the Pearson correlation coefficients (R) exceeding 0.6 for N_d , r_{eff} and CF (Extended Data Fig. 2; higher than 0.5 for LWP). By contrast, the 90% probability of R being below 0.6 for all cloud properties in volcano-perturbed conditions indicates large influences of the volcanic aerosol on cloud properties. We estimate a volcanic aerosol-induced increase in N_d of 28% over the region (Fig. 3, showing that the ratio between ML-MODIS and MODIS is 1.27 with volcano versus 0.99 without volcano), which is clearly statistically significant because the perturbation lies outside the range of uncertainty of the machine-learning method. This increase is similar to the ~32% increase in N_d from pre-industrial to present day according to multimodel estimates¹⁴, suggesting that the results from our analysis may be a reasonable proxy for anthropogenic aerosols in terms of the strength in perturbing clouds since pre-industrial times.

Twomey effect and liquid water path adjustment

We first use our machine-learning approach to examine the Twomey r_{eff} effect and LWP adjustment. We observe a consistent spatial pattern of volcano-induced increase in N_d and an average reduction in r_{eff} (Fig. 2a,b) from 15.2 μm to 13.9 μm. The spatial pattern is also consistent with the climatological MODIS anomaly analysis¹⁵ (Extended Data Fig. 3), but with some difference in the strength of ACI signal. This further demonstrates the viability of our machine-learning approach in identifying cloud changes created by volcanic aerosols above those expected due to meteorological variability. Climatological anomalies may identify regions influenced by the Holuhraun plume¹⁵ but may not be robust in quantifying ACI signals arising from Holuhraun because the ACI signal is confounded by meteorology where 2014 conditions are not necessarily equal to climatological average. Indeed, while Malavelle et al.¹⁵ developed a robust method for removing the meteorological variability in the modelled response, they also cautioned that meteorological differences from the long-term mean could cause some of the observed response (their supplementary figs. S6.1 and S6.2). Our machine-learning approach overcomes these issues (Methods, see also Supplementary Discussion sections 1 and 2). We estimate an 8% decrease in r_{eff} as a response to a 28% increase in N_d on average (and median) over the geographical region (Fig. 3). In line with previous studies^{8,17,31}, no notable LWP response is found when examining the region as a whole (Fig. 3 and Extended Data Fig. 4). This may be due to the cancellation of the LWP adjustment-induced increase⁶ by entrainment-induced decrease of LWP²², as suggested by Toll et al.⁸, who examined over 10,000 globally representative aerosol-perturbation tracks of small scale in liquid clouds.

Cloud fraction response

So far, results from our large-scale machine-learning approach agree with previous analyses: a distinct and robust Twomey r_{eff} effect but a weak LWP adjustment (for example, refs. ^{8,17,31}). We now examine the adjustment of liquid-phase CF, which is a macro-property of cloud and difficult to examine using small-scale aerosol-induced tracks⁸. Our results of volcanic aerosol-perturbed conditions show an overall increase of zonal CF at all latitudes of our domain and a clear shift of probability distribution from a median value of 0.36 to 0.39 (Fig. 2c). The CF increase exhibits a spatial pattern that is consistent with the Twomey r_{eff} effect (Fig. 2b,c). This strongly suggests that it is the aerosol perturbation that leads to increased cloud cover since the Twomey r_{eff} effect has been well documented as an ACI indicator^{8,9,15,18}.

We present the response of CF and other cloud properties over the geographical region using the Monte Carlo method (Fig. 3). For all non-perturbed cloud properties, the validation shows the median and average values on the 1:1 line. For volcano-induced changes in

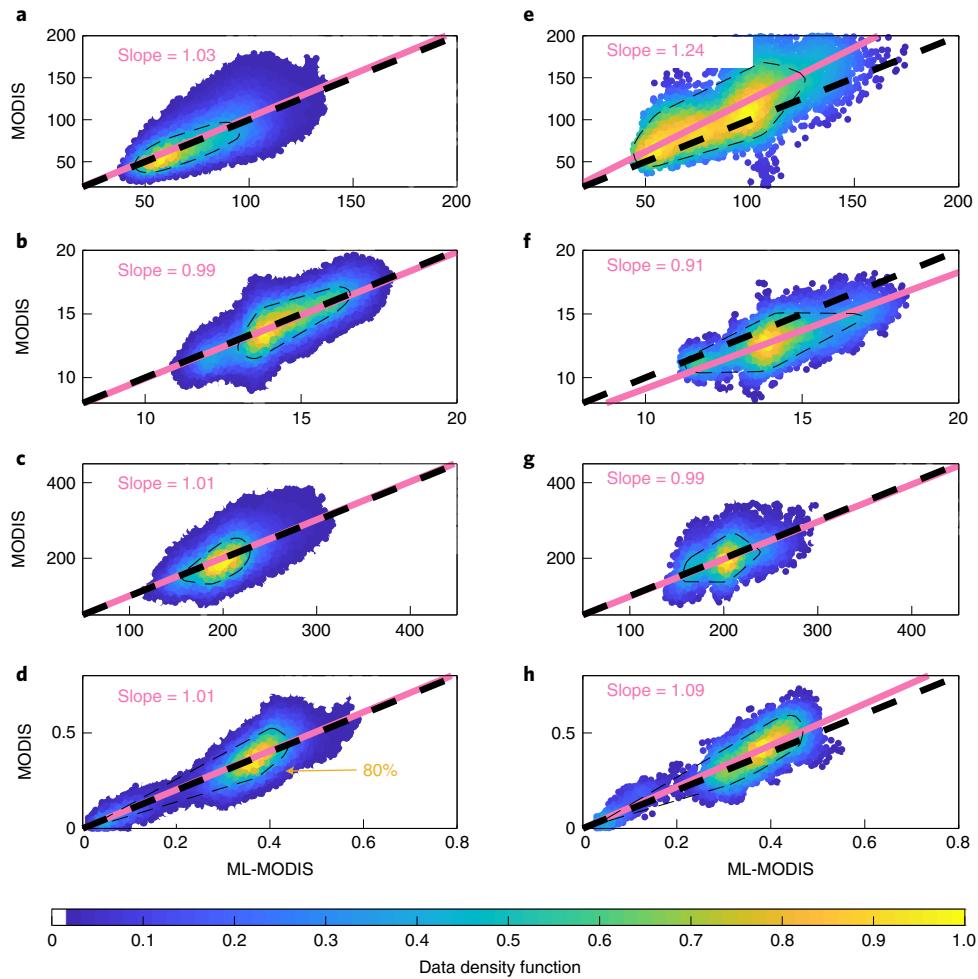


Fig. 1 | Comparison between ML-MODIS predictions and MODIS observations. **a–d**, Validation against non-perturbed observations (excluding 2014) of cloud properties: N_d (cm^{-3}) (**a**), r_{eff} (μm) (**b**), LWP (g m^{-2}) (**c**) and CF (unitless) (**d**). **e–h**, Volcanic perturbation signals in October 2014, indicated by the difference between the machine-learning predictions and the observations: N_d (cm^{-3}) (**e**), r_{eff} (μm) (**f**), LWP (g m^{-2}) (**g**) and CF (unitless) (**h**). October MODIS observations from Aqua (2002–2020) and Terra (2001–2020) are analysed. Colour indicates the normalized data density function with a maximum value of 1.0, with 80% of the data being contained within the black dashed area.

N_d and r_{eff} we confirm the expected increase and decrease, respectively, but see little LWP response. For CF, we observe a statistically significant median (and average) relative increase of 11%, with the signal variability range lying outside the uncertainty. We estimate $\text{dlnCF/dln}N_d = 0.41$ [~ -0.05 –1.53, 90% confidence interval], indicating a strong susceptibility of CF to aerosol-induced perturbation in N_d . Ref. ³⁰ found a similar strong susceptibility using a climatological approach, but for the convective cores of southern ocean liquid clouds. This strong susceptibility is also consistent with other studies (for example refs. ^{16,29,31,36}), although unlike the present study, their results are probably either influenced by the confounding meteorology associated with the climatological correlation approach ^{33,36} or limited by relatively small-scale Lagrangian trajectories ³³. For example, ref. ¹⁴ showed that climatological correlation analysis differs greatly from perturbation analysis across multiple GCMs despite efforts to classify and isolate different meteorological regimes.

To back up our finding of CF increase, we perform a traditional climatological anomaly analysis, which shows a similar spatial pattern for the CF response (Extended Data Fig. 3c). In addition, we investigate the impact of the unusually low sea surface temperature that developed to the south of the region (Extended Data Fig. 5a) owing to factors that appear to be independent from the eruption ⁴². While this could affect CF, it cannot be accounted for in the climatological

anomaly analysis using only MODIS data. Our machine-learning approach, however, accounts for this variability (Extended Data Fig. 6 and Supplementary Discussion section 2). We are therefore in a position to better quantify a weaker CF increase over the corresponding region (~ 45 – 60°N , ~ 20 – 45°W ; compare Fig. 2c with Extended Data Fig. 3c). We also find 14% fewer cloud-free high-resolution (1 km^2) MODIS pixels during October 2014 compared with the long-term October mean. Again, this implies CF increases in response to the volcanic aerosol. Any conceivable increase in cloud cover from ice clouds is also investigated and cannot be discerned (Extended Data Fig. 5b); this suggests that any potential confounding effect from ice cloud or transition to ice cloud is small and that our results regarding ACI of liquid clouds are robust.

