

RESEARCH ARTICLE

Cold-pool-driven convective initiation: using causal graph analysis to determine what convection-permitting models are missing

Mirjam Hirt¹  | George C. Craig¹ | Sophia A. K. Schäfer² | Julien Savre¹ | Rieke Heinze³

¹Meteorologisches Institut,
Ludwig-Maximilians Universität,
München, Germany

²Deutscher Wetterdienst, Offenbach,
Germany

³Max Planck Institute for Meteorology,
Hamburg, Germany

Correspondence

M. Hirt, Meteorologisches Institut,
Ludwig-Maximilians Universität,
München 80333, Germany
Email: m.hirt@lmu.de

Funding information

German Research Foundation (DFG)
German Federal Ministry of Education
and Research, Grant/Award Numbers:
SFB/TRR 165 A6, 01LK1507A and
01LK1507D

Abstract

Cold-pool-driven convective initiation is investigated in high-resolution, convection-permitting simulations with a focus on the diurnal cycle and organization of convection and the sensitivity to grid size. Simulations of four different days over Germany were performed using the ICON-LEM model with grid sizes from 156 to 625 m. In these simulations, we identify cold pools, cold-pool boundaries and initiated convection. Convection is triggered much more efficiently in the vicinity of cold pools than in other regions and can provide as much as 50% of total convective initiation, in particular in the late afternoon. By comparing different model resolutions, we find that cold pools are more frequent, smaller and less intense in lower-resolution simulations. Furthermore, their gust fronts are weaker and less likely to trigger new convection. To identify how model resolution affects this triggering probability, we use a linear causal graph analysis. In doing so, we postulate a graph structure with potential causal pathways and then apply multi-linear regression accordingly. We find a dominant, systematic effect: reducing grid sizes directly reduces upward mass flux at the gust front, which causes weaker triggering probabilities. These findings are expected to be even more relevant for km-scale, numerical weather prediction models. We thus expect that a better representation of cold-pool-driven convective initiation will improve forecasts of convective precipitation.

KEYWORDS

causal effect estimation, numerical weather prediction, convection organization

1 | INTRODUCTION

Convection-permitting models have become increasingly prominent for numerical weather prediction in recent years (Baldauf *et al.*, 2011; Clark *et al.*, 2016). Such

models have a grid size of several hectometres to a few kilometres which is sufficient to run without the use of a convection parametrization. Due to the many underlying approximations and systematic biases of convection schemes (e.g., Gentine *et al.*, 2018), being able to simulate

deep convective overturning without the need for such a parametrization significantly improves forecasts of convective precipitation (Clark *et al.*, 2016; Baldauf *et al.*, 2011). Nonetheless, there remain systematic deficiencies in convection-permitting simulations: the diurnal cycle of convection is still often not correct (Baldauf *et al.*, 2011); a strong sensitivity has been observed of convective cell sizes, intensity and initiation time to the mixing-length parameter of the subgrid turbulence scheme (Hanley *et al.*, 2015); and frequent grid cell storms and a lack of persistent organization into the night was identified by Rasp *et al.* (2018) and confirmed by Hirt *et al.* (2019). Significantly, the diurnal cycle of precipitation is improved during the night when resolution is increased to a few hundred metres (Stevens *et al.*, 2020).

A substantial part of these deficiencies is likely related to planetary boundary-layer processes and the initiation of convection (Baldauf *et al.*, 2011; Kober and Craig, 2016; Yano *et al.*, 2018; Hirt *et al.*, 2019), processes which are still inadequately represented in these models. On summertime convective days, large amounts of convective available potential energy (CAPE) may be present, a necessary prerequisite for deep convection. Yet the atmospheric boundary layer is often capped by a temperature inversion which inhibits the initiation of convection (convective inhibition, CIN). For air parcels to overcome this CIN, local triggering processes are necessary, such as lifting by boundary-layer turbulence, low-level convergence, orography or convective cold pools (e.g., Doswell, 1987; Johns and Doswell, 1992). These triggering processes are often sub-grid processes even in convection-permitting models, which restricts the models' ability to simulate convective initiation (CI). Consequently, the quality of simulated convection itself is affected. Kober and Craig (2016) addressed these issues by introducing stochastic perturbations to reintroduce missing variability of boundary-layer turbulence. This approach was extended by Hirt *et al.* (2019) who also accounted for lifting by subgrid-scale orography. In this work we will focus on the role of cold pools in driving CI and their representation in convection-permitting models.

Convective cold pools are volumes of negatively buoyant air that originate from precipitating downdraughts. Evaporation of precipitation in the sub-cloud layer, combined with the weight of the condensed water, creates negative buoyancy and accelerates the downdraught. When these cold, moist air masses hit the surface, they spread in near-circular patterns and interact strongly with the surface (Gentine *et al.*, 2016; Drager and van den Heever, 2017; Peters and Hohenegger, 2017; Grant and van den Heever, 2018). This spreading is often led by a gust front, where lifting occurs which fosters the initiation of new convection. In recent decades, three major

mechanisms have been proposed as explanations for this secondary CI. First, Rotunno *et al.* (1988) and Weisman and Rotunno (2004) proposed that horizontal vorticity linked to cold-pool-induced buoyancy gradients at the gust front interacts with the vorticity generated by the background wind shear to yield approximately vertical updraughts. This pattern of convergence and ascent is optimal for initiating new convection and thereby allowing the reinforcement and propagation of squall lines (Rotunno *et al.*, 1988; Weisman and Rotunno, 2004; Markowski and Richardson, 2011). We will refer to this mechanism as mechanical lifting. Second, collisions of cold pools have been identified as a major contributor to new CI (e.g., (Purdom, 1976; Droegemeier and Wilhelmson, 1985; Feng *et al.*, 2015; Cafaro and Rooney, 2018; Haerter *et al.*, 2018; Torri and Kuang, 2019)). The collision of two cold pools can cause a superposition of their updraught velocities, resulting in stronger updraughts. Collisions of three cold pools can trap non-cold-pool air between the colliding fronts, which results in even stronger vertical motion (Haerter *et al.*, 2018). Third, Tompkins (2001) proposed a mechanism based on buoyancy-driven initiation: decaying cold pools accumulate moisture from the perished precipitation downdraughts at their boundaries. This increased moisture can compensate the effect of low temperatures on buoyancy (especially for "old" cold pools, where entrainment has depleted the cold air) and thereby reduce the CIN, which supports the initiation of new convection. Torri *et al.* (2015) investigated the relative importance of the mechanical and thermodynamic mechanisms in idealized studies, and found that mechanical lifting drives the vertical motion near the surface, whereas higher in the boundary layer, moisture seems to be relevant for vertical acceleration. Independently of the specific mechanism, the effect of cold pools is to trigger convection in the vicinity of already existing convection, leading to organization of convective cells and a more clustered precipitation pattern. Cold pools may also influence the diurnal cycle of convection since they are a relevant triggering mechanism only after initial precipitation events have occurred. Especially in the late afternoon/evening, when other mechanisms like surface heating are no longer strong, cold-pool triggering may dominate. For instance, Maurer *et al.* (2017) found that 35% of all CI in the afternoon in the Sahel was attributable to cold-pool gust fronts.

Given the weaknesses of convective forecasts from convection-permitting models, the question arises whether such models can simulate cold-pool-related triggering processes with sufficient accuracy. Grant and van den Heever (2016) have revealed a resolution dependence of cold-pool gust fronts in idealized, two-dimensional large-eddy simulations (LESs) with grid sizes of 50–400 m, i.e., weaker maximum vertical velocities in coarser

models. This is consistent with studies showing a general sensitivity of vertical velocity to model resolution (e.g., Heinze *et al.*, 2017; Jeevanjee, 2017). Squitieri and Galus (2020) recently identified a resolution sensitivity of cold pools in mesoscale convective systems with a higher frequency and larger areas in higher-resolution simulations. However, these studies have not investigated how well the thermodynamic triggering mechanisms are represented in convection-permitting models. Perhaps as a result of these errors, convection-permitting models struggle with two aspects of convection that can potentially be ascribed to cold pools, namely the diurnal cycle and organization of convection. The difficulty of capturing the diurnal cycle of convection has been documented in a variety of convection-permitting km-scale models (Baldauf *et al.*, 2011; Hanley *et al.*, 2015; Clark *et al.*, 2016). In contrast to convection-parametrized models, where convection occurs too early during the day (Leutwyler *et al.*, 2017), convection in convection-permitting models starts too late during the day (Baldauf *et al.*, 2011; Clark *et al.*, 2016). This error can be improved by tuning the model, for example, with the turbulent mixing length, but at the expense of other biases (Baldauf *et al.*, 2011; Schraff *et al.*, 2016; Leutwyler *et al.*, 2017). Using a more physically based approach, Kober and Craig (2016) and Hirt *et al.* (2019) were able to improve the onset of convection using stochastic perturbations to account for boundary-layer turbulence. However, a lack of precipitation in the late afternoon and evening remained. A similar behaviour was documented by Rasp *et al.* (2018) based on a 12-day summertime convective period over Germany. In addition, a lack of convective organization and structural deficits in cloud sizes were found (Rasp *et al.*, 2018; Senf *et al.*, 2018; Panosetti *et al.*, 2019). Considering the characteristics of cold pools as described in the preceding paragraph, we expect cold pools to be relevant for this lack of late afternoon/evening precipitation and organization. A quantitative understanding of how cold pools and the associated circulations depend on model resolution could potentially provide the basis for an improved treatment in km-scale models, similar to the improved representation of marginally resolved boundary-layer turbulence given by the stochastic schemes of Kober and Craig (2016) and Hirt *et al.* (2019).

In this study we address the following three research questions:

1. Are cold pools a significant factor for the diurnal cycle and the organization of convection for midlatitude, continental weather?
2. Can convection-permitting models adequately represent cold-pool-driven CI processes?

3. If not, which misrepresented processes are responsible for the deficiencies in cold-pool-driven CI?

Since observations of cold pools are sparse and often incomplete, we base our work on high-resolution simulations of realistic weather situations over Germany. Simulations with 156 m grid sizes are available from the High-Definition Clouds and Precipitation for advancing Climate Prediction project (HD(CP)²; Heinze *et al.*, 2017). The simulations have been thoroughly compared to observations, confirming that their turbulence profiles or cloud distributions are sufficiently comparable to observations (Heinze *et al.*, 2017). We assume that cold-pool dynamics will be sufficiently resolved as well. The 156 m simulation can be compared with additional simulations at 312 and 625 m grid spacing. By comparing cold-pool and CI characteristics between these resolutions, we draw conclusions regarding the ability of convection-permitting models to represent cold-pool-driven CI. In doing this, we assume that higher resolution produces more realistic cold pools and cold-pool-driven CI, and that the simulations with 625 m grid spacing are sufficiently representative of convection-permitting simulations – or at least its higher-resolution end of the spectrum. Due to the immense computational cost of the high-resolution simulations, we are only able to analyse four summertime convective days. These days have been selected with a range of large-scale meteorological conditions which we believe are representative of the diversity typically found in nature. To address the question of which processes are responsible for the differences between the simulations at different resolutions, we apply linear causal effect estimation based on Pearl's causality (Pearl, 2009) and the graphical representation of causal structures (Wright, 1921; Pearl, 2013; Chen and Pearl, 2014). This enables a formalized estimation of different direct or indirect causal effects in the model data. While this framework is not yet commonly applied in atmospheric sciences, some climate studies have applied and developed related causality concepts (e.g., Ebert-Uphoff and Deng, 2012; Runge *et al.*, 2015; Kretschmer *et al.*, 2016; Samarasinghe *et al.*, 2019). A recent review of causality inference in Earth system science can be found in Runge *et al.* (2019).

The paper is structured as follows. In Section 2 we lay out relevant details of the numerical simulations and the days to be investigated, how we identify cold pools, cold-pool boundaries and CI, and the applied diagnostic and causal methods. Section 3 presents the results of this study organized around the three research questions. In Section 4 we provide a summary of our results and discuss their underlying assumptions and implications.

2 | DATA AND METHODS

2.1 | ICON-LEM simulations

The basis of this study is a set of high-resolution summertime simulations over Germany performed with the large-eddy version of the ICOSahedral Non-hydrostatic atmosphere model (ICON; Dipankar *et al.*, 2015; Zängl *et al.*, 2015) in limited-area model configuration (Heinze *et al.*, 2017) as part of the HD(CP)² project. Simulations have been performed at horizontal resolutions of 156, 312 and 625 m on a number of days with a range of meteorological conditions. Details of the model set-up and a comprehensive evaluation of the simulations can be found in Heinze *et al.* (2017). The meteorological situations, thermodynamic conditions and cloud and precipitation properties seen in observations are generally well captured in the ICON-LEM simulations, and show improvements when compared to km-scale simulations (Heinze *et al.*, 2017). For comparability, the model output is interpolated onto a common 1.2 km latitude–longitude grid for all three resolutions. We select a region of 5.503–14.496°E and 47.599–54.496°N for our analysis, for all dates and resolutions. The simulation output is available from 0600 to 2400 UTC with 5-min output frequency (simulations start at 0000 UTC). The simulated data have been analysed using the Python environment with frequently used packages specified in Appendix A.

2.2 | Selected days and their synoptic situations

We have selected four days that represent diverse large-scale conditions and exhibit summertime convection and cold pools, namely 5 July 2015, 29 May, 6 June and 1 August 2016. Snapshots of buoyancy intensity (as defined in Section 2.3.1), precipitation and cold-pool boundaries are shown for each day in Figure 1. Time series of selected domain-aggregated variables that capture the large-scale convective situation are displayed in Figure 2.

