JAMES | Journal of Advances in Modeling Earth Systems

RESEARCH ARTICLE
10.1029/2022MS003048

Predicting Convective Downdrafts From Updrafts and Environmental Conditions in a Global Storm Resolving Simulation

J. M. Windmiller[1], J. Bao[1], S. C. Sherwood[2,3], T. D. Schanzer[2,3], and D. Fuchs[2,3]

[1]Max Planck Institute for Meteorology, Hamburg, Germany, [2]Climate Change Research Centre, University of New South Wales, Sydney, NSW, Australia, [3]ARC Centre of Excellence for Climate Extremes, University of New South Wales, Sydney, NSW, Australia

Key Points:

- The best predictors of downdraft mass flux and velocity are rain amount and rate, respectively
- Updraft properties impact downdraft properties through their control on rain formation
- For a given rain amount and rate, environmental conditions add little skill to downdraft prediction

Correspondence to:

J. M. Windmiller,
julia.windmiller@mpimet.mpg.de

Citation:

Windmiller, J. M., Bao, J., Sherwood, S. C., Schanzer, T. D., & Fuchs, D. (2023). Predicting convective downdrafts from updrafts and environmental conditions in a global storm resolving simulation. *Journal of Advances in Modeling Earth Systems*, *15*, e2022MS003048. https://doi.org/10.1029/2022MS003048

Received 14 FEB 2022
Accepted 20 JAN 2023

Abstract One important component of precipitating convection is the formation of convective downdrafts. They can terminate the initial updraft, affect the mean properties of the boundary layer, and cause strong winds at the surface. While the basic forcing mechanisms for downdrafts are well understood, it is difficult to formulate general relationships between updrafts, environmental conditions, and downdrafts. To better understand what controls different downdraft properties, we analyze downdrafts over tropical oceans in a global storm resolving simulation. Using a global model allows us to examine a large number of downdrafts under naturally varying environmental conditions. We analyze the various factors affecting downdrafts using three alternative methods. First, hierarchical clustering is used to examine the correlation between different downdraft, updraft, and environmental variables. Then, either random forests or multiple linear regression are used to estimate the relationships between downdraft properties and the updraft and environmental predictors. We find that these approaches yield similar results. Around 75% of the variability in downdraft mass flux and 37% of the variability in downdraft velocity are predictable. Analyzing the relative importance of our various predictors, we find that downdrafts are coupled to updrafts via the precipitation generation argument. In particular, updraft properties determine rain amount and rate, which then largely control the downdraft mass flux and, albeit to a lesser extent, the downdraft velocity. Among the environmental variables considered, only lapse rate is a valuable predictor: a more unstable environment favors a higher downdraft mass flux and a higher downdraft velocity.

Plain Language Summary Once a cloud begins to rain, the air inside or below the cloud can gain negative buoyancy and sink to the ground. This downward movement of air is called a downdraft. Downdrafts can end the life cycle of a cloud and also result in strong, sometimes destructive, wind gusts at the surface. The basic driving forces for downdrafts are well understood. For example, we know that evaporation of rain and the associated latent cooling of air is usually critical in causing the air to become negatively buoyant. Even though the basic driving forces are known, many interrelated processes contribute simultaneously to the strength of the downdraft, making it difficult to predict the strength of a downdraft under specific conditions. In this study, we use an atmospheric simulation whose model domain spans the globe and can explicitly resolve rain clouds. Compared to previous studies, the use of a global domain allows us to study a very large number of rain clouds, and their associated downdrafts, which form under very different, naturally varying environmental conditions. Machine learning techniques and traditional statistical methods agree on the result that the strength of the downdraft can be well predicted if we know the strength of the updraft that caused the downdraft or, even better, if we know the amount of rain that an updraft produced. Surprisingly, we have found that downdrafts can be predicted only slightly better if we also know other environmental conditions of the air surrounding the downdraft, such as the temperature and/or humidity profiles.

1. Introduction

The onset of precipitation in a convective updraft is usually followed by the formation of a convective downdraft. If the downdraft cuts off the supply of buoyant air, it ends the original convective updraft (Byers & Braham, 1949). In addition to their central role in the life-cycle of precipitating convection, other important aspects of downdrafts are the injection of free tropospheric air into the boundary layer as well as the downdraft-induced outflow at the surface. The mass flux of free-tropospheric air into the boundary layer has been proposed to play a central

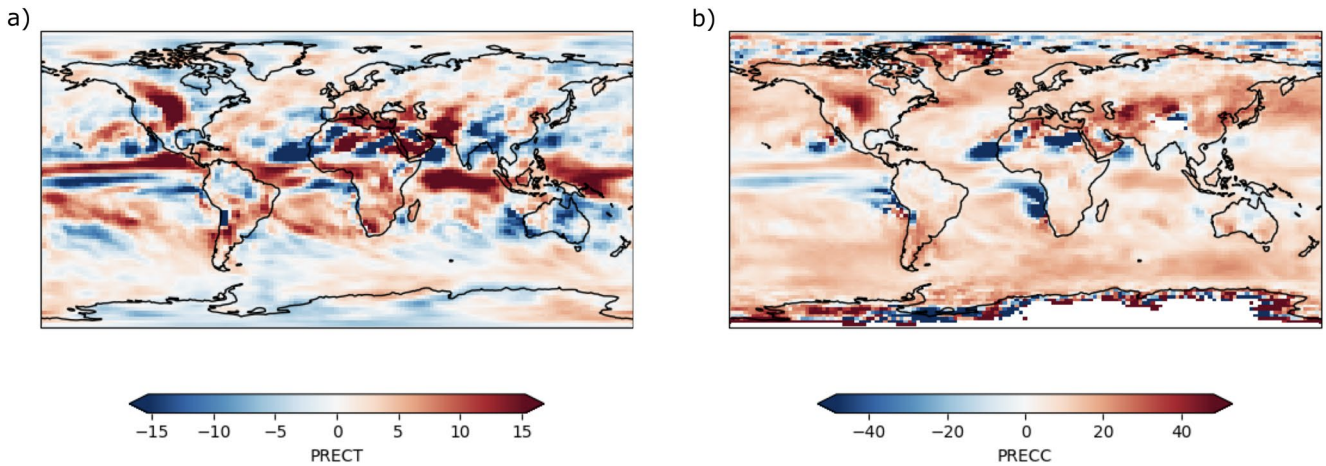

Figure 1. Relative changes in global precipitation between a simulation with and without downdraft parameterization. The simulations are run under Atmospheric Modeling Intercomparison Project conditions with prescribed sea surface temperature. Panel (a) shows the change on total precipitation (PRECT) and panel (b) shows the corresponding change in the precipitation as produced by the convection scheme itself (PRECC).

role for the equilibrium state of the tropical boundary layer (Raymond, 1995), though the importance of downdrafts compared to other processes, in particular entrainment at the boundary layer top, is still under discussion (Hansen et al., 2020; Thayer-Calder & Randall, 2015; Torri & Kuang, 2016). The downdraft-induced outflow at the surface has been shown to aid the development of new convection at its leading edge, see for example, Purdom (1976), Tompkins (2001), Schlemmer and Hohenegger (2014), Torri et al. (2015), and can cause severe damage due to strong winds (e.g., Fujita & Wakimoto, 1981). Therefore, to correctly parameterize convection it is crucial to represent the effects of downdrafts as well as the effects of updrafts. A simple way of showing the impact of downdraft parameterization is to analyze the effect of turning it off. For example, turning off the downdraft parameterization in the Zhang-McFarlane convection scheme (Zhang & McFarlane, 1995) in the setup of Fuchs et al. (2022), leads to significant changes in the distribution of precipitation as well as in the mechanism causing precipitation, see Figure 1. This also demonstrates the importance of predicting downdraft properties correctly.

The basic forcing mechanisms for downdrafts are given by the vertical momentum equation; the significance of the individual terms is summarized in the literature review by Wakimoto (2001). He notes that the vertical gradient of perturbation pressure and the perturbation pressure buoyancy term were, in most cases, less important than the precipitation induced changes in buoyancy. In particular, precipitation changes the buoyancy of downdrafts through evaporative cooling (i.e., the thermal buoyancy) and through condensate loading. Of the two effects, evaporative cooling is thought to make the larger contribution (see also Torri & Kuang, 2016). The negative buoyancy of a downdraft depends not only on the air inside the downdraft, but also on the surrounding air. The buoyancy is given by the density difference between the downdraft and the surrounding air, which is additionally influenced by the degree of mixing between the two air masses. Numerical calculations, investigating how downdraft strength depends on varying environments (e.g., Proctor, 1989; Srivastava, 1985, 1987), show a sensitivity to ambient environmental stability as well as humidity. More unstable lapse rates favor stronger downdrafts as unstable conditions help accelerate negatively buoyant air downward in the same way they help accelerate positively buoyant air upward. Srivastava (1985) notes that in the extreme case of an environmental lapse rate close to a dry adiabatic lapse rate, even light precipitation can lead to significant downdrafts. The dependence on ambient environmental humidity is less clear (see also Wakimoto, 2001). While a moister environment increases the virtual temperature difference between downdraft and environment, thus strengthening the downdraft, this effect can be reversed when the effect of entrainment of environmental air is considered. In particular, entraining dry air can enhance downdrafts as it can increase the evaporation rate. While Srivastava (1985) finds stronger downdrafts at higher environmental humidity, Proctor (1989) notes that this sensitivity is height dependent: stronger downdrafts are favored by drier environments near the melting level but moister environments near the surface.

What complicates the interpretation of these results from numerical calculations is that they exclude the effect of the environmental conditions on the preceding updraft, even though this might impact the amount of available

precipitation. Trapp and Woznicki (2017) and Marion and Trapp (2019) therefore used a "top-down" approach to link updraft strength to outflow strength. Using high-resolution and small-domain atmospheric simulations, they studied the evolution of supercell thunderstorms under different environmental conditions, focusing on the relationship between environment, updrafts, and downdrafts. They found that a change in environmental conditions led to changes in the updraft mass flux, which had a proportional effect on the downdraft mass flux. Trapp and Woznicki (2017) and Marion and Trapp (2019) argue that this coupling between the updraft and downdraft mass fluxes happens via the *precipitation-generation argument* (Trapp & Woznicki, 2017). In particular, large updraft areas lead to the formation of large precipitating areas and thus large areas over which evaporative cooling and condensate loading lead to the formation of downdrafts. While the "top-down" approach expressly includes the indirect effect of environment on the downdraft mass flux, that is, the effect of the environment on the downdraft strength via the changing updraft mass flux, this approach makes it difficult to distinguish between the direct and indirect effects of the environment on the downdraft strength.

Formulating a dependence of downdraft properties on updraft and environmental properties has, therefore, long been a challenge for convective parameterization schemes. Tiedtke (1989), for example, assumes downdraft mass flux to be 20% of the updraft mass flux at cloud base (see their Equation 16) and constant with height (their Equation 17). This assumption makes the downdraft mass flux independent of any direct impact of the environment. This is different in Kain and Fritsch (1993), where the downdraft mass flux is again proportional to the updraft mass flux, but by a factor that depends on environmental conditions, in particular on the vertical shear of the horizontal wind and cloud-base height. In Zhang and McFarlane (1995), the proportionality constant between updraft and downdraft mass flux depends, among other factors, on ambient humidity. Based on the assumption that downdrafts are kept in a saturated state by evaporation, the proportionality constant decreases in drier environments, where more evaporation is needed to keep downdrafts saturated (see their Equation 11). Thus, all else being equal, this parameterization leads to weaker downdrafts in drier environments. Finally, in a much more recently developed parameterization by Suselj et al. (2019), downdraft mass flux is, at the initialization level, equal in magnitude and opposite in direction to the corresponding updraft mass flux. The corresponding thermodynamic variables of the downdraft at the initialization level are set to the grid-mean values. This means, for example, that in a drier environment, downdrafts are, at least initially, drier. How much this environmental dependence at the initialization level impacts the downdraft at lower levels depends on their formulation of the entrainment and the evaporation rate. Downdrafts are, therefore, parameterized very differently in different convective parameterization schemes. Given the central role of downdrafts in the convective life cycle, and given that climate models rely on convective parameterization schemes, the goal of this study is to improve our understanding of how best to predict downdraft properties as a function of updraft and environmental properties.

The preceding discussion shows that it is difficult to formulate general relationships between updrafts, environmental conditions and downdrafts even though the fundamental drivers for downdrafts are well understood. In this study, we therefore extend the "top-down" approach used by Trapp and Woznicki (2017) and Marion and Trapp (2019) to a more general setup by using a global atmospheric simulation to study precipitating downdrafts over tropical oceans. While a global simulation puts limits on the horizontal resolution as well as the output frequency, it provides three key advantages:

1. A wide variety of different types of precipitating convection and environments is studied
2. The variability in convection type and environment emerges "naturally," that is, from the simulated cross-correlation functions instead of having to be fixed externally
3. The large number of downdrafts allows us to explicitly address the question of how downdrafts depend on the environment in a direct and an indirect manner

This low-resolution, global-domain study therefore complements the high-resolution, small-domain studies and numerical calculations described above. In particular, it allows us to test the results of these studies under naturally varying environmental conditions.

