



Global estimates of damaging hail hazard

Andreas F. Prein*, Greg J. Holland

National Center for Atmospheric Research (NCAR), 3090 Center Green Drive, Boulder, CO, 80301, USA



ARTICLE INFO

Keywords:

Hail hazard
Algorithm
Global
Severe convection

ABSTRACT

A single hailstorm can cause losses in the billion-dollar range if it occurs over a densely populated area. Property losses from hailstorms are rising with time mainly due to a combination of increases in population density and wealth. Report based observational hail data alone are highly inhomogeneous and unable to discriminate between climate and societal changes. Here we present a statistical approach that estimates hail hazard from large-scale environmental conditions. Using daily ERA-Interim reanalysis data and large hail observations (diameter larger than 2.5 cm) from the conterminous United States (CONUS) we show that four predictors enable skillful discrimination of large hail frequencies on a regional scale.

The predictors include atmospheric instability, freezing level height, and 0–3 km wind shear and storm relative helicity. These variables are used to develop a hail algorithm, which provides the probabilities for large hail occurrence from regional to global scales and from daily to climate timescales. The algorithm skill is tested over the CONUS and with independent hail observations from Australia and Europe. It skillfully captures the frequency, annual cycle, spatial patterns, and interannual variability of observed large hail records in a large variety of climate regions. Deficiencies are found in regions with strong orographic forcing and low shear environments. The algorithm outperforms established severe convection indices in terms of more accurately predicting absolute hail frequencies and the annual cycles of large hail in all tested regions. The code is open-source and is applicable to a variety of tasks including daily to seasonal forecasting and assessing climate change influences on hail hazard.

1. Introduction

Globally the losses due to convective extremes such as large hail (diameter larger than 2.5 cm), tornadoes, wind gusts, and flash floods are increasing (Munich, 2016; Changnon, 2009). In the US average annual loss from severe convective storms are \$11.23 billion (in 2016 USD) compared to \$11.28 billion from hurricanes (Gunturi and Tippett, 2017). The increasing losses mainly result from an increase in population density and wealth and changes while the contribution from changes in the intensity and frequency of convective storms are more uncertain (Changnon, 2009). This uncertainty is due to short and inconsistent observational records (Allen and Tippett, 2015), the high cost and limited capacity of climate model simulations able to resolve severe hailstorms, and missing physical understanding of interactions between the climate system and severe convection (Tippett et al., 2015).

Three approaches have been used to assess the impact of climate variability and climate change on large hail occurrence.

- 1) Hail observations with hail pads or from damage reports. Observing hail is difficult because hail is highly localized. Station based hail observations with hail pads can provide homogeneous time records but there are only a few countries that have long records and data access is often limited (e.g., Xie et al., 2008). Observations from either hail reports (e.g., Changnon and Changnon, 2000; Xie et al., 2008) or economic loss reports (Changnon et al., 2000; Changnon, 2009; Barthel and Neumayer, 2012) are an alternative to hail pad observations but effects of socioeconomic development, observational practices, and climatic changes have to be decomposed. Report based hail datasets combine information about hail occurrence from a variate of sources including the general public (Hand and Cappelluti, 2011; Allen and Tippett, 2015). These datasets have higher spatial resolution than station-based records but are highly inhomogeneous with hail frequencies varying with populated density and reporting practice (Allen and Tippett, 2015).
- 2) The dynamical simulation of hail with numerical models (Mahoney et al., 2012; Brimelow et al., 2017; Adams-Selin and Ziegler, 2016) provides a physical understanding of the mechanisms that are

* Corresponding author.

E-mail address: prein@ucar.edu (A.F. Prein).

URL: <http://preina03.wixsite.com/andreas-prein> (A.F. Prein).

- associated with changing hail frequencies but comes at high computational costs and large uncertainties in the parameterization of the involved processes.
- 3) Empirical relationships between environmental properties and observed damaging hail are used to relate large-scale environmental conditions (Brooks, 2009; Mohr and Kunz, 2013; Allen et al., 2015a; Westermayer et al., 2017; Mohr et al., 2015), remote sensing data (Witt et al., 1998; Féral et al., 2003; Cecil, 2009; Mroz et al., 2017; Ni et al., 2017), or a combination of both (Punge et al., 2017; Bedka et al., 2018) to hail occurrence. This approach is flexible and computationally efficient but assumes that the environmental conditions for large hail development are not changing with climate warming and that all favorable environments are equally likely to produce large hail. Recently, machine learning algorithms were developed that show promising results to improve traditional empirical forecasting methods (Gagne et al., 2017).

Our approach belongs to the third category and uses large-scale environmental ingredients to estimate large hail probabilities. Large hail formation is often related to supercell thunderstorms (Moller et al., 1994), strong mesoscale convective systems (Houze, 2004), and pulse storms (Miller and Mote, 2017). Most of the giant hail occurrences –diameter larger than 5 cm– are related to supercells, which provide the ideal environmental conditions such as strong and persistent updrafts that exist for 30 min or more (Blair et al., 2017). These storms typically develop under strong convective instability, sufficient moisture at the low levels, strong wind shear that usually varies with height, and a triggering mechanism that can release the instability (Newton, 1963). Many of these basic characteristics are included in current hail parameters that are used for severe weather forecasting such as the Significant Hail Parameter (SHIP; http://www.spc.noaa.gov/exper/mesoanalysis/help/help_sigh.html), the Large Hail Parameter (LHP; http://www.spc.noaa.gov/exper/mesoanalysis/help/help_lghl.html), the Significant Severe Parameter (Craven et al., 2004), or the Severe Thunderstorm Environment Index (Brooks et al., 2003).

The aim of this study is to develop a synthetic hail algorithm that derives hail occurrence probabilities from large-scale environmental conditions from regional-to global-scale and from daily to climate timescales. This is an advantage over existing hail indices due to the global scope and higher temporal resolution (cf. Allen et al., 2015a), due to only including predictors that are observational- and not model-based (cf. Large Hail Parameter; Allen et al., 2015a), due to the tailoring to large hail instead of general severe convection (Brooks et al., 2003; Craven et al., 2004), and using a smooth probabilistic instead of step-wise transition from unfavorable to favorable hail environments (Brooks et al., 2003).

We assume that environments that lead to large hail are consistent worldwide, which means that e.g., environments that lead to large hail in the US will also lead to large hail in e.g., Australia or Europe as shown in Brooks (2009) and Allen et al. (2011). The global capability is an advantage to previous environmental-based approaches, which typically focused on environmental conditions of smaller regions for hail hazard assessment (e.g., Mohr and Kunz, 2013; Allen and Karoly, 2014; Allen et al., 2015a). Space-borne hail observations are also able to provide a global coverage but often have limited time series length for climate change assessments (Cecil and Blankenship, 2012). Our algorithm avoids assumptions about the distribution of variables, does not prescribe specific relationships between the variables, and needs no weighting or normalization functions.

The hail observations that are used to build and evaluate our approach are presented in section 2. Section 3 describes the synthetic hail algorithm, together with a sensitivity analysis and the algorithm evaluation is shown in section 4. The paper closes with a global assessment of large hail hazards in section 5 and with summary and conclusions in section 6.

2. Hail observations

We consider hail observations from three sources: 1) the Storm Prediction Center's Storm Events dataset (Schaefer and Edwards, 1999), 2) the Australian Bureau of Meteorology's (BoM) Severe Storms Archive (<http://www.bom.gov.au/australia/stormarchive/>), and 3) the European Severe Storm Laboratory's (ESSL) European Severe Weather Database (Dotzek et al., 2009). All data sets cover a common period from 1979 to 2015 and provide the date and time, location, and the maximum diameter of observed hailstones. We only consider hail reports with hail diameters of 25 mm or larger (large hail hereafter) to account for sampling issues and changes in observational practices (Allen and Tippett, 2015). The BoM and ESSL datasets have severe issues with temporal homogeneity. The ESSL data has most records over Central Europe and shows a steep increase in extremes after its unification in 2006 (Groenemeijer et al., 2017). The BoM data is limited due to the sparse population density and the non-existence of a formal collection mechanism for severe storm reports before 1987 (Allen and Allen, 2016).

Based on the hail observations, we create a regular $0.7^\circ \times 0.7^\circ$ daily gridded data set for each region on a grid that is identical to the ERA-Interim reanalysis grid (Dee et al., 2011). A grid cell value cannot exceed one hail event per day even if multiple events were reported. This is to avoid double counting of the same event and to reduce the effects of varying population density on the gridded data set.

The average annual large hail occurrences from the gridded datasets in Fig. 1 show some of the issues with hail data sets that are based on hail reports (Allen and Tippett, 2015). All data sets have higher large hail frequencies in big cities compared to the surrounding areas due to the higher detection probability of large hail in urbanized areas (Fig. 1a–c). Allen et al. (2015a) showed that large hail in the US also is more frequently observed along the road network. A further issue in the European data set is that different countries have different reporting practices leading to artificial spatial gradients such as seen between Germany and France in Fig. 1c.

Also, the time series of observed large hail are inhomogeneous due to changes in population densities and reporting practices. The time series in the conterminous United States (CONUS; Fig. 1d) shows a non-physical linear increase from 1979 to ~2005 (Allen et al., 2015b; Allen and Tippett, 2015), Australian records show a step increase in hail frequency in the late 80s due to the introduction of common reporting practices (Fig. 1e), and the European data show an exponential increase in observed large hail frequency starting in the mid 90s. Detecting climate change impacts from these time series is very difficult due to the large influence of the spatiotemporal inhomogeneities.

3. Example existing severe thunderstorm indices

We compare the here developed hail algorithm to three existing severe thunderstorm indices that can provide severe convective hazard estimates on a daily basis. The first one is the significant severe parameter, which is the product of mean layer CAPE and 0–6 km shear (Craven et al., 2004). According to Craven et al. (2004) values larger than $20\,000\text{ m}^3\text{ s}^{-3}$ indicate environments that produce significant hail/wind. The second index is the severe thunderstorm environment index, which is one if $\text{CAPE} \leq 100$, if the 2–4 km lapse rate $< 6.5\text{ K km}^{-1}$, and if the environmental conditions are right of the line $2.86 \cdot \log(0 - 6\text{ km shear}) + 1.79 \cdot \log(\text{CAPE}) = 8.36$ and zero otherwise (Brooks et al., 2003). The third index is the Significant Hail Parameter (SHIP), which is defined as

$$\begin{aligned} \text{SHIP} = & [(\text{MUCAPE J/kg}) * (\text{Mixing Ratio of MU parcel g/kg}) * (700 \\ & - 500 \text{ hPa lapse rate K/km}) * (-500 \text{ mb TEMP C}) * (0 \\ & - 6 \text{ km Shear m/s})] / \\ & 42,000,000. \end{aligned}$$

The 0–6 km shear must be within the range $7\text{--}27\text{ m s}^{-1}$, the mixing ratio is confined to $11\text{--}13.6\text{ g kg}^{-1}$, and the 500 hPa