On 5 July 2015, Germany lay in the warm sector of a frontal system that had approached from France. The approaching front was led by a convergence line and relatively strong southwesterly winds were present. High CIN, CAPE, bulk Richardson number (BRN)¹ and significant wind shear were present over the simulated domain (Figure 2a,b,e,f). Squall line-like structures developed in the simulations, travelling over Germany from west to

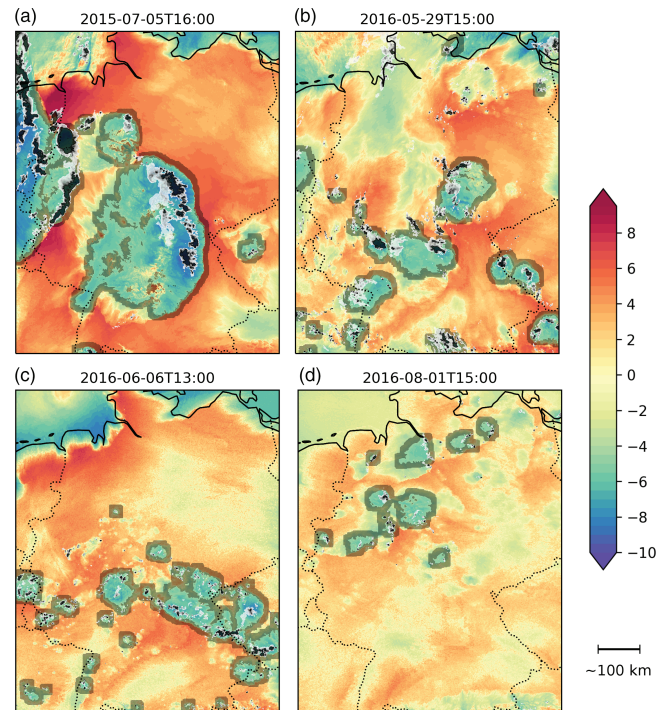


FIGURE 1 Example snapshots of buoyancy intensity ($\text{m}\cdot\text{s}^{-1}$; colour shading; Section 2.3.1), cold-pool boundary regions (grey shading) and precipitation intensity (white $1\text{ mm}\cdot\text{hr}^{-1}$ to black $20\text{ mm}\cdot\text{hr}^{-1}$) are displayed for each day from the 156 m grid size simulations. Dotted and solid black lines show country borders and coastlines respectively

east, accompanied by two enormous cold pools (Figure 1a; the second enormous cold pool entered the domain from the west at this time). On 29 May 2016, a low pressure system connected to an omega block was centred over Germany. Cyclonic rotation, large-scale lifting (Figure 2d) and intense and damaging flooding occurred, especially in Braunschweig, Baden-Württemberg. Several cold pools are found in the simulations. The high pressure system of the omega block slowly shifted subsequently to fully enclose Germany on 6 June 2016, suppressing large-scale lifting (Figure 2d) and wind shear (Figure 2f), but leaving high values of CAPE and BRN (Figure 2b,e). During the day, widespread scattered convection was simulated, and an organized band formed from merged cold pools and spread over southeastern Germany during the afternoon. Due to the intense convective activity and high variability in underlying forcing, the period spanning 29 May to 6 June 2016 has been considered in a number of studies on convection (Hirt *et al.*, 2019; Rasp *et al.*, 2018; Keil *et al.*, 2019; Baur *et al.*, 2018; Bachmann *et al.*, 2020) and analyzed in detail by Piper *et al.* (2016). On 1 August 2016, a low pressure system over Scandinavia and a small high pressure system over Switzerland dominated the synoptic situation over Germany. Relatively small CAPE values (Figure 2b), a deep boundary layer, some large-scale subsidence, and

¹The bulk Richardson number is computed as specified by Markowski and Richardson (2011). For the 0–6 km mean wind, the model levels 76 to 150 are used, and for the 0–500 m mean wind, model levels 139 to 150 are used.

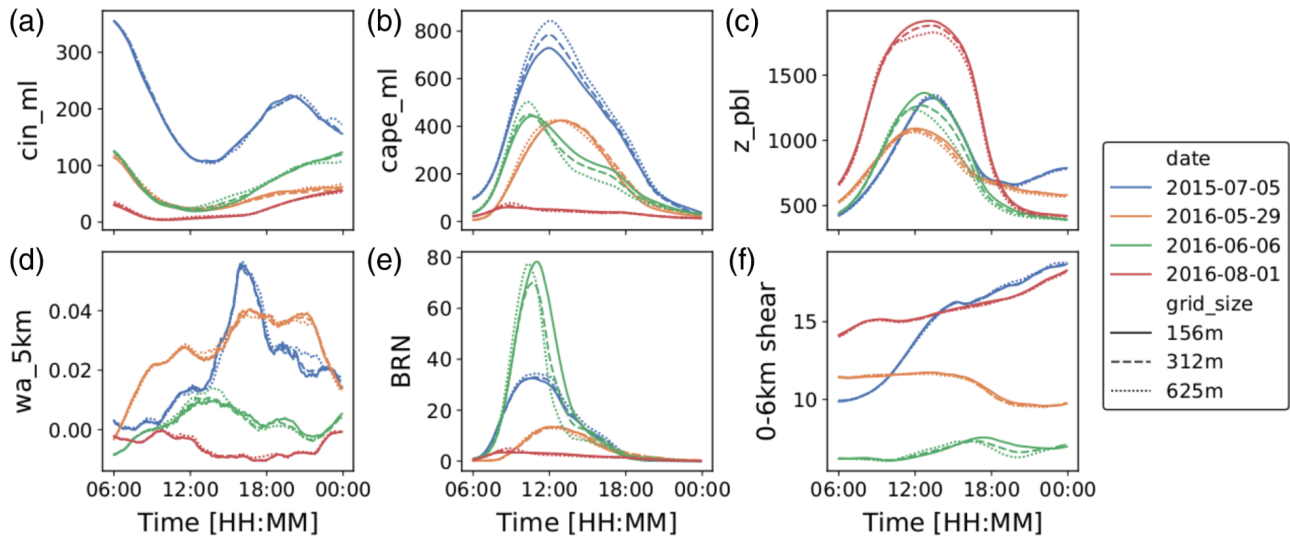


FIGURE 2 Time series of domain-aggregated variables describing the synoptic situation: (a) mean layer convective inhibition ($\text{J}\cdot\text{kg}^{-1}$, direct model output); (b) mean layer convective available potential energy ($\text{J}\cdot\text{kg}^{-1}$, direct model output); (c) height of the planetary boundary layer (m a.s.l., direct model output); (d) vertical velocity at approximately 5 km ($\text{m}\cdot\text{s}^{-1}$); (e) Bulk Richardson number (BRN, computed according to Markowski and Richardson, 2011); (f) 0–6 km wind shear ($\text{m}\cdot\text{s}^{-1}$, using model levels 150 and 76). For BRN, the domain median has been computed due to the occurrence of high outlier values. Otherwise, domain averages are displayed

significant shear were simulated (Figure 2c,d,f). The simulations produced many small cold pools over northern Germany (Figure 1d).

2.3 | Detection of cold pools and their edges

2.3.1 | Cold pool detection

We identify cold pools using the density potential temperature perturbation and precipitation at the lowest layer above the surface. Density potential temperature is defined as

$$\theta_\rho = \theta \left(1 + 0.608r_v - r_w - r_i - r_r - r_s - r_g - r_h \right)$$

(equations (4.3.2) and (4.3.6) of Emanuel, 1994, using an approximation to first order in r analogous to equation (4.3.1)). Here θ denotes the potential temperature, r_v the water vapour mixing ratio and r_w , r_i , r_r , r_s , r_g and r_h the mixing ratios of liquid cloud water, cloud ice, rain, snow, graupel and hail respectively.

We compute the local perturbation in density potential temperature, $\theta'_{\rho,0}$, relative to a moving average, $\overline{\theta_{\rho,0}}$, of filter size 166 pixels (≈ 200 km) horizontally and 8 hr in time. The moving average is computed with reflection at the boundaries. This combination of spatial and temporal filtering enables the identification of large cold pools (especially on 5 July 2015) while background θ_ρ gradients

at the coast are still sufficiently resolved to detect cold pools there. A small sensitivity analysis with different spatial filtering scales and with/without temporal filtering showed no strong influence on the qualitative behaviour of our results. We further calibrate the perturbations by subtracting the domain-mean density potential temperature perturbation: $\theta'_\rho = \theta'_{\rho,0} - \overline{\theta'_{\rho,0}}$. We then identify regions where the density potential temperature perturbation θ'_ρ is below -2 K and apply the watershed-merge segmentation method of Senf *et al.* (2018) to identify cold pools. The segmentation algorithm partitions these cold areas into cold-pool objects by first shrinking the objects to identify object cores and assigning each edge pixel to the nearest core. If part of the interface between two objects thus found is farther from the environment than half of the maximum distance to the environment in one of the objects (i.e., the interface passes through an inner part of the object), they are merged again, otherwise (if the interface lies close to the outer edge), they are kept separate (appendix A in Senf *et al.*, 2018). Cold-pool objects smaller than 20 pixels, that is, with an equivalent diameter of ≈ 6 km, are excluded. Following Senf *et al.* (2018), this criterion has been chosen to include only cold pools with an equivalent diameter larger than the effective model resolution of the coarsest model (≈ 5 km; Heinze *et al.*, 2017). In order to exclude cold areas like sea breezes or valley inversions that are not connected to convection, we stipulate that precipitation in at least one pixel of the cold-pool object must exceed $10 \text{ mm}\cdot\text{hr}^{-1}$, and discard objects otherwise. This method successfully identifies roughly convex

cold-pool areas connected to convective precipitation. It is not very sensitive to the choice of thresholds, and is more robust than more complex cold-pool identifiers suggested by Drager and van den Heever (2017). However, given our selected criteria, we will not be able to identify very weak cold pools (with regard to buoyancy anomaly), very small cold pools or non-/weakly precipitating cold pools, which may especially exclude very young and old cold pools. This may have some implications for our results, but due to the small size and the weakness of young and old cold pools, we expect to have captured most cold pools that are relevant for CI. The expectation that dissipating cold pools trigger fewer new convective cells is also supported by results from Fournier and Haerter (2019).

For each air column, we define the cold pool *intensity* I using the integrated buoyancy anomaly b_{int} over the lowest five layers:

$$I = \text{sign}(b_{\text{int}}) \sqrt{|2b_{\text{int}}|}, \quad \text{with } b_{\text{int}} = \int_{0\text{m}}^{\approx 150\text{m}} b \, dz \quad \text{and buoyancy } b = g \frac{\theta'_\rho}{\theta_\rho}.$$

This definition is chosen to agree with the cold-pool intensity used by Feng *et al.* (2015), except for the sign which we choose to be positive for positively buoyant air (and therefore negative within cold pools). We choose fixed model layers as cold-pool depth for simplicity. The lowest five model layers (~ 150 m) were selected as they include the buoyancy perturbation of most cold pools; integrating slightly higher than the cold-pool top does not affect the result, since the buoyancy anomaly is small at these levels. Figure 1 shows snapshots of intensity for a selected time on each of the four days.

2.3.2 | Cold-pool boundary regions

It is not the cold pool itself but the leading gust front with thermodynamic or mechanical lifting that drives the initiation of new convection. However, detecting the gust front is not trivial in an automated framework and associating it to a specific cold pool is also ambiguous. For instance, using horizontal wind speed or vertical velocity above a threshold to identify cold-pool gust fronts will also capture other phenomena like land–sea breezes, orography or convergence lines. Recently, Fournier and Haerter (2019) and Henneberg *et al.* (2020) developed two approaches for detecting cold-pool gust fronts based on horizontal wind speeds which may provide better alternatives in the future. Using the above cold-pool definition, the gust fronts tend to lie just outside the identified cold pools (Figure 3). To avoid the difficulties of objectively identifying gust fronts, we define a cold-pool boundary region as a zone around

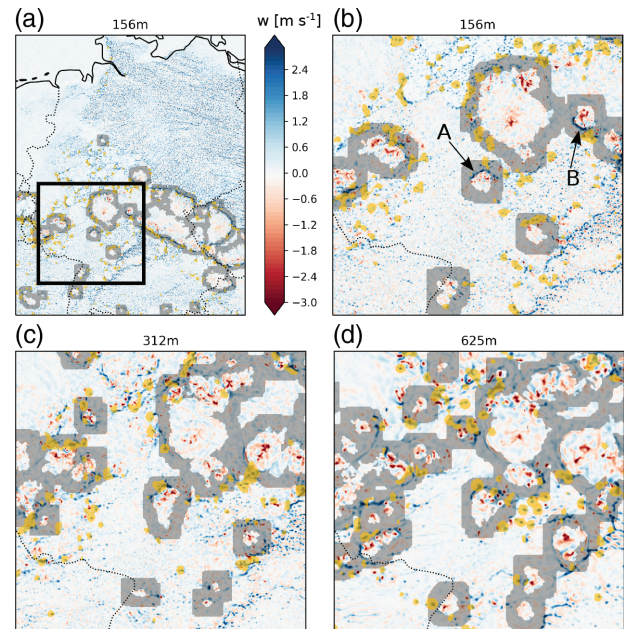


FIGURE 3 Example snapshots of vertical velocity ($\text{m}\cdot\text{s}^{-1}$, colour shading) at ≈ 1 km, cold-pool boundaries in grey shading and locations of convective initiation in yellow for each model resolution on 6 June 2016 at 1300 UTC. The black box in (a) marks the displayed region for (b), (c) and (d). In (b), A and B mark cold-pool gust fronts

each cold pool with ≈ 25 km width, which is the region most likely to contain the gust front. With this definition, large buoyancy gradients and forced ascent will occupy only a fraction of this region. To characterize the gust front itself, we consider 95th percentile values, rather than averages over the entire gust front region. An advantage of considering this cold-pool boundary region is that the thermodynamic lifting will be included, which may not necessarily be located directly at the gust front. Grid points belonging to the original cold pool or other cold pools are excluded from the boundary region. However, overlap with other cold-pool boundary regions is possible. We will consider overlapping boundary regions as a proxy for cold-pool collisions.