The structure of our paper is as follows. In Section 2, we introduce the atmospheric simulation, our method for detecting downdrafts as well as the three methods we use to analyze the resulting data set: hierarchical clustering, random forests, and multiple linear regression. In Section 3, we discuss the mean properties of the detected up- and downdrafts as well as the relationship we find between the updraft properties, environmental conditions and the downdraft properties. A possible explanation for the surprisingly small impact of environmental humidity on downdraft mass flux and velocity is given in Section 4. Our overall results are summarized in Section 5.

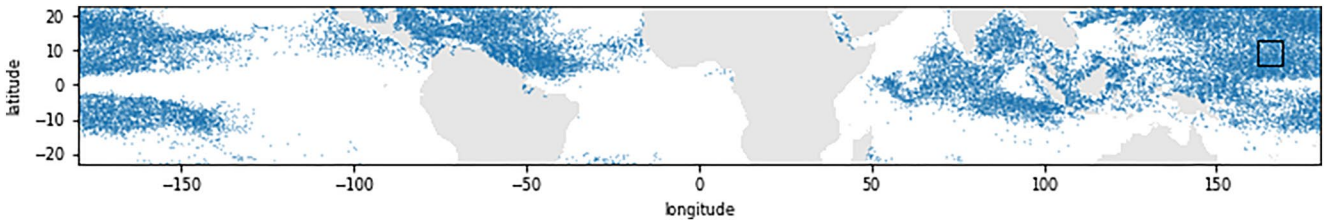

Figure 2. Location of convective events in the analysis domain. Each dot represents the geographic location of the center of mass of one of the 31,340 identified convective events. The black box indicates the outline of the sub-domain shown in Figure 3. Each event is considered as one instance in the machine-learning and multiple linear regression analyses.

2. Methods

2.1. Numerical Model Description

We use the ICON model version 2.1.02 (Icosahedral Nonhydrostatic Weather and Climate Model; Zängl et al., 2015) to perform a global atmospheric simulation for 20 days (1 August 2016 to 20 August 2016). To be able to simulate a global domain we use a quasi-uniform horizontal mesh of 5 km and 90 vertical levels with a grid spacing that increases from the surface to the top of the atmosphere at 75 km. While this is too coarse to fully resolve deep convection, the key advantage of a global domain is that convective events emerge naturally in a system where all scales from convective to global are modeled consistently. This allows us not only to obtain a large sample but, even more importantly, to sample convective systems in very different but realistically varying environments. In contrast, finer-scale simulations will better resolve convection but normally have unrealistic boundary, initial conditions or nudging, which could bias the results even though the convection is better resolved, and would provide a smaller sample size of independent drafts/storms.

In general, the setup of the simulation follows the protocol of the DYAMOND model intercomparison project (Stevens et al., 2019) and is identical to the setup used in Hohenegger et al. (2020) and the control setup in Bao and Windmiller (2021) but with an increased output frequency of 15 min (instantaneous fields) for the last 5 days of the simulation. Our main reason to increase the output frequency for only 5 days was to limit the size of the output data. Additional analysis shows that the model has not fully equilibrated after 15 days, but we expect that the characteristics of convection have reached a natural state by then. As in Hohenegger et al. (2020), the atmosphere in the simulation is initialized using the atmospheric state of the European Centre for Medium-Range Weather Forecasts on the 1 August 2016 at 00UTC and the bottom boundary is initialized based on the daily observed sea surface temperature and sea ice cover. The parameterizations used in the model are a microphysics scheme (Baldauf et al., 2011), a radiation scheme (Rapid Radiative Transfer Model; Mlawer et al., 1997), a turbulent kinetic energy scheme for turbulent mixing (Raschendorfer, 2001), an interactive surface flux scheme as well as a soil model (Reick et al., 2013; Schrodin & Heise, 2002). The model does not use a parameterization scheme for gravity wave drag, nor one for convection, either shallow or deep. Note that the distribution of convective cloud heights in this setup has been shown to depend on horizontal resolution. In particular, Hohenegger et al. (2020) show that when systematically increasing the horizontal resolution from 80 to 2.5 km, the population of deep convective clouds increases at the expense of the population of shallow clouds. While this sensitivity is a limitation of our simulation that should be noted, the largest differences are found between simulations with a coarser or higher resolution than 20 km.

In the following, we use the model output from the days with increased output frequency, that is, the last 5 days of the simulation, and focus on tropical oceanic convection. We therefore exclude land and limit our analysis to 23°S–23°N. The analysis domain is shown in Figure 2 along with the locations of the convective events, which are determined as described below. For simplicity, we have chosen a domain that is symmetric with respect to the equator, even though the position of the ITCZ is shifted significantly to the north in August. We obtain very similar results using only the northern half of this domain (not shown).

2.2. Extraction of Downdraft, Updraft, and Environmental Properties

In this study, we use the surface precipitation field to locate deep convection, the corresponding up- and downdrafts, and environmental conditions. In the following, we illustrate the details of our method for a small

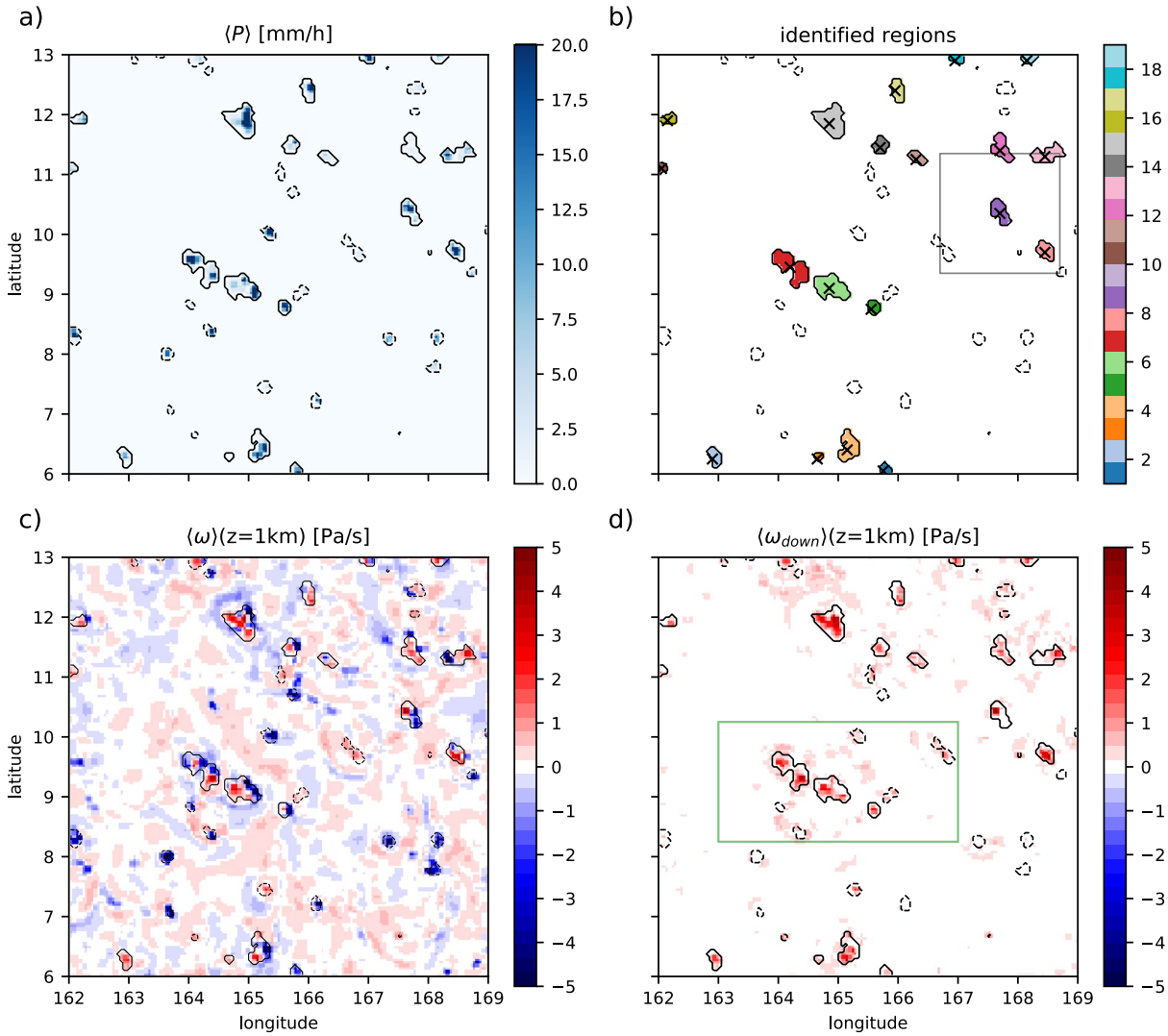

Figure 3. Illustration of the method to determine the convective events and the associated downdrafts (on 20 August 2016). (a) Mean surface precipitation in units of mm h^{-1} (time average calculated from 02:30 to 03:30 UTC), (b) areas of the identified convective events (colors) and the "environment" of the convective event at 167.7°E and 10.35°N (gray box), (c) pressure vertical velocity at a height of 1 km at 03:30 UTC in units of Pa s^{-1}, (d) the corresponding downdraft field, also in units of Pa s^{-1} and the sub-domain further analyzed in Figure 4 (green box). The areas of the convective events rejected due to our condition for coherent downdrafts are shown with dashed lines (see text for details).

sub-domain in the Pacific Ocean (6° to 13° north and 162° to 169° east) during a period of 6 hr (00 UTC to 06UTC 20 August 2016). The location of the sub-domain is indicated in Figure 2.

The basic idea of our method is to use the surface precipitation field to identify convective events and then, for each identified event, calculate the mean profiles of various properties of interest (e.g., updraft and downdraft velocities). We proceed in three steps to identify the relevant convective events:

- Use the surface precipitation field to identify potential convective events
- Split the vertical velocity field in a downdraft field and an updraft field
- Restrict our analysis to convective events with coherent downdrafts

2.2.1. Precipitation

We start by applying a threshold to the 1-hourly averaged precipitation field and then use a cluster algorithm (four-neighborhood kernel) to detect connected areas. As wind shear may lead to a spatial displacement between updraft, downdraft, and surface precipitation, we use a small precipitation threshold (0.1 mm hr^{-1}) to capture

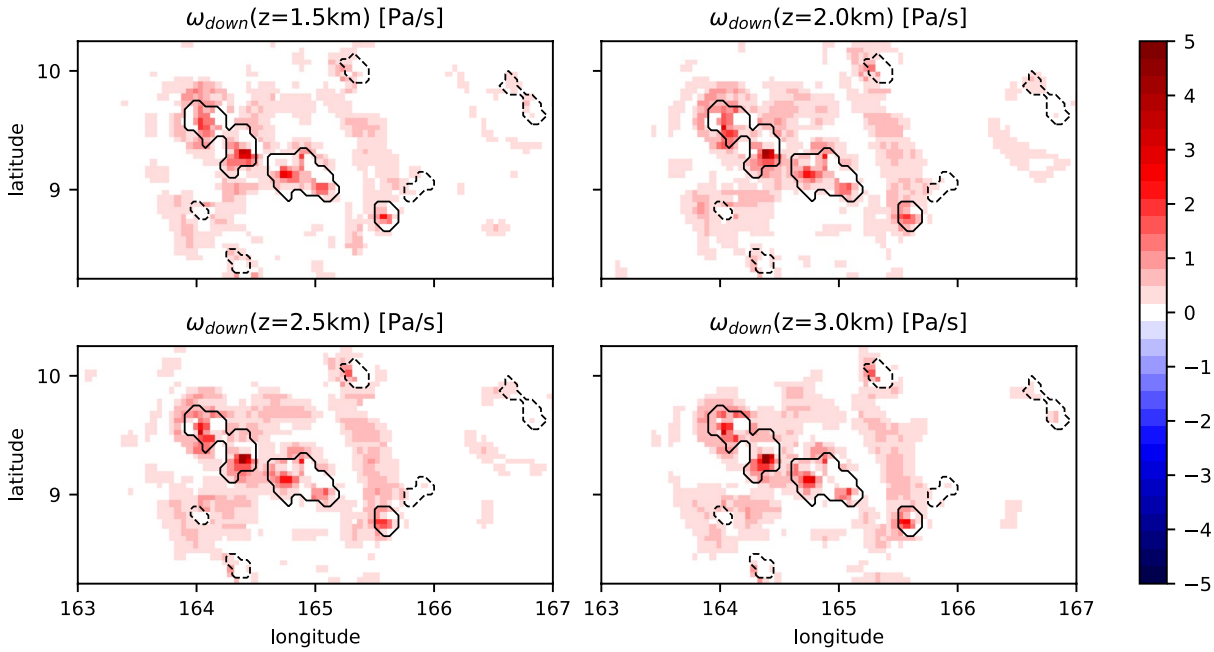

Figure 4. Illustration of the criterion for coherent downdrafts. Time averaged downdraft fields (see text for details) at four different heights with events matching our criterion marked by solid lines while events not matching our criterion are marked by dashed lines.

comparatively large "convective" areas. For our sample data set, the averaged precipitation field and the corresponding identified events are shown in Figures 3a and 3b.

2.2.2. Vertical Velocity

Figure 3c shows the hourly averaged vertical velocity field as well as the outline of the areas of the convective events corresponding to Figures 3a and 3b. Note that the vertical velocity field is only an approximation of a true hourly average, as it is calculated from averaging the quarter-hourly instantaneous vertical velocity fields. In Figure 3c, strong absolute vertical winds, which we identify as downdrafts and updrafts, occur mainly within the areas of the convective events. Outside these areas, vertical winds are generally much weaker and probably due to gravity waves. Visual inspection suggests that by applying a threshold of 1.0 Pa s^{-1} or higher for downdrafts and -1.0 Pa s^{-1} or lower for updrafts to the instantaneous velocity field *before* calculating the temporal average, as indicated by $\langle\cdot\rangle$, we can split the vertical velocity field in a downdraft and an updraft field and, at the same time, remove most of the background signal (see e.g., Figure 3d).