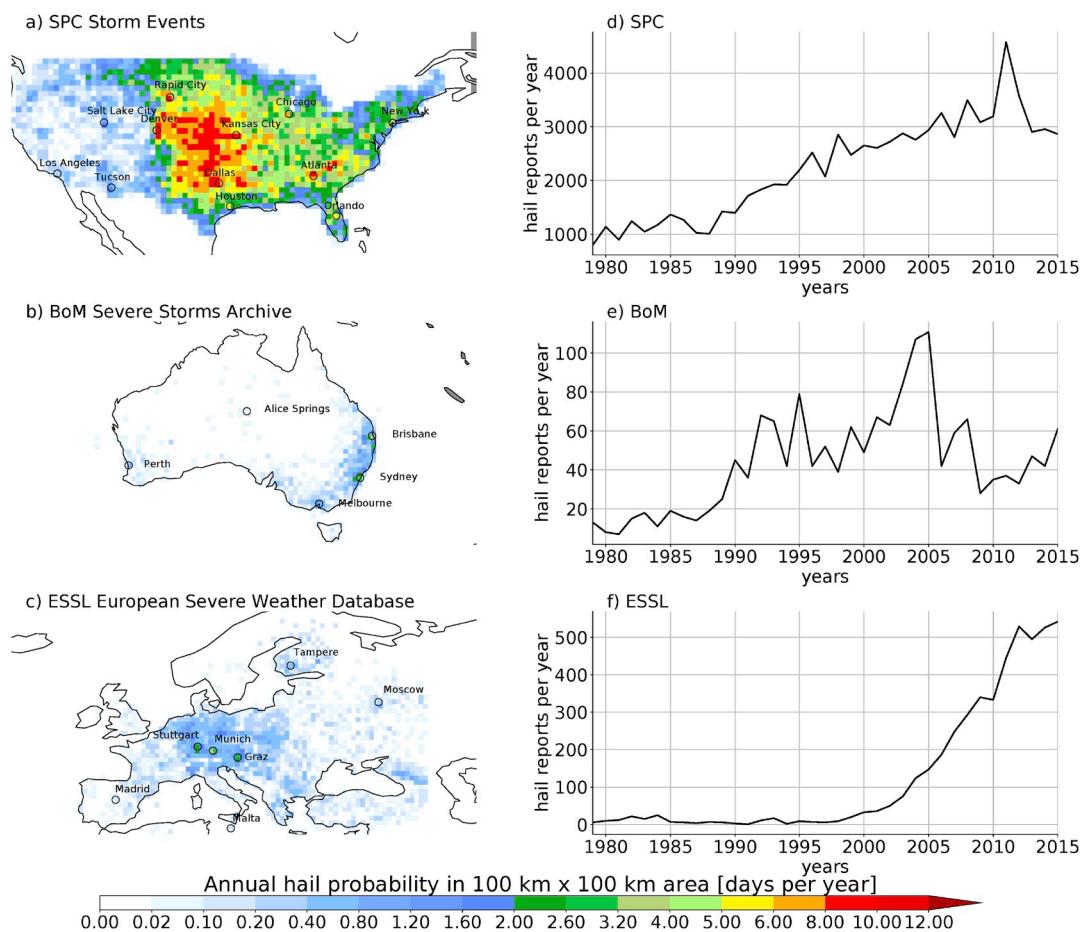


Fig. 1. Observed annual average large hail day frequency for the period 1979 to 2015 in a 100.100 km area for the CONUS (a), Australia (b), and Europe (c). The corresponding annual time series of accumulated large hail occurrences are shown in d–f.

temperature is set to -5.5°C for any warmer values. After calculating the initial version of SHIP its values are modified if one of the following three criteria are met: 1) If $\text{MUCAPE} < 1300 \text{ J kg}^{-1}$, $\text{SHIP} = \text{SHIP} * (\text{MUCAPE}/1300)$; 2) if $700\text{--}500 \text{ hPa}$ lapse rate $< 5.8 \text{ K km}^{-1}$, $\text{SHIP} = \text{SHIP} * (700 - 500 \text{ hPa} \text{ lapse rate}/5.8)$; and 3) if the freezing level is $< 2400 \text{ m AGL}$, $\text{SHIP} = \text{SHIP} * (\text{freezing level}/2400)$.

4. Hail algorithm

We decided to avoid the parametric fitting of predictors to the raw hail observations. Instead, we use the Storm Prediction Center's observations over the CONUS to identify environmental conditions that were present at the time of large hail occurrence. We use the Storm Prediction Center's data only for the hail algorithm development and evaluate the algorithm's performance with independent observations in Australia and Europe. The Storm Prediction Center dataset is best suited for algorithm development since it covers a large range of climate regions and includes an order of magnitude more recorded hail observations than the European and Australian dataset combined (see Fig. 1d–f).

The environmental predictors are derived from ERA-Interim reanalysis 6-hourly model level data within the period 1979–2015. ERA-Interim has a $0.7^{\circ} \times 0.7^{\circ}$ horizontal grid spacing ($\sim 78 \text{ km}$ at the equator), 60 vertical levels, and provides data in six-hourly intervals (Dee et al., 2011). Twenty-two predictors are tested for their ability in adding predictive skill to the algorithm (see Table 1). The predictors include measures of buoyancy, vertical wind shear, atmospheric moisture, and the freezing level height. Similar predictors are frequently used for hail detection such as in the SHIP or LHP parameter

Table 1
List of tested single predictors for the hail algorithm.

Variables	Acronym	Unit
0–1 km vector wind shear	VS0-1 ^L	m s^{-1}
0–3 km vector wind shear	VS0-3 ^L	m s^{-1}
0–6 km vector wind shear	VS0-6 ^L	m s^{-1}
0–12 km vector wind shear	VS0-12 ^L	m s^{-1}
maximum CAPE	CAPE ^L	J kg^{-1}
minimum CIN	CIN ^L	J kg^{-1}
lifting condensation level	LCL ^{LU}	m
700–500 hPa average lapse rate	LRml ^{LU}	K m^{-1}
0–1 km storm relative helicity	SRHO-1 ^L	$\text{m}^2 \text{s}^{-2}$
0–3 km storm relative helicity	SRHO-3 ^L	$\text{m}^2 \text{s}^{-2}$
mean PBL relative humidity	RHpb ^{LU}	%
700–500 hPa mean relative humidity	RHml ^U	%
mean mixing ratio in PBL	MRpb ^{LU}	g kg^{-1}
700–500 hPa mean mixing ratio	MRml ^{LU}	g kg^{-1}
freezing level height above surface	FLH ^U	m
mixing ratio at 1st model-level	MR1l ^{LU}	g kg^{-1}
mixing ratio at freezing level	MRfl ^{LU}	g kg^{-1}
mean THETA-E in PBL	THETA-Epb ^L	K
700–500 hPa mean THETA-E	THETA-Em ^L	K
dewpoint temperature at 1st model-level	DT1l ^U	K
mean temperature in PBL	DTpb ^{LU}	K
700–500 hPa mean dewpoint temperature	DTml ^U	K

and in studies by Allen et al. (2015a) and Brooks et al. (2003).

All variables are based on 6-hourly ERA-Interim model-level soundings. The exponents at the end of the acronyms indicate if predictors have an upper limit (^U), a lower limit (^L), or an upper and lower limit (^{LU}) for large hail development. Bold acronyms show the variables

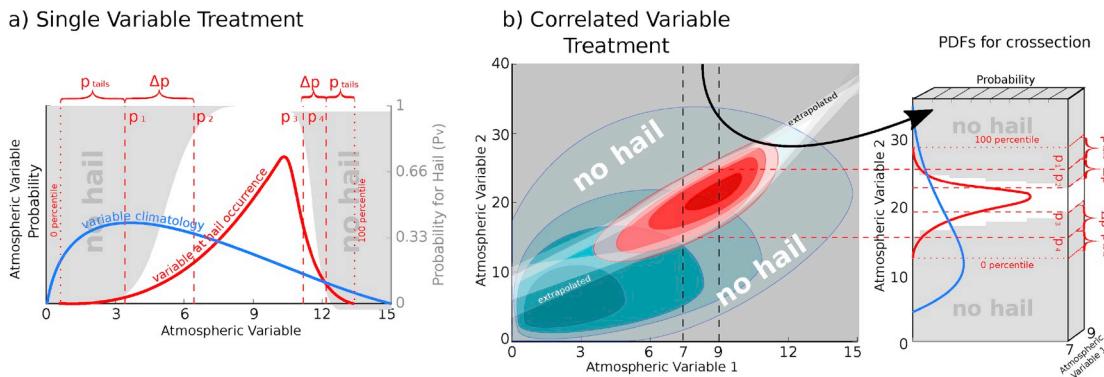


Fig. 2. Schematic of constraining large hail environmental conditions for single predictors (a) and for joint predictors (b). For single predictors, the climatological daily record of a predictor (blue line) and the PDF of the same predictor conditioned on days with large hail observations (red solid line) is shown. The location of the percentile values p_1 , p_2 , p_3 , and p_4 of the conditioned PDF, which are used in equation (1), is shown with red dashed lines, and the zero and 100 percentiles are shown with red dotted lines. The probabilities for large hail occurrence ($P_{(v)}$) is shown in gray shades (secondary y-axis). For joint variables (b) the blue/red contours show the 2D PDF of the climatological/conditioned environments. The transition between hail and non-hail environments is visualized by increasingly dark gray contours. The boundaries of these contours are derived by binning the x-axis predictor into 35 bins of equal sample size. Percentile thresholds are calculated for each bin similar to single predictors but a discrete step function is used for joint predictor transitions to increase the computational efficiency (inlay on the right and equation (2)). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

that are included in the hail algorithm.

To define environmental conditions that are favorable for large hail development we calculate conditioned predictors, which means that we condition predictors on days and locations where large hail was observed. This conditioning consists of three steps. 1) The ERA-Interim grid cell that includes the location of the hail observation is selected. 2) To account for spatial displacements of hail environments in ERA-Interim and numerical diffusion a region of ± 2 grid cells around the hail observation is selected. 3) The time slice with maximum instability is selected at each day and all predictors are calculated from the corresponding sounding. We use local time rather than UTC time to select the four-time slices that occur during a day.

Fig. 2 shows a schematic of the climatological probability density functions (PDFs) of a predictor variable (blue line in Fig. 2) and the conditioned PDF of the same predictor (red line in Fig. 2). In this example hail occurs favorably when the predictor is higher than its climatological average. Based on the conditioned PDF we define the probability for large hail ($P_{(v)}$) for variable v and the variable values x as a smooth transition between environments with low ($P_{(v)}$) and environments with higher ($P_{(v)}$):

$$P_{(v)} = \begin{cases} 0.5 + 0.5 \cdot \tanh\left(\frac{x - \frac{x_{p_1} + x_{p_2}}{2}}{(x_{p_2} - x_{p_1}) \cdot c}\right) & x < \frac{x_{p_2} + x_{p_3}}{2} \\ 1 - \left[0.5 + 0.5 \cdot \tanh\left(\frac{x - \frac{x_{p_3} + x_{p_4}}{2}}{(x_{p_4} - x_{p_3}) \cdot c}\right)\right] & x \geq \frac{x_{p_2} + x_{p_3}}{2}. \end{cases} \quad (1)$$

In Equation (1), p_i denotes percentile values of the conditioned PDF and x_{p_i} stands for the p_i percentile of x . The subscripts i can have 4 values. $i = 1$, and $i = 2$ stand for the lower first and lower second percentile that define the lower limit for hail occurrence. Subscripts $i = 3$, and $i = 4$ stand for the higher first and higher second percentile defining the upper limit for hail occurrence with $p_1 < p_2 < p_3 < p_4$ (see Fig. 2a). The percentiles must be $0 < p < 100$ with $p = 0$ being the minimum of x and $p = 100$ being the maximum.

The positive constant c controls the smoothness of the transition. The transition is a step function for $c = 0$ and is very smooth for c larger than one. In this study we choose $c = 0.3$, which leads to $P_{(v)}$ close to zero for $x_{p_1} > x > x_{p_4}$ and $P_{(v)}$ close to one for $x_{p_2} < x < x_{p_3}$. To exclude the same amount of hail environments we truncate the hail environmental PDF equally on both sides by using $p_{tail} = p_1 = 100 - p_4$ (i.e., if p_1 is the 4th percentile p_4 is the 96th). Similarly, we use $\Delta p = p_2 - p_1 = p_4 - p_3$ to achieve the same smoothness of transition

from hail to non-hail environments on both tails of the distribution. Truncating hail environments can be beneficial due to errors in the ERA-Interim data but can also result in excluding a-typical environments that produce large hail. For example, we will show later that hail was observed at locations were ERA-Interim has no buoyancy, which is unphysical. We perform extensive testing to find an optimal setting for p_{tail} and Δp in section 4.1. We assume that the p values are the same for all predictors although the optimal values might vary according to ERA-Interim's quality in simulating a predictor and the predictors contribution to the hail-producing environment. This assumption was made to improve the computational efficiency of the hail algorithm.

Not all predictors in Table 1 have lower and upper bounds. To determine if there is a physical reason to limit large hail hazard at the upper or lower end of the conditioned distribution we calculated the hail production efficiency and the average hail size (Supplementary Fig. 1). Hail production efficiency is defined as the ratio between the conditioned density function and the unconditioned density function of a predictor – e.g., the hail production efficiency of CAPE = 4000 J kg^{-1} is $\sim 5\%$, which means that large hail is (observed and) reported one out of 20 times when CAPE is that high in the CONUS. Lower and/or upper bounds for hail occurrences are indicated by a decrease in average hail size and hail efficiency towards the tails of the distribution. Predictors that have lower and/or upper bounds are highlighted in Table 1.

