Long-term trends of hail-related weather types in an ensemble of regional climate models using a Bayesian approach

M.-L. Kapsch,[1,2] M. Kunz,[1] R. Vitolo,[3] and T. Economou[3]

Received 18 November 2011; revised 20 June 2012; accepted 25 June 2012; published 9 August 2012.

[1] This paper investigates the long-term variability of specific weather types that are associated with damaging hailstorms in Germany for past (1971–2000) and future (2011–2050) time periods. Forty large-scale weather types are determined by the objective weather type classification scheme of German Weather Service. This scheme is applied to both reanalyses (ERA-40) and eight different regional climate model (RCM) simulations. It is shown that the RCMs are able to approximately reproduce the distribution of weather type occurrences obtained from the reference of ERA-40. Using additional insurance loss data, the weather types are further identified as hail-related or hail-unrelated. Hailstorms are neither captured comprehensively by existing observation systems nor can they be modeled reliably and the large-scale weather types are here considered as proxies for hail occurrence. Four weather types that are most likely associated with damaging hailstorms show a slight increase both during the past and future period according to the RCM simulations. A novel statistical model is developed for the probabilistic prediction of the fraction of hail damage days conditional on the weather types. The model is Bayesian and uses a Markov Chain Monte Carlo approach. For the ERA-40 reanalysis the model prediction agrees well with fraction of hail damage days observed in the insurance data. For most of the RCM projections, the statistical model predicts a slight increase in the number of hail days in the future (2031–2045), with relative changes between 7 and 15% compared to the period 1971–2000.

Citation: Kapsch, M.-L., M. Kunz, R. Vitolo, and T. Economou (2012), Long-term trends of hail-related weather types in an ensemble of regional climate models using a Bayesian approach, *J. Geophys. Res.*, *117*, D15107, doi:10.1029/2011JD017185.

1. Introduction

[2] Severe hailstorms frequently cause significant damage to buildings, crops, or automobiles in several regions of Central Europe, such as southern Germany, Austria, or northern Italy. In the federal state of Baden-Württemberg (southwest Germany), for example, almost 40% of all damage to buildings by natural hazards can be attributed to large hail [*Kunz and Puskeiler*, 2010]. Hailstorms usually have a limited spatial extent. From a survey of various damage-related hailstreaks in the US, *Changnon* [1970] found that 80% had areas of less than 40 km^2. In case of strong upper tropospheric forcing, the hailstreaks associated with long-traveling storms may also expand over a meso-scale region with typical areas of more than 500 km^2 [*Dessens*, 1986].

[3] In light of global warming [*IPCC*, 2007], it is still a matter of debate whether the number and/or severity of hailstorms has increased over the last decades and is expected to increase in the future. Due to both their local-scale extent and insufficient direct monitoring systems, hail swaths are not captured accurately and uniquely over a long period of time. Remote-sensing systems such as radars are able to detect convection signals in a basic way, but they lack the ability to discern a clear relation between measured intensity and rain or hail on the ground [*Sauvageot*, 1992]. On the other hand, although operational weather forecast or regional climate models are accessible for a sufficiently long time period, their temporal and spatial resolutions are too coarse to explicitly and reliably simulate deep convection [*Noppel et al.*, 2010]. This lack of appropriate data sets hampers the study of the temporal variability of hailstorm occurrence and the finding of observational evidence for any long-term changes.

[4] Damage data from insurance companies is the only available source of information displaying the direct impact of hail on vulnerable structures. Based on loss data of an agricultural insurance company in Switzerland (1920–1999), *Schiesser* [2003] found high temporal variability in the annual number of severe hailstorms, with an increase between 1980 and 1994. This finding is more or less confirmed by *Kunz et al.* [2009], who report a significant increase in the number of

[1]Institute for Meteorology and Climate Research, Karlsruhe Institute of Technology, Karlsruhe, Germany.
[2]Now at Department of Meteorology, Stockholm University, Stockholm, Sweden.
[3]College of Engineering, Mathematics and Physical Sciences, University of Exeter, Exeter, UK.

Corresponding author: M.-L. Kapsch, MISU, Stockholm University, Svante Arrhenius vg 16C, Stockholm SE-10691, Sweden. (marie@misu.su.se)

0148-0227/12/2011JD017185

hail days over the last two decades by analyzing insured damage to buildings in southwest Germany. Over Ontario, Canada, *Cao* [2008] identifies a robust ever-increasing trend in the frequency of severe hail events over the last decades. However, insurance loss data strongly depends on features, such as population density and/or vulnerability of the insured assets, which may substantially change with time. Therefore, a robust trend in the frequency of convective storms cannot be inferred from the positive trends found in the insurance data.

[5] A way of overcoming the lack of direct hail observations is to connect thunderstorm occurrence to meteorological parameters that are measured or modeled consistently in space and time. If a relationship exists between appropriate parameters and severe weather occurrence, it can be expected that changes in the parameter values also impact the intensity and/or number of thunderstorms. Although the exact mechanisms for hail formation within a cumulus cloud are very complex and not yet fully understood, the prerequisites for severe convective storms are well known: atmospheric convective instability, low-level moisture, vertical wind shear, and an additional trigger mechanism [e.g., *Doswell*, 1987; *Emanuel*, 1994]. Whereas the first two requirements describe the energy content of the prevailing environment, the third mechanism is relevant for determining the kind of convective organization, and the fourth for the release of atmospheric instability.

[6] The convective potential of the atmosphere can be expressed by various convective parameters and indices. Several studies established a relation between thunderstorm probability of occurrence and specific convective parameters and their thresholds [e.g., *Haklander and van Delden*, 2003; *Kunz*, 2007]. This connection provides the basis for statistical analyses of long-term variability including trends in the convective parameters over past [e.g., *DeRubertis*, 2006; *Cao*, 2008] and future decades [*Leslie et al.*, 2007; *Van Klooster and Roebber*, 2009; *Marsh et al.*, 2009; *Hanafin et al.*, 2011]. In summary, most studies found evidence for a slight increase in the convective potential of the atmosphere both for past and future.

[7] Severe convective storms may also be attributed to specific weather situations, namely, those favoring prerequisites for the storms: prevailing moist, unstable air masses and differential positive vorticity advection. These conditions are often associated with a typical spatial arrangement of the most relevant synoptic systems, for example a weak ridge over the investigation area and a pronounced trough upstream. Several approaches exist to classify either the general circulation or the prevailing synoptic systems from meteorological parameters at different levels [*Yarnal et al.*, 2001]. Examples include the large-scale circulation and weather patterns according to *Hess and Brezowsky* [1977], or the Objective Weather Type Classification (OWTC) [*Bissolli and Dittmann*, 2001] of the German Weather Service (Deutscher Wetterdienst (DWD)). While the former rely on manual analyses of the locations of prevailing main pressure and frontal systems, the latter combines three meteorological parameters that can be computed from weather forecast/analyses or climate models. *Bissolli et al.* [2007], for example, attributed the majority of tornadoes in Germany to three specific OWTC weather types. By comparing the two classification schemes, *Ehmann* [2009] found a higher skill for OWTC to estimate hail damage days.

[8] Built upon these studies, this paper investigates the relation between significant hail damage days and large-scale weather types (WTs) determined from OWTC, also examining the long-term variability of those WTs which are mostly associated to hail occurrence. In this approach, atmospheric parameters relevant for the occurrence of severe thunderstorms (see above), are only implicitly considered by the parameters defining the different WTs. The investigation area is Germany with a special focus on a meso-scale region in the southwest, where detailed insurance loss data are available. The overall goals are:

[9] 1. to evaluate to what extent specific WTs can be related to damage causing hailstorms;

[10] 2. to evaluate whether Regional Climate Model (RCM) simulations are able to reliably reproduce the statistics of the hail-relevant synoptic systems in terms of WT distribution;

[11] 3. to develop a model for the probabilistic prediction of the number of hail damage days using all WTs, and

[12] 4. to study both long-term variability and changes in the occurrence of those WTs mostly associated with significant hail.

[13] The weather classification is applied to both reanalysis (ERA-40) and climate model data. For reasons of consistency and to obtain higher spatial variance of the input fields, all data were downscaled with the Consortium for Small-Scale Modeling (COSMO) model in climate mode (hereinafter referred to as CCLM). Eight different RCM simulations for the period 1971–2050 are used to account for model (aleatory) and scenario (epistemic) uncertainty. These data sets include two different CCLM versions, two driving Atmosphere–ocean General Circulation Models (AOGCM), three realizations in terms of pre-industrial initial conditions and two emissions scenarios (A1B and B1).

[14] In the first step, WTs derived from the RCMs are evaluated against reanalysis and insurance data. With this step, we assess the skill of the RCMs to reliably reproduce the observed distributions of the WTs. Furthermore, we identify a few specific weather patterns that are either frequently or never associated with severe hailstorms. Two statistical tools are applied to investigate the long-term variability and changes in the number of hail-related WTs. One is the classical linear trend analysis including the rank-based Mann-Kendall significance test. The other tool, which represents a central point in this paper, is a statistical model for the probabilistic prediction of the number of hail days based on all WTs that occurred within a certain time period. Such a model has never before been applied to estimate hail damage days from time series of synoptic-scale weather patterns. Application of this model to different RCM simulations allows us to estimate changes in the number of hail damage days in future decades including quantification of uncertainty in such predictions.

[15] The article is organized as follows. Section 2 describes the different data sets considered in this study: reanalyses, climate simulations, and insurance data. In the next section, we outline the methods of the OWTC scheme and describe in detail the probability model used to predict the number of hail days. Section 4 evaluates the WTs derived from reanalyses and climate model data and examines long-term variability of

Table 1. Regional Climate Models Assessed in This Study

Model (Version)	Global Forcing	Horizontal Resolution	Emission Scenario	Number of Model Runs	Referred to as
COSMO-CLM (4.8)	ECHAM5	0.44°	A1B	3	CCLM-ECHAM5
COSMO-CLM (4.8)	CGCM3	0.44°	A1B	1	CCLM-CGCM3
COSMO-CLM (3.1)	ECHAM5	0.167°	A1B	2	CCLM-C
COSMO-CLM (3.1)	ECHAM5	0.167°	B1	2	CCLM-C

specific WTs that were found to favor hailstorm development. Section 5 estimates changes in the annual number of hail damage days for future decades by applying the probability model to the RCM simulations. Finally, section 6 presents our discussion with some concluding remarks.

2. Data Sets

[16] Large-scale weather types (WTs) according to the OWTC scheme are computed from climate model simulations. As the OWTC requires meteorological parameters at different pressure levels, the number of appropriate and available RCM realizations is limited. To identify both hail-related and hail-unrelated WTs, insurance loss data are used additionally. All data sets will be briefly described in the following.

2.1. Regional Climate Model Data

[17] The non-hydrostatic numerical model CCLM is based on the fully compressible Navier–Stokes equations using finite differences [*Rockel et al.*, 2008]. The horizontal grid is a rotated latitude/longitude Arakawa C-grid with a generalized terrain-following vertical coordinate. An overview of the eight different CCLM experiments considered in this study is shown in Table 1.

[18] The CCLM runs are initialized and driven by two different AOGCMs. One is the ECHAM5/MPI-OM, developed and operated at the Max-Planck Institute for Meteorology (MPI-M), Hamburg, Germany [*Roeckner et al.*, 2003, 2006]. The initial conditions of the three realizations of ECHAM5 are different states of the 500-yr pre-industrial control integration based on constant greenhouse gas concentrations. The other AOGCM is the third version of the Canadian Centre for Climate Modeling and Analysis (CCCma) Coupled Global Climate Model (CGCM3). Both AOGCMs have a T63 spectral resolution, corresponding to 1.87° (≈200 km).