2.2.3. Coherent Downdrafts

Note that we limit our analysis to deep convective events with coherent downdrafts, thus not all the areas with high enough precipitation rate are actually identified as convective events of interest (the areas corresponding to the rejected events are marked with dashed lines in Figure 3). To identify coherent downdrafts, we require at least one horizontal grid location to have hourly averaged vertical velocity of at least 1.0 Pa s^{-1} on all model levels from 1 to 3 km. An illustration of this criterion is shown in Figure 4, where we visualized the downdraft velocity at different heights for the sub-domain shown in Figure 3d. Note that this criterion could lead to a rejection of tilted downdrafts in particular. To test the sensitivity of our results to this criterion, we repeated our analysis using an alternative criterion which is at same time stricter in identifying downdrafts and updrafts within one level but less strict concerning the vertical overlap. To this end, we increased the threshold for downdrafts and updrafts at a single level but only require the consistent downward motion to be larger or equal to 0.5 Pa s^{-1}. While using this alternative criterion increased the number of detected downdrafts by almost 50%, it had otherwise almost no impact on the results (not shown).

2.2.4. Properties Extracted for Each Convective Event

After identifying the convective events of interest, the corresponding downdraft, updraft and environmental profiles can be calculated for each event. An overview of these properties and how we determine them is given

Table 1
Properties Extracted for Each Convective Event and, if Applicable, the Location and Time at Which the Different Features Are Analyzed Using Hierarchical Clustering, Random Forests, and Multiple Linear Regression

Feature	Physical field(s)	Horizontal	Rep. time	Rep. height	Method
Downdraft area (m^2)	$\omega \geq 1.0\ Pa\,s^{-1}$	(see Section 2)	t.s. at which maximal	1 km	Size of area where the vertical velocity is larger than the threshold
Downdraft velocity ($Pa\,s^{-1}$)	$\omega \geq 1.0\ Pa\,s^{-1}$	$\langle\cdot\rangle_{\text{updraft area}}$	t.s. at which maximal	1 km	
Downdraft mass-flux ($kg\,s^{-1}$)	$\omega \geq 1.0\ Pa\,s^{-1}$	(see Section 2)	t.s. at which maximal	1 km	Downdraft area times downdraft velocity divided by standard gravity
Updraft area (m^2)	$\omega \leq 1.0\ Pa\,s^{-1}$	(see Section 2)	t.s. at which maximal	5 km	Size of area where the vertical velocity is smaller than the threshold
Updraft velocity ($Pa\,s^{-1}$)	$\omega \leq 1.0\ Pa\,s^{-1}$	$\langle\cdot\rangle_{\text{downdraft area}}$	t.s. at which maximal	5 km	
Updraft mass-flux ($kg\,s^{-1}$)	$\omega \leq 1.0\ Pa\,s^{-1}$	(see Section 2)	t.s. at which maximal	5 km	Updraft area times updraft velocity divided by standard gravity
q_r ($g\,kg^{-1}$)	q_r	$\langle\cdot\rangle_{\text{convective event}}$	t.s. at which maximal	3 km	
q_r^{tot} ($g\,kg^{-1}\,m^2$)	q_r	$\int_{\text{convective event}}\cdot$	t.s. at which maximal	3 km	
CAPE/CIN ($J\,kg^{-1}$)	T, p, and RH	$\langle\cdot\rangle_{\text{environmental area}}$	$\langle\cdot\rangle_{-1\text{ to }-0.5\text{ hr}}$	–	Most unstable parcel found between the surface and 400 hPa
DCAPE ($J\,kg^{-1}$)	T, p, and RH	$\langle\cdot\rangle_{\text{environmental area}}$	$\langle\cdot\rangle_{-1\text{ to }-0.5\text{ hr}}$	–	Based on the min. 100 hPa layer averaged Θ_e found in the lowest 400 hPa
RH (–)	RH	$\langle\cdot\rangle_{\text{environmental area}}$	$\langle\cdot\rangle_{-1\text{ to }-0.5\text{ hr}}$	$\langle\cdot\rangle_{1\text{ km–}3\text{ km}}$	
CWV (mm)	CWV	$\langle\cdot\rangle_{\text{environmental area}}$	$\langle\cdot\rangle_{-1\text{ to }-0.5\text{ hr}}$	–	
Γ_v ($K\,km^{-1}$)	T_v	$\langle\cdot\rangle_{\text{environmental area}}$	$\langle\cdot\rangle_{-1\text{ to }-0.5\text{ hr}}$	$\langle\cdot\rangle_{1\text{ km–}3\text{ km}}$	Vertical gradient of T_v times minus one
Lat/lon (∘)					c.o.m.

Note. The environmental area is set to a 200 km × 200 km box centered around the center of mass, and the reference time for averaging is the time where surface precipitation peaks. Note that the center of mass of the convective area is denoted as c.o.m and time-step as t.s.

in Table 1. In particular, downdraft (updraft) profiles are calculated by horizontal averaging of the respective thresholded fields at each altitude level within the areas corresponding to the different convective events. As we consider the time evolution of updraft and downdraft velocity, we use the updraft and downdraft fields from 2.5 hr before to 2.5 hr after the 1 hr time interval used for averaging the precipitation field. Note that we limit our analysis to the last 5 days of the simulation because the increased output frequency during this period allows us to calculate up- and downdraft profiles every 15 min. In addition to the updraft and downdraft profiles, we calculate profiles of the environmental conditions for each event. To this end, we calculate the center of mass of the area of each convective event and horizontally average the environmental fields at each height within a 200 km × 200 km box centered around this point, for an example see Figure 3b. Note that the chosen size of the box represents a trade-off between a box that is too small, for which we would largely capture conditions within the convective clusters themselves, and a box that is too large, for which environmental conditions may not be representative of the environment experienced by the downdraft. To estimate the largest distance from the center of mass of a convective event within which the environmental conditions are consistent, we calculated the radially averaged auto-correlation function for the relative humidity (RH) as well as for the temperature field within the region shown in Figure 3. Based on this, we find that the humidity and temperature of two locations are, on average, uncorrelated when they are farther than about 100 km apart. We therefore chose a box size that extends ±100 km around the center of mass.

We consider here downdraft (updraft) velocity, area and the resulting mass flux separately. Downdraft and updraft area are defined as the area of one grid cell (25.0 km^2) times the number of grid cells where the velocity is larger/smaller than the corresponding threshold value. Downdraft and updraft velocity are averaged within the corresponding areas and given in units of $Pa\,s^{-1}$. Finally, the total mass flux is calculated by dividing either downdraft or updraft velocity by standard gravity (9.81 $m\,s^{-2}$) and then multiplying it with the respective area. We define downdraft mass flux to be positive and updraft mass flux to be negative. For rain-water mixing ratio, we use the area of the convective event (A_c) to calculate the average rain-water mixing ratio, q_r, within this area at each level.

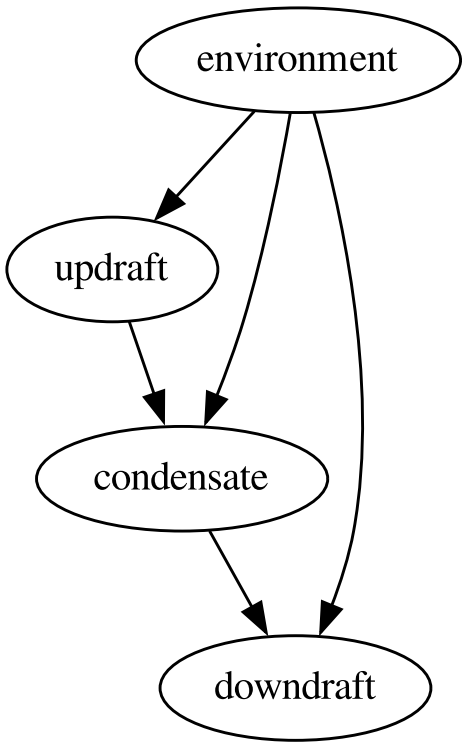

Figure 5. Conceptual model to illustrate the possible causal links investigated here for downdraft formation including direct and indirect environmental effects.

The total rain-water mixing ratio q_r^{tot} is then given by $q_r \cdot A_c$. To characterize the environment, we focus on environmental RH, virtual temperature lapse rate ($\Gamma_v = -\partial_z T_v$), column water vapor (CWV), convective available potential energy (CAPE), convective inhibition (CIN), and downdraft convective available potential energy (DCAPE). CAPE, CIN, and DCAPE are calculated from environmental temperature, pressure and humidity profiles using the SHARPpy module, version 1.4.0, from Blumberg et al. (2017). For the calculation of CAPE and CIN we use the most unstable parcel found between the surface and 400 hPa, that is, we set the argument *flag* in the parcel routine of SHARPpy equal to three. DCAPE in SHARPpy is calculated based on the minimum 100 hPa layer averaged Θ_e found in the lowest 400 hPa of the sounding. Finally, we also consider the geographic location of the center of mass of the event (*lon* and *lat*).

2.3. Methods to Analyze Different Controls on Downdraft Strength

Based on the literature review in Section 1, we have constructed a causal model for downdraft formation, shown in Figure 5. In this model environmental conditions lead to the formation of updrafts, which then—again depending on environmental conditions—lead to the formation of condensate. Finally, condensate and the given environmental conditions can cause the formation of downdrafts. We hereby refer to the effect of the environment through its effect on the updraft and condensate as an *indirect* environmental effect, and its effect directly on the downdraft as a *direct* environmental effect. Before discussing the physical principles on which this model is based, it is important to note that this is already a simplification of the actual process that occurs in nature. For example, the model does not account for any changes in the environmental conditions caused by the updraft which then affect the environmental conditions for the downdraft.

A few causal relationships are suggested by past work (see Section 1). First, higher environmental instability, as measured for example, by the lapse rate, should clearly favor stronger updrafts and downdrafts. However, the influence of ambient humidity is less straightforward. It has been shown that deeper clouds and stronger updrafts tend to form in more humid environments. A simple explanation is that entrainment of moist air has much less negative impact on updrafts than entrainment of dry air. However, two opposing processes have been suggested for a direct effect of humidity on downdrafts. On the one hand, drier environments reduce the density difference between the environment and the downdraft (leading to weaker downdrafts); on the other hand, entrainment of dry air promotes evaporation, which increases the density of the downdraft air (leading to stronger downdrafts). While we focus here only on these two environmental characteristics (measured in various quantities), this is not an exclusive list. For example, other studies have investigated the effect of wind shear (e.g., Marion & Trapp, 2019). In addition to these known unknowns, there is of course always the possibility of unknown unknowns. Here, we attempt to address the problem of unknown unknowns, at least in part, by also considering the spatial location of the convective events. If we found that the location of the convective event is an important predictor, this would indicate that we have missed an important environmental variable that varies with longitude and latitude.

In addition to the environmental effects we also briefly summarize the expected relationship between updraft and condensate on the one hand and downdraft on the other, based on the results of Trapp and Woznicki (2017) and Marion and Trapp (2019). Following the precipitation formation argument presented in Trapp and Woznicki (2017), we expect that a crucial link between updraft, condensate, and precipitation is the updraft area. The reason is that a larger updraft area leads to a larger precipitation area, which in turn leads to a more extended downdraft. Since this explanation focuses only on the extent of the downdraft and not its velocity, we will examine the dependence on downdraft velocity separately. We expect that more condensate, that is, a higher rainwater mixing ratio, will lead to more evaporation and condensate loading and thus to stronger downdrafts.

In the following, we apply two different machine learning algorithms as well as multiple linear regression to analyze how the detected downdrafts depend on the respective updraft and the environment. In accordance with the parlance of machine-learning, we refer to each convective event as one *instance* of our data set and to the

properties of a given convective event as the *features* of this instance. We first use a method called *Hierarchical Clustering* to identify clusters of strongly correlated features. Next, using an alternative method called *Random Forests*, we derive a model that allows us to predict the features of the downdrafts (i.e., our target features) from the updraft and environmental features (i.e., our input features or predictors) and examine their relationship. We will briefly discuss each of these methods here; for more details see for example, Howard and Gugger (2020). Finally, we use multiple linear regression, that is, a well-known method from classical statistics, to test the machine learning results and derive a simple, linear model for predicting downdraft features. Feature hierarchy, random forests, and multiple linear regression are calculated using the scikit-learn library version 0.22.1 (Pedregosa et al., 2011) in Python version 3.6.10 (e.g., Sanner, 1999).