Cloud fraction adjustment dominates radiative forcing

We revisit the relative contributions to ACI-induced radiative forcing from the Twomey effect, LWP and CF adjustments (Radiative forcing). In line with previous studies ^{8,31}, we find a weak contribution ($2 \pm 17\%$) from the LWP adjustment. However, in contrast to recent studies reporting that the Twomey r_{eff} effect dominates ($>70\%$) the ACI radiative forcing ^{8,19,31} we show that, for this large-scale study across a wide range of meteorological and cloud regimes, the CF adjustment ($61 \pm 23\%$) surpasses the Twomey r_{eff} effect ($37 \pm 18\%$)

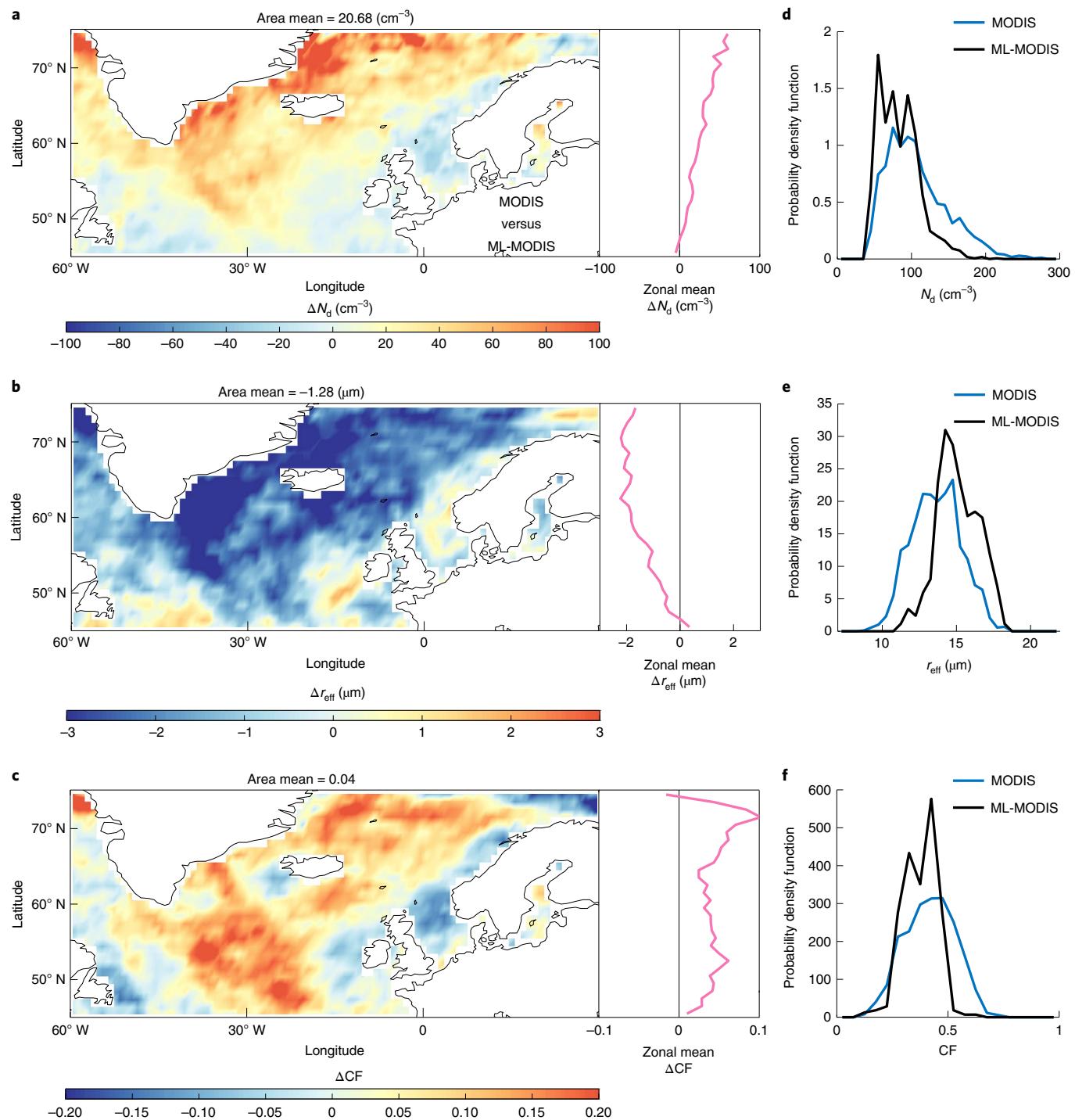


Fig. 2 | Changes in cloud properties caused by the volcanic perturbation estimated using machine-learning predictions and MODIS observations for October 2014. **a–c**, The spatial distribution and zonal means of the changes (Δ) in N_d (**a**), r_{eff} (**b**) and CF (**c**). **d–f**, Probability density functions (so that the areas under the curves are equivalent) for MODIS and ML-MODIS for N_d (**d**), r_{eff} (**e**) and CF (**f**).

in terms of ACI cooling (Fig. 3). This new finding may be due to the much larger spatiotemporal scale of our investigation, which extends up to tens of thousands of kilometres with perturbations lasting for months. Given the large range of meteorological conditions and cloud regimes included (Extended Data Fig. 1), our study appears arguably more suitable for constraining large-scale climate models and ACI associated with anthropogenic emissions, which themselves persist across many geographical areas and are associated with a wide variety of cloud regimes.

Our results suggest that cooling caused by a CF increase is notably underestimated in current climate projections¹⁰. A recent multimodel assessment of the susceptibility of $\text{dlnCF}/\text{dln}N_d$ versus $-\text{dln}r_{\text{eff}}/\text{dln}N_d$ (fig. 1 in ref. ¹⁴) suggests ratios of approximately 1:3. Our results suggest that the CF adjustment is possibly larger than the Twomey r_{eff} effect since the ratio of their susceptibilities is around 5:4. It is possible that GCMs compensate for the lack of CF response with overly strong LWP adjustment^{8,10,15,19,34}; that is, they estimate the ‘right’ cooling but for manifestly the wrong reasons.

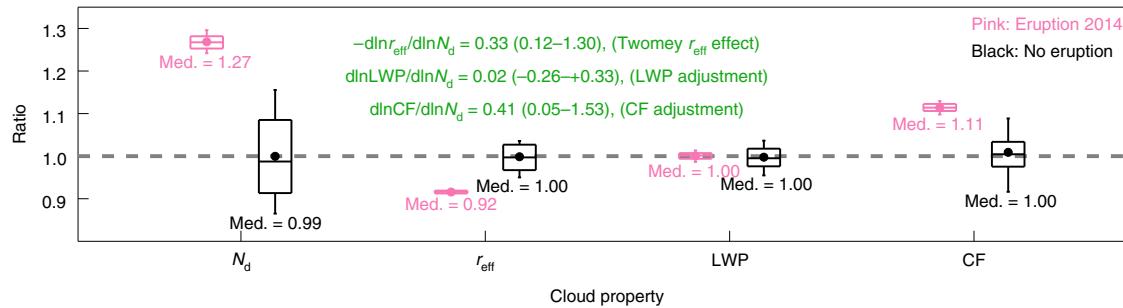


Fig. 3 | Responses of cloud properties to the volcanic aerosol perturbation in October 2014. The ACI signals of responses (fingerprints) are indicated as the ratios between MODIS (Aqua and Terra) observations and machine-learning predictions: ratio = MODIS divided by ML-MODIS. Uncertainties of non-perturbed baseline references are estimated using a Monte Carlo method and are shown in black (Methods; based on non-volcanic October datasets spanning 2001–2020). The variability of the cloud responses to the Holuhraun volcanic aerosol perturbation is shown in pink. The box plots show 10th, 25th, 75th and 90th percentiles and median (Med.); the mean value is indicated by a dot. The susceptibilities of r_{eff} , LWP and CF to changes in N_d are given in a green colour, median [90% confidence interval]. Area (in units of km^2) weighted averaging is used to calculate average cloud properties over the geographical region (Fig. 2), in order to estimate an unbiased large-scale response signal. Therefore, the ratios shown here are slightly different from the slopes shown in Fig. 1, in which area-weighted averaging is not applied.

This work illuminates certain aspects of ACI conventionally thought to follow the following route: an increase in aerosols gives rise to (1) an increase in N_d leading to (2) a larger number of smaller cloud droplets, leading to (3) a decrease in the collision-coalescence growth rate of cloud droplets, leading to (4) a reduction in precipitation, leading to (5) an increase in LWP, leading to (6) an increase in cloud lifetime, leading to (7) an increase in CF. Ref. ¹⁵ suggested that (4) and (5) do not operate as expected, while this new study provides strong evidence for (6) and (7). This conundrum needs to be addressed in further research. Suggestions for how to approach this in future work include performing large eddy model simulation of the Holuhraun event to identify difference in the ACI causal chain between the heavily parameterized GCM representations and the more explicit cloud-resolving models. Identifying any changes in cloud regimes (for example, refs. ^{31,40}) might also provide further clues in solving this puzzle. We maintain that because clouds are such a fundamentally important component of Earth's hydrological cycle and energy flows, the underlying reasons of deficient model performance need to be urgently addressed. Our findings appear to provide robust new constraints for climate models despite the uncertainties associated with machine-learning and MODIS retrievals. We acknowledge that the cold sea surface temperature anomaly in October 2014 could potentially introduce more uncertainty in the machine-learning representation of cloud conditions, but this influence appears insignificant in this study (Supplementary Discussion section 2). ACI signals are statistically significant, lying outside the uncertainty range of the machine-learning approach (Fig. 3). Uncertainty in the MODIS retrievals can be decomposed into systematic errors and random errors. Random errors are greatly suppressed by averaging over a geographical region of thousands of kilometres¹³, while systematic errors are largely cancelled when taking differences between MODIS and ML-MODIS⁸.

The quantified constraints from our machine-learning study pave the way to advance our current understanding of physical ACI processes, and point to new directions and challenges towards future improvement of climate models. With advances in both areas, we expect that our large-scale constraints on ACI will lead to reduced uncertainty in climate projections and future estimates of climate sensitivity.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of

author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41561-022-00991-6>.

Received: 18 October 2021; Accepted: 14 June 2022;