[19] Simulation results of two different CCLM versions are considered in this study. The older version 3.1 was used by MPI-M to calculate the so-called CCLM-consortium runs (hereinafter referred to as CCLM-C) [*Lautenschlager et al.*, 2009a, 2009b].These runs were initialized and driven by the first two realizations of ECHAM5 for the IPCC emissions scenarios (SRES) A1B and B1 [*IPCC*, 2007]. The horizontal resolution of the data is 0.167° (≈18 km).

[20] Furthermore, the Institute for Meteorology and Climate Research (IMK-TRO) at the Karlsruhe Institute of Technology (KIT) performed several climate runs with CCLM version 4.8 for the scenario A1B. These simulations were forced by all three realizations of ECHAM5 (referred to as CCLM-ECHAM5) and by CGCM3 (referred to as CCLM-CGCM3). The spatial resolution of the first nesting step of the two considered in this study is 0.44° (≈50 km).

[21] While CCLM-C and CCLM-CGCM3 runs are available for the complete investigation period from 1971 to 2050, CCLM-ECHAM5 runs exhibit a gap between 2001 and 2010. Therefore, the long-term variability of hail-related WTs is analyzed separately over the two periods.

2.2. Reanalysis Data

[22] For evaluation purposes, we used ERA-40 reanalyses released by the European Centre for Medium-Range Weather Forecast (ECMWF) [*Uppala et al.*, 2005]. The reanalyses are represented as spherical harmonics at a TL159 spectral resolution or as Gaussian data on a quasi-regular N80 linear Gaussian grid (≈1.125°). The ERA-40 project covers the period from mid-1957 to 2001. Additionally, ERA-40 was dynamically downscaled with CCLM 4.8 by IMK-TRO for the period from 1971 to 2000. The simulation results of the first nesting step with a resolution of 0.44° are considered as a reference (hereinafter referred to as CCLM-ERA40).

2.3. Insurance Loss Data

[23] Hail damage days over the federal state of Baden-Württemberg (southwest Germany) are determined from building damage data provided by the insurance company *SV SparkassenVersicherung AG* (hereinafter referred to as SV) for the period from 1986 to 2008 (C20). In Baden-Württemberg, building insurance against elementary events was mandatory until 1994 and exclusively offered by one company. After the abolition of the insurance obligation, the market share of the SV portfolio decreased from 100% (1994) to approximately 70% (2010). The high proportion of settlement areas (12.1%) in Baden-Württemberg, extensively spread over the area, ensures that almost all severe hailstorms are captured by the SV data. Resolved in 5-digit postal code zones, the data include the number of claims and amount of reimbursement for each hail damage day.

[24] Direct application of insurance loss data for meteorological investigations is constrained by their dependence on settlement density, by changes of the portfolio, and by construction related changes in the building's vulnerability [*Egli*, 2007]. To compensate for those effects, we considered only hail damage days defined by a threshold of ten claims that were settled on a day according to the study of *Kunz et al.* [2009]. Changes in the portfolio were considered by multiplying the number of claims on a specific day with a correction factor which is given by the ratio of the average number of contracts between 1986 and 2008 and the number of contracts in the specific year. Application of these methods yields a total of 223 hail damage days during C20, that is 14.8 days per year on average.

3. Methods

[25] The investigation area extends from 44.33° to 57.67°N and from 0.00° to 18.67°E, covering Germany and

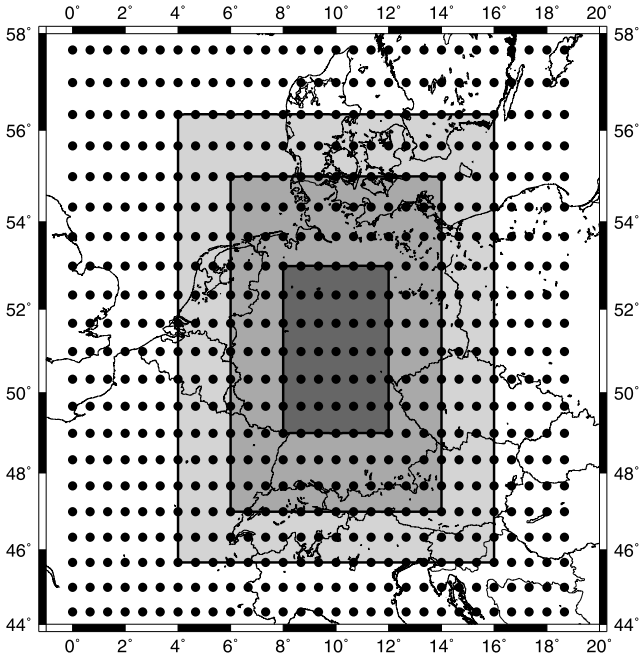

Figure 1. Investigation area with interpolated grid used for the objective weather type classification. Within the central frame (dark grey), the grid points are weighted three times, within the surrounding frame it is two times (medium grey), and near the boundary (light grey) it is unity.

some adjacent countries (see Figure 1). All model data were re-rotated and interpolated on a uniform, cylindric equidistant grid with a resolution of 0.66° × 0.66° (≈50 km). Even though insurance loss data are only available for Baden-Württemberg, we decided to center the investigation area over Germany to avoid bias of the large-scale weather type classification (see section 3.1) due to elevated terrain. Furthermore, Baden-Württemberg is a significant hot spot in hailstorm probability. Because damage-producing hailstorms in Germany occur almost exclusively in the summer half-year with a peak maximum during the second half of the day [*Kunz and Puskeiler*, 2010], we considered only model data at 1200 UTC for April to September.

3.1. Objective Weather Type Classification

[26] The OWTC method differentiates among 40 WTs by evaluating three integral model parameters at different levels in a dichotomous scheme. The classification scheme includes:

[27] 1. Mean flow direction *AA* in 700 hPa;

[28] 2. Cyclonality *CY* in the lower and middle troposphere;

[29] 3. Humidity index *HI* as precipitable water.

[30] A weighting scheme is applied to emphasize grid points that are within the border of Germany, but to disregard those that are located near or outside the borders of the investigation area. As shown in Figure 1, grid points in the center of the domain are weighted three times, whereas those of the surrounding frame are double weighted [see *Bissolli and Dittmann*, 2001].

[31] The mean flow direction *AA* is calculated by splitting the wind directions in 700 hPa into four sectors: ([0°–90°), [90°–180°), [180°–270°), and [270°–360°)). If at least two-thirds of the weighted grid points falls into one of these sectors, the corresponding direction is selected. Thus, only four wind directions are possible: northeast (NE), southeast (SE), southwest (SW) or northwest (NW). In cases that less than two-thirds of the weighted grid points exhibit the same direction, the flow remains unclassified (XX). Comparisons of wind maps for single days with the results obtained by this method revealed that application of the scheme gives more realistic results compared to the original DWD scheme, where the 90°-sectors successively increase by 10°, that is 36 × 90° (see *Bissolli and Dittmann* [2001] for further details).

[32] On the cylindric equidistant grid, cyclonality *CY* related to the geostrophic vorticity ζ_g in the *p*-system (pressure vertical coordinate) is computed by

$$CY = f \cdot \zeta_g = \nabla^2 \phi = \frac{\partial^2 \phi}{\partial x^2} + \frac{\partial^2 \phi}{\partial y^2} \approx \frac{\phi(i+1,j) + \phi(i-1,j) - 2\phi(i,j)}{(\Delta x)^2} + \frac{\phi(i,j+1) + \phi(i,j-1) - 2\phi(i,j)}{(\Delta y)^2}, \quad (1)$$

where *f* is the Coriolis parameter, $\phi(i, j)$ is the geopotential at a grid point in zonal (i) and meridional (j) direction and Δx and Δy are the geometrical distance between the grid points. At the boundaries of the investigation area, only values of the nearest available grid points are used. In the original OWTC routine, *CY* is calculated on the 950 and 500 hPa pressure levels. Because the former is not available in most of the RCM data sets, we used the 1000 hPa level instead. It should be noted that differences in *CY* between 950 and 1000 hPa are only marginal. Finally, *CY* is integrated over the whole domain by applying the weighting routine and separated into two classes: cyclonality/positive vorticity ($CY \geq 0$, denoted as *C*) and anticyclonality/negative vorticity ($CY < 0$, denoted as *A*).

[33] The last parameter in the scheme, the humidity index *HI*, is computed from the weighted areal mean of the vertically integrated water vapor content, referred to as precipitable water:

$$PW = -\frac{1}{g} \int_{p_1}^{p_2} r \; dp, \quad (2)$$

where p_1 and p_2 are 1000 and 500 hPa, respectively. Since *PW* has a distinct annual cycle like the temperature, it does not directly enter to determine *HI.* Rather the actual value *PW′* on a specific day is compared with the climatological 18-yr daily mean $\overline{PW}$, which is calculated separately for each RCM run. If $PW' > \overline{PW}$, the atmosphere is denoted as wet (*HI* = W) on that day. Otherwise it is denoted as dry (*HI* = D).

[34] In summary, the WTs in the OWTC scheme are defined by a 4-digit identifier:

$$AACY_{1000\text{hPa}} \; CY_{500\text{hPa}} HI.$$

[35] An example of the classification is the pattern SWCAW, which is one of the most important hail damage related WT. It indicates southwesterly mean flow direction, cyclonic in 1000, but anticyclonic in 500 hPa, and higher moisture content compared to the 18-yr mean.

3.2. Linear Trend Analysis

[36] In the first step, changes in the time series of hail-related (and unrelated) WTs are inferred through a linear regression. To test the statistical significance of the monotonic trends, the non-parametric, rank-based Mann-Kendall (MK) [*Mann*, 1945; *Kendall*, 1975] test is applied to the time series of the hail-related WTs. As shown, for example, by *Yue et al.* [2002], positive serial correlation in the time series increases the variance of the MK statistic, which, in turn, increases the probability for detecting a significant trend. That applies in particular for small sample sizes and small trends. To account for auto-correlation, the time series are first prewhitened with the trend-free pre-whitening (TFPW) approach according to *Yue et al.* [2002]. This approach yields a blended time series including the original trend, but without any autocorrelation, where the MK test is applied afterwards.

3.3. Probabilistic Forecast of Hail Damage Days

[37] Consider first the C20 period (1986–2000). Since we restrict the analysis to the period from April to September (183 days), this gives a total number of $N = 2745$ of days. We construct a time series of $N = 2745$ ones or zeroes, according to whether hail occurred or not on each day in the SV data set (see section 2.3 for the threshold criterion). The total number of hail days in the observational period is $h = 223$. Denote as $\{WT_i, i = 1, \ldots, 40\}$ the set of 40 WTs, by N_i the number of days in C20 where weather type WT_i occurred and by x_i the number of days, out of the N_i with weather type WT_i, where hail occurred. We then have $\sum N_i = N = 2745$ and $\sum x_i = 223$ and the numbers

$$f_i = x_i/N_i, \quad f = \sum_i x_i/N \tag{3}$$

are the empirical frequency f_i of hail occurrence for weather type WT_i and the overall fraction f of hail damage days.