So far we have described the probability for large hail $P_{(v)}$ dependent on single predictors but interactions between predictors can also be important. Due to computational limitations, we are not able to test all 231 possible combinations of predictors. Instead, we select the top 20 predictors that have a minimum overlap in the volume of their joint distribution and their conditional joint distribution. Predictor combinations that have small overlaps have the ability to well differentiate hail environments from the background climatology. The top 20 predictor combinations according to this criterion are shown in Table S1 in the supplement. The joint predictors are included in the hail algorithm in a similar way as single predictors. The conditioned 2D distribution for joint predictors (red contours in Fig. 2b) is binned into 35 bins with equal sample sizes – equal numbers of points in each bin – along the x-axis predictor. 35 bins are a compromise between sampling resolution and sample size for statistical robustness.

The y-axis data in each bin is used to calculate probabilities for large hail ($P_{(v_1 v_2)}$) similar to individual variables. To increase the computational efficiency we discretize the smooth transition function in equation (1) into a step function with five steps (right inlay in Fig. 2b) resulting in a stepwise increase/decrease of hail hazard instead of a

smooth transition.

$$P_{(v_1, v_2)} = \begin{cases} 0 & x < x_{p_1} \\ 0.12 & x_{p_1} \leq x < x_{p_1} + \Delta x \\ 0.34 & x_{p_1} + \Delta x \leq x < x_{p_1} + 2\Delta x \\ 0.66 & x_{p_1} + 2\Delta x \leq x < x_{p_1} + 3\Delta x \\ 0.88 & x_{p_1} + 3\Delta x \leq x < x_{p_1} + 4\Delta x \\ 1 & x_{p_2} \leq x < x_{p_3} \\ \dots & \end{cases} \quad (2)$$

Here $\Delta x = (x_{p_2} - x_{p_1})/5$. Equation (2) only shows the increasing probability between x_{p_1} and x_{p_2} – from non-supportive to supportive environments. The transition from supportive to non-supportive environments between x_{p_3} and x_{p_4} is equivalent.

We use a linear extrapolation of the discrete hail hazard probabilities at the tails of the 2D conditioned distribution to estimate hail hazard outside the observed predictor space. We fit the linear model to the lowest/highest five bins to guarantee a smooth transition between calculated and extrapolated hail hazard probabilities.

The probability of an environment producing large hail (P_H) is given by the product of the considered $P_{(v)}$ and $P_{(v_1, v_2)}$ probabilities. This means, P_H will be zero even if only one of the environmental conditions is not favorable for large hail development. This can be problematic since slightly varying combinations of variables can compensate each other and increase the environments ability to produce large hail. The smooth transitions between favorable and non-favorable environments partly account for this effect but large hail that is produced by atypical environments is not captured by the algorithm as we will show later.

4.1. Model optimization

The goal of the algorithm optimization is to achieve the most skillful hail estimates with a minimum number of predictors. We use three metrics to optimize the algorithm's ability to capture 1) the shape of the annual cycle of hail occurrence, 2) the sum of hit minus miss events, and 3) the climatological spatial pattern of large hail in the CONUS.

The first metric is the root-mean-squared-error (RMSE) between the normalized observed ($\|O_m\|$) and modeled ($\|P_m\|$) annual cycle of monthly mean CONUS-wide hail frequency:

$$\|P_m\| = \frac{P_{(H)m}}{\sum P_{(H)}}, \quad O_m = \frac{O_m}{\sum O} \\ RMSE_{AC} = \sum_{m=1}^{12} \sqrt{\|O_m\|^2 - \|P_m\|^2}. \quad (3)$$

Here $P_{(H)}$ is the hail hazard probability and the index m indicates the monthly mean. We use the normalized frequency to minimize the impact of undersampled hail observations on the statistic and assess if the algorithm can reproduce the shape of the observed annual cycle, which is less affected by under observing and artificial trends (Allen and Tippett, 2015). The lower the $RMSE_{AC}$ the better. This metric ensures that the seasonal cycle of hail hazard is well represented in the algorithm.

The second metric is the normalized sum of hit minus miss events (NHM).

$$NHM = \sum_{i=0}^N \sum_{j=0, t=0}^t \begin{cases} 1 & 0.25 \leq \sum_{k=i-2}^{i+2} \sum_{l=j-2}^{j+2} P_{(H)k,l,t} \\ -1 & 0.25 > \sum_{k=i-2}^{i+2} \sum_{l=j-2}^{j+2} P_{(H)k,l,t} \end{cases} \\ \|NHM\| = \frac{|NHM|}{\sum P_{(H)}} \quad (4)$$

Here the subscripts i, j , and t stand for the latitude, longitude, and day of observed hail events ($t = 0, 1, 2, \dots, N$). If the sum of hail probabilities $P_{(H)}$ in a square of 5×5 grid cells centred around the hail observation is larger or equal to 0.25 we count this as a hit event (plus one). If it is lower than 0.25 the algorithm has failed to capture the observed hail event (minus one). If more hail events have been captured than missed NHM is positive. Finally, we divide the NHM by the sum of

hail probabilities $\sum P_{(H)}$ to penalize overprediction. A perfect algorithm would have a $\|NHM\| = 1$ and an algorithm that has more hit than missed events has a positive $\|NHM\|$. Note that a $\|NHM\|$ score of one is not desirable due to the under observation of hail. This metric ensures that the algorithm is not overpredicting hail hazard and that individual hail events are captured.

The third metric is Spearman's R rank correlation coefficient between the observed and modeled climatological average large hail frequency in the CONUS (Wilks, 2011). A perfect algorithm would have a Spearman's R of one and a skill-less algorithm would be close to zero. This metric ensures that the climatological average spatial pattern of hail hazard is captured in the algorithm.

In the optimization, we test p_{tail} values between one and eleven and Δp values between three and twenty-five. Increasing the p_{tail} value results in excluding more observed hail environments while increasing Δp values results in more diffusive hail hazard estimates. For each p_{tail} and Δp setting, all three skill score metrics are calculated for each predictor in Table S1 and each predictor combination in Table S1 (in supplement). The goal is to find the $P_{(v)}$ and $P_{(v_1, v_2)}$ combination that leads to the highest possible overall skill. The overall skill is calculated by ranking the skill-scores for each of the three metrics and summing the ranks. The predictor that has the lowest sum of ranks $P_{(O1)}$ is considered to be the most skillful in predicting hail hazard.

In the next step the remaining predictor probabilities $P_{(v)}$ and predictor combinations $P_{(v_1, v_2)}$ are multiplied to $P_{(O1)}$ to find the predictors that lead to the highest skill scores. This procedure is performed until all predictors and predictor combinations are included:

$$P_{(H)} = P_{(O1)} \cdot P_{(O2)} \cdot P_{(O3)} \cdots P_{(ON)}, \quad (5)$$

where N is 42 – the number of predictors or predictor combinations in Table 1 and Table S1. At a certain point adding more predictors leads to deteriorating skill in predicting large hail. Therefore, we search for the number of predictors that lead to the most skillful $P_{(H)}$ estimates and stop adding new predictors when skill increases become small (i.e., when the overall skill reached 95% of the maximum possible skill).

Fig. 3 shows the most skillful parameter settings for algorithms with different number of predictors. In general, settings with large p_{tail} lead to low skills but need only a few variables to reach their optimum performance and settings with small Δp have lower skills and demand more input variables.

We decided to use $p_{tail} = 3$ and $\Delta p = 25$ as setting for the hail algorithm, which has a very good overall performance and needs four predictors which are: a joint predictor between maximum convective available potential energy (CAPE) and freezing level height (FLH), CAPE as a single predictor, and surface to 3 km storm-relative helicity (SRHO-3), and vector shear (VS0-3). Adding additional predictors does not improve the overall skill of the algorithm. The resulting percentile thresholds that are used for single predictors (see equation (1)) are shown in Table 2 and the hail environments are shown in Fig. 4.

$$P_{(H)} = P_{(CAPE, FLH)} \cdot P_{(CAPE)} \cdot P_{(SRHO-3)} \cdot P_{(VS0-3)}. \quad (6)$$

Other settings that lead to similar high skill scores result in a similar set of predictors and show comparable performances in the CONUS and other mid-latitude regions, but hail estimates can vary in amplitude (see Supplementary Fig. S2 for an example).

The first predictor selected by the optimization algorithm is the joint conditional distribution of the most unstable CAPE and FLH (Fig. 4d). CAPE is directly related to the potential updraft strength of a thunderstorm and has proven its value in various severe weather indices (Brooks et al., 2003; Allen and Karoly, 2014). CAPE is by far the best single discriminator between hail days and non-hail days since its climatological and conditional density distribution only have an overlap of 21% (Fig. 4a). Since it is not obvious which parcel should be used in calculating CAPE we decided to use the most unstable CAPE to account for elevated convection and shallow boundary layers. Previous studies that focused mainly on the central US (Craven et al., 2002) showed that

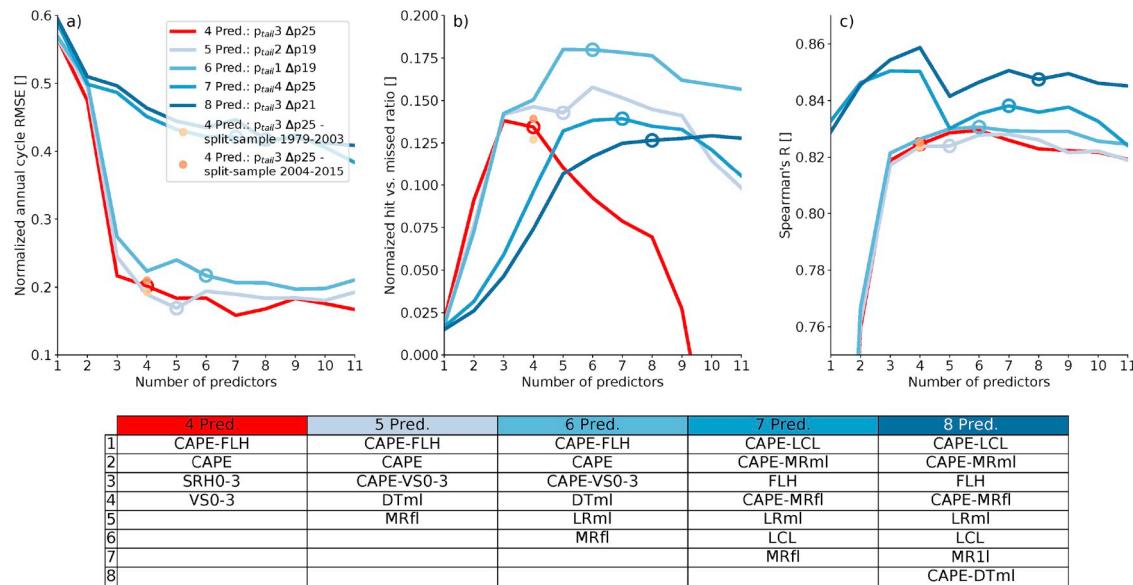


Fig. 3. Three skill-scores used for the algorithm optimization. Panel a shows the root-mean-squared-error of the normalized monthly mean annual cycle of hail frequencies (lower is better). Panel b shows the normalized hit minus miss event score (higher is better) and panel c shows the Spearman rank correlation coefficient for climatological average hail frequencies (higher is better). The different colored lines show the best performing algorithm settings depending on the number of predictors. The colored circles show the skill-scores where the algorithm reaches 95% of its global optimum setting, which denotes the point where the addition of additional predictors does not lead to significant improvements. The final algorithm setting with four predictors is shown in red. Filled orange circles show the skill scores from a split sample analysis and indicate no over-fitting issues. The table at the bottom shows the predictors for algorithms with different number of predictors. The predictor acronyms are explained in Table 1. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

Table 2

Parameters used in equation (1). The values are based on the conditioned PDFs in the CONUS with $q_{tail} = 3$ and $\Delta q = 25$.