Using the method described in Section 2.2, most features in Table 1 are available for a range of time and height values. To simplify the subsequent analysis, we represent each feature for each instance by a representative value as summarized in Table 1, where $\langle\cdot\rangle_{t_{\min}\ \text{to}\ t_{\max}}$ again denotes temporal averages for the time interval $[t_{\min}, t_{\max}]$ and $\langle\cdot\rangle_{h_{\min}\ \text{to}\ h_{\max}}$ denotes vertical averages for the height range $[h_{\min}, h_{\max}]$. For each downdraft, updraft, and condensate feature, we choose, for example, the maximum (absolute) value at a specified height level, that is, the value at the time step where the absolute value of the respective variable is maximum. We test the sensitivity of our results to the chosen height of the downdraft features in Section 3.2. The environmental and parcel variables are averaged in time and (if applicable) in height. In particular, we use time averages of the 30 min time interval before the interval considered for the precipitation field, to ensure that the environmental variables are not modified by the convection itself. To ensure that our results are not sensitive to the chosen averaging times and heights for the environmental and parcel features, we also used a range of different starting times ($t_{\min}$ in the range −3.0, −2.0, and −1.0 hr) and, if applicable, heights ($h_{\min}$ in the range 1.0, 2.0, and 3.0 km and $h_{\max} = h_{\min} + 1.0$ km). Including all the differently averaged environmental and parcel quantities as additional input features for predicting the downdraft features, we found neither a significant increase in skill nor a change in the importance of the key type of input feature.

2.3.1. Hierarchical Clustering

The aim of hierarchical clustering is to create a ranking of how similar different features are to each other. In our case, the similarity of two features is given by the absolute value of the Spearman's rank correlation coefficient $|r_s|$. The clustering is performed in a successive and agglomerative manner. First, the two features which have the highest correlation coefficient are combined in one cluster. Second, for all remaining features and the cluster created in the first step, the correlation coefficients are calculated for each possible pair and, again, the pair with the highest correlation is combined into one cluster. Note that the r_s of a feature and a cluster is calculated as the average of the pairwise calculated r_s between the single feature and the individual features in the cluster. This process is continued until all features have been combined into a single cluster. The key information provided by hierarchical clustering is usually in the highly correlated clusters, that is, the clusters that form at high $|r_s|$. If we want to understand how a target feature (e.g., downdraft velocity) depends on a set of input features (e.g., updraft and environmental properties), it is useful to distinguish between highly correlated clusters with and without the target feature. If the target feature is part of a cluster, one can conclude that the input features in that cluster are important to the target feature. In contrast, if the target feature is not in a cluster, we can try to reduce the total number of input features (and thus the complexity of the problem) by selecting a representative feature from that cluster. Finally, it can also be interesting to test our expectations by noting whether the features we expect to be highly correlated actually end up in the same, highly correlated cluster. In the following, for instance, we use the results of hierarchical clustering to decide which downdraft properties need to be investigated separately and which are so closely correlated that they can be considered together.

2.3.2. Random Forest

A random forest is a model which is trained to predict a target feature (in our case downdraft properties) from a set of input features (updraft and environmental properties). To understand how a random forest is trained and how it makes its predictions, we first need to look at decision trees, as they are the key ingredient of the model. In a decision tree, a prediction is made by successively answering a set of *yes* and *no* questions with respect to the input features. Each question is referred to as a branch of the decision tree. This can best be illustrated using the example of a very small decision tree, see Figure 6, where we predict the downdraft velocity from the updraft velocity by asking two questions. The first question is always "Is the updraft velocity smaller or equal to −9.5 Pa s^{-1}?" If the answer to this question is *True*, the second question is "Is the updraft velocity smaller or

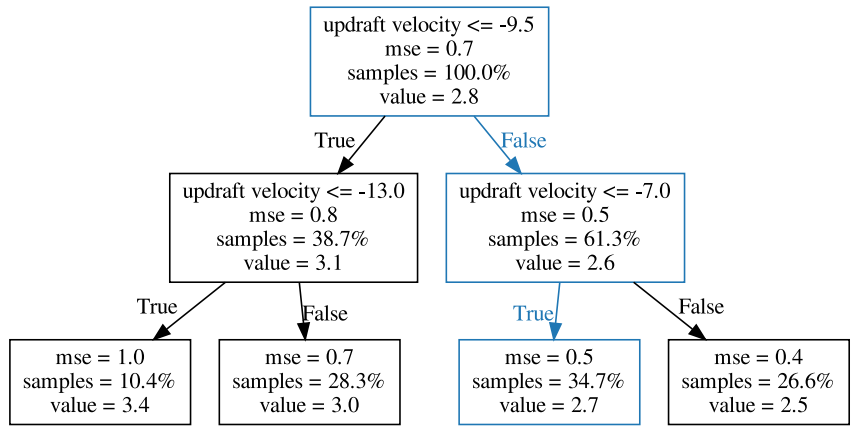

Figure 6. A sample decision tree which predicts downdraft velocity from updraft velocity, where *value* is the predicted downdraft velocity, *samples* is the fraction of instances which fall into this category and *mse* stands for the total mean squared error of all instances in this sample.

equal to −13.0 Pa s^{-1}?" If the answer to the first question is *False*, the second question is "Is the updraft velocity smaller or equal to −7.0 Pa s^{-1}?" Depending on the updraft velocity, this tree will predict a downdraft velocity of either 3.4, 3.0, 2.7, or 2.5 Pa s^{-1}. For example, a downdraft velocity equal to 2.7 Pa s^{-1} would be predicted for all convective events which have an updraft velocity larger than −9.5 Pa s^{-1} (answer to first question is *False*) and smaller than −7.0 Pa s^{-1} (answer to second question is *True*). This scenario has been highlighted in blue in Figure 6.

If we know what the true downdraft value is, we can test how well the downdraft was predicted by the model by comparing it to the true value, for example, by calculating the squared difference between the two. While the number of questions to ask, the so called depth of the tree d_{tree}, is specified in advance, which questions to ask in which order is learned by the model during training. During training, we pass a subset of all instances, that is, convective events, to the model (the so called *training data set*). The resulting decision tree is the model where the mean squared error between all the predicted values of the target feature, that is, the values we obtain after answering all questions for each element in the training data set, and the actual values of the target feature is smallest. The mean squared error of the final model can, for example, be calculated by multiplying the mean squared error within each leaf with the fraction of instances in this leaf and then summing over all leaves. For the decision tree shown in Figure 6, for example, the mean squared error is approximately 0.6. The simple decision tree therefore reduced the mean squared error of the climatological forecast, here equal to 0.7, by about 15% ($0.6/0.7 \approx 85\%$). To finally assess the performance of the model, we use all instances not used during training, the so called *validation data set*. For each instance in the validation data set, we use the model to predict the value of the target feature and then compare the predicted values to the true values. Note that either in addition to, or even instead of, d_{tree} we can also specify a minimum number of convective events (N_{min}) which have to fall into each of the final categories or "leaves." If only N_{min} is specified, the maximum number of branches is reached when each new question would lead to a final leaf containing fewer than N_{min} instances.

While a single decision tree can already predict the target feature from a set of input features, it has been shown by Breiman (2001) that a more robust model is obtained by training an ensemble of decision trees and using the mean outcome as prediction. A model made up of an ensemble of decision trees, trained on randomly chosen subsets of the data, is known as a random forest. Breiman (2001) showed that, in general, an ensemble approach can improve the accuracy of any prediction method, if this method is unstable, that is, when small changes in the training data set can lead to different predictions. In the case of decision trees, for example, this is the case when many trees fit the training data set about equally well. This problem is circumvented in the ensemble approach by training each of the decision trees in the ensemble (N_{trees}) on a given number of instances from the training data set ($N_{samples}$) using random sampling. As noted above, the prediction of a random forest is then given by the average prediction of all individual trees. As for a single decision tree, the performance of the model is assessed by calculating the total mean squared error between the predicted and the actual values of the target feature in the validation set.

To train a random forest on a given data set one therefore has to make the following decisions. First, one has to specify the properties of the decision trees by choosing their depth d_{tree} and/or minimum number of instances in each leaf (N_{min}). Second, one has to decide on the number of decision trees in the random forest (N_{trees}), the number of training instances used for the individual trees (N_{samples}) and how these are chosen (e.g., random sampling with or without replacement). While this seems like a large set of choices, one key advantage of random forests is that they are very robust with respect to these choices (e.g., Howard & Gugger, 2020); in particular, increasing the number of trees does not lead to overfitting.

For our purpose, it is important to note that random forests can not only be used to make predictions but, by understanding how these predictions were made, we can try to learn something about how the target feature depends on the input features. To this end, we will assess the *feature importance* as well as the *partial dependence* once we have trained our random forest. Feature importance can be used to assess which input features are important for predicting the target feature and which are not. In particular, feature importance quantifies how much improvement in the model results from decisions made regarding each of the input features separately. This is quantified by determining for each input feature, what fraction of the total reduction in the mean squared error of the model, compared to the climatological forecast, results from decisions made regarding this feature. While feature importance reveals which of the input features are most relevant for predicting the target feature, *partial dependence* quantifies how the target feature depends on the input features. The key idea is to measure how the mean value of the target feature, as predicted by the random forest, changes if we artificially change the values of one of the input features.

Specifically we use the trained model, keeping the values of all input features as they were, except for the values of the one input feature whose influence we want to discover. For this selected input feature, we replace the original values throughout the data sample by a fixed value stepped through a range. The resulting changes in prediction reveal the impact of that input feature with others held fixed.

2.3.3. Multiple Linear Regression

In multiple linear regression, the residual sum of squares between the input features, that is, the updraft and environmental properties, and the target features, that is, the downdraft properties, is minimized by fitting a linear model. Whether linearity can be assumed between the input and target variables can be seen, at least to some degree, from the partial dependence plots described above.

How much of the variation in downdraft properties can be explained by the variation in input features is again indicated by the coefficient of determination. The contribution of the various updraft and environmental properties to the coefficient of determination is also a robust measure of their individual importance in predicting the downdraft properties. In particular, one can assess how much an individual feature contributes to the coefficient of determination by removing that feature from the set of input features and determining by how much the coefficient of determination decreases as a result. Note, however, that the presence of highly correlated predictors makes it difficult to generalize these results. For example, removing a single input feature but keeping features that are highly correlated with that feature means that the contribution of the single input feature can be compensated, at least to some extent, by the remaining features. In the following, we will therefore also remove entire sets of highly correlated input features as identified by hierarchical clustering in order to evaluate their relative importance.

3. Results

3.1. Mean Properties of Updrafts and Downdrafts

Applying the above described method to our data set, we detect more than 30,000 convective events. The location of the detected events is shown in Figure 2. As expected, the convective events are found in areas known to have strong deep convective activity/high precipitation rates.

The time evolution of surface precipitation rate, rain water mixing ratio, updraft mass flux, and downdraft mass flux during a composite event is shown in Figure 7. Note that for each detected convective event we define the time step of peak precipitation as time zero before averaging. By construction, peak precipitation rates therefore occur at time zero, which is also the time at which rain water mixing ratio maximizes at a height of about 3.5 km (Figure 7a). As expected, updraft mass flux peaks before, and downdraft mass flux peaks after, surface

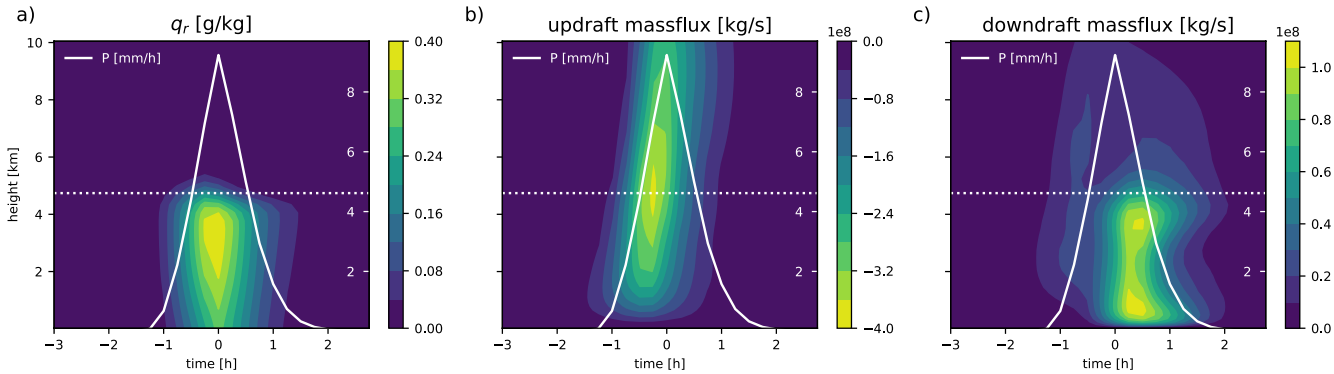

Figure 7. Composite average time evolution of the precipitation rate (solid white line) and (a) rain water mixing ratio, (b) updraft mass flux, and (c) downdraft mass flux (colors), with respect to the time of peak precipitation (0 hr) calculated over all detected events. The dotted white line is the average height of the freezing level (4.7 km).

precipitation. In fact, maximum updraft mass flux most frequently peaks 15 min (corresponding to one output time-step) before and maximum downdraft mass flux 15 min after the time of peak precipitation. The updraft mass flux (Figure 7b) shows a single, pronounced maximum just above the freezing level at a height of about 5 km. The vertical profile of the downdraft mass flux shows a two peaked structure (Figure 7c). One peak occurs close to a height of 3.7 km, about 1 km below the freezing level, and a second, more pronounced peak, closer to the surface at about 1.5 km.