[38] Given an RCM simulation, one could predict the future occurrence of hail days in the following simple way: denote by N_i^{model} the number of days where weather type WT_i occurs in the climate simulation. Then one could define

$$X_i^{model} = f_i N_i^{model}, \tag{4}$$

essentially using the inverse of equation (3) to predict the number X_i^{model} of hail days for weather type WT_i in the climate simulation. The problem is that this approach would not provide estimates of the uncertainty in the prediction. To account for uncertainty, we construct a statistical model for the probability p_i of hail occurrence for weather type WT_i according to the following specification:

$$X_i \sim Bin(p_i, N_i), \quad i = 1, \ldots, 38, \tag{5}$$

where we stop at 38 since 2 of the 40 WTs have never occurred in the data. Equation (5) states that the number X_i of hail days for weather type WT_i is modeled as a random variable having a binomial distribution, denoted by $Bin(p_i, N_i)$: The probability that the number X_i of hail days out of N_i days is an integer x is therefore

$$P[X_i = x] = \binom{N_i}{x} p_i^x (1 - p_i)^{N_i - x}. \tag{6}$$

This is the probability distribution of the number of "successes" in a sequence of N_i independent experiments, each of which yields "success" (1, corresponding to hail) with the probability p_i or "failure" (0, corresponding to no hail) with probability $1 - p_i$.

[39] This model's appealing feature is the direct formulation in terms of the quantity of interest: The probability p_i of hail occurrence for weather type WT_i. However, direct estimation of this model is problematic. Indeed, the model contains the same number of data points (namely, pairs X_i, N_i) and unknown parameters (namely, p_i). This is called a *saturated* model in the statistical literature. Although such a model will describe the data perfectly, it lacks the ability of inference and prediction. To overcome this problem, we adopt a Bayesian hierarchical model [*Gelman et al.*, 2004] where the unknown parameters p_i are themselves modeled stochastically:

$$X_i \,|\, p_i \sim Bin(p_i, N_i), \quad i = 1, \ldots, 38, \tag{7}$$

$$p_i \sim Beta(\alpha, \beta), \quad i = 1, \ldots, 38, \tag{8}$$

$$\alpha, \beta \sim Gamma(0.5,\ 0.005). \tag{9}$$

Conditional on the p_i, the random variables X_i are binomially distributed. The effective number of model parameters is now much less than in model (5) since the variability of the probabilities p_i are themselves modeled by a common Beta distribution with only two unknown parameters α and β, thus pooling information across the p_i while reducing model dimensionality. Furthermore, α and β are themselves random variables (hyperpriors) which follow a gamma distribution with large mean and variance (uninformative prior).

[40] An appealing property of this model is that the probability of hail for the two unobserved WTs can be predicted by the model from the other p_i, $i = 1, \ldots, 38$. Such predictions will include pooled information from the other 38 WTs, which will be reflected in both the estimates and the uncertainty for (p_{39}, p_{40}).

[41] The model is estimated using Markov Chain Monte Carlo [*Gelman et al.*, 2004] and specifically with the software WinBUGS [*Spiegelhalter et al.*, 2003]. The output of the software is a sample of 3000 values for each random quantity which includes the probabilities p_i, resulting in a matrix p_i^k with $i = 40$ columns and $k = 3000$ rows. Each column contains a sample

$$\left\{p_i^k, \quad k = 1, \ldots, 3000\right\} \tag{10}$$

representing the posterior probability distribution for the p_i. This is one way of representing the epistemic uncertainty in the estimation of p_i. Details on the estimation are provided in Appendix A.

4. Climatology of Weather Types

4.1. Weather Types Derived From Reanalysis

[42] To identify WTs that are frequently (or very rarely) associated with damaging hail during C20, the OWTC scheme (see section 3.1) is applied to CCLM-ERA40 reanalysis and the results are evaluated using SV data. The distribution of WTs quantified from CCLM-ERA40 is close

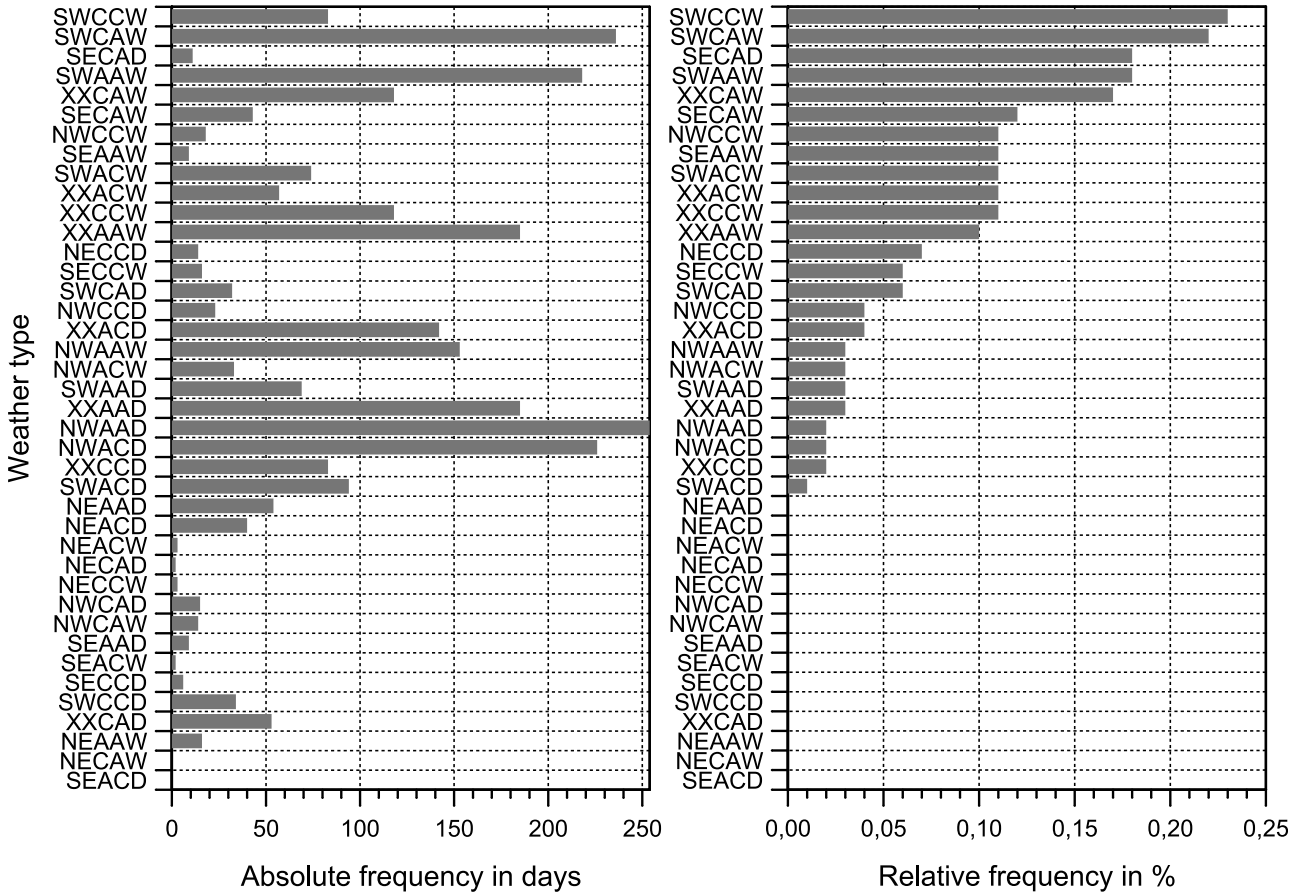

Figure 2. (left) Number of days each WT occurred and (right) relative frequency of WTs on hail damage days according to CCLM-ERA40 for the time period 1986 to 2000.

to the results of DWD and to those derived from the original ERA-40 reanalysis (not shown). However, slight discrepancies on single days can be ascribed to the sensitivity of the WT parameters (*AA*, *CY*, *HI*) matching the dichotomous scheme. For example, in cases where the vorticity (cyclonality) is close to zero or the humidity close to the mean value slight variations in the model simulation fields can lead to a different parameter value. In such cases, the spatial resolution of the simulations may modify the results of the OWTC.

[43] During the whole period of C20 (1986–2000), where both SV and CCLM-ERA40 data are available, nearly all 40 possible WTs occurred (Figure 2, left). The highest frequency is found for NWAAD prevailing on more than 250 days. By contrast, the two patterns NECAW and SEACD never occurred (recall that we consider only the summer half-year). Furthermore, several days remained unclassified in the wind direction indicated by *XX*, revealing a potential weakness of the classification scheme.

[44] According to the relative fraction of WTs on hail damage days (Figure 2, right), five of the 40 can be identified as the most hail-related WTs (hereinafter referred to as HWT). With a frequency of more than 15%, the types SWCCW, SWCAW, SECAD, SWAAW and XXCAW are notably related to hail occurrence. For example, the SWAAW pattern prevailed on 218 days, from which 18% were hail damage days. We find a distribution of hail damage days that is very similar to the distribution of tornado days over WTs found by *Bissolli et al.* [2007]. This is not surprising since atmospheric conditions which are favorable for tornado formation are similar to a certain extent to those conducive to significant hail. *Brooks et al.* [2003], for example, showed that severe thunderstorms in the United States are in general associated with high convective available potential energy (CAPE), high lapse rate and high shear. They determined a discrimination between tornadic environments and significant but non-tornadic environments as a function of shear over the lowest 1 km of the atmosphere, height of the lifted condensation level and station (grid point) elevation.

[45] Three of the five HWTs feature SW flow, while one exhibit an indifferent type which may also include SW. This direction implies lower tropospheric warm air advection from the Mediterranean or, occasionally, from the Atlantic that creates potential instability. Due to the origin, the air masses usually have a high moisture content as indicated by the humidity index which is "wet" (*W*) on four of the five HWTs. Higher temperature in combination with higher moisture at low levels increase the probability for thunderstorm occurrence. Three of the five patterns are characterized by cyclonic flow at 1000 hPa and anticyclonic flow at 500 hPa. However, the frequent occurrence of the patterns

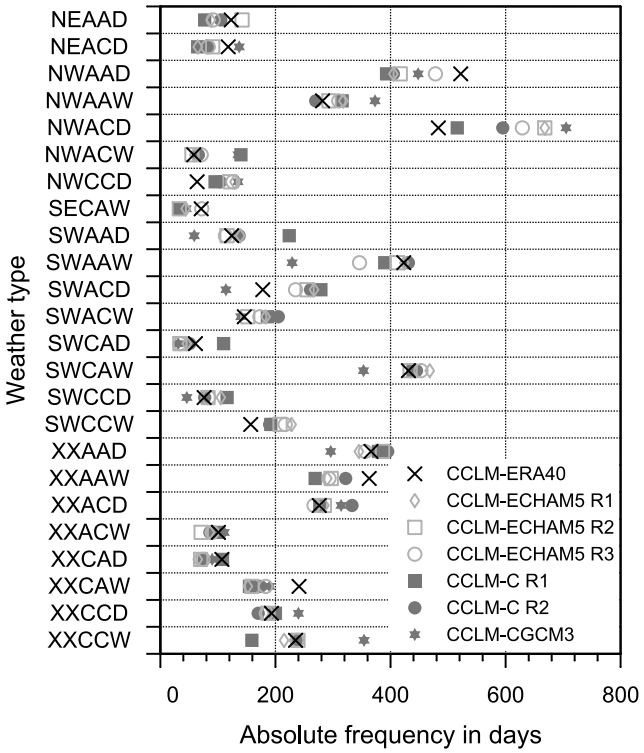

Figure 3. Absolute frequency of weather types derived from CCLM-ERA40 (cross) and climate model realizations (Table 1) during the hailstorm season 1971 to 2000. Only weather types that prevailed more than 50 times according to CCLM-ERA40 are shown.

SWAAW and SWCCW on hail damage days suggests that cyclonality may not be as important as mean flow direction and humidity - at least in the classification scheme.