Published online: 1 August 2022

References

1. Lohmann, U. & Feichter, J. Global indirect aerosol effects: a review. *Atmos. Chem. Phys.* **5**, 715–737 (2005).
2. L'Ecuyer, T. S., Hang, Y., Matus, A. V. & Wang, Z. Reassessing the effect of cloud type on Earth's energy balance in the age of active spaceborne observations. Part I: top of atmosphere and surface. *J. Clim.* **32**, 6197–6217 (2019).
3. Latham, J. et al. Global temperature stabilization via controlled albedo enhancement of low-level maritime clouds. *Phil. Trans. R. Soc.* **366**, 3969–3987 (2008).
4. Chen, Y.-C., Christensen, M. W., Stephens, G. L. & Seinfeld, J. H. Satellite-based estimate of global aerosol–cloud radiative forcing by marine warm clouds. *Nat. Geosci.* **7**, 643–646 (2014).
5. Twomey, S. Pollution and the planetary albedo. *Atmos. Environ.* **8**, 1251–1256 (1974).
6. Albrecht, B. A. Aerosols, cloud microphysics, and fractional cloudiness. *Science* **245**, 1227–1230 (1989).
7. Boucher, O. et al. in *Climate Change 2013: The Physical Science Basis* (eds Stocker, T. F. et al.) 571–657 (Cambridge Univ. Press, 2013).
8. Toll, V., Christensen, M., Quaas, J. & Bellouin, N. Weak average liquid-cloud-water response to anthropogenic aerosols. *Nature* **572**, 51–55 (2019).
9. Bellouin, N. et al. Bounding global aerosol radiative forcing of climate change. *Rev. Geophys.* **58**, e2019RG000660 (2020).
10. IPCC *Climate Change 2021: The Physical Science Basis* (eds Masson-Delmotte, V. et al.) (Cambridge Univ. Press, 2021).
11. Rogelj, J., Forster, P. M., Kriegler, E., Smith, C. J. & Séférian, R. Estimating and tracking the remaining carbon budget for stringent climate targets. *Nature* **571**, 335–342 (2019).
12. Andreae, M. O., Jones, C. D. & Cox, P. M. Strong present-day aerosol cooling implies a hot future. *Nature* **435**, 1187–1190 (2005).
13. Seinfeld, J. H. et al. Improving our fundamental understanding of the role of aerosol–cloud interactions in the climate system. *Proc. Natl. Acad. Sci. USA* **113**, 5781–5790 (2016).
14. Ghan, S. et al. Challenges in constraining anthropogenic aerosol effects on cloud radiative forcing using present-day spatiotemporal variability. *Proc. Natl. Acad. Sci. USA* **113**, 5804–5811 (2016).
15. Malavelle, F. F. et al. Strong constraints on aerosol–cloud interactions from volcanic eruptions. *Nature* **546**, 485–491 (2017).
16. Kaufman, Y. J., Koren, I., Remer, L. A., Rosenfeld, D. & Rudich, Y. The effect of smoke, dust, and pollution aerosol on shallow cloud development over the Atlantic Ocean. *Proc. Natl. Acad. Sci. USA* **102**, 11207–11212 (2005).
17. McCoy, D. T. & Hartmann, D. L. Observations of a substantial cloud–aerosol indirect effect during the 2014–2015 Bárðarbunga-Veiðivötn fissure eruption in Iceland. *Geophys. Res. Lett.* **42**, 409–410,414 (2015).

18. Toll, V., Christensen, M., Gassó, S. & Bellouin, N. Volcano and ship tracks indicate excessive aerosol-induced cloud water increases in a climate model. *Geophys. Res. Lett.* **44**, 492–412,500 (2017).
19. Diamond, M. S., Director, H. M., Eastman, R., Possner, A. & Wood, R. Substantial cloud brightening from shipping in subtropical low clouds. *AGU Adv.* **1**, e2019AV000111 (2020).
20. Gryspenrdt, E. et al. Constraining the aerosol influence on cloud liquid water path. *Atmos. Chem. Phys.* **19**, 5331–5347 (2019).
21. Possner, A., Eastman, R., Bender, F. & Glassmeier, F. Deconvolution of boundary layer depth and aerosol constraints on cloud water path in subtropical stratocumulus decks. *Atmos. Chem. Phys.* **20**, 3609–3621 (2020).
22. Ackerman, A. S., Kirkpatrick, M. P., Stevens, D. E. & Toon, O. B. The impact of humidity above stratiform clouds on indirect aerosol climate forcing. *Nature* **432**, 1014–1017 (2004).
23. Stevens, B. & Feingold, G. Untangling aerosol effects on clouds and precipitation in a buffered system. *Nature* **461**, 607–613 (2009).
24. Lebo, Z. J. & Feingold, G. On the relationship between responses in cloud water and precipitation to changes in aerosol. *Atmos. Chem. Phys.* **14**, 11817–11831 (2014).
25. Koren, I., Dagan, G. & Altaratz, O. From aerosol-limited to invigoration of warm convective clouds. *Science* **344**, 1143–1146 (2014).
26. Seifert, A., Heus, T., Pincus, R. & Stevens, B. Large-eddy simulation of the transient and near-equilibrium behavior of precipitating shallow convection. *J. Adv. Modeling Earth Syst.* **7**, 1918–1937 (2015).
27. Mauger, G. S. & Norris, J. R. Meteorological bias in satellite estimates of aerosol–cloud relationships. *Geophys. Res. Lett.* <https://doi.org/10.1029/2007GL029952> (2007).
28. Kaufman, Y. J. & Koren, I. Smoke and pollution aerosol effect on cloud cover. *Science* **313**, 655–658 (2006).
29. Gryspenrdt, E., Quaas, J. & Bellouin, N. Constraining the aerosol influence on cloud fraction. *J. Geophys. Res. Atmos.* **121**, 3566–3583 (2016).
30. Rosenfeld, D. et al. Aerosol-driven droplet concentrations dominate coverage and water of oceanic low-level clouds. *Science* **363**, eaav0566 (2019).
31. Christensen, M. W., Jones, W. K. & Stier, P. Aerosols enhance cloud lifetime and brightness along the stratus-to-cumulus transition. *Proc. Natl. Acad. Sci. USA* **117**, 17591–17598 (2020).
32. Breen, K. H., Barahona, D., Yuan, T., Bian, H. & James, S. C. Effect of volcanic emissions on clouds during the 2008 and 2018 Kilauea degassing events. *Atmos. Chem. Phys.* **21**, 7749–7771 (2021).
33. Glassmeier, F. et al. Aerosol–cloud–climate cooling overestimated by ship-track data. *Science* **371**, 485–489 (2021).
34. Christensen, M. W. et al. Opportunistic experiments to constrain aerosol effective radiative forcing. *Atmos. Chem. Phys.* **22**, 641–674 (2022).
35. Bender, F. A. M., Frey, L., McCoy, D. T., Grosvenor, D. P. & Mohrmann, J. K. Assessment of aerosol–cloud–radiation correlations in satellite observations, climate models and reanalysis. *Clim. Dyn.* **52**, 4371–4392 (2019).
36. Fuchs, J., Cermak, J. & Andersen, H. Building a cloud in the southeast Atlantic: understanding low-cloud controls based on satellite observations with machine learning. *Atmos. Chem. Phys.* **18**, 16537–16552 (2018).
37. Oreopoulos, L., Cho, N. & Lee, D. A global survey of apparent aerosol–cloud interaction signals. *J. Geophys. Res. Atmos.* **125**, e2019JD031287 (2020).
38. Fan, J., Wang, Y., Rosenfeld, D. & Liu, X. Review of aerosol–cloud interactions: mechanisms, significance, and challenges. *J. Atmos. Sci.* **73**, 4221–4252 (2016).
39. Gettelman, A., Schmidt, A. & Egill Kristjánsson, J. Icelandic volcanic emissions and climate. *Nat. Geosci.* **8**, 243–243 (2015).
40. Oreopoulos, L., Cho, N., Lee, D. & Kato, S. Radiative effects of global MODIS cloud regimes. *J. Geophys. Res. Atmos.* **121**, 2299–2317 (2016).
41. Mastrandrea, M.D., et al. Guidance Note for Lead Authors of the IPCC Fifth Assessment Report on Consistent Treatment of Uncertainties. *Intergovernmental Panel on Climate Change (IPCC)*. Available at https://www.ipcc.ch/site/assets/uploads/2017/08/AR5_Uncertainty_Guidance_Note.pdf (2010).
42. Grist, J. P. et al. Extreme air–sea interaction over the North Atlantic subpolar gyre during the winter of 2013–2014 and its sub-surface legacy. *Clim. Dyn.* **46**, 4027–4045 (2016).
43. Grosvenor, D. P. et al. Remote sensing of droplet number concentration in warm clouds: a review of the current state of knowledge and perspectives. *Rev. Geophys.* **56**, 409–453 (2018).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

© The Author(s), under exclusive licence to Springer Nature Limited 2022, corrected publication 2022

Methods

MODIS observations. We used the Level 3 products of MODIS Collection 6.1, that is, MYD08 for 2002–2020 from Aqua and MOD08 for 2001–2020 from Terra. The reported retrieval bias due to instrument degradation in Terra-MODIS Collection 5.1 datasets¹⁵ has been rectified in Collection 6.1. An inadvertent artefact in the calculations of cloud fraction (derived from cloud optical property) in Collection 5.1 has also been removed in Collection 6.1¹⁴, and both Terra-MODIS and Aqua-MODIS now show consistent results^{45,46}. Cloud droplet r_{eff} , in-cloud LWP, cloud optical thickness and cloud phase are retrieved from observed radiances using a radiative transfer model at 1 km nadir resolution in Level 2 products and aggregated to the $1^\circ \times 1^\circ$ Level 3 products⁴⁶. The Level 3 Cloud Optical Property Cloud Fraction product for the liquid phase (dataset name Cloud Retrieval Fraction Liquid)⁴⁶ is used in the CF analysis because this CF product can distinguish between clouds of liquid and ice phase and is consistent with the other microphysical retrievals of cloud properties used in this study. Note that ref. ¹⁷ used MODIS Collection 6 data and used cloud fraction derived from the cloud mask¹⁷ multiplied by the fractional liquid cloud and found a more modest increase in cloud fraction of ~1.7% in October. Differences in the climatological analysis between our findings and those from ref. ¹⁷ probably arise from a combination of the use of different CF products, the extension of the MODIS data to include data from the 2015–2020 period and differences in the areas of investigation. Monthly mean products are used in this study, with differences being negligible when aggregating Level 3 daily products into monthly means^{15,32}. An exception is liquid cloud N_d , which is derived from r_{eff} and cloud optical thickness assuming adiabatic conditions^{8,20,31,48}, and because of nonlinear dependencies, N_d is first obtained on a daily basis and then averaged to monthly means^{48,49}. Only pixels with r_{eff} between 4 μm and 30 μm and cloud optical thickness between 4 and 70 are used for the most reliable N_d retrievals⁴⁹. The uncertainty of the derived N_d is discussed in detail in ref. ⁴³, which estimated that the uncertainty can be largely reduced to about 50% when averaged over $1^\circ \times 1^\circ$ regions. The uncertainty is expected to be even smaller in our study since we average across a geographical region of about 3,000 km × 6,000 km.

To further back up our finding of increased CF, we also analysed the frequency of cloud-free conditions in arguably the most stringent MODIS product, namely, pixels with retrieved AOD at 550 nm, which are used as a proxy of cloud-free pixels. These pixels are most stringent because any thin or sub-grid-scale cloud is screened out to prevent contamination of AOD retrievals. Level 3 monthly MODIS AOD products record the number of validated 1 km² pixels used in the Level 2 products when performing aggregation. These statistics are used to calculate the relative reduction of cloud-free pixels in our region in October 2014 relative to the long-term 2001–2020 October mean excluding 2014. While the number of pixels with AOD retrievals does not have a one-to-one correspondence with the number of cloud-free pixels because factors such as sun glint in cloud-free pixels can reduce the number of AOD pixels, it is still a good relative (rather than absolute) proxy for cloud-free pixels.