3.2. Controls on Downdraft Strength

3.2.1. Hierarchical Clustering

The result of hierarchical clustering (see Section 2.3) applied to our data set can be seen in form of a dendrogram in Figure 8. The first thing we note is that the downdraft, updraft, and condensate variables are all contained in two, basically uncorrelated, clusters. In the following discussion we restrict ourselves to the discussion of these two clusters, highlighted in blue and orange. First, the blue cluster which contains total updraft mass flux, total amount of rain-water $\left(q_r^{\mathrm{tot}}\right)$, updraft area, downdraft mass flux, and downdraft area. We note that this cluster contains, of all the considered variables, the most strongly correlated variables with particularly strong correlation between, on the one hand, the updraft and condensate variables, and, on the other hand, the downdraft variables. Second, the orange cluster which contains updraft velocity, mean rain-water mixing ratio and downdraft velocity. As in the blue cluster, updraft velocity is best correlated with the rain related variable, followed by the downdraft variable. Compared to the blue cluster, however, the correlation coefficients are smaller. Note that the blue cluster contains all updraft/downdraft variables which depend on the area of the convective event while the orange cluster contains all updraft/downdraft variables independent of the area. In analogy to thermodynamics, we refer to the variables in the blue cluster as *extensive* variables and the variables in the orange cluster as *intensive* variables. Thus, Figure 8 shows that in our study the extensive and the intensive variables are almost uncorrelated to each other. We also note that within the two clusters the updraft variable(s) are most correlated to the precipitation variable and, together, they are then most correlated to the downdraft variable. The degree of correlation is, however, significantly higher for the extensive than the intensive variables.

This analysis suggests two things regarding our initial question of what determines downdraft mass flux. First, downdraft mass flux and area are most correlated with the variables updraft area, mass flux and the total amount of rain-water, q_r^{tot}. This result agrees well with the precipitation generation argument by Trapp and Woznicki (2017). Second, downdraft mass flux and velocity (as well as updraft mass flux and velocity) are not strongly correlated to each other. As the total mass flux is the product of velocity and area, this result suggests that the coupling between updraft and downdraft happens via the area, again in agreement with Trapp and Woznicki (2017). To further analyze the respective importance of the downdraft area and velocity on the downdraft mass flux, we trained a random forest that predicts the downdraft mass flux from the downdraft area and the downdraft velocity. We find that the downdraft area has a feature importance of 87% and the downdraft velocity of 13%. This confirms that in our study the downdraft mass flux is mainly controlled by the area rather than the velocity. In the

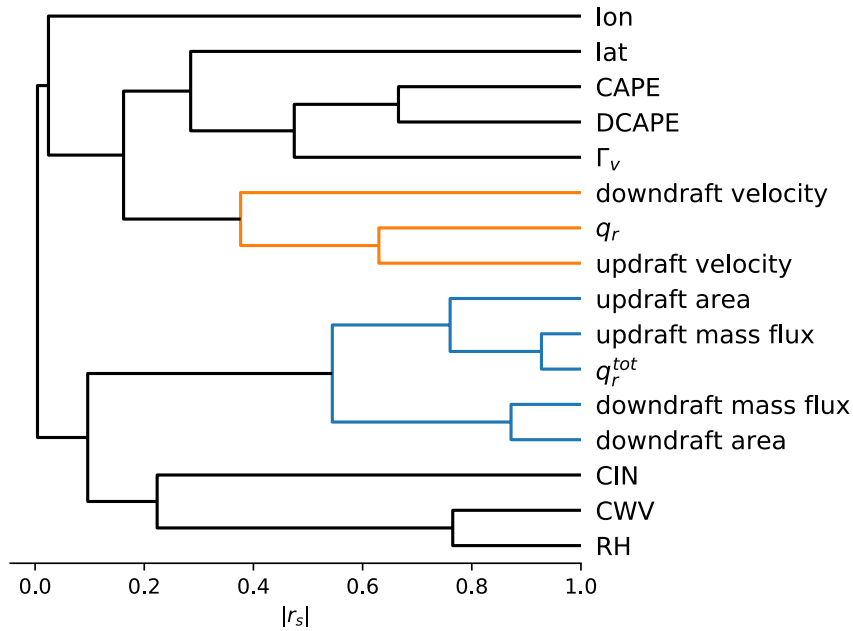

Figure 8. Dendrogram showing the hierarchy of clusters for the different features detected for each convective event and the corresponding absolute value of the Spearman's rank correlation coefficient $|r_s|$. The cluster containing the extensive (intensive) updraft/downdraft properties is marked in blue (orange).

following, we will use a random forest approach to investigate what controls downdraft mass flux and velocity. Based on the results above, we will investigate what controls these two downdraft properties separately.

3.2.2. Random Forests

We train two random forests, as described in the previous section, one for downdraft mass flux and one for downdraft velocity. Excepting the downdraft properties, all properties in Table 1 are used as input features. We start by randomly splitting the $N = 31{,}340$ convective events into a training data set, $N_{\text{train}} = 21{,}938$ events (70% of the data), and a validation data set, $N_{\text{valid}} = 9{,}402$ events (30% of the data). The two random forests we train consist of 50 decision trees each, that is, $N_{\text{trees}} = 50$. For the decision trees, we do not prescribe the depth of the tree but only require that each of the final categories contains at least $N_{\text{min}} = 50$ instances from the training data set. For sampling, we use random sampling with replacement (i.e., bootstrap sampling) and the number of samples drawn from the training data set for each tree is $N_{\text{samples}} = N_{\text{train}}$. This means that the estimated fraction of unique events used for each tree is about 63%, as the expected fraction of unique elements when drawn k times from a sample of size N is, following standard statistics, equal to $1 - (1 - 1/N)^k$ with, in our case, $N = N_{\text{samples}}$ and $k = N_{\text{train}}$. Using the resulting random forests to predict the target feature in the validation data set we find that the random forests can predict downdraft mass flux with a coefficient of determination of 76% and downdraft velocity of 37%. Thus, while there is only moderate predictability in downdraft velocity from the input features, there is high predictability for downdraft mass flux.

The feature importance factors of the individual input features for the two random forests are shown in Figure 9. By far the most important input feature for predicting downdraft mass flux is the total amount of rain water, and for downdraft velocity the average rain water mixing ratio. This result does not suggest that the mass flux or velocity of the updraft is irrelevant to the corresponding downdraft quantity, but rather shows that the essential role of the updraft is to determine how much rain water is available. One way of showing this more explicitly is by removing q_r^{tot} as an input feature from the random forest predicting downdraft mass flux, after which we find that the value of R^2 decreases only by 3%. In contrast, if we train a random forest which has neither q_r^{tot}, nor the updraft mass flux, nor the area as input feature, R^2 drops to 27% (−49%). Taken together, this indicates that when predicting downdraft mass flux it is better to know how much rain water is available than to know how strong the updraft was, but also, that much of the information about q_r^{tot} is already contained in the updraft mass flux. A similar, though less pronounced, picture emerges when removing q_r from the set of input features for the

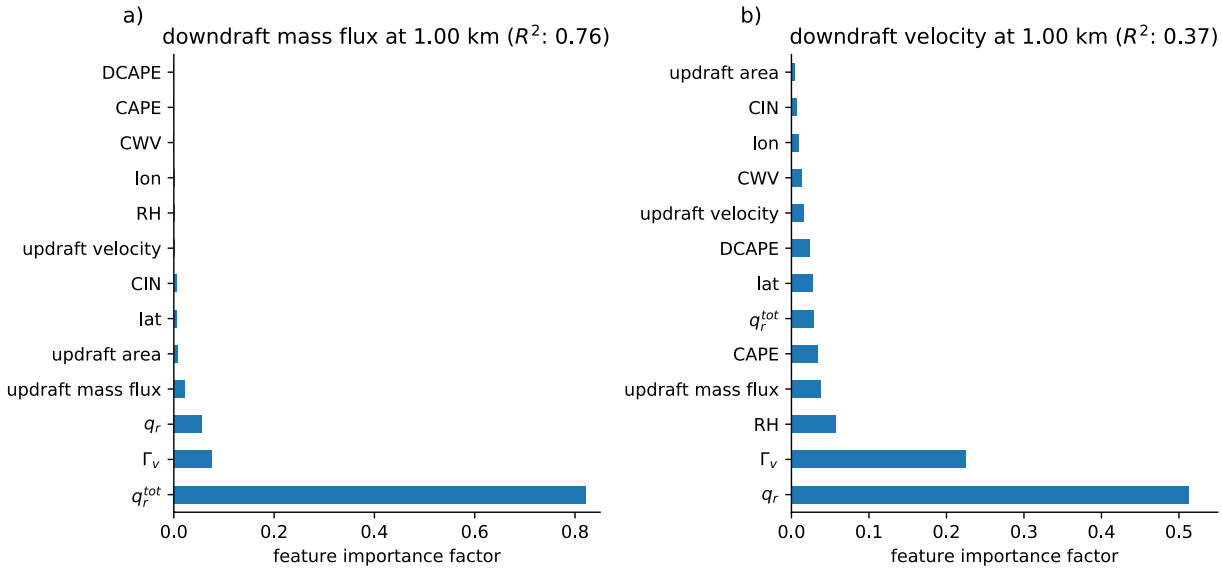

Figure 9. Feature importance factors for all input features for the random forest predicting (a) downdraft mass flux and (b) downdraft velocity.

random forest predicting downdraft velocity. Again, R^2 decreases but again only by a small fraction of 5%, while removing all updraft and precipitation related properties at the same time leads to a R^2 value of only 25%. Thus, the following picture emerges: downdrafts are mainly controlled by the availability of rain water, which, in turn, strongly depends on the preceding updraft. These results, therefore, support the precipitation generation argument from Trapp and Woznicki (2017). Considering the environmental properties, the lapse rate appears to play some role in controlling downdraft mass flux as well as velocity, while we find environmental RH to be of very little importance. Strikingly, there is also very little dependence on any of the parcel related properties. Finally, note that the weak dependence on the location of deep convection suggests that we do not miss any important environmental property which systematically varies with geographic location.

To understand how downdraft mass flux and velocity depend on a set of input features, Figure 10 shows the partial dependence on the amount of rain water, environmental lapse rate, and environmental RH. To facilitate comparison of the results for downdraft mass flux and downdraft velocity, we normalize each of the two variables by first subtracting the mean value and then dividing by the corresponding standard deviation. As expected from the feature importance plot, the dominant relationship is between rain water and downdraft mass flux. In particular, more rain water present at a height of 3 km means stronger downdrafts at a height of 1 km. This relationship is much stronger for downdraft mass flux compared to downdraft velocity. The dependence on environmental lapse rate is shown in Figures 10b and 10e. While we do find that stronger downdrafts form in regions with higher lapse rates, that is, more unstable environments, the dependence is weak compared to the dependence on rain water. For environmental RH, no dependence of downdraft mass flux is visible in Figure 10c and only a very weak dependence of downdraft velocity (faster downdrafts in drier environments) can be seen in Figure 10f.

3.2.3. Multiple Linear Regression

Next, we use multiple linear regression to fit two linear models, one for downdraft mass flux and one for downdraft velocity. Note that the base MLR assumption of linear relationships is supported by the partial dependence results in Figure 10, at least for the shown variables. The resulting R^2 values are 75% for downdraft mass flux and 36% for downdraft velocity (Figure 11). These values are, therefore, almost identical to the ones we found for the decision tree method above (76% for downdraft mass flux and 37% for downdraft velocity). The corresponding regression coefficients are shown in Table 2. Considering only the signs of the coefficients, we find that we again get stronger downdraft mass flux and velocities for more unstable and drier environments. Stronger downdraft mass flux for larger values of the extensive updraft variables and stronger downdraft velocities for the intensive updraft variables.

As described in Section 2.3, we determine the relative importance of the different variables by assessing how much they contribute to the coefficient of determination. Considering all variables individually, we find that the strongest reduction in R^2 for downdraft mass flux results from the removal of q_r^{tot} (3.1%) and the strongest

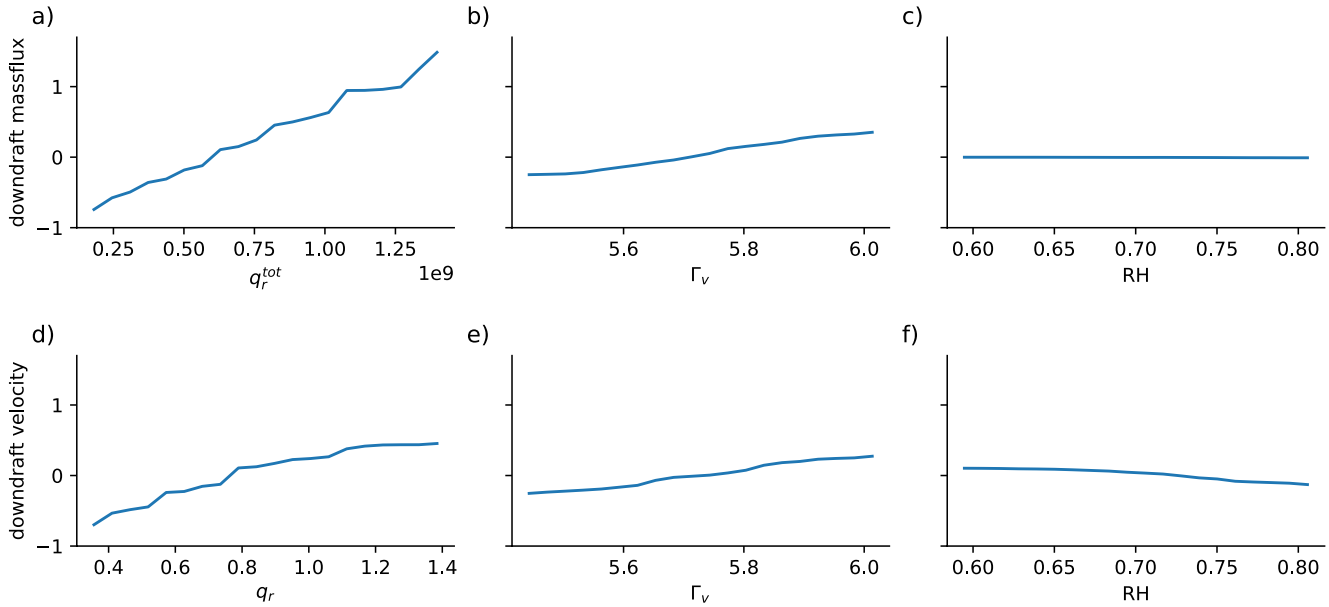

Figure 10. Partial dependence plots for the dependence of normalized downdraft mass flux on (a) total amount of rain water, (b) environmental lapse rate, (c) environmental humidity. The same for normalized downdraft velocity on (d) rain water mixing ratio, panels (e and f) same as panel (b and c).

reduction in R^2 for downdraft velocity from the removal of q_r (5.5%). As the results from hierarchical clustering show, some variables are so strongly correlated that we should really assess their combined effect on R^2. To this end, we have divided the dependent variables into four categories of strongly correlated variables. In particular, we divide our variables into *extensive* variables (updraft mass flux, updraft area, q_r^{tot}), *intensive* variables (q_r, updraft velocity), *parcel* variables (CAPE, CIN, and DCAPE), *humidity* variables (RH, CWV), and the *stability* variable (Γ_v). This division is, of course, not unique but it is best suited to test the key results found above.