[46] During the SWCAW, SWCCW and XXCAW patterns, an eastward moving upper air ridge is followed by a low pressure system from the west. Large-scale upward motion caused by differential vorticity advection and/or warm air advection can reduce convective inhibition, which may lead to an increase in the probability that deep convection is triggered by meso-scale processes. However, this constellation is not the driver for the SWAAW and SECAD patterns. Note, however, that the latter prevailed on 11 days only, considerably restricting the significance of the result. In the same way as for HWTs, we identified specific WTs that are not or very infrequently accompanied by hail, namely NWAAW, XXAAD, NWAAD and NWACD (hereinafter referred to as NWT). These four patterns prevailed on more than 150 days during C20, but only 3% are associated with hail damage. This result is physically plausible since northwesterly flow directions and dry air masses prevent deep convection to occur.

4.2. Weather Types Derived From Regional Climate Model Simulations

[47] A central point in the discussion of the temporal variability of WTs is to what extent the RCMs are able to reliably simulate the dominant synoptic systems. Because a RCM run cannot be expected to keep track with the real atmospheric state on a daily or annual basis, the different RCMs are evaluated by examining the statistics of all WTs between 1971 and 2000 compared to CCLM-ERA40 results. In this evaluation, only WTs that occurred on more than 50 days (i.e., 1.7 days p.a.) in the reanalyses are considered, in total 26 WTs.

[48] In general, all RCMs produce reliable results in the absolute number of the various WTs, which are more or less close to the reference CCLM-ERA40 (Figure 3). In particular the large variability in the frequency of specific WTs is reproduced well by all climate simulations. Single patterns such as SWACW, SWCAW, or NWAAW show a remarkable accordance between most of the simulations. Larger deviations are evident only for infrequently occurring WTs such as NWCCD and NWACW, which are characterized by discrepancies of up to 140 days. On average, the deviations amount values between 20 $\pm$ 17 days for CCLM-C R2 and 35 $\pm$ 30 days for CCLM-CGCM3, where the second value denotes the standard deviation in days. In general, CCLM-CGCM3 mainly underestimates the advection types of SE and SW. The statistical distributions of all meteorological parameters considered in the OWTC reveal that the vorticity in 500 hPa is higher in all RCM realizations compared to the reference. Consequently, less WTs with cyclonic flow conditions prevail in the different realizations. Therefore, the number of WTs such as NWACD, NWCCD, and SWCCW are overestimated, while WTs with anticyclonic flow conditions in 500 hPa are underestimated (e.g., NWAAD) by the climate simulations.

[49] Comparing all RCM results, the largest deviations in the number of WTs is found for CCLM-CGCM3 - at least in the mean. Any systematic deviation for the other simulations cannot be detected - neither with regards to the CCLM version nor to the realization of ECHAM5 (Run 1–3). This suggests that WT occurrence is determined primarily by the AOGCM and to a lesser extent by the RCM. This result is plausible since the OWTC classification scheme relies on meteorological fields on the synoptic scale, while dynamical downscaling using an RCM add only regional-scale variability (see, e.g., the review of *Feser et al.* [2011]).

[50] In view of the fact that the RCMs do not know the real state of the atmosphere when driven by an AOGCM, the good agreement of the RCM results with the reanalyses is somewhat surprising. It also documents the robustness of the OWTC scheme in general and its applicability for further investigation such as hail-related patterns.

4.3. Temporal Variability of Specific Weather Types

[51] The different results obtained from the six RCM runs (cf. Figure 3) illustrate the necessity to consider not only one but an ensemble of different models when exploring temporal variability and trends in the occurrence of the WTs. In our miniensemble approach, each RCM realization (scenario) is treated equally without any weighting. Note that the ensemble comprises the six RCM runs used in Figure 3 for C20 and eight RCM realizations for the projection period, as both CCLM-C runs are available for two emission scenarios.

[52] Within the two periods, from 1971 to 2000 and from 2011 to 2048, both the ensemble mean and the standard deviation of the annual number of HWT (without the rare SECAD) show a high temporal variability (Figure 4). In the

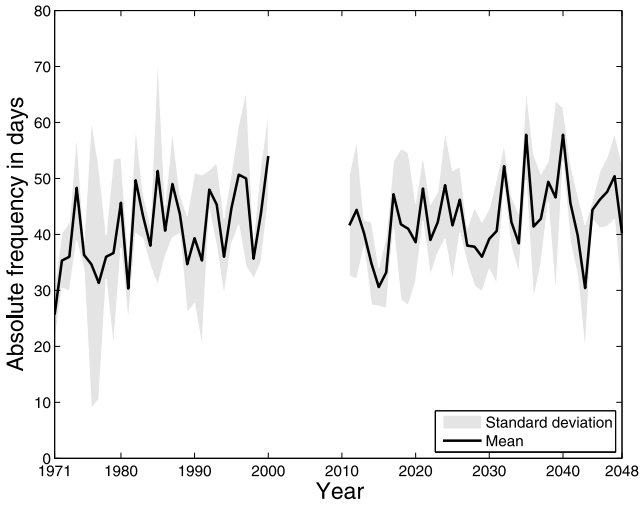

Figure 4. Frequency of HWT for C20 and the future projection period derived from all realizations of CCLM-ECHAM5, CCLM-C, CCLM-CGCM3 with mean values (black) and standard deviation (gray).

first period, the ensemble mean is between 26 and 54 days, whereas it fluctuates between 30 and 58 days in the second period. Large standard deviation indicates the differences of the RCMs to reproduce the HWTs as discussed in the previous section. By examining the parameters considered in the OWTC scheme, we found that the high standard deviation in the late 1990s is caused in particular by unrealistic high PW values in the CCLM-C R1 run. This issue will be discussed in detail in the next section. Compared to the first period, the standard deviation of the second (projection) period is generally lower. This seems to be counterintuitive as uncertainties caused by the underlying emission scenario are expected to cause additional variability between the simulations. However, as shown by *Hawkins and Sutton* [2009], uncertainty in regional climate predictions for the next decades is dominated by model uncertainty in the AOGCM/RCM hierarchy, whereas emission scenario uncertainty increases only slowly with lead time. Therefore, the latter uncertainty becomes important only for longer lead times and not for the time horizon considered here.

[53] The trend in the number of HWTs computed by linear regression from the ensemble mean is approximately 0.5 days per year during both periods. To further examine the robustness and the statistical significance of the trends, we split the whole time series into various shorter sub-time series. Both the start and end year are successively altered with a minimum duration of 10 years. For each sub-time series, linear monotonic trends and statistical significance are determined by applying the rank based non-parametric Mann-Kendall test (see section 3.2). To account for autocorrelation, the time series were prewhitened prior to the application of the statistical tests. For further details of the Mann-Kendall test and the trend-free pre-whitening (TFPW) approach it is referred to the description of *Yue et al.* [2002].

[54] The results of the trend estimates are visualized in terms of trend matrices, where the x-axis indicate the start and the y-axis the end of the various sub-time series (Figure 5). The shading shows the trend (number of HWT p.a.) estimated for each sub-time series. Note that only the ensemble mean and not the standard deviation is considered here. Changes with a statistical significance of less than 90% are plotted as white boxes. In all cases, the (significant) trends are positive, except for some series in the middle of the projection period. Significant positive trends are evident for the two entire time series. Accordingly, the number of HWT increased for about 12 days in the past and about 7 days in the projection period.

[55] Almost all sub-time series starting in the 1970s exhibit positive trends, in particular those from the mid-1970s to the mid-1980s, where the trend also exhibits a maximum. After 1980, all changes (positive as well as negative) are insignificant. During the projection period, mainly longer series of more than 20 years exhibit significant positive trends. However, some shorter trends are also statistically significant, for example those ending in the mid-2020s or early 2040s. In summary, there are some indications of an increase in the number of HWTs for recent and future decades - despite the high temporal variability and standard deviation. These changes will be further investigated by

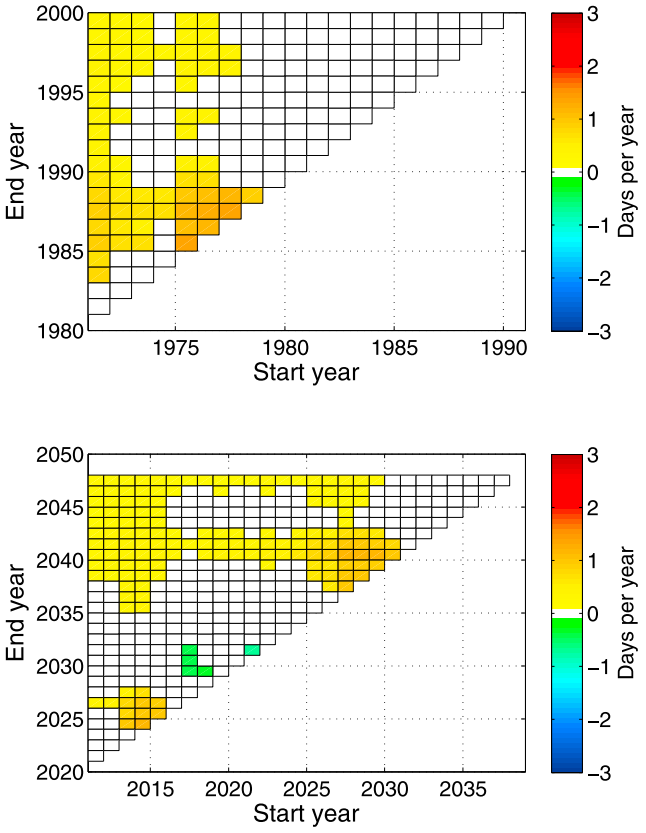

Figure 5. Trend matrices and significance of the ensemble mean for (top) C20 and (bottom) the projection period. The x-axis represents the start and the y-axis the end of the sub-time series. The color shading indicates the trend (number of HWT p.a.) estimated for each sub-time series. Changes with a statistical significance of less than 90% are plotted as white boxes.

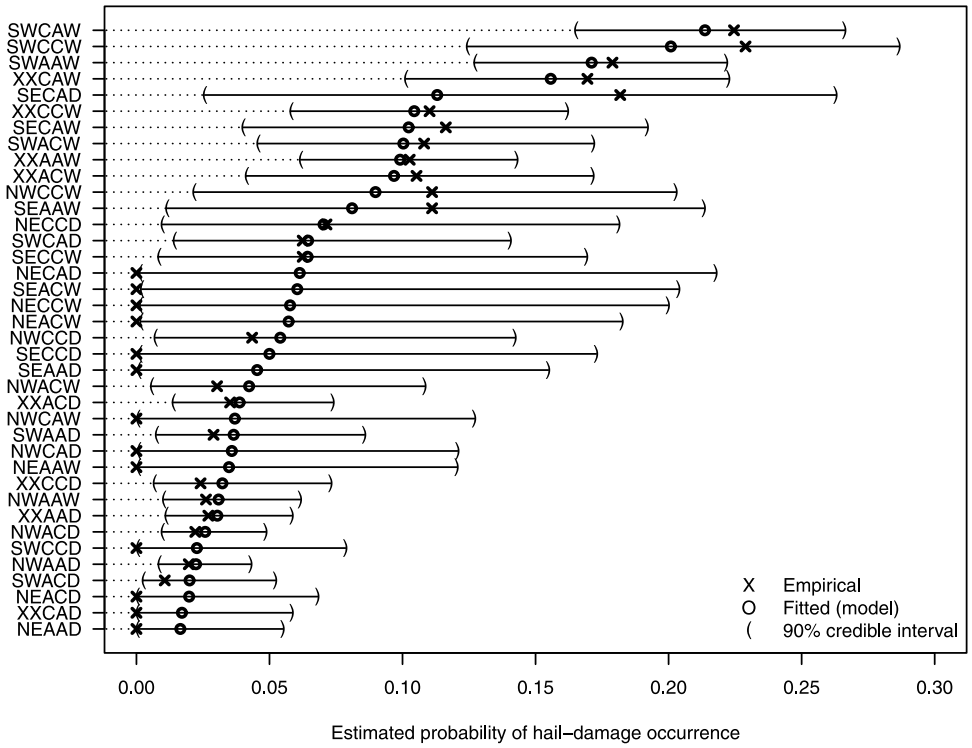

Figure 6. Point estimates (dots) and 95% credible intervals for the probabilities p_i of having hail during days with the *i-th* WT. The estimates are computed from WTs derived from CCLM-ERA40 and hail damage days of the SV. Crosses indicate empirical probabilities derived from the same data sets.

applying the statistical model to the RCM results in the next section.