Variable and unit	x_{p1}	x_{p2}	x_{p3}	x_{p4}
CAPE [J kg^{-1}]	105	646	–	–
VS0-3 [m s^{-1}]	6.9	12.7	–	–
SRH0-3 [$\text{m}^2 \text{s}^{-2}$]	35	138	–	–

CAPE for a mean layer parcel averaged over the lowest 100 hPa is skillful in predicting large hail occurrences but other regions such as Australia typically have shallower moisture reservoirs and calculating CAPE for average parcels over the lowest 50 hPa is beneficial (Allen et al., 2011; Allen and Karoly, 2014). Using maximum CAPE, therefore, is a compromise to account for a wider variety of hail-producing environments. FLH is important for hail development since too high freezing levels can lead to melting of hailstones before they reach the surface (Dessens et al., 2015) while too low melting levels might limit the amount of supercooled water in areas with strong updraft speed. This joint predictor shows that large hail can occur in low CAPE environments if FLH is low. With increasing FLH larger CAPE also is necessary. This joint predictor does not constrain environments with low or zero CAPE (Fig. 4d), which makes it beneficial to add CAPE as the second predictor to the algorithm (Fig. 4a). The third predictor in our algorithm is SRH0-3, which leads to a substantial improvement in all three considered skill scores (Fig. 3). SRH0-3 is a proxy for cyclonic updraft rotation that occurs in supercell thunderstorms (Davies and Ponal, 1990). Large values of SRH0-3 have been shown to be favorable for large hail development (Rasmussen and Blanchard, 1998). SRH0-3 was calculated using the wrf.srhel function of the wrf-python package (Ladwig, 2018), which is based on Kain et al. (2008). Fourth, adding a lower limit for VS0-3 improves the normalized annual cycle and Spearman's R rank correlation. Wind shear plays an important role in amplifying and sustaining the updraft through storm organization (Weisman and Klemp, 1982; Rasmussen and Blanchard, 1998; Brooks

et al., 2003; Dennis and Kumjian, 2017). Adding VS0-3 and SRH0-3 to the algorithm leads to a shift in peak annual hail frequency from mid-summer to late spring. Using large-eddy-simulations (Dennis and Kumjian, 2017) show that VS0-3 can, in some cases, reduce the hail mass in a storm and that deep-level, 0–6 km shear, could be a more reliable predictor. This is in contrast to our findings, which favor low-mid-level over deep-level shear for hail prediction.

In statistical modeling, care has to be taken to avoid model overfitting. Overfitting denotes the formulation of an overly complex algorithm that corresponds too closely to a dataset, and may, therefore, fail to reliably model additional data or to predict future observations (Wilks, 2011). To ensure that the derived algorithm is not overfitted we use two tests. 1) Split-sample cross-validation only uses half of the observed hail observations for the model training and uses the other half for independent validation. In our case, this means splitting the time series at the beginning of 2002. The split sample hail estimates have very similar skill than the original estimates (orange circles in Fig. 3a-c), which means that the algorithm is not overfitted. 2) The evaluation of the algorithm with independent hail observations from Europe and Australia, shown in section 5, does not show signs of overfitting.

4.2. Regional dependency of hail environments

The algorithm development is solely based on hail observations from the CONUS. To test the robustness of the derived hail environments we investigate if hail environments in Europe and Australia are similar to those found in the CONUS. A comparison between the 10th percentiles for CAPE, VS0-3, and SRH0-3 of conditioned PDFs in sub-regions of the CONUS, Europe, and Australia (Fig. 5a-c) shows that there can be differences in the hail environments. CAPE is clearly higher in most regions of the CONUS. The conditional PDFs for Australia and central Europe agree well especially for values larger than 2 kJ kg^{-1} , however, Allen and Karoly (2014) showed that ERA-Interim systematically underestimates CAPE in Australia. VS0-3 is similar in Australia and in the CONUS but is lower in European hail storms. This

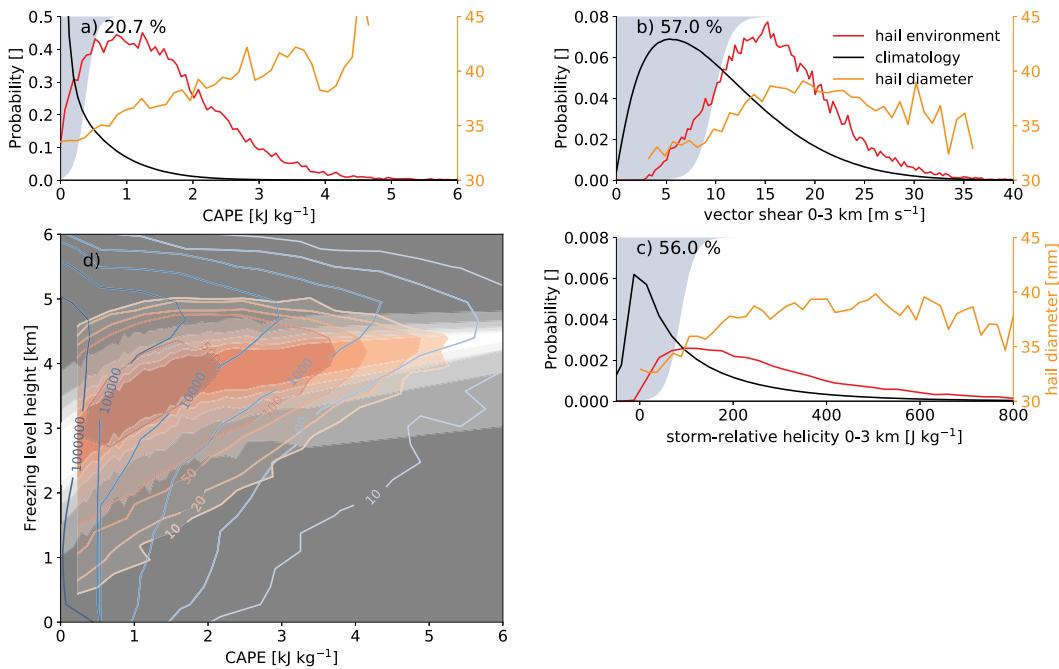


Fig. 4. Climatological PDFs for all days within 1979–2015 (black line) and conditioned PDFs (red solid line) for CAPE, MRfl, and DTml (panels a–c respectively) for the CONUS environments. Values in the upper left corner show the overlapping area between the two PDFs. Additionally, the average observed maximum hail size conditioned on the corresponding predictor is shown as an orange line (right y-axis). Panel d shows the joint distribution between CAPE and FLH for the climatology (blue contours) and conditioned on hail days (red contours). In panel a–c, gray areas in the background show the probability for hail development from zero to one (bottom to top of y-axis) according to equation (1). The dark gray to light gray contours in panel d show the transition of hail probabilities for joint predictors according to equation (2). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

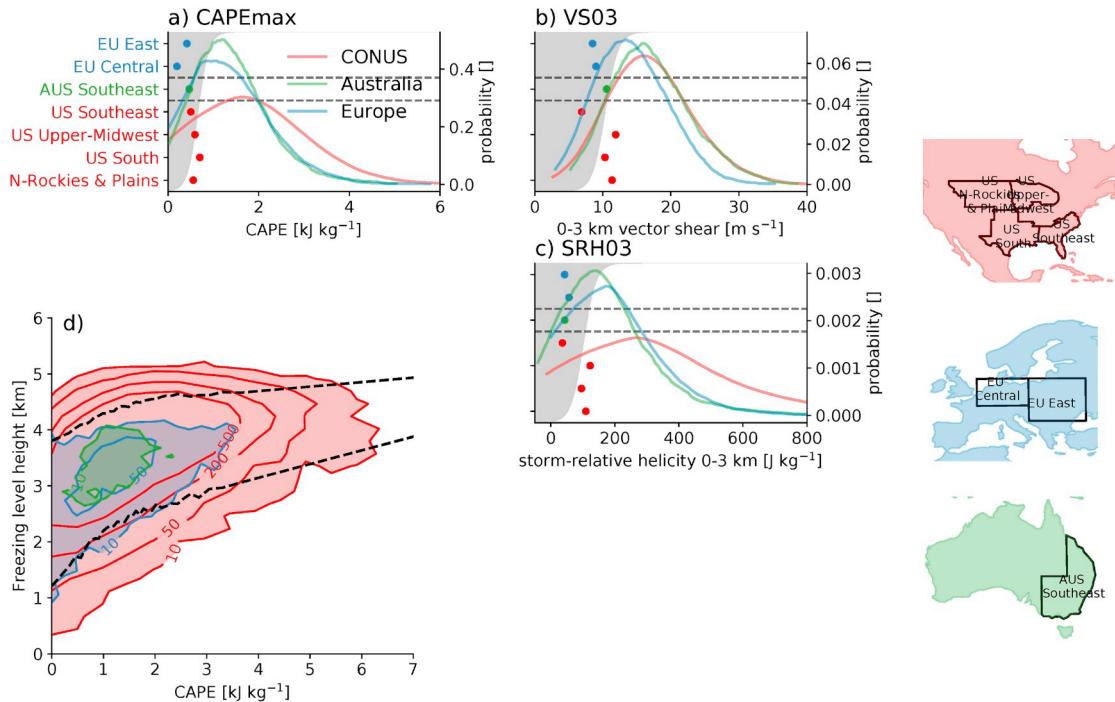


Fig. 5. Similar to Fig. 4 but showing conditioned PDFs for the CONUS (red), Australia (green), and Europe (blue). A 25 point moving average smoothing has been applied to remove noise from the PDFs. Colored markers in Panels a–c show the 11 percentile of the conditioned PDF for different climate regions in Europe, Australia, and the CONUS. The regions are highlighted in maps on the right. Panel d shows the conditional joint distribution between CAPE and FLH where red, blue, and green contours show hail producing environments for the CONUS, Europe, and Australia respectively. The black dashed lines outline the area in which large hail can occur in our algorithm. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

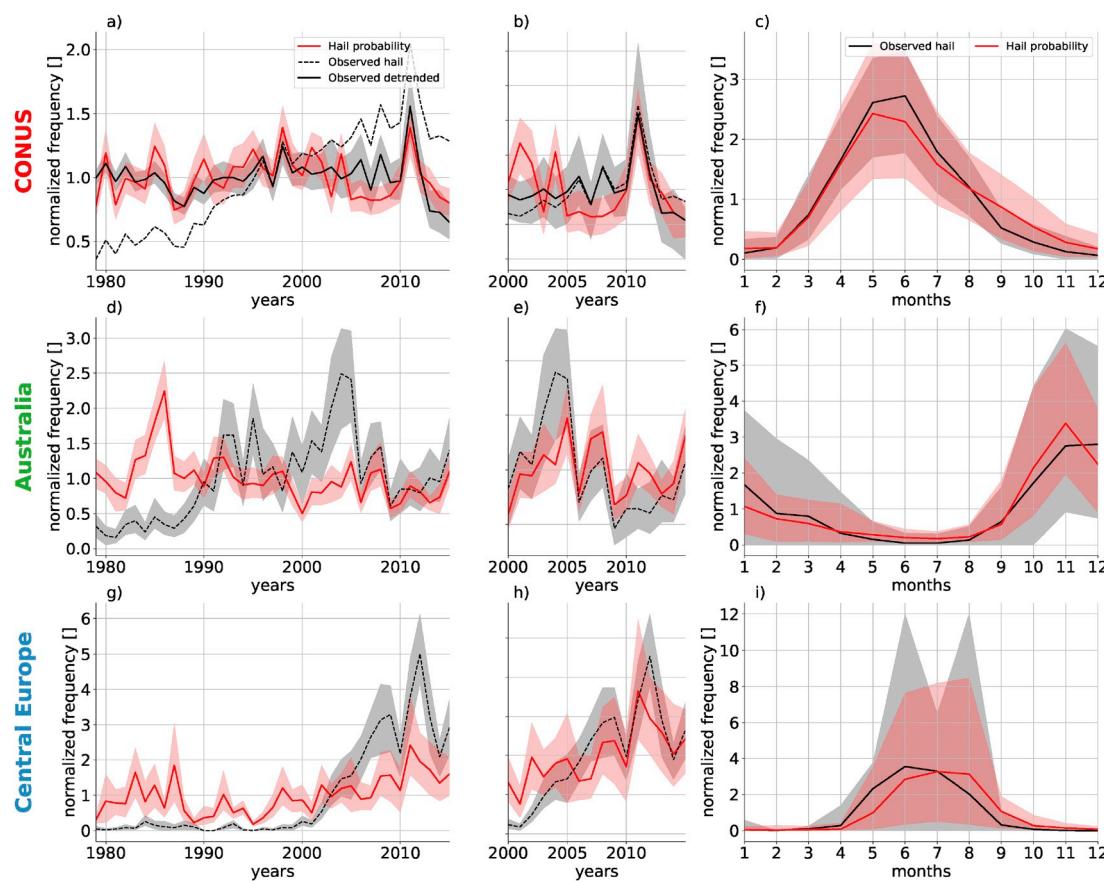


Fig. 6. Annual hail frequencies normalized by the climatological average (1979–2015) for the CONUS (top row), Australia (middle row), and Central Europe (bottom row). A zoom-in on 2000–2015 is shown in panel b, e, and h. Black dashed lines show the normalized hail observations, black solid lines the detrended normalized observations (for the CONUS only), and red lines show the modeled hail probabilities. Shaded show the 5–95 percentile of the daily sampling uncertainty in each year of 100 bootstrap samples. The right panels (c, f, and i) show the normalized annual cycle of observed and modeled hail frequencies. The contours show the 5–95 percentile spread in the interannual variability. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