The results are shown in Figure 11. In agreement with our results above, we find that the extensive updraft variables are the most important variables for downdraft mass flux, the intensive updraft variables the most

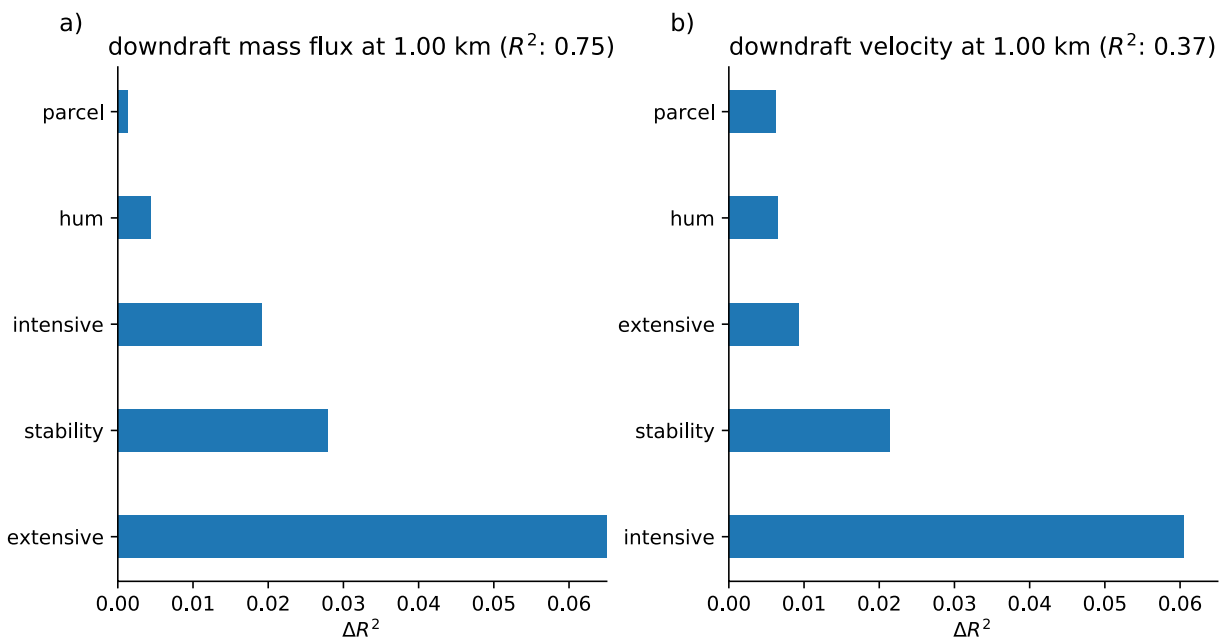

Figure 11. Reduction of R^2 caused by excluding different categories of input features in the multiple linear regression model for (a) downdraft mass flow and (b) downdraft velocity.

Table 2
Multiple Linear Regression Coefficients

	Unit	Downdraft mass flux	Downdraft velocity
Γ_v	$\mathrm{K\,km^{-1}}$	1.23	1.08
q_r^{tot}	$\mathrm{g\,kg^{-1}\,m^2}$	1.52×10^{-9}	-6.42×10^{-10}
q_r	$\mathrm{g\,kg^{-1}}$	-5.75×10^{-1}	9.81×10^{-1}
CAPE	$\mathrm{J\,kg^{-1}}$	-7.11×10^{-5}	4.45×10^{-4}
CIN	$\mathrm{J\,kg^{-1}}$	-3.59×10^{-3}	-9.47×10^{-3}
CWV	mm	2.57×10^{-2}	-2.42×10^{-3}
DCAPE	$\mathrm{J\,kg^{-1}}$	-7.66×10^{-5}	6.19×10^{-5}
RH	–	−1.65	−1.54
Updraft area	$\mathrm{m^2}$	7.15×10^{-10}	-3.99×10^{-10}
Updraft mass flux	$\mathrm{kg\,s^{-1}}$	2.50×10^{-10}	-1.46×10^{-9}
Updraft velocity	$\mathrm{Pa\,s^{-1}}$	-2.20×10^{-2}	3.50×10^{-3}

important variables for downdraft velocity, and the environmental stability is again the most important environmental variable. This confirms the robustness of the results from the random forest approach, but also suggests that the key dependencies of downdraft mass flux and downdraft velocity can be well approximated by linear models.

We therefore derived a set of strongly simplified linear models for downdraft mass flux

$$m_{\mathrm{down}} = -0.3 m_{\mathrm{up}} + 1.7 \times 10^7 \mathrm{kg\ s^{-1}} \qquad \text{with } R^2 = 0.59 \tag{1}$$

$$m_{\mathrm{down}} = -0.3 m_{\mathrm{up}} + 9.4 \times 10^7 \mathrm{kg\ km\ K^{-1}\ s^{-1}} \Gamma_v - 51.6 \times 10^7 \mathrm{kg\ s^{-1}} \qquad \text{with } R^2 = 0.62 \tag{2}$$

and downdraft velocity

$$\omega_{\mathrm{down}} = -0.1 \omega_{\mathrm{up}} + 1.9 \mathrm{Pa\ s^{-1}} \qquad \text{with } R^2 = 0.13 \tag{3}$$

$$\omega_{\mathrm{down}} = -0.1 \omega_{\mathrm{up}} + 1.5 \mathrm{Pa\ km\ K^{-1}\ s^{-1}} \Gamma_v - 6.6 \mathrm{Pa\ s^{-1}} \qquad \text{with } R^2 = 0.25. \tag{4}$$

Starting with a simple linear regression on a small subset of the considered input features to model the relationship between downdraft mass flux and updraft mass flux, we can already explain nearly 60% of the variability in downdraft mass flux, see Equation 1. Using multiple linear regression, we calculate a model which includes environmental stability in addition to updraft mass flux, see Equation 2. The corresponding minimal model for the dependence of downdraft velocity on updraft velocity is given by Equation 3 and the minimal model for the dependence of downdraft velocity on both, updraft velocity and environmental stability, is given by Equation 4.

3.2.4. Downdraft Properties at Different Heights

So far, we have restricted our discussion on the prediction of downdraft mass flux and velocity at $z = 1$ km. We now partly repeat the analysis above for downdraft properties at heights of 1.5, 2.0, and 2.5 km. The results, including the respective values of R^2, are shown in Figure 12. Note that, apart from the considered height of the downdraft properties, we use the same settings for the random forests and also use the same convective events for the training and the validation data set.

The most striking feature comparing Figures 9–12 is that the relative importance of rainwater and updraft mass flux for predicting downdraft mass flux changes dramatically with downdraft height. For downdrafts above 1.5 km, the most important predictor for downdraft mass flux becomes the updraft mass flux, with q_r^{tot} showing relatively little importance by 2.5 km, whereas it was quite dominant at 1.0 km. Strangely however, for downdraft velocity, q_r dominates at all levels with almost no height-dependence. These results imply that downdraft area is less predictable at high levels, potentially because mechanical forcing of downward motion outside of updrafts is important at higher levels but difficult to predict. At lower levels the limiting role of rain water may become more important as the effect of evaporation has had time to accumulate, leading to an increase in predictability for downdraft mass flux as it becomes more and more closely linked to the available rain water $\left(q_r^{\mathrm{tot}}\right)$. It is not clear why the downdraft velocity does not also become more predictable; it may depend on unmeasured factors that vary much less with height.

For the downdraft velocity, although q_r is equally dominant at all heights, we note that the relative importance of the (secondary) environmental properties does change with height. While the importance of the atmospheric lapse rate decreases with height, the importance of both CAPE and DCAPE increases with height. One explanation for the decreased importance of the environmental lapse rate is that, for consistency, we still average Γ_v between $h_{\mathrm{min}} = 1$ km and $h_{\mathrm{max}} = 3$ km. We expect, however, that it is the environmental stability *above* the level of the downdraft that is most important for downdraft velocity. As Γ_v does not necessarily represent the conditions above 3 km, in particular as the variability in Γ_v is expected to be greatest at low levels, this might increase the relative importance of the more general measures of environmental stability, CAPE and DCAPE.

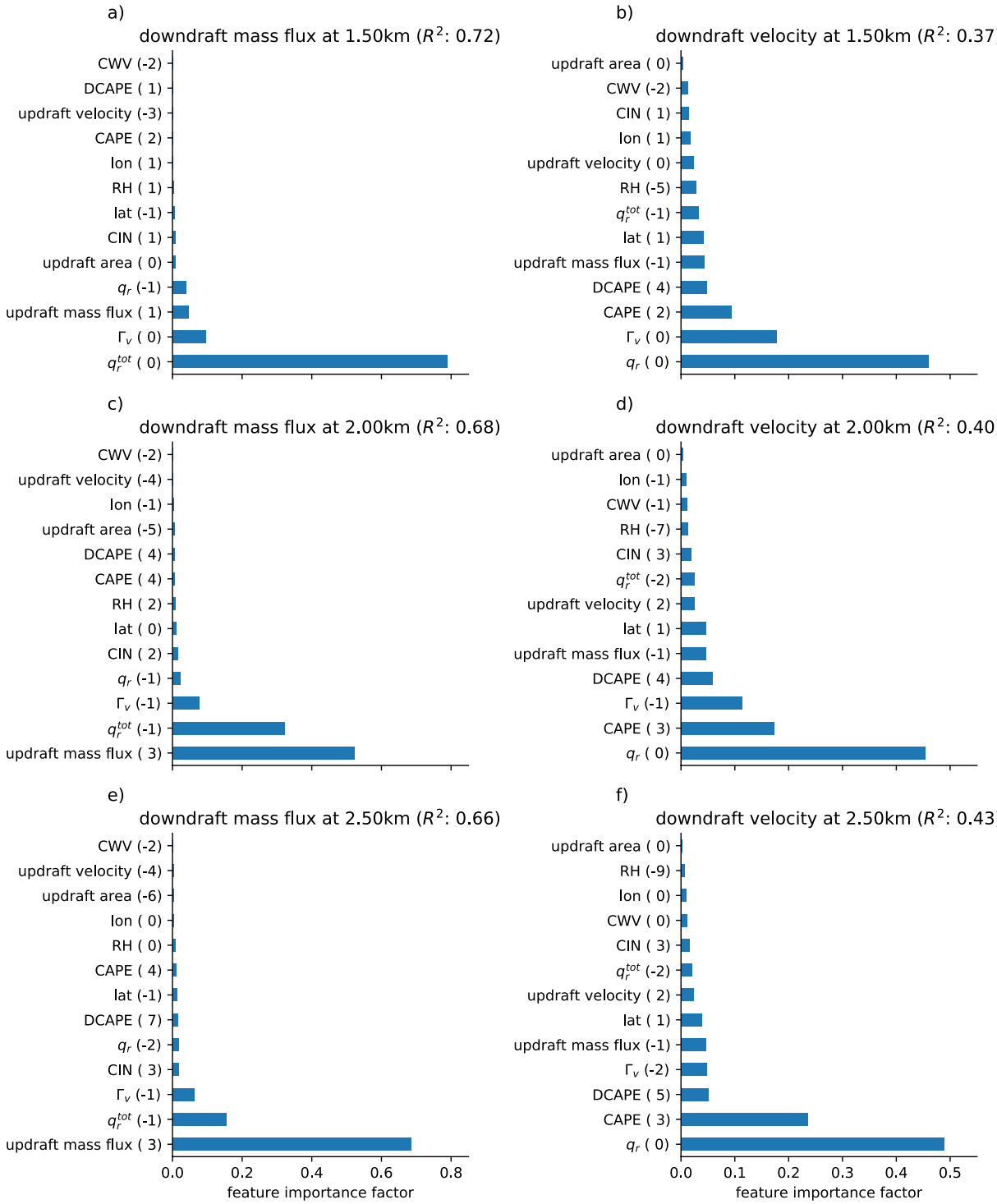

Figure 12. As in Figure 9 but for downdraft mass flux and velocity measured at a height of 1.5, 2.0 and 2.5 km (top row, middle row, and bottom row). Limited to the 10 most important input features. The numbers in brackets indicate how the rank of each variable changed compared to the rank at height level 1 km.