Comparing cold-pool boundaries with vertical velocity in Figure 3, many line structures of strong vertical velocity can be identified as cold-pool gust fronts (e.g., A and B in Figure 3b) and generally they are positioned somewhere within the defined cold-pool boundary region. This confirms that such cold pool-boundaries are a useful proxy for their gust fronts.

2.4 | Defining convective initiation

There is no standard definition for identifying CI. In this study, our goal is to relate the initiation of a convective cloud to properties in the boundary layer, such as the gust

front vertical velocity. Hence, to obtain an independent identification of vertical velocity at the gust front and of CI, we refrain from using vertical velocity as an indicator for CI. We also avoid using quantities such as precipitation particles, which occur later in the convective cloud life cycle and may be difficult to associate with boundary-layer features due to advection of the cloud to neighbouring grid points. With these conditions in mind, we choose the decisive variable for CI to be the temporal change of cloud water content in the mid-troposphere. The onset of deep convection is associated with a rapid generation of cloud droplets extending through a large part of the tropospheric depth; hence large values of the temporal change of cloud water content $\partial r_w / \partial t = \partial_t r_w$ are associated with CI. The following definition for CI is used:

$$r_w^* = G \left[\sum_{z=2.5 \text{ km}}^{8 \text{ km}} \partial_t r_{w_z} \right] > b_{r_w^*}. \quad (1)$$

The temporal change is computed offline at 5 min time intervals. To eliminate the effect of high cirrus clouds and low stratocumulus, we consider only mid-tropospheric levels and compute the sum of $\partial_t r_w$ over the model levels 60–110, corresponding approximately to 2.5–8 km. To reduce noise and to allow for some spatial displacement, we apply a Gaussian filter G using two grid boxes as standard deviation for the Gaussian kernel. CI is determined to have occurred at grid points where r_w^* exceeds a threshold $b_{r_w^*}$ which depends on day and resolution. The threshold value is defined as the α -percentile of the r_w^* field, where $\alpha = 1 - f_{\text{prec}}$ with f_{prec} being the relative frequency of precipitating grid points for that day and model resolution. A grid point is identified as precipitating if the rain rate exceeds $1 \text{ mm} \cdot \text{hr}^{-1}$. In this way, we identify as many grid points as having CI as we have precipitating gridpoints for that day and model resolution. We note that single grid points can be identified as CI for subsequent time steps as long as the criterion holds.

Example fields of CI are displayed in Figure 3. Further visual comparisons with precipitation fields (not shown) reveal that most identified CI is indeed related to an onset of precipitation at later times and that advection of deep convection is only rarely identified as CI.

2.5 | Diagnostic approaches

2.5.1 | Grid point classification

With the criteria defined above, we can partition the horizontal domain into four categories, namely (a) *no cold pools*, (b) *cold pools*, (c) *cold-pool edges*, and (d) *overlapping cold-pool edges*, as schematically illustrated in Figure 4a.

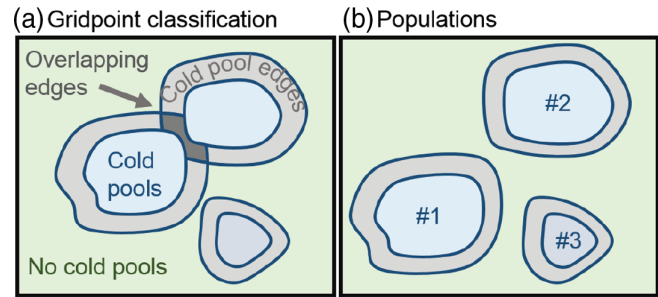


FIGURE 4 Schematic illustration of (a) classification of gridpoints into cold pools, cold-pool boundaries and no cold pools; and (b) cold-pool populations

Overlapping cold-pool edge regions are also included in the cold-pool edge category, as discussed below.

2.5.2 | Cold-pool objects

We can further identify each single cold pool as an object and compute cold-pool characteristics averaged or integrated over its cold-pool region (Figure 4b). To each cold pool we associate a cold-pool edge, which represents the 25 km wide boundary region defined above. Recall that gridpoints belonging to a cold pool are excluded from the cold-pool edge. Regions that overlap with another cold pool's edge region are included. This means that some gridpoints may be counted as part of the cold-pool edge region for more than one cold pool. Again, for each cold-pool edge, area-averaged or -integrated quantities can be computed, but also 95th percentiles to represent the more extreme behaviour of the actual gust fronts. As defined in Table 1, eight quantities are used to describe the essential characteristics of cold pools and their boundaries. For the cold pool, these are intensity, downdraught mass flux, radius and precipitation, while for the cold-pool edge region, we use vertical velocity, mass flux, and buoyancy. The relative frequency of gridpoints where CI occurs defines the probability of CI, $P[\text{CI}]$, or *triggering probability*.

2.5.3 | Cold-pool populations

Given these characteristics, we can compare the distributions of cold-pool properties (populations) between model resolutions. For this purpose we will generally show the median of the cold-pool distributions and vertical bars for the 95% confidence interval which is computed via bootstrapping with 1,000 draws. We note that cold pools from 5 min output are used and that the same temporally correlated cold pools appear more than once in the cold-pool

TABLE 1 Quantities that represent essential characteristics of cold pools and their gust fronts

Variables	Region	Description
Intensity	Cold pool	Buoyancy anomaly θ'_p integrated vertically over the five lowest model levels (≈ 150 m). Area averages (mean) or integrals (sum) are considered.
D mass flux	Cold pool	Downward mass flux, i.e., negative values of ρw at ~ 1 km, integrated over each cold-pool area (sum).
Radius	Cold pool	Equivalent radius, estimated as the radius of a circle with the same area as the cold pool.
Precipitation	Cold pool	Precipitation inside cold pools, integrated over the cold-pool area.
Gust front vertical velocity	Edges	95th percentile of vertical velocity in the cold-pool boundary region at ~ 1 km height. This height was visually identified as the height of most pronounced gust fronts.
Gust front mass flux	Edges	Upward mass flux inside the cold-pool edge regions at 1 km height. Both the area integrated (sum) and area average (mean) are used.
Buoyancy	Edges	95th percentile of buoyancy at ~ 150 m height. Potential temperature, moisture or surface fluxes are not considered, as buoyancy is expected to hold most of their information relevant for convective initiation.
Triggering probability, $p[CI]$	Edges	Number of grid points with convective initiation divided by the number of grid points in the cold-pool edge region.

populations. Hence the data are not independent and the significance of our results will likely be overestimated. This possibility has been briefly tested and confirmed by sub-sampling the data every full hour.

2.6 | Estimating linear causal effects

To determine the processes through which model resolution affects cold-pool-driven CI, we apply linear causal effect estimation. In general, associational expressions like regression estimates do not prove causal relationships – correlation is not causation² – because common drivers may produce associations between two variables without one causing the other. Furthermore, in complex systems, there may be a chain of causality through different variables, and it is of interest to quantify and compare different pathways from the original cause to the effect, since they may represent different physical mechanisms. To do this, we follow the framework reviewed by Chen and Pearl (2014) and Pearl (2013), based on Pearl's causality (Pearl, 2009) and the graphical representation of causal structures, first proposed by Wright (1921). This framework is based on the causal graph of the underlying phenomenon which has to be determined using *a priori* process understanding. We will assume that all variables are linearly related, so that multiple linear regressions can be used to determine single causal effects. The graphical

analysis can also be applied for nonlinear relationships, but for this work we will avoid the additional complexity. Importantly, the required predictors for the multiple linear regressions are selected based on the structure of the causal graph by the so-called *single-door criterion*. The remainder of this section will more carefully define a causal effect, and illustrate the analysis method with some simple examples of causal graphs. To define the single-door criterion, we will need to introduce additional concepts, including directed acyclic graphs, collider and d-separation. Note that these latter concepts are not needed to understand the application of causal analysis to our cold-pool experiments in Section 3.3, provided that the reader is willing to accept the structure of the multiple linear regressions that results from the analysis. For a more thorough introduction to linear causal analysis, we refer the reader to Chen and Pearl (2014).

A **causal effect** of X on Y can be defined as the change of the expected value of Y when X is changed actively (by intervention) from $X = x$ to $X = x + \Delta x$. The active intervention is crucial because it eliminates any factors that otherwise impact X . Consequently, randomized experiments or model sensitivity studies are commonly performed to quantify causal effects. For example, X can be the wetness of grass and Y the wetness of the adjacent street. If one wants to test the hypothesis that the wet grass causes the wet street (Figure 5a), simply computing the correlation of the wetness of the street with the wetness of the grass will give a biased estimate, because the common cause of rain has not been included. One way to avoid such a spurious correlation is by active intervention. If the grass is

²But they are correlated.

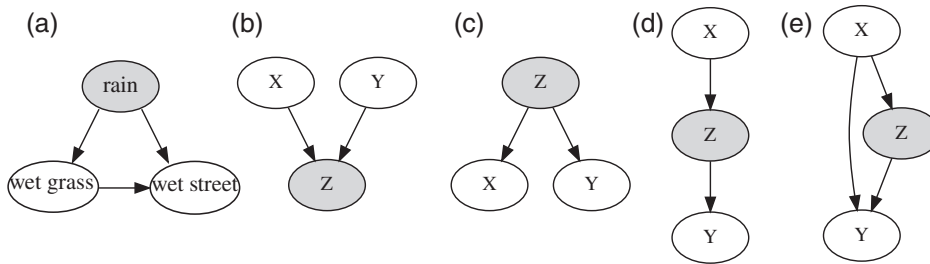


FIGURE 5 Examples of directed acyclic graphs illustrating the concepts of a collider, d-separation or the single-door criterion. The text gives details

actively moistened, for example by using a sprinkler, the correlation with a wet street will be reduced.

Fortunately, under certain circumstances, it is possible to determine causal effects solely from associations, without doing any intervention. In the simple example of wet grass and street, the solution would be to condition on the presence of rain, e.g. by including rain as an additional predictor in a linear regression. On rainy days, everything will be wet, but on dry days, the wetness of the grass and street is determined by sprinklers and street cleaners and will not be correlated, thus reducing the overall spurious correlation. For more complex systems, it can be difficult to determine a set of parameters that can be conditioned on to allow a causal connection to be identified, or even if such a set of parameters exists. Here we identify these circumstances with the so called *single-door criterion*, which makes use of the following concepts.

Directed acyclic graphs are graphs with directed edges and no cyclic structures. They are powerful tools to visualize causal structures and enable graphical criteria to estimate causal effects. Example graphs are shown in Figure 5. In this application, the nodes of a directed acyclic graph represent variables of a causal model, which are causally influenced by their parent nodes and themselves cause changes in child nodes. Parents and children of a node are determined by the direction of the arrows connecting them: arrows pointing towards a node emanate from its parent nodes and arrows emanating from a node point to its child nodes. In the example of Figure 5b, Z is a child of both X and Y, whereas in Figure 5c, Z is a parent of X and Y. The concepts of ancestors and descendants of a node follow naturally. Assuming linearity, a directed edge connecting a parent and a child node can be associated with single *path coefficient* representing the direct causal effect, that is, the rate of change of the child variable, when the parent is actively perturbed. It is these path coefficients which we wish to estimate by appropriate multiple linear regressions.

A **collider** is a special element on a path in a causal graph, where two arrow heads collide at one node. If a path between X and Y passes through a collider, that path is said to be *blocked*: despite its connection by the colliding edges, no information will pass through this pathway. In

Figure 5b, for instance, Z is a collider for the path connecting X and Y through Z. Collider nodes will have important consequences as described in the following paragraph.

Directed separation, or **d-separation** encapsulates the graphical conditions that correspond to a missing causal relationship. There are three rules for two variables to be d-separated (Chen and Pearl, 2014):

1. If there is no *active* path between X and Y, the variables X and Y are d-separated. An active path refers to any path that is not blocked by a collider. In Figure 5b X and Y are d-separated as the only path is blocked by the collider Z. In Figure 5c–e, X and Y are not d-separated as active paths exist.
2. X and Y can also be d-separated *conditioned on a set of nodes Z*, if there is no active path between X and Y without traversing Z. Meaning, if Z is included in all active paths from X to Y, X and Y are d-separated conditioned on Z. In Figure 5c,d, X and Y are only d-separated if conditioned on Z, as the only active path between X and Y is via Z.
3. If a conditional set Z contains a collider or a child of a collider, X and Y cannot be d-separated conditioned on Z. Hence, X and Y in Figure 5b are not d-separated conditioned on Z.

If two variables X and Y are d-separated conditioned on a set of nodes Z, the partial correlation coefficients between X and Y, where Z is accounted for, vanish (e.g., Chen and Pearl, 2014). By applying this concept, causal graphs can be linked to statistical (conditional) independence.

With these concepts, we can now state the conditions under which we can determine causal effects from non-interventional data. In a linear framework this is formally given by the **single-door criterion**. It states that the direct causal effect of X on Y within a causal graph G is *identifiable* if a set of variables Z exists with the following criteria:

- (a) Z contains no descendent of Y, and
- (b) X and Y are d-separated conditioned on Z in the hypothetical graph G' where the direct path from X to Y is removed (this implies that Z does not contain a collider or a child of a collider).