5. Probabilistic Forecast of Hail Damage Days

5.1. Goodness of Fit of the Model for Probabilistic Prediction

[56] As explained in section 3.3, the main output of our statistical model (7)–(9) is a collection of samples p_i^k, $k = 1, \ldots, K = 3000$ of the posterior probability distributions of the frequencies of hail occurrence for each WT_i (see equation 10). For each WT_i, we obtain a point estimate for the hail occurrence probability p_i by taking the posterior mean:

$$\hat{p}_i = \frac{1}{K}\sum_{k=1}^{K} p_i^k. \tag{11}$$

Also, we obtain a quantification of the uncertainty in the point estimate of p_i by taking the interval bounded by the 2.5% and 97.5% quantiles of the empirical (posterior) distribution $\{p_i^k,\ k = 1, \ldots, K\}$. In the Bayesian terminology, such interval is called a 95% credible interval for the parameter p_i [see *Gelman et al.*, 2004], because the probability of the parameter value being in this range is 0.95.

[57] Figure 6 illustrates the fit of our model to the WTs derived from CCLM-ERA40 and SV loss data during C20. The crosses indicate the fractions f_i of hail damage days according to the SV data during each WT, as defined in equation (3), further referred to as empirical values. The circles refer to the point estimates $\hat{p}_i$ given by the posterior mean defined in equation (11) and the corresponding 95% credible intervals are denoted by solid lines bounded by brackets. The figure highlights a number of features of our model. First of all, four weather types SWCAW, SWCCW, SWAAW and XXCAW are most likely associated with damaging hailstorms, yielding an estimated probability $\hat{p}_i$ between 15% and 25%. For these patterns, the point estimates provided by the posterior means (circles) are close to the empirical values f_i (crosses). There is a fifth pattern SECAD with a large empirical value for f_i. However, the statistical model produces a much larger credible interval for the corresponding probability p_i and the point estimate provided by the posterior mean $\hat{p}_i$ is sensibly smaller than the empirical value f_i. This is due to the rare occurrence of the SECAD pattern, which implies high statistical uncertainty. Note that these five patterns were also identified in section 4.1 as the most hail-related. However, we clearly see the advantage of working with a statistical model, rather than with a simpler empirical formula as in equation (4). The statistical model provides an explicit quantification of the uncertainty in the estimate $\hat{p}_i$, which would be harder to obtain for (4).

[58] Qualitatively, the model identifies almost the same WTs as hail-related as we already found in the relative fraction of WTs on hail damage days (Figure 2, right). For large values of $\hat{p}_i$, the average probability predicted by the model (posterior mean) is slightly smaller than the empirical values. Similarly the average predicted probability is slightly larger than zero for those WTs without hail in the record,

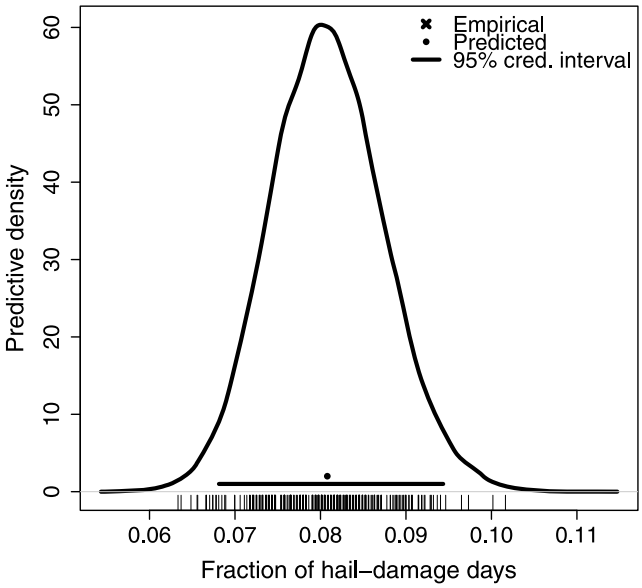

Figure 7. Predictive distribution for the fraction of hail damage days derived from CCLM-ERA40 during the observed record of the SV loss data (1986–2000). The 95% prediction interval is shown as black, bold line and the point estimate and the empirical value are shown as black dot and cross, respectively.

where the empirical values are zero. A related feature is that the estimated $\hat{p}_i$'s vary smoothly for patterns below SECAD in Figure 6. This is due to the structure of our model (7)–(9), where the probabilities p_i come from one and the same Beta distribution, equation (8). This induces a tendency of the point estimates (posterior means) to exhibit a reduced variability with respect to the empirical frequencies. For this reason, the predicted probabilities of hail are nonzero for several of those patterns for which the empirical frequencies are zero. For a few patterns such as XXCAD and NEAAD, both the predicted probabilities and the prediction intervals are rather small, concentrated near zero. For other patterns, such as NECAD and SEACW, the credible intervals are very large, indicating substantial uncertainty in the possible value of the corresponding probability of hail occurrence.

[59] In summary, our model is slightly conservative, in that the predicted probabilities do not blindly follow the empirical frequencies of hail occurrence. This may be viewed as a shortcoming in terms of model fit however this is in fact a strong feature of the model. The fact that non-zero probabilities are assigned to those weather types with a zero empirical frequency of occurrence is a desirable feature and this is owed to the ability of the model to pool information across WTs through the common model for the p_i. In other words, the model does not rule out the possibility that hail occurs also with such weather types, although it assigns a small probability to such occurrences. These discrepancies between the empirical and probability values, however, are only marginal: not only are all the empirical values contained within the credible intervals, but, as we shall see, the differences do not affect the overall goodness of fit of the model.

[60] To further evaluate the goodness of fit of our model, we now compare a probabilistic prediction for the fraction of hail days obtained by our model with the empirical value obtained from the SV loss data. Consider the CCLM-ERA40 reanalysis. For each WT_i, we randomly sample one value p_i^k from the predictive sample (10). We then sample $B = 200$ integers x_i^j, $j = 1, \ldots, 200$ from the binomial distribution $Bin(N_i, p_i)$. These integers represent possible values for the number X_i of hail damage days, compatible with the aleatory uncertainty which we have assumed in our model for the hail-generating process (7)–(9). This process is then repeated $P = 200$ times, each time sampling one of the p_i^k (in practice, we sample with repetition an index k from 1 to $K = 3000$). This step accounts for our epistemic uncertainty in the precise value of p_i.

[61] The whole sampling process of P values for p_i^k and of the B integers is then repeated for each WT_i. This produces sequences of integers

$$\left(x_1^j, x_2^j, \ldots, x_{40}^j\right), \quad \text{for } j = 1, 2, \ldots, P \cdot B, \tag{12}$$

which represent realizations of the number of hail damage days for each WT in the C20 period, conditional on the observed sequence of weather types. This collection of realizations represents a set of possible "hail damage histories", alternative to the actual historical sequence, which incorporate both the epistemic uncertainty and the aleatory variability of our model for the hail generation process. Each synthetic hail history gives a total number of hail days. So the set

$$x^j = \sum_{i=1}^{40} x_i^j, \quad \text{for } j = 1, 2, \ldots, P \cdot B, \tag{13}$$

forms a sample from the predictive distribution of hail damage days. Dividing this by the total number N of days in the C20 period yields a sample

$$f^j = \frac{x^j}{N}, \quad \text{for } j = 1, 2, \ldots, P \cdot B, \tag{14}$$

of the predictive distribution for the fraction f of hail damage days during the observed record, see equation (3). The mean of this sample gives a pointwise estimate of $\hat{f} = 8.2\%$, corresponding to approximately 15 days per year (see Figure 7). This estimate is very close to the empirical (observed) value of 14.8. To obtain a measure of uncertainty in our prediction, we take the 2.5% and 97.5% quantiles of the predictive sample, yielding $(8.2 \pm 1.3)\%$. This is called a *prediction interval* (as opposed to the credible intervals, as those displayed for the p_i in Figure 6), because it incorporates both the epistemic uncertainty in the parameters (the p_i's) and the aleatory variability of the underlying binomial distribution [see *Gelman et al.*, 2004]. Several additional prediction tests have been carried out with the ERA-40 and ERA-Interim reanalysis as input and they all gave similar results (not shown). This demonstrates the applicability of our statistical model to different data sets.

5.2. Prediction of Hail Damage Days

5.2.1. Validation of Climate Model Realizations

[62] The statistical model is now applied to investigate whether the RCM simulations are capable to reproduce the statistics of all WTs in relation to hail damage days during

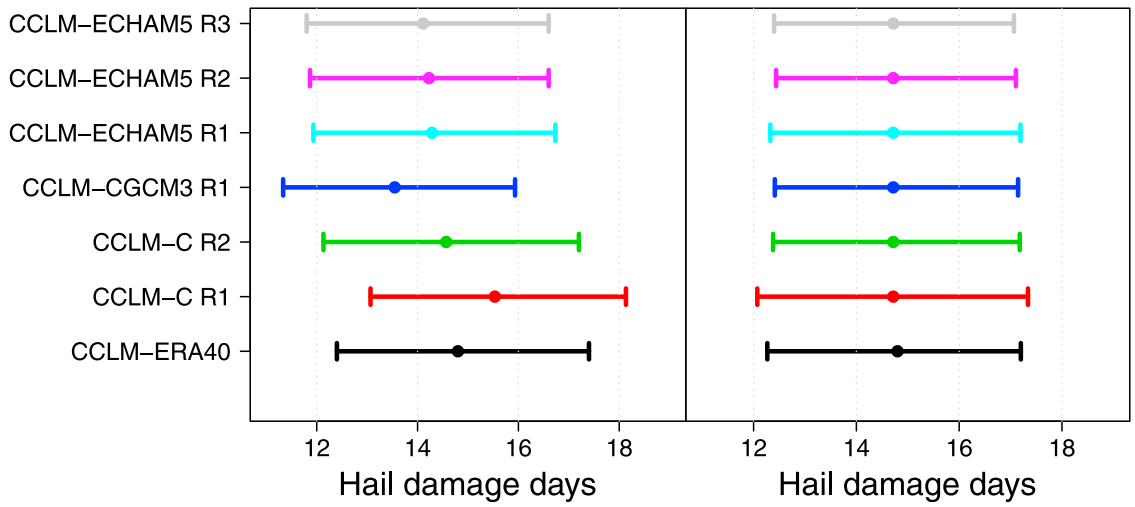

Figure 8. Predictive distribution of the number of hail damage days derived from CCLM-ERA40 (black), CCLM-C Run1 (red) and Run2 (green), CCLM-CGCM3 Run1 (blue), CCLM-ECHAM5 Run1 (light blue) and Run2 (purple) and Run3 (gray) for the time period 1986 to 2000. The right panel shows the probability distributions after a bias-correction.

the C20 period. To do so, for each RCM run we repeat the simulation process as described in section 5.1. For the probabilities p_i, we again use those of the model estimated from the CCLM-ERA40 reanalysis, that is we reuse the predictive sample (10). However, for the number of days with WT_i we use those obtained for the RCM simulation at hand, rather than those occurring in the CCLM-ERA40 reanalysis. That is, we sample from the binomial distribution $Bin(N_i^{model}, p_i^k)$, where N_i^{model} is the number of days with WT_i in the RCM run. This amounts to assume that the probabilistic relation between WT_i and hail occurrence is the same in the CCLM-ERA40 reanalysis and in all RCM runs.