4. Discussion

Considering the above result that the mass flux and velocity of the downdraft strongly depend on the availability of rain water, it is surprising that the RH of the environment has such a small influence on the properties of the downdraft. As explained in the introduction, the influence of environmental humidity on the strength of the

downdraft is unclear. In particular, we noted that two opposing mechanisms have been discussed in the literature: more humid air entrained into the parcel means less evaporation and latent cooling, but also reduces the environmental air density relative to the downdraft. While in the cited studies the initial downdraft conditions were set independently of the environmental humidity, we note that the environmental humidity could also affect the initial downdraft properties. Specifically, for a given rainfall rate, one would expect higher environmental humidity to result in a higher initial moisture content or size of the downdraft and thus enhance the downdraft properties. There are, therefore, at least three mechanisms by which environmental humidity affects downdraft properties: one mechanism through which higher environmental humidity reduces downdraft properties, and two mechanisms through which environmental humidity enhances downdraft properties, one more relevant near the formation point and one more relevant in the boundary layer. As these different effects are of different importance at different altitudes, one could argue that they cancel each other out when averaging over a certain layer. In this case, however, we would expect a stronger (and opposite) signal when considering environmental humidity at different heights, which is not the case (not shown).

One possible explanation is simply that the humidity through which the raindrops fall is not well captured by the ambient humidity measure used above. In particular, the local moisture content relevant for the downdraft could be strongly influenced by the preceding updraft rather than the large-scale RH prior to the formation of that updraft. This should be different for the virtual temperature (or virtual lapse rate), as gravity waves can efficiently remove density differences. To test this hypothesis, Figure 13 shows the time evolution of local and environmental RH and lapse-rate from the numerical simulations. The environmental properties are calculated as previously described in Table 1 and the local properties are calculated by averaging over the area of the respective convective event. To test how the difference between the environmental and local properties compares to the overall variability of the considered variable, we scale the *y*-axis in both plots in Figure 13 such that it ranges from the 3rd to the 97th percentile of the environmental variable.

Figure 13a shows that the environmental RH remains approximately constant over time, while the RH in the area of the convective event increases around the time of maximum precipitation. While it is not surprising that the formation of a convective cloud is accompanied by an increase in RH, it highlights how different the local humidity content, that is, the humidity "seen" by the downdraft, is from the environmental humidity. This pronounced local increase in humidity associated with convective clouds also raises the question of how the presence of other convective clouds affects the humidity encountered by the downdraft. Although beyond the scope of this study, it would be interesting to investigate how the characteristics of the surrounding convective events impact the downdraft properties. Compared to the time evolution of the two RH variables, the difference between environmental and local lapse rates is much weaker (see Figure 13b), that is, the lapse rate relevant to the downdraft is well captured by the environmental variable. To further test this result, we used a random forest to predict the

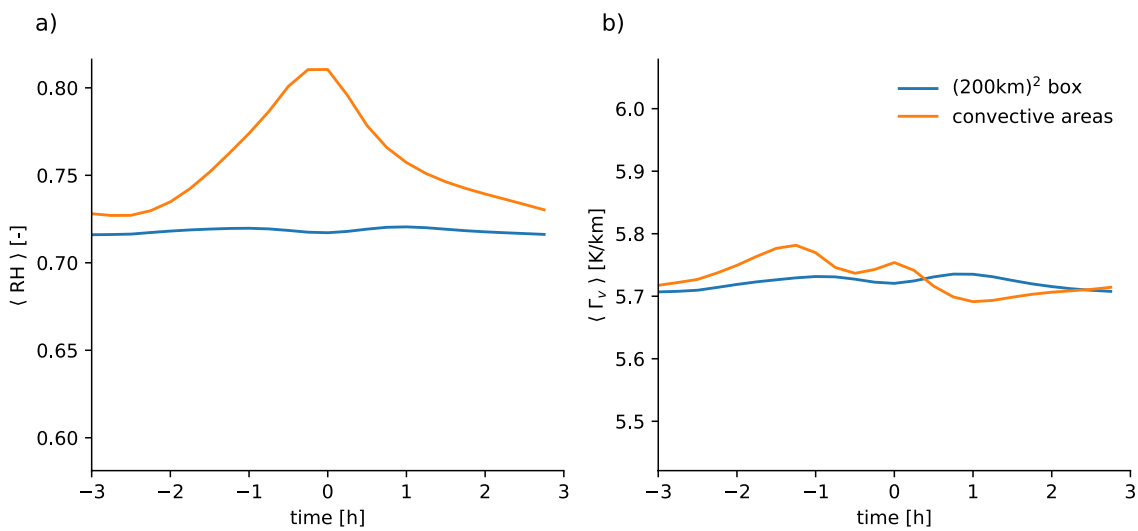

Figure 13. Time evolution of the average (a) relative humidity and (b) lapse rate calculated over all detected events. The blue lines indicate "environmental" properties averaged over a 200 km × 200 km box, the orange lines indicate environmental properties averaged within the area of the respective convective event.

local humidity from the environmental humidity and a random forest to predict the local lapse rate from the environmental lapse rate, where local humidity and lapse rate are averaged over the same layer as the environmental variable but within the time interval 0–0.5 hr. We find that the random forest for local humidity content only explains 36% of the variability from environmental RH. In contrast, the corresponding random forest for local lapse rate explains 63% of the variability if based soley on the environmental lapse rate.

Being aware that the local RH variable may still not be a good predictor of the RH relevant to the downdraft, we question whether the inclusion of this variable (together with the local lapse rate variable) leads to increased predictability of the downdraft properties. We therefore repeat the analysis shown in Section 3 and find that only the downdraft velocity can be explained to a slightly larger extent (see Figure 14). It is, however, interesting to note that local RH is now the second most important feature in predicting downdraft velocity. Together, these results seem to imply that the local humidity's influence can be predicted indirectly by the larger-scale environmental predictors, though not by the larger-scale humidity alone.

While the discussion above has focused on the effect of evaporative cooling due to entrainment of environmental air, we also note that environmental air could play an important role for the initial characteristics of a downdraft. In particular, if downdrafts originate from mixtures that contain a significant fraction of environmental air rather than detrained air (whose properties might be more independent from environmental conditions), the environmental humidity could influence the strength of the initial temperature anomaly by enhancing evaporative cooling, but could also reduce the amount of available rainwater through the same process. Testing a number of cases in a parcel-based downdraft model (https://github.com/climate-enigma/dparcel, version 0.1), we found that downdrafts become either stronger or weaker with increasing environmental humidity, depending on how the initial properties of the downdraft parcel were linked to environmental humidity. This result is particularly interesting in light of modern convective parameterization schemes (e.g., D'Andrea et al., 2014; Gerard, 2007; Suselj et al., 2019; Wu et al., 2020), which explicitly represent many of the downdraft-environment interactions but make comparatively crude assumptions about the initial thermodynamic variables of the downdrafts.

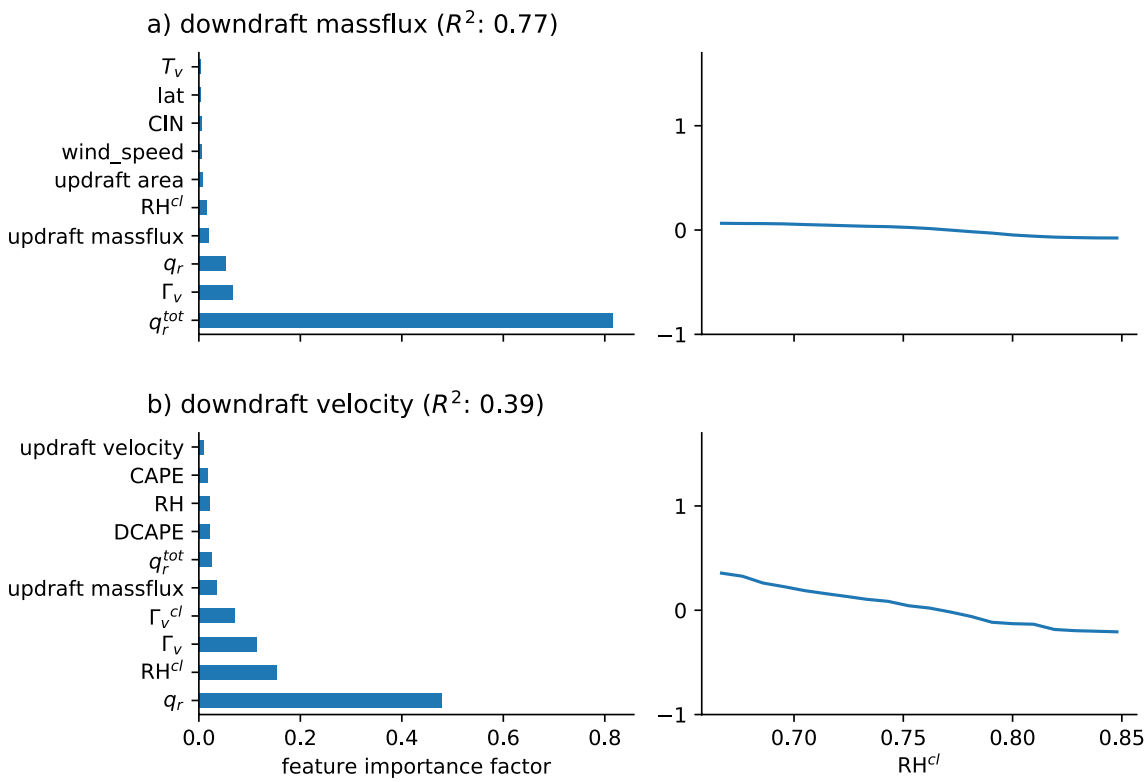

Figure 14. (Left) Feature importance (as in Figure 9) of the 10 most important predictors and (right) partial dependence (as in Figure 10c), but now including local relative humidity (RH^{cl}) and local lapse rate $\left(\Gamma_v^{cl}\right)$ as additional input features for (a) downdraft mass flux and (b) downdraft velocity.

In summary, therefore, there exist at least three ways in which environmental humidity can impact downdraft strength. In a drier environment:

- the virtual temperature difference between the downdraft and the environment is smaller leading to weaker downdrafts
- the entrainment of environmental air enhances evaporation leading to stronger downdrafts
- more rainwater can evaporate during downdraft formation, causing an initially more negative temperature perturbation but also a depletion of rainwater which, taken together, can lead to either stronger or weaker downdrafts

This collection of partly competing effects might explain why the random forest approach above does not show a clearer dependence of downdraft strength on environmental humidity, either because the effects differ, for example, between different types of convection, or because these effects tend to approximately balance in general.

5. Conclusion

Using a global storm-resolving model, we address the question of what determines the properties of convective downdrafts over tropical oceans. In particular, we investigate how the properties of downdrafts are related to the properties of the preceding updrafts as well as to the environmental conditions. To this end, we use hierarchical clustering to examine the degree of correlation between different properties, and random forests as well as multiple linear regression to determine how well downdraft properties can be predicted from updraft properties and environmental conditions.

The downdraft properties we focus on are the downdraft area, the downdraft velocity, and the resulting downdraft mass flux. Analyzing the relationship between these three properties, we find that the variability of the downdraft mass flux is largely determined by the variability of the downdraft area rather than the downdraft velocity. Because both downdraft mass flux and downdraft velocity are properties of interest, we examine their relationship to the preceding updraft and environmental conditions separately.

Using two random forests to predict downdraft velocity and downdraft mass flux, respectively, we find that in both cases the updraft properties are far more important than the environmental properties. In particular, we find that the most important predictors of downdraft mass flux are the *extensive* updraft properties, that is, updraft properties that scale with area: updraft area, updraft mass flux, and total amount of available rain water. Of all these variables, the amount of available rain water is the best predictor of downdraft mass flux. For downdraft velocity, the best predictors are the *intensive* updraft properties, that is, updraft velocity and rainfall amount. Again, the rainfall dependent variable is the best predictor for downdraft velocity. A schematic summary of these results is shown in Figure 15. These results suggest that downdrafts are mainly controlled by the availability of rain water, which in turn is strongly dependent on the preceding updraft.