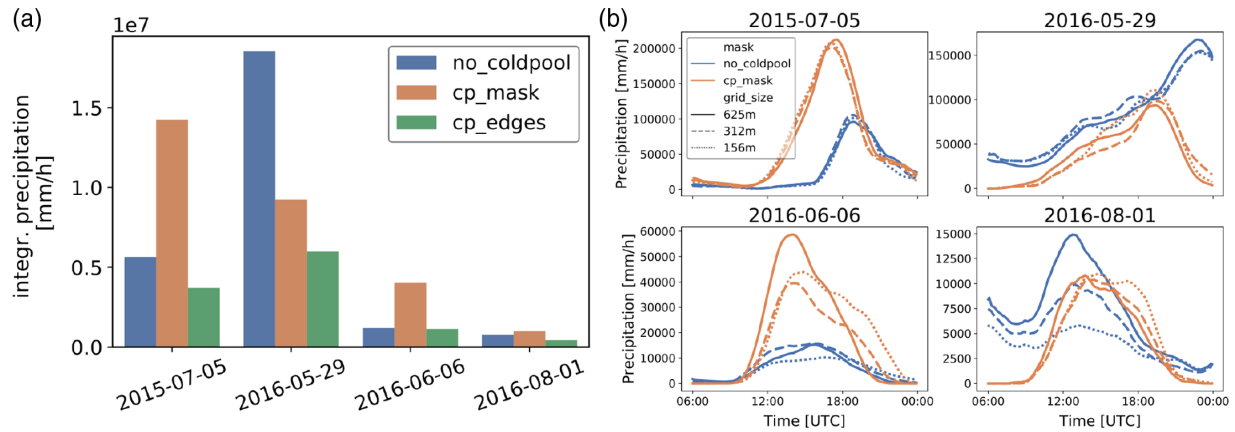


FIGURE 6 (a) Domain- and time-integrated precipitation for the 156 m resolution simulation outside cold pools, within cold pools and within cold-pool boundaries. (b) Time series of domain-integrated precipitation within cold pools and outside cold pools. A 3 hr running mean has been applied to smooth the data

Most importantly, such identifiable path coefficients are equal to the regression coefficient that is obtained from multiple linear regression predicting Y using not just X but also Z as predictors. We will refer to Z as the *adjustment set*. The aggregated causal effects of parallel or sequential paths can then be derived from single-path estimates by addition or multiplication, respectively.

Considering the example graph in Figure 5e, Z d-separates X from Y in the graph where the direct X – Y path is removed (here corresponding to Figure 5d). Hence the direct effect of X on Y is identifiable and can be estimated via the regression estimate of X from a linear regression where both X and Z are used as predictors and Y is the predictand. To estimate the direct causal effects between X and Y in Figure 5b,c,d (which are all zero due to missing, direct links) Z has to be included in Figure 5c,d, whereas it must not be included in Figure 5b. Returning to the example in Figure 5a, the single-door criterion shows that, to estimate the causal effect of the wetness of grass on the wetness of the street from observational data (no intervention), the rain amount has to be included in the linear regression.³

In summary, this framework enables a graphical criterion, namely the single-door criterion, which specifies if and how causal conclusions can be drawn from purely statistical associations. However, the causal, graphical structure must be provided using *a priori* understanding

of the governing processes. We apply this framework to identify the impact of model resolution on cold pools and CI in Section 3.3.

3 | RESULTS

3.1 | Linking precipitation and convective initiation with cold pools

The first question we address is the relevance of cold pools for precipitation and especially for its diurnal cycle and organization. We focus on the 156 m grid size simulations as the qualitative results are mostly similar for the other two resolutions. We compare temporally and spatially integrated precipitation amounts inside cold pools (*cp_mask*) and outside cold pools (*no coldpool*) in Figure 6a for each day using the gridpoint classification described previously. First, we consider some general behaviour of precipitation inside and outside cold pools on these four days. The magnitude of total precipitation clearly depends on the day, with 29 May 2016 showing the highest total amplitude and 1 August 2016 the lowest. But the relative importance of precipitation inside cold pools also varies: on 6 June 2016 precipitation inside cold pools (*cp_mask*) was approximately three times greater than outside cold pools whereas on 29 May 2016 precipitation outside cold pools dominates. Nonetheless, precipitation inside cold pools contributes substantially for all days, ranging from approximately 50 to 300% of precipitation outside cold pools. This corroborates that precipitation predominantly produces cold pools and indicates the potential importance of cold pools. Figure 6a also shows that significant amounts of precipitation occur in cold-pool edge regions, which provides a first indication of the role of cold pools in initiating convection.

³The single-door criterion is a sufficient, but not a necessary, criterion to identify causal effects. Further criteria – the backdoor criteria and instrumental variables – enable additional identification (Chen and Pearl, 2014). However, in our case, the single-door criterion is adequate. It is also possible to identify different indirect causal effects for nonlinear or non-parametric models (Pearl, 2014). However, the necessary conditions for identification are more restrictive and the required estimation methods more complex.

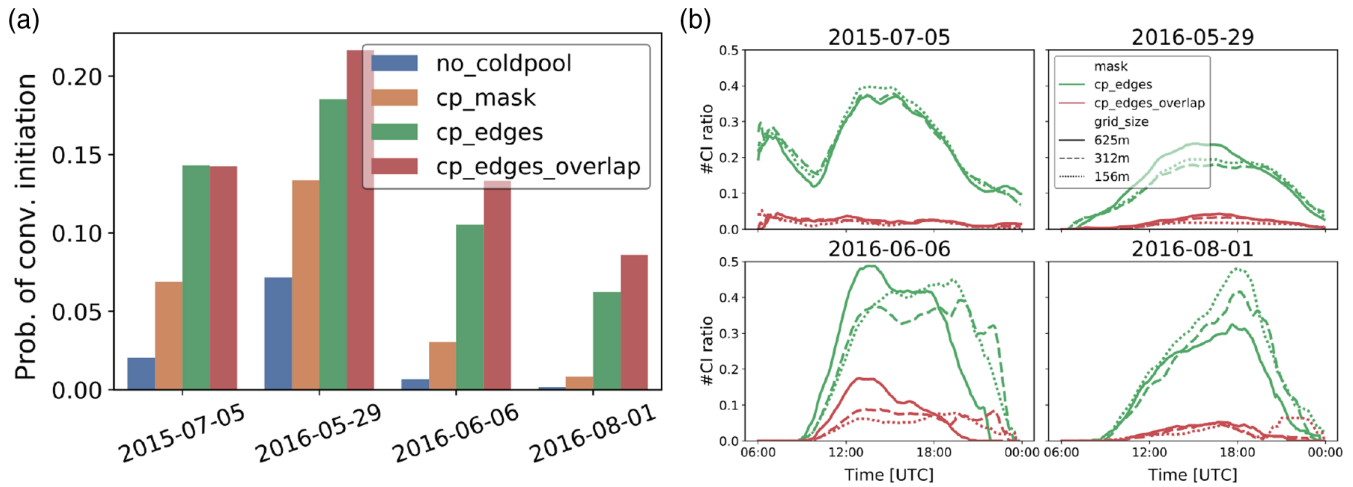


FIGURE 7 (a) Triggering probability in the four grid point categories. (b) Ratio of convective initiation inside cold-pool boundaries or overlapping cold-pool boundaries to total convective initiation for each day and model resolution. A 3 hr running mean has been applied to reduce noise

Time series of domain-integrated precipitation for the different model resolutions are shown in Figure 6b. Again, we find a strong dependence of the diurnal cycle on the synoptic conditions. In general, cold-pool-related precipitation, but also *no coldpool* precipitation, tends to be more dominant during the day and decreases in the evening. Also, the phase and amplitudes of the diurnal cycles do not completely coincide with the diurnal cycle of precipitation outside cold pools. On 5 July 2015, when two enormous cold pools develop, cold-pool precipitation increases until the early afternoon without noteworthy precipitation outside cold pools. On 29 May 2016, the increase in precipitation during the day occurs both inside and outside cold pools. However, precipitation outside cold pools continues to increase throughout the night, likely related to synoptically driven ascent. Interestingly, on the other two days, the rapid evening decline of cold-pool-related precipitation tends to occur later than for precipitation outside cold pools, again suggesting a role for cold pools in maintaining convection.

We now focus on the connection between cold-pool gust fronts and convective initiation, which indicates the influence of cold pools on precipitation. Figure 7a shows the triggering probability as a function of gridpoint classification and day. In *no_coldpool* regions, we find the lowest probability of convective initiation for all days. Triggering is substantially enhanced inside cold pools. Note that this behaviour is based on our definition of convective initiation as strong increase in cloud water content and several possible explanations exist for convective initiation inside cold pools.⁴ As expected, the cold-pool edge regions

have even higher triggering probability. Depending on the day, from 7% to almost 20% of gridpoints inside cold-pool boundaries initiate convection. The highest triggering probability on most days is found for overlapping edges, with an increase in probability of CI of approximately 3% in comparison to all cold-pool edges. These findings are consistent with previous studies showing that cold pools are very efficient at initiating new convection nearby.

To estimate the absolute impact of cold pools on total convective initiation, Figure 7b displays the diurnal cycle of convective initiation within cold-pool edges and overlapping edges relative to total convective initiation. We find that triggering by cold-pool edges account for up to 50% of total triggering. For the last two days this is especially strong in the afternoon and evening. Although we have seen a high efficiency in triggering by overlapping edges, their contribution to the total convective initiation is small (Figure 7b), being usually well under 10%.

The results address the first research question, showing that cold-pool-driven convective initiation plays a substantial role for the diurnal cycle and the organization of convection. Strong links between precipitation and cold pools are observed, with cold pool triggering contributing up to 50% of total convective initiation with a particular relevance in the evening in some synoptic situations. Cold-pool boundaries are found to be particularly important for triggering of convection. In two cases this impact is felt strongly during the evening, when non-cold-pool triggering has decreased in strength.

⁴These explanations include: (a) merged cold pools which may still have active gust fronts within the identified cold-pool regions, (b) gust fronts

at the edge, but still inside identified cold-pool regions, and (c) rapid growth or propagation (not advection) of existing cells within the cold pool. A further reason could be the wrong identification of advected cells.

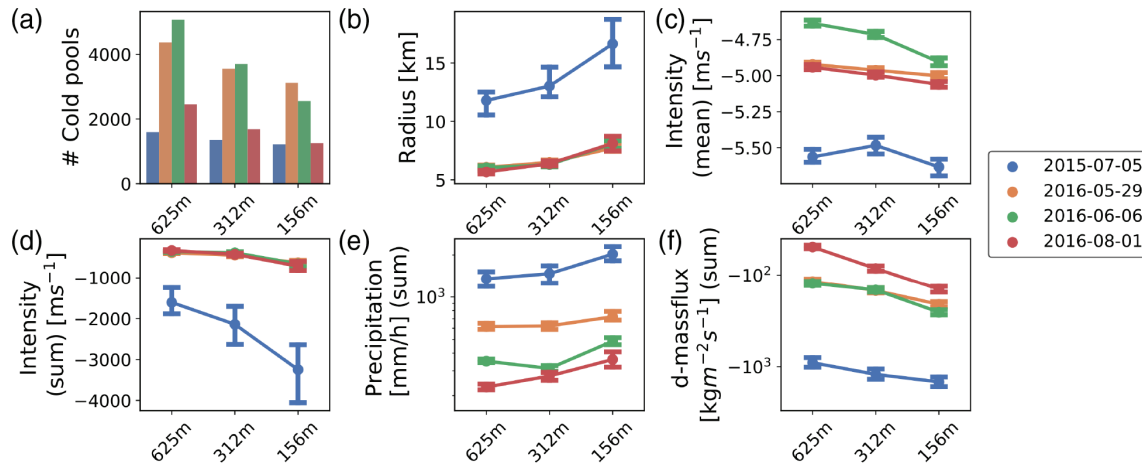


FIGURE 8 Overview of differences in cold-pool populations: (a) number of cold pools, (b) equivalent cold-pool radius (derived from cold-pool area), (c) cold-pool mean intensity, (d) cold-pool integrated intensity, (e) cold-pool integrated precipitation and (f) cold-pool integrated downward mass flux. Except for the cold-pool frequency, the median of the distributions is displayed with 95% confidence intervals estimated by bootstrapping

3.2 | Sensitivity of cold-pool properties to resolution

Having established strong indications for the influence of cold pools on the initiation and diurnal cycle of convection, it follows that the appropriate representation of cold-pool dynamics in convection-permitting models is vital. We will evaluate whether cold pools and the related CI are indeed appropriately represented at all resolutions and hence insensitive to model resolution.

For the first two days in Figure 6b, no systematic differences in the diurnal cycle of precipitation between model resolutions are visible. For the third case, 6 June 2016, the coarsest model shows a higher peak in the early afternoon and increased model resolution results in a shift of cold-pool-related precipitation towards the evening. A similar, but less pronounced behaviour is also visible for the last day, 1 August 2016. Surprisingly, on this day, precipitation outside cold pools shows a relatively strong sensitivity to model resolution, related to many small, scattered precipitation cells over Northern Germany in the coarsest model, which become less frequent at higher resolutions.

To investigate the resolution dependence of cold pools more systematically, we now examine distributions of cold-pool properties. Figure 8 summarizes the most evident differences between resolutions. For all four days, the number of cold pools is reduced at higher resolutions. The size of cold pools (i.e., the equivalent radius), on the other hand, increases with resolution. These observations are also in accordance with the visual impression given in Figure 3. Cold pools are generally more intense (integrated or averaged over their areas), and have stronger

precipitation (mainly on the first and last days) and integrated downward mass flux at higher resolutions. Some of these differences between model resolutions are relatively small compared to the day-to-day variability. Especially the first day (5 July 2017) is different, since it has only a few, enormous, cold pools (Figure 1). Overall, the dependencies of cold-pool properties are quite systematic, despite the large differences in their absolute values from day to day.

Having seen systematic dependencies of cold-pool properties on resolution, we now look for corresponding sensitivities in CI and cold-pool edge properties that may impact initiation. Comparing the model resolutions in Figure 7b, the last two days show a substantial sensitivity to the relative importance of cold-pool-driven CI in the cold-pool boundary region. This sensitivity becomes especially evident in the evening, when the higher resolutions are associated with more persistent convective activity (6 June 2016), or simply a higher peak (1 August 2016). Again we consider cold-pool objects and focus now on the characteristics of their edges. Medians of CI-related variables are displayed in Figure 9, namely triggering probability, mean and total upward mass flux, and 95th percentiles of vertical velocity and buoyancy. Figure 9 shows a modest tendency towards increasing triggering probability with higher model resolution. Mean and total upward mass flux and 95th percentile of vertical velocity increase systematically with resolution, although the strength of the increase varies between days. In contrast, the 95th percentile of buoyancy at the cold-pool edge does not show a systematic dependence. While the first day shows enhanced buoyancy in higher resolutions, the buoyancy on the other three days is relatively insensitive, or even decreases slightly with model resolution.