Meteorological reanalyses. Meteorological reanalyses represent the best estimate of global atmospheric conditions⁵⁰ and are available from the European Centre for Medium-Range Weather Forecasts Reanalysis v.5 (ERA5) products. To train ML-MODIS, we use the monthly averaged ERA5 reanalysis from the surface up to 550 hPa level at $0.25^\circ \times 0.25^\circ$ horizontal resolution and 50 hPa vertical resolution. This vertical range covers most of the low-level liquid clouds. In total, 114 meteorological parameters are re-gridded to MODIS grid cells and used in the training, details of which are provided in Supplementary Table 1. The ERA5 monthly reanalysis products at 11:00 and 13:00 Icelandic time (same as UTC) are closest to the daytime Terra and Aqua overpass times and are paired with the respective MODIS products from these satellites for the training.

Machine learning. Previous studies that use machine learning to investigate the statistical correlation between cloud properties and aerosol (for example, refs. ^{36,51}) can possibly be affected by confounding meteorological co-variability that would prevent confirmation of the causal processes of ACIs^{43,46}. Here we use a random-forest algorithm⁵² to train an ML-MODIS that diagnoses cloud properties for given meteorological conditions but is unperturbed by volcanic aerosol. This allows comparisons of cloud properties between conditions with and without volcanic aerosol perturbation but otherwise alike, therefore quantifying cloud responses only to volcanic aerosol, that is, signals of ACI. Note that this machine-learning approach is not designed to calculate the temporal evolution of cloud properties and cannot predict the development of meteorological systems. The latter is obtained from the ERA5 reanalysis, which provides the best estimate of atmospheric state⁵⁰.

The random-forest algorithm is chosen because of its excellent performance in dealing with relatively small sample sizes and high-dimensional feature spaces and in avoiding over-fitting^{52,53}. Random-forest-based machine learning has been successfully applied to isolate the confounding meteorological variability in air-quality assessments and has been shown to perform much better than multinomial regression models^{54–56}. A regression mode forest of 100 trees is trained independently for each cloud property (N_d , r_{eff} , LWP and CF) and for each month (October and September), with a minimal leaf size of seven for each tree without merge leaves. Each tree samples ~60% of the input data with replacement for

the training data, and the remaining data are used as out-of-bag observations. With larger forests, we find a negligible reduction in out-of-bag mean squared error and a negligible increase in out-of-bag coefficient of determination (a more informative estimate of performance than mean squared error⁵⁷) of up to 0.87 for CF prediction. This indicates a good stability and avoidance of over-fitting⁵⁸. The number of randomly selected predictors is 38 (one-third of the total number of features), and the interactive-curvature method is used to select split predictors. The ERA5 meteorological reanalysis is independent of the MODIS datasets, which are not assimilated in the reanalysis⁵⁹, and provides the explanatory variables in the ML-MODIS training. The dependent variables are the corresponding cloud properties observed by MODIS with no volcanic eruption. The successful training of ML-MODIS is enabled by the large MODIS dataset from continuous observations over the past 20 years on two satellite platforms. We employ the ‘out-of-bag permuted predictor delta error’ method^{52,59} to measure the importance of each explanatory feature in predicting cloud properties. The results for CF are shown in Extended Data Fig. 6.

The performance of ML-MODIS as a surrogate of the MODIS observations under conditions without the volcanic perturbation is evaluated using leave-one-year-out cross validation⁶⁰ for each cloud property, as shown on Fig. 1a–d and Extended Data Fig. 7. This involves training ML-MODIS using randomly selected sets of 18 years of ERA5–MODIS dataset pairs and then evaluating ML-MODIS against the remaining 19th year of MODIS observations. This evaluation is carried out for each non-eruption year during 2001–2020. The uncertainty of ML-MODIS is further estimated using a Monte Carlo method, and the variability of the reference baselines is shown as black box plots in Fig. 3 and Extended Data Fig. 8a. For the Monte Carlo uncertainty estimate, we randomly perform leave-one-year-out validation 500 times for each cloud property by excluding both Terra and Aqua datasets of the randomly selected year over the entire region from machine-learning training but use them for validation. A test for N_d using the validation of a 700-member Monte Carlo ensemble showed negligible differences. The ratios of cloud properties between the ML-MODIS prediction (without volcano perturbation) and MODIS observations in 2014 (with volcano perturbation) are in pink in Fig. 3 for October and in Extended Data Fig. 8a for September, with the pink box plots showing the variability of all decision trees within the random-forest Monte Carlo ensembles, that is, the variability of the ACI signals.

Radiative forcing. We estimate the relative contributions from the Twomey r_{eff} , LWP adjustment and CF adjustment to ACI-induced radiative forcing using the susceptibilities of r_{eff} , LWP and CF to N_d perturbations. The radiative forcing arising from cloud albedo brightening can be described as equation (1) at a constant CF^{8,9,61}, and the forcing arising from CF enhancement can be described as equation (2) at a constant cloud albedo α_{cld} :

$$\left. \frac{dS_{\text{TOA}}}{d \ln \text{AOD}} \right|_{\text{CF}} = -SW_{\text{down}} \times CF \times \alpha_{\text{cld}} \times (1 - \alpha_{\text{cld}}) \times \frac{d \ln N_d}{d \ln \text{AOD}} \quad (1)$$

$$\times \left(\frac{1}{3} + \frac{5}{6} \frac{d \ln \text{LWP}}{d \ln N_d} \right)$$

$$\left. \frac{dS_{\text{TOA}}}{d \ln \text{AOD}} \right|_{\text{albedo}} = -SW_{\text{down}} \times (\alpha_{\text{cld}} - \alpha_{\text{cs}}) \times \frac{dCF}{d \ln \text{AOD}} = -SW_{\text{down}} \times (\alpha_{\text{cld}} - \alpha_{\text{cs}}) \times CF \times \frac{d \ln CF}{d \ln N_d} \times \frac{d \ln N_d}{d \ln \text{AOD}} \quad (2)$$

where dS_{TOA} is the short-wave radiative forcing at the top of the atmosphere, SW_{down} is the incoming short-wave solar radiation at the top of the atmosphere, α_{cld} is the average broadband short-wave cloud albedo with a global mean of 0.38 for liquid clouds⁶² and α_{cs} is clear-sky broadband ocean surface albedo, which is about 0.07 for representative of global average (solar zenith angle of 60°)⁶². The total ACI-induced short-wave radiative forcing is the sum of equation (1) and equation (2), as shown in equation (3).

$$\left. \frac{dS_{\text{TOA}}}{d \ln \text{AOD}} \right|_{\text{albedo}} = \left. \frac{dS_{\text{TOA}}}{d \ln \text{AOD}} \right|_{\text{albedo}} + \left. \frac{dS_{\text{TOA}}}{d \ln \text{AOD}} \right|_{\text{CF}} = -SW_{\text{down}} \times CF \times \frac{d \ln N_d}{d \ln \text{AOD}} \times [\frac{1}{3} \alpha_{\text{cld}} (1 - \alpha_{\text{cld}}) + \alpha_{\text{cld}} (1 - \alpha_{\text{cld}}) \times \frac{5}{6} \frac{d \ln \text{LWP}}{d \ln N_d} + (\alpha_{\text{cld}} - \alpha_{\text{cs}}) \frac{d \ln CF}{d \ln N_d}] \quad (3)$$

The radiative forcing contributions from the Twomey r_{eff} effect, LWP adjustment and CF adjustment are described as the three terms in the square bracket from left to right, respectively.

Data availability

The MODIS cloud and aerosol products from Aqua (MYD08_L3) and Terra (MOD08_L3) used in this study are available from the Atmosphere Archive and Distribution System Distributed Active Archive Center of National Aeronautics and Space Administration (LAADS-DAAC, NASA), <https://ladsweb.modaps.eosdis.nasa.gov>. ERA5 datasets are available from the European Centre for Medium-range Weather Forecast (ECMWF) archive, <https://cds.climate.copernicus.eu>. Source data are provided with this paper.

Code availability

Code is available from the corresponding author on reasonable request.

References

44. Platnick, A. S. et al. *MODIS Cloud Optical Properties: User Guide for the Collection 6/6.1 Level-2 MOD006/MYD006 Product and Associated Level-3 Datasets*. https://atmosphere-imager.gsfc.nasa.gov/sites/default/files/ModAtmo/MODISCloudOpticalPropertyUserGuideFinal_v1.1.pdf (NASA, 2018).
45. Platnick, S. et al. The MODIS cloud optical and microphysical products: collection 6 updates and examples from Terra and Aqua. *IEEE Trans. Geosci. Remote Sens.* **55**, 502–525 (2017).
46. Hubanks, P., Platnick, A. S., King, M. & Ridgway, B. *MODIS Atmosphere L3 Gridded Product Algorithm Theoretical Basis Document (ATBD) & Users Guide*. https://modis-images.gsfc.nasa.gov/_docs/L3_ATBD_C6.pdf (NASA, 2016).
47. Maddux, B. C., Ackerman, S. A. & Platnick, S. Viewing geometry dependencies in MODIS cloud products. *J. Atmos. Ocean. Technol.* **27**, 1519–1528 (2010).
48. Quaas, J., Boucher, O., Bellouin, N. & Kinne, S. Satellite-based estimate of the direct and indirect aerosol climate forcing. *J. Geophys. Res. Atmos.* <https://doi.org/10.1029/2007JD008962> (2008).
49. Quaas, J., Boucher, O. & Lohmann, U. Constraining the total aerosol indirect effect in the LMDZ and ECHAM4 GCMs using MODIS satellite data. *Atmos. Chem. Phys.* **6**, 947–955 (2006).
50. Hersbach, H. et al. The ERA5 global reanalysis. *Q. J. R. Meteorol. Soc.* **146**, 1999–2049 (2020).
51. Andersen, H., Cermak, J., Fuchs, J., Knutti, R. & Lohmann, U. Understanding the drivers of marine liquid-water cloud occurrence and properties with global observations using neural networks. *Atmos. Chem. Phys.* **17**, 9535–9546 (2017).
52. Breiman, L. Random forests. *Mach. Learn.* **45**, 5–32 (2001).
53. Biau, G. & Scornet, E. A random forest guided tour. *Test* **25**, 197–227 (2016).
54. Grange, S. K., Carslaw, D. C., Lewis, A. C., Boleti, E. & Hueglin, C. Random forest meteorological normalisation models for Swiss PM₁₀ trend analysis. *Atmos. Chem. Phys.* **18**, 6223–6239 (2018).
55. Shi, Z. et al. Abrupt but smaller than expected changes in surface air quality attributable to COVID-19 lockdowns. *Sci. Adv.* **7**, eabd6696 (2021).
56. Yang, J. et al. From COVID-19 to future electrification: assessing traffic impacts on air quality by a machine-learning model. *Proc. Natl. Acad. Sci. USA* **118**, e2102705118 (2021).
57. Chicco, D., Warrens, M. J. & Jurman, G. The coefficient of determination *R*-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *PeerJ Comput. Sci.* **7**, e623 (2021).
58. Cutler, A., Cutler, D. R. & Stevens, J. R. in *Ensemble Machine Learning* (eds Zhang, C. & Ma, Y.), 157–175, (Springer, 2012); https://doi.org/10.1007/978-1-4419-9326-7_5
59. Bonte, S., Goethals, I. & Holen, R. V. Individual prediction of brain tumor histological grading using radiomics on structural MRI. In *Proc. 2017 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC)*, pp 1–3, <https://doi.org/10.1109/NSSMIC.2017.8532793> (2017).
60. Bastos, L. S. & O'Hagan, A. Diagnostics for Gaussian process emulators. *Technometrics* **51**, 425–438 (2009).
61. Ackerman, A. S. et al. Effects of aerosols on cloud albedo: evaluation of Twomey's parameterization of cloud susceptibility using measurements of ship tracks. *J. Atmos. Sci.* **57**, 2684–2695 (2000).
62. Jin, Z., Charlock, T. P., Smith, W. L. Jr & Rutledge, K. A parameterization of ocean surface albedo. *Geophys. Res. Lett.* <https://doi.org/10.1029/2004GL021180> (2004).