While the properties of the preceding updraft are critical for the downdraft properties, additional knowledge of environmental conditions is not as important. Considering environmental humidity, environmental stability, and several parcel-related variables (CAPE, CIN, and DCAPE), we find that only environmental lapse rate is a valuable predictor, with a more unstable environment favoring higher downdraft mass flux and velocity. The small

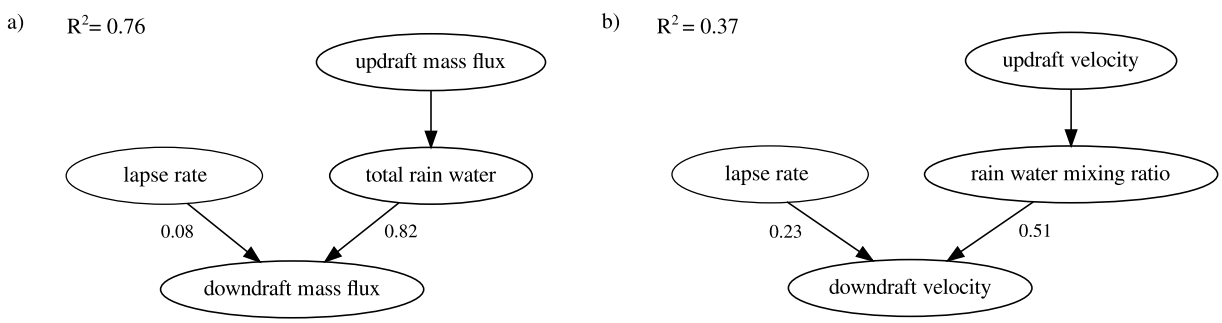

Figure 15. Schematic representation of key results, showing the coefficient of determination, the key predictors and the corresponding feature importance factors for (a) downdraft mass flux and (b) downdraft velocity as determined by the random forests.

importance of environmental humidity may be surprising, since entrainment of dry air increases evaporation, which should increase the strength of the downdraft. One possible explanation for this result is that the mean humidity on larger scales is not a good indicator of the RH of the air entrained into the downdraft. An other explanation is that, also considering the impact of environmental humidity on the starting properties of convective downdraft, a number of competing effects exist which might work in opposite direction for different convective events or tend to balance in general.

Interestingly, all the important dependencies of downdraft mass flux and downdraft velocity are well approximated by linear relationships. Repeating the analysis by fitting two linear models to downdraft mass flux and downdraft velocity, respectively, we confirm the results from the random forest approach. In particular, multiple linear regression yields very similar coefficients of determination and identify the same key input features for predicting downdraft properties. Based on these results, we additionally computed two highly simplified linear models for each downdraft variable. For each variable, we calculated one model based solely on the key updraft variable, and one that also accounts for the influence of the environmental lapse rate. These models provide a simple way to parameterize the downdraft mass flux based on the results of our simulations.

One limitation of our approach is that we are constrained in both horizontal and temporal resolution by the use of a global storm resolution model. The horizontal resolution of 5 km used here is much coarser than the resolutions commonly used to study downdrafts, since neither updrafts nor downdrafts are well resolved at this resolution. Also, the temporal resolution of 15 min is comparatively low, especially considering that the lifetime of a deep convective cloud in the tropics is usually on the order of an hour. While it is reasonable to expect that a higher resolution simulation would result in quantitative differences, we expect that the dependence on environmental conditions will still be qualitatively similar. Thus, this study does not replace higher resolution studies, but complements them in one important aspect: global simulations correctly reproduce cross-correlation functions observed in nature and therefore allow us to test predictions based on higher resolution simulations in as natural an environment as possible.

Data Availability Statement

The model source code is available on https://mpimet.mpg.de/en/science/modeling-with-icon/code-availability. The simulation run scripts and code for reproducing the plots are available through http://hdl.handle.net/21.11116/0000-0009-A854-B.

Acknowledgments
The authors greatly appreciate Monika Esch's help in performing the simulation. The authors are grateful to Hauke Schulz for the internal review of the manuscript. This study was supported by public funding to the Max Planck Society and the ARC Centre of Excellence for Climate Extremes. Computing resources were provided by the German Climate Computing Center (DKRZ). Open Access funding enabled and organized by Projekt DEAL.

References

Baldauf, M., Seifert, A., Förstner, J., Majewski, D., Raschendorfer, M., & Reinhardt, T. (2011). Operational convective-scale numerical weather prediction with the COSMO model: Description and sensitivities. *Monthly Weather Review*, *139*(12), 3887–3905. https://doi.org/10.1175/MWR-D-10-05013.1

Bao, J., & Windmiller, J. M. (2021). Impact of microphysics on tropical precipitation extremes in a global storm-resolving model. *Geophysical Research Letters*, *48*(13), e2021GL094206. https://doi.org/10.1029/2021GL094206

Blumberg, W. G., Halbert, K. T., Supinie, T. A., Marsh, P. T., Thompson, R. L., & Hart, J. A. (2017). SHARPpy: An open-source sounding analysis toolkit for the atmospheric sciences. *Bulletin of the American Meteorological Society*, *98*(8), 1625–1636. https://doi.org/10.1175/BAMS-D-15-00309.1

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Byers, H. R., & Braham, R. R. (1949). The thunderstorm: Report of the thunderstorm project (a joint project of four U.S. Government Agencies: Air Force, Navy, National Advisory Committee for Aeronautics, and Weather Bureau) (Tech. Rep.). https://doi.org/10.1002/qj.49707733225

D'Andrea, F., Gentine, P., Betts, A. K., & Lintner, B. R. (2014). Triggering deep convection with a probabilistic plume model. *Journal of the Atmospheric Sciences*, *71*(11), 3881–3901. https://doi.org/10.1175/JAS-D-13-0340.1

Fuchs, D., Sherwood, S. C., Waugh, D., Dixit, V., England, M. H., Hwong, Y.-L., & Geoffroy, O. (2022). Midlatitude jet position spread linked to atmospheric convective types. *Journal of Climate*, 1–44. https://doi.org/10.1175/JCLI-D-21-0992.1

Fujita, T. T., & Wakimoto, R. M. (1981). Five scales of airflow associated with a series of downbursts on 16 July 1980. *Monthly Weather Review*, *109*(7), 1438–1456. https://doi.org/10.1175/1520-0493(1981)109<1438:FSOAAW>2.0.CO;2

Gerard, L. (2007). An integrated package for subgrid convection, clouds and precipitation compatible with meso-gamma scales. *Quarterly Journal of the Royal Meteorological Society*, *133*(624), 711–730. https://doi.org/10.1002/qj.58

Hansen, Z. R., Back, L. E., & Zhou, P. (2020). Boundary layer quasi-equilibrium limits convective intensity enhancement from the diurnal cycle in surface heating. *Journal of the Atmospheric Sciences*, *77*(1), 217–237. https://doi.org/10.1175/JAS-D-18-0346.1

Hohenegger, C., Kornblueh, L., Klocke, D., Becker, T., Cioni, G., Engels, J. F., et al. (2020). Climate statistics in global simulations of the atmosphere, from 80 to 2.5 km grid spacing. *Journal of the Meteorological Society of Japan*, *98*(1), 73–91. https://doi.org/10.2151/jmsj.2020-005

Howard, J., & Gugger, S. (2020). Tabular modeling deep dive. In *Deep learning for coders with fastai and PyTorch*. O'Reilly Media, Inc. Retrieved from https://learning.oreilly.com/library/view/deep-learning-for/9781492045519/

Kain, J. S., & Fritsch, J. M. (1993). Convective parameterization for mesoscale models: The Kain-Fritsch scheme. In K. A. Emanuel & D. J. Raymond (Eds.), *The representation of cumulus convection in numerical models* (Meteorolog ed., pp. 165–170). American Meteorological Society. https://doi.org/10.1007/978-1-935704-13-3_16

Marion, G. R., & Trapp, R. J. (2019). The dynamical coupling of convective updrafts, downdrafts, and cold pools in simulated supercell thunderstorms. *Journal of Geophysical Research: Atmospheres*, *124*(2), 664–683. https://doi.org/10.1029/2018JD029055

Mlawer, E. J., Taubman, S. J., Brown, P. D., Iacono, M. J., & Clough, S. A. (1997). Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the longwave. *Journal of Geophysical Research*, *102*(D14), 16663–16682. https://doi.org/10.1029/97jd00237

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830. Retrieved from http://jmlr.org/papers/v12/pedregosa11a.html

Proctor, F. H. (1989). Numerical simulations of an isolated microburst. Part II: Sensitivity experiments. *Journal of the Atmospheric Sciences*, *46*(14), 2143–2165. https://doi.org/10.1175/1520-0469(1989)046<2143:NSOAIM>2.0.CO;2

Purdom, J. F. W. (1976). Some uses of high-resolution GOES imagery in the mesoscale forecasting of convection and its behavior. *Monthly Weather Review*, *104*(12), 1474–1483. https://doi.org/10.1175/1520-0493(1976)104<1474:SUOHRG>2.0.CO;2

Raschendorfer, M. (2001). The new turbulence parameterization of LM. *Consortium for Small Scale Modelling Newsletter*, *1*, 89–97. Retrieved from http://www.cosmo-model.org/content/model/documentation/newsLetters/newsLetter01/newsLetter_01.pdf

Raymond, D. J. (1995). Regulation of moist convection over the west Pacific warm pool. *Journal of the Atmospheric Sciences*, *52*(22), 3945–3959. https://doi.org/10.1175/1520-0469(1995)052<3945:ROMCOT>2.0.CO;2

Reick, C. H., Raddatz, T., Brovkin, V., & Gayler, V. (2013). Representation of natural and anthropogenic land cover change in MPI-ESM. *Journal of Advances in Modeling Earth Systems*, *5*(3), 459–482. https://doi.org/10.1002/jame.20022

Sanner, M. F. (1999). Python: A programming language for software integration and development. *Journal of Molecular Modeling*, *17*(1), 57–61.

Schlemmer, L., & Hohenegger, C. (2014). The formation of wider and deeper clouds as a result of cold-pool dynamics. *Journal of the Atmospheric Sciences*, *71*(8), 2842–2858. https://doi.org/10.1175/JAS-D-13-0170.1

Schrodin, R., & Heise, E. (2002). A new multi-layer soil-model. *Consortium for Small Scale Modelling Newsletter*, *2*, 149–151. Retrieved from http://www.cosmo-model.org/content/model/documentation/newsLetters/newsLetter02/default.htm

Srivastava, R. C. (1985). A simple model of evaporatively driven downdraft: Application to microburst downdraft. *Journal of the Atmospheric Sciences*, *42*(10), 1004–1023. https://doi.org/10.1175/1520-0469(1985)042<1004:ASMOED>2.0.CO;2

Srivastava, R. C. (1987). A model of intense downdrafts driven by the melting and evaporation of precipitation. *Journal of the Atmospheric Sciences*, *44*(13), 1752–1773. https://doi.org/10.1175/1520-0469(1987)044<1752:amoidd>2.0.co;2

Stevens, B., Satoh, M., Auger, L., Biercamp, J., Bretherton, C. S., Chen, X., et al. (2019). DYAMOND: The DYnamics of the atmospheric general circulation modeled on non-hydrostatic domains. *Progress in Earth and Planetary Science*, *6*, 1–17. https://doi.org/10.1186/s40645-019-0304-z

Suselj, K., Kurowski, M. J., & Teixeira, J. (2019). A unified Eddy-diffusivity/mass-flux approach for modeling atmospheric convection. *Journal of the Atmospheric Sciences*, *76*(8), 2505–2537. https://doi.org/10.1175/JAS-D-18-0239.1

Thayer-Calder, K., & Randall, D. (2015). A numerical investigation of boundary layer quasi-equilibrium. *Geophysical Research Letters*, *42*(2), 550–556. https://doi.org/10.1002/2014GL062649

Tiedtke, M. (1989). A comprehensive mass flux scheme for cumulus parameterization in large-scale models. *Monthly Weather Review*, *117*(8), 1779–1800. https://doi.org/10.1175/1520-0493(1989)117<1779:ACMFSF>2.0.CO;2

Tompkins, A. M. (2001). Organization of tropical convection in low vertical Wind shears: The role of cold pools. *Journal of the Atmospheric Sciences*, *58*(13), 1650–1672. https://doi.org/10.1175/1520-0469(2001)058<1650:OOTCIL>2.0.CO;2

Torri, G., & Kuang, Z. (2016). A Lagrangian study of precipitation-driven downdrafts. *Journal of the Atmospheric Sciences*, *73*(2), 839–854. https://doi.org/10.1175/JAS-D-15-0222.1

Torri, G., Kuang, Z., & Tian, Y. (2015). Mechanisms for convection triggering by cold pools. *Geophysical Research Letters*, *42*(6), 1943–1950. https://doi.org/10.1002/2015GL063227

Trapp, R. J., & Woznicki, J. M. (2017). Convectively induced stabilizations and subsequent recovery with supercell thunderstorms during the Mesoscale Predictability Experiment (MPEX). *Monthly Weather Review*, *145*(5), 1739–1754. https://doi.org/10.1175/MWR-D-16-0266.1

Wakimoto, R. M. (2001). Convectively driven high wind events. In *Severe convective storms* (pp. 255–298). American Meteorological Society. https://doi.org/10.1007/978-1-935704-06-5_7

Wu, E., Yang, H., Kleissl, J., Suselj, K., Kurowski, M. J., & Teixeira, J. (2020). On the parameterization of convective downdrafts for marine stratocumulus clouds. *Monthly Weather Review*, *148*(5), 1931–1950. https://doi.org/10.1175/MWR-D-19-0292.1

Zängl, G., Reinert, D., Rípodas, P., & Baldauf, M. (2015). The ICON (ICOsahedral Non-hydrostatic) modelling framework of DWD and MPI-M: Description of the non-hydrostatic dynamical core. *Quarterly Journal of the Royal Meteorological Society*, *141*(687), 563–579. https://doi.org/10.1002/qj.2378

Zhang, G. J., & McFarlane, N. A. (1995). Sensitivity of climate simulations to the parameterization of cumulus convection in the Canadian Climate Centre general circulation model. *Atmosphere-Ocean*, *33*(3), 407–446. https://doi.org/10.1080/07055900.1995.9649539