might be related to orographic effects that compensate for lower shear environments in European hail storms. Low shear environments can also be seen in the US Southeast where pulse storms can produce hail with diameter ≥ 2.5 cm in low shear environments (Miller and Mote, 2017). Large differences occur also for SRH0-3 where the CONUS clearly has larger values than Europe and Australia. Only in the US Southeast is the 10th percentile of the SRH0-3 PDFs similar low to the European and Australian regions. The joint conditional PDFs of CAPE and FLH agree well with each other and the hail environment limits that are based on the CONUS hail environments include European and Australian conditions (Fig. 5d). Although the conditional PDFs can be different between the three continents, the 10th percentiles (circles in Fig. 5a–c) generally fall within the transition region from non-favorable to favorable environments (gray shading in background). This is in agreement with previous studies that found similarities in extreme convective storm environments (Brooks, 2009; Allen and Karoly, 2014). However, these differences could introduce systematic biases in the hail hazard estimates.

5. Hail algorithm evaluation

5.1. Annual cycle and interannual variability

Although we cannot use the raw hail observations for algorithm evaluation due to the considerable spatiotemporal inhomogeneities, the normalized annual cycle of monthly hail frequencies and the detrended interannual variability can provide more unbiased information. Similar

to Allen et al. (2015a) we linearly detrend the CONUS observed annual mean hail frequencies to remove artificial trends and to simplify the comparison between observed and modeled interannual variability. We only show the raw observations for Europe and Australia due to their highly non-linear temporal development.

For the CONUS the hail algorithm is able to capture the main variability of the observed annual time series, especially the peak years in 1980, 1998, and 2011 and the low years of 1987, 1988, 2007, 2014, and 2015 (Fig. 6a and b). The variability in large hail frequency in the CONUS has partially been attributed to the El Niño–Southern Oscillation cycles (Allen et al., 2015b) and our algorithm is able to pick up this signal (not shown). Minor differences between the modeled and observed annual frequencies should not be over-interpreted due to the large uncertainties in hail observations. Also, the modeled normalized annual cycle of monthly mean hail frequency is well captured (Fig. 6c). The biggest difference occurs in September and October where the algorithm overestimates the observed normalized frequency. These differences can also be found in other hail models (Allen et al., 2015a) and their origin will be discussed later.

The observed annual time series in Australia is more variable than the one in the CONUS (Fig. 6d), which is captured by the hail algorithm. Focusing on the more recent period from 2000 to 2015 shows that the algorithm is able to capture the main mode of the normalized interannual variability (Fig. 6e). Also, the normalized annual cycle of hail frequency is well captured (Fig. 6f). The overall good performance of the algorithm in Australia is encouraging since the algorithm development is solely based on the CONUS data.

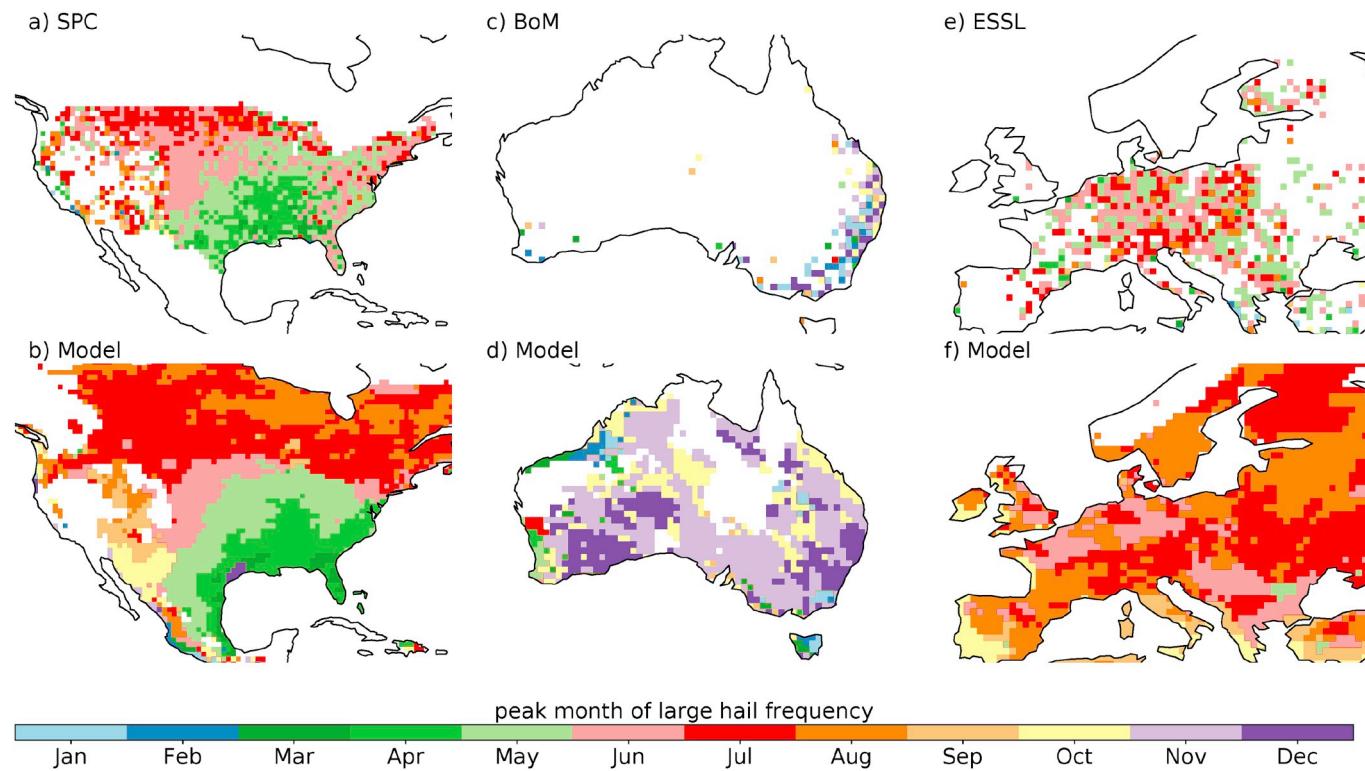


Fig. 7. Color coded month of peak observed (top panels) and simulated (bottom panels) large hail frequencies for the CONUS, Australia, and Europe (left to right). Areas with at least 2 large hail days per year in a 100×100 km area within the period 2000–2015 are shown. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

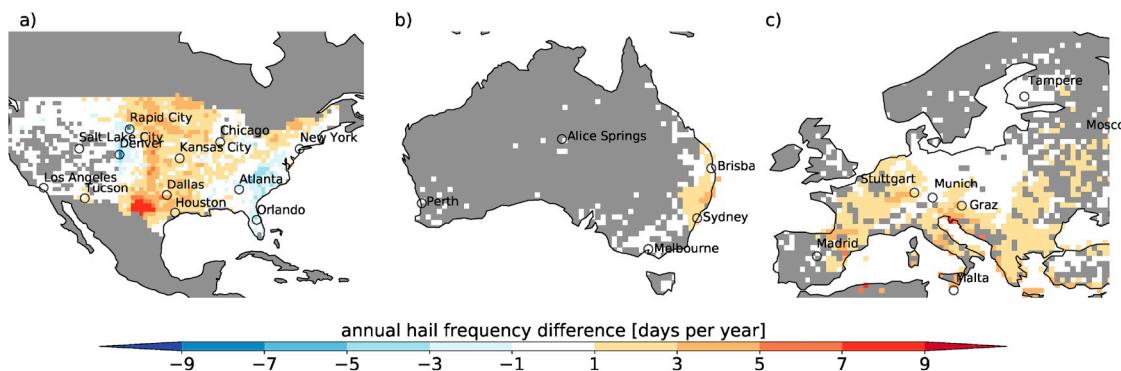


Fig. 8. Difference – simulated minus observed – annual average hail days for the period 2005–2015. Areas where no large hail was observed in this eleven year period are shown in gray.

The algorithm evaluation in Europe is hardest due to the substantial inhomogeneity of the hail observations (Fig. 6g) and the time series are mainly shown for consistency than for skill analysis. We choose to focus on the Central Europe region (Fig. 5 in blue map) where observational coverage is relatively high. The algorithm is able to qualitatively capture the main mode of the observed variability in recent years (Fig. 6h) and has a reasonably good representation of the annual cycle concerning the interannual variability, although the modeled median cycle is shifted by one month. One explanation for this shift could be the higher frequency of very damaging and large hail stones in June and July compared to more frequent but smaller events in May and June (Punge and Kunz, 2016) assuming that our algorithm is more effective in identifying large and giant hail.

The observed and modeled peak months of large hail frequencies in the CONUS agree well between the observations and the algorithm and show a seasonal northward propagation (Fig. 7a and b). March and April peaks occur in the Southeast while mid-to late-summer peaks are

found in the Plains, Rocky Mountains, and Northeast. The largest differences occur along the southern Atlantic coast, from North Carolina to Florida, where observations partly show later peak months. In the western half of the CONUS, the pattern of observed peak months is noisy due to the low large hail frequency, which complicates evaluations. In Australia modeled and observed results agree well, although a comparison is complicated by the under-sampling of hail frequencies in large parts of the country (Fig. 7c and d). Large hail occurs most frequently during early summer in most parts of the country and late summer along the west coast. The algorithm seems to shift peak hail occurrence by one month in Europe consistent with Fig. 6i, although a detailed analysis is not possible due to the noisy and sparse observational record (Fig. 7e and f).