Acknowledgements

We acknowledge the support of the UK Natural Environment Research Council (NERC) funded ADVANCE project (NE/T006897/1), which funded J.H., Y.C., D.P., A.S., D.G. and P.F. J.H., G.J. and F.M. were also partly funded by the EU's Horizon 2020 research and innovation programme under the CONSTRAIN grant agreement 820829. J.H., P.F., G.J. and F.M. are supported by the Joint UK BEIS/Defra Met Office Hadley Centre Climate Programme (GA01101). Y.W. is grateful for support from P. Sarasin and the ETH Zurich Foundation (ETH Fellowship project: 2021-HS-332). D.G. is funded by the National Centre for Atmospheric Science (NCAS), one of the UK NERC's research centres. J.d.L. acknowledges funding from the NERC funded V-PLUS grant NE/S00436X/1. N.C., L.O. and S.P. are funded by USA NASA programmes. The machine-learning training is performed using the Statistics and Machine Learning Toolbox in MATLAB (version R2019b, MathWorks Inc.). We thank K. Carslaw (University of Leeds) for co-developing and co-leading the Leeds aspect of ADVANCE project, and A. Jones (UK Met Office) for helpful discussions.

Author contributions

Y.C. and J.H. conceived the study. Y.W. and Y.C. designed and developed the machine-learning approach used in this study with help from J.H. and J.F. J.H. led the ADVANCE project funded by UK NERC. Y.C., F.M., G.J. and J.H. performed the analysis of MODIS data with help from D.G., N.C., L.O. and S.P. N.C., L.O. and S.P. performed the cloud regime analysis. Y.C., J.H., Y.W., D.G., U.L., P.F., L.O., S.P., J.d.L., A.S., D.P. and J.F. contributed to the uncertainty discussion. Y.C. and J.H. performed the analyses and interpreted the results with inputs from all co-authors. Y.C. and J.H. led the manuscript writing with specific inputs and edits from D.G., L.O. and U.L.. All co-authors discussed the results and commented on the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

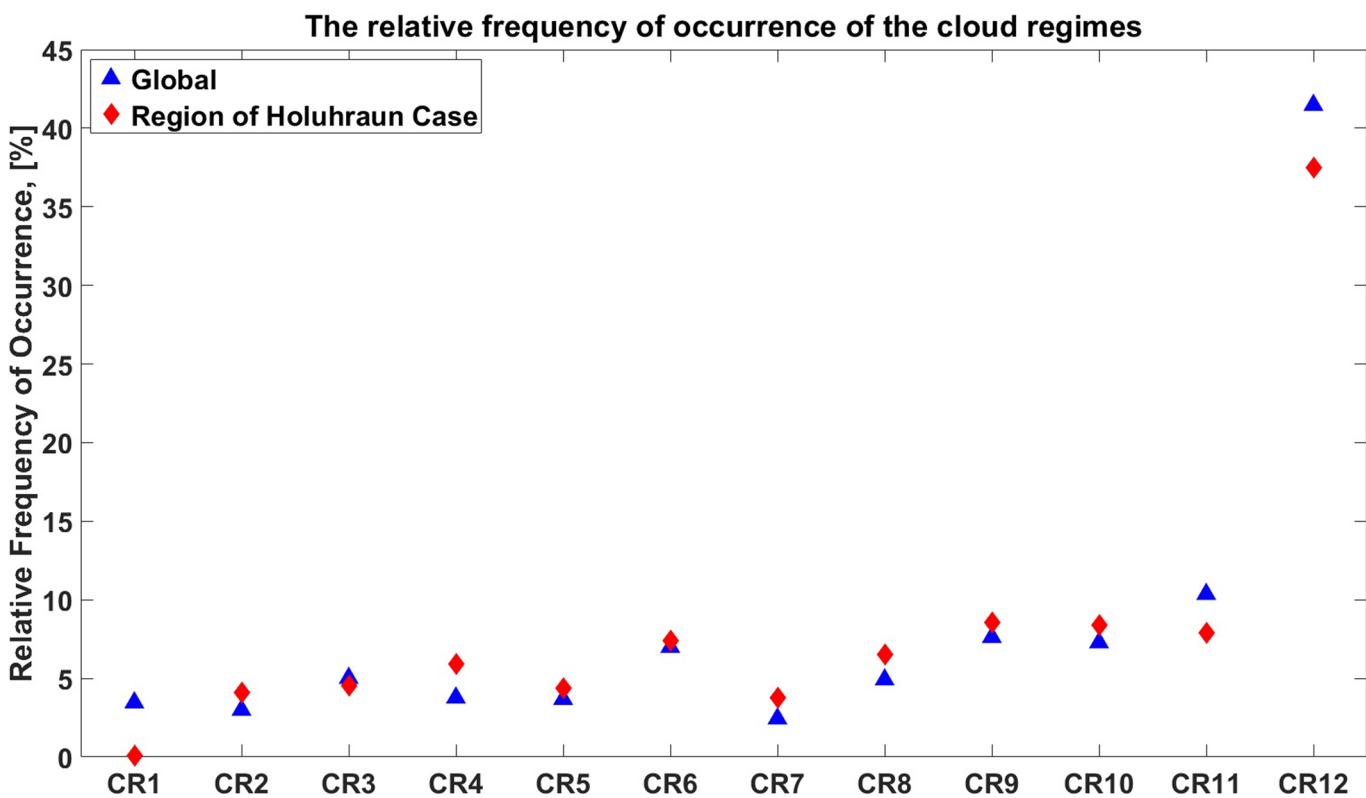
Extended data is available for this paper at <https://doi.org/10.1038/s41561-022-00991-6>.

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41561-022-00991-6>.

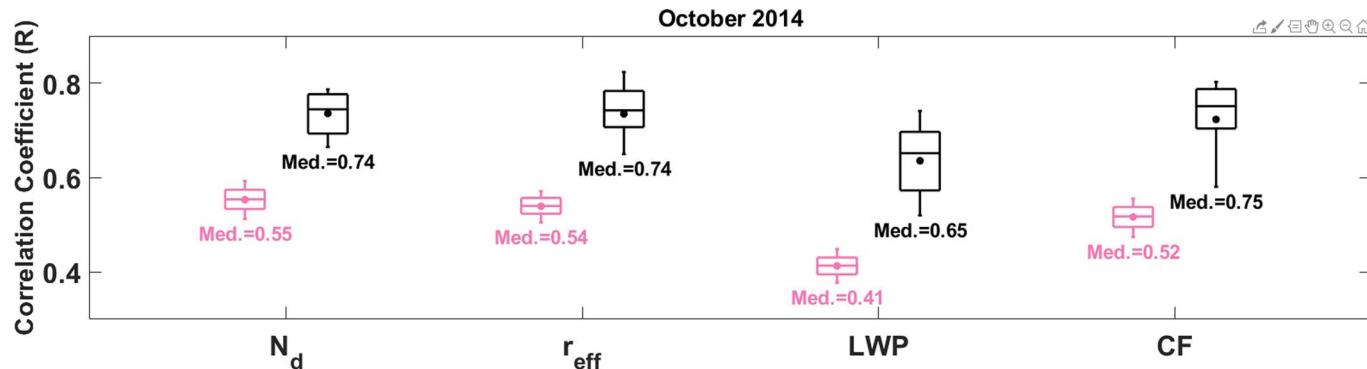
Correspondence and requests for materials should be addressed to Ying Chen.

Peer review information *Nature Geoscience* thanks Julia Fuchs, Velle Toll and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Primary Handling Editor: Tom Richardson, in collaboration with the *Nature Geoscience* team.