[63] The predictive distributions for the annual numbers of hail damage days are shown in Figure 8 (left), summarized by point estimates (posterior means) and 95% prediction intervals. Predictions for the annual numbers of hail damage days are easily obtained by multiplying the predicted frequencies of hail days by the number of days (183) in a season. Predictive samples for the frequencies are again obtained as for Figure 6 (see section 5.1). In general, all RCMs give very similar predictions both in terms of the mean and of the uncertainty (prediction interval). Considering the CCLM-ERA40 results as a reference, largest deviations are apparent for CCLM-CGCM3, for which the posterior mean is 1.4 days less than the reference. Except for CCLM-C R1, the point estimates and prediction intervals show a shift toward lower values for all other RCMs. The variation, however, is very small, ranging between 0.2 and 0.8 days per year. Only the prediction derived from CCLM-C R1 has posterior mean number of hail damage days larger than the reference (by ≈0.7 days).

[64] For ease of visualization and comparison, we account for the deviations of the different RCM simulations from the reference simulation CCLM-ERA40 by a simple linear bias correction. For each RCM simulation, the whole predictive distribution is shifted by an amount equal to the difference between the mean predictions of the RCM and of the reference, where the difference is calculated during the C20 period only (Figure 8). This bias correction is then applied to each RCM over the whole period, including C20 and future projections. In doing so, it is assumed that the deviations are of systematic nature and caused by different model simulation characteristics or initial times. The bias correction allows us to quantitatively assess future changes in the probability of hail damage days from the RCM projections. Indeed, the linear shift is applied uniformly to the whole period: therefore, differences between the predictive distributions for two different time slices of the same RCM simulation are not affected by the bias correction.

[65] As shown in Figure 8 (right), the bias-corrected prediction intervals from all RCMs are highly concordant with each other. Only that derived from CCLM-C R1 exhibit a slightly larger uncertainty. These differences, however, are relatively small and do not allow us to draw any conclusions concerning different RCM simulation characteristics.

5.2.2. Prediction of Hail Damage Days in the Future

[66] The probabilistic model, together with the bias correction, was then applied to predict the distribution of hail damage days in the future, based on the sequences of WTs projected by the RCMs runs. Predictive distributions of hail damage days per year were computed for the entire future projection period (2001–2050), as well as for three disjoint 15-year time slices: 2001–2015, 2016–2030, and 2031–2045. We recall that CCLM-ECHAM5 projections are available only after 2011 and, therefore, these are omitted in the first time slice 2001–2015. Furthermore, the CCLM-C A1B scenario is available and considered in the analyses only until 2048. However, as the probability distributions are calculated from time series that comprise at least 40 years, they are assumed to be comparable even though the times differ.

[67] Most RCM simulations show an increase in the predicted number of hail damage days for the entire projection period 2001–2050 (see Figure 9, top left). However, the magnitude of this increase is not large: for most RCM runs the difference in the mean predicted value for the future and for the C20 period lies between one and two days. On the other hand,

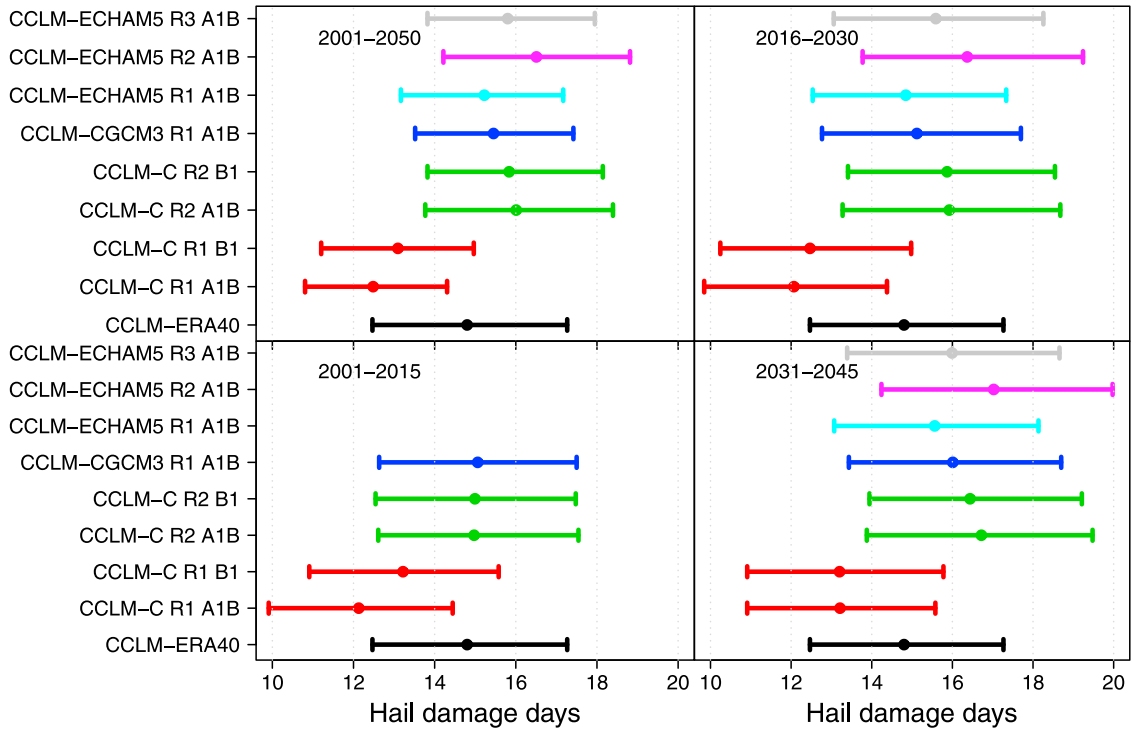

Figure 9. Same as Figure 8 but for the future projection period (top left) 2001 to 2050, (bottom left) 2001 to 2015, (top right) 2016 to 2030 and (bottom right) 2031–2045. Note that for CCLM-ECHAM5 only the period between 2011 and 2050 and for CCLM-C A1B only 2001 to 2048 is available and, hence, the time periods are adjusted, respectively.

the prediction derived from CCLM-C R1 shows a decrease. Reasons for this will be discussed later on in this section. Neglecting the CCLM-C R1 runs, the current-to-future differences in hail damage days is larger than the variability amongst the different models for all periods shown. For example, the difference between the entire projection period 2001–2050 and C20 is +1.0 ± 0.47 days (mean ± standard deviation), whereas model variability is only 0.33 ± 0.3 days. As can be seen in all panels of Figure 9, the spread between the two differences increases with increasing lead time.

[68] To test whether the trends in the frequency distributions of hail damage days between past and future projections are significant, several statistical tests were performed. First, a two-sided Kolmogorov-Smirnov test (KS) [*Clarke and Cooke*, 2004] was used to test if the frequency distributions are themselves significant different. Using a bootstrap resampling method (1000 bootstraps) allows to account for discontinuity in the distributions (distribution of hail days is based on counts). Applying the test, we find that all distributions for the time period 2001–2050, derived from the different climate model simulations, are significantly different from the distribution sampled from CCLM-ERA40 derived HWTs, with a probability larger than 99.99%. This test gives information on the overall distributional difference but not specifically on the significance of the trends of hail damage days. We investigate the significance in the trends by considering the posterior distribution of the difference in means between past and future, that is $\Delta f = f_{past} - f_{future}$. The posterior of Δf is obtained by repeatedly sampling (5000 times) from the posterior distribution of past and future hail days, and then calculating the difference in the means (f_{past} and f_{future}) of each sample. This results in 5000 samples from the posterior of Δf. The 95% prediction interval for Δf, constructed taking the 2.5% and 97.5% quantiles of the posterior sample, does not contain zero indicating it is not a value likely to occur. This implies that the means of the sampled distributions for C20 (1986–2000) and future projections (2001–2050) are not equal. Note that this Bayesian way of 'testing' does not make any distributional assumptions unlike, for example, the conventional *t*-test. We find that even though the magnitudes of the trends are small, all of the trends between C20 and the entire projection period are statistically significant on a 95% significance level (Table 2).

[69] To examine the temporal variability in the predictions, we now consider the three separate slices of the future projection period. Note that the length of each slice is 15 years, which is the same as the record length of C20. No clear-cut signal is detectable in the first time slice (2001 to 2015) across the RCM estimates (Figure 9, bottom left). The predictions derived from CCLM-CGCM3 and CCLM-C R2 (A1B and B1) only show a small significant increase and almost coincide with the reference (CCLM-ERA40; black line and dot). Only CCLM-C R1 shows a decrease in the number of hail days during this and all other time slices; reasons will be discussed below. For the next time slice (2016 to 2030), the signal becomes clearer (Figure 9, top right). For all RCM projections, the predictions are significantly shifted toward higher numbers of hail days (Table 2). Most interestingly, in the last time slice (2031 to 2045) all simulations project an increase of one or two in the number

Table 2. Mean Number of Hail Damage Days (hdd) According to the Predictive Distributions for Each Model Simulation and Time Slice[a]

Model Simulation		Time Period	Mean of hdd	Bayesian Statistics m [2.5;97.5]	Time Period	Mean of hdd	Bayesian Statistics m [2.5;97.5]
CCLM-ERA40		1989–2000	14.79	-	-	-	-
CCLM-C	R1 A1B	2001–2050	12.59	2.20 [2.17;2.22]	2001–2015	12.57	2.21 [2.19;2.24]
	R1 B1	2001–2050	13.06	1.72 [1.70;1.75]	2001–2015	13.12	1.67 [1.64;1.70]
	R2 A1B	2001–2050	16.00	−1.21 [−1.24;−1.18]	2001–2015	16.09	−1.30 [−1.33;−1.27]
	R2 B1	2001–2050	15.79	−0.98 [−1.04;−1.01]	2001–2015	15.79	−1.01 [−1.03;−0.98]
CCLM-CGCM3	R1 A1B	2001–2050	15.42	−0.64 [−0.67;−0.61]	2001–2015	15.45	−0.66 [−0.69;−0.63]
CCLM-ECHAM5	R1 A1B	2001–2050	15.16	−0.38 [−0.40;−0.35]	2001–2015	-	-
	R2 A1B	2001–2050	16.54	−1.75 [−1.78;−1.73]	2001–2015	-	-
	R3 A1B	2001–2050	15.70	−0.92 [−0.94;−0.89]	2001–2015	-	-
CCLM-C	R1 A1B	2016–2030	12.86	1.93 [1.90;1.96]	2031–2045	13.41	1.37 [1.34;1.40]
	R1 B1	2016–2030	13.04	1.75 [1.72;1.78]	2031–2045	13.56	1.23 [1.20;1.26]
	R2 A1B	2016–2030	16.44	−1.66 [−1.69;−1.63]	2031–2045	16.87	−2.08 [−2.11;−2.05]
	R2 B1	2016–2030	16.17	−1.38 [−1.41;−1.35]	2031–2045	16.44	−1.65 [−1.68;−1.63]
CCLM-CGCM3	R1 A1B	2016–2030	15.54	−0.75 [−0.78;−0.73]	2031–2045	15.81	−1.02 [−1.05;−0.99]
					2031–2045		
CCLM-ECHAM5	R1 A1B	2016–2030	15.52	−0.73 [−0.76;−0.70]	2031–2045	15.76	−0.97 [−1.00;−0.95]
	R2 AB1	2016–2030	16.76	−1.97 [−2.00;−1.94]	2031–2045	17.08	−2.29 [−2.32;−2.27]
	R3 A1B	2016–2030	15.87	−1.08 [−1.11;−1.06]	2031–2045	16.11	−1.33 [−1.36;−1.30]

[a]Further, the test statistics of the Bayesian trend test are given in form of the mean (m) of the distribution Δf and the 2.5% and 97.5% percentiles (for further details, see section 5.2.2). For the trend test each time slice is compared to the preceding one.

of hail days (again except of CCLM-C R1). Compared to the previous two time slices, the mean values display a slightly smaller scatter but the prediction intervals are generally wider, indicating a higher predictive uncertainty.