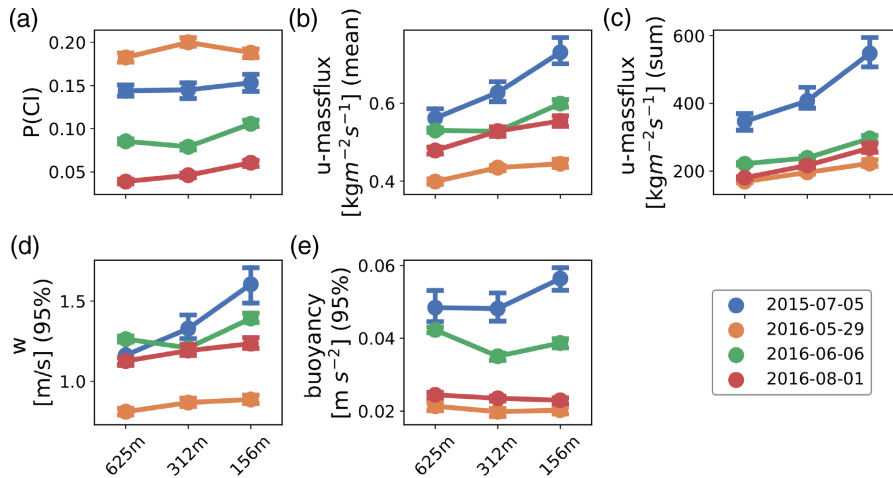


FIGURE 9 Overview of differences in cold-pool populations with regard to their boundary regions and convective initiation: (a) triggering probability, (b) upward mass flux averaged over each cold-pool boundary (1 km), (c) upward mass flux integrated over the cold-pool boundary area, (d) 95th percentile vertical velocity (1 km) at cold-pool boundaries, and (e) 95th percentile of buoyancy (150 m) at cold-pool boundaries. Median of the distributions is displayed with 95% confidence intervals estimated by bootstrapping

Concerning our second research question, we find that coarser models have more, but smaller and less intense, cold pools. Their cold pools also have weaker gust fronts (with respect to upward mass flux/vertical velocity) and a tendency towards reduced initiation of convection. The combination of these differences are also evident in diurnal cycles of total cold-pool precipitation.

3.3 | Identifying causes of the resolution dependence of convective initiation

Given the results from the previous section, can we conclude that the weaker vertical velocity in coarser models causes a reduction of CI? Figure 10a shows that higher mean upward mass flux is related to higher triggering, which would endorse such a conclusion. However, the triggering differences might be caused by differences in cold-pool intensities. And while we observe differences between gust front strengths for all days, the difference in CI is only clear for the last two days. Figure 10b further seems to suggest that an increase in buoyancy at the cold-pool boundaries is related to more CI. Can we still neglect buoyancy because differences between resolutions are not as systematic?

To disentangle such different aspects, we follow the approach of Pearl (2013) and Chen and Pearl (2014) to investigate the causal pathways under the assumption of linearity (as described in Section 2 for a short introduction). Our goal is to identify how model resolution R impacts convective initiation $P[CI]$. We distinguish between the formulation of our causal model (Section 3.3.1) and the estimation of causal path coefficients (Section 3.3.2). In subsection 3.3.1, we describe the design of a causal graph representing our conceptual understanding of the relevant pathways through which changes in model resolution may impact CI. We identify

which indirect effects are reasonable and which confounding variables exist, and specify a causal graph structure with several acyclic, directed pathways from resolution R to triggering probability $p[CI]$. In subsection 3.3.2, we quantify these different pathways by applying multi-linear regression. After normalizing the data by subtracting the mean and dividing by standard deviation, we estimate each path coefficient using a linear regression and include additional variables as predictors, as specified by the single-door criterion. Aggregated pathways are then derived by multiplication or addition of sequential or parallel paths. For statistical uncertainty quantification, we apply bootstrapping ($N=1,000$). The magnitudes of the different, aggregated pathways allow us to draw conclusions about the relative importance of each pathway. To evaluate the robustness of our causal model and the subsequent estimated causal effects, we further evaluate the sensitivity of our model to some modifications. (Section 3.3.3).

3.3.1 | Causal structure

To determine the dominant effect of model resolution on triggering probability by cold pools, we propose a causal structure as illustrated in Figure 11. The selected variables of interest are model resolution (R), cold-pool integrated intensity (I), cold-pool boundary 95th percentile buoyancy (B), mean upward mass flux at cold-pool boundaries (G), and triggering probability (P) at cold-pool boundaries.

Model resolution Our goal is to identify the main pathway through which model resolution (R) impacts convective initiation (P). A direct effect is represented by the arrow from R to P . This pathway could also incorporate indirect pathways that are not represented in the causal model, for example, differences in vertical acceleration,

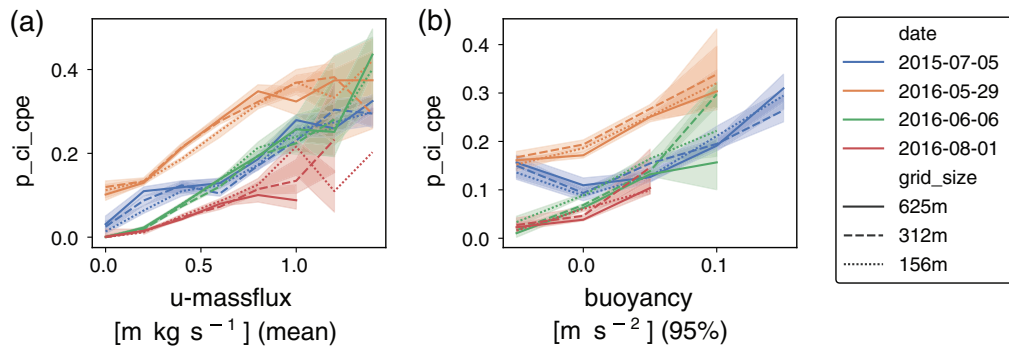


FIGURE 10 Dependence of triggering probability at cold-pool boundaries on (a) mean upward mass flux (1 km) and (b) 95th percentile buoyancy (150 m). Triggering probability has been binned according to mass flux and buoyancy. To reduce sparse data ranges, values below or above the displayed ranges of mass flux and buoyancy have been set to the threshold values

condensation processes or entrainment/detrainment due to model resolution. We have further observed that model resolution affects cold-pool intensity and vertical velocity. These dependencies can be related to microphysical processes, surface fluxes or vertical acceleration. For buoyancy, the influence was not found to be very systematic, but we evaluate its strength in this model anyway. These causal relationships are represented by the arrows from R to I, G and B.

Cold-pool intensity It is reasonable to assume that stronger cold pools have different buoyancy (B) or vertical velocity (G) at the cold-pool boundary. Hence, we draw two paths IB and IG. We have further included a direct arrow from cold-pool intensity to CI to evaluate whether we are missing some important processes that are not captured by buoyancy or vertical velocity (IP).

Buoyancy Buoyancy certainly influences vertical velocity, but there might also be a direct effect on triggering. These effects are represented by the arrows BG and BP.

Gust front The impact of gust front upward mass flux (mean) on CI is represented by the arrow GP.

Triggering We don't consider any impact of triggering on other variables. Hence there is no feedback and we consequently have a directed acyclic graph. This is a necessary feature to enable the following estimation of causal pathways.

Furthermore, we have included the synoptic situation and diurnal cycle (D), as represented by the variables given in Figure 2. This is necessary because we expect them to be common drivers for cold-pool intensity, buoyancy, gust front and CI. However, time of day is not taken into account explicitly.

Several noteworthy assumptions have been made with regards to the causal structure. First, no temporal evolution or time lag is considered. Hence we assume that the autocorrelation time-scale of the variables of interest

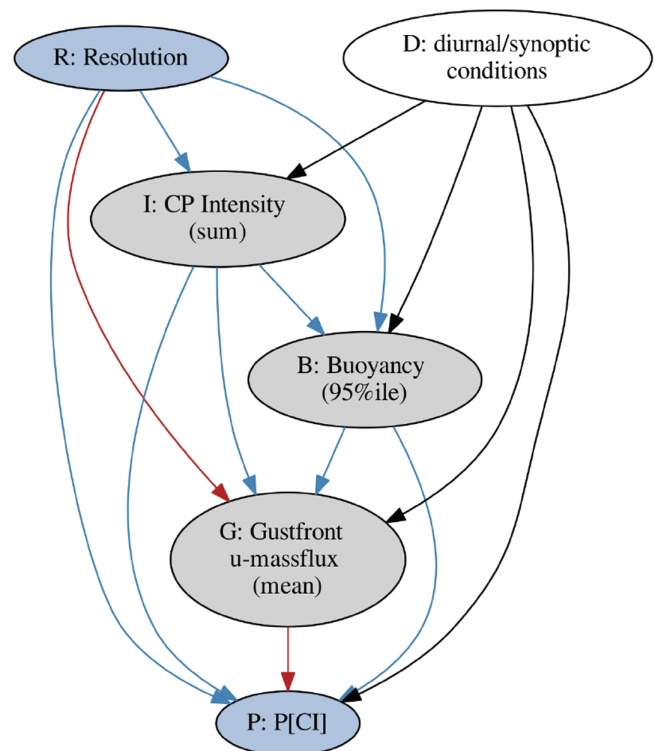


FIGURE 11 Conceptual causal model of how model resolution can impact convective initiation. Black arrows are used to emphasize their role as common driver. Blue/red arrows are used for paths that influence how model resolution (R) impacts convection initiation (P). The red paths display the identified dominant pathway. The text gives further explanation

is longer than the time it takes for stronger cold-pool intensities to affect gust fronts, buoyancy or CI. This also means that the data are not independent, as subsequent time steps may include the same, temporally correlated, cold pool. The causal time series approach introduced by Runge *et al.* (2015) is likely better suited for time series and we consider its application to tracked cold pools

in the future. We also neglect any feedback cycles, for example, that stronger CI might lead to a strengthening of cold pools. This is justified if the autocorrelation time-scale of the variables of interest is expected to be smaller than the 15–30 min time-scale for precipitation development. Furthermore, no impact of model resolution on synoptic conditions is assumed (no link between R and D). This is justified by the small resolution sensitivity of the domain-aggregated, large-scale variables in Figure 2. Finally, we assume that we have captured all major variables that have a strong influence on at least two of the variables of interest (assumption of no unobserved common drivers).

3.3.2 | Estimating causal effects

To estimate the path coefficients of the causal graph, we apply multiple linear regression based on the single-door criterion using all identified cold pools at all time steps. The predictor sets considered for each path are listed in Appendix B. The resulting average path estimates are illustrated in Figure 12a. As expected, strong paths are observed between buoyancy and gust front upward mass flux, and also between gust front mass flux and triggering probability. However the impact of buoyancy on triggering probability seems to be completely mediated by the gust front. More intense cold pools lead to an increase in buoyancy and a small decrease in upward mass flux (the signs of these paths arise from the fact that intensity is negative). The graph further shows that, in comparison to synoptic conditions, model resolution is of secondary importance for the variables of interest. Hence, all models are able to capture the dominant, physical processes. This is to be expected and does not undermine the importance of comparably small, systematic biases due to model resolution.

To evaluate dominant direct and indirect pathways between resolution and triggering probability, aggregated path estimates are displayed in Figure 12b. The * in $R * B * P$ denotes that all possible direct or indirect paths between the two specified variables are accounted for. This shows that the RP , RIP , $RIGP$ and $R * B * P$ paths are relatively small. Hence, the dependence of the cold-pool intensity and the buoyancy on resolution do not play an important role for triggering probability. The RGP path clearly dominates: larger grid sizes decrease upward mass flux at the cold-pool boundaries, which consequently reduces CI. The displayed, de-normalized value of -0.017 means that if model grid size is increased from 156 to 625 m, triggering probability will be reduced by 0.017. This is somewhat less than some of the differences observed in Figure 9, but is still of comparable magnitude. Furthermore, this

TABLE 2 Estimation of direct path coefficients. To allow the estimation of path coefficients by multiple linear regression, adjustment sets need to be included

Path	Adjustment set
R→I	[D]
R→B	I, D
I→B	R, D
R→G	I, B, D
I→G	R, B, D
B→G	R, I, D
R→P	I, B, G, D
I→P	R, B, G, D
B→P	R, I, G, D
G→P	R, I, B, D

Note: The necessary adjustment sets are given by the single-door criterion. They are then used as additional predictors (together with the parent of the direct path) within the multiple linear regression for predicting the child of the given path. For the R→I path, the adjustment set [D] is not strictly necessary, since no link between RD is assumed.

corresponds to a difference of 10–30% in median triggering probability.

3.3.3 | Robustness

The statistical robustness of the results and their sensitivity to several of the assumptions made in constructing the causal graph are examined in Appendix C3. This includes tests of the statistical significance of the linear regression results, the linearity of the response to changing model resolution, selecting different variables as proxies for gust front or cold-pool intensity, sensitivity to stratification by days, and of the impact of model resolution on the synoptic conditions. One noteworthy concern arises from sometimes relatively small R^2 values, which can imply that several informative variables may not be included in our causal graph. This can increase the likelihood of missing a common driver, which causes biased causal effect estimates (Appendix C3. gives more details).