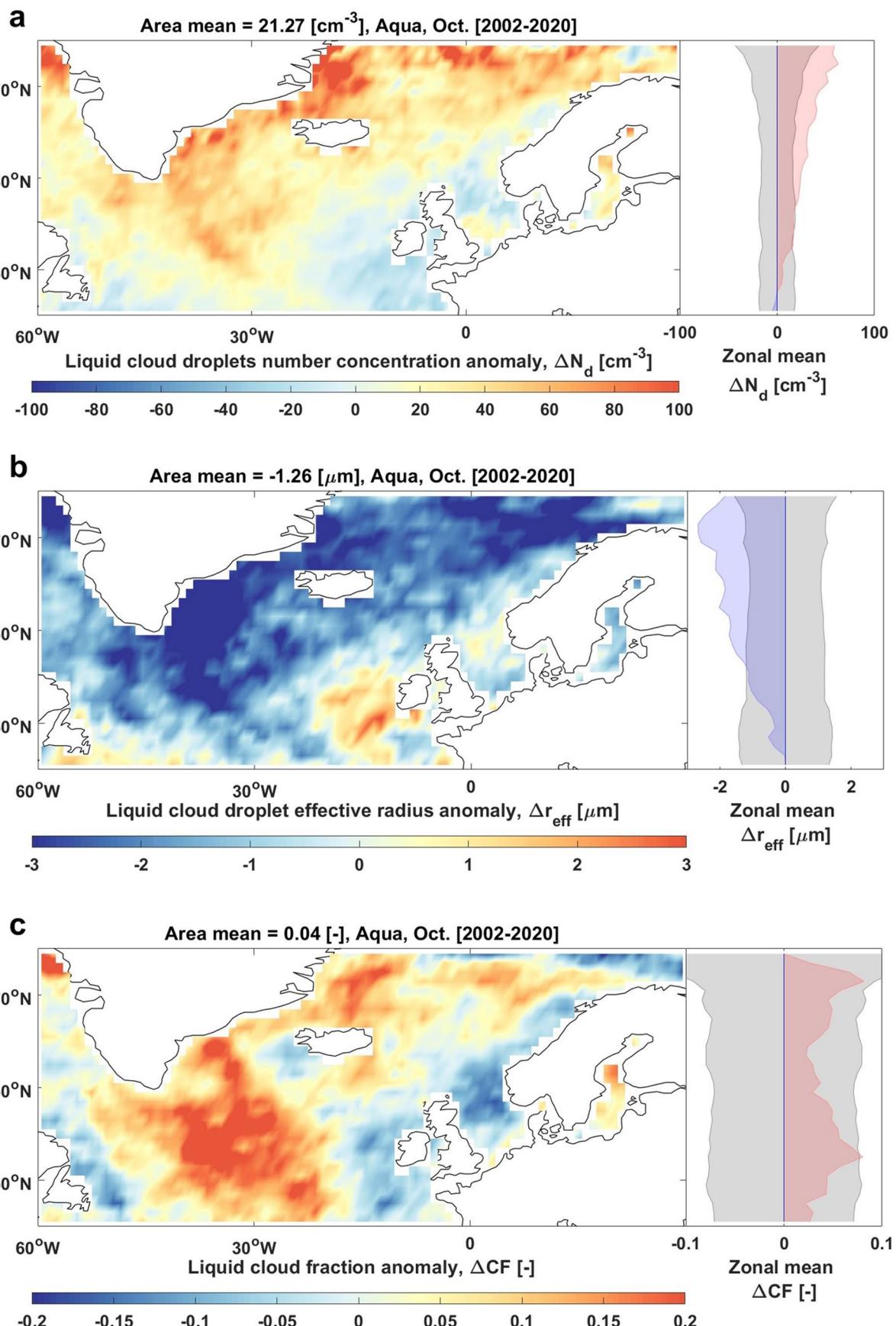
Reprints and permissions information is available at www.nature.com/reprints.



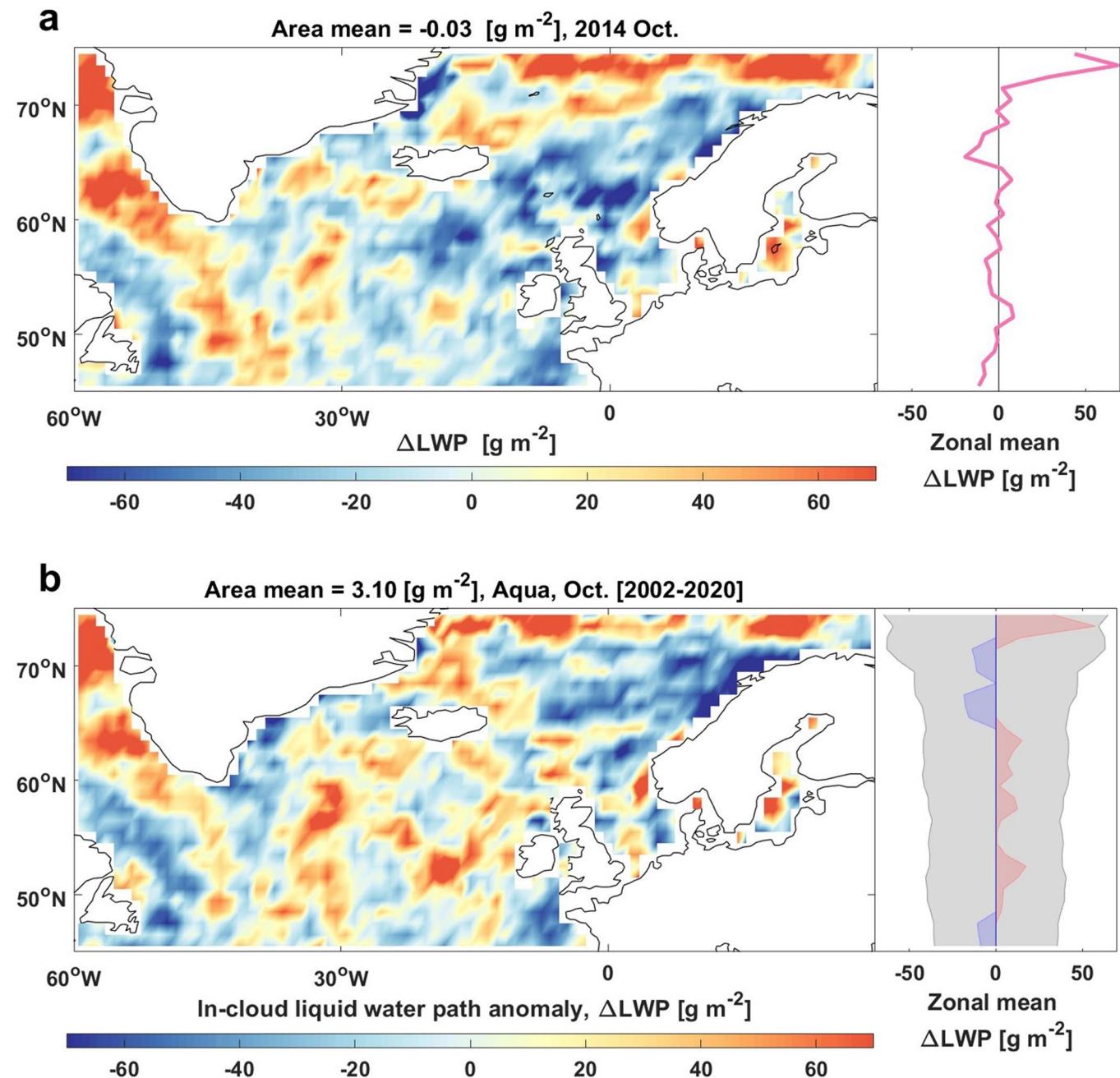
Extended Data Fig. 1 | Relative frequency of occurrence (RFO) of cloud regimes. The RFO values of the region studied here in September–October 2014 are given in red diamonds, data sourced from Malavelle et al.¹⁵. The RFO values during 2002–2014 globally are given in blue triangles, data sourced from Oreopoulos et al.⁴⁰. CR6–CR11 are liquid-dominated cloud regimes, and the others are ice-dominated cloud regimes. The details of each cloud regime are given in the above references accordingly.



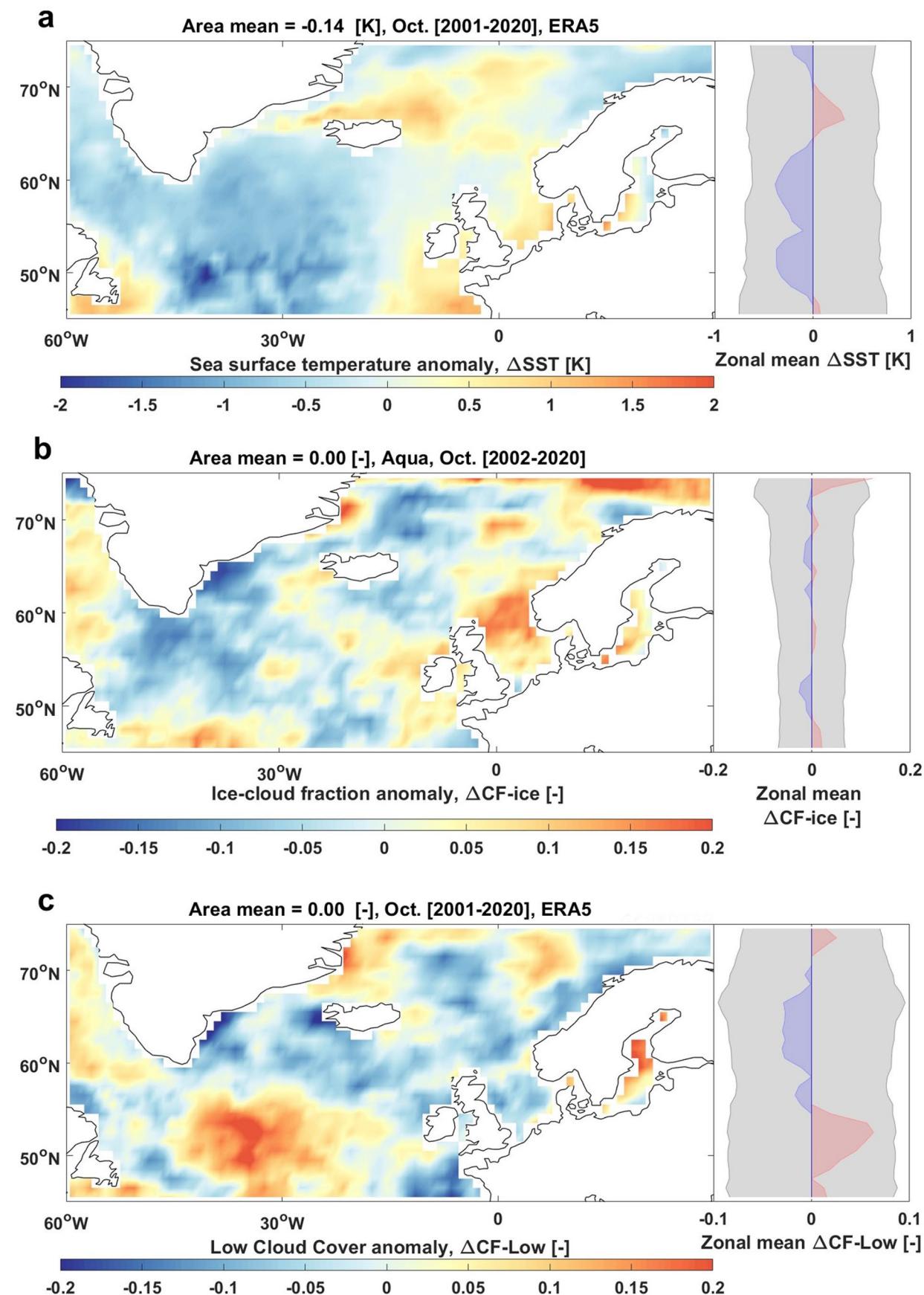
Extended Data Fig. 2 | Correlation coefficient between machine-learning predictions and MODIS observations of cloud properties, including liquid cloud droplet number concentration (N_d), liquid droplet effective radius (r_{eff}), liquid water path (LWP) and liquid cloud fraction (CF). The Monte Carlo results of ML-MODIS validation against MODIS observations without volcanic aerosol-perturbation are given in black. The variations of comparisons with volcanic aerosol-perturbation in October 2014 are given in pink. The boxplot shows 10th, 25th, median (Med.), 75th and 90th percentiles with the mean value indicated by a dot.