[70] As a general consideration, however, the predictions obtained from simulation CCLM-C R1 differ from all others. This applies to the emission scenarios A1B and B1 and to all time slices - even for the first slice, where the other simulations only display small variations. This raises the question, why the CCLM-C R1 simulation displays such an abrupt decrease and whether this could be caused by the RCM/AOGCM characteristics, the OWTC routine, or the statistical model.

[71] In the time series of hail-unrelated WTs (NWT), a substantial decrease of more than 50% is found in the late 1990s for CCLM-C R1, but not for CCLM-ECHAM5 (Figure 10). By analyzing the distribution of the parameters considered in the OWTC scheme, we found a strong positive

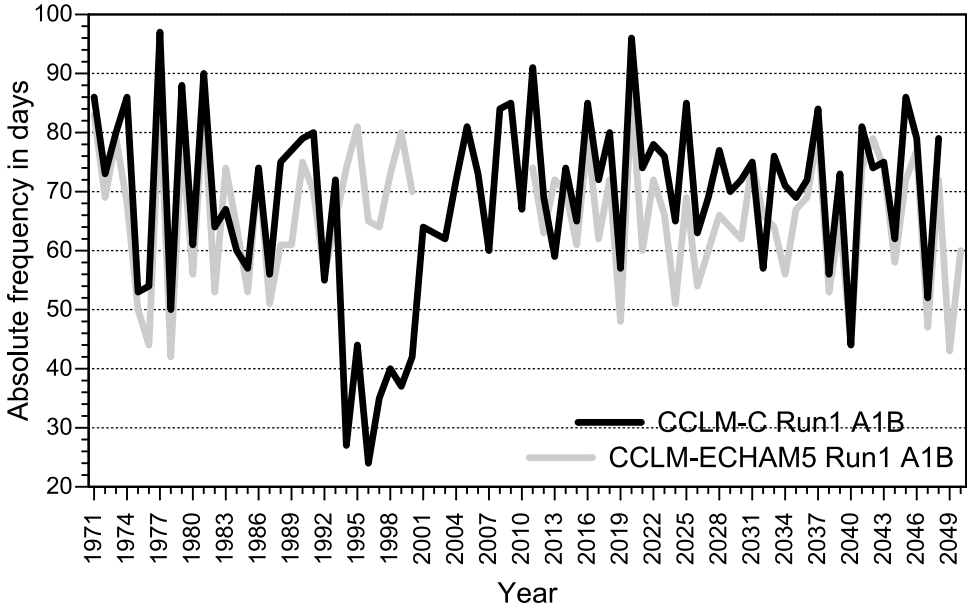

Figure 10. Absolute frequency of NWT (SWACD, NWAAW, NWAAD, NWACD, XXAAD) per summer half-year according to CCLM-C Run1 A1B (black) and CCLM-ECHAM5 Run1 A1B (gray) from 1971 to 2048.

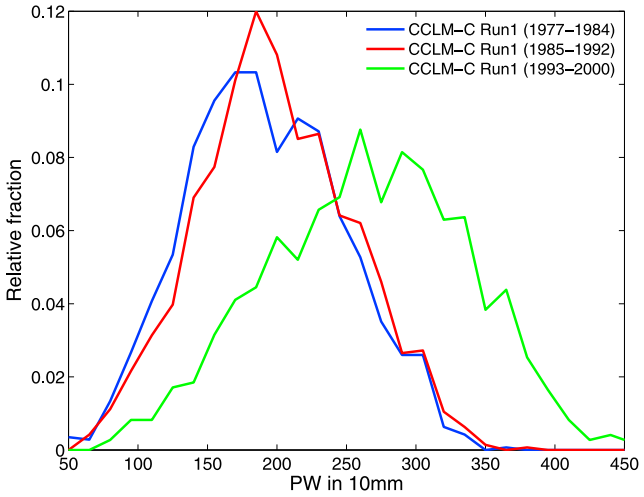

Figure 11. Distribution of PW for three sequenced time periods of eight years from CCLM-C Run 1. Blue: 1977–1984; Red: 1985–1992; Green: 1993–2000.

anomaly of *PW* between 1993 and 2000 (mean: 26.6 mm; Figure 11) that coincides with the decrease in the time series of NWT. No other model simulation shows such high water vapor amounts during that period. By contrast, the distributions of *PW* during the other two periods, from 1977 to 1984 (mean: 18.8 mm) and from 1985 to 1992, are almost similar and in agreement to those of the other simulations. The apparent overestimation of *PW* leads to an overestimation in the number of hail damage days during the whole C20 period (cf. Figure 8). The applied bias correction, based on the estimates for C20, which include the large anomaly in *PW*, shifts the predictive distributions of the number of hail days toward lower values for all time slices. Consequently, the number of hail days in future projections is significantly underestimated. For this reason, the results of the two CCLM-C R1 runs should be interpreted with care. The reasons behind the abnormal high amounts of *PW* are not clear as the model physics and the forcing AOGCM are the same for all three periods.

[72] The fact that two of eight model simulations show a significant decrease in the number of hail damage days illustrates the uncertainty that is connected to the chosen ensemble members. If these two model simulation are regarded in the ensemble, a positive trend of hail damage days in the order of 6% can be retrieved when comparing the C20 period with the last time slice (2030–2045). In comparison, disregarding them in the ensemble leads to a positive trend of almost 11%. These different results indicate the importance of a large ensemble, as model/simulation specific randomness carries less weight within the ensemble and thus, the results.

6. Summary and Conclusions

[73] The present study investigates whether there is evidence suggesting changes in the frequency of hailstorms for past (1971–2000) and future (2011–2050) time periods. Since hail is neither captured comprehensively by existing observation systems nor can it be modeled reliably even by high-resolution RCMs, large-scale weather types determined by the objective weather type classification scheme from DWD were used here as proxy data for hail occurrence. We presented a new statistical model that predicts the probability for hail days by including pooled information from all WTs and not only those most likely associated with hailstorms. The model is Bayesian and uses a Markov Chain Monte Carlo approach. The examinations are based on ERA-40 reanalyses and several RCM runs, which consider different driving global models and their realizations in terms of different initial conditions, different RCM model versions, and different emission scenarios. In our study, we used model data from the first of two nesting steps with a spatial resolution of $0.44° \times 0.44°$, which is sufficient for the WT classification scheme. To get information about hail occurrence, loss data from an insurance company were used in addition.

[74] Based on ERA-40 reanalyses, we identified specific WTs that are most likely associated with damaging hailstorms and evaluated to what extent the considered RCMs are able to reliably reproduce the distribution of all WTs within a 15-yr control period (1986–2000). Long-term variability of specific WTs that are frequently (or never/very rarely) associated with damaging hail is investigated for past (1971–2000) and future (2011–2050) time periods using both conventional linear regression and the statistical model.

[75] From the results obtained in this study, the following major issues and conclusions can be inferred:

[76] 1. A relation can be established between days with damaging hail and specific WTs. This means that even if hail is a local-scale phenomena, the atmospheric preconditions for hailstorm occurrence are large-scale. The identified WTs of SWCAW, SWCCW and XXCAW are associated with an eastward moving upper air ridge followed by a low pressure system from the west. Large-scale upward motion can reduce convective inhibition leading to an increase in the probability of deep convection to be triggered by meso-scale processes. This configuration is not the driver for the SWAAW and SECAD patterns, which, however, prevailed only on a few days.

[77] 2. The RCMs considered are able to reliably reproduce the distribution of WTs determined from ERA-40 reanalysis data, in view of the fact that the RCMs do not know the real state of the atmosphere when driven by an AOGCM. The good agreement also documents the robustness of the OWTC scheme in general and its applicability to derive proxy data for investigation such as hail-related patterns.

[78] 3. According to the classical trend analysis the number of HWTs slightly increased from 1979 to 2000, considering the ensemble mean of all RCM simulations. The results of the statistical model suggest similar trends for the future projection period. An average increase in the number of hail damage days of 5% between C20 and 2030–2045 is found for the ensemble. However, the trends between the different model simulations vary substantially. While two of the ensemble members show a negative trend of hail damage days in the order of −10 to −11%, all other members indicate positive trends between 6–13%. Largest discrepancies between the RCMs results are due to different realizations of the driving global model (ECHAM-5), whereas the emission scenario is found to be of minor importance. This is in agreement to the findings of *Hawkins and Sutton* [2009],

who showed that the overall uncertainty in RCM projections for the next decades is dominated by model uncertainty in the AOGCM–RCM chain. Uncertainty due to the emission scenario, on the other hand, is of minor importance and increases only slowly with lead time.

[79] 4. In general, the results obtained with the standard linear regression model and the Bayesian model are somewhat similar. However, the Bayesian model allows for direct estimation of hail days including uncertainty ranges, which is not possible by the linear regression method. By incorporating valuable information possibly inherent in each WT, the results obtained from the Bayesian model provide a much more comprehensive quantification of uncertainty than those obtained by the trend analysis. Using a Bayesian model to predict hail days in a probabilistic way is a novel and promising approach that will be considered also for further research.

[80] The differences between the climate models, in particular the trends in the future as shown in the last section, confirm the need to consider not only one climate model, but an ensemble of models considering different driving global models and different RCMs. Of course, an ensemble comprising only eight members is far from being ideal [*Hagedorn et al.*, 2005; *Harris et al.*, 2010]. However, the input of meteorological fields on different levels required by the OWTC scheme strongly limits the availability of appropriate data sets. Furthermore, we are aware that using claim data of an building insurance company leads to an underestimation of the actual number of hail events, since hailstones have to reach a diameter of more than 2.5 cm to cause damage to buildings. However, to our best knowledge, no data set exists at the moment that provides information about hail occurrence over such a large area, with such a high coverage and over such a long time period (from 1986).

[81] To estimate the number of hail-damage days from the RCMs by applying the statistical model, we assumed that the probabilistic relation between hail occurrence and the WT_is is the same as for CCLM-ERA40. As the evaluation of hail-related WTs derived from reanalyses and climate simulations showed a good agreement, the deviations of the probability distributions are assumed to be systematic and a bias correction toward the CCLM-ERA40 results is justified and thus, applied to all RCM data. For the future projections, we observe an increase in the minimum number of hail-damage days of up to 2 days (2030–2045) in most of the climate simulations. Only the results from CKLCR1 showed a decrease of hail-damage days, which can be ascribed to unrealistic precipitable water patterns in the 1990s.

[82] The investigations demonstrated that statistical modeling of hail days is a promising approach and that the model is applicable to various data sets. The differences between the empirical values estimated from the insurance loss data and the results of the statistical model showed that the model is slightly conservative, in that the predicted probabilities do not blindly follow the empirical frequencies of hail occurrence. The mean prediction of hail damage days is slightly larger than the empirical value (mean of the historical observations) for those WTs that never occurred, while it is slightly smaller for the WTs with highest observed frequency of hail occurrence. In all cases, however, the empirical value is contained within the credible interval, thus indicating that the statistical model is consistent with the empirical value.