5.2. Annual large hail frequency

As we discussed in section 2 it is difficult to derive the frequency of

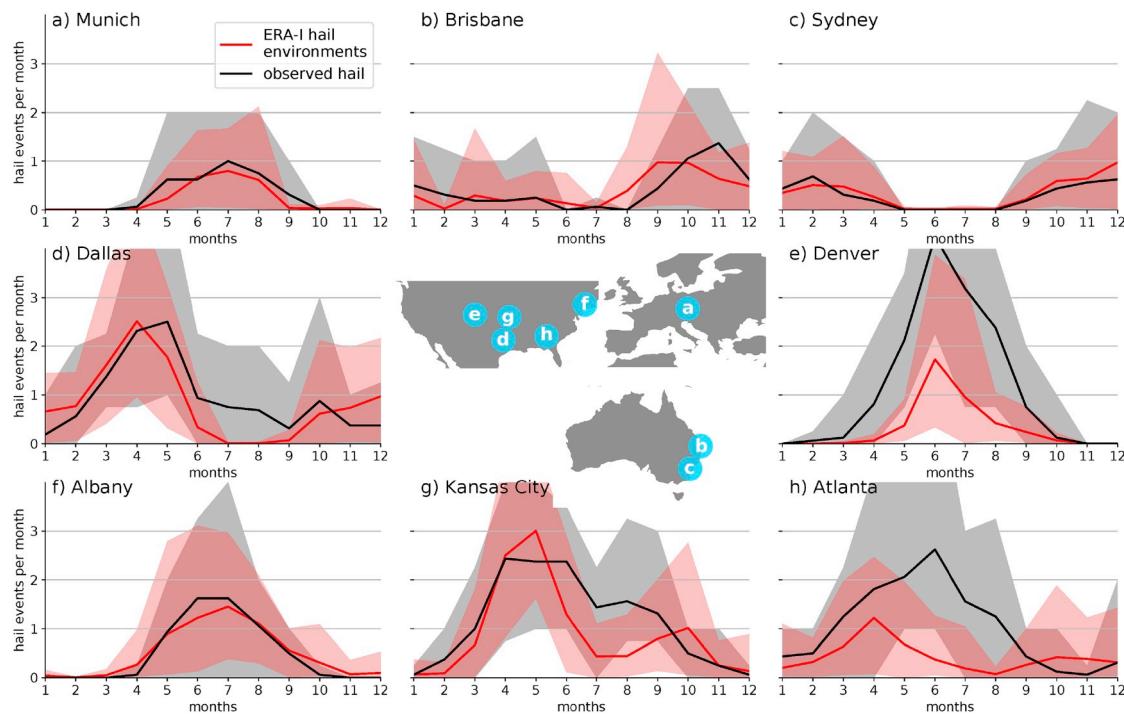


Fig. 9. Average annual cycles of observed (black) and simulated (red) absolute hail frequencies for selected cities in the CONUS, Australia, and Europe within the period 2000–2015 (city locations are shown in maps). Shown is the maximum monthly hail frequency from the closest 3×3 grid-cell centred on each city and its surrounding land-cells. Contours show the 5–95 percentile spread in the interannual variability. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

large hail occurrence from observations due to undersampling and artificial trends (Allen and Tippett, 2015). This means that a certain degree of overprediction of estimated hail hazard is desirable and can be seen in most regions in the CONUS, Australia, and Europe (Fig. 8). We focus the analysis on the most recent 11-years (2005–2015) due to the improved coverage of hail observations (see Fig. 1). In the CONUS only the front range region in Colorado and Wyoming, and the region east of the Appalachians shows lower than observed large hail frequencies. This is likely related to missing orographic effects on hail development due to the low spatial resolution of the ERA-Interim data. Largest positive differences occur in south-east Texas. The overestimation in this region is likely related to strong convective inhibition, which is not well simulated in reanalysis products (Brooks et al., 2003; Gensini and Ashley, 2011). Differences are generally smaller in densely populated areas where observational under catch is low. Similar characteristics can be seen in Australia and Europe.

We take advantage of the higher accuracy of hail observations in populated areas for evaluating the algorithm's performance in capturing the frequency, annual cycle, and interannual variability on regional scales within the period 2005–2015. The algorithm shows acceptable performance for many cities around the world (Fig. 9). Largest differences occur for Denver and Atlanta, which lie in areas that have been already identified as low biased in Fig. 8a. Also, Dallas and Kansas City show low biases in late summer. Other hail models that use reanalysis data as input show similar differences in these regions (Allen et al., 2015a).

To better understand why the hail algorithm fails to reproduce observed hail frequencies we investigate the difference in predictor variables during observed and modeled hail days in three focus-regions (Fig. 10). We first focus on an area between Minnesota and Iowa where we see an overprediction of 42.5% in the hail estimates during fall (e.g., Fig. 8). The PDF of CAPE conditioned on observed and modeled hail days are very similar (Fig. 10a). VS0-3 and SRH0-3 (not shown) is slightly lower in observed than in modeled hail events but both PDFs overlap substantially (Fig. 10b). Also, other environmental variables

such as CIN, and humidity do agree, which means that we cannot attribute the overprediction to differences in the investigated predictors. Possible bias sources are an under-observation of hail in this region, erroneous ERA-Interim soundings, and/or a missing trigger mechanism that inhibits the realization of the potential for large hail development. The second region is focused on Denver in June where the peak in hail frequency is underestimated by 59.3% in our algorithm (Fig. 9e). Observed hail partly occurs in environments with atypically low FLH, CAPE, and SRH0-3 compared to the rest of the US (Fig. 10c–e). A possible reason for the low bias in this region is missing orography effects on hail producing storms due to the poor representation of terrain in ERA-Interim. The third region is focusing on Atlanta during June where the hail algorithm underestimates hail frequencies by 88%. Although CAPE is sufficiently high in this area (Fig. 10f) the very low VS0-3 and SRH0-3, in which observed hailstorms develop, leads to a significant underestimation. The southeastern US experiences a high frequency of pulse thunderstorms during summer, which are known to occur in weakly sheared environments but can produce large hail (Miller and Mote, 2017). Contributing to the low bias are known issues with the reliability of hail size reports in this region (Cintineo et al., 2012; Allen et al., 2015a). For example, Allen and Tippett (2015) show that changes in the classification of hail size in 2010 lead to an overnight decrease in hail of 0.75" diameter, and a corresponding increase in hail in excess of 1".

5.3. Global large hail hazard and comparison to other severe weather indices

Since similar large-scale environmental conditions lead to large hailstorms in the CONUS, Europe, and Australia we feel confident in applying the hail hazard algorithm on a global scale. Fig. 11 highlights several hot spots for hail hazard. The most prominent is located on the lee side of the Rocky Mountains in the U.S. Plains. Additional hot spots are on the lee side of the Andes in Argentina, in the southern foothills of the Himalayas, in central Africa, and in the mountains of the Arabian

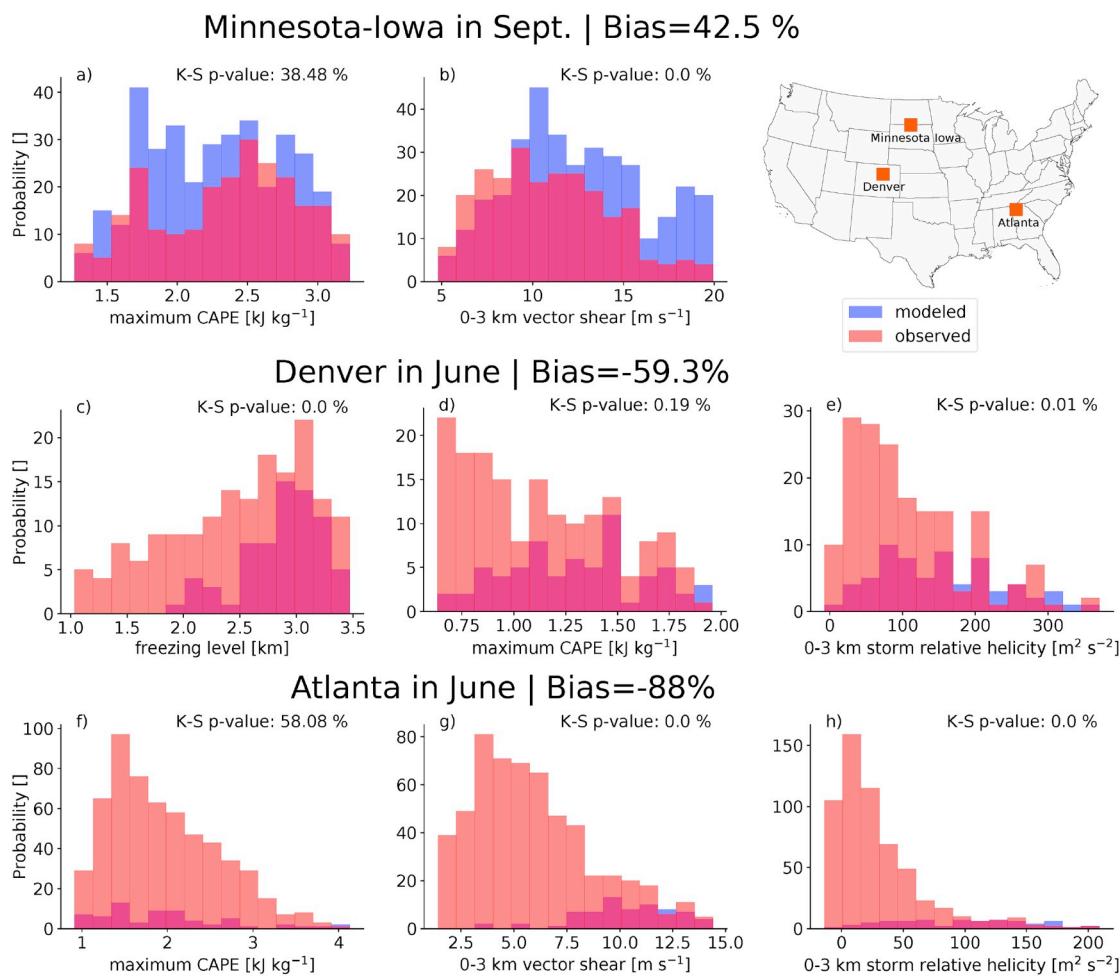


Fig. 10. Difference in the density functions of large-hail predictors between observed (red) and modeled hail events (blue). Shown are results for 2000–2015 at the Minnesota-Iowa border in September where hail estimates are 42.5% too high (panel a and b) and areas focused on Denver (panel c–e) and Atlanta (panel f–h) during June where the algorithm underestimates observed hail frequencies by 59.3% and 88% respectively. *P*-values from a two-sided Kolmogorov-Smirnov (K–S) test are, shown in the top right corner in each panel. Values lower than 5% indicate significantly different distributions. The investigated areas are shown in the top right map. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

Peninsula. The map clearly shows the importance of orography and land-sea contrast in hail formation. It also shows the zonal dependence of hail hazards with low hazard in the equatorial region, highest hazard around $\pm 30^\circ$, and low-hazard poleward of $\pm 60^\circ$. Further, the east sides of the continents typically have higher large hail hazards than the west sides.

The hail hazard map in Fig. 11a partly highlights similar hot spot regions compared to the significant severe parameter (Fig. 11b; Craven et al., 2004), the severe thunderstorm environments index (Fig. 11c; Brooks et al., 2003), and the SHIP parameter (11d). Although a direct comparison is complicated since the first two of these indices are rather tailored to severe convective storms and not specifically to hail environments or have different hail diameter thresholds, it is obvious that they have much higher risk for severe convection in the tropics over Africa (except for SHIP), the Arabian Sea, the Gulf of Mexico, and Northern Australia. Using satellite data Cecil and Blankenship (2012) and (Mroz et al., 2017) generally show high hail hazards in the tropics and subtropics including Northwestern Australia where freezing levels are high. Hail observations from Australia (Fig. 1b) do not confirm these high frequencies, which is supported by Allen and Allen (2016) and Bedka et al. (2018). However, due to a lack of ground observations, tropical and subtropical large hail hazards are not well understood.

The SHIP parameter is designed to capture severe hail environments and can, therefore, be more directly compared to our estimates. We consider a SHIP index larger than 0.5 as likely to produce large hail.

SHIP captures the hail hazard hot spots in the CONUS and Argentina well but tends to substantially overestimate hail risk especially in subtropical coastal regions and has steep gradients between high and low hazard areas. SHIP is tuned to capture large hail occurrences in the CONUS and significantly under-predicts hail frequencies in Europe and Southeast Australia compared to observations.

On a regional scale, our results compared well to hail estimates by Bedka et al. (2018) over Australia and highlight the Southeast Coast as the main area at risk with peak frequencies of up to 6 hail events per year in a 100^2 area in northern New South Wales and Southern Queensland. This area is also highlighted in other severe weather indices but the number of severe days is two to three times higher than indicated by severe hail observations, which means that their applicability for hail hazard assessments is difficult.