Extended Data Fig. 3 | Anomalies in MODIS-Aqua cloud properties for October 2014. The spatial distributions and zonal means of anomalies in N_d , r_{eff} and CF are shown in the panels a-c. Anomalies correspond to the deviation from the 2002–2020 climatology (excluding the 2014 eruption year). The positive anomalies are shown in red and negative ones in blue. The standard deviation is shown by the grey shading.

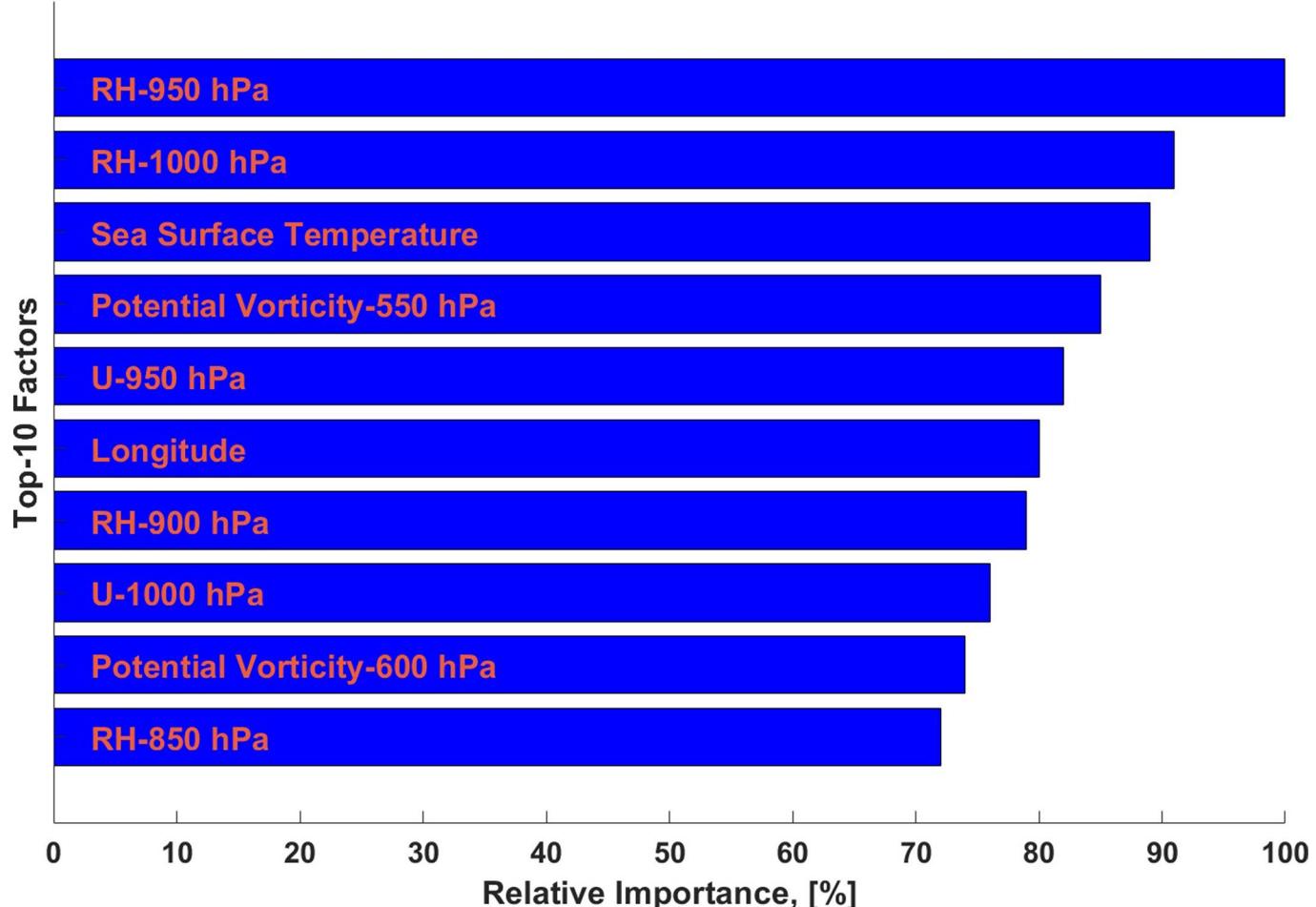


Extended Data Fig. 4 | Change (a) and anomaly (b) in liquid water path (LWP). **a**) Similar to Fig. 2, changes are detected using machine-learning; **b**) similar to Extended Data Fig. 3, anomaly corresponds to the deviation from 2002–2020 climatology.

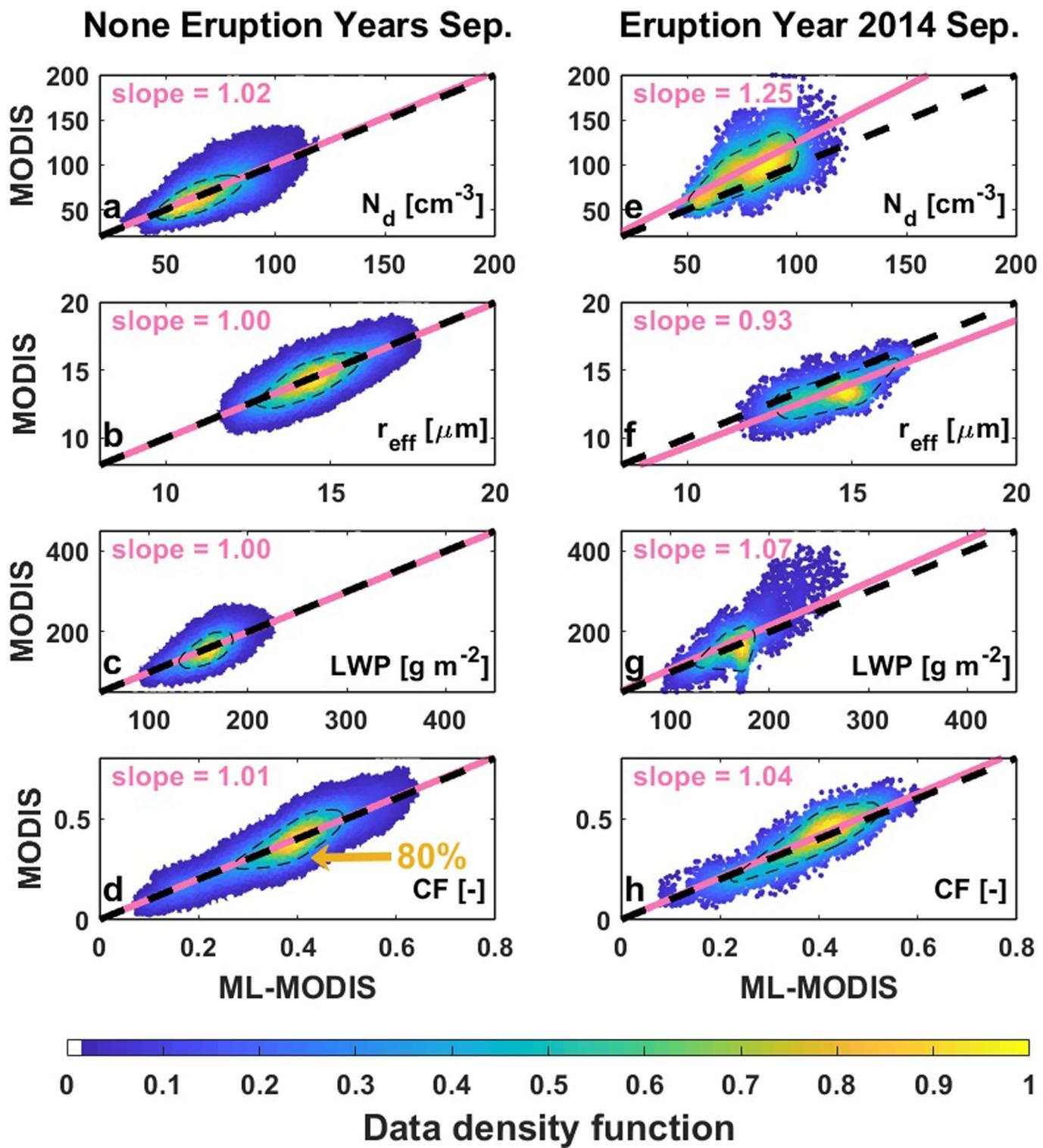


Extended Data Fig. 5 | Similar to Extended Data Fig. 3. but show anomaly in sea-surface temperature (a), anomaly in ice-cloud fraction in October 2014 (b), and climatological anomaly of low-level cloud cover in October 2014 using ERA5 reanalysis (c).

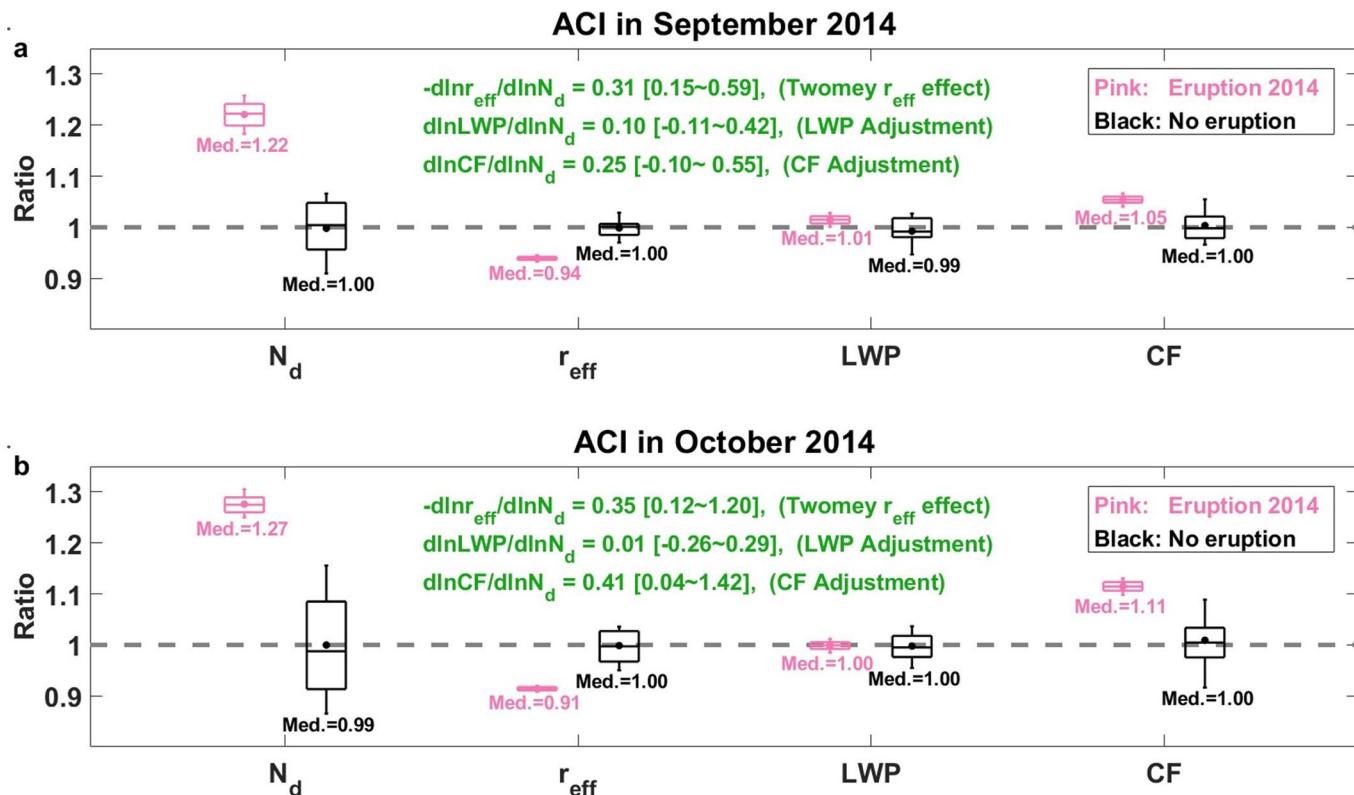
Machine Learning, AQUA & TERRA, Oct., CF