[83] The fact that non-zero probabilities are assigned to those WTs with a zero empirical frequency of occurrence, however, is a desirable feature and owed to the ability of the model to pool information across all WTs through the common model for the p_i. So the model does not rule out the possibility that hail occurs also with such weather types, although it assigns a small probability to such occurrences. These discrepancies between the empirical and probability values, however, are only marginal. Further, the modeled number of hail-damage days and the empirical values (*SV Sparkassenversicherung*) are in good agreement and indicate that the suggested statistical model is sufficient for our purpose.

[84] A potential weakness of our approach is the underlying assumption that the different WTs are the sole predictor for hail damage days. This assumption neglects possible changes in atmospheric parameters that are relevant for the occurrence of severe thunderstorms, such as conditional or latent instability, vertical wind shear, or low-level moisture. Even though these parameters are included implicitly in the parameters defining the specific WTs (e.g., PW is considered, but not only at low levels), they are not addressed explicitly in the categorization approach of the OWTC. A few recent studies estimated changes in the potential for convection initiation and severity based upon large-scale meteorological parameters well resolved by RCM simulations [*Trapp et al.*, 2011]. They found evidence for a small increase in favorable severe thunderstorm environments caused by an increase in CAPE and high deep layer shear for several regions in the United States [*Van Klooster and Roebber*, 2009] and Europe [*Marsh et al.*, 2009]. In contrast to this, *Mahoney et al.* [2012] showed in a most recent study that hail may be nearly eliminated in the future over Colorado's mountains due to an increase in the height of the environmental melting level. The contrasting results confirm the need to further scrutinize the regional effects of climate change on the probability of hazardous convective weather. In that sense, we are aware that our study highlights only one part of the complex interaction of various processes leading to hail on the ground.

[85] In the future, we intend to consider other meteorological parameters in a similar dichotomous scheme as the OWTC to further improve the relation between WTs and hail damage. First calculations using Lifted Index and vorticity advection showed promising results. Furthermore, we intend to include additional RCMs to consider a larger variety of possible future scenarios.

Appendix A: Summary of the Markov Chain Monte Carlo Estimation Runs

[86] Two MCMC chains were ran with different initial values, each for 25000 iterations. Of those, 10000 were used as burn-in and the remaining 15000 were thinned by 10 to reduce auto-correlation in the samples. This resulted in 1500 samples from each chains and convergence was assessed using the Gelman and Rubin $\hat{R}$ values for each parameter (for details, see *Gelman et al.* [2004, chap. 11]). For convergence, all $\hat{R}$ values need to be approximately close to 1 and

here the highest value for $\hat{R}$ was 1.12 indicating convergence of all parameters (p_i, α and β).

[87] **Acknowledgments.** The authors thank the German Weather Service for providing the OWTC software, the SV SpakassenVersicherung AG for the provision of detailed loss data, and the team of the CERA data base for the provision of CCLM data. We thank Hendrik Feldmann, Peter Berg, and Hans-Jürgen Panitz from IMK-TRO for technical support and helpful hints. The Willis Research Network (WRN), London, is acknowledged for networking and sponsorship. A research visit of the first author at the University of Exeter, UK, was funded by the Deutscher Akademischer Austauschdienst (DAAD).

References

Bissolli, P., and E. Dittmann (2001), The objective weather types classification of the German Weather Service and its possibilities of application to environmental and meteorological investigations, *Meteorol. Z.*, *10*, 253–260.

Bissolli, P., J. Grieser, N. Dotzek, and M. Welsch (2007), Tornadoes in Germany 1950–2003 and their relation to particular weather conditions, *Global Planet. Change*, *57*, 124–138.

Brooks, H. E., J. W. Lee, and J. Craven (2003), The spatial distribution of severe thunderstorm and tornado environments from global reanalysis data, *Atmos. Res.*, *67*, 73–94.

Cao, Z. (2008), Severe hail frequency over Ontario, Canada: Recent trend and variability, *Geophys. Res. Lett.*, *35*, L14803, doi:10.1029/2008GL034888.

Changnon, S. A. (1970), Hailstreaks, *J. Atmos. Sci.*, *27*, 109–125.

Clarke, G. M., and D. Cooke (2004), *A Basic Course in Statistics*, 492 pp., John Wiley, London.

DeRubertis, D. (2006), Recent trends in four common stability indices derived from US radiosonde observations, *J. Clim.*, *19*, 309–323.

Dessens, J. (1986), Hail in southwestern France: Hailfall characteristics and hailstorm environment, *J. Clim. Appl. Meteorol.*, *25*, 35–47.

Doswell III, C. A. (1987), The distinction between large-scale and mesoscale contributions to severe convection: A case study example, *Weather Forecast.*, *2*, 3–16.

Egli, T. (2007), Elementarschutzregister Hagel, technical report, 35 pp., Kantonale Gebäudeversicherungen, Bern.

Ehmann, C. (2009), Analyse der vorherrschenden Großwetterlagen während Hagelereignissen in Baden-Württemberg, seminar thesis, 72 pp. Inst. for Meteorol. and Clim. Res., Karlsruhe Inst. for Technol., Karlsruhe, Germany.

Emanuel, K. (1994), *Atmospheric Convection*, 580 pp., Oxford Univ. Press, Oxford, U. K.

Feser, F., B. Rockel, H. von Storch, J. Winterfeldt, and M. Zahn (2011), Regional climate models add value to global model data–A review and selected examples, *Bull. Am. Meteorol. Soc.*, *92*, 1181–1192.

Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin (2004), *Bayesian Data Analysis*, 689 pp., Chapman and Hall, Boca Raton, Fla.

Hagedorn, R., F. J. Doblas-Reyes, and T. N. Palmer (2005), The rationale behind the success of multi-model ensembles in seasonal forecasting–I. Basic concept, *Tellus, Ser. A*, *57*, 219–233.

Haklander, A. J., and A. van Delden (2003), Thunderstorm predictors and their forecast skill for the Netherlands, *Atmos. Res.*, *67–68*, 273–299.

Hanafin, J., R. McGrath, T. Semmler, S. Wang, P. Lynch, S. Steele-Dunne, and P. Nolan (2011), Air flow and stability indices in GCM future and control runs, *Int. J. Climatol.*, *31*, 1240–1247.

Harris, G. R., M. Collins, D. M. H. Sexton, J. M. Murphy, and B. B. B. Booth (2010), Probabilistic projections for 21st century European climate, *Nat. Hazards Earth Syst. Sci.*, *10*, 2009–2020.

Hawkins, E. and R. Sutton (2009), The potential to narrow uncertainty in regional climate predictions, *Bull. Am. Meteor. Soc.*, *90*, 1095–1107.

Hess, P., and H. Brezowsky (1977) Katalog der Grosswetterlagen Europas, *Tech. Rep. 113*, 14 + 54 pp., Ber. Dtsch. Wetterdienst., Potsdam, Germany.

IPCC (2007), *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, edited by S. Solomon et al., 996 pp., Cambridge Univ. Press, Cambridge, U. K.

Kendall, M. (1975), *Rank Correlation Methods*, 260 pp., Griffin, London.

Kunz, M. (2007), The skill of convective parameters and indices to predict isolated and severe thunderstorms, *Nat. Hazards Earth Syst. Sci.*, *7*, 327–342.

Kunz, M., and M. Puskeiler (2010), High-resolution assessment of the hail hazard over complex terrain from radar and insurance data, *Meteorol. Z.*, *19*, 427–439.

Kunz, M., J. Sander, and C. Kottmeier (2009), Recent trends of thunderstorm and hailstorm frequency and their relation to atmospheric characteristics in southwest Germany, *Int. J. Climatol.*, *29*, 2283–2297.

Lautenschlager, M., K. Keuler, C. Wunram, E. Keup-Thiel, M. Schubert, A. Will, B. Rockel, and U. Boehm (2009a), Climate Simulation with CLM. Climate of the 20th Century run no. 1. Data Stream 3: European region MPI-M/MaD, http://dx.doi.org/10.1594/WDCC/CLM_C20_1_D3, World Data Cent. for Clim., Hamburg, Germany.

Lautenschlager, M., K. Keuler, C. Wunram, E. Keup-Thiel, M. Schubert, A. Will, B. Rockel, and U. Boehm (2009b), Climate Simulation with CLM. Climate of the 20th Century run no. 2. Data Stream 3: European region MPI-M/MaD, http://dx.doi.org/10.1594/WDCC/CLM_C20_2_D3, World Data Cent. for Clim., Hamburg, Germany.

Leslie, L. M., M. Leplastrier, and B. W. Buckley (2007), Estimating future trends in severe hailstorms over the Sydney Basin: A climate modelling study, *Atmos. Res.*, *87*, 37–51.

Mahoney, K., M. A. Alexander, G. Thompson, J. J. Barsugil, and J. D. Scott (2012), Changes in hail and flood risk in high-resolution simulations over Colorado's mountains, *Nat. Clim. Change*, *2*, 125–131.

Mann, H. B. (1945), Nonparametric tests against trend, *Econometrica*, *13*, 245–259.

Marsh, P. T., H. E. Brooks, and D. J. Karoly (2009) Preliminary investigation into the severe thunderstorm environment of Europe simulated by the Community Climate System Model 3, *Atmos. Res.*, *93*, 607–618.

Noppel, H., U. Blahak, A. Seifert, and K. D. Beheng (2010), Simulations of a hailstorm and the impact of CCN using an advanced two-moment cloud microphysical scheme, *Atmos. Res.*, *96*, 286–301.

Rockel, B., A. Will, and A. Hense (2008), The Regional Climate Model COSMO-CLM (CCLM), *Meteorol. Z.*, *17*, 347–348.

Roeckner, E., et al. (2003), The atmospheric general circulation model ECHAM5. Part I: Model description. *MPI-Rep. 349*, Max-Planck-Inst. für Meteorol., Hamburg, Germany.

Roeckner, E., R. Brokopf, M. Esch, M. Giorgetta, S. Hagemann, L. Kornblueh, E. Manzini, U. Schlese, and U. Schulzweida (2006), Sensitivity of simulated climate to horizontal and vertical resolution in the ECHAM5 atmosphere model, *J. Clim.*, *19*, 3771–3791.

Sauvageot, H. (1992), *Radar Meteorology*, 366 pp., Artech House, Norwood, Mass.

Schiesser, H. (2003), *Extremereignisse und Klimaänderung*, chap. 2.6, pp. 65–68, OcCC, Bern.

Spiegelhalter, D. J., A. Thomas, N. G. Best, and D. Lunn (2003), WinBUGS Version 1.4 user manual, MRC Biostat. Unit, Inst. of Public Health, Cambridge, U. K. [Available at http://www.mrc-bsu.cam.ac.uk/bugs].

Trapp, R. J., E. D. Robinson, M. E. Baldwin, N. S. Diffenbaugh, and B. R. J. Schwedler (2011), Regional climate of hazardous convective weather through high-resolution dynamical downscaling, *Clim. Dyn.*, *37*, 677–688.

Uppala, S. M., et al. (2005), The ERA-40 re-analysis, *Q. J. R. Meteorol. Soc.*, *131*, 2961–3012.

Van Klooster, S. L., and P. J. Roebber (2009), Surface-based convective potential in the contiguous United States in a business-as-usual future climate, *J. Clim.*, *22*, 3317–3330.

Yarnal, B., A. C. Comrie, B. Frakes, and D. P. Brown (2001) Review: Developments and prospects in synoptic climatology, *Int. J. Climatol.*, *21*, 1923–1950.

Yue, S., P. Pilon, B. Phinney, and G. Cavadias (2002) The influence of autocorrelation on the ability to detect trend in hydrological series, *Hydrol. Process.*, *16*, 1807–1829.