Our algorithm shows higher hail frequencies in the tropical regions than estimates by Bedka et al. (2018). Observed large hail frequencies in China agree well with our estimates (see Li et al., 2018, Fig. 7) with highest frequencies in Northeast and North China, South Central China, and in the far Northwest. However, the algorithm tends to underestimate frequencies in Central China and parts of the Tibetan Plateau. Large hail frequencies and spatial patterns agree well with studies in South Africa (Blamey et al., 2017) although Le Roux and Olivier (1996) shifts the area of largest risk further inland. Hail frequency estimates over Europe clearly highlight the surrounding of the Alps, Pyrenees, and the Carpathians as the areas under highest risk (Punge et al., 2014,

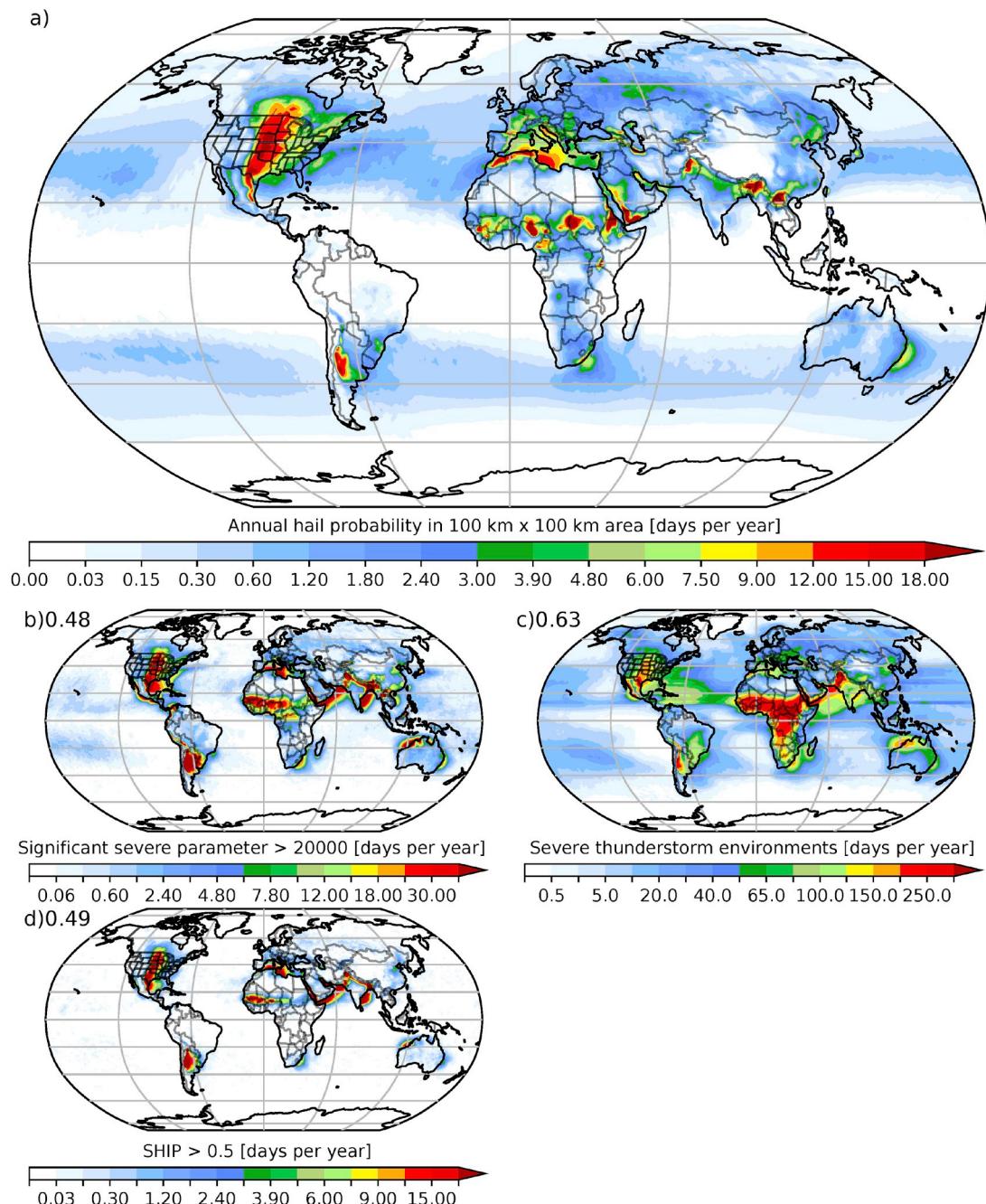


Fig. 11. Global annual average large hail probability normalized to a 100×100 km area from 1979 to 2015 (panel a). Annual number of days with significant severe parameters larger 20 000 (panel b), severe thunderstorm environments (panel c), and SHIP parameters > 0.5 (panel d). The numbers in the top left corner in panel b-d show Spearman's rank correlation coefficient compared to the hail frequency estimates in panel a.

2017; Mohr et al., 2015; Punge and Kunz, 2016) in agreement with our estimates. Punge et al. (2017) include northern Africa in their hail frequency estimates, which seems to be over-predicted in our algorithm. Martins et al. (2017) assess destructive hail storms in Brazil and highlight southern Brazil as the primary and southeastern Brazil as the secondary hot-spot in good agreement with our estimates. Also, the spatial patterns of hail frequencies in Argentina agree well with observations except for a low bias in the southern part of the country (Mezher et al., 2012).

Comparing the normalized annual cycle of our large hail frequency estimate with the significant severe parameter, the severe thunderstorm environments index, and the SHIP (see Section 3 for a definition of these indices) shows that our estimates are closest to the observations in

the CONUS (Fig. 12a). The other indices underestimate the hail risk in spring and overestimate the risk in late summer and fall. Our hail algorithm also outperforms the other indices in Australia (Fig. 12b). Over Central Europe, all parameters show a shifted hail hazard by approximately one month (Fig. 12c). The significant severe and SHIP parameters have a pronounced maximum in June. Our hail frequency estimates have the lowest normalized annual cycle RMSE in all regions (Fig. 12d). They particularly outperform the other indices in Australia. The SHIP index performs fairly well over the CONUS but has the lowest skill in the other regions, which shows that it is tuned to capture US hailstorms.

It is important to notice that the other three indices that we compare our algorithm with, were not specifically developed to capture hail

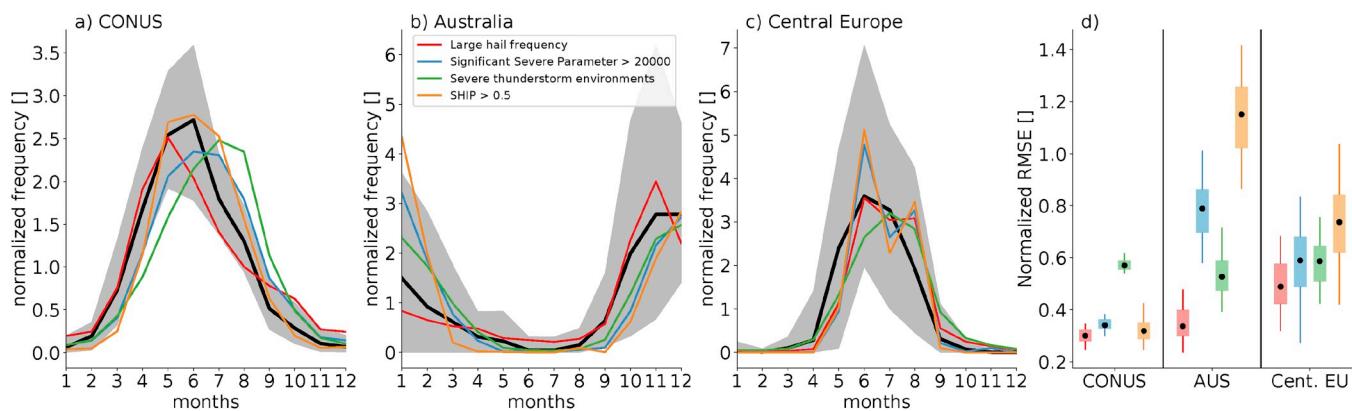


Fig. 12. Normalized annual cycles of observed large hail frequency (black) and four hail indices: 1) our hail algorithm's large hail frequency (red), 2) days with significant severe index values larger 20 000 (blue), 3) severe thunderstorm environments (green), and 4) SHIP > 0.5 (orange). Shown are results for the CONUS, Australia, and Central Europe (panel a–c) for the period 2000–2015. Panel d shows box-whisker statistics for the normalized annual cycle RMSE according to a 100 member bootstrap ensemble. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

frequency larger than 25 mm in diameter and that an improved performance of our hail frequency estimates should be expected to some degree.

6. Summary and conclusion

The presented hail algorithm is based on four environmental ingredients that are present during large hail observations. The algorithm is developed with US hail observations (Schaefer and Edwards, 1999) and its performance is tested over the US, Australia, and Europe. The algorithm has skill in:

- reproducing the shape of the annual cycle of hail frequencies from regional to continental scales,
- simulating the observed frequency of hail occurrence,
- reproducing observed spatial patterns of hail frequencies and their seasonal evolution, and
- simulating the observed interannual variability of large hail frequencies.

The largest deficiencies, which result in a low-biased hail hazard estimate, are found in areas with strong local scale forcing and in areas and seasons where hailstorms develop in atypical conditions. The former deficiency affects particularly region on the leeward side of mountains. This is likely due to the coarse grid spacing and systematic biases in the ERA-Interim input variables. The latter affects the Southeastern US where pulse storms are the dominant mode of severe convection during summer. In areas with strong inhibition, an overprediction of hail frequencies can occur.

The algorithm compares well with existing hail risk estimates from hail pads, satellite data. It also agrees well with estimates of other environmental ingredient-based indices concerning the spatial locations of the major hail hazard regions in mid-latitudes. Major differences occur in tropical and subtropical regions where large hail hazards are not well observed.

The advantage, compared to existing approaches, is its flexibility to work with a wide variety of sounding data (i.e., observed, forecast, and climate model soundings), its ability to provide daily and sub-daily hail risk estimates, its global scope, and its non-parametric approach that does not demand scaling parameters and is not based solely on parameterized families of probability distributions. We show that our algorithm has advantages over three widely used environmental based severe convective indices that provide hazard assessments on a daily basis in terms of capturing the annual cycle of hail hazards and its accuracy in predicting large hail frequency. For monthly averages over the US, the algorithm performance similarly to the hail model by Allen et al. (2015a), which also uses four predictor variables but has a simpler algorithm.

Our algorithm has a wide range of applications including hail forecasting and climate assessments on regional to global scales. Its statistical approach and input variables are different from existing hail models and indices that focus on multiple convective extremes (Gensini and Ashley, 2011; Allen et al., 2015a; Brooks et al., 2003) but the two leading predictor variables – CAPE and SRH – are similar to the hail model by Allen et al. (2015a). It should also be mentioned that the hail algorithm's skill is strongly dependent on the quality of the ERA-Interim input data, which varies in time and between locations. To derive local point hail frequencies for an individual exposed asset it is necessary to transfer the presented large-hail frequencies.

A further caveat is that storm triggering mechanisms (e.g., orographic lifting, convergent flows) are not included in the algorithm, resulting in an underestimation of hail hazards in areas where these processes are important or an overestimation of hail risk in areas with strong inhibition, e.g., south-east Texas. However, even sophisticated convection-resolving models struggle with simulating this process (Gagne et al., 2017).

Future work will aim to assess the algorithm's sensitivity to different input data sets such as different reanalysis, radiosonde data, and climate models. A future development of the algorithm could include a probabilistic maximum hail size distribution estimate based on the relationships between the observed hail environments and observed maximum hail sizes. Furthermore, future studies will focus on trends in historic hail hazard assessments, the global impacts of large-scale climate variability such as the El Niño–Southern Oscillation, and potential changes in large hail hazards in future climates. In another direction, using satellite observations in combination with hail environments as in Punge et al. (2017) or Bedka et al. (2018) is promising to enhance our understanding of global hail hazards and to increase the spatiotemporal accuracy of the presented data.

Acknowledgments

The National Center for Atmospheric Research (NCAR) is funded by the US National Science Foundation (NSF) and this work was partially supported by the NSF grant 1048829 "Developing a Next-Generation Approach to Regional Climate Prediction at High Resolution". We thank the Storm Prediction Center (SPC), the Australian Bureau of Meteorology (BoM), and the European Severe Storm Laboratory (ESSL) for making available their data sets. Computer resources were provided by the Computational and Information Systems Laboratory (NCAR Community Computing; <http://n2t.net/ark:/85065/d7wd3xhc>). We thank Dr. Charles Knight, Dr. Andrew Heymsfield, Dr. Stanley Trier, Dr. Hugh Morrison, Dr. James Done and Mr. Bruce Buckley for their constructive comments.