Nonetheless, the causal effect estimates vary moderately under most modifications with no strong dependence on the specific resolution range or the selected variables. Only the stratification of the data into the four individual days substantially influences the causal effects. The path estimates computed for each of the four days are

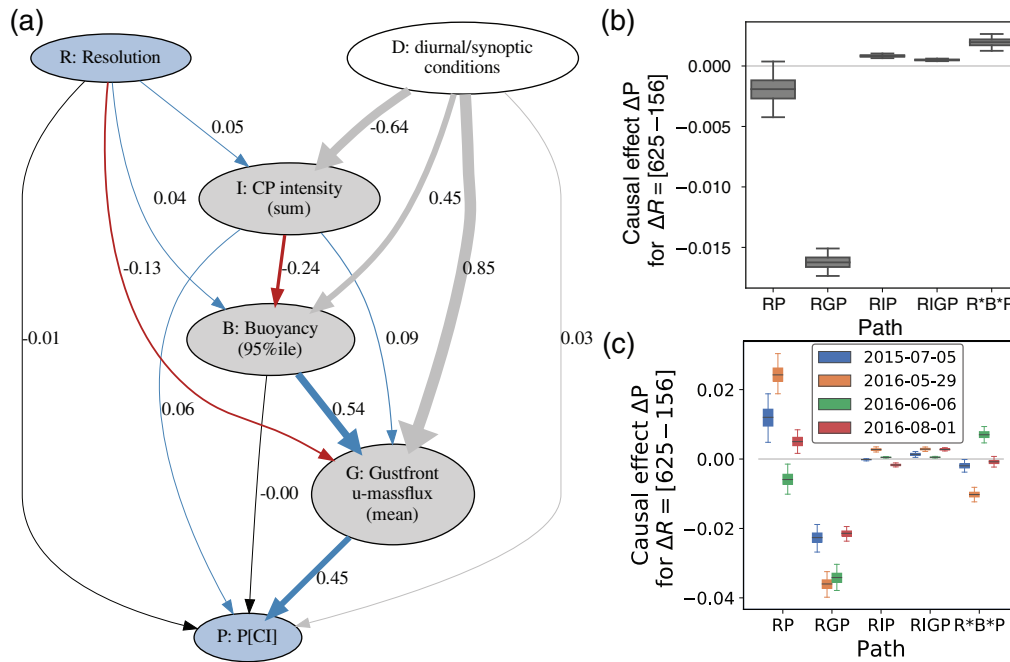


FIGURE 12 (a) Causal structural model of how model resolution can impact convective initiation. Single estimated path coefficients (normalized) are displayed and qualitatively visualized by the thickness of the various paths. Negative arrows (red) indicate that an increase in the parent variable results in a decrease of the child variable. The paths from the diurnal/synoptic variables display the sums over all diurnal/synoptic variables for better visualization and are coloured in grey. (b) Aggregated path estimates, de-normalized to represent the causal effect of changing model resolution from 625 to 156 m on triggering probability. (c) as is (b), but with stratification by days. Note that (b) cannot be obtained simply as the average over the four days of (c). The whiskers of the box plots correspond to the 95% confidence intervals from bootstrapping

illustrated in Figure 12c. The estimates for the paths *RIP* and *RIGP* show some variability between days, but the values are all close to zero. The path estimates of the *R*B*P* path show stronger variability among the days, although centred around zero. The *RGP* path, which we have earlier identified as the dominant path, shows moderate variability between the days, with values ranging from -0.02 to -0.04 . Thus, this effect is relatively robust to the stratification by days and is consistently stronger than other indirect effects. On the other hand, the direct effect of resolution on triggering (*RP*) shows a strong sensitivity to the day, ranging from 0.03 to -0.01 . Since the direct *RP* path includes influences of causal processes not represented explicitly in the graph, it is interesting to note that it is strongest on 29 May 2016, when the diurnal cycle of precipitation is disturbed by strong synoptic forcing. In such situations the initiation of convection may be less influenced by cold-pool effects.

Overall, we conclude that the relatively robust *RGP* path predominates in most situations. Hence, to answer our third research question, the reduction in cold-pool-driven CI in coarser models is a direct result of reduced upward mass flux at cold-pool boundaries. However, the eventual impact on precipitation is very dependent on the weather situation, since the

dominant *RGP* path can be partially compensated by other processes.

4 | SUMMARY AND DISCUSSION

The aim of this work is to identify the relevance of cold pools for the diurnal cycle and the organization of convection; whether convection-permitting models can adequately represent cold-pool-driven CI; and which misrepresented processes dominate the deficiencies in cold-pool-driven CI. To do so, we compared available high-resolution ICON simulations with grid sizes 156, 312 and 625 m of four selected days over Germany and identified cold pools, cold-pool boundaries and CI locations.

The first research question was to determine whether cold pools are a significant factor for the diurnal cycle and organization of convection. To address this, we have evaluated identified cold pools and quantified the related precipitation and CI. We found strong links between precipitation and cold pools. Although day-to-day variability is high, cold-pool-driven CI can be as much as 50% of the total CI. For two of the four days, this is especially prominent in the late afternoon and evening. Cold-pool

boundary regions, where gust fronts occur, were identified as very efficient convection initiators, with triggering probabilities several times higher than outside cold pools. Consistent with previous studies (e.g., Böing *et al.*, 2012; Schlemmer and Hohenegger, 2014; Feng *et al.*, 2015), these results demonstrate the importance of cold pools for convection organization and the diurnal cycle of convection, but also the strong role of the synoptic environment.

The second research question asks whether convection-permitting models can accurately represent the CI process. This was addressed by evaluating the sensitivity of cold-pool-related properties to model resolution, using simulations with grid sizes from 156 m, where there is some hope of resolving the relevant processes, to 625 m, where the characteristic errors of convection-permitting models should be apparent. For two of the four days, we identified a significant sensitivity of the diurnal cycle of precipitation and CI to model resolution. The cold pools themselves were found to be smaller and less intense at coarser resolutions, which is in agreement with the results from Squitieri and Gallus (2020). Regarding the influence of model resolution on CI, a reduced triggering probability with increasing grid size was found for two of the four simulated days. Vertical velocity and upward mass flux at the gust fronts was reduced with increasing grid sizes, while buoyancy in this region did not show a systematic sensitivity. These findings show that coarser models have difficulties in representing cold-pool-driven CI. However the specific character seems to depend on the large-scale situation.

The third research question is concerned with which errors in the coarse-resolution simulations lead to the observed differences in CI with changing model resolution. To evaluate the relative importance of different causal mechanisms, we applied a linear causal effect estimation based on causal graphs. Again the day-to-day variability is large, but a systematic and dominant causal effect was identified: coarser models yield weaker updraughts at the cold-pool gust fronts which then cause reduced CI. Since the intensity of the cold pools and the buoyancy of the lifted air do not appear to play a role in the sensitivity, it is likely that the reduced upward motion at the boundary is simply a result of increased numerical diffusion at the gust front.

The causal analysis is based on several assumptions of the underlying model, which have been mentioned and tested to some extent in Sections 3.3.1, 3.3.3 and Appendix C3. Open issues include the possibility of missing common drivers, for example, due to neglecting the temporal evolution of cold pools or due to missing other variables. This possibility is strengthened by moderately high R^2 values (Appendix C3.). Furthermore, by using identified cold pools from 5 min model output, subsequent time steps

include the same cold pools and the data are not independent. This may cause an overestimation of the significance. Extensive testing of the robustness of the causal analysis revealed one important sensitivity, namely the dependence on the different synoptic conditions present on the four days. This resulted in large variability of the direct effect of model resolution on triggering probability (RP). On some days, this direct effect partially compensates the effect of reduced CI by weaker gust fronts (RGP), whereas on other days, the effect is negligible or complements the RGP path. Since the direct path includes the effects of mechanisms not represented in the causal graph (which was focused on cold-pool processes), it may become important when strong synoptic forcing leads to convective quasi-equilibrium and local triggering processes are less relevant (Keil *et al.*, 2014). To investigate such processes thoroughly will require a much larger number of simulation days than are currently available. However, it is worth recalling that there was no day in the current analysis where the RP path became as strong as the dominant RGP path.

While we investigated the causes for the reduced triggering probability for our third research question, we have not addressed the causes for the differences in cold-pool sizes and intensities. We speculate that the following processes may be relevant. First, the representation of cloud microphysics can have a substantial influence on precipitating downdraughts or on sub-cloud layer evaporation, which can in turn influence cold-pool numbers, sizes and intensities. Entrainment or detrainment into or from the downdraught or the cold pool can also be relevant. They are mostly determined by turbulence parametrizations, which show substantial deficiencies even in models with resolutions of a hundreds of metres, often denoted as the grey zone of turbulence parametrization (e.g., Honnert, 2016). Surface fluxes have also been found to have a significant impact on cold pools (Gentine *et al.*, 2016; Grant and van den Heever, 2018) and can be sensitive to the resolution of small-scale surface features. Finally, stronger CI at the gust front can result in reinforcement of the original cold pool and thereby increase cold-pool size and intensity (Böing *et al.*, 2012; Schlemmer and Hohenegger, 2014). Further analysis of the statistics of cold pools and the mechanisms that influence them within the ICON simulations is planned for a future publication.

The original motivation for this study came from systematic errors observed in km-scale numerical weather prediction models. The extrapolation of the present results to such models is based on two critical assumptions. First, the highest-resolution simulation examined here must behave sufficiently similarly to the real atmosphere to serve as reference. A thorough evaluation of the ICON configuration used here (Heinze *et al.*, 2017) confirms that

the highest-resolution configuration is able to reproduce observed turbulence profiles well, and generate cloud-size distributions that are more realistic than for the coarser resolutions. Pscheidt *et al.* (2019) still find some deficits in precipitation and cloud coverage, but convective organization seems to be well captured. Second, the errors in the coarser resolution runs must be representative of errors found in the NWP models, if the results are to provide guidance for correcting those errors. There is one concern regarding the different approaches for turbulence parametrizations. While a three-dimensional turbulence closure (Heinze *et al.*, 2017) was employed in the hectometer-scale simulations analysed, most km-scale models are equipped with a one-dimensional planetary boundary-layer scheme (e.g., Kealy, 2019). However, both hectometer- and km-scale models suffer from similar deficiencies: the lag in timing of the diurnal cycle of convection and a relative lack of convective organization, especially late in the day (e.g., Baldauf *et al.*, 2011; Clark *et al.*, 2016; Rasp *et al.*, 2018). Hence we conclude that the biases obtained here in cold-pool-driven CI for hectometre simulations are valid and likely even more pronounced in km-scale models, and that they are partially responsible for biases in precipitation forecasts. However, given the large day-to-day variability seen here, future studies using much longer time periods are desirable.

One feature of this study is the relatively uncommon application of linear causal analysis in atmospheric sciences. The main advantage lies in the formal framework, which specifies how causal effects can be estimated from non-interventional data without the need for complex methods. As we have demonstrated, such methods are also helpful for simulation datasets where focused interventions are often not feasible. The increasing availability of computationally expensive high-resolution model output (Heinze *et al.*, 2017; Stevens *et al.*, 2019; Stevens *et al.*, 2020) and very large ensembles of simulations (e.g., Necker *et al.*, 2020) will provide new opportunities to explore the physical processes controlling cumulus convection and many other weather phenomena across a wide range of scales. Causal methods such as applied here have great potential to give insight into these complex datasets.

ACKNOWLEDGEMENTS

The research leading to these results has been done within the subproject A6 of the Transregional Collaborative Research Center SFB/TRR 165 “Waves to Weather” (<http://www.wavestoweather.de>; accessed 1 April 2020) funded by the German Research Foundation (DFG). We further acknowledge funding through the German Federal Ministry of Education and Research for the project “High definition clouds and precipitation for advancing

climate prediction HD(CP)²” within the framework program “Research for Sustainable Development (FONA)”, under 01LK1507A (RH) and 01LK1507D (SS). The simulations were performed at the German Climate Computing Center (DKRZ) and assistance is acknowledged.

ORCID

Mirjam Hirt  <https://orcid.org/0000-0002-5650-7594>

REFERENCES

- Bachmann, K., Keil, C., Craig, G.C., Weissmann, M. and Welzbacher, C.A. (2020) Predictability of deep convection in idealized and operational forecasts: Effects of radar data assimilation, orography and synoptic weather regime. *Monthly Weather Review*, 148, 63–81. <https://doi.org/10.1175/MWR-D-19-0045.1>
- Baldauf, M., Seifert, A., Förstner, J., Majewski, D., Raschendorfer, M. and Reinhardt, T. (2011) Operational convective-scale numerical weather prediction with the COSMO model: description and sensitivities. *Monthly Weather Review*, 139, 3887–3905.
- Baur, F., Keil, C. and Craig, G.C. (2018) Soil moisture–precipitation coupling over central Europe: Interactions between surface anomalies at different scales and the dynamical implication. *Quarterly Journal of the Royal Meteorological Society*, 144, 2863–2875.
- Böing, S.J., Jonker, H.J., Siebesma, A.P. and Grabowski, W.W. (2012) Influence of the subcloud layer on the development of a deep convective ensemble. *Journal of the Atmospheric Sciences*, 69, 2682–2698.
- Cafaro, C. and Rooney, G.G. (2018) Characteristics of colliding density currents: a numerical and theoretical study. *Quarterly Journal of the Royal Meteorological Society*, 144, 1761–1771.
- Chen, B. and Pearl, J. (2014). Graphical Tools for Linear Structural Equation Modeling. Technical Report R-432, Department of Computer Science, University of California, Los Angeles, CA.
- Clark, P.A., Roberts, N., Lean, H.W., Ballard, S.P. and Charlton-Perez, C. (2016) Convection-permitting models: a step-change in rainfall forecasting. *Meteorological Applications*, 23, 165–181. <https://doi.org/10.1002/met.1538>
- Dipankar, A., Stevens, B., Heinze, R., Moseley, C., Zängl, G., Giorgetta, M. and Brdar, S. (2015) Large-eddy simulation using the general circulation model ICON. *Journal of Advances in Modeling Earth Systems*, 7, 963–986. <https://doi.org/10.1002/2015MS000431>
- Doswell, C.A. (1987) The distinction between large-scale and mesoscale contribution to severe convection: A case study example. *Weather and Forecasting*, 2, 3–16. [https://doi.org/10.1175/1520-0434\(1987\)002<0003:TDBLSA>2.0.CO;2](https://doi.org/10.1175/1520-0434(1987)002<0003:TDBLSA>2.0.CO;2)
- Drager, A.J. and van den Heever, S.C. (2017) Characterizing convective cold pools. *Journal of Advances in Modeling Earth Systems*, 9, 1091–1115. <https://doi.org/10.1002/2016MS000788>
- Droegemeier, K.K. and Wilhelmson, R.B. (1985) Three-dimensional numerical modeling of convection produced by interacting thunderstorm outflows. Part I: control simulation and low-level moisture variations. *Journal of the Atmospheric Sciences*, 42, 2381–2403. [https://doi.org/10.1175/1520-0469\(1985\)042<2381:TDNMOC>2.0.CO;2](https://doi.org/10.1175/1520-0469(1985)042<2381:TDNMOC>2.0.CO;2)