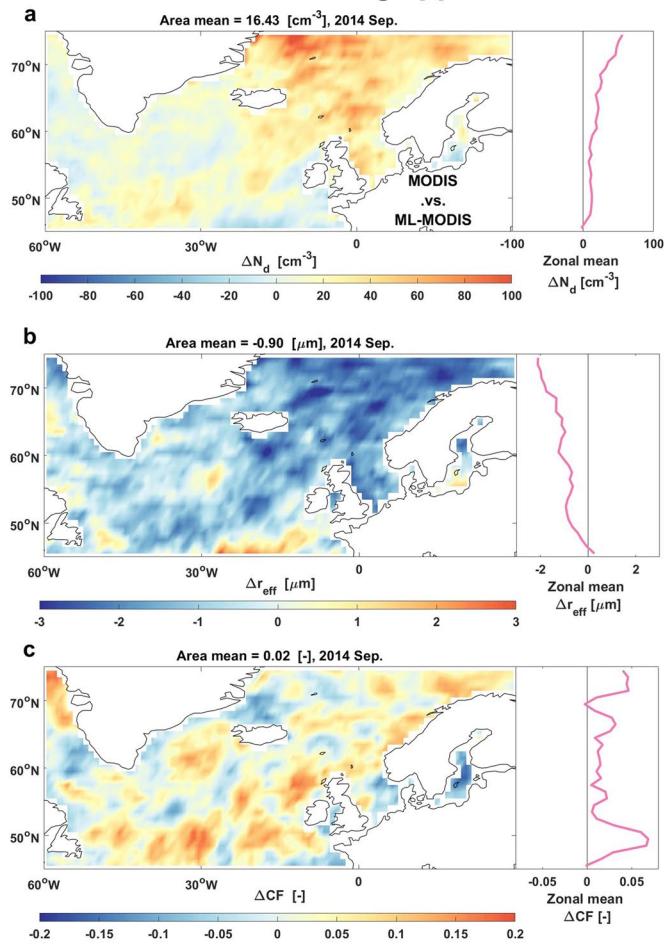
Extended Data Fig. 6 | The top-10 most important features for machine-learning to predict unperturbed liquid cloud fraction in October. The feature importance is normalized with the maximum as 100%. The value of these features in 2014 are entirely within the variation range of machine-learning training dataset, see Extended Data Fig. 10.



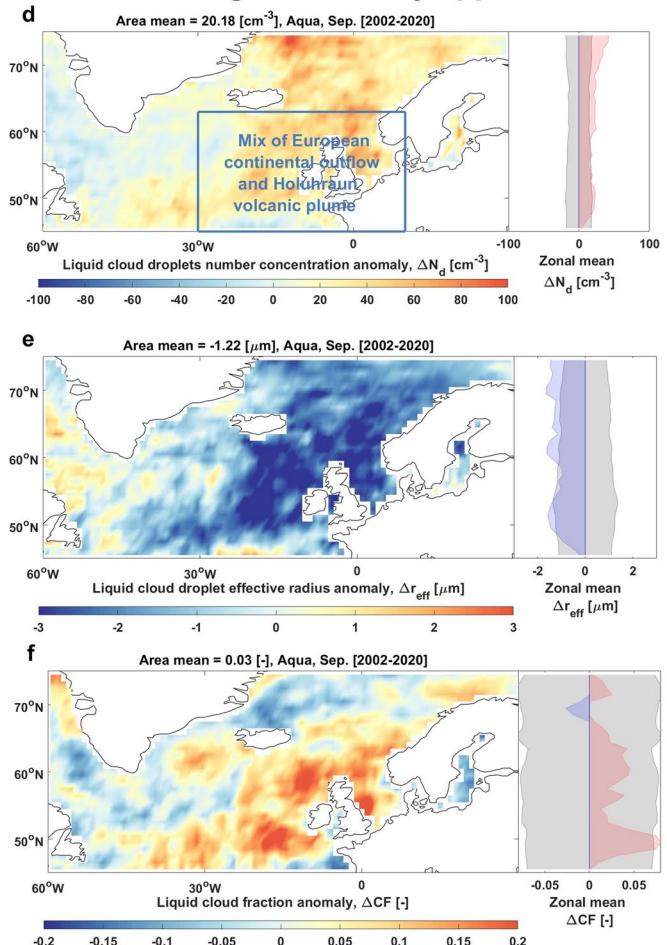
Extended Data Fig. 7 | Similar to Fig. 1, but show results in September 2014.



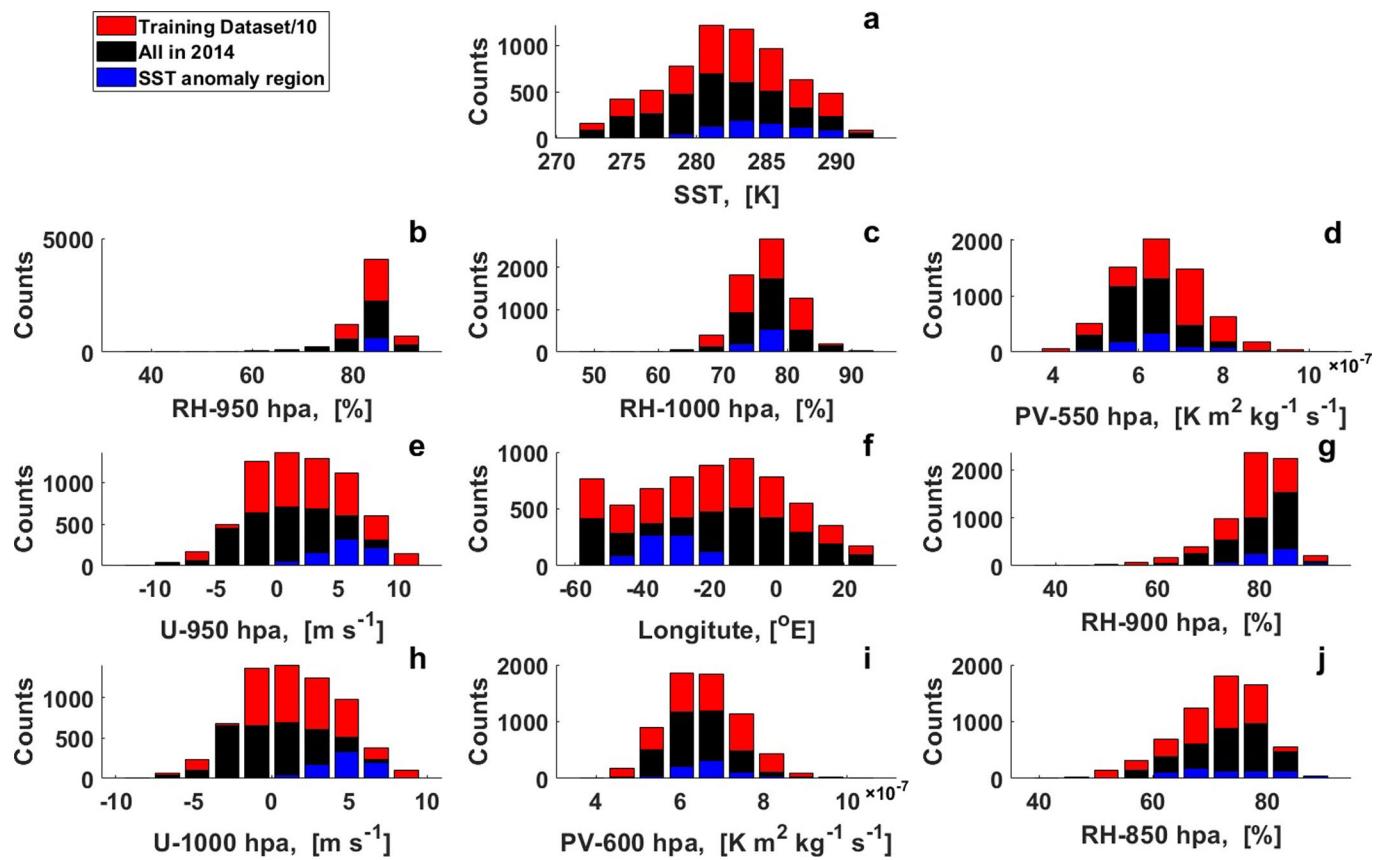
Machine-learning approach



Climatological anomaly approach



Extended Data Fig. 9 | Cloud responses to Holuhraun volcanic aerosol in September 2014. Left panels **a-c** (similar to Fig. 2 but for September 2014) show cloud responses to volcanic aerosol using machine-learning (ML) approach. Right panels **d-f** (similar to Extended Data Fig. 3 but for September 2014) show anomalies in cloud properties.



Extended Data Fig. 10 | Probability distribution of the top-10 most important features. as shown in Extended Data Fig. 6. Red bars indicate the counts (scaled by 0.1 to fit the display range) of the training data in each bin, which covers the entire variability range of black and blue bars; black bars indicate the data counts from the entire studied region in October 2014; and blue bars indicate the counts from the sea surface temperature anomaly only. Note that the counts per longitude are different, because we only consider data over the oceans.