ERA-Interim data can be derived from ECMWF's MARS archive (<http://apps.ecmwf.int/datasets/data/interim-full-daily/levtype=ml/>). Storm Prediction Center's Storm Events data can be downloaded from <https://www.ncdc.noaa.gov/stormevents/ftp.jsp>. The BOM Severe Storms Archive data can be derived from <http://www.bom.gov.au/australia/stormarchive/> and ESSL's European Severe Weather Database is located at <https://www.eswd.eu/>. The daily hail hazard estimates can be freely downloaded under <https://doi.pangaea.de/10.1594/PANGAEA.888881?format=html#download> and the hail hazard algorithm is available from the authors upon request.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.wace.2018.10.004>.

References

- Adams-Selin, R.D., Ziegler, C.L., 2016. Forecasting hail using a one-dimensional hail growth model within WRF. *Mon. Weather Rev.* 144, 4919–4939.
- Allen, J.T., Allen, E.R., 2016. A review of severe thunderstorms in Australia. *Atmos. Res.* 178, 347–366.
- Allen, J.T., Karoly, D.J., 2014. A climatology of Australian severe thunderstorm environments 1979–2011: inter-annual variability and ENSO influence. *Int. J. Climatol.* 34, 81–97.
- Allen, J.T., Karoly, D.J., Mills, G.A., 2011. A Severe Thunderstorm Climatology for Australia and Associated Thunderstorm Environments.
- Allen, J.T., Tippett, M.K., 2015. The characteristics of United States hail reports: 1955–2014. *E.J. Severe Storms Meteorol.* 10.
- Allen, J.T., Tippett, M.K., Sobel, A.H., 2015a. An empirical model relating US monthly hail occurrence to large-scale meteorological environment. *J. Adv. Model. Earth Syst.* 7, 226–243.
- Allen, J.T., Tippett, M.K., Sobel, A.H., 2015b. Influence of the El Niño/southern oscillation on tornado and hail frequency in the United States. *Nat. Geosci.* 8, 278.
- Barthel, F., Neumayer, E., 2012. A trend analysis of normalized insured damage from natural disasters. *Climatic Change* 113, 215–237.
- Bedka, K.M., Allen, J.T., Punge, H.J., Kunz, M., Simanovic, D., 2018. A long-term overshooting convective cloud-top detection database over Australia derived from MTSA Japanese advanced meteorological imager observations. *J. Appl. Meteorol. Climatol.* 57, 937–951.
- Blair, S.F., Laflin, J.M., Cavanaugh, D.E., Sanders, K.J., Currens, S.R., Pullin, J.I., Cooper, D.T., Deroche, D.R., Leighton, J.W., Fritchie, R.V., et al., 2017. High-resolution hail observations: implications for NWS warning operations. *Weather Forecast.* 32, 1101–1119.
- Blamey, R., Middleton, C., Lennard, C., Reason, C., 2017. A climatology of potential severe convective environments across South Africa. *Clim. Dynam.* 49, 2161–2178.
- Brimelow, J.C., Burrows, W.R., Hanesiak, J.M., 2017. The changing hail threat over North America in response to anthropogenic climate change. *Nat. Clim. Change* 7, 516–522.
- Brooks, H.E., 2009. Proximity soundings for severe convection for Europe and the United States from reanalysis data. *Atmos. Res.* 93, 546–553.
- Brooks, H.E., Lee, J.W., Craven, J.P., 2003. The spatial distribution of severe thunderstorm and tornado environments from global reanalysis data. *Atmos. Res.* 67, 73–94.
- Cecil, D.J., 2009. Passive microwave brightness temperatures as proxies for hailstorms. *J. Appl. Meteorol. Climatol.* 48, 1281–1286.
- Cecil, D.J., Blankenship, C.B., 2012. Toward a global climatology of severe hailstorms as estimated by satellite passive microwave imagers. *J. Clim.* 25, 687–703.
- Changnon, S.A., 2009. Increasing major hail losses in the US. *Climatic Change* 96, 161–166.
- Changnon, S.A., Changnon, D., 2000. Long-term fluctuations in hail incidences in the United States. *J. Clim.* 13, 658–664.
- Changnon, S.A., Piehl Jr., R.A., Changnon, D., Sylves, R.T., Pulwarty, R., 2000. Human factors explain the increased losses from weather and climate extremes. *Bull. Am. Meteorol. Soc.* 81, 437–442.
- Cintineo, J.L., Smith, T.M., Lakshmanan, V., Brooks, H.E., Ortega, K.L., 2012. An objective high-resolution hail climatology of the contiguous United States. *Weather Forecast.* 27, 1235–1248.
- Craven, J.P., Brooks, H.E., Hart, J.A., 2004. Baseline climatology of sounding derived parameters associated with deep, moist convection. *Natl. Weather Digest* 28, 13–24.
- Craven, J.P., Jewell, R.E., Brooks, H.E., 2002. Comparison between observed convective cloud-base heights and lifting condensation level for two different lifted parcels. *Weather Forecast.* 17, 885–890.
- Davies, J., Ponaldi, B., 1990. In: Test of Helicity as Tornado Forecast Parameter Preprint of the 16th Conference on Severe Local Storm. Kananaskis. Amer. Meteor. Soc, Alta, Canada, pp. 588–593.
- Dee, D.P., Uppala, S., Simmons, A., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M., Balsamo, G., Bauer, P., et al., 2011. The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Q. J. R. Meteorol. Soc.* 137, 553–597.
- Dennis, E.J., Kumjian, M.R., 2017. The impact of vertical wind shear on hail growth in simulated supercells. *J. Atmos. Sci.* 74, 641–663.
- Dessens, J., Berthet, C., Sanchez, J., 2015. Change in hailstone size distributions with an increase in the melting level height. *Atmos. Res.* 158, 245–253.
- Dotzek, N., Groenemeijer, P., Feuerstein, B., Holzer, A.M., 2009. Overview of ESSL's severe convective storms research using the European Severe Weather Database ESWD. *Atmos. Res.* 93, 575–586.
- Féral, L., Sauvageot, H., Soula, S., 2003. Hail detection using S-and C-band radar reflectivity difference. *J. Atmos. Ocean. Technol.* 20, 233–248.
- Gagne, D.J., McGovern, A., Haupt, S.E., Sobash, R.A., Williams, J.K., Xue, M., 2017. Storm-based probabilistic hail forecasting with machine learning applied to convection-allowing ensembles. *Weather Forecast.* 32, 1819–1840.
- Gensini, V.A., Ashley, W.S., 2011. Climatology of potentially severe convective environments from the North American regional reanalysis. *E.J. Severe Storms Meteorol.* 6.
- Groenemeijer, P., Púčik, T., Holzer, A.M., Antonescu, B., Riemann-Campe, K., Schultz, D.M., Kühne, T., Feuerstein, B., Brooks, H.E., Doswell III, C.A., et al., 2017. Severe convective storms in Europe: ten years of research and education at the European severe storms laboratory. *Bull. Am. Meteorol. Soc.* 98, 2641–2651.
- Gunturi, P., Tippett, M., 2017. Managing Severe Thunderstorm Risk: Impact of ENSO on US Tornado and Hail Frequencies. Technical Report Willis Re - Tech. Rep.
- Hand, W.H., Cappelluti, G., 2011. A global hail climatology using the UK Met Office convection diagnosis procedure (CDP) and model analyses. *Meteorol. Appl.* 18, 446–458.
- Houze, R.A., 2004. Mesoscale convective systems. *Rev. Geophys.* 42.
- Kain, J.S., Weiss, S.J., Bright, D.R., Baldwin, M.E., Levit, J.J., Carbin, G.W., Schwartz, C.S., Weisman, M.L., Drogemeyer, K.K., Weber, D.B., et al., 2008. Some practical considerations regarding horizontal resolution in the first generation of operational convection-allowing NWP. *Weather Forecast.* 23, 931–952.
- Ladwig, B., 2018. wrf-python Documentation. Technical Report Release 1.2.0.
- Le Roux, N., Olivier, J., 1996. Modelling hail frequency using a generalized additive-interaction technique. *S. Afr. Geogr. J.* 78, 7–12.
- Li, X., Zhang, Q., Zou, T., Lin, J., Kong, H., Ren, Z., 2018. Climatology of hail frequency and size in China, 1980–2015. *J. Appl. Meteorol. Climatol.* 57, 875–887.
- Mahoney, K., Alexander, M.A., Thompson, G., Barsugli, J.J., Scott, J.D., 2012. Changes in hail and flood risk in high-resolution simulations over Colorado's mountains. *Nat. Clim. Change* 2, 125.
- Martins, J.A., Brand, V.S., Capucim, M.N., Felix, R.R., Martins, L.D., Freitas, E.D., Gonçalves, F.L., Hallak, R., Dias, M.A.S., Cecil, D.J., 2017. Climatology of destructive hailstorms in Brazil. *Atmos. Res.* 184, 126–138.
- Mezher, R.N., Doyle, M., Barros, V., 2012. Climatology of hail in Argentina. *Atmos. Res.* 114, 70–82.
- Miller, P.W., Mote, T.L., 2017. Standardizing the definition of a pulse thunderstorm. *Bull. Am. Meteorol. Soc.* 98, 905–913.
- Mohr, S., Kunz, M., 2013. Recent trends and variabilities of convective parameters relevant for hail events in Germany and Europe. *Atmos. Res.* 123, 211–228.
- Mohr, S., Kunz, M., Keuler, K., 2015. Development and application of a logistic model to estimate the past and future hail potential in Germany. *J. Geophys. Res.: Atmosphere* 120, 3939–3956.
- Moller, A.R., Doswell III, C.A., Foster, M.P., Woodall, G.R., 1994. The operational recognition of supercell thunderstorm environments and storm structures. *Weather Forecast.* 9, 327–347.
- Mroz, K., Battaglia, A., Lang, T.J., Cecil, D.J., Tanelli, S., Tridon, F., 2017. Hail-detection algorithm for the GPM Core Observatory satellite sensors. *J. Appl. Meteorol. Climatol.* 56, 1939–1957.
- Munich, R.E., 2016. Natural Catastrophes 2015 Analyses, Assessments, Positions. 2016 Issue. TOPICS GEO.
- Newton, C.W., 1963. Dynamics of severe convective storms. In: Severe Local Storms. Springer, pp. 33–58.
- Ni, X., Liu, C., Cecil, D.J., Zhang, Q., 2017. On the detection of hail using satellite passive microwave radiometers and precipitation radar. *J. Appl. Meteorol. Climatol.* 56, 2693–2709.
- Punge, H., Bedka, K., Kunz, M., Reinbold, A., 2017. Hail frequency estimation across Europe based on a combination of overshooting top detections and the ERA-INTERIM reanalysis. *Atmos. Res.* 198, 34–43.
- Punge, H., Bedka, K., Kunz, M., Werner, A., 2014. A new physically based stochastic event catalog for hail in Europe. *Nat. Hazards* 73, 1625–1645.
- Punge, H.J., Kunz, M., 2016. Hail observations and hailstorm characteristics in Europe: a review. *Atmos. Res.* 176, 159–184.
- Rasmussen, E.N., Blanchard, D.O., 1998. A baseline climatology of sounding-derived supercell and tornado forecast parameters. *Weather Forecast.* 13, 1148–1164.
- Schaefer, J., Edwards, R., 1999. The SPC tornado/severe thunderstorm database. In: Preprints, 11th Conf. On Applied Climatology, vol. 6 Amer. Meteor. Soc., Dallas, TX.
- Tippett, M.K., Allen, J.T., Gensini, V.A., Brooks, H.E., 2015. Climate and hazardous convective weather. *Current Climate Change Reports* 1, 60–73.
- Weisman, M.L., Klemp, J.B., 1982. The dependence of numerically simulated convective storms on vertical wind shear and buoyancy. *Mon. Weather Rev.* 110, 504–520.
- Westermayer, A., Groenemeijer, P., Pistotnik, G., Sausen, R., Faust, E., 2017. Identification of favorable environments for thunderstorms in reanalysis data. *Meteorol. Z.* 26, 59–70.
- Wilks, D.S., 2011. Statistical Methods in the Atmospheric Sciences, vol. 100 Academic press.
- Witt, A., Eilts, M.D., Stumpf, G.J., Johnson, J., Mitchell, E.D.W., Thomas, K.W., 1998. An enhanced hail detection algorithm for the WSR-88D. *Weather Forecast.* 13, 286–303.
- Xie, B., Zhang, Q., Wang, Y., 2008. Trends in hail in China during 1960–2005. *Geophys. Res. Lett.* 35.