- Ebert-Uphoff, I. and Deng, Y. (2012) Causal discovery for climate research using graphical models. *Journal of Climate*, 25, 5648–5665.
- Emanuel, K.A. (1994) *Atmospheric Convection*. Oxford University Press, Oxford, UK.
- Feng, Z., Hagos, S., Rowe, A.K., Burleyson, C.D., Martini, M.N. and de Szoeke, S.P. (2015) Mechanisms of convective cloud organization by cold pools over tropical warm ocean during the AMIE/DYNAMO field campaign. *Journal of Advances in Modeling Earth Systems*, 7, 357–381. <https://doi.org/10.1002/2014MS000384>
- Fournier, M.B. and Haerter, J.O. (2019) Tracking the gust fronts of convective cold pools. *Journal of Geophysical Research: Atmospheres*, 124, 11103–11117. <https://doi.org/10.1029/2019JD030980>
- Gentine, P., Garelli, A., Park, S.-B., Nie, J., Torri, G. and Kuang, Z. (2016) Role of surface heat fluxes underneath cold pools. *Geophysical Research Letters*, 43, 874–883.
- Gentine, P., Pritchard, M., Rasp, S., Reinaudi, G. and Yacalis, G. (2018) Could machine learning break the convection parameterization deadlock?. *Geophysical Research Letters*, 45, 5742–5751. <https://doi.org/10.1029/2018GL078202>
- Grant, L.D. and van den Heever, S.C. (2016) Cold pool dissipation. *Journal of Geophysical Research: Atmospheres*, 121, 1138–1155.
- Grant, L.D. and van den Heever, S.C. (2018) Cold pool–land surface interactions in a dry continental environment. *Journal of Advances in Modeling Earth Systems*, 10, 1513–1526. <https://doi.org/10.1029/2018MS001323>
- Haerter, J.O., Boeing, S.J., Henneberg, O. and Nissen, S.B. (2018) Reconciling cold pool dynamics with convective self-organization. arXiv: 1810.05518
- Hanley, K.E., Plant, R.S., Stein, T.H., Hogan, R.J., Nicol, J.C., Lean, H.W., Halliwell, C. and Clark, P.A. (2015) Mixing-length controls on high-resolution simulations of convective storms. *Quarterly Journal of the Royal Meteorological Society*, 141, 272–284.
- Heinze, R., Dipankar, A., Henken, C.C., Moseley, C., Sourdeval, O., Trömel, S., Xie, X., Adamidis, P., Ament, F., Baars, H., Barthlott, C., Behrendt, A., Blahak, U., Bley, S., Brdar, S., Brueck, M., Crewell, S., Deneke, H., Di Girolamo, P., Evaristo, R., Fischer, J., Frank, C., Friederichs, P., Göcke, T., Gorges, K., Hande, L., Hanke, M., Hansen, A., Hege, H.-C., Hoose, C., Jahns, T., Kalthoff, N., Klocke, D., Kneifel, S., Knippertz, P., Kuhn, A., van Laar, T., Macke, A., Maurer, V., Mayer, B., Meyer, C.I., Muppa, S.K., Neggers, R.A.J., Orlandi, E., Pantillon, F., Pospichal, B., Röber, N., Scheck, L., Seifert, A., Seifert, P., Senf, F., Siligam, P., Simmer, C., Steinke, S., Stevens, B., Wapler, K., Weniger, M., Wulfmeyer, V., Zängl, G., Zhang, D. and Quaas, J. (2017) Large-eddy simulations over Germany using ICON: a comprehensive evaluation. *Quarterly Journal of the Royal Meteorological Society*, 143, 69–100. <https://doi.org/10.1002/qj.2947>
- Henneberg, O., Meyer, B. and Haerter, J.O. (2020) Particle-based tracking of cold pool gust fronts. *Journal of Advances in Modeling Earth Systems*. <https://doi.org/10.1029/2019MS001910>
- Hirt, M., Rasp, S., Blahak, U. and Craig, G. (2019) Stochastic parameterization of processes leading to convection initiation in kilometer-scale models. *Monthly Weather Review*, 147, 3917–3934.
- Hirt, M. (2020) HirtM/cold pool driven convection initiation: release repository for publication. <https://doi.org/10.5281/zenodo.3611048>
- Honnert, R. (2016) Representation of the grey zone of turbulence in the atmospheric boundary layer. *Advances in Science and Research*, 13, 63–67. <https://doi.org/10.5194/asr-13-63-2016>
- Hoyer, S. and Hamman, J. (2017) xarray: N-D labeled arrays and datasets in Python. *Journal of Open Research Software*, 5. <https://doi.org/10.5334/jors.148>
- Hunter, J.D. (2007) Matplotlib: A 2d graphics environment. *Computing in Science & Engineering*, 9, 90–95. <https://doi.org/10.1109/MCSE.2007.55>
- Jeevanjee, N. (2017) Vertical velocity in the gray zone. *Journal of Advances in Modeling Earth Systems*, 9, 2304–2316. <https://doi.org/10.1002/2017MS001059>
- Johns, R.H. and Doswell, C.A. (1992) Severe local storms forecasting. *Weather and Forecasting*, 7, 588–612. [https://doi.org/10.1175/1520-0434\(1992\)007<0588:SLSF>2.0.CO;2](https://doi.org/10.1175/1520-0434(1992)007<0588:SLSF>2.0.CO;2)
- Kealy, J.C. (2019) Probing the ‘grey zone’ of NWP – is higher resolution always better?. *Weather*, 74, 246–249. <https://doi.org/10.1002/wea.3506>
- Keil, C., Heinlein, F. and Craig, G.C. (2014) The convective adjustment time-scale as indicator of predictability of convective precipitation. *Quarterly Journal of the Royal Meteorological Society*, 140, 480–490. <https://doi.org/10.1002/qj.2143>
- Keil, C., Baur, F., Bachmann, K., Rasp, S., Schneider, L. and Barthlott, C. (2019) Relative contribution of soil moisture, boundary-layer and microphysical perturbations on convective predictability in different weather regimes. *Quarterly Journal of the Royal Meteorological Society*, 145, 3102–3115. <https://doi.org/10.1002/qj.3607>
- Kober, K. and Craig, G.C. (2016) Physically based stochastic perturbations (PSP) in the boundary layer to represent uncertainty in convective initiation. *Journal of the Atmospheric Sciences*, 73, 2893–2911.
- Kretschmer, M., Coumou, D., Donges, J.F. and Runge, J. (2016) Using causal effect networks to analyze different arctic drivers of midlatitude winter circulation. *Journal of Climate*, 29, 4069–4081.
- Leutwyler, D., Lüthi, D., Ban, N., Fuhrer, O. and Schär, C. (2017) Evaluation of the convection-resolving climate modeling approach on continental scales. *Journal of Geophysical Research: Atmospheres*, 122, 5237–5258. <https://doi.org/10.1002/2016JD026013>
- Markowski, P. and Richardson, Y. (2011) *Mesoscale Meteorology in Midlatitudes*. John Wiley & Sons, Chichester, UK.
- Mirjam, H., George, C.C., Sophia, S., Julien, S. and Rieke, H. (2020) Cold pool driven convective initiation: using causal graph analysis to determine what convection permitting models are missing. *Open Data LMU*. <https://doi.org/10.5282/ubm/data.178>
- Maurer, V., Bischoff-Gauß, I., Kalthoff, N., Gantner, L., Roca, R. and Panitz, H.-J. (2017) Initiation of deep convection in the Sahel in a convection-permitting climate simulation for Northern Africa. *Quarterly Journal of the Royal Meteorological Society*, 143, 806–816. <https://doi.org/10.1002/qj.2966>
- McKinney, W. (2010). Data structures for statistical computing in Python, pp. 51–56 in Proceedings of the 9th Python in Science Conference. Austin TX. van der Walt, S., Millman, J. (eds).
- Necker, T., Geiss, S., Weissmann, M., Ruiz, J., Miyoshi, T. and Lien, G.-Y. (2020) A convective-scale 1000-member ensemble simulation and potential applications. *Quarterly Journal of the Royal Meteorological Society* 146(728), 1423–1442. <https://doi.org/10.1002/qj.3744>

- Panosetti, D., Schlemmer, L. and Schär, C. (2019) Bulk and structural convergence at convection-resolving scales in real-case simulations of summertime moist convection over land. *Quarterly Journal of the Royal Meteorological Society*, 145, 1427–1443. <https://doi.org/10.1002/qj.3502>
- Pearl, J. (2009) *Causality*. Cambridge University Press, Cambridge, UK.
- Pearl, J. (2013) Linear models: A useful “microscope” for causal analysis. *Journal of Causal Inference*, 1, 155–170.
- Pearl, J. (2014) Interpretation and identification of causal mediation. *Psychological Methods*, 19, 459
- Peters, K. and Hohenegger, C. (2017) On the dependence of squall-line characteristics on surface conditions. *Journal of the Atmospheric Sciences*, 74, 2211–2228.
- Piper, D., Kunz, M., Ehmele, F., Mohr, S., Mühr, B., Kron, A. and Daniell, J. (2016) Exceptional sequence of severe thunderstorms and related flash floods in May and June 2016 in Germany—part 1: Meteorological background. *Natural Hazards and Earth System Sciences*, 16, 2835–2850.
- Pscheidt, I., Senf, F., Heinze, R., Deneke, H., Trömel, S. and Hohenegger, C. (2019) How organized is deep convection over Germany?. *Quarterly Journal of the Royal Meteorological Society*, 145, 2366–2384. <https://doi.org/10.1002/qj.3552>
- Purdum, J.F.W. (1976) Some uses of high-resolution goes imagery in the mesoscale forecasting of convection and its behavior. *Monthly Weather Review*, 104, 1474–1483. [https://doi.org/10.1175/1520-0493\(1976\)104<1474:SUOHRG>2.0.CO;2](https://doi.org/10.1175/1520-0493(1976)104<1474:SUOHRG>2.0.CO;2)
- Rasp, S., Selz, T. and Craig, G.C. (2018) Variability and clustering of midlatitude summertime convection: Testing the Craig and Cohen theory in a convection-permitting ensemble with stochastic boundary-layer perturbations. *Journal of the Atmospheric Sciences*, 75, 691–706. <https://doi.org/10.1175/JAS-D-17-0258.1>
- Rotunno, R., Klemp, J.B. and Weisman, M.L. (1988) A theory for strong, long-lived squall lines. *Journal of the Atmospheric Sciences*, 45, 463–485.
- Runge, J., Bathiany, S., Bollt, E., Camps-Valls, G., Coumou, D., Deyle, E., Glymour, C., Kretschmer, M., Mahecha, M.D., Muñoz-Mari, van Nes, E.H., Peters, J., Quax, R., Reichstein, M., Scheffer, M., Schölkopf, B., Spirtes, P., Sugihara, G., Sun, J., Zhang, K. and Zscheischler, J. (2019) Inferring causation from time series in earth system sciences. *Nature Communications*, 10(2553). <https://doi.org/10.1038/s41467-019-10105-3>
- Runge, J., Petoukhov, V., Donges, J.F., Hlinka, J., Jajcay, N., Vejmelka, M., Hartman, D., Marwan, N., Paluš, M. and Kurths, J. (2015) Identifying causal gateways and mediators in complex spatio-temporal systems. *Nature Communications*, 6(8502). <https://doi.org/10.1038/ncomms9502>
- Samarasinghe, S.M., McGraw, M.C., Barnes, E.A. and Ebert-Uphoff, I. (2019) A study of links between the Arctic and the midlatitude jet stream using Granger and Pearl causality. *Environmetrics*, 30(4). <https://doi.org/10.1002/env.2540>
- Schlemmer, L. and Hohenegger, C. (2014) The formation of wider and deeper clouds as a result of cold-pool dynamics. *Journal of the Atmospheric Sciences*, 71, 2842–2858.
- Schraff, C., Reich, H., Rhodin, A., Schomburg, A., Stephan, K., Periañez, A. and Potthast, R. (2016) Kilometre-scale ensemble data assimilation for the COSMO model (KENDA). *Quarterly Journal of the Royal Meteorological Society*, 142, 1453–1472.
- Senf, F., Klocke, D. and Brueck, M. (2018) Size-resolved evaluation of simulated deep tropical convection. *Monthly Weather Review*, 146, 2161–2182.
- Squitiere, B.J. and Gallus, W.A. (2020) On the forecast sensitivity of MCS cold pools and related features to horizontal grid spacing in convection-allowing WRF simulations. *Weather and Forecasting*, 35, 325–346. <https://doi.org/10.1175/WAF-D-19-0016.1>
- Stevens, B., Satoh, M., Auger, L., Biercamp, J., Bretherton, C., Düben, P., Judt, F., Khairoutdinov, M., Klocke, D., Kornbluh, L., Kodama, C., Neumann, P., Lin, S., Putman, W.M., Röber, N., Shibuya, R., Vidale, P. and Wedi, N. (2019) DYAMOND: the Dynamics of the Atmospheric general circulation Modeled On Non-hydrostatic Domains. *Progress In Earth And Planetary Science*, 6(61). <https://doi.org/10.1186/s40645-019-0304-z>
- Stevens, B., Acquistapace, C., Hansen, A., Heinze, R., Klinger, C., Klocke, D., Schubotz, W., Windmiller, J., Adamidis, P., Arka, I., Barlakas, V., Biercamp, J., Brueck, M., Brune, S., Buehler, S., Burkhardt, U., Cioni, G., Costa-Suros, M., Crewell, S., Crüger, S., Deneke, H., Friedrichs, P., Henken, C.C., Hohenegger, C., Jacob, M., Jakub, F., Kalthoff, N., Köhler, M., Van Laar, T.W., Li, P., Löhnert, U., Macke, A., Madenbach, N., Mayer, B., Nam, C., Naumann, A.K., Peters, K., Poll, S., Quaas, J., Röber, N., Rochetin, N., Scheck, L., Schemann, V., Schnitt, S., Seifert, A., Senf, F., Shapkalijevski, M., Simmer, C., Singh, S., Sourdeval, O., Spickermann, D., Strandgren, J., Tessiot, O., Vercauteren, N., Vial, J., Voigt, A. and Zängl, G. (2020) The added value of large-eddy and storm-resolving models for simulating clouds and precipitation. *Journal of the Meteorological Society of Japan, Series II*. <https://doi.org/10.2151/jmsj.2020-021>
- Tompkins, A.M. (2001) Organization of tropical convection in low vertical wind shears: the role of cold pools. *Journal of the Atmospheric Sciences*, 58, 1650–1672.
- Torri, G. and Kuang, Z. (2019) On cold pool collisions in tropical boundary layers. *Geophysical Research Letters*, 46, 399–407.
- Torri, G., Kuang, Z. and Tian, Y. (2015) Mechanisms for convection triggering by cold pools. *Geophysical Research Letters*, 42, 1943–1950. <https://doi.org/10.1002/2015GL063227>
- van der Walt, S., Colbert, S.C. and Varoquaux, G. (2011) The numpy array: A structure for efficient numerical computation. *Computing in Science Engineering*, 13, 22–30. <https://doi.org/10.1109/MCSE.2011.37>
- Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S.J., Brett, M., Wilson, J., Jarrod Millman, K., Mayorov, N., Nelson, A.R.J., Jones, E., Kern, R., Larson, E., Carey, C., Polat, İ., Feng, Y., Moore, E.W., Van der Plas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E.A., Harris, C.R., Archibald, A.M., Ribeiro, A.H., Pedregosa, F., van Mulbregt, P. and Contributors, S. (2019) SciPy 1.0—fundamental algorithms for scientific computing in Python. arXiv: 1907.10121
- Waskom, M., Botvinnik, O., O’Kane, D., Hobson, P., Ostblom, J., Lukauskas, S., Gemperline, D.C., Augspurger, T., Halchenko, Y., Cole, J.B., Warmenhoven, J., de Ruiter, J., Pye, C., Hoyer, S., Vanderplas, J., Villalba, S., Kunter, G., Quintero, E., Bachant, P., Martin, M., Meyer, K., Miles, A., Ram, Y., Brunner, T., Yarkoni, T., Williams, M.L., Evans, C., Fitzgerald, C.B. and Qalieh, A. (2018) mwaskom/seaborn: v0.9.0 (July 2018). <https://doi.org/10.5281/zenodo.1313201>

- Weisman, M.L. and Rotunno, R. (2004) "A theory for strong long-lived squall lines" revisited. *Journal of the Atmospheric Sciences*, 61, 361–382.
- Wilks, D.S. (2011) *Statistical Methods in the Atmospheric Sciences*. Academic Press, Cambridge, MA.
- Wright, S. (1921) Correlation and causation. *Journal of Agricultural Research*, 20, 557–585.
- Yano, J.-I., Ziemiański, M.Z., Cullen, M., Termonia, P., Onvlee, J., Bengtsson, L., Carrassi, A., Davy, R., Deluca, A., Gray, S.L., Homar, V., K'ohler, M., Krichak, S., Michaelides, S., Phillips, V.T.J., Soares, P.M.M. and Wyszogrodzki, A.A. (2018) Scientific challenges of convective-scale numerical weather prediction. *Bulletin of the American Meteorological Society*, 99, 699–710.
- Zängl, G., Reinert, D., Rípodas, P. and 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, 563–579. <https://doi.org/10.1002/qj.2378>

How to cite this article: Hirt M, Craig GC, Schäfer SAK, Savre J, Heinze R. Cold-pool-driven convective initiation: using causal graph analysis to determine what convection-permitting models are missing. *Q J R Meteorol Soc.* 2020;146:2205–2227. <https://doi.org/10.1002/qj.3788>

APPENDICES

A Computational set-up and data processing

The simulated data have been analysed using mostly open source tools within the Python environment. The packages used include numpy (van der Walt *et al.*, 2011), scipy (Virtanen *et al.*, 2019), xarray (Hoyer and Hamman, 2017), pandas (McKinney, 2010), seaborn (Waskom *et al.*, 2018), matplotlib (Hunter, 2007) and enstools⁵. The final data analysis has been conducted within two Jupyter notebooks which are available on github (Hirt, 2020). The data relevant for these notebooks are published as (Mirjam *et al.*, 2020).

B Single-door criterion for the causal model

We provide the outcome of applying the single-door criterion (Section 2) to our causal structure (Figure 11). Each set of adjustment variables that are necessary to estimate the direct causal effects by means of linear regression is given in Table 2. The D in brackets for the RI path marks a special situation: It is not necessary to include the

diurnal/synoptic variables for the RI path, because we have postulated that model resolution does not impact the large scale situation. However, including it should not affect the regression estimate either, unless this hypothesis is violated. In the following section, we will also evaluate this hypothesis by comparing the regression estimates of RI when the diurnal/synoptic conditions (D) are included or excluded.

C Sensitivity and robustness of the causal analysis

In this section we test the robustness of our causal analysis to some of the more important assumptions used in constructing the graph.

C.1 Evaluation of linear regression

By evaluating some standard properties of the applied linear regressions, some weak conclusions for our causal model can be drawn. We focus here on the p -values, which contain similar information as the bootstrapping, and the coefficients of determination R^2 (Wilks, 2011).

The p -values, displayed in Figure 1a,b, are mostly small (below 0.05), which suggest that the estimated correlation coefficients differ significantly from zero. This implies that the corresponding variables, which we have identified, are indeed correlated. It follows that the variables are not d-separated under the adjustment sets given in Table 2. Consequently, there must be some causal connection between the selected variables. However, this causal connection does not have to correspond to our causal structure (Pearl, 2009). Note that the RP and BP path (Figure 1a) and a few path coefficients for single days (Figure 1b), do not yield small p -values and no causal connection can be inferred.

The R^2 values are given in Figure 1c,d and characterize the relative amount of variability that can be explained by the predictors. For gust front mass flux and triggering probability as predictands, the R^2 values are approximately 0.4 (Figure 1c), which implies that a substantial part of their variability is included in our model. If separate days are considered, the highest R^2 values are ~ 0.8 (Figure 1d). Such high values confirm that a substantial part of the information necessary to determine the predictor has been included. It is consequently less likely that we have missed influential common drivers. Unfortunately, not all cases show such high R^2 values. This can be problematic, if the missing informative variables also influence other variables of our causal model (common driver).

C.2 Stratification by model resolution

We can further evaluate the causal effects by using only two of three model resolutions. Corresponding causal effects are displayed in Figure 2. We find that no substantial differences exist whether 156–312 m, 312–625 m or

⁵Enstools is a python package for processing of atmospheric, ensemble simulations currently under development within the Waves to Weather project (www.wavestoweather.de). It can be made available on request.

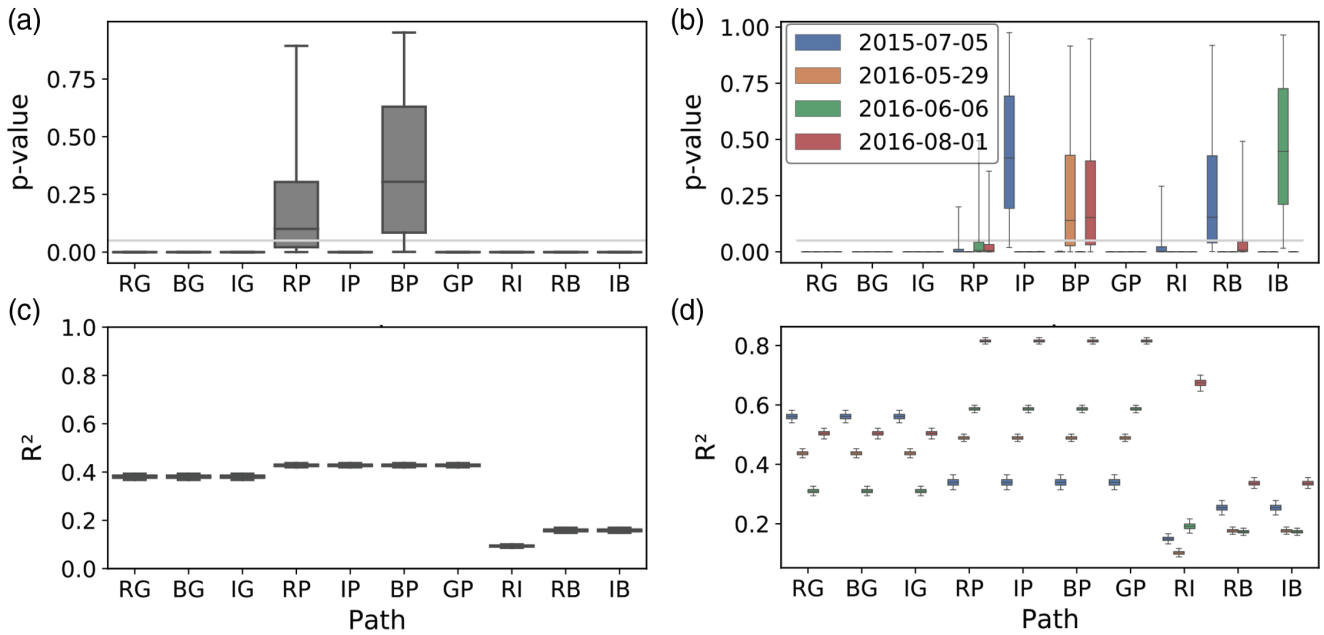


FIGURE A1 The (a, b) p and (c, d) R^2 values of the estimated path coefficients. (a, c) are for all cold pools and (b, d) when stratified by the four different days. The whiskers of the box plots correspond to the 95% confidence intervals from bootstrapping

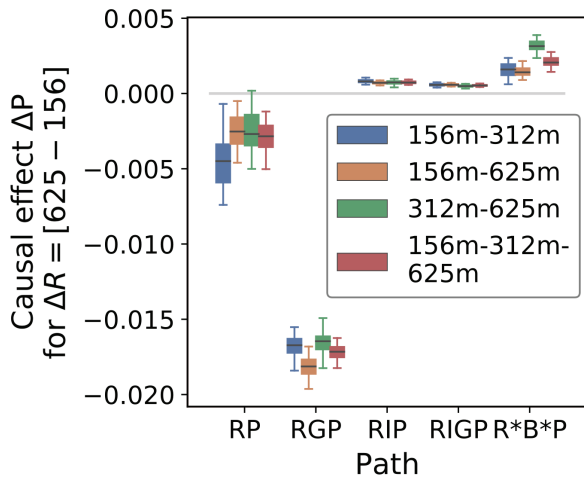


FIGURE A2 Aggregated path coefficients computed using only two of the three model resolutions. The whiskers of the box plots correspond to the 95% confidence intervals from bootstrapping

156–625 m are compared. Hence, within the given range of model resolution, the sensitivity to model resolution is indeed sufficiently linear.

C.3 Stratification by days

As described in Sections 3 and 4, the causal effects can also be evaluated for each day. The high variability in the RP path implies that the large-scale conditions influence the RP effect, violating our linearity assumption. Since this is an important sensitivity, it is discussed extensively in the main text.

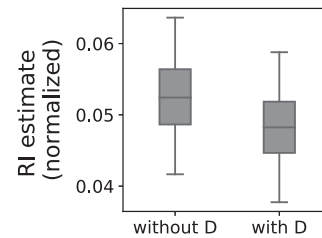


FIGURE A3 Normalized path estimates for the impact of model resolution on cold-pool intensity (RI path) by including or excluding the diurnal/synoptic variables (D) as adjustment set for the linear regression. The whiskers of the box plots correspond to the 95% confidence intervals from bootstrapping

C.4 Sensitivity of chosen synoptic variables

We have specified six synoptic variables which we use to describe the synoptic and diurnal situation. We have also considered several other variables (e.g., specific humidity, virtual potential temperature) and used each variable separately. However, these modifications in the synoptic variables did not lead to any substantial deviations in the estimated direct and indirect path coefficients from model resolution to triggering probability (not shown).

C.5 Different variables

The choice of variables used to specify the I- and G-node is not unambiguous. Consequently we have evaluated several reasonable possibilities. We have determined the I-node by the area-averaged intensity, area-integrated

intensity or the effective radius of the cold pools. To measure the gust-front strength, we have considered 95th percentile vertical velocity and area-averaged or -integrated upward mass flux within the cold-pool boundaries. However the different combinations of variables did not substantially impact the results of the causal path estimates (not shown).

C.6 Impact of resolution on synoptic conditions

Properly evaluating the effect of model resolution on the diurnal/synoptic conditions (RD) is not trivial, because links between the different large-scale variables are to be expected (this complicates the identification).

An indication that this RD effect is small is provided by the small, mostly unsystematic, differences between resolutions of each large-scale variable (Figure 2). Additionally, we can include or exclude the large-scale variables when estimating the RI path. If there is indeed no link between model resolution and diurnal/synoptic conditions, then both estimates will be equal. The corresponding two estimates for the RI path are displayed in Figure 3. The strongly overlapping distributions of the bootstrapping sample suggest that the two estimates are not significantly different. Thus, it is reasonable to neglect a causal connection between model resolution and synoptic